A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 131, 1993

*Władysław Milo**

CHARACTERIZATIONS OF EFFICIENCY AND PRECISION. PART 1

1. INTRODUCTION

In the previous paper (see M i l o 1989) we tried to analyze the characterization of estimators efficiency and precision in the case of models with one explanatory variable and scalar parameter. In this paper the more complex situation of vector parameter is considered.

Our goal is to formulate some quantitative characterizations of non-asymptotic efficiency of statistics. They do not pretend to be final and immediately usable in planning of experiments for evaluation of efficiency and precision of studied statistics.

In § 2 we propose characterizations of non-asymptotic efficiency by using characteristics of dispersion of probability distribution.

In § 3 we show how these propositions should be exposed to further deep studies in order to evaluate the indicators invariance on such important operations as scaling and translation of random variables. Some preliminary results concern the commonly known statistics: least squares estimator and Hoerl-Kennard ridge estimator.

* Lecturer at the University of Łódź.

2. CHARACTERIZATION OF NON-ASYMPTOTIC EFFICIENCY

Let $g(\Theta)$ be a vector-valued function of parameter vector Θ, $\Theta \in R^k$. In particular $g(\Theta) = \Theta$. The function $g(\Theta)$ may be called the parameter space reduction function.

It is convenient to accept "aprg(Θ)" as a name of stochastic approximate for $g(\Theta)$. Further we consider only $aprg(\Theta) = g(T_n)$, where T_n is a statistic for Θ and the size of g is $k_g \times 1$. In particular $g(T_n) = T_n$. Particular and general problems of stochastic approximation can be found, among others, in the works of M. Wasan, J. Blum, V. Fabian, J. Koronacki, H. Kushner, M. Newelson, R. Chasminski, H. Robbins, L. Schmetterer, R. Zieliński, P. Neumann. The formal bases of approximation are given in works of f.e., A. Smoluk, N. Achiezer, G. Lorentz.

The possible characterizations use:

a) dispersion characteristics ($Dg(T_n)$, $MSEg(T_n)$);

b) concentration characteristics (concentration function for $g(T_n)$, scattering function for $g(T_n)$);

c) parametric, nonparametric and entropy information measures about $g(\Theta)$ provided by $g(T_n)$;

d) characteristics of statistical cost of obtaining values of (a)-(c) for given two functions $g(T_{1,\kappa_n})$, $g(T_{2,n})$;

e) characteristics of numerical cost of calculating $g(T_n)$.

In this paper we consider only the case (a).

Some comments about further notations and ideas may prove to be useful. We shall use the ratio form of efficiency measures. In some sense a numerator expresses an effect of using an input in the form of denumerator. More specifically let $g_1 \equiv g(T_{1,n})$ denote a function g defined on the statistic $T_{1,n}$ with the index 1. Then the dispersion matrix (or variance-covariance matrix) of g_1 is equal $D_{g_1} \equiv D_g(T_1)$ and the dispersion matrix of Y is equal DY. These two matrices are basic arguments of our efficiency and precision type indicators.

2.1. CHARACTERIZATIONS USING CHARACTERISTICS OF DISPERSION AND PRECISION

The general form of efficiency indicator is

$$Ed \equiv \frac{ed}{id} \tag{1}$$

where ed is the value of a scalar real function of dispersion matrix $Dg(T_{1,n})$ of $g(T_{1,n})$ and id is the value of scalar real function of dispersion matrix DY for $Y = (Y_1, \ldots, Y_n)$.

We will use the following scalar functions

$$trDY, detDY, \lambda_{max}(DY), \lambda_i(DY), i = 1, \ldots, n \quad (2)$$

$$trg(T_{1,n}), detg(T_{1,n}), \lambda_{max}(Dg(T_{1,n})) \quad (2.1)$$

$j = 1, \ldots, k_g$, $\lambda_j \equiv \lambda_j(A)$ is the j-th eigenvalue of A, k_g is the order of matrix Dg(), $g(\theta) = \theta$, $\theta \in R^k$.

The most often practice is to use the efficiency indicators based on det function and sometimes on the trace function. However, when we want to study numerical stability of statistic $\{T_{1,n}\}$, or $g\{T_{1,n}\}$, then the indicators based on λ_{max} or $\{\lambda_i\}$ are also attractive options.

Further we shall extensively use some well know facts from linear algebra on matrix diagonalization and simultaneous diagonalization of two non-negative commutative matrices. Due to non--negative definiteness of matrix A there exists an orthogonal matrix O_A: $O_A' A O_A = \Lambda_A$, where $A \equiv DY, D_g(\ )$. For commutative matrices DY, $D_g(.)$ there exists a matrix O_A that simultaneously diagonalizes both dispersion matrices. It is easy to see this in the case of linear model

$$LM \equiv (R^{nxk}, S, S_Y, Y = x\beta + U, P_Y = N(x\beta, o^2I)),$$

where: S, S_Y are the known probability spaces, R^{nxk} is the set of real n x k matrices, the probability distribution P_Y is gaussian with the mean $EY = X\beta$ and dispersion (variance-covariance) matrix $DY = o^2I$ and $x \in R^{nxk}$, $\beta \in R^k$. In the case of LM the least - squares estimator (l.s.e.) is $T_{1,n} = x^+Y \equiv g(T_{1,n})$.

In general, due to known properties of trace and determinant we have

$$trDY = \sum_{i=1}^{n} \lambda_i(DY), \quad det\ DY = \prod_{i=1}^{n} \lambda_i(DY) \quad (3)$$

$$tr\ Dg(T_{1,n}) = trDg_1 = \sum_{i=1}^{k_g} \lambda_i(Dg_1), \quad det\ Dg_1 = \prod_{i=1}^{k_g} \lambda_i(Dg_1) \quad (3.1)$$

Using (2)-(3.1) we formulate the following, non-relative w.r.t. other functions g_1, dispersional indicators of efficiency of $g_1 = g(T_{1,n})$:

$$Ed1gl \equiv \frac{tr\ Dg_1}{tr\ DY}, \quad Ed2gl \equiv \frac{det\ Dg_1}{det\ DY} \tag{4}$$

$$Ed3gl \equiv \frac{\lambda_{max}(Dg_1)}{\lambda_{max}(DY)} \quad Ed4gl \equiv w_1 Ed1gl \tag{4.1}$$

where:

$$w_1 = (1 + \nu_1)/(1 + \nu_o), \quad \nu \equiv \nu_{DY}, \nu_1 \equiv \nu_{Dg_1} \quad \text{and } \nu_A \equiv \frac{\lambda_{min}(A)}{\lambda_{max}(A)}$$

is the bad-conditioning index of matrix A,

$$Ed5gl \equiv \frac{(1 + \nu_1)^{k_g}}{(1 + \nu_o)^n} Ed2gl, \quad Ed6gl \equiv w_1 Ed3gl \tag{4.2}$$

It is obvious that the indicators Ed1gl - Ed6gl are applicable if $Eg(T_{1,n}) \equiv m_{g_1} = g(\Theta)$. For the case $m_{g_1} \neq g(\Theta)$ instead of D_{g_1} we should use the matrix $\bar{D}g_1 = Dg_1 + \text{bias } g_1(\text{bias } g_1)'$, where

$$\text{bias } g_1 = Eg_1 - g(\Theta)$$

As in (3) we have

$$tr\ \bar{D}g_1 = \sum_{i=1}^{k_g} \lambda_i(\bar{D}g_1), \quad det\ \bar{D}g_1 = \prod_{i=1}^{k_g} \lambda_i(Dg_1)$$

It is known that

$$MSE\ g_1 \equiv tr\ \bar{D}g_1, \quad MSEY \equiv tr\ \bar{D}Y = tr\ DY \tag{4.3}$$

Hence the indicators of non-relative precision-type efficiency, i.e. eff g_1 in the sense of second order moment's calculated w.r.t.g(Θ), are

$$E^*d1gl \equiv \frac{MSEg_1}{MSEY}, \quad E^*d2gl \equiv \frac{det\ \bar{D}g_1}{det\ \bar{D}Y} \tag{4.1*}$$

$$E^*d3gl \equiv \frac{\lambda_{max}(\bar{D}g_1)}{\lambda_{max}(DY)}, \quad E^*d4gl \equiv w_1^* E^*d1gl \tag{4.2*}$$

$$E^*d5gl \equiv \frac{(1 + \nu_1^*)^{k_g}}{(1 + \nu_o^*)^n} E^*d2gl, \quad E^*d6gl \equiv w_1^* E^*d3gl \tag{4.3*}$$

where $w_1^* = (1 + \nu_1^*)$, $\nu_1^* = \nu^*_{\bar{D}g_1}$, $\nu_o^* = \nu^*_{DY}$ and $\nu_A^* = \lambda_{min}(A)/$ $/\lambda_{max}(A)$ is the index of bad-conditioning of A.

N o t e: For $Eg_1 = g(\theta)$ we have $MSEg_1 = tr\ \bar{D}g_1 = tr\ Dg_1$, bias $g_1 = 0$, $\bar{D}g_1 = Dg_1$.

In the case of LM we get $\theta = \beta$, $g(\theta) = x\beta$, $MSEY = \sigma^2 n$, $T_{1,n} = x^+Y$, $g(T_{1,n}) = xT_{1,n}$, $ET_{1,n} = \theta$, $Eg(T_{1,n}) = x\theta = g(\theta)$, $MSEg_1 = \sigma^2 tr(x'x)^{-1}$, $\bar{D}T_{1,n} = DT_{1,n}$, $Dg_1 = Dg_1$. So one uses (4)-(4.2).

The indicators in (4)-(4.2), (4.1*)-(4.3*) are computable for each statistic $l \in N$, N is the set of natural numbers.

For comparison of statistics between themselves it is necessary to formulate relative dispersion-type indicators of efficiency of g_l w.r.t. g_{l^o}, $l, l^o \in N$, where l is the index of studied statistic and l^o is the index of the dispersion non relatively most efficient statistic.

In the case of indicators from (4)-(4.3*), the index l_j^o is such that

$$Edjgl_j^o = \min_{l \in N} Edjgl, \quad j = \overline{1, 6}$$

Another two options are $\underline{l}^o$ and $\bar{l}^o$ that satisfy:

$$Edg\underline{l}^o = \min_j Edjgl_j^o, \quad j = \overline{1, 6},$$

$$Edg\bar{l}^o = \max_j Edjgl_j^o, \quad j = \overline{1, 6}.$$

To these three options of fixing l^o, $\underline{l}^o$, $\bar{l}^o$ correspond the following indicators of relative dispersional efficiency of g_l, l, l^o, $\bar{l}^o$, $\underline{l}^o \in N$, $j = \overline{1, 6}$:

$$Edjgl|l_j^o = \frac{Edjgl_j^o}{Edjgl}, \quad Edjgl|l^o = \frac{Edg\underline{l}^o}{Edjgl} \tag{5}$$

$$\mathrm{Edjgl}|\underline{l}^{-o} = \frac{\mathrm{Edjgl}}{\mathrm{Edg}\bar{l}^{o}} \tag{5.1}$$

When norming principle is not so important we can put instead of l_j or $\underline{l}^o$, $\bar{l}^o$ any index $l^* \neq l$ and we get

$$\mathrm{Edjgl}|l^* = \frac{\mathrm{Edjgl}^*}{\mathrm{Edjgl}} \tag{5.2}$$

On the basis of (4.1*)-(4.3*) we formulate the following counterparts of (5)-(5.2) for l, l_j^o, $\underline{l}^o$, $\bar{l}^o \in N$, $j = \overline{1, 6}$:

$$E^*\mathrm{djgl}|l_j^o = \frac{E^*\mathrm{djgl}_j^o}{E^*\mathrm{djgl}} \tag{6}$$

$$E^*\mathrm{djgl}|\underline{l}^o = \frac{E^*\mathrm{dg}\underline{l}_j^o}{E^*\mathrm{djgl}} \tag{6.1}$$

$$E^*\mathrm{djgl}|\bar{l}^o = \frac{E^*\mathrm{djg}\bar{l}^o}{E^*\mathrm{dgl}} \tag{6.2}$$

$$E^*\mathrm{djgl}|l^* = \frac{E^*\mathrm{djgl}^*}{E^*\mathrm{djgl}} \tag{6.3}$$

where indices l_j^o, $\underline{l}^o$, $\bar{l}^o$ are fixed as above and $l^* \in N$.

The above efficiency and precision indicators will be further studied in terms of their invariance with respect to scale and translation changes.

3. INVARIANCE OF EFFICIENCY AND PRECISION INDICATORS

We are aware that the proposed indicators have different importance in studying properties of statistics by the use of Monte--Carlo studies. However, they provide the framework for more detailed evaluation of them. What kind of evaluation is needed we now show by considering the most simple dispersion type indicators. We try now to check two important properties: scale and translation invariance. Let us begin with (4). Assume that $g_1 \equiv (X'X)^{-1}X'Y \equiv X^+Y$. By well known facts about LSE of the form X^+Y we write for LM

$DY = \sigma^2 I_n$, $Dg = \sigma_1^2 (x'x)^{-1}$, $trDY = n\sigma^2$, $trDg_1 = \sigma^2 tr(x'x)^{-1}$

Let now $Y \Rightarrow aY$, $a \in R$. By properties of operator D we have

$DaY = a^2\sigma^2 I_n$, $Dg_1(a) = a^2\sigma^2(x'x)^{-1}$,

$tr\ DaY = a^2\sigma^2 n$, $tr\ Dg_1(a) = a^2\sigma^2 tr(x'x)^{-1}$.

Hence $Ed1g1 = n^{-1}\ tr(x'x)^{-1} = Ed1g1(a)$.

Suppose now that $Y \Rightarrow Y-c$. Therefore by the properties of Y, D we have $D(Y-c) = DY$, $Dg_1(c) = Dg_1$

Hence

COROLLARY 1. For Y from LM and $g_1 = x^+Y$ being the LSE, the indicator Ed1g1 is scale and translation invariant.

Take now Ed2g1. It is known on the grounds of assumptions of LM and properties of det, D, that

$\det DY = \det \sigma^2 I_n = \sigma^{2n}$, $\det Dg_1 = \sigma^{2k}\det(x'x)^{-1}$

$Ed2g1 = \sigma^{k/n}\det(x'x)^{-1}$.

Let $Y \Rightarrow aY$, $a \in R$. Then

$Ed2g1(a) = (a\sigma)^{k/n}\det(x'x)^{-1}$ since $\det DaY = (a\sigma)^{2n}$,

$\det Dg_1(a) = (a\sigma)^{2k}\det(x'x)^{-1}$

If $Y \Rightarrow Y - c$, then

$\det D(Y - c) = \det DY = \sigma^{2n}$ $\det Dg_1(c) = \det Dg_1$ $\sigma^{2k}\det(x'x)^1$.

So $Ed2g1(c) = \sigma^{k/n}\det(x'x)^{-1} = Ed2g1$. Therefore

COROLLARY 2. For Y and g_1 fulfilling Corollary 1. The indicator Ed2g1 is not scale but it is translation invariant.

Recall $\lambda_{max}(DY) = \sigma^2$, $\lambda_{max}(Dg_1) = \sigma^2\lambda_1^{-1}$, λ_1 is the smallest eigenvalue of $x'x$. So $Ed3g1 = \lambda_1^{-1}$. Now let $Y \Rightarrow aY$.

It is obvious that $\lambda_{max}(DaY) = a^2\sigma^2$, $\lambda_{max}(Dg_1(a)) = a\sigma^2\lambda_1^{-1}$. In the case $Y \Rightarrow Y - c$ we obtain $\lambda_{max}(D(Y - c)) = \lambda_{max}(DY)$, $\lambda_{max}(Dg_1(c)) = \lambda_{max}(Dg_1)$ and $ED3g1 = \lambda_1^{-1}$. Therefore

COROLLARY 3. For Y, g_1 as above the indicator Ed3gl is scale and translation invariant.

By the definitions of Ed4gl, Ed5gl, Ed6gl, and the above corollaries we have.

COROLLARY 4. For Y, g_1 from LM the indicator Ed4gl is scale and translation invariant.

COROLLARY 5. For Y, g_1 from LM the indicator Ed5gl is not scale but translation invariant.

COROLLARY 6. For Y, g_1 from LM the indicator Ed6gl is scale and translation invariant.

N o t e 1. We have omitted in § 2-4 another natural option. Let

$$\text{Ed7gl} \equiv \frac{||Dg_1||^2}{||DY||^2},$$ where $|| \ ||$ denote Euclidean norm. It is obvious that for the above Y, g_1

$$||DY||^2 = \sigma^2 n, \quad ||Dg_1||^2 = \sigma^2 \sum_{i=1}^{k} \lambda_i^{-2}$$

$$\text{Ed7gl} = n^{-1} \sum_{i=1}^{k} \lambda_i^{-2}$$

Due to $DaY = a^2\sigma^2 I_n$, $Dg_1(a) = a^2\sigma^2 (x'x)^{-1}$

$$||DaY||^2 = a^2\sigma^2 n, \quad ||Dg_1(a)||^2 = a^2\sigma^2 \sum_{i=1} \lambda_i^{-2},$$

we have

$$||Dg_1(a)||^2/||DaY||^2 = n^{-1} \sum \lambda_1^{-2},$$

and it means that Ed7gl is scale invariant.

Because of $D(Y - c) = D(Y)$, $Dg_1(c) = Dg_1$, the indicator Ed7gl is translation invariant.

Let us now consider the biased ridge estimator of Hoerl-Kennard

$$T_n \equiv B(c) = (X'X + cI)^{-1}x'Y.$$

It is well known that in the case of LM and $c > 0$

$$DT_n = \sigma^2(X'X + cI)^{-1} X'X(X'X + cI)^{-1},$$

$$trDT_n = \sigma^2 \sum_1^k \frac{\lambda_i}{(\lambda_i + c)^2}, \quad trDY = \sigma^2 n,$$

Let $Y \to Y - d$, $d \in R$. Then $trD(Y - d) = \sigma^2 n$,

$$trDT(d) = trDT_n.$$

Suppose $Y \to aY$. Then $tr\, DaY = a^2\sigma^2 n$,

$$trDT(a) = a^2\sigma^2 \Sigma \frac{\lambda_i}{(\lambda_i + c)}.$$

Hence

COROLLARY 7. The dispersion type indicator Edlgl for the Hoerl-Kennard estimator (HKE) is invariant with respect to scale and translation changes.

Recall that B(c) is biased with bias $B(c) \equiv \text{bias}\, T_n = (X'X + cI)^{-1}x'x\, \beta - \beta$. Hence

$$MSE\, T_n = trDT_n + \text{bias}^2 T_n = \sigma^2 \Sigma \frac{\lambda_i}{(\lambda_i + c)^2} + \beta' A\beta, \quad MSEY = \sigma^2 n$$

$$A = [(X'X + cI)^{-1}x'x - I]'[(X'x + cI)^{-1}X'X - I].$$

In the case of $Y \to Y - d$ we have

$$MSE\, T(d) = trDT(d) + \beta' AB + d^2 1'X(X'X + cI)^{-2}X'1 +$$
$$- 2d\beta'(X'X(X'X + cI)^{-1} - I)(X'X + cI)^{-1}X'1.$$

Since $\frac{MSE\, T_n}{MSE\, Y_n} \neq \frac{MSE\, T_n(d)}{MSE\, Y(d)}$, therefore

COROLLARY 8. The precision type indicator E^*dlgl, when l is HKE, is not scale and translation invariant.

4. FINAL REMARKS

The results obtained above are only a small part of a long list of results on invariance. In this paper we wanted to establish some preliminary invariancy results for the derived indicators of efficiency and precision. They would be a starting point for further studies. It appeared that there are situations, even under the most simple model LM, that some of the analysed indicators are not always invariant with respect to scale and translation operations.

REFERENCES

B a r r a J. (1982), *Matematyczne podstawy statystyki*, PWN, Warszawa.

D a n i e l s H. (1954), *Saddlepoint Approximation in Statistics*, "Annals of Mathematical Statistics", p. 631-640.

F e r e n t i o n o s K., P a p a i n o n n a u J. (1980), *New Parametric Measures of Information*, Technical Report, No. 41, University of Ioannina.

F i s h e r R. (1925), *Theory of Statistical Estimation*, Proceedings Cambridge Philosophical Society, p. 700-725.

K a g a n A. (1963), *On the Theory of Fisher's Amount of Information*, "Sovet Matematicheskie Dokłady", No. 4, p. 991-993.

L i n n i k J. (1960), *O nekotoreh svyazah*, Transaction 2 Prague Conference on Information Theory Prague, p. 313-327.

M i l o W. (1983), *Efektywnościowa odporność estymatorów*, "Technical Report", RIII.9, Łódź UŁ, p. 47.

M i l o W. (1989), *MSE-robustness of estimators*, "Acta Universitatis Lodziensis" (in press).

S m o l u k A. (1974), *Podstawy teorii aproksymacji i s-funkcje*, PWE, Warszawa.

Władysław Milo

CHARAKTERYZACJE EFEKTYWNOŚCI I PRECYZJI
CZĘŚĆ 1

Celem artykułu jest podanie nowych charakteryzacji wskaźników efektywności i precyzji wektorowych statystyk. Dla dwu popularnych wzorów estymatorów (MNK i estymatora obciążonego Hoerla-Kennarda) przeprowadzono analizę niezmienniczości niektórych zaproponowanych wskaźników. Pokazano, które wskaźniki są niezmiennicze i dla jakich statystyk.