Decomposing the Watson efficiency in a linear statistical model

Jarkko Isotalo, Simo Puntanen

Tampere, Finland

Ka Lok Chu, George P. H. Styan

McGill University, Montréal, Canada

Tartu, 27 June 2007

ABSTRACT

- In **Chips1** (2004), we introduced a particular decomposition for the Watson efficiency of the OLS estimator $\hat{\beta}$: *product of three factors*.

- **Chips2** (2005) shows:

  all three factors are related to the efficiencies of particular submodels or their *transformed versions*.

- There is:

  an interesting connection between a particular reduction of the Watson efficiency and the concept of *linear sufficiency*.

- The efficiency and specific *canonical correlations*.

- Decomposition for the *Bloomfield–Watson commutator criterion*, and a condition for its specific reduction.

- Efficiency of $\mathbf{K}'\hat{\beta}$

TOC

1. Introduction

Consider the partitioned linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

$$\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \sigma^2\mathbf{V}\},$$

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \quad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}.$$

- $\mathbf{y}$ is an $n \times 1$ observable random vector,
- $\boldsymbol{\varepsilon}$ is an $n \times 1$ random error vector,
- $\mathbf{X}$ is a known $n \times p$ model matrix,
- $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters.

Denote

$$\mathbf{H} = \mathbf{P}_{\mathbf{X}}, \quad \mathbf{M} = \mathbf{I} - \mathbf{H}.$$

The ordinary least squares estimator of $\mathbf{X}\boldsymbol{\beta}$:

$$\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{y}} = \mathbf{H}\mathbf{y} = \mathbf{P}_{\mathbf{X}}\mathbf{y}.$$

An unbiased estimator $\mathbf{G}\mathbf{y}$ is the **best linear unbiased estimator** (BLUE) of $\mathbf{X}\boldsymbol{\beta}$ if

$$\mathbf{G}\mathbf{V}\mathbf{G}' \leq \mathbf{B}\mathbf{V}\mathbf{B}' \quad \forall\, \mathbf{B} : \mathbf{B}\mathbf{X} = \mathbf{X},$$

i.e.,

$$\mathbf{B}\mathbf{V}\mathbf{B}' - \mathbf{G}\mathbf{V}\mathbf{G}' \text{ is nnd} \quad \forall\, \mathbf{B} : \mathbf{B}\mathbf{X} = \mathbf{X}.$$

In addition to $\mathscr{M}_{12}$, we consider

$$\mathscr{M}_1 = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1,\, \mathbf{V}\},$$

$$\mathscr{M}_{1\mathrm{H}} = \{\mathbf{Hy},\, \mathbf{X}_1\boldsymbol{\beta}_1,\, \mathbf{HVH}\},$$

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{M}_1\mathbf{y},\, \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2,\, \mathbf{M}_1\mathbf{V}\mathbf{M}_1\}.$$

$\mathscr{M}_1 =$ a *small model*, $\quad \mathscr{M}_{12\cdot 1} =$ a *reduced model*.

- $\mathscr{M}_{12\cdot 1}$ is obtained by premultiplying $\mathscr{M}_{12}$ by $\mathbf{M}_1$.
- $\mathscr{M}_{1\mathrm{H}}$ is obtained by premultiplying $\mathscr{M}_1$ by $\mathbf{H}$.

We consider a *weakly singular model* which means that $\mathbf{V}$ may be singular but

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V}). \tag{*}$$

Under this model,

$$\begin{aligned}
\tilde{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{+}\mathbf{y}, \\
\operatorname{cov}(\tilde{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-1} \\
&= \mathbf{U}^{-1}[\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}'\mathbf{V}\mathbf{X}]\mathbf{U}^{-1},
\end{aligned}$$

where $\mathbf{U} = \mathbf{X}'\mathbf{X}$.

Hence,

$$\phi_{12} = \operatorname{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12}) = \frac{|\operatorname{cov}(\tilde{\boldsymbol{\beta}})|}{|\operatorname{cov}(\hat{\boldsymbol{\beta}})|}$$

$$= \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{V}^+\mathbf{X}|}$$

$$= \frac{|\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^-\mathbf{Z}'\mathbf{V}\mathbf{X}|}{|\mathbf{X}'\mathbf{V}\mathbf{X}|}$$

$$= |\mathbf{I}_p - \mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^-\mathbf{Z}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}|,$$

where $\mathbf{Z} = \mathbf{X}^\perp$.

Chips1 introduced a new decomposition for ϕ_{12}:

$$\mathrm{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12}) = \mathrm{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_1) \cdot \mathrm{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12}) \cdot \alpha_1,$$

where $\mathrm{eff}(\cdot \mid \cdot)$ refers to the Watson efficiency under a particular model, and α_1 is a specific determinant ratio.

- One of the key results in **Chips2**:
  $1/\alpha_1 =$ the efficiency of $\hat{\boldsymbol{\beta}}_1$ under $\mathscr{M}_{1\mathrm{H}}$.
- Another interesting observation: the reduction

$$\mathrm{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12}) = \mathrm{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12}), \tag{1.2}$$

  is closely connected to the concept of linear sufficiency.

- Formally,

$$\mathbf{Fy} \text{ is defined to be } \textbf{linearly sufficient} \text{ for } \mathbf{X}\boldsymbol{\beta}$$

$$\text{under } \mathscr{M} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}\}$$

$$\text{if}$$

$$\text{there exists a matrix } \mathbf{A} \text{ such that}$$

$$\mathbf{AFy} \text{ is the BLUE of } \mathbf{X}\boldsymbol{\beta}.$$

. . . that was the the Introduction . . .

I told you it was supposed to be a serious one . . .

2. Decomposing the Efficiency

$$\operatorname{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_1) = \frac{|\mathbf{X}_1'\mathbf{X}_1|^2}{|\mathbf{X}_1'\mathbf{V}\mathbf{X}_1| \cdot |\mathbf{X}_1'\mathbf{V}^{+}\mathbf{X}_1|} := \phi_{1/1},$$

$$\begin{aligned} \operatorname{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12}) &= \operatorname{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12\cdot 1}) \\ &= \frac{|\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2|^2}{|\mathbf{X}_2'\mathbf{M}_1\mathbf{V}\mathbf{M}_1\mathbf{X}_2| \cdot |\mathbf{X}_2'\dot{\mathbf{M}}_1\mathbf{X}_2|} \\ &:= \phi_{2/12}, \end{aligned}$$

where

$$\dot{\mathbf{M}}_1 = \mathbf{M}_1(\mathbf{M}_1\mathbf{V}\mathbf{M}_1)^{-}\mathbf{M}_1.$$

- IMPORTANT NOTE:

The BLUE of $\boldsymbol{\beta}_2$ under model $\mathscr{M}_{12}$

coincides with

the BLUE of $\boldsymbol{\beta}_2$ under model $\mathscr{M}_{12\cdot 1}$, i.e.,

$$\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12\cdot 1}) = \tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}).$$

THEOREM **1** *The total Watson efficiency ϕ_{12} of the* $\mathrm{OLSE}(\boldsymbol{\beta})$ *under the partitioned weakly singular linear model* $\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}\}$*, can be expressed as the product*

$$\begin{aligned}&\mathrm{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12})\\ &= \mathrm{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_1) \cdot \mathrm{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12}) \cdot \frac{1}{\mathrm{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_{1\mathrm{H}})}\,,\end{aligned}$$

where

$$\mathrm{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_{1\mathrm{H}}) = \frac{|\mathbf{X}_1'\mathbf{X}_1|^2}{|\mathbf{X}_1'\mathbf{V}\mathbf{X}_1| \cdot |\mathbf{X}_1'(\mathbf{HVH})^-\mathbf{X}_1|}\,,$$

$$\operatorname{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_{1\mathrm{H}}) := \phi_{1\mathrm{H}}$$

$$= \frac{|\mathbf{X}_1'\mathbf{X}_1|^2}{|\mathbf{X}_1'\mathbf{V}\mathbf{X}_1| \cdot |\mathbf{X}_1'(\mathbf{HVH})^{-}\mathbf{X}_1|}$$

$$= |\mathbf{I}_{p_1} - \mathbf{X}_1'\mathbf{V}\mathrm{M}_1\mathbf{X}_2(\mathbf{X}_2'\mathrm{M}_1\mathbf{V}\mathrm{M}_1\mathbf{X}_2)^{-1} \cdot \mathbf{X}_2'\mathrm{M}_1\mathbf{V}\mathbf{X}_1'(\mathbf{X}_1'\mathbf{V}\mathbf{X}_1)^{-1}|.$$

Next we take a look at the conditions for a particular **reduction** of the total Watson efficiency.

THEOREM **2** *Let* $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *be a partitioned weakly singular linear model. Then the following statements are equivalent:*

(a) $\operatorname{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12}) = \operatorname{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12})$,

(b) $\mathscr{C}(\mathbf{X}_1) \subset \mathscr{C}(\mathbf{VX})$,

(c) $\mathbf{Hy}$ *is linearly sufficient for* $\mathbf{X}_1\boldsymbol{\beta}_1$ *under* $\mathscr{M}_1$.

3. The commutator criterion

THEOREM **3** *The Bloomfield–Watson efficiency*

$$\mathrm{eff}_{\mathrm{BW}}(\hat{\beta} \mid \mathscr{M}_{12}) = \psi_{12} = \tfrac{1}{2}\|\mathbf{HV} - \mathbf{VH}\|^2 = \|\mathbf{HVM}\|^2$$

has the decomposition

$$\psi_{12} = \psi_{1/1} + \psi_{2/12} - \psi_{1\mathrm{H}};$$

$$\psi_{1/1} = \mathrm{eff}_{\mathrm{BW}}(\hat{\beta}_1 \mid \mathscr{M}_1) = \|\mathbf{P}_1\mathbf{VM}_1\|^2,$$

$$\psi_{2/12} = \mathrm{eff}_{\mathrm{BW}}(\hat{\beta}_2 \mid \mathscr{M}_{12}) = \|\mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\mathbf{VM}\|^2,$$

$$\psi_{1\mathrm{H}} = \mathrm{eff}_{\mathrm{BW}}(\hat{\beta}_1 \mid \mathscr{M}_{1\mathrm{H}}) = \|\mathbf{P}_1\mathbf{VP}_{\mathbf{M}_1\mathbf{X}_2}\|^2.$$

Consider the condition under which

- ψ_{12} reduces into $\psi_{2/12}$.

It is interesting that this condition differs from the corresponding condition regarding the Watson efficiency.

THEOREM **4** *The B-W efficiency has property*

$$\psi_{12} = \psi_{2/12} \tag{3.2}$$

if and only if

$$\mathscr{C}(\mathbf{V}\mathbf{X}_1) \subset \mathscr{C}(\mathbf{X}). \tag{3.3}$$

Under a weakly singular linear model (3.3) *becomes*

$$\mathscr{C}(\mathbf{X}_1) \subset \mathscr{C}(\mathbf{V}^+\mathbf{X}).$$

References

Baksalary, J.K. and Kala, R. (1981). *Ann. Statist.* **4**, 913–916.

Baksalary, J.K. & Kala, R. (1986). *JSPI*, **14**, 331–338.

Barnard, G.A. (1963). *JRSS, B* **25**, 124–127.

Bartmann, F.C. & Bloomfield, P. (1981). *Biometrika*, **68**, 67–71.

Bloomfield, P. & Watson, G.S. (1975). *Biometrika*, **62**, 121–128.

Chu, K.L., Isotalo, J., Puntanen, S. & Styan, G.P.H. (2004). (**Chips1**) *Sankhyā*, **66**, 634–651.

Chu, K.L., Isotalo, J., Puntanen, S. & Styan, G.P.H. (2005). (**Chips2**) *Sankhyā*, **67**, 74–89.

Drygas, H. (1983). *Sankhyā, Ser. A* **45**, 88–98.

Groß, J. & Puntanen, S. (2000). *LAA* **321**, 131–144.

Rao, C.R. & Mitra, S.K. (1971). Wiley, New York.

Watson, G.S. (1955). *Biometrika* **42**, 327–341.

Zyskind, G. & Martin, F.B. (1969). *SIAM J. Appl. Math.* **17**, 1190–1202.