Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT

Journal of Multivariate Analysis 86 (2003) 242–253

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

# Best affine unbiased response decomposition

Geert Dhaene,* Erik Schokkaert, and Carine Van de Voorde

*Department of Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium*

Received 13 April 2000

**Abstract**

Given two linear regression models $y_1 = X_1\beta_1 + u_1$ and $y_2 = X_2\beta_2 + u_2$ where the response vectors $y_1$ and $y_2$ are unobservable but the sum $y = y_1 + y_2$ is observable, we study the problem of decomposing $y$ into components $\hat{y}_1$ and $\hat{y}_2$, intended to be close to $y_1$ and $y_2$, respectively. We develop a theory of best affine unbiased decomposition in this setting. A necessary and sufficient condition for the existence of an affine unbiased decomposition is given. Under this condition, we establish the existence and uniqueness of the best affine unbiased decomposition and provide an expression for it.


*AMS 2000 subject classifications:* 62H12; 62J05

*Keywords:* Decomposability; Decomposition of response; Best affine unbiasedness; Gauss–Markov theory

## 1. Introduction

The idea of best linear unbiasedness in parameter estimation, response prediction and prediction of disturbances has a long history in statistics. The theory of best linear unbiased estimation by the method of least-squares originated with Gauss [4]. See Stigler [10] and Farebrother [3] for historical details. Subsequent contributions were made by Aitken [2] and Rao [8], inter alia. Goldberger [5] developed a theory of best linear unbiased prediction in the linear model. Theil [11] and other authors (e.g. Abrahamse and Koerts [1], Neudecker [7]) considered best linear unbiased prediction of the disturbances, subject to constraints on the covariance matrix of the predictors.

*Corresponding author.

*E-mail addresses:* geert.dhaene@econ.kuleuven.ac.be (G. Dhaene), erik.schokkaert@econ.kuleuven.ac.be (E. Schokkaert), carine.vandevoorde@econ.kuleuven.ac.be (C. Van de Voorde).


doi:10.1016/S0047-259X(03)00023-X

This paper extends the idea of best linear, or affine, unbiasedness to the decomposition of the response vector in a linear model into two (or more) additive components. Each of these components is attributed to a specific set of regressors and a specific disturbance term. Another way of looking at the same problem is as follows. Suppose we have two linear models with unobservable response vectors, but where the sum of the two response vectors is observable. The question is then, how to retrieve the original response vectors. We note that the additivity assumption can, at least partly, be tested by the inclusion of interaction terms between regressors. Should an interaction term be required, it would simply assume the role of an additional regressor and thus be associated with one of the components that one is seeking to distinguish.

We have in mind applications in the social sciences such as, for example, the decomposition of school results of pupils into an environmental component, on the one hand, and a component related to innate characteristics of the pupil, on the other. For an application in health economics, see Schokkaert et al. [9]. Individual medical expenditures are decomposed into a component related to the health condition of the individual and a component consisting of cultural factors and wealth. Potential applications are not limited to the social sciences, however.

The paper is organized as follows. A formal statement of the problem is given in Section 2, along with definitions of unbiased decomposition, decomposability and best affine unbiased decomposition. Section 3 gives a necessary and sufficient condition for the existence of an affine unbiased decomposition. Section 4 establishes that if an affine unbiased decomposition exists, then there exists a unique best affine unbiased decomposition. Section 5 extends the theory to cover the cases where: (i) a linear combination of the response vectors is given instead of their sum; (ii) the decomposition into $p$ components is sought; (iii) some parts of the response vectors are separately observable, while for the remaining parts only their sum is observable. Section 6 discusses the issue of identifiability of the covariance matrices appearing in the decomposition. Section 7 concludes.

## 2. Statement of the problem

Consider the standard linear regression models

$$y_1 = X_1\beta_1 + u_1, \qquad y_2 = X_2\beta_2 + u_2,$$

with disturbance vectors $u_1$ and $u_2$ satisfying $E(u_i) = 0$ and $E(u_i u_j') = \sigma^2 V_{ij}$ $(i,j = 1,2)$. Suppose that one only observes the $n \times k_1$ and $n \times k_2$ non-stochastic regressor matrices $X_1$ and $X_2$, and the sum of the response vectors

$$y = y_1 + y_2 = X\beta + u, \tag{1}$$

where $X = (X_1 : X_2)$, $\beta = (\beta_1' : \beta_2')'$ and $u = u_1 + u_2$. Let $k = k_1 + k_2$. We assume here that the matrices $V_{ij}$ $(i,j = 1,2)$ are known. On the other hand, $\sigma^2$ (assumed positive) and $\beta$ are unknown, although extraneous information on $\beta$ may be present in the form

$$R\beta = r, \tag{2}$$

where the $m \times k$ matrix $R$ and the $m$ vector $r$ are known and non-stochastic. Thus the parameter space of $\beta$, denoted $\mathscr{B}$, is the Euclidean space $\Re^k$, or the affine subspace of $\Re^k$ determined by (2). When no extraneous information is present, $R = 0$ and $r = 0$.

The problem that we address consists of finding a *decomposition* of $y$, that is, a pair $(\hat{y}_1, \hat{y}_2)$ such that $\hat{y}_1 + \hat{y}_2 = y$. The obvious interest lies in finding a decomposition such that $\hat{y}_1$ and $\hat{y}_2$ are close to $y_1$ and $y_2$, respectively. We seek an *affine* decomposition, that is, one of the form

$$\hat{y}_1 = a + Ay, \qquad \hat{y}_2 = y - \hat{y}_1,$$

where $a$ and $A$ are non-stochastic. Further, we shall say that a decomposition $(\hat{y}_1, \hat{y}_2)$ of $y$ is *unbiased* if

$$E(\hat{y}_1 - y_1) = 0, \qquad E(\hat{y}_2 - y_2) = 0 \qquad \text{for all } \beta \in \mathscr{B}.$$

If there exists an affine unbiased decomposition of $y$, we say that $y$ is *decomposable*. A necessary and sufficient condition for $y$ to be decomposable will be given in the next section. If $y$ is decomposable, it is natural to look for a *best* affine unbiased decomposition, say $(\hat{y}_1, \hat{y}_2)$, defined by the requirement that

$$\mathrm{Var}\begin{pmatrix} \tilde{y}_1 - y_1 \\ \tilde{y}_2 - y_2 \end{pmatrix} - \mathrm{Var}\begin{pmatrix} \hat{y}_1 - y_1 \\ \hat{y}_2 - y_2 \end{pmatrix}$$

be positive semidefinite for all affine unbiased decompositions $(\tilde{y}_1, \tilde{y}_2)$ of $y$. Note that, since $\tilde{y}_2 - y_2 = -(\tilde{y}_1 - y_1)$ and $\hat{y}_2 - y_2 = -(\hat{y}_1 - y_1)$, an equivalent condition is that

$$\mathrm{Var}(\tilde{y}_1 - y_1) - \mathrm{Var}(\hat{y}_1 - y_1)$$

be positive semidefinite for all affine unbiased decompositions $(\tilde{y}_1, \tilde{y}_2)$ of $y$. We will show that, if $y$ is decomposable, there exists a *unique* best affine unbiased decomposition of it.

So far we have not imposed rank conditions on $R$, $X$ nor $V_{ij}$ $(i,j = 1,2)$. It is natural, though, to require that the constraints (2) be consistent, i.e. that $\mathrm{rank}(R, r) = \mathrm{rank}(R)$. More generally, the constraints need to be consistent with the model specification (1). It is therefore necessary (and sufficient) to require that

$$\begin{pmatrix} y \\ r \end{pmatrix} \in \mathscr{M}\begin{pmatrix} X & V \\ R & 0 \end{pmatrix} \qquad \text{a.s.,} \tag{3}$$

where $V = V_{11} + V_{12} + V_{21} + V_{22} = \sigma^{-2}E(uu')$ and $\mathscr{M}(A)$ denotes the column space of the matrix $A$. For conciseness, we shall use the triplet $(y, X_1\beta_1 + X_2\beta_2, \sigma^2(V_{11} + V_{12} + V_{21} + V_{22}))$ to denote the composite linear regression model (1) together with the observability assumptions, and say that it is consistent with the linear constraints $R\beta = r$ if (3) holds. In the absence of linear constraints, the linear regression model (1) is consistent if $y \in \mathscr{M}(X : V)$ a.s.

## 3. Existence of an affine unbiased decomposition

The following proposition gives a necessary and sufficient condition for $y$ to be decomposable. The proofs of this and the following propositions are to a large extent inspired by the constructive methods of proof of Magnus and Neudecker [6, Chapter 13].

For any matrix $Z$ and positive semidefinite (hence symmetric) matrix $W$ such that $\mathcal{M}(Z)\subset\mathcal{M}(W)$, define the orthogonal projector $P_Z = ZZ^+$ and the oblique projector $P_Z^W = Z(Z'W^+Z)^+Z'W^+$, and let $M_Z = I - P_Z$ and $M_Z^W = I - P_Z^W$. For any vector $a$, $P_Z a = P_Z^W a = a$ if and only if $a \in \mathcal{M}(Z)$.

**Proposition 3.1.** *Let the composite linear regression model* $(y, X_1\beta_1 + X_2\beta_2, \sigma^2(V_{11} + V_{12} + V_{21} + V_{22}))$ *be consistent with the linear constraints* $R\beta = r$. *Then,* $y$ *is decomposable if and only if*

$$\mathcal{M}\begin{pmatrix} X_1' \\ 0 \end{pmatrix} \subset \mathcal{M}(X' : R'), \tag{4}$$

*where the matrix of zeroes has the same order as* $X_2'$.

**Proof.** The unbiasedness requirement for the affine decomposition

$$\hat{y}_1 = a + Ay, \qquad \hat{y}_2 = y - \hat{y}_1,$$

is $E(\hat{y}_1 - y_1) = 0$ or, equivalently,

$$a + AX\beta - X_1\beta_1 = 0 \qquad \text{for all } \beta \text{ such that } R\beta = r. \tag{5}$$

Solving $\beta$ from $R\beta = r$ yields $\beta = R^+r + M_{R'}q$ where $q$ is an arbitrary $k$ vector. Hence (5) is equivalent to

$$a + [AX - (X_1 : 0)](R^+r + M_{R'}q) = 0 \qquad \text{for all } q,$$

and, in turn, to

$$a + [AX - (X_1 : 0)]R^+r = 0, \qquad [AX - (X_1 : 0)]M_{R'} = 0.$$

This pair of equations has a solution in $a$ and $A$ if and only if the latter equation has a solution in $A$. This will be the case if and only if we can ensure that, for some matrix $B$,

$$AX - (X_1 : 0) = BR.$$

Thus, a necessary and sufficient condition for $y$ to be decomposable is that the rows of $(X_1 : 0)$ be linear combinations of the rows of $X$ and $R$. □

The condition for $y$ to be decomposable is equivalent to the condition that $X_1\beta_1$ be estimable, in the sense that an affine unbiased estimator of $X_1\beta_1$ has to exist. See Magnus and Neudecker [6, Proposition 13.3]. When $R = 0$, $y$ is decomposable if and only if

$$\text{rank}(X) = \text{rank}(X_1) + \text{rank}(X_2). \tag{6}$$

To see this, note that $R = 0$ implies that $y$ is decomposable if and only if

$$\mathcal{M}\begin{pmatrix} X_1' \\ 0 \end{pmatrix} \subset \mathcal{M}\begin{pmatrix} X_1' \\ X_2' \end{pmatrix}. \tag{7}$$

Clearly,

$$\mathcal{M}\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \subset \mathcal{M}\begin{pmatrix} X_1' & X_1' \\ X_2' & 0 \end{pmatrix}.$$

Hence (7) is equivalent to

$$\mathcal{M}\begin{pmatrix} X_1' & X_1' \\ X_2' & 0 \end{pmatrix} = \mathcal{M}\begin{pmatrix} X_1' \\ X_2' \end{pmatrix}.$$

Thus,

$$\begin{aligned} \operatorname{rank}(X') &= \operatorname{rank}\begin{pmatrix} X_1' & X_1' \\ X_2' & 0 \end{pmatrix} = \operatorname{rank}\begin{pmatrix} X_1' - X_1' & X_1' \\ X_2' & 0 \end{pmatrix} \\ &= \operatorname{rank}\begin{pmatrix} 0 & X_1' \\ X_2' & 0 \end{pmatrix} = \operatorname{rank}(X_1') + \operatorname{rank}(X_2'). \end{aligned}$$

When $R \neq 0$, (6) is a sufficient condition for $y$ to be decomposable. To see this, note that

$$\mathcal{M}\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \subset \mathcal{M}\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix}.$$

Hence (6) implies

$$\mathcal{M}\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} = \mathcal{M}\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix}.$$

Moreover,

$$\mathcal{M}\begin{pmatrix} X_1' \\ 0 \end{pmatrix} \subset \mathcal{M}\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix}.$$

Thus,

$$\mathcal{M}\begin{pmatrix} X_1' \\ 0 \end{pmatrix} \subset \mathcal{M}\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} = \mathcal{M}(X') \subset \mathcal{M}(X' : R').$$

Finally, note that the properties of the matrices $V_{ij}$ $(i,j = 1,2)$ do not matter for the existence of an affine unbiased decomposition.

## 4. Uniqueness of the best affine unbiased decomposition

We establish below the existence and uniqueness of a best affine unbiased decomposition of $y$ if it is decomposable. The following lemma will be useful.[1]

**Lemma 4.1.** *Let $V_{ij}$ $(i,j=1,2)$ be $n\times n$ matrices such that the matrix*

$$U=\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

*is positive semidefinite. Then*

$$\mathcal{M}(V_{11}+V_{21})\subset\mathcal{M}(V_{11}+V_{12}+V_{21}+V_{22}).$$

**Proof.** We have

$$V=V_{11}+V_{12}+V_{21}+V_{22}=(I:I)U\begin{pmatrix} I \\ I \end{pmatrix}$$

and consequently

$$\mathcal{M}(V)=\mathcal{M}[(I:I)U]=\mathcal{M}(V_{11}+V_{21}:V_{12}+V_{22})\supset\mathcal{M}(V_{11}+V_{21}),$$

using the identity $\mathcal{M}(ABA')=\mathcal{M}(AB)$ when $B$ is positive semidefinite. □

We first consider the decomposition of $y$ in the special case where no linear constraints on $\beta$ are given.

**Proposition 4.1.** *Let the composite linear regression model $(y, X_1\beta_1+X_2\beta_2, \sigma^2(V_{11}+V_{12}+V_{21}+V_{22}))$ be consistent, and let $W=V+XX'$, where $V=V_{11}+V_{12}+V_{21}+V_{22}$ and $X=(X_1:X_2)$. Then, if $y$ is decomposable, there exists a unique best affine unbiased decomposition. It is given by*

$$\hat{y}_1=Ay, \qquad \hat{y}_2=y-\hat{y}_1,$$

*where*

$$A=(X_1:0)(X'W^+X)^+X'W^+ + (V_{11}+V_{12})W^+M_X^W. \tag{8}$$

**Proof.** From the proof of Proposition 3.1 we retain the fact that, in the case where $R=0$ and $r=0$, the affine decomposition $(\hat{y}_1,\hat{y}_2)=(a+Ay, y-a-Ay)$ is unbiased if and only if $a=0$ and $AX=(X_1:0)$. We shall first find the (unique) minimum-trace affine unbiased decomposition of $y$ by minimizing $\frac{1}{2}\operatorname{tr}\operatorname{Var}(\hat{y}_1-y_1)$ subject to $a=0$ and $AX=(X_1:0)$, and then show that the unique minimum-trace affine unbiased decomposition is also the unique best affine unbiased decomposition

[1] We owe the formulation of this lemma and its proof to one of the referees.

of $y$. We have

$$\begin{aligned}\mathrm{Var}(\hat{y}_1 - y_1) &= \mathrm{Var}[A(u_1 + u_2) - u_1] \\ &= \sigma^2[AVA' - A(V_{11} + V_{21}) - (V_{11} + V_{12})A' + V_{11}].\end{aligned}$$

Define the Lagrangian function $\Lambda$ by

$$\begin{aligned}\Lambda(A) =& \tfrac{1}{2}\,\mathrm{tr}[AVA' - A(V_{11} + V_{21}) - (V_{11} + V_{12})A' + V_{11}] \\ &- \mathrm{tr}L'[AX - (X_1 : 0)],\end{aligned}$$

where $L$ is a matrix of Lagrange multipliers. Differentiating $\Lambda$ with respect to $A$ gives

$$d\Lambda = \mathrm{tr}VA'(dA) - \mathrm{tr}(V_{11} + V_{21})(dA) - \mathrm{tr}XL'(dA).$$

The first-order conditions for a constrained minimum of $\Lambda$ are

$$VA' - (V_{11} + V_{21}) - XL' = 0,$$

$$AX - (X_1 : 0) = 0,$$

or, in matrix form,

$$\begin{pmatrix} V & X \\ X' & 0 \end{pmatrix}\begin{pmatrix} A' \\ -L' \end{pmatrix} = \begin{pmatrix} V_{11} + V_{21} \\ (X_1 : 0)' \end{pmatrix}.$$

By Theorem 3.23 of Magnus and Neudecker [6], this matrix equation has a solution in $A$ and $L$ if and only if

$$\mathcal{M}(V_{11} + V_{21}) \subset \mathcal{M}(V : X) \qquad \text{and} \qquad \mathcal{M}\begin{pmatrix} X_1' \\ 0 \end{pmatrix} \subset \mathcal{M}(X').$$

By Lemma 4.1, the first of these conditions is always satisfied, and the second one is satisfied by the assumption that $y$ is decomposable. The general solution for $A$ is

$$A = (X_1 : 0)(X'W^+X)^+X'W^+ + (V_{11} + V_{12})W^+M_X^W + QM_W, \tag{9}$$

where $Q$ is an arbitrary matrix of appropriate order. Since $M_W y = 0$ a.s., it follows that the minimum-trace affine unbiased decomposition of $y$ is unique and is given by $(Ay, y - Ay)$ with $A$ as in (8). Note that $M_W X = (I - WW^+)X = 0$, because $\mathcal{M}(X) \subset \mathcal{M}(W)$, and that $W^+M_X^W X = W^+X - W^+X(X'W^+X)^+X'W^+X = 0$, because $C^{1/2}X - C^{1/2}X(X'CX)^+X'CX = C^{1/2}X - C^{1/2}X = 0$ for $C$ positive semidefinite. Therefore, as required, $AX = (X_1 : 0)(X'W^+X)^+X'W^+X = (X_1 : 0)X^+X = (X_1 : 0)$, by Theorem 3.20(iii) of Magnus and Neudecker [6] and the assumption that $y$ is decomposable. To show that $(Ay, y - Ay)$ is also a best affine unbiased decomposition of $y$, consider the problem of minimizing $\mathrm{tr}\,\mathrm{Var}[c'(\hat{y}_1 - y_1)]$ subject to $a = 0$ and $AX = (X_1 : 0)$ for an arbitrary $n$ vector $c$. Now,

$$\begin{aligned}\mathrm{Var}[c'(\hat{y}_1 - y_1)] &= c'[\mathrm{Var}(\hat{y}_1 - y_1)]c \\ &= \sigma^2 c'[AVA' - A(V_{11} + V_{21}) - (V_{11} + V_{12})A' + V_{11}]c.\end{aligned}$$

The solution of this constrained minimization problem is $c'\hat{y}_1 = c'Ay$ with $A$ as in (8), implying that

$$c'[\mathrm{Var}(\tilde{y}_1 - y_1)]c \geqslant c'[\mathrm{Var}(\hat{y}_1 - y_1)]c$$

for all affine unbiased decompositions $(\tilde{y}_1, \tilde{y}_2)$ of $y$. Since $c$ is arbitrary, $\mathrm{Var}(\tilde{y}_1 - y_1) - \mathrm{Var}(\hat{y}_1 - y_1)$ is positive semidefinite and thus $(\hat{y}_1, y - \hat{y}_1)$ is a best affine unbiased decomposition of $y$. Now, any best affine unbiased decomposition is also a minimum-trace affine unbiased decomposition, since for any matrices $B$ and $C$, if $B - C$ is positive semidefinite, then $\mathrm{tr}\, B \geqslant \mathrm{tr}\, C$. Therefore, the uniqueness of the minimum-trace affine unbiased decomposition implies the uniqueness of the best affine unbiased decomposition. □

When no linear constraints on $\beta$ are given, the best affine unbiased decomposition of $y$ turns out to be linear. We now consider the general case.

**Proposition 4.2.** *Let the composite linear regression model* $(y, X_1\beta_1 + X_2\beta_2, \sigma^2(V_{11} + V_{12} + V_{21} + V_{22}))$ *be consistent with the linear constraints* $R\beta = r$, *and let*

$$Z = \begin{pmatrix} X \\ R \end{pmatrix}, \qquad v = \begin{pmatrix} y \\ r \end{pmatrix}, \qquad W = \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} + ZZ',$$

*where* $V = V_{11} + V_{12} + V_{21} + V_{22}$ *and* $X = (X_1 : X_2)$. *Then, if* $y$ *is decomposable, there exists a unique best affine unbiased decomposition. It is given by*

$$\hat{y}_1 = Av, \qquad \hat{y}_2 = y - \hat{y}_1,$$

*where*

$$A = (X_1 : 0)(Z'W^+Z)^+Z'W^+ + (V_{11} + V_{12} : 0)W^+M_Z^W.$$

**Proof.** Partition $R$ into $(R_1 : R_2)$ conformably with $X = (X_1 : X_2)$, and let $r_i = R_i\beta_i$ and

$$Z_i = \begin{pmatrix} X_i \\ R_i \end{pmatrix}, \qquad v_i = \begin{pmatrix} y_i \\ r_i \end{pmatrix}, \qquad U_{ij} = \begin{pmatrix} V_{ij} & 0 \\ 0 & 0 \end{pmatrix}.$$

Then, $(y, X_1\beta_1 + X_2\beta_2, \sigma^2(V_{11} + V_{12} + V_{21} + V_{22}))$, together with $R\beta = r$, is equivalent to $(v, Z_1\beta_1 + Z_2\beta_2, \sigma^2(U_{11} + U_{12} + U_{21} + U_{22}))$. Moreover, the first model is consistent with the constraints if and only if the latter model is consistent, and $y$ is decomposable if and only if $v$ is decomposable. Proposition 4.1 yields the unique best affine unbiased decomposition of $v$ and, as appropriate subvectors thereof, that of $y$. □

With linear constraints on $\beta$, the best affine unbiased decomposition of $y$ is in general affine, not linear. It can also be written as

$$\hat{y}_1 = \widehat{X_1\beta_1} + (V_{11} + V_{12})(W^+)_{11}\hat{u}, \qquad \hat{y}_2 = \widehat{X_2\beta_2} + (V_{22} + V_{21})(W^+)_{11}\hat{u},$$

where $\widehat{X_1\beta_1}$ and $\widehat{X_2\beta_2}$ are the best affine unbiased estimators of $X_1\beta_1$ and $X_2\beta_2$, respectively, $\hat{u} = y - \widehat{X_1\beta_1} - \widehat{X_2\beta_2}$, and $(W^+)_{11}$ is the leading $n \times n$ submatrix of $W^+$.

## 5. Generalizations

It is easy to see that if $(\hat{y}_1, \hat{y}_2)$ is the best affine unbiased decomposition of $y$, then, for arbitrary scalars $c_1$ and $c_2$, the best affine unbiased prediction of $(c_1y_1, c_2y_2)$ is $(c_1\hat{y}_1, c_2\hat{y}_2)$. This leads to the following generalization. Suppose we do not observe $y_1 + y_2$, but instead the linear combination $y = c_1y_1 + c_2y_2$, where $c_1$ and $c_2$ are known and non-zero. The methods of the previous section yield the unique best affine unbiased decomposition $(\widehat{c_1y_1}, \widehat{c_2y_2})$ of $y$ under the necessary and sufficient condition that $c_1X_1\beta_1$ be estimable. Here, however, we are interested in an affine decomposition $(\hat{y}_1, \hat{y}_2) = (a_1 + A_1y, a_2 + A_2y)$ such that $c_1\hat{y}_1 + c_2\hat{y}_2 = y$. Unbiasedness is imposed by requiring $E(\hat{y}_1 - y_1) = 0$ and $E(\hat{y}_2 - y_2) = 0$ for all $\beta \in \mathscr{B}$. The best affine unbiased decomposition, say $(\hat{y}_1, \hat{y}_2)$, exists if and only if $X_1\beta_1$ is estimable, in which case it is unique and follows directly from $(\widehat{c_1y_1}, \widehat{c_2y_2}) = (c_1\hat{y}_1, c_2\hat{y}_2)$.

The extension to the $p$-components decomposition of $y$ is straightforward. In obvious notation, $y$ is decomposable if and only if $X_1\beta_1, \ldots, X_p\beta_p$ are all estimable. Furthermore, the unique best affine unbiased decomposition $(\hat{y}_1, \ldots, \hat{y}_p)$ can be obtained, for example, from the first components of the 2-component best affine unbiased decompositions $(\hat{y}_i, y - \hat{y}_i)$, $i = 1, \ldots, p$.

Consider now the situation where, in addition to $y$, a subset of the elements of $y_1$ and $y_2$ is observable, say $Sy_1$ and $Sy_2$, where $S$ is obtained from the $n \times n$ identity matrix after deleting the rows that correspond to the unobservable elements of $y_1$ and $y_2$. An affine decomposition of $y$ may now be defined as a pair $(\hat{y}_1, \hat{y}_2) = (a + Ay + A_1Sy_1, y - a - Ay - A_1Sy_1)$. The dependence on $Sy_2$ is implicit, through the dependence on $y$ and $Sy_1$. The decomposition is unbiased if $E(\hat{y}_1 - y_1) = 0$ or, equivalently,

$$a + AX\beta + (A_1S - I)(X_1 : 0)\beta = 0 \qquad \text{for all } \beta \in \mathscr{B}.$$

An argument similar to that used in the proof of Proposition 3.1 shows that $y$ is decomposable if and only if

$$\mathscr{M}\begin{pmatrix} X_1' \\ 0 \end{pmatrix} \subset \mathscr{M}\left(\begin{pmatrix} X_1'S' \\ 0 \end{pmatrix} : X' : R'\right).$$

An equivalent condition is that $X_1\beta_1$ be estimable. If $y$ is decomposable, it is straightforward to show that there exists a unique best affine unbiased decomposition, given by

$$\hat{y}_1 = M_{S'}Av + S'Sy_1, \qquad \hat{y}_2 = y - \hat{y}_1,$$

where $A$ and $v$ are defined in Proposition 4.2. Observe that, since $S' = S^+$, we have $S\hat{y}_1 = Sy_1$ and $S\hat{y}_2 = Sy_2$, as it should be.

## 6. Identifiability of the covariance matrices

Applying the best affine unbiased decomposition hinges on knowledge of the covariance matrices of the disturbances. More precisely, it requires $V_{11} + V_{12}$ and $V_{11} + V_{12} + V_{21} + V_{22}$, apart from a common multiplicative scalar. Such knowledge cannot be (fully) extracted from the available observations, and hence should (partly) come from extraneous sources. This problem is similar to the problem encountered by best affine unbiased parameter estimation, i.e. by the method of generalized least squares, which requires knowledge of the covariance matrix of the vector of disturbances. To be able to apply the method of generalized least squares, one has to put some structure on the disturbance covariance matrix in order to reduce its number of unknowns to a smaller number of identifiable parameters. Something similar applies to the best affine unbiased decomposition, although here the problem is more severe since we have two disturbances with unknown covariance structures instead of a single one. The following examples show that in some cases the covariance matrices can be inferred from the data and the structure imposed on them.

The least-squares residuals $\hat{u} = M_X y$ have second moment

$$E(\hat{u}\hat{u}') = \sigma^2 M_X (V_{11} + V_{12} + V_{21} + V_{22}) M_X. \tag{10}$$

Suppose now that the matrices $\sigma^2 V_{ij}$ $(i,j = 1,2)$ are known functions of non-stochastic observables (perhaps other than $X$) and possibly unknown parameters (perhaps other than $\beta$ and $\sigma^2$). Then, in some cases the unknown parameters, and hence $\sigma^2 V_{ij}$ $(i,j = 1,2)$, are identified from (10). A simple example is the heteroskedastic model

$$\sigma^2 V_{ij} = \sigma_{ij}\, \mathrm{diag}|X_i\beta_i|\, \mathrm{diag}|X_j\beta_j|, \qquad i,j = 1,2,$$

where $\mathrm{diag}|X_i\beta_i|$ is a diagonal matrix with the absolute values of $X_i\beta_i$ on the diagonal. Then, the matrices $\sigma^2 V_{ij}$ $(i,j = 1,2)$ are identified from (10) if and only if $M_X \neq 0$ and $X_1\beta_1$ is estimable. Estimation can proceed in a variety of ways, for example by minimizing with respect to $\sigma_{ij}$ $(i,j = 1,2)$ the $L_1$ or $L_2$ norm of

$$\hat{u}\hat{u}' - M_X \left( \sum_{i,j=1,2} \sigma_{ij}\, \mathrm{diag}|\widehat{X_i\beta_i}|\, \mathrm{diag}|\widehat{X_j\beta_j}| \right) M_X$$

or of its diagonal, where $\widehat{X_1\beta_1}$ and $\widehat{X_2\beta_2}$ are first-step estimates of $X_1\beta_1$ and $X_2\beta_2$. The above example generates many others. Observe that, in order to achieve identifiability, it is crucial that the matrices $\sigma^2 V_{ij}$ $(i,j = 1,2)$ depend on non-constant observables.

Another structure that is identified from (10) is as follows. Let

$$\sigma^2 V_{11} = \sigma_{11} \begin{pmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_G \end{pmatrix}, \qquad \sigma^2 V_{22} = \sigma_{22} I,$$

$$V_{12} = V_{21} = 0,$$

where $J_g$ $(g = 1, \dots, G)$ is an $n_g \times n_g$ matrix of ones (with $n_g$ known) and $n_1 + \dots + n_G = n$. The following interpretation can be given. Suppose $X_1$ and $u_1$ are environmental regressors and disturbances, while $X_2$ and $u_2$ have some other interpretation, unrelated to environment. It is assumed that $u_2$ is homoskedastic and uncorrelated across observations, while for the $n_g$ observations within any group $g$, the environment is kept constant. This implies that the rows of $(X_1 : u_1)$ corresponding to observations that belong to the same group are identical. Assuming homoskedasticity, the above structure of $\sigma^2 V_{11}$ follows. It can then be seen that $\sigma^2 V_{11}$ and $\sigma^2 V_{22}$ are identified from (10), unless $M_X = 0$ or $n_g = 1$ for all $g$.

We note that the covariance structure $\sigma^2 V_{ij} = \sigma_{ij} I$ $(i, j = 1, 2)$ is not identified from (10) nor from any other device that is solely based on the observable data $y$, $X_1$ and $X_2$, even when $\sigma_{12}$ is known to be zero. For identifiability, it is necessary (and sufficient) to observe also at least one element of $y_1$ (and hence of $y_2$).

It should be mentioned that replacing $V_{ij}$ $(i, j = 1, 2)$ with estimates in formulae for the best affine unbiased decomposition yields a decomposition of $y$ which is, in general, neither affine nor unbiased.

Finally, when insufficient extraneous information is available to put forward a realistic and identifiable disturbance covariance structure, we may consider the decomposition

$$\left( \widehat{X_1 \beta_1} + \frac{1}{2}\hat{u}, \widehat{X_2 \beta_2} + \frac{1}{2}\hat{u} \right),$$

which is affine unbiased anyway, but not necessarily best affine unbiased.

## 7. Conclusion

We have proposed a method for decomposing the response vector in a linear model into two or more additive components, each of which is related to a specific set of regressors and a specific disturbance term. We have accounted for arbitrary regressor matrices, covariance matrices of the disturbance terms, and linear constraints on the parameters. In this setting, a necessary and sufficient condition for the existence of a sensible, i.e. unbiased, decomposition has been given. Furthermore, the existence and uniqueness of the best affine unbiased decomposition has been proven, and an expression has been given to calculate it.

## Acknowledgments

We are grateful to Richard William Farebrother, Jan Magnus, two anonymous referees and the handling editor for helpful comments. This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The first author acknowledges financial support from the Flemish Fund for Scientific Research (Grant G.0366.01).

## References

[1] A.P.J. Abrahamse, J. Koerts, New estimators of disturbances in regression analysis, J. Amer. Statist. Assoc. 66 (1971) 71–74.

[2] A.C. Aitken, On least squares and linear combinations of observations, Proc. Roy. Soc. Edinburgh 55 (1935) 42–48.

[3] R.F. Farebrother, Fitting Linear Relationships: A History of the Calculus of Observations 1750–1900, Springer, New York, 1999.

[4] C.F. Gauss, Theoria Combinationis Observationum Erroribus Minimis Obnoxiae: Pars Prior, Pars Posterior, Supplementum, Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores 5 (1823) 33–90, and 6 (1828), 57–93. Reprinted with English translation by G.W. Stewart, Theory of the Combination of Observations Least Subject to Errors. Part One, Part Two, Supplement, Classics Appl. Math. 11, Siam, Philadelphia, 1995.

[5] A.S. Goldberger, Best linear unbiased prediction in the generalized linear regression model, J. Amer. Statist. Assoc. 57 (1962) 369–375.

[6] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, rev. ed., Wiley, Chichester, 1999.

[7] H. Neudecker, Abrahamse and Koerts' 'New Estimator' of disturbances in regression analysis, J. Econometrics 5 (1977) 129–133.

[8] C.R. Rao, Unified theory of linear estimation, Sankhyā 33 (1971) 371–394.

[9] E. Schokkaert, G. Dhaene, C. Van de Voorde, Risk adjustment and the trade-off between efficiency and risk selection: an application of the theory of fair compensation, Health Econom. 7 (1998) 465–480.

[10] S.M. Stigler, The History of Statistics: The Measurement of Uncertainty before 1900, Harvard University Press, Cambridge, MA, 1986.

[11] H. Theil, The analysis of disturbances in regression analysis, J. Amer. Statist. Assoc. 60 (1965) 1067–1079.