A Note on Estimation Via Linearly Combining Two Given Statistics

Jürgen Groß

Department of Statistics
University of Dortmund
Vogelpothsweg 87
D-44221 Dortmund, Germany

Abstract. Linear combination of two statistics is considered when some prior knowledge about their expectation and complete knowledge about their joint dispersion is available. The considered setup is more general than those already known in the literature, in the sense that the expectation of one of the statistics is not necessarily assumed to be completely known when estimation of the expectation of the other statistic is of interest.

AMS 1991 subject classification: 62H12.

Key words: Linear combination of statistics, Gauss–Markov model, minimum dispersion linear unbiased estimation, covariance adjustment estimation, prediction.

1. Introduction. Let us be given two statistics $\mathbf{U}_1$ and $\mathbf{U}_2$ with expectation vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ respectively. Suppose that we wish to estimate the k–dimensional parameter vector $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_1 \in \mathcal{X}_1$ where $\mathcal{X}_1$ is a known subspace of the space of all k–dimensional real vectors. Additionally the l–dimensional vector $\boldsymbol{\theta}_2$ is known to lie in $\mathcal{X}_2$, where $\mathcal{X}_2$ is a given subspace of the space of all l–dimensional real vectors, and the joint dispersion matrix of $\mathbf{U}_1$ and $\mathbf{U}_2$ is known apart from a positive scalar.

Now, it is quite natural to ask whether it is possible to combine the statistics $\mathbf{U}_1$ and $\mathbf{U}_2$ in such a way that the additional information will lead to improved estimation of $\boldsymbol{\theta}_1$. Baksalary and Kala [2] derived explicit formulas for the best linear combination of $\mathbf{U}_1$ and $\mathbf{U}_2$ as an unbiased estimator for $\boldsymbol{\theta}_1$ for the special cases that $\boldsymbol{\theta}_2 = \mathbf{0}$, $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ being a subvector of $\boldsymbol{\theta}_1$. In this note we wish to hint to the general case, $\boldsymbol{\theta}_2 \in \mathcal{X}_2$, by adopting the method of Baksalary and Kala [2], viz identifying the best linear combination of $\mathbf{U}_1$ and $\mathbf{U}_2$ with *minimum dispersion linear unbiased estimation* under an appropriate Gauss–Markov model. However, we do not assume that some relationship is known a priori between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, and therefore our results will not cover two of the cases investigated in [2]. As it will be seen subsequently, our discussion will lead to an easy interpretation of the so called *covariance adjustment estimator*, that is the best linear combination of $\mathbf{U}_1$ and $\mathbf{U}_2$ when $\boldsymbol{\theta}_2 = \mathbf{0}$.

2. Preliminaries. Let $I\!R_{m\times n}$, $I\!R_m^s$, and $I\!R_m^{\geq}$ denote the set of $m \times n$ real matrices, the subset of $I\!R_{m\times m}$ consisting of symmetric matrices, and the subset of $I\!R_m^s$ consisting of nonnegative definite matrices, respectively. The symbols $\mathbf{A}'$, $\mathbf{A}^-$, $\mathbf{A}^+$ and $\mathcal{R}(\mathbf{A})$ will stand for the transpose, any generalized inverse, the Moore-Penrose inverse, and the range of $\mathbf{A} \in I\!R_{m\times n}$. Recall that $\mathbf{A}^-$ is a generalized inverse of $\mathbf{A}^-$ if it is a solution to $\mathbf{AGA} = \mathbf{A}$ with respect to $\mathbf{G}$, whereas $\mathbf{A}^+$ is the unique solution to the four equations $\mathbf{AGA} = \mathbf{A}$, $\mathbf{GAG} = \mathbf{G}$, $\mathbf{AG} = (\mathbf{AG})'$ and $\mathbf{GA} = (\mathbf{GA})'$ with respect to $\mathbf{G}$.

Consider the general Gauss–Markov model denoted by

$$M = \{\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\},$$

where $\mathbf{Y}$ is an $n \times 1$ observable random vector with $\mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{D}(\mathbf{Y}) = \sigma^2\mathbf{V}$. The operators $\mathrm{E}(\cdot)$ and $\mathrm{D}(\cdot)$ stand for expectation vector and dispersion matrix, respectively, of a random vector argument. The matrices $\mathbf{X} \in I\!R_{n\times p}$ and $\mathbf{V} \in I\!R_n^{\geq}$ are known, whereas $\boldsymbol{\beta} \in I\!R_{p\times 1}$ and $\sigma^2 > 0$ are unknown parameters.

If $\mathbf{K} \in I\!R_{r\times p}$, then the vector of parametric functions $\mathbf{K}\boldsymbol{\beta}$ is known to be unbiasedly estimable under M if and only if $\mathcal{R}(\mathbf{K}') \subseteq \mathcal{R}(\mathbf{X}')$. In that case, $\mathbf{LY} + \mathbf{l}$ is the minimum dispersion linear unbiased estimator (MDLUE) of $\mathbf{K}\boldsymbol{\beta}$ under M if and only if

$$(\mathbf{LX} : \mathbf{LVX}^{\perp} : \mathbf{l}) = (\mathbf{K} : \mathbf{0} : \mathbf{0}), \tag{2.1}$$

where $\mathbf{X}^{\perp}$ denotes any matrix such that $\mathcal{R}(\mathbf{X}^{\perp}) = \mathcal{N}(\mathbf{X}')$, cf. Rao [13] or Drygas [8]. The term 'minimum dispersion' is understood in the usual sense of the nonnegative definite (Löwner) ordering between dispersion matrices of estimators. Explicit solutions to the system of equations (2.1) are well known and widely discussed in the literature. For example, an appropriate choice for $\mathbf{L}$ would be

$$\mathbf{L} = \mathbf{K}(\mathbf{X}'\mathbf{T}^-\mathbf{X})^-\mathbf{X}'\mathbf{T}^-, \tag{2.2}$$

where $\mathbf{T} = \mathbf{V} + \mathbf{X}\mathbf{X}'$, cf. [12].

Suppose now that $\mathbf{Y} = (\mathbf{Y}_1' : \mathbf{Y}_2')'$, where $\mathbf{Y}_1$ is of dimension $k \times 1$ and $\mathbf{Y}_2$ is of dimension $l \times 1$ and $k + l = n$. Moreover, assume that

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix},$$

where $\mathbf{X}_1 \in I\!R_{k \times p_1}$, $\mathbf{X}_2 \in I\!R_{l \times p_2}$, and $p_1 + p_2 = p$. Then partitioning of $\boldsymbol{\beta}$ and $\mathbf{V}$ accordingly leads to the Gauss–Markov model

$$\left\{\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \sigma^2\begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}\right\}. \tag{2.3}$$

Under the above model, estimation of $\mathrm{E}(\mathbf{Y}_1) = \mathbf{X}_1\boldsymbol{\beta}_1$ means estimation of the vector of parametric functions $\mathbf{K}\boldsymbol{\beta}$ with $\mathbf{K} = (\mathbf{I}_k : \mathbf{0})\mathbf{X}$. By choosing

$$\mathbf{X}^{\perp} = \mathbf{I}_n - \mathbf{X}\mathbf{X}^{+} = \begin{pmatrix} \mathbf{I}_k - \mathbf{X}_1\mathbf{X}_1^{+} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_l - \mathbf{X}_2\mathbf{X}_2^{+} \end{pmatrix},$$

it is easily seen from (2.1) that $\mathbf{L}_1\mathbf{Y}_1 + \mathbf{L}_2\mathbf{Y}_2$ is the MDLUE for $\mathrm{E}(\mathbf{Y}_1)$ if and only if $\mathbf{L}_1$ and $\mathbf{L}_2$ are two solutions to the system of equations

$$(\mathbf{L}_1\mathbf{X}_1 : \mathbf{L}_2\mathbf{X}_2) = (\mathbf{X}_1 : \mathbf{0}), \tag{2.4}$$

$$\mathbf{L}_1\mathbf{V}_{11}(\mathbf{I}_k - \mathbf{X}_1\mathbf{X}_1^{+}) + \mathbf{L}_2\mathbf{V}_{21}(\mathbf{I}_k - \mathbf{X}_1\mathbf{X}_1^{+}) = \mathbf{0}, \tag{2.5}$$

and

$$\mathbf{L}_1\mathbf{V}_{12}(\mathbf{I}_l - \mathbf{X}_2\mathbf{X}_2^{+}) + \mathbf{L}_2\mathbf{V}_{22}(\mathbf{I}_l - \mathbf{X}_2\mathbf{X}_2^{+}) = \mathbf{0}. \tag{2.6}$$

Explicit solutions to (2.4), (2.5) and (2.6) may for example be obtained by evaluating (2.2). Clearly, if $\mathbf{V}_{12} = \mathbf{0}$, then we may choose $\mathbf{L}_2 = \mathbf{0}$ and $\mathbf{L}_1$ such that $\mathbf{L}_1\mathbf{Y}_1$ is the MDLUE of $\mathbf{X}_1\boldsymbol{\beta}_1$ under a model $\{\mathbf{Y}_1, \mathbf{X}_1\boldsymbol{\beta}_1, \sigma^2\mathbf{V}_{11}\}$.

As a further preliminary result we introduce the following lemma due to Albert [1].

LEMMA. *Let* $\mathbf{A} \in I\!R^{s}_{k+l}$ *be partitioned as*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}.$$

Then $\mathbf{A} \in I\!R^{\geq}_{k+l}$ *if and only if* $\mathbf{A}_{22} \in I\!R^{\geq}_{l}$, $\mathcal{R}(\mathbf{A}_{21}) \subseteq \mathcal{R}(\mathbf{A}_{22})$ *and*

$$\mathbf{A}_{*} := \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}^{-}_{22}\mathbf{A}_{21} \in I\!R^{\geq}_{k}$$

for some (and hence every) choice of generalized inverse $\mathbf{A}^{-}_{22}$.

Note that the invariance of $\mathbf{A}_{*}$ with respect to the choice of the generalized inverse $\mathbf{A}^{-}_{22}$ is implied by $\mathcal{R}(\mathbf{A}_{21}) \subseteq \mathcal{R}(\mathbf{A}_{22})$, since the latter may also be expressed as $\mathbf{A}_{21} = \mathbf{A}_{22}\mathbf{G}$ for some matrix $\mathbf{G}$. Therefore $\mathbf{A}_{12} = \mathbf{A}'_{21} = \mathbf{G}'\mathbf{A}_{22}$, and $\mathbf{A}_{*} = \mathbf{A}_{11} - \mathbf{G}'\mathbf{A}_{22}\mathbf{A}^{-}_{22}\mathbf{A}_{22}\mathbf{G} = \mathbf{A}_{11} - \mathbf{G}'\mathbf{A}_{22}\mathbf{G}$. The matrix $\mathbf{A}_{*}$ is also known as the *generalized Schur complement* of $\mathbf{A}_{22}$ in $\mathbf{A}$, cf. [7, 14].

3. Best linear combination of two statistics. Now suppose that $\mathbf{U}_1$ and $\mathbf{U}_2$ are two statistics with $\mathrm{E}(\mathbf{U}_1) = \boldsymbol{\theta}_1$ and $\mathrm{E}(\mathbf{U}_2) = \boldsymbol{\theta}_2$, where $\boldsymbol{\theta}_1 \in \mathcal{X}_1$ and $\boldsymbol{\theta}_2 \in \mathcal{X}_2$ and the $k \times 1$ and $l \times 1$ subspaces $\mathcal{X}_1$ and $\mathcal{X}_2$ are known. Clearly we can identify $\mathcal{X}_1$ and $\mathcal{X}_2$ with the column spaces of two known matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, say, i.e. $\mathcal{X}_1 = \mathcal{R}(\mathbf{X}_1)$ and $\mathcal{X}_2 = \mathcal{R}(\mathbf{X}_2)$. In addition the dispersion matrix of the joint vector $(\mathbf{U}'_1 : \mathbf{U}'_2)'$,

$$\mathrm{D}\begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{pmatrix} = \sigma^2 \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix},$$

is assumed to be known apart from the positive scalar σ^2.

If we are interested in estimation of one of the parameter vectors, $\boldsymbol{\theta}_1$ say, then it is quite natural to ask whether also the information delivered by $\mathbf{U}_2$ can be incorporated into the estimation procedure. The resulting estimator should be a linear combination of $\mathbf{U}_1$ and $\mathbf{U}_2$ which is unbiased for $\boldsymbol{\theta}_1$. This means the estimator should lie in the class

$$\mathcal{U} \equiv \{\mathbf{U} = \mathbf{L}_1\mathbf{U}_1 + \mathbf{L}_2\mathbf{U}_2 : \mathbf{L}_1 \in I\!R_{k\times k}, \mathbf{L}_2 \in I\!R_{k\times l}, \mathrm{E}(\mathbf{U}) = \boldsymbol{\theta}_1\, \forall \boldsymbol{\theta}_1 \in \mathcal{R}(\mathbf{X}_1)\}.$$

The optimal choice of estimator $\widehat{\mathbf{U}}$ from this class should have minimum dispersion, that is $\mathrm{D}(\mathbf{U}) - \mathrm{D}(\widehat{\mathbf{U}}) \in I\!R_k^{\geq}$ for all $\mathbf{U} \in \mathcal{U}$. It is quite obvious that the best linear combination of $\mathbf{U}_1$ and $\mathbf{U}_2$ in this sense is equal to the MDLUE of $\mathbf{X}_1\boldsymbol{\beta}_1$ under the Gauss–Markov model (2.3). Thus we may state:

PROPOSITION 1. *The best linear combination of the statistics* $\mathbf{U}_1$ *and* $\mathbf{U}_2$ *for estimating* $\boldsymbol{\theta}_1$ *is given by* $\mathbf{L}_1\mathbf{U}_1 + \mathbf{L}_2\mathbf{U}_2$, *where* $\mathbf{L}_1$ *and* $\mathbf{L}_2$ *are any solutions to the equations* (2.4), (2.5) *and* (2.6).

It is clear that when $\mathbf{U}_1$ and $\mathbf{U}_2$ are uncorrelated, that is $\mathbf{V}_{12} = \mathbf{0}$ and $\mathbf{V}_{21} = \mathbf{V}_{12}' = \mathbf{0}$, then we cannot expect any advantage from using $\mathbf{U}_2$ in estimating $\boldsymbol{\theta}_1$. Indeed, as mentioned in the previous section, in that case $\mathbf{L}_2 = \mathbf{0}$ and $\mathbf{L}_1$ is such that $\mathbf{L}_1\mathbf{U}_1$ is the MDLUE of $\mathbf{X}_1\boldsymbol{\beta}_1$ under the model $\{\mathbf{U}_1, \mathbf{X}_1\boldsymbol{\beta}_1, \sigma^2\mathbf{V}_{11}\}$.

Our prior knowledge about the parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ is comprised in the subspaces $\mathcal{X}_1 = \mathcal{R}(\mathbf{X}_1)$ and $\mathcal{X}_2 = \mathcal{R}(\mathbf{X}_2)$, respectively. If $\mathcal{X}_1 = I\!R_{k\times 1}$ (or $\mathcal{X}_2 = I\!R_{l\times 1}$), then we may say that $\boldsymbol{\theta}_1$ (or $\boldsymbol{\theta}_2$) is completely unknown. If $\mathcal{X}_2 = \{\boldsymbol{\vartheta}_2\}$, where $\boldsymbol{\vartheta}_2$ is a known $l \times 1$ vector, then $\boldsymbol{\theta}_2$ is completely known. In this case we may assume without loss of generality that $\boldsymbol{\theta}_2 = \mathbf{0}$ since we may as well consider the statistic $\mathbf{U}_2 - \boldsymbol{\vartheta}_2$ instead of $\mathbf{U}_2$.

If we consider the special case that $\boldsymbol{\theta}_2$ is completely unknown, then we may choose $\mathbf{X}_2 = \mathbf{I}_l$, and it is easily seen that solutions to (2.4), (2.5) and (2.6) are again given by $\mathbf{L}_2 = \mathbf{0}$ and $\mathbf{L}_1$ being such that $\mathbf{L}_1\mathbf{U}_1$ is the MDLUE of $\mathbf{X}_1\boldsymbol{\beta}_1$ under the model $\{\mathbf{U}_1, \mathbf{X}_1\boldsymbol{\beta}_1, \sigma^2\mathbf{V}_{11}\}$. If in addition $\boldsymbol{\theta}_1$ is completely unknown, that is $\mathbf{X}_1 = \mathbf{I}_k$, then $\mathbf{L}_1 = \mathbf{I}_k$. Thus we may state:

PROPOSITION 2. *If the two statistics* $\mathbf{U}_1$ *and* $\mathbf{U}_2$ *are uncorrelated, or* $\boldsymbol{\theta}_2$ *is completely unknown, then every linear combination of* $\mathbf{U}_1$ *and* $\mathbf{U}_2$ *for estimating* $\boldsymbol{\theta}_1$ *is worse than the MDLUE of* $\mathbf{X}_1\boldsymbol{\beta}_1$ *under the model* $\{\mathbf{U}_1, \mathbf{X}_1\boldsymbol{\beta}_1, \sigma^2\mathbf{V}_{11}\}$. *The latter is* $\mathbf{U}_1$ *itself if in addition* $\boldsymbol{\theta}_1$ *is completely unknown.*

We note that the situation when both $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are completely unknown does not cover Corollary 2.2 from [2]. This is so because in our approach no relationship between the parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ is assumed to be a priori known. Therefore our Proposition 1 does not cover the cases $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ being a subvector of $\boldsymbol{\theta}_1$ which were investigated in [2].

As mentioned above, the case that $\boldsymbol{\theta}_2$ is completely known may be regarded as the case of $\boldsymbol{\theta}_2 = \mathbf{0}$ and hence $\mathbf{X}_2 = \mathbf{0}$. In the literature estimation in this case is referred as *covariance adjustment estimation*, cf. [11, 10, 2]. When $\mathbf{X}_2 = \mathbf{0}$, then the equations (2.4), (2.5) and (2.6) reduce to

$$\mathbf{L}_1\mathbf{X}_1 = \mathbf{X}_1, \tag{3.1}$$

$$\mathbf{L}_1\mathbf{V}_{11}(\mathbf{I}_k - \mathbf{X}_1\mathbf{X}_1^+) + \mathbf{L}_2\mathbf{V}_{21}(\mathbf{I}_k - \mathbf{X}_1\mathbf{X}_1^+) = \mathbf{0}, \tag{3.2}$$

$$\mathbf{L}_1\mathbf{V}_{12} + \mathbf{L}_2\mathbf{V}_{22} = \mathbf{0}. \tag{3.3}$$

From the Lemma in Section 2 we know that $\mathcal{R}(\mathbf{V}_{21}) \subseteq \mathcal{R}(\mathbf{V}_{22})$, which may be written as $\mathcal{R}(\mathbf{V}'_{12}) \subseteq \mathcal{R}(\mathbf{V}'_{22})$, or equivalently $\mathbf{V}_{12}\mathbf{V}_{22}^-\mathbf{V}_{22} = \mathbf{V}_{12}$ for any choice of generalized inverse $\mathbf{V}_{22}^-$. Thus a special solution to (3.3) with respect to $\mathbf{L}_2$ is given by

$$\mathbf{L}_2 = -\mathbf{L}_1\mathbf{V}_{12}\mathbf{V}_{22}^-$$

for an arbitrary choice of $\mathbf{V}_{22}^-$. Substituting this special solution into (3.2) shows that (3.1) and (3.2) become equivalent to

$$[\mathbf{L}_1\mathbf{X}_1 : \mathbf{L}_1\mathbf{V}_*(\mathbf{I}_k - \mathbf{X}_1\mathbf{X}_1^+)] = (\mathbf{X}_1 : \mathbf{0}), \tag{3.4}$$

where $\mathbf{V}_* = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^-\mathbf{V}_{21}$ is the generalized Schur complement of $\mathbf{V}_{22}$ in $\mathbf{V}$. Since by the Lemma we have $\mathbf{V}_* \in I\!R_k^{\geq}$, $\mathbf{L}_1$ is a solution to (3.4) if and only if $\mathbf{L}_1\mathbf{U}_1$ is the MDLUE of $\mathbf{X}_1\boldsymbol{\beta}_1$ under the model $\{\mathbf{U}_1, \mathbf{X}_1\boldsymbol{\beta}_1, \sigma^2\mathbf{V}_*\}$, see also (2.1). Thus we may state:

PROPOSITION 3. *If $\boldsymbol{\theta}_2$ is completely known (that is $\boldsymbol{\theta}_2 = \mathbf{0}$), then the best linear combination of the statistics $\mathbf{U}_1$ and $\mathbf{U}_2$ for estimating $\boldsymbol{\theta}_1$ is given by $\mathbf{L}_1\mathbf{U}_1 - \mathbf{L}_1\mathbf{V}_{12}\mathbf{V}_{22}^-\mathbf{U}_2$, where the choice of $\mathbf{V}_{22}^-$ is arbitrary and $\mathbf{L}_1$ is such that $\mathbf{L}_1\mathbf{U}_1$ is the MDLUE of $\mathbf{X}_1\boldsymbol{\beta}_1$ under the model $\{\mathbf{U}_1, \mathbf{X}_1\boldsymbol{\beta}_1, \sigma^2\mathbf{V}_*\}$. The latter is $\mathbf{U}_1$ itself if in addition $\boldsymbol{\theta}_1$ is completely unknown.*

The addendum of Proposition 3 is easy to prove. It states that if $\boldsymbol{\theta}_1$ is completely unknown and $\boldsymbol{\theta}_2 = \mathbf{0}$, then the best linear combination is

$$\mathbf{U}_1 - \mathbf{V}_{12}\mathbf{V}_{22}^{-}\mathbf{U}_2.$$

This has also been observed in [2, Corollary 1.2]. Actually, also the remainder of Proposition 3 could follow by appropriate combination of Lemma 1, Lemma 3 and Theorem 1 in [2].

4. Other concepts of linear combination of two statistics. Assuming nonsingularity of $\mathbf{V}$, Baksalary and Trenkler [6] investigated linear combinations of the form

$$\mathbf{L}_1\mathbf{U}_1 - \mathbf{L}_1\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{U}_2 \tag{4.1}$$

with $\mathbf{L}_1 = \mathbf{F}(\mathbf{F}'\mathbf{V}_*^{-1}\mathbf{F})^{-}\mathbf{F}'\mathbf{V}_*^{-1}$, $\mathbf{V}_* = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$, and $\mathbf{F}$ being any known matrix such that $\mathcal{R}(\mathbf{F}) \subseteq \mathcal{R}(\mathbf{X}_1) = \mathcal{X}_1$. Such combinations were considered by the authors [6] as estimators for $\boldsymbol{\theta}_1$ when $\boldsymbol{\theta}_2$ is completely unknown. Although, by Proposition 2, these estimators are not best linear combinations for estimating $\boldsymbol{\theta}_1$, the authors [6] demonstrated that under certain conditions they can outperform the best estimator $\mathbf{L}_1\mathbf{U}_1$ being the MDLUE of $\mathbf{X}_1\boldsymbol{\beta}_1$ under $\{\mathbf{U}_1, \mathbf{X}_1\boldsymbol{\beta}_1, \sigma^2\mathbf{V}_{11}\}$. The motivation for the derivations is the observation that any estimator of the form (4.1) is admissible for $\boldsymbol{\theta}_1$ among the set of all linear combinations of $\mathbf{U}_1$ and $\mathbf{U}_2$ and with respect to the mean square error criterion. Note that this criterion would lead to a greater reduction of the class of linear estimators than the mean square error matrix criterion, cf. [5].

However, the results might be generalized by considering $\boldsymbol{\theta}_2$ to be restricted to some subspace $\mathcal{X}_2$ instead of assuming $\mathcal{X}_2 = I\!R_{l\times 1}$. Essentially, this would require the identification of all linearly admissible estimators under model (2.3), which could be done by applying the results of Baksalary and Markiewicz [3, 4]. Once a certain admissible linear combination of $\mathbf{U}_1$ and $\mathbf{U}_2$ is identified, conditions for outperforming the best linear combination, the latter being $\mathbf{L}_1\mathbf{U}_1 + \mathbf{L}_2\mathbf{U}_2$ with solutions $\mathbf{L}_1$ and $\mathbf{L}_2$ to (2.4), (2.5) and (2.6), can be investigated.

5. Using combination of statistics for prediction. Let us be given two statistics $\mathbf{U}_1$ and $\mathbf{U}_2$ as in Section 3. Suppose that the realization of an unobservable random variable w is to be predicted, where

$$\mathrm{D}\begin{pmatrix} \mathbf{U}_1 \\ w \end{pmatrix} = \sigma^2 \begin{pmatrix} \mathbf{V}_{11} & \boldsymbol{v}_{12} \\ \boldsymbol{v}_{21} & v_{22} \end{pmatrix}$$

is known apart from σ^2. For simplicity let $\mathbf{V}_{11}$ be nonsingular. Moreover, $\mathrm{E}(w)$ is assumed to be a known linear combination of the elements of $\mathrm{E}(\mathbf{U}_1) = \boldsymbol{\theta}_1 \in \mathcal{X}_1$, that is $\mathrm{E}(w) = \mathbf{c}'\boldsymbol{\theta}_1$, where $\mathbf{c} \in I\!R_{k\times 1}$ is known. This situation describes state 3 of knowledge as introduced in [9]. Following Harville [9, Sec. 2.3], a possible predictor is given by

$$\widehat{\tau} + \boldsymbol{v}_{21}\mathbf{V}_{11}^{-1}\mathbf{U}_1,$$

where $\widehat{\tau}$ is an estimator for

$$\tau = (\mathbf{c}' - \boldsymbol{v}_{21}\mathbf{V}_{11}^{-1})\boldsymbol{\theta}_1.$$

Thus, there is a one-to-one correspondence between estimators of τ and predictors of the realization of w. Since estimators of τ may rely upon estimators of $\boldsymbol{\theta}_1$, it is reasonable to use the best linear combination of $\mathbf{U}_1$ and $\mathbf{U}_2$ from Proposition 1 for estimating $\boldsymbol{\theta}_1$, in order to obtain a better predictor for the realization of w.

6. Conclusion. If the two statistics $\mathbf{U}_1$ and $\mathbf{U}_2$ are correlated and the dimension of the subspace $\mathcal{X}_2$ of the $l \times 1$ parameter vector $\boldsymbol{\theta}_2 \in \mathcal{X}_2$ is less than l, then we can expect an advantage from linearly combining $\mathbf{U}_1$ and $\mathbf{U}_2$. The best linear combination for estimating $\boldsymbol{\theta}_1$ is $\widehat{\mathbf{U}} = \mathbf{L}_1\mathbf{U}_1 + \mathbf{L}_2\mathbf{U}_2$, where $\mathbf{L}_1$ and $\mathbf{L}_2$ are any solutions to (2.4), (2.5) and (2.6). It is reasonable to conjecture that the advantage from using $\widehat{\mathbf{U}}$ grows with our prior knowledge about $\boldsymbol{\theta}_2$, which becomes more when the dimension of $\mathcal{X}_2$ becomes smaller.

Acknowledgements. This work was motivated by the research project 'Complexity Reduction in Multivariate Data Structures', supported by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich 475. Financial support was provided by DFG under grant Tr 253/2-3.

REFERENCES

1. A. Albert, Conditions for positive and nonnegative definiteness in terms of pseudoinverses, *SIAM Journal on Applied Mathematics*, 17:434–440.

2. J.K. Baksalary and R. Kala, Estimation via linearly combining two given statistics, *The Annals of Statistics* 11:691–696 (1983).

3. J.K. Baksalary and A. Markiewicz, Admissible linear estimators in the general Gauss–Markov model, *Journal of Statistical Planning and Inference* 19:349–359 (1988).

4. J.K. Baksalary and A. Markiewicz, A matrix inequality and admissibility of linear estimators with respect to the mean square error matrix criterion, *Linear Algebra and Its Applications* 112:9–18 (1989).

5. J.K. Baksalary, C.R. Rao and A. Markiewicz, A study of the influence of the 'natural restrictions' on estimation problems in the singular Gauss-Markov model. *Journal of Statistical Planning and Inference* 31:335–351 (1992).

6. J.K. Baksalary and G. Trenkler, Covariance adjustment in biased estimation, *Computational Statistics & Data Analysis*, 12:221–230 (1991).

7. D. Carlson, What are Schur complements anyway?, *Linear Algebra and Its Applications* 59:188–193 (1984).

8. H. Drygas, *The Coordinate–Free Approach to Gauss–Markov Estimation*, Springer, Berlin, 1970.

9. D.A. Harville, Decomposition of prediction error, *Journal of the American Statistical Association*, 80:132–138 (1985).

10. T.O. Lewis and P.L. Odell, *Estimation in Linear Modells*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

11. C.R. Rao, Least squares theory using an estimated dispersion matrix and its application to measurement of signals, in: L.M. Le Cam and J. Neyman (Eds.), *Proceedings of the Fifth Berkely Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, CA, 355–372 (1967).

12. C.R. Rao, Unified theory of linear estimation, *Sankhyā, Ser. A* 33:371–394 (1971); Corrigenda, 34:194, 477 (1972).

13. C.R. Rao, Representations of best linear estimators in the Gauss–Markoff model with a singular dispersion matrix, *Journal on Multivariate Analysis* 3:276–292 (1973).

14. G.P.H. Styan, Schur complements and linear statistical models, in: T. Pukkila and S. Puntanen (Eds.), *Proceedings of the First Tampere Seminar on Linear Models*, Department of Mathematical Sciences, University of Tampere, Tampere, 37–75 (1985).