Best Affine Unbiased Response Decomposition*

Geert Dhaene
Universiteit Gent

Erik Schokkaert
Katholieke Universiteit Leuven

Carine Van de Voorde
Katholieke Universiteit Leuven

March 28, 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 reponse vectors y_1 and y_2 are unobserved but the sum $y = y_1 + y_2$ is observed, 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 explicit expression for it.

Keywords: Decomposition, best affine unbiasedness, Gauss-Markov theory.
JEL classification: C13, C53.

*We are grateful to Jan Magnus for helpful comments. Address correspondence to: Geert Dhaene, Universiteit Gent, Faculteit Economie en Bedrijfskunde, Hoveniersberg 24, B-9000 Gent, Belgium. Tel.: (+32) 9 2648978. Fax: (+32) 9 2648995. E-mail: geert.dhaene@rug.ac.be.

1 Introduction

The idea of best linear unbiasedness in parameter estimation, response prediction, and estimation of disturbances has a long history in statistics. The theory of best linear unbiased estimation by the method of least squares originated with Gauss (1821–1823). See Plackett (1949) for historical details. Subsequent contributions were made by Aitken (1934) and Rao (1971), inter alia. Goldberger (1962) developed a theory of best linear unbiased prediction in the linear model. Theil (1965) and subsequent authors considered best linear unbiased estimation of the disturbances, subject to constraints on the covariance matrix of the estimators.

This paper applies 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 formal problem is as follows. Suppose we have two linear models with unobserved response vectors, but where the sum of the two response vectors is observed. The question is then, how to retrieve the original response vectors.

We have in mind applications in the social sciences like, for example, the decomposition of school results of pupils into an environmental component on the one hand, and a component determined by innate characteristics of the pupil on the other hand. For an application in health economics, see Schokkaert, Dhaene and Van de Voorde (1998). 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, there exists a unique best affine unbiased decomposition. The extension to the decomposition into p components is also given in Section 4. Section 5 concludes.

2 Statement of the problem

In 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 $E(u_i u_j') = V_{ij}$ $(i, j = 1, 2)$, 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 that the matrices V_{ij} $(i, j = 1, 2)$ are known. On the other hand, β is unknown, although extraneous information may be present in the form

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

where the $m \times k$ matrix R and the $m \times 1$ vector r are known and non-stochastic. Thus the parameter space of β, denoted $\mathcal{B}$, is the Euclidean space $\mathbb{R}^k$, or the affine subspace of $\mathbb{R}^k$ 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

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

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 is given in the next section. If y is decomposable, it is natural to look for the *best* affine unbiased decomposition, say $(\hat{y}_1, \hat{y}_2)$, defined by the property that

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

is positive semidefinite for all affine unbiased decompositions $(\tilde{y}_1, \tilde{y}_2)$ of y. Note that an equivalent condition is that

$$\operatorname{tr}\operatorname{Var}(\tilde{y}_1 - y_1) \geq \operatorname{tr}\operatorname{Var}(\hat{y}_1 - y_1)$$

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

We have not imposed rank conditions on R, X nor V_{ij} $(i, j = 1, 2)$. Nevertheless, the constraints (2) have to be consistent with the model specification (1). It is therefore necessary (and sufficient) to require that

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

where $V = V_{11} + V_{12} + V_{21} + V_{22} = E(uu')$ and $\mathcal{S}(A)$ denotes the column space of the matrix A. For conciseness, we shall use the triplet $(y, X_1\beta_1 + X_2\beta_2, V_{11} + V_{12} + V_{21} + V_{22})$ 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 $y \in \mathcal{S}(X : V)$ a.s.

3 Existence of an affine unbiased decomposition

The following proposition gives a necessary and sufficient condition for y to be decomposable. The proof of this and the following propositions are to a large extent inspired by the constructive methods of proof of Magnus and Neudecker (1999, Chapter 13).

Proposition 1 *Let the composite linear regression model $(y, X_1\beta_1 + X_2\beta_2, V_{11} + V_{12} + V_{21} + V_{22})$ be consistent with the linear constraints $R\beta = r$. Then, y is decomposable if and only if*

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

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 β from $R\beta = r$ yields $\beta = R^+r + (I - R^+R)q$ where q is an arbitrary $k \times 1$ vector. Hence (5) is equivalent to

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

and, in turn, to

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

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 $(X_1 : 0)$ are linear combinations of the rows of X and R. *(Q.E.D.)*

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. Note that $\mathrm{rank}(X) = \mathrm{rank}(X_1) + \mathrm{rank}(X_2)$ is a sufficient condition for y to be decomposable. It is, moreover, a necessary condition if $R = 0$ and $r = 0$. 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

The following propositions show that, if y is decomposable, the best affine unbiased decomposition of y exists and is unique. We first prove a lemma that will be needed.

Lemma 1 *Let u_1 and u_2 be $n \times 1$ vectors satisfying $E(u_i) = 0$ and $E(u_i u_j') = V_{ij}$ $(i, j = 1, 2)$. Then*

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

Proof. We can always write $u_1 = Ax$ and $u_2 = Bx$ for some vector x with properties $E(x) = 0$ and $E(xx') = I$, and some matrices A and B satisfying $AA' = V_{11}$ and $BB' = V_{22}$. Then,

$$V_{11} + V_{22} + V_{21} + V_{12} = (A + B)(A + B)',$$
$$(V_{11} + V_{21})(V_{11} + V_{21})' = (A + B)A'A(A + B)'.$$

We see that, if $(V_{11} + V_{22} + V_{21} + V_{12})z = 0$ for some vector z, then also $(V_{11} + V_{21})(V_{11} + V_{21})'z = 0$. Hence

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

The proof is complete by noting that $\mathcal{S}(V_{11} + V_{21}) = \mathcal{S}(V_{11} + V_{21})(V_{11} + V_{21})'$. *(Q.E.D.)*

We now establish the uniqueness of the best affine unbiased response decomposition of y in the special case where no linear constraints on β are given.

Proposition 2 *Let the composite linear regression model $(y, X_1\beta_1 + X_2\beta_2, V_{11} + V_{12} + V_{21} + V_{22})$ 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, the best affine unbiased decomposition of y exists and is given by*

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

where

$$A = (X_1 : 0)(X'W^+X)^+X'W^+ + (V_{11} + V_{12})W^+[I - X(X'W^+X)^+X'W^+].$$

Proof. From the proof of Proposition 1 we retain 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 $AX = (X_1 : 0)$. The best affine unbiased decomposition of y is found by minimizing $\frac{1}{2} \operatorname{tr} \operatorname{Var}(\hat{y}_1 - y_1)$ subject to $AX = (X_1 : 0)$. Now,

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

Define the Lagrangian function Λ by

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

where L is a matrix of Lagrange multipliers. Differentiating Λ with respect to A gives

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

The first order conditions for a constrained minimum are

$$\begin{aligned} VA' - (V_{11} + V_{21}) - XL' &= 0 \\ AX - (X_1 : 0) &= 0, \end{aligned}$$

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

According to Magnus and Neudecker (1999, Theorem 3.23), this matrix equation has a solution in A and L if and only if

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

By Lemma 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

$$A = (X_1 : 0)(X'W^+X)^+X'W^+ + (V_{11}+V_{12})W^+[I - X(X'W^+X)^+X'W^+] + Q(I - WW^+),$$

where Q is an arbitrary matrix of appropriate order. The proof is complete by noting that $WW^+y = y$ a.s. *(Q.E.D.)*

The following proposition considers the general case.

Proposition 3 *Let the composite linear regression model* $(y, X_1\beta_1 + X_2\beta_2, V_{11} + V_{12} + V_{21} + V_{22})$ *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, the best affine unbiased decomposition of y exists and is given by

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

where

$$A = (X_1 : 0)(Z'W^+Z)^+Z'W^+ + (V_{11} + V_{12} : 0)W^+[I - Z(Z'W^+Z)^+Z'W^+].$$

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, $(y, X_1\beta_1 + X_2\beta_2, V_{11} + V_{12} + V_{21} + V_{22})$, together with $R\beta = r$, is equivalent to $(v, Z_1\beta_1 + Z_2\beta_2, U_{11} + U_{12} + U_{21} + U_{22})$. 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 2 yields the unique best affine unbiased decomposition of v and, as appropriate subvectors thereof, that of y. *(Q.E.D.)*

The best affine unbiased decomposition of y can also be written as

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

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, and $\hat{u} = y - \widehat{X_1\beta_1} - \widehat{X_2\beta_2}$. Finally, we note that

$$\begin{aligned} \operatorname{Var}\begin{pmatrix} \hat{y}_1 - y_1 \\ \hat{y}_2 - y_2 \end{pmatrix} &= \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}[(Z'W^+Z)^+ - I]\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \\ &\quad + QW^+[I - Z(Z'W^+Z)^+Z']W^+Q', \end{aligned}$$

where

$$Q = \begin{pmatrix} V_{11} + V_{12} & 0 \\ 0 & V_{22} + V_{21} \end{pmatrix}.$$

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 $(\hat{y}_i, y - \hat{y}_i)$, $i = 1, \ldots, p$.

Note that applying the best affine unbiased decomposition hinges on knowledge of the covariance matrices of the disturbances. Such a knowledge cannot be extracted from the available observations, and hence should come from extraneous sources.[1] When no

[1]This problem is, of course, similar to the problem faced 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.

such sources are available and we are completely ignorant, we might opt to assume that all disturbances are uncorrelated and have equal variances, yielding the p-component decomposition

$$\left(\widehat{X_1\beta_1} + \frac{1}{p}\,\hat{u}, \ldots, \widehat{X_p\beta_p} + \frac{1}{p}\,\hat{u}\right),$$

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

5 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 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 was given. Furthermore, the existence and uniqueness of the best affine unbiased decomposition was proven, and an explicit expression to calculate it was given.

References

[1] Aitken, A. C. (1935). "On Least Squares and Linear Combinations of Observations", *Proceedings of the Royal Society of Edinburgh*, **55**, 42–48.

[2] Gauss, K. F. (1821–23). *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae. Pars Prior. Pars Posterior. Supplementum.* Reprinted with English translation by G. W. Stewart, *Theory of the Combination of Observations Least Subject to Errors. Part One, Part Two, Supplement*, Classics in Applied Mathematics Series, **11**, Philadelphia, Society for Industrial and Applied Mathematics, 1995.

[3] Goldberger, A. S. (1962). "Best Linear Unbiased Prediction in the Generalized Linear Regression Model", *Journal of the American Statistical Association*, **57**, 369–375.

[4] Magnus, J. R. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edition, Chichester, Wiley.

[5] Plackett, R. L. (1949). "A Historical Note on the Method of Least Squares", *Biometrika*, **36**, 458–460.

[6] Rao, C. R. (1971). "Unified Theory of Linear Estimation", *Sankhyā*, **33**, 371–394.

[7] Schokkaert, E., G. Dhaene and C. Van de Voorde (1998). "Risk Adjustment and the Trade-Off between Efficiency and Risk Selection: an Application of the Theory of Fair Compensation", *Health Economics*, **7**, 465–480.

[8] Theil, H. (1965). "The Analysis of Disturbances in Regression Analysis", *Journal of the American Statistical Association*, **60**, 1067–1079.