On Hsu's Model in Regression Analysis

FRIEDRICH PUKELSHTEIM

Summary. The paper exemplifies with Hsu's model a general pattern as how to derive results of variance component estimation from well known results of mean estimation, as far as linear model theory is concerned. This 'dispersion-mean-correspondence' provides new and short proofs for various theorems from the literature, concerning unbiased invariant quadratic estimators with minimum Bayes risk or minimum variance. For pure variance component models, unbiased non-negative quadratic estimability is characterized in terms of the design matrices.

1. Introduction

The purpose of this communication is to exemplify with Hsu's model a general pattern as how to derive results of variance component estimation from known results of mean estimation.

The dispersion-mean-correspondence (Sect. 2) introduces a derived model such that mean regression in the derived model corresponds to estimating the variance components in the original model. Hsu's model specifies the fourth moments via the kurtosis γ, thus opening the way for minimum variance estimation. Sect. 3 collects some matrix algebra for convenient reference. In Sect. 4, some known results of J. Kleeke and R. Pincus [8], P. L. Hsu [6], H. Drygas [2], and C. R. Rao [14] are proved by persistently applying the dispersion-mean-correspondence. This approach extends insight and understanding of linear model theory, providing short proofs, slight generalizations, and alternative characterizations. Sect. 5 is concerned with the existence of unbiased non-negative definite quadratic estimates of variance components and presents an estimability criterion in terms of the design matrices that specify the model.

Previous Work. S. K. Mitra [10] suggests an approach that is very close to the dispersion-mean-correspondence as presented here or in [12], he, however, stops exploitation at an intermediate stage. J. Seely [16] and other authors, cf., H. Drygas [3], S. Gnot, W. Kloncki and R. Zmysloný [4], J. Kleeke [7], R. Zmysloný [18], reduce variance component estimation to mean estimation in coordinate free terms which seems to provoke ready application somewhat less than the dispersion-mean-correspondence. Most of the examples in Sect. 4 were originally proved by explicitly minimizing the risk function, though the connection to mean estimation is hinted at (H. Drygas [2, p. 382]) or

1 Institut für Mathematische Stochastik der Universität, Hermann-Herder-Str. 10, D-7800 Freiburg im Breisgau, Federal Republic of Germany.
implicit (C. R. Rao [14, p. 451]). The algebraic notions as introduced below have successfully been utilized by other authors as well, cf., J. Kleffe and P. Pincus [8], S. K. Mitra [10], G. P. H. Styan [17], R. Zmyslony [18].

**Notations.** For a matrix $A$, let $A'$, $A^+$, $\Re A$, $\Re^2 A$ denote its transposed matrix, Moore-Penrose inverse [15, p. 26], range (column space), and the range’s orthogonal complement, respectively. Let $\text{vec} \ A$ be the vector obtained from $A$ by ordering its entries lexicographically. The Kronecker product [15, p. 29] is denoted by $\otimes$. The function vec is an inner product and tensor product preserving vector space isomorphism:

$$
(\text{vec} \ A)' \text{vec} \ B = \text{trace} \ AB',
$$

$$
\text{vec} \ xy' = x \otimes y,
$$

as follows by considering the standard basis vectors $e_r$ with $r$-th component 1 and zeroes elsewhere, and the basis matrices $E_{r\mu} = e_r e_{\mu}'$ with $(r, \mu)$-entry 1 and zeroes elsewhere. For $(\text{vec} \ A)'$ we shall also write $\text{vec}' A$. $I$ denotes an identity matrix, its order following from the context.

## 2. Model Set-up and the Dispersion-Mean-Correspondence

The General Linear Model. For an $\mathbb{R}^n$-valued random vector $Y$, a linear model is specified by linear decompositions of both the mean vector $EY$ and the dispersion matrix (variance-covariance matrix) $DY$:

$$
EY = \sum_{s=1}^p b_s x_s = Xb, \quad DY = \sum_{s=1}^k t_s V_s, \quad M = I - XX^+,
$$

where the $(n, p)$-matrix $X = [x_1, \ldots, x_p]$ and the $k$ symmetric $(n, n)$-matrices $V_s$ are known, whereas $b = (b_1, \ldots, b_p)'$ and $t = (t_1, \ldots, t_k)'$ are to be estimated; $M$ is the orthogonal projector onto $\Re X$.

For estimating $t$ or linear functions of $t$, we choose, as usual, quadratic estimators $Q(Y)$ which, by definition, are derived from bilinear functions $E(\cdot, \cdot)$ by setting both arguments equal to $Y$: $Q(Y) = E(Y, Y)$. A maximal invariant statistic with respect to all ‘mean translations’ $y \to y + Xb$, $b \in \mathbb{R}^p$, is $MY$ (cf., J. Seely [16, p. 1646], J. Kleffe [7]). Thus, $Q(Y)$ is an invariant quadratic estimator (IQE) iff $Q(Y) = Q(MY)$.

E.g., when estimating a linear form $q't$, $q \in \mathbb{R}^k$, the set of all IQEs is $\{Y'AY\}$, where $A$ is an arbitrary symmetric $(n, n)$-matrix satisfying $A = MAM$, or, equivalently, $AX = 0$.

Invariance is a natural statistical requirement: All of $Y$ which is in the range $\Re X$ may be explained by mean regression, leaving the residuals $MY$ for inference on the dispersion parameter $t$. Technically speaking, invariant estimates of $t$ are free from the mean parameter, cf., R. R. Corbeil and S. R. Searle [1], J. Kleffe [7]. Finally, the expectation of a quadratic estimate $Y'AY$ does not depend on the mean parameter iff $X'AX = 0$. In most applications, as in Hsu’s model, $A$ should be non-negative definite (NND). This, however, and $X'AX = 0$ imply $AX = 0$, i.e., invariance.
The Dispersion-Mean-Correspondence. For a linear model (3), consider the derived random \( R^{a\times} \)-vector \( MY \otimes MY \). By (3) and (2), \( EMY \otimes MY = M \otimes M \cdot \text{vec} \sum t U \), and \( MY \otimes MY \) gives rise to a linear model for mean estimation. Introducing the \((n^2, k)\)-matrices

\[
D = [\text{vec} V_1; \cdots; \text{vec} V_k], \quad D_M = M \otimes M \cdot D,
\]

we arrive at

\[
EMY \otimes MY = D_M t.
\]

Thus, \( t \) may be looked at as the dispersion parameter in the original model (3), or as the mean parameter in the derived model (5); this we call the dispersion-mean-correspondence.

It remains to be shown that the class of natural estimators of \( t \) is not changed by the dispersion mean-correspondence. Clearly, for any \((k, n^2)\)-matrix \( L \),

\[
Q(Y) = L \cdot MY \otimes MY
\]

as a linear estimator of \( t \) in the derived model (5), is an IQE of \( t \) in the original model (3). Conversely, for every bilinear function \( B(x, y) \) there exists a (unique) linear function \( L \) such that \( B(x, y) = L \cdot x \otimes y \). This implies that any IQE \( Q(Y) \) is representable as in (6), and thus is a linear estimator in the derived model. E.g., for an IQE \( YA'Y \) of a linear form \( q't \) we get from (1) and (2) \( Y'AY = \text{trace} A \cdot MY'(MY)' = \text{vec} A' \cdot MY' \otimes MY, i.e., L = vec A'. \)

In Sect. 4 we shall be concerned with minimum variance estimation of \( t \). To this end, we introduce the \((n^2, n^2)\)-matrices

\[
F = D Y \otimes Y, \quad F_M = M \otimes M \cdot F \cdot M \otimes M, \quad N = I - D_M D_M^*,
\]

i.e., the matrix of all central mixed fourth moments of \( Y \), the dispersion matrix of \( MY \otimes MY \), and the orthogonal projector onto \( R^k \cdot D_M \), respectively. The following lemma is stated for later reference. It applies the celebrated Lehmann-Scheffé theorem [15, p. 317] to the derived model; the alternative representations in part (ii) follow by (1), (4), and (12).

**Lemma 2.1.** Let a linear model be given by (3) and (7); let \( L \) be a \((k, n^2)\)-matrix, \( A \) be a symmetric \((n, n)\)-matrix, and \( q \in R^k \). Then:

(i) \( L \cdot MY \otimes MY \) is an unbiased IQE of \( t \) with minimum variance under \( F_M \) (among all other unbiased IQE) iff \( LD_M = I \) and \( LF_M \cdot N = 0 \).

(ii) \( YA'Y \) is an unbiased IQE of \( q't \) with minimum variance under \( F_M \) iff \( AX = 0 \),

\[
q = D_M \cdot \text{vec} A = (\text{trace} V_1 A, \ldots, \text{trace} V_k A)', \quad \text{and}
\]

\[
F_M \cdot \text{vec} A \in \mathbb{R} D_M = \{ \text{vec} \sum \lambda_M Y \cdot M / \lambda \in R^k \}.
\]

Hsu's Model. Hsu's model specifies the fourth moments \( F \) via (quasi-)independence and the kurtosis \( \gamma \) of \( k \) random effects \( \xi \), cf., H. Drygas [2], P. L. Hsu [6], J. Klemm and R. Pincus [8], C. R. Rao [14]. Following C. R. Rao [14, p. 446] we assume a linear decomposition according to

\[
Y - EY = \sum_{k=1}^k U_k \xi_k = U \xi,
\]

where the \( U_k \) are known \((n, c)\)-matrices, \( c = \sum c \), \( U = [U_1; \cdots; U_k] \) is of order \((n, c)\), and \( \xi' = [\xi_1'; \cdots; \xi_k']' \) is a random \( R^c \)-vector whose independent subvectors

\footnote{Statistics, Bd. 8, H. 3/1977}
\(\xi_n\) have independent components \(\xi_{n,v}\) satisfying
\[
E\xi_{n,v} = 0, \quad E\xi_{n,v}^2 = \sigma_v^2, \quad E\xi_{n,v}^3 = (r_n + 3) \sigma_v^3, \quad v = 1, \ldots, c_n.
\] (9)

In general, a random \(R^c\)-vector \(\xi\) whose independent components satisfy \(E\xi_v = 0, E\xi_v^2 = \sigma_v^2, E\xi_v^3 = (r_v + 3) \sigma_v^3\), has mixed fourth moments
\[
D\xi \otimes \xi = \sum_{v,r=1}^c \delta_v^2 \delta_r^2 (E_{v,r} \otimes E_{v,r} + E_{v,r} \otimes E_{r,v} + \delta_{v,r} \delta_{v,r}^2 E_{v,v} \otimes E_{r,r}).
\] (10)

Finally, we introduce
\[
H = \begin{bmatrix}
  1 & 0 & \cdots & 0 \\
  c_1 & \ddots & \ddots & \vdots \\
  0 & \cdots & 1 & c_k
\end{bmatrix}, \quad \tilde{\sigma}^2 = H'\sigma^2, \quad \tilde{\gamma} = H'\gamma,
\] (11)

where the vectors \(1_{c_n}\) consist of \(c_n\) ones, so that \(H\) is of order \((k, c)\). Thus, (10) and (11) yield the mixed fourth moments under assumption (9). We are now ready to precisely define a Hsu-model:

**Definition.** A Hsu-model is a linear model as specified by (3), (4), (7), (10), (11), where \(V = U'U\), \(z = 1, \ldots, k\); \(U\) and the \(U'\)'s are as in (8), \(t = \sigma^2 = (\sigma_1^2, \ldots, \sigma_k^2)'\), \(\gamma = (\gamma_1, \ldots, \gamma_k)'\), and \(F_M = F_{\gamma_\gamma}(\sigma^2, \gamma) = MU \otimes MU \cdot (D\xi \otimes \xi) \cdot U'M \otimes U'M\).

In order to appealingly display \(F_M(\sigma^2, \gamma)\), we now collect some matrix algebra.

### 3. Some Matrix Algebra

The **Separating Property** of \(\vee\). For any 3 matrices \(A, B, C\) of appropriate order one has
\[
\text{vec } ABC = A \otimes C' \cdot \text{vec } B.
\] (12)

Taking \(B = E_{v,r}\), this follows at once from (2).

**Diagonalizer.** For a square matrix \(A\), \(\text{Diag } A\) denotes the diagonal matrix with diagonal entries copied from \(A\). For a vector \(\gamma\), \(\text{Diag } \gamma\) is the diagonal matrix with diagonal equal to \(\gamma\).

Introducing the \((c^2, c)'\-matrix \(D_\gamma = [\text{vec } E_{11} \cdot \cdots \cdot \text{vec } E_{c\gamma}]\), where \(E_{11}\) etc. are basis \((c, c)'\-matrices, yields, by (2), \(D_\gamma D_\gamma' = \sum \text{vec } E_{v,r} \cdot \text{vec } E_{r,v} = \sum (e_v \otimes e_r) \times (e_r \otimes e_v) = \sum E_{v,r} \otimes E_{r,v}\), and, by (12), \(D_\gamma D_\gamma' \cdot \text{vec } A = \text{vec } \text{Diag } A\). More general, for every diagonal \((c, c)'\-matrix \(A\), and for every \((c, c)'\-matrix \(A\), we have
\[
D_\gamma D_\gamma' \cdot \text{vec } A = \text{vec } \Delta \text{Diag } A, \quad \text{where } D_\gamma = [\text{vec } E_{11} \cdot \cdots \cdot \text{vec } E_{c\gamma}].
\] (13)

**Hadamard's Produkt.** When diagonalizer are used, **Hadamard's Produkt** \(A \ast B = ((A_{ij}, B_{ij}))\) is not far. This is due to \(\text{Diag } A = I \ast A\) and the following lemma.

**Lemma 3.1.** Let \(D_\gamma, D_\delta\) be defined as in (13) of order \((c^2, c)\), \((n^2, n)\), respectively. Let \(A, B\) two \((c, n)'\-matrices, and \(a \in \mathbb{R}^c, b \in \mathbb{R}^n\). Then:

(i) \[A \ast B = D_\gamma' \cdot A \otimes B \cdot D_\delta.\]

(ii) \[a' \cdot A \ast B \ast b = \text{trace } \text{Diag } a \cdot A \cdot \text{Diag } b \cdot b'.\]
Proof. Verification is immediate when taking \( A, B, a, b \) to be basis matrices (vectors).

Using Lemma 3.1, many properties of the Hadamard product, cf., e.g., G. P. H. Styan [17], may easily be inferred from corresponding properties of the Kronecker product.

4. On Estimates in the Hsu-Model

The Mixed Fourth Moment. In the next two lemmas, we study the mixed fourth moment as a linear operator.

**Lemma 4.1.** Let \( \xi \) be a random \( R^r \)-vector whose independent components \( \xi_i \) satisfy \( E\xi_i = 0, \) \( E\xi_i^2 = \sigma_i^2 \), and \( E\xi_i = (\gamma_i + 3) \sigma_i^3. \) For a fixed \( \sigma_0^2, \) put \( \Lambda = E\xi = \text{Diag} \{ \sigma_i^2 \}, \) \( \Gamma = \text{Diag} \{ \gamma_i \}, \Delta_1 = \Delta_1, \) \( \Gamma \Delta_1, \) and let \( D_i \) be defined as in (13). Then, for every \((n, n)\)-matrix \( A,\)

\[
(D_i \otimes \Lambda) \cdot \text{vec} A = (2\Delta_1 \otimes \Delta_1 + D_i \Delta_i D_i^T) \cdot \text{vec} \frac{A + A'}{2}.
\]

(14)

Proof. The assertion is a consequence of formulae (10), (12), and (13).

**Lemma 4.2.** Assume a Hsu-model. For a fixed \( \sigma_0^2, \) put \( V_n = \sum \sigma_0^2 V_{n} = U_{n} \Delta_1 U'.\) Then:

(i) For every \((n, n)\)-matrix \( A,\)

\[
F_M(\sigma_0^2, \gamma) \cdot \text{vec} A = (2MV_{\ast} M \otimes MV_{\ast} M + MU \otimes MU - D_i \Delta_i \Delta_i^T) \cdot \text{vec} \frac{A + A'}{2}.
\]

(15)

(ii) For every symmetric \((n, n)\)-matrix \( A \) satisfying \( A = MAM,\)

\[
F_M(\sigma_0^2, \gamma) \cdot \text{vec} A = \text{vec} (2MV_{\ast} AM + MU \Delta_i \cdot \text{Diag} U'AU \cdot U'M).
\]

(16)

Proof. Part (i) follows from Lemma 4.1, and, by (12), implies part (ii).

With these preparations we are now ready to characterize optimal estimates in the Hsu-model.

Bayes estimates. J. Kleffe and R. Pincus [8, Th. 3.8] consider unbiased IQEs with minimum Bayes risk:

**Theorem 4.1.** Assume a Hsu-model with a priori distribution \( P \) for \( \sigma^2, \) put

\[
\bar{R} = E_{\sigma^2 \sigma^2} Y'Y, \quad \bar{S} = \text{Diag} \bar{H}' \bar{H},
\]

(17)

and let \( q \in R^k. \) Then \( Y'AY \) is an unbiased IQE of \( q't \) with minimum Bayes risk at \( \gamma \) iff \( AX = 0, q' = (\text{trace } V_1 A, \ldots, \text{trace } V_n A)', \) and \( MU \times (2H' \bar{H} + \bar{S}) \ast (U'AU) \) \( U'M \) is a linear combination of \( MV_{\ast}M, \ldots, MV_{\ast} M.\)

Proof. By Lemma 4.2.ii, the risk operator is \( \Phi(A) = \text{vec} \ E_{\sigma^2} (2MU \Delta_1 U'AU \Delta_1 \cdot U'M + MU \Delta_i^2 \Gamma \cdot \text{Diag} U'AU \cdot U'M). \) But, cf., [17], \( \Lambda_1 = \text{Diag} \{ \sigma^2 \}, \) \( \Delta_1 = (H' \sigma^2 \sigma^2 H) \ast (U'AU) \) and \( \Delta_2 = (H' \sigma^2 \sigma^2 H) \ast I = \text{Diag} \{ \sigma^2 \}, \) \( H. \) Thus \( E_{\sigma^2} \Delta_1 = \bar{S}, \) and we get \( \Phi(A) = \text{vec} \ MU \times (2H' \bar{H} + \bar{S}) \ast (U'AU) \cdot U'M. \) The assertion now follows by Lemma 2.1.ii, mutatis mutandis.

MV estimates. The rest of this section is concerned with minimum variance unbiased IQE (\( MV - UB - IQE \)) of a linear form \( q't. \)
P. L. Hsu was the first in this area, and his problem [6, Th. 2] reads in terms of the derived model: When is the smallest least squares estimate of minimum variance? Hsu assumes independent components of \( \gamma \) with equal variances \( \sigma^2 \) and possibly unequal kurtosis \( \gamma_1, \ldots, \gamma_n \).

**Theorem 4.2.** Assume a Hsu-model with \( c = n, U = I, V = \sigma^2 I, \) and \( s = \text{rank } X < n \). Put \( M_1 = M \ast M, \) and \( m = (M_{11}, \ldots, M_{nn})' \). Then \((n - s)^{-1} Y'MY \) is a MV - UB - 1QE of \( \sigma^2 \) iff

\[
M_1 m = q m, \quad q = (n - s)^{-1} m' T m.
\]

(18)

Proof. Here, \( D_M = \text{vec } M \), so that the smallest least squares estimate is \((D_M^T D_M)^{-1} D_M^T MY = (\text{trace } M)^{-1} trace M \cdot MY(MY)' = (n - s)^{-1} Y'MY \), by (1) and (2). This is of minimum variance [11, p. 148] iff \( D_M \) is invariant under \( F_{M}(\sigma_0^2, \gamma) \). Lemma 4.2.ii yields \( F_{M}(\sigma_0^2, \gamma) \cdot \text{vec } M = \sigma_0^2 \cdot \text{vec } (2M + \Gamma \cdot \text{Diag } M' M) \). Thus, a necessary and sufficient condition is

\[
M \cdot \text{Diag } T m \cdot M = q M, \quad q = (n - s)^{-1} m'Tm.
\]

(19)

By considering the diagonal elements, (19) implies (18). The converse follows, with Lemma 3.1.ii, from \( \|M \cdot \text{Diag } T m \cdot M - q M\|^2 = \| \text{Diag } T m \cdot M \cdot \text{Diag } T m \cdot M - 2q \text{trace } \text{Diag } m'T \cdot \text{Diag } m + q^2 \text{trace } M = m'T M m - 2q \text{trace } T m + \),

\[
+ q^2 m'T m = m'T M m - q m m.
\]

The next result, due to H. Drygas [2, Th. 3.5.a], does also lead to eq. (19).

**Theorem 4.3.** Let \( \gamma' = \gamma A \) be an IQE of \( \sigma^2 \) for the Hsu-model in Th. 4.2. Then \( \gamma' = \gamma A \) is a MV - UB - 1QE of \( \sigma^2 \) iff \( \text{trace } A = 1 \) and \( 2A + \Gamma \cdot \text{Diag } A' M \) is a scalar multiple of \( M \).

Proof. Using Lemmas 2.1.ii, 4.2.ii, we get \( 1 = D_M \text{vec } A = \text{trace } A \), and \( F_{M}(\sigma_0^2, \gamma) \text{vec } A = \sigma_0^2 \text{vec } (2MAM + \Gamma \cdot \text{Diag } A' M) \). For the general Hsu-model, C. R. Rao [14, Th. 1] derives the following estimates. J. Kleffe [7, Th. 2] gives similar representations.

**Theorem 4.4.** Let a Hsu-model be given. For a fixed \( \sigma_0^2 \) introduce the \((n, n)\)-matrices \( V_a = \sum \sigma_0^2 V_a = U \Delta_1 U' \), and \( R_a = (MV_a M)' \), and the \((c, c)\)-matrices \( M_1 = U'R_a U, \ M_2 = M_1 \ast M_1 \). Let \( q \in R^k \). If \( V_a \) is positive definite, and the \( R \)-vector \( \Theta \) and the \( R^k \)-vector \( \lambda \) satisfy

\[
H \Theta = q, \quad \Theta = M_2 \Delta_2 \Theta - M_1 \Delta_1 \Theta,
\]

then an unbiased IQE of \( q' \) with minimum variance at \( (\sigma_0^2, \gamma) \) is given by

\[
Y'R_a U \cdot \text{Diag } \frac{1}{2}(H \lambda' - \Delta_1 \Theta) \cdot U'R_a Y' .
\]

(20)

Proof. The assertion follows from Lemma 2.1.ii. Firstly, since \( R_a = U a M \), we have \( A = \frac{1}{2} R_a U \cdot \text{Diag } (H \lambda' - \Delta_1 \Theta) \cdot U'R_a \). Now, \( \text{vec } U \cdot \text{Diag } H' \Theta \cdot U' = \text{vec } \sum \delta_a V_a = D \delta_a \), and, by (13), \( \Delta_2 \text{Diag } \Theta = D \Delta_2 \Theta \). Thus, from (12), \( L' = \text{vec } A = \frac{1}{2} R_a \ominus R_a \cdot D \delta_a - \frac{1}{2} R_a U \ominus R_a U \cdot D \Delta_2 \Theta \). Using \( D = U \ominus U \cdot D H' \) and Lemma 3.1, this yields

\[
D'^2 L' = \frac{1}{2} H \delta' \cdot U \ominus U' \cdot R_a \ominus R_a \cdot U \ominus U \cdot D H' \lambda - \frac{1}{2} HD' \cdot M_1 \Omega
\]

(21)
\( \odot M_1 \cdot D_1 \Delta_2 \Theta = \frac{1}{2} HM_1 H' \lambda - \frac{1}{2} HM_2 \Delta_2 \Theta, \) which equals \( q \) under the assumption (20).

Secondly, the positive definiteness of \( V_\ast \) implies \( R_\ast R_\ast = M_\ast \). Lemma 4.2.1, then, yields \( F_M(\sigma_\ast^2, \gamma) \cdot L' = M \otimes M \cdot D \lambda - M U \otimes M U \cdot D_1 \Delta_2 \Theta + \frac{1}{2} M U \otimes M U \cdot D_2 \Delta_2 \Theta' \cdot U R_\ast \otimes U R_\ast \cdot D \lambda - \frac{1}{2} M U \otimes M U \cdot D_1 \Delta_1 \Theta = D M_\ast \lambda - M_\ast H' \lambda + M_\ast \Delta_2 \Theta. \) Under the assumption (20), this is in \( R D_M \). \[ \blacksquare \]

**Uniformity Criteria.** Essentially, the last theorem of this section is also due to C. R. Rao [14, p. 453–454]. The first part characterizes those situations when the estimates are independent of the kurtosis \( \gamma \); the second part assumes quasi normality, i.e., \( \gamma = 0 \), and investigates independence from \( \sigma_\ast^2 \).

**Theorem 4.5.** Assume the Hsu-model and notation of Th. 4.4. Then:

(i) All MV – UB – IQEs at \( (\sigma_\ast^2, 0) \) are of minimum variance at \( (\sigma_\ast^2, \gamma) \) iff \( \Re D_M \) is invariant under \( M U \otimes M U \cdot D_1 \Delta_1 \Theta \cdot \text{Diag} \ H' \gamma \cdot D_1 \cdot U R_\ast \otimes U R_\ast \), or, equivalently, iff \( \Re M_\ast H' \) is invariant under \( M_\ast \Delta_2 \text{Diag} H' \gamma \).

(ii) All MV – UB – IQEs at \( (1, 0) \) are of minimum variance at \( (\sigma_\ast^2, 0) \) iff \( \Re D_M \) is invariant under \( M U \otimes M U \cdot R \otimes U R_\ast \), or, equivalently, iff for every \( \lambda \in \Re^k \) there exist a \( \mu \in \Re^k \) such that \( \sum \lambda \nu \nu R \nu \nu R = \sum \mu \nu \nu R \). Here, \( V = \sum V_\ast \) is assumed positive definite, \( \Re = (M V M)^+ \), and \( V_\ast \) is assumed non-negative definite.

**Proof.** (i) We have to check [11, p. 147] when \( \Re F_M(\sigma_\ast^2, \gamma) \cdot N \subset \Re F_M(\sigma_\ast^2, 0) \cdot N. \) This is the case, by Lemma 4.2.3 after premultiplying with \( R_\ast \otimes R_\ast \), iff the range of \( A = R_\ast U \otimes R_\ast U \cdot D_1 \Delta_1 \Theta \cdot U M \otimes U M \otimes U M \otimes U M \cdot N \) is contained in \( \Re M \otimes M \otimes M \otimes M \cdot N \). But \( M \otimes M \otimes M \otimes M \) is a projector, so \( N \cdot M \otimes M \cdot A = A \) yields \( D_1 \cdot R_\ast U \otimes R_\ast U \cdot D \Delta_1 \Theta \cdot U M \otimes U M \cdot N = 0. \) This is true iff \( \Re M \otimes M \cdot D_1 \Delta_1 \Theta \cdot U R_\ast \otimes U R_\ast \cdot D M \subset \Re D_M \), (22)

which is the first characterization. Using \( D = U \otimes U \cdot D H' \) and premultiplying (22) with \( D_1 \cdot U R_\ast \otimes U R_\ast \) yields the second characterization, [14, eq. 5.8].

For part (ii), check \( \Re F_M(\sigma_\ast^2, 0) \cdot N \subset \Re F_M(1, 0) \) analogously to get the first characterization. Then premultiply with \( R \otimes R \), use \( D_3 \lambda = \nu \nu \sum \lambda \mu \nu \nu M \cdot M \) and (12) to get the second characterization [14, eq. 6.4]. \[ \blacksquare \]

As indicated in the proof of part (ii), formulae (12) and (13) may be used to reformulate the above criteria in matrix space. The formulation above parallels that of mean estimation: The BLUE under \( V \) is BLUE under \( V_\ast \) iff \( \Re X \) is invariant under \( V_\ast V^{-1} \) (cf., [11, p. 149]).

For further results on uniform MV – UB – IQE see H. Drygas [3] and S. Gnot, W. Klonecki and R. Zmyslony [4]. The dispersion-mean-correspondence may also be used to get MINQUE, weighted least squares, or Ridge-type estimates of variance components, cf., F. Pukelsheim [12], [13]. Maximum likelihood estimates are considered, e.g., by R. R. Correll and S. R. Searle [1].
5. Unbiased NND Quadratic Estimability

Though one should not dispense with requiring that a quadratic estimate \( Y'AY \) of a single variance component \( \sigma^2 \) be non-negative definite (NND), one does, and there are only few investigations of this subject, cf., H. Drygas [2], L. R. La Motte [9]. The next lemma is implicitly given by L. R. La Motte [9, p. 728], the formulation below is to stress that NND estimates shift the problem from linearity into convexity.

**Lemma 5.1.** Assume a linear model as given by (3), let \( q \in \mathbb{R}^k \). Then:

(i) There exists an unbiased IQE of \( q' t \) iff
\[
q \in \text{linear hull } \{ D'M : y \otimes y \mid y \in \mathbb{R}^n \}.
\]

(ii) There exists an unbiased NND quadratic estimator of \( q' t \) iff
\[
q \in \text{convex hull } \{ D'M : y \otimes y \mid y \in \mathbb{R}^n \}.
\]

**Proof.** Unbiasedness and non-negative definiteness imply invariance, see above Sect. 2. \( Y'AY \) is unbiased for \( q' t \) iff \( q = D'M \cdot \text{vec} \ A \). Assertions i, ii then follow from the spectral representation [15, p. 39] of symmetric matrices, NND matrices, respectively.

What has the model to look like such that a single component \( t_\alpha \) be unbiasedly NND estimable? For a pure variance components model, e.g., a Hsu-model, we finally prove as a necessary and sufficient condition: The \( \alpha \)-th dispersion design must properly contribute to the explanation of the error space \( \mathbb{R} M \sum \mathcal{V}_\alpha M \), (cf., [15, p. 297]).

**Theorem 5.1.** For a linear model (3), let all \( \mathcal{V}_\alpha \), \( \alpha = 1, \ldots, k \), be NND, and fix \( \alpha \). Then there exists an unbiased NND quadratic estimator of \( t_\alpha \) iff
\[
\mathbb{R} \mathcal{V}_\alpha M \subset \mathbb{R} M \sum \mathcal{V}_\alpha M.
\]

**Proof.** Put \( \mathfrak{L} = \text{convex hull } \{ D'M : y \otimes y \mid y \in \mathbb{R}^n \} \). Clearly, \( \mathfrak{L} = \{ \sum D'y \cdot y \otimes y \mid y_1, \ldots, y_k \in \mathbb{R}^n \} \). With all \( \mathcal{V}_\alpha \) NND, it is easily shown that \( e_\alpha \in \mathfrak{L} \) iff \( e_\alpha \in \{ D'M : y \otimes y \mid y \in \mathbb{R}^n \} \). The latter means that the nullspace of \( M \sum \mathcal{V}_\alpha M \) be not contained in the nullspace of \( \mathcal{V}_\alpha M \). This is the orthogonal dual of the assertion.

Acknowledgement

The author would like to thank Prof. H. Witting for his encouragement and permanent support. He is also grateful to Prof. H. Drygas and the referee for suggesting improvements and extensions.

References

On Hsu's Model in Regression Analysis


**Zusammenfassung**


**Résumé**

Dans ce travail on démontre de nouveau quelques théorèmes concernant l’estimation des composants de la variance dans le modèle linéaire de Hsu: Par l’application d’une ‘correspondance dispersion-moyenne’ la question posée se trouve réduite aux problèmes déjà connus et résolus dans la théorie d’estimation de la moyenne. Dans les théorèmes cités on étudie des estimateurs quadratiques, invariants, sans biais, qui minimisent soit le risque bayesien soit la variance. En outre, l’existence d’un estimateur quadratique, non-négatif, sans biais d’un seul composant de la variance est caractérisée par des matrices qui, elles, déterminent le modèle.

Received July 1975. Revised May 1976.