Estimating Variance Components in Linear Models

FRIEDRICH PUKELSHEIM

Institut für Mathematische Stochastik der Universität Freiburg im Breisgau, Germany

Estimation of variance components in linear model theory is presented as an application of estimation of the mean by introducing a dispersion–mean correspondence. Without any further computations, this yields most general representations of minimum variance–minimum bias–invariant quadratic estimates, estimates from MINQUE theory, and Ridge-type estimates of the variance components.

1. INTRODUCTION

This note aims to emphasize that in linear model theory linear estimation of the mean and quadratic estimation of the variance components pose the same problem when taking a suitable point of view (Section 2). This *dispersion–mean correspondence* yields most general representations of estimates of the variance components (Section 3) for both minimum variance unbiased estimation (Theorem 1) and MINQUE theory (Theorem 2), at the same time exhibiting when the estimate of one procedure is optimal in the sense of the other. Finally, the approach suggested here is used to derive Ridge-type estimates of the variance components. All results follow from the theory of mean estimation, no further computations being necessary.

The present paper extends the works of Mitra [2] and Seely [6]: no rank assumptions are made, unbiasedness is replaced by the more general concept of minimum bias, and the close relatedness to multilinear algebra is stressed.

2. THE DISPERSION–MEAN CORRESPONDENCE

Let a linear model be characterized by linear decompositions of the expectation vector $\mathscr{E}\mathbf{Y}$ and the dispersion matrix $\mathscr{D}\mathbf{Y}$:

$$\mathscr{E}\mathbf{Y} = \mathbf{Xb} = \sum_{\pi=1}^{p} b_\pi \mathbf{x}_\pi , \qquad \mathscr{D}\mathbf{Y} = \sum_{\kappa=1}^{k} t_\kappa \mathbf{V}_\kappa , \tag{1}$$

where $\mathbf{Y}$ is a $\mathbb{R}^n$-valued random vector, $\mathbf{X} = (\mathbf{x}_1 \vdots \cdots \vdots \mathbf{x}_p)$, $\mathbf{x}_\pi(\mathbf{V}_\kappa)$ are known $\mathbb{R}^n$-vectors (symmetric (n, n)-matrices), and $\mathbf{b} = (b_1, \ldots, b_p)'$ and $\mathbf{t} = (t_1, \ldots, t_k)'$ are to be estimated.

Quadratic estimates $Q(\mathbf{Y})$ of $\mathbf{t}$ are, by definition, derived from bilinear functions $B(\cdot, \cdot)$ from $\mathbb{R}^n \times \mathbb{R}^n$ into $\mathbb{R}^k$ by setting both arguments equal to $\mathbf{Y}$. Since the Kronecker product $\mathbf{x} \otimes \mathbf{y}$ [5, p. 29] is a tensorproduct [1, p. 12], any quadratic estimate factorizes according to $Q(\mathbf{Y}) = B(\mathbf{Y}, \mathbf{Y}) = \mathbf{L} \cdot \mathbf{Y} \otimes \mathbf{Y}$ with a (k, n^2)-matrix $\mathbf{L}$. Another tensorproduct is $\mathbf{xy}'$, being related to $\mathbf{x} \otimes \mathbf{y}$ by the inner product and tensor product preserving isomorphism vec:

$$(\operatorname{vec} \mathbf{A})' \operatorname{vec} \mathbf{B} = \operatorname{trace} \mathbf{AB}', \quad \operatorname{vec} \mathbf{xy}' = \mathbf{x} \otimes \mathbf{y},$$

where vec $\mathbf{A}$ is the column vector obtained from the matrix $\mathbf{A}$ by ordering its entries lexicographically. Further, estimation of $\mathbf{t}$ is restricted to estimates which are *invariant* under all mean translations $\mathbf{y} \to \mathbf{y} + \mathbf{Xb}$, $\mathbf{b} \in \mathbb{R}^p$. A maximal invariant statistic with respect to these translations is $\mathbf{MY}$, where $\mathbf{M} = \operatorname{Proj}(\mathscr{R}\mathbf{X})^\perp$ is the orthogonal projector onto the orthogonal complement of the range (column space) of $\mathbf{X}$. Hence, an estimate $Q(\mathbf{Y})$ is invariant iff $Q(\mathbf{Y}) = Q(\mathbf{MY})$. In summary, an arbitrary *invariant quadratic estimate* (IQE) of $\mathbf{t}$ is given by a (k, n^2)-matrix $\mathbf{L}$ according to $Q(\mathbf{Y}) = \mathbf{L} \cdot \mathbf{MY} \otimes \mathbf{MY}$.

Clearly,

$$\begin{aligned} \mathscr{E}\mathbf{MY} \otimes \mathbf{MY} &= \mathbf{M} \otimes \mathbf{M} \cdot \mathscr{E}(\mathbf{Y} - \mathbf{Xb}) \otimes (\mathbf{Y} - \mathbf{Xb}) \\ &= \mathbf{M} \otimes \mathbf{M} \cdot \operatorname{vec} \mathscr{D}\mathbf{Y} = \mathbf{M} \otimes \mathbf{M} \cdot \textstyle\sum t_\kappa \operatorname{vec} \mathbf{V}_\kappa . \end{aligned}$$

By introducing the (n^2, k)-matrices $\mathbf{D} = (\operatorname{vec} \mathbf{V}_1 \vdots \cdots \vdots \operatorname{vec} \mathbf{V}_k)$ and $\mathbf{D}_\mathbf{M} = \mathbf{M} \otimes \mathbf{M} \cdot \mathbf{D}$, $\mathbf{MY} \otimes \mathbf{MY}$ gives rise to the *derived linear model*

$$\mathscr{E}\mathbf{MY} \otimes \mathbf{MY} = \mathbf{M} \otimes \mathbf{M} \cdot \mathbf{Dt} = \mathbf{D}_\mathbf{M}\mathbf{t}. \tag{2}$$

Estimating dispersion components in a linear model (1) is a question of point of view: $\mathbf{t}$ may be looked at as regression parameter for the dispersion matrix in the original model (1) or of the mean vector in the derived model (2).

3. Estimates

Theorem 1 derives various properties of a *minimum variance–minimum bias–IQE* (MV–MB–IQE) $\hat{\mathbf{t}}$ of $\mathbf{t}$ from the theory of mean estimation (cf., [5, p. 307]). Note that given $\hat{\mathbf{t}}$ and $\mathbf{q} \in \mathbb{R}^k$, the MV–MB–IQE of $\mathbf{q}'\mathbf{t}$ is $\mathbf{q}'\hat{\mathbf{t}}$, and that $\hat{\mathbf{t}}$, $\mathbf{q}'\hat{\mathbf{t}}$ are unbiased whenever possible, i.e., rank $\mathbf{D}_\mathbf{M} = k$, $\mathbf{q} \in \mathscr{R}\mathbf{D}_\mathbf{M}'$, respectively.

THEOREM 1. *Let* $\mathbf{F} = \mathscr{D}(\mathbf{Y} - \mathscr{E}\mathbf{Y}) \otimes (\mathbf{Y} - \mathscr{E}\mathbf{Y})$ *be the known matrix of fourth moments in the linear model* (1), *put* $\mathbf{M} = \mathrm{Proj}(\mathscr{R}\mathbf{X})^{\perp}$, $\mathbf{N} = \mathrm{Proj}(\mathscr{R}\mathbf{D_M})^{\perp}$, $\mathbf{F_M} = \mathbf{M} \otimes \mathbf{M} \cdot \mathbf{F} \cdot \mathbf{M} \otimes \mathbf{M}$. *Then*: (i) *The ordinary least squares estimate* $\mathbf{D_M}^{+} \cdot \mathbf{Y} \otimes \mathbf{Y}$ *is a* MV–MB–IQE *of* $\mathbf{t}$ *iff* $\mathscr{R}\mathbf{F_M}\mathbf{D} \subset \mathscr{R}\mathbf{D_M}$. (ii) *The normal equations* $\mathbf{D}'\mathbf{F_M}^{+}\mathbf{D}\hat{\mathbf{t}} = \mathbf{D}'\mathbf{F_M}^{+} \cdot \mathbf{Y} \otimes \mathbf{Y}$ *yield a* MV–MB–IQE *of* $\mathbf{t}$ *iff* $\mathscr{R}\mathbf{D_M} \subset \mathscr{R}\mathbf{F_M}$. (iii) *Every* MV–MB–IQE *of* $\mathbf{t}$ *when the fourth moments are* $\mathbf{G}$, *is a* MV–MB–IQE *of* $\mathbf{t}$ *when the fourth moments are* $\mathbf{F}$, *iff* $\mathscr{R}\mathbf{F_M}\mathbf{N} \subset \mathscr{R}\mathbf{G_M}\mathbf{N}$.

Proof. The results follow when the theory of mean estimation is applied to the derived model, cf., [7, p. 654, 658; 3, p. 148].

EXAMPLE. When $\mathbf{Y}$ is normally distributed with zero mean one has $\mathbf{F} \cdot \mathrm{vec}\,\mathbf{A} = (\mathscr{D}\mathbf{Y} \otimes \mathbf{Y}) \cdot \mathrm{vec}\,\mathbf{A} = (\mathscr{D}\mathbf{Y}) \otimes (\mathscr{D}\mathbf{Y}) \cdot \mathrm{vec}(\mathbf{A} + \mathbf{A}')$ for any (n, n)-matrix $\mathbf{A}$. Let $\mathscr{B}_1$ be the set of dispersion matrices admissible in model (1), and let $\mathscr{B}$ be the subspace of symmetric matrices spanned by the $\mathbf{V}_\kappa$. Then $\mathscr{R}\mathbf{D} = \mathrm{vec}\,\mathscr{B}$, and $\mathscr{R}\mathbf{F}\mathbf{D} = \mathrm{vec}(\mathscr{D}\mathbf{Y})\,\mathscr{B}(\mathscr{D}\mathbf{Y})$. When $\mathbf{I}_n \in \mathscr{B}_1$, then the least squares estimate is of minimum variance with respect to every $\mathbf{V} \in \mathscr{B}_1$ iff $\mathbf{V}\mathscr{B}\mathbf{V} \subset \mathscr{B}$ for all $\mathbf{V} \in \mathscr{B}_1$, by Theorem 1(i). This leads to quadratic subspaces as introduced and discussed by Seely [6, p. 714]. Applications to Hsu's model are given in [4].

Theorem 2 concerns Rao's MINQUE theory [5, p. 302–305] where unbiasedness is relaxed to minimum bias. Assume $\mathbf{T}$ to be nonnegative definite. When estimating a linear form $\mathbf{q}'\mathbf{t}$ one has to minimize

$$\mathrm{trace}\ \mathbf{ATAT} = \| \mathbf{T}^{1/2}\mathbf{MAMT}^{1/2} \|^2 = \|(\mathrm{vec}\,\mathbf{A})' \cdot \mathbf{M} \otimes \mathbf{M} \cdot (\mathbf{T} \otimes \mathbf{T})^{1/2} \|^2$$

among all MB–IQEs $\mathbf{Y}'\mathbf{AY} = (\mathrm{vec}\,\mathbf{A})' \cdot \mathbf{MY} \otimes \mathbf{MY}$. When estimating $\mathbf{t}$ this generalizes to minimizing $\| \mathbf{L} \cdot \mathbf{M} \otimes \mathbf{M} \cdot (\mathbf{T} \otimes \mathbf{T})^{1/2} \|^2$ among all MB–IQEs $Q(\mathbf{Y}) = \mathbf{L} \cdot \mathbf{MY} \otimes \mathbf{MY}$. A resulting estimate will be called *minimum norm–minimum bias–IQE* (MN–MB–IQE) of $\mathbf{t}$.

THEOREM 2. *Assume the notation of Theorem* 1 *and let* $\mathbf{T}$ *be nonnegative definite. Then*: (i) *Parts* (i) *and* (ii) *of Theorem* 1 *remain true for* MN–MB–IQEs *instead of* MV–MB–IQEs *if* $\mathbf{F}$ *is replaced by* $\mathbf{T} \otimes \mathbf{T}$. (ii) *There is a unique* MN–MB–IQE *for* $\mathbf{t}$ *iff* $\mathbf{My} \otimes \mathbf{My} \in \mathscr{R}\mathbf{D_M} + \mathscr{R}\mathbf{F_M}$ *for all* $\mathbf{y} \in \mathbb{R}^n$. (iii) *Every* MN–MB–IQE *is a* MV–MB–IQE *when the fourth moments are* $\mathbf{F}$, *iff* $\mathscr{R}\mathbf{F_M}\mathbf{N} \subset \mathscr{R}\mathbf{MTM} \otimes \mathbf{MTM} \cdot \mathbf{N}$.

Proof. Minimizing variances of $\mathbf{L} \cdot \mathbf{MY} \otimes \mathbf{MY}$ would mean minimizing $\mathrm{trace}\ \mathscr{D}\mathbf{L} \cdot \mathbf{MY} \otimes \mathbf{MY} = \| \mathbf{L} \cdot \mathbf{M} \otimes \mathbf{M} \cdot \mathbf{F}^{1/2} \|^2$. The theorem thus follows by formal identification of MN–MB–IQEs with MV–MB–IQEs when the fourth moments are $\mathbf{F} = \mathbf{T} \otimes \mathbf{T}$, cf., [5, p. 305]. Part (ii) follows since the MV–MB–IQE has a unique representation iff all possible observations are in $\mathscr{R}\mathbf{D_M} + \mathscr{R}\mathbf{F_M}$ [7, p. 658].

EXAMPLE. Let $\mathbf{T}$ be positiv definite, $\mathbf{R} = \mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{T}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{T}^{-1}$, and $\mathbf{S} = \mathbf{D}' \cdot \mathbf{R} \otimes \mathbf{R} \cdot \mathbf{D}$, whence $\mathbf{R} = (\mathbf{MTM})^{+}$, $\mathscr{R}\mathbf{R} = \mathscr{R}\mathbf{MTM} = \mathscr{R}\mathbf{M}$, $\mathscr{R}\mathbf{D_M}' = \mathscr{R}\mathbf{S}$. By Theorem 2(ii), Theorem 1(ii), the unique MN–MB–IQE of $\mathbf{t}$ is $\mathbf{S}^{-}\mathbf{D}' \cdot \mathbf{RY} \otimes \mathbf{RY}$. For $\mathbf{q} \in \mathbb{R}^k$, $\mathbf{q}'\mathbf{t}$ is estimable iff $\mathbf{q} \in \mathscr{R}\mathbf{D_M}'$, i.e., $\mathbf{q} \in \mathscr{R}\mathbf{S}$, and in this case its MINQUE is

$$\mathbf{q}'\mathbf{S}^{-}\mathbf{D}' \cdot \mathbf{RY} \otimes \mathbf{RY} = \boldsymbol{\lambda}'\mathbf{D}' \cdot \mathbf{RY} \otimes \mathbf{RY} = \mathbf{Y}'\mathbf{R}\textstyle\sum \lambda_\kappa \mathbf{V}_\kappa \mathbf{RY}, \ \mathbf{S}\boldsymbol{\lambda} = \mathbf{q},$$

as given by Rao [5, p. 304].

Finally, the derived model (2) may successfully be employed to derive Ridge-type estimates for the variance components $\mathbf{t}$. For example, an IQE $\mathbf{L} \cdot \mathbf{MY} \otimes \mathbf{MY}$ of a linear form $\mathbf{q}'\mathbf{t}$, $\mathbf{q} \in \mathbb{R}^k$, minimizes the weighted sum $s = \| \mathbf{L} \cdot \mathbf{M} \otimes \mathbf{M} \cdot \mathbf{F}^{1/2} \|^2 + \tau^2 \| \mathbf{L} \cdot \mathbf{D_M} - \mathbf{q}' \|^2$, $\tau^2 > 0$, of variances and bias iff $\mathbf{L} = \mathbf{q}'\mathbf{D_M}'(\tau^{-2}\mathbf{F_M} + \mathbf{D_M}\mathbf{D_M}')^{-}$ [5, p. 305–306]. Note that the mean square error of $\mathbf{L} \cdot \mathbf{MY} \otimes \mathbf{MY}$ at $\mathbf{t}$ is $\| \mathbf{L} \cdot \mathbf{M} \otimes \mathbf{M} \cdot \mathbf{F}^{1/2} \|^2 + \|(\mathbf{LD_M} - \mathbf{q}')\mathbf{t}\|^2$, with maximum value s when $\mathbf{t}$ varies subject to $\| \mathbf{t} \| \leqslant \tau$; hence an estimate as mentioned above minimizes the maximal mean square error when $\| \mathbf{t} \| \leqslant \tau$.

ACKNOWLEDGMENT

The author would like to thank Prof. Hermann Witting for his encouragement and continued support.

REFERENCES

[1] GREUB, W. H. (1967). *Multilinear Algebra*. Springer–Verlag, Berlin.
[2] MITRA, S. K. (1971). Another look at Rao's MINQUE of variance components. *Bull. Inst. Internat. Statist.* **44**,2 279–283.
[3] MITRA, S. K. AND MOORE, B. J. (1973). Gauss–Markov estimation with an incorrect dispersion matrix. *Sankhyā Ser. A* **35** 139–152.
[4] PUKELSHEIM, F. (1976). On Hsu's model in regression analysis. *Math. Operationsforsch. Statist.* To appear.
[5] RAO, C. R. (1973). *Linear Statistical Inference and its Applications, 2nd ed.* Wiley, New York.
[6] SEELY, J. (1971). Quadratic subspaces and completeness. *Ann. Math. Statist.* **42** 710–721.
[7] ZYSKIND, G. (1975). Error structures, projections and conditional inverses in linear model theory. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, Ed.), pp. 647–663. North Holland, Amsterdam.