Modelling Examination Marks, II

D. J. DALEY

Statistics Department
University of North Carolina at Chapel Hill
(visiting)[1]

ABSTRACT

In a given population $\mathscr{C}$ suppose that student i studies a subset $\mathscr{A}_i$ of the subjects offered in a curriculum and that marks or assessment scores $\{X_{ij}: i \in \mathscr{C}_j\}$ are given reflecting the (relative) achievements of the candidature $\mathscr{C}_j$ of the students taking subject j . Various estimates of scale parameters $\{\beta_j\}$ in the one-factor model

$$\alpha_j + \beta_j X_{ij} \equiv Y_{ij} = v_i + e_{ij}$$

for uncorrelated error variables $\{e_{ij}\}$ are examined for unbiasedness: those based on a method of moments approach appear to be asymptotically optimal. Further, for the range of values of $\mathrm{var}(v_i)$ and $\mathrm{var}(e_{ij})$ encountered in practice, the same estimators are fairly robust against the two-factor model

$$Y_{ij} = v_{i1} + \gamma_j v_{i2} + e'_{ij}$$

in which verbal/quantitative contrast factor measures $\{v_{i2}\}$ supplement the general factors $\{v_{i1}\}$, while the sizes of the residuals $\{e'_{ij}\}$ are then close to the known sizes of the measurement errors they incorporate.

---

[1]Permanent address: Statistics Dept. (IAS), Australian National University, G.P.O. Box 4, Canberra A.C.T. 2601, Australia.

1 Introduction

This paper is a sequel to an earlier joint work (Daley and Seneta, 1986), referred to below as (I), in which a one-factor model was proposed to describe a data set $\{X_{ij}\}$ such as the examination marks obtained by students at the Year 12 level at which they complete their secondary schooling and seek entry to a tertiary institution on the basis of their exam. marks as their academic record. Within the several systems operating in Australasia, such data sets have for some one to two decades been the source for the prime or even sole determinant of entry to university or college of advanced education. The earlier paper and this are concerned with examining the basis for such determination which, for all that it has been accepted at large, is neither well understood nor administered with the degree of impartiality and sophistication that might be hoped for. This is said because the analyses that follow from the discussion below lead to the conclusion that, if the existing procedures were brought into line with what is being attempted, then for admission to some tertiary level courses, proportions of up to some 10 to 20 or even 30 per cent. of students who at present gain admission would be replaced by others. Unquestionably, existing procedures are technically sloppy: worse, the sloppiness exists to an extent that there are observable consequences of appreciable size. However, it would be wrong to attach blame to some authorities, because the technical sloppiness has first to be exposed, and that in part is what this paper is about.

The data set $\{X_{ij}\}$ need not necessarily consist simply of examination marks. For example, the marks or scores may arise from school-based assessments, or from a combination of them with exam. marks, or from reference test scores such as Aptitude tests (e.g. SAT scores in USA or Australian Scholastic Aptitude Test scores). It will be convenient to call all such measures *scores* or *marks*, and to call the "subject" or "course" area from which they are derived a *subject* or *course*, even though there may not necessarily be a uniform and precisely defined "subject" for the individuals i given scores X_{ij} in the subject or course j.

The dominant issues in this paper are the consequences of estimating parameters (α_j, β_j) in the linear transformation

$$Y_{ij} = \alpha_j + \beta_j X_{ij} \tag{1.1}$$

and the resultant properties of the average score

$$Y_{i\cdot} \equiv \text{ave}_i(Y_{ij}) \tag{1.2}$$

(formal first and second moment operators like $\text{ave}_i(\cdot)$ are defined around (1.3)–(1.7) below). In particular we note the different properties of $Y_{i\cdot}$ that arise from a variety of possible estimation procedures for (α_j, β_j) and from a variety of possible model assumptions for $\{Y_{ij}\}$.

The basic notation used here is consistent with that of Daley (1988). The set $\mathscr{C}$ of all individuals with scores $\{X_{ij}\}$ is called the *candidature*. In general it will be a *sub-candidature*

$$\mathscr{C}_j \equiv \{i\colon i \text{ has a score } X_{ij}\}$$

of individuals taking a particular course j, because usually each student i is expected to choose only a subset $\mathscr{S}_i$ from all the subjects $\mathscr{S}$ in the curriculum. Student i has $n_i \equiv \#(\mathscr{S}_i)$ course scores and $N_j \equiv \#(\mathscr{C}_j)$ students take course j. Formal moments are defined as in

$$\text{ave}_j(X_{ij}) = \Sigma_{i\in\mathscr{C}_j} X_{ij}/N_j\,, \tag{1.3}$$

$$[\text{s.d.}_j(X_{ij})]^2 = \text{var}_j(X_{ij}) = \Sigma_{i\in\mathscr{C}_j} [X_{ij} - \text{ave}_j(X_{ij})]^2/N_j\,, \tag{1.4}$$

$$\text{ave}_{jk}(X_{ij}) = \Sigma_{i\in\mathscr{C}_{jk}} X_{ij}/N_{jk}\,, \tag{1.5}$$

$$\text{cov}_{jk}(X_{ij}, X_{ik}) = \Sigma_{i\in\mathscr{C}_{jk}} [X_{ij} - \text{ave}_{jk}(X_{ij})][X_{ik} - \text{ave}_{kj}(X_{ik})]/N_{jk}\,, \tag{1.6}$$

where $\mathscr{C}_{jk} = \mathscr{C}_j \cap \mathscr{C}_k$, $N_{jk} = \#(\mathscr{C}_{jk})$, and

$$\text{corr}_{jk}(X_{ij}, X_{ik}) = \text{cov}_{jk}(X_{ij}, X_{ik})/\{[\text{s.d.}_{jk}(X_{ij})][\text{s.d.}_{kj}(X_{ik})]\} . \tag{1.7}$$

Under (1.1),

$$\text{ave}_j(Y_{ij}) = \alpha_j + \beta_j \text{ave}_j(X_{ij}) , \tag{1.8}$$

$$\text{var}_j(Y_{ij}) = \beta_j^2 \text{var}_j(X_{ij}) . \tag{1.9}$$

A major aim here is to elucidate what is entailed in basing ranking decisions on statistics like (cf. (1.2))

$$Y_{i\cdot} = \Sigma_{j \in \mathscr{S}_i} Y_{ij}/n_i \tag{1.10}$$

or more generally, for some subset $\mathscr{S}_i'$ of $\mathscr{S}_i$ that may depend on $\{Y_{ij}: j \in \mathscr{S}_i\}$.

$$Y_{i\cdot}'' = \Sigma_{j \in \mathscr{S}_i'} Y_{ij}/\#(\mathscr{S}_i') . \tag{1.11}$$

Whether it is specifically stated or not, Australasian practice has reflected as an act of faith that, no matter what scaling procedure has been used, a representation of the form

$$Y_{ij} = v_i + e_{ij} \tag{1.12}$$

then holds for certain error terms e_{ij} with zero mean, and that this is a valid unbiased representation. If this is so, then

$$Y_{i\cdot}'' = v_i + e_i'' \tag{1.13}$$

for some error term e_i'' that does not necessarily have zero mean but does have smaller variance than (almost all) e_{ij} . Empirically, a representation such as (1.12) does hold as a crude first approximation, implying that, when a recipe such as at (1.11) is followed, the representation at (1.13) necessarily holds in this crude sense. Because of this implication, it follows that when the parameter v_i in (1.12) is replaced by an estimate, unbiasedness of that estimate will be a highly desirable property.

There is an important practical reason for considering *linear* transformations such as at (1.1), relating to an invariance property of rankings as at (1.10) and (1.11). It is simply this, that if $\{(\alpha_j, \beta_j), v_j\}$ as in (1.1) and (1.12) are replaced by

$$\{(\alpha_j', \beta_j'), v_j'\} \equiv \{(A + \alpha_j B, \beta_j B), A + Bv_j\} \tag{1.14}$$

for any real A and real positive B , so that in place of Y_{ij} we should have

$$Y_{ij}' = A + \alpha_j B + \beta_j B X_{ij} = A + BY_{ij}, \tag{1.15}$$

the ranking as follows from (1.10) or (1.11) is unchanged, and the implication that (1.13) holds is likewise unchanged.

2 One-factor model

Whether unknowingly or explicitly as reported for example in Daley (1987) and Seneta (1987), existing Australasian mark-scaling procedures are based on an assumption that the one-factor model as in (I) provides a satisfactory description of the scores $\{X_{ij}\}$ concerned. The linearly transformed scores Y_{ij} as at (1.1), or more generally the transforms $f_j(X_{ij})$ for some family of monotonic increasing functions $f_j(\cdot)$, are assumed to have the structure

$$Y_{ij} = v_i + e_{ij} \tag{2.1}$$

for some unknown common (achievement) measure v_i and residual error terms e_{ij} that may embody both model-fit and measurement errors, such that, when viewed as random variables (r.v.s), the set $\{e_{ij}: i \in \mathscr{C}_j\}$ has

$$E(e_{ij}) = 0 , \quad \mathrm{Var}(e_{ij}) = \sigma_j^2 , \tag{2.2}$$

and is independent of both $\{v_i\}$ and sets $\{e_{ik}\}$ for $k \neq j$. (The use of $\mathrm{Var}(\cdot)$ as distinct from e.g. $\mathrm{var}_j(\cdot)$ is deliberate.) From these assumptions it follows that

$$\mathrm{Ave}_j(Y_{ij}) = \mathrm{ave}_j(v_i) , \tag{2.3}$$

$$\mathrm{Var}_j(Y_{ij}) = \mathrm{var}_j(v_i) + \sigma_j^2 , \tag{2.4}$$

$$\mathrm{Cov}_{jk}(e_{ij}, e_{ik}) = 0 , \tag{2.5}$$

$$E(\mathrm{cov}_j(v_i, e_{ij})) = 0 , \tag{2.6}$$

$$\mathrm{Cov}_{jk}(Y_{ij}, Y_{ik}) = \mathrm{var}_{jk}(v_i) . \tag{2.7}$$

3 Estimation in the one-factor model

The concern of this section is with the following questions:

If random variables $\{X_{ij}\}$ ***are such that*** $\{Y_{ij}\} \equiv \{\alpha_j + \beta_j X_{ij}\}$ ***satisfy the assumptions of section 2, what estimators of*** $\{(\alpha_j, \beta_j), v_i\}$ ***might be used, and what are their properties?***

3.1 Maximum likelihood estimation

Suppose additionally, only in this sub-section, that the r.v.s e_{ij} are independently and normally distributed like $N(0, \sigma_j^2)$ r.v.s. Then the likelihood of the data set $\{X_{ij}\}$ is well-defined by

$$L \equiv \prod_{j \in \mathscr{S}} \prod_{i \in \mathscr{C}_j} (\sigma_j \sqrt{2\pi})^{-1} \exp[-(\alpha_j + \beta_j X_{ij} - v_i)^2 / 2\sigma_j^2] . \tag{3.1}$$

Suppose that for one particular j we have $(\alpha_j, \beta_j) = (0, 1)$ and $v_i = X_{ij}$. Then the term $\exp[\cdot]$ in (3.1) is identically one for this j , irrespective of σ_j^2 , and L is maximized by setting $\sigma_j^2 = 0$ and takes the value $L = \infty$.

In comparison with the three other estimators from (I), the assumption of normality in order to have an expression for L is unnecessarily strong. Also, as in (I), it is unreasonable to assume that $\sigma_j^2 = 0$ for any j . Accordingly, this approach will be considered no further here.

3.2 Least squares estimation

It is frequently the case that there are close connections between least squares and maximum likelihood estimators. Accordingly, in reverting to the general one-factor model assumptions as outlined earlier without any assumptions of normality, we start by seeking estimators via the minimization of

$$S^2 \equiv \sum_{j\in\mathscr{S}} \sum_{i\in\mathscr{C}_j} (Y_{ij} - v_i)^2 = \sum_{j\in\mathscr{S}} \sum_{i\in\mathscr{C}_j} (\alpha_j + \beta_j X_{ij} - v_i)^2 . \tag{3.2}$$

Inspection shows that S^2 is minimized with the value 0 by setting $\alpha_j = \beta_j = v_i = 0$ for all $j \in \mathscr{S}$ and $i \in \mathscr{C}_j$. Now the only conceivable boundary condition involving these parameters concerns the positivity of β_j, so we can expect that any other minimum of S^2 will be a solution $\{(\bar{a}_j, \bar{b}_j), \bar{v}_i\}$ of the least squares normal equations

$$\bar{v}_i = \Sigma_{j\in\mathscr{S}_i} (\bar{a}_j + \bar{b}_j X_{ij})/n_i \equiv \mathrm{ave}_i(\bar{a}_j + \bar{b}_j X_{ij}) = \mathrm{ave}_i(Y_{ij}) , \tag{3.3}$$

$$\bar{a}_j + \bar{b}_j \mathrm{ave}_j(X_{ij}) = \mathrm{ave}_j(\bar{v}_i) , \tag{3.4}$$

$$\bar{b}_j \mathrm{var}_j(X_{ij}) = \mathrm{cov}_j(\bar{v}_i, X_{ij}) . \tag{3.5}$$

Given any non-trivial solution $\{(\bar{a}_j, \bar{b}_j), \bar{v}_i\}$, inspection shows that for any real A and real positive B, the set $\{(A + B\bar{a}_j, B\bar{b}_j), A + B\bar{v}_i\}$ is also a solution. The all-zero solution is included in this set at the point $(A, B) = (0, 0)$.

Suppose without loss of generality that the data set $\{X_{ij}\}$ is generated by the model

$$X_{ij} = v_i + e_{ij} , \tag{3.6}$$

and that the set $\{(\bar{a}_j, \bar{b}_j), \bar{v}_i\}$ satisfies the equations (3.3)–(3.5) with $(\bar{a}_s, \bar{b}_s) = (0, 1)$ for some s (this last condition simply determines values of the parameters A, B within the class of solutions of the equations). Observe that the model quantities have $(\alpha_j, \beta_j) = (0, 1)$ for all j: we shall show that in general we cannot expect the least squares estimators to have these values as their long-run average values, and thereby conclude that

the least squares estimators $\{(\bar{a}_j, \bar{b}_j), \bar{v}_i\}$ ***are in general biased.*** (3.7)

Start by assuming that all N_j are sufficiently large for the strong law of large numbers to hold, so that

$$\text{ave}_j(X_{ij}) = \text{ave}_j(v_i) + O_p(N^{-\frac{1}{2}}) , \tag{3.8}$$

$$\text{var}_j(X_{ij}) = \text{var}_j(v_i) + \sigma_j^2(1 + O_p(N^{-\frac{1}{2}})) . \tag{3.9}$$

Then, correct to terms that are $o_p(1)$ in $N_{.}$,

$$\bar{v}_i = \text{ave}_i(\bar{a}_j + \bar{b}_j X_{ij}) = \text{ave}_i(\bar{a}_j + \bar{b}_j(v_i + e_{ij}))$$

$$= \text{ave}_i(\bar{a}_j) + v_i \text{ave}_i(\bar{b}_j) + \text{ave}_i(\bar{b}_j e_{ij}) ,$$

$$\text{Var}(\bar{v}_i) = o(1) + o(1) + \text{ave}_i(\bar{b}_j^2 \sigma_j^2) ,$$

$$\text{Var}_j(\bar{v}_i) \approx o(1) + [\text{ave}_j(\text{ave}_i(\bar{b}_k))]^2 \text{var}_j(v_i) + \text{ave}_j[\text{ave}_i(\bar{b}_k^2 \sigma_k^2)] .$$

In practice, there is interest in finding transformations that satisfy one of the two sets of constraints:

[1] $\{X_{is}\} = \{Y_{is}\}$ for some particular subject s , equivalently, $(\alpha_s, \beta_s) = (0, 1)$:

[2] $\Sigma_j \Sigma_i (X_{ij} - Y_{ij}) = 0 = \Sigma_j \Sigma_i (X_{ij}^2 - Y_{ij}^2)$, equivalently,

$$X_{..} \equiv \text{ave}_{\text{all}}(X_{ij}) \equiv \frac{\Sigma_{j \in \mathcal{S}} N_j \text{ave}_j(X_{ij})}{\Sigma_{j \in \mathcal{S}} N_j} = \frac{\Sigma_{j \in \mathcal{S}} N_j \text{ave}_j(Y_{ij})}{\Sigma_{j \in \mathcal{S}} N_j} \equiv Y_{..} \tag{3.10}$$

and

$$\text{var}_{\text{all}}(X_{ij}) \equiv \frac{\Sigma_j \Sigma_i (X_{ij} - X_{..})^2}{\Sigma_{j \in \mathcal{S}} N_j} = \frac{\Sigma_{j \in \mathcal{S}} N_j[\text{var}_j(X_{ij}) + (\text{ave}_j(X_{ij}) - X_{..})^2]}{\Sigma_{j \in \mathcal{S}} N_j}$$

$$= \frac{\Sigma_{j \in \mathcal{S}} N_j[\text{var}_j(Y_{ij}) + (\text{ave}_j(Y_{ij}) - Y_{..})^2]}{\Sigma_{j \in \mathcal{S}} N_j} \equiv \text{var}_{\text{all}}(Y_{ij}) . \tag{3.11}$$

Either of these sets of constraints leads to seeking solutions $\{(\bar{a}_j, \bar{b}_j), \bar{v}_i\}$ of a modified set of equations. In the case of [1] for example, we should seek to minimize

$$S^2 + \lambda_1 \alpha_s + \lambda_2(\beta_s - 1) \tag{3.12}$$

for Lagrangian multipliers λ_1 , λ_2 , leading to equations (3.3)–(3.5) as above for $j \neq s$

while for $j = s$ we have instead

$$\bar{a}_s + \bar{b}_s \mathrm{ave}_s(X_{is}) + \lambda_1 = \mathrm{ave}_s(\bar{v}_i) ,$$

$$\bar{b}_s \mathrm{var}_s(X_{is}) + \lambda_2 = \mathrm{cov}_s(\bar{v}_i, X_{is}) .$$

In view of the assumed identity $(\alpha_s, \beta_s) = (0, 1)$, these yield

$$\lambda_1 = \mathrm{ave}_s(\bar{v}_i) - \mathrm{ave}_s(X_{is}) , \tag{3.13}$$

$$\lambda_2 = \mathrm{cov}_s(\bar{v}_i, X_{is}) - \mathrm{var}_s(X_{is}) , \tag{3.14}$$

which with the other equations at (3.3)–(3.5) yield a set of $2\#(\mathscr{A}) + \#(\mathscr{C})$ linear equations in as many unknowns. In what follows, it is assumed that this set has a unique solution (cf. (I)). Observe that the convergence of any iterative routine for solving the equations is prima facie evidence for the existence though not necessarily the uniqueness of a solution.

The identities that follow from a similar treatment of the constraints at [2] are more suggestive because they can be written (with $X_{\cdot j} = \mathrm{ave}_j(X_{ij})$) in the forms

$$\Sigma_j N_j[\alpha_j + (\beta_j - 1)X_{\cdot j}] = 0 , \tag{3.15}$$

$$\Sigma_j N_j[(\beta_j^2 - 1)\mathrm{var}_j(X_{ij}) + (\alpha_j + \beta_j X_{\cdot j})^2 - X_{\cdot j}^2] = 0 , \tag{3.16}$$

from which we may anticipate that if the scores X_{ij} satisfy (3.6) with $\mathrm{ave}_{all}(v_i) = 0$, then when $n_i \approx n$ independent of i and v_i , both

$$\Sigma_j N_j(\beta_j - 1) \approx 0 \qquad \text{and} \qquad \Sigma_j (\beta_j - 1) \approx 0 . \tag{3.17}$$

The expression to be minimized, with Lagrangian parameters λ_1 and λ_2 , is now

$$S^2 + 2\lambda_1 \Sigma_j N_j[\alpha_j + (\beta_j - 1)X_{\cdot j}] + \lambda_2 \Sigma_j N_j[(\beta_j^2 - 1)\mathrm{var}_j(X_{ij}) + (\alpha_j + \beta_j X_{\cdot j})^2 - X_{\cdot j}^2] .$$

The resulting normal equations can be written in the form

$$[\bar{a}_j + \bar{b}_j X_{\cdot j} - \text{ave}_j(\bar{v}_i)] + \lambda_1 + \lambda_2[\bar{a}_j + \bar{b}_j X_{\cdot j}] = 0 , \qquad (3.18)$$

$$\Sigma_{i \in \mathscr{C}_j} X_{ij}[\bar{a}_j + \bar{b}_j X_{ij} - \bar{v}_i] + \lambda_1 N_j X_{\cdot j} + \lambda_2 N_j[\bar{b}_j \text{var}_j(X_{ij}) + X_{\cdot j}(\bar{a}_j + \bar{b}_j X_{\cdot j})] = 0 , \qquad (3.19)$$

or equivalently, for each $j \in \mathscr{J}$,

$$\lambda_1 + (1 + \lambda_2)(\bar{a}_j + \bar{b}_j X_{\cdot j}) = \text{ave}_j(\bar{v}_i) , \qquad (3.18)'$$

$$(1 + \lambda_2)\bar{b}_j \text{var}_j(X_{ij}) = \text{cov}_j(\bar{v}_i, X_{ij}) . \qquad (3.19)'$$

Compare the Lagrangian multipliers here with the particular solution as below (3.5) with $A = \lambda_1$ and $B = 1 + \lambda_2$.

What now follows amplifies remarks[2] in (I) concerning the relative sizes of the least squares (LS) estimators $\bar{b}_j$ of β_j . Assume without loss of generality that the relations (3.6) hold and that LS estimators $\{(\bar{a}_j, \bar{b}_j), \bar{v}_i\}$ have been determined satisfying (3.3)–(3.5) subject to the constraints (3.10) and (3.11). What then is

$$E(\bar{Y}_{ij}) = E(\bar{a}_j + \bar{b}_j X_{ij}) = E(\bar{a}_j) + E(\bar{b}_j)v_i \ ?$$

Observe that for large N_j ,

$$\text{var}_j(X_{ij}) \approx \text{var}_j(v_i) + \sigma_j^2 ,$$

$$\text{cov}_j(\bar{v}_i, X_{ij}) = \text{cov}_j\left[\Sigma_{k \in \mathscr{J}_i} (\bar{a}_k + \bar{b}_k X_{ik})/n_i , X_{ij}\right]$$

$$= \text{cov}_j\left[\Sigma_{k \in \mathscr{J}_i} (\bar{a}_k + \bar{b}_k v_i + \bar{b}_k e_{ik})/n_i , v_i + e_{ij}\right]$$

$$\approx \text{ave}_j(\text{ave}_i(\bar{b}_k))\text{var}_j(v_i) + \bar{b}_j \sigma_j^2/n$$

where $1/n = \text{ave}_{\text{all}}(1/n_i)$. Consequently,

$$\bar{b}_j \approx \frac{\text{ave}_j(\text{ave}_i(\bar{b}_k)) + \bar{b}_j/(n\Gamma_j)}{(\lambda_2 + 1)(1 + 1/\Gamma_j)} \qquad (3.20)$$

[2]Masters and Beswick (1986), in quoting remarks from (I) about least squares estimators at their §2.49, erroneously inferred that they apply to method-of-moment estimators.

where

$$\Gamma_j = \mathrm{var}_j(v_i)/\sigma_j^2 . \tag{3.21}$$

In practice, $\mathrm{cov}_j(\overline{v}_i, X_{ij}) > 0$ so $1 + \lambda_2 > 0$, $n \approx 5$, and $\Gamma_j \approx 3$ to 6 . Noting that $E(\overline{b}_j) = 1$ for unbiased $\overline{b}_j$, we ask how much does $\overline{b}_j$ then differ from 1 ? Consider two scenarios. First, suppose the iterated average in (3.20) equals 1 ; this requires $1 + \lambda_2 \approx 5/6$ and the range for $\overline{b}_j$, which is then a function of Γ_j , is about 0.96 to 1.07 . For the second scenario, noting that it tends to be the case that students have more courses with Γ_k in common, and hence $\overline{b}_k$ in common, replace the double average by $(1 + \overline{b}_j)/2$; using $1 + \lambda_2 \approx 5/6$ again, $\overline{b}_j$ is now about 0.92 to 1.14. In either case, the assertion at (3.7) is supported, and it is supported more strongly by the example involving what appear to be the more realistic approximations.

3.3 Estimation via mean and variance equating

Suppose given just two sets of scores, $\{X_{i0}\}$ and $\{X_{i1}\}$ say, and suppose that the latter set is such that, after some unknown linear transformation as at (1.1), the resulting scores $\{Y_{i1}\}$ have the same structure (3.6) as $\{X_{i0}\}$ with $\mathrm{Var}(X_{i0} - v_i) = \mathrm{Var}(Y_{i1} - v_i)$ for all i . In this bivariate context, the following estimation procedure is asymptotically appropriate, remaining so in the multivariate context *provided* that the variance terms σ_j^2 are constant for all j , a condition which is not met in practice.

Notwithstanding the absence of such justification in terms of consistency with any model, it has been common practice to use as estimators of $\{(\alpha_j, \beta_j), v_i\}$ the attempted "solution" to the set of equations

$$\overline{\overline{v}}_i = \mathrm{ave}_i(\overline{\overline{a}}_j + \overline{\overline{b}}_j X_{ij}) , \tag{3.22}$$

$$\overline{\overline{b}}_j = [\mathrm{var}_j(\overline{\overline{v}}_i)/\mathrm{var}_j(X_{ij})]^{\frac{1}{2}} = \mathrm{s.d.}_j(\overline{\overline{v}}_j)/\mathrm{s.d.}_j(X_{ij}) , \tag{3.23}$$

$$\bar{\bar{a}}_j + \bar{\bar{b}}_j X_{\cdot j} = \mathrm{ave}_j(\bar{\bar{v}}_i) . \tag{3.24}$$

Both in theory and in practice, these equations when iterated converge to the degenerate (null) solution because the ratio at (3.23) is < 1. This inconsistency has been resolved in practice by fixing the estimated scale parameters $\bar{\bar{b}}_j$ after one or two iterations and determining $\bar{\bar{a}}_j$ for such fixed $\bar{\bar{b}}_j$. In view of the invariance properties around (1.14) and (3.5), an alternative is to impose one of the sets of conditions at [1] and [2] above.

Provided now that both σ_j^2 and $\mathrm{var}_j(v_i)$ are independent of j, the procedure is consistent with the model as outlined. To see this, observe as earlier that if any consistent non-degenerate solution of these equations exists, then a family of such solutions will exist consistent with the invariance property already noted.

Next, suppose as earlier that the data set is generated as at (3.6). Then for

$$E(\bar{Y}_{ij}) \equiv E(\bar{\bar{a}}_j) + E(\bar{\bar{b}}_j)v_i$$

it again suffices to consider just the scale parameter estimator. For large N_j, much as in Section 3.2, (3.9) holds while

$$\mathrm{var}_j(\bar{\bar{v}}_i) = \mathrm{var}_j\Big[\Sigma_{k\in \mathscr{S}_i} (\bar{\bar{a}}_k + \bar{\bar{b}}_k X_{ik})/n_i \Big] = \mathrm{var}_j\Big[\Sigma_{k\in \mathscr{S}_i} (\bar{\bar{a}}_k + \bar{\bar{b}}_k v_i + \bar{\bar{b}}_k e_{ik})/n_i \Big]$$

$$\approx \mathrm{ave}_j\{[\mathrm{ave}_i(\bar{\bar{b}}_k)]^2\}\mathrm{var}_j(v_i) + \mathrm{ave}_j[\mathrm{ave}_i(\bar{\bar{b}}_k^2\sigma_k^2)]/n .$$

Consequently,

$$\bar{\bar{b}}_j^2 \approx \frac{\mathrm{ave}_j\{[\mathrm{ave}_i(\bar{\bar{b}}_k)]^2\}[1 + 1/(n\Gamma_j)]}{1 + 1/\Gamma_j} \tag{3.25}$$

and, as in the previous section, $\bar{\bar{b}}_j$ again varies with Γ_j, and in general is asymptotically biased on each side of 1. However, the bias is about half that of using LS estimators, so this method is preferable to LS estimation.

3.4 Method of Moments estimation

Refer back to the equations (2.2)–(2.7) where it was noted that, when (2.1) holds,

$$E(\alpha_j + \beta_j \mathrm{ave}_j(X_{ij})) = \mathrm{ave}_j(v_i) , \tag{3.26}$$

$$E(\beta_j^2 \mathrm{var}_j(X_{ij})) = \mathrm{var}_j(v_i) + \sigma_j^2 , \tag{3.27}$$

$$E(\mathrm{ave}_j(\alpha_k + \beta_k X_{ik})) = v_i . \tag{3.28}$$

Furthermore, from $v_i(\alpha_j + \beta_j X_{ij}) = v_i^2 + v_i e_{ij}$ we have

$$E(\beta_j \mathrm{cov}_j(v_i, X_{ij})) = \mathrm{var}_j(v_i) . \tag{3.29}$$

Method of moment estimation entails replacing the unknown parameters in (3.26)–(3.29) by their estimators $\{(\tilde{a}_j, \tilde{b}_j), \tilde{v}_i\}$ say, and solving for them. Again, since as earlier the equations are no longer linear in the unknowns, an iterative solution scheme is adopted, of which a more extended account with a particular data set has been detailed in Daley (1987). For this, it is not necessary to use (3.27) other than to estimate σ_j^2 after finding all the other parameters, that is, ***the method of moment estimators*** $\{(\tilde{a}_j, \tilde{b}_j), \tilde{v}_i\}$ ***satisfy the equations***

$$\tilde{a}_j + \tilde{b}_j \mathrm{ave}_j(X_{ij}) = \mathrm{ave}_j(\tilde{v}_i) , \tag{3.30}$$

$$\tilde{b}_j \mathrm{cov}_j(\tilde{v}_i, X_{ij}) = \mathrm{var}_j(\tilde{v}_i) , \tag{3.31}$$

$$\tilde{v}_i = \mathrm{ave}_j(\tilde{a}_k + \tilde{b}_k X_{ik}) , \tag{3.32}$$

while the (biased) estimator $\tilde{s}_j^2$ of σ_j^2 is then given by

$$\tilde{s}_j^2 = \tilde{b}_j^2 \mathrm{var}_j(X_{ij}) - \mathrm{var}_j(\tilde{v}_i) . \tag{3.33}$$

In practice, initial estimates such as $(\alpha_j, \beta_j) = (0, 1)$ are taken and successively iterated through (3.32), (3.31), and (3.30), as a perturbation analysis shows that under the conditions usually encountered, such a scheme then has satisfactory convergence properties.

To study the bias properties of the estimators of $\{\beta_j\}$ it again entails no loss of generality to assume that (3.6) holds. Then, asymptotically as before,

$$\begin{aligned}\mathrm{var}_j(\tilde{v}_i) &= \mathrm{var}_j\left[\Sigma_{k\in\mathscr{S}_i}(\tilde{a}_k + \tilde{b}_k X_{ik})/n_i\right] = \mathrm{var}_j\left[\Sigma_{k\in\mathscr{S}_i}(\tilde{a}_k + \tilde{b}_k v_i + \tilde{b}_k e_{ik})/n_i\right] \\ &\approx \mathrm{ave}_j\{[\mathrm{ave}_i(\tilde{b}_k)]^2\}\mathrm{var}_j(v_i) + \mathrm{ave}_j[\mathrm{ave}_i(\tilde{b}_k^2\sigma_k^2)]/n \ , \end{aligned} \tag{3.34}$$

$$\begin{aligned}\mathrm{cov}_j(\tilde{v}_i, X_{ij}) &= \mathrm{cov}_j\left[\Sigma_{k\in\mathscr{S}_i}(\tilde{a}_k + \tilde{b}_k X_{ik})/n_i, X_{ij}\right] \\ &= \mathrm{cov}_j\left[\Sigma_{k\in\mathscr{S}_i}(\tilde{a}_k + \tilde{b}_k v_i + \tilde{b}_k e_{ik})/n_i,\ v_i + e_{ij}\right] \\ &= \mathrm{ave}_j[\mathrm{ave}_i(\tilde{b}_k)]\mathrm{var}_j(v_i) + \tilde{b}_j\sigma_j^2/n \ . \end{aligned} \tag{3.35}$$

Thus,

$$\tilde{b}_j \approx \frac{\mathrm{ave}_j[\mathrm{ave}_i(\tilde{b}_k)]^2\mathrm{var}_j(v_i) + \mathrm{ave}_j[\mathrm{ave}_i(\tilde{b}_k^2\sigma_k^2)]/n}{\mathrm{ave}_j[\mathrm{ave}_i(\tilde{b}_k)]\mathrm{var}_j(v_i) + \tilde{b}_j\sigma_j^2/n} \ , \tag{3.36}$$

which is closer to being unbiased than either of the two previous estimators. In particular, it is much less affected by variability of Γ_j. Note also that

$$\begin{aligned}\mathrm{Var}(\tilde{v}_i) &= \mathrm{Var}\left[\Sigma_{k\in\mathscr{S}_i}(\tilde{a}_k + \tilde{b}_k v_i + \tilde{b}_k e_{ik})/n_i\right] \\ &\approx \mathrm{Var}\left[\Sigma_{k\in\mathscr{S}_i}(\tilde{a}_k + \tilde{b}_k v_i)/n_i\right] + \Sigma_{k\in\mathscr{S}_i}\tilde{b}_k^2\sigma_k^2/n_i^2 \\ &= O(N^{-1}) + n_i^{-1}\mathrm{ave}_i(\tilde{b}_k^2\sigma_k^2) \ . \end{aligned} \tag{3.37}$$

3.5 External reference measure

Suppose finally that a further set of scores $\{\tilde{\tilde{v}}_i\}$ is provided as estimators of $\{v_i\}$, so that

$$\tilde{\tilde{v}}_i = v_i + e_{iV} \tag{3.38}$$

where $E(e_{iV}) = 0$, $E(e_{iV}^2) = \sigma_V^2$, $\mathrm{cov}(v_i, e_{iV}) = 0$. It has been common in the educational measurement literature to use $\{\tilde{\tilde{v}}_i\}$ for "reference score equating" by requiring that

$$\text{ave}_j(Y_{ij}) = \text{ave}_j(\tilde{V}_i) , \quad \text{var}_j(Y_{ij}) = \text{var}(\tilde{V}_i) , \tag{3.39}$$

in spite of its being known that biased estimators of the scale parameter β_j then ensue (see e.g. Cooney (1974, 1977), Hasofer (1977), and Potthoff (1982)), largely as a result it would appear that a model such as (2.1) was not in view, and in particular there was no suggestion of an approach via the estimation of v_i . In view of the model assumptions, equations (3.39) are equivalent to assuming that $\sigma_j^2 = \sigma_V^2$ for each j concerned. This assumption is similar to that of the mean and variance equating estimation procedure already outlined. On the one hand, it recognizes that both sets of scores $\{X_{ij}\}$ and $\{\tilde{V}_i\}$ are subject to error (i.e., imprecise determination), whether coming from model-fitting or actual measurement or both. On the other hand, it assumes that these errors are of the same size for all scores, whether from courses or the reference test, when in practice these are known to vary considerably (cf. the range 3 to 6 for Γ_j: for evidence, see Daley (1985) and Daley and Eyland (1987)).

Observe also that the error variance of the estimator is now proportional to σ_V^2/N_j rather than $\text{ave}_j[\text{ave}_i(\sigma_k^2)]/n \approx \sigma_j^2/(nN_j)$, and the increases from both $\sigma_j^2 < \sigma_V^2$ and $1/n < 1$ introduce appreciable errors into the estimation of v_i unless N_j is large.

Of even more concern is that estimates of the location parameters α_j are now prone to bias within the error variables e_{iV} . Broadly speaking these can be regarded as cultural biases, as for example concerning ethnicity and gender with SAT scores in USA and of gender in both Australia and UK (*MATHEF* (1986) refers to a survey paper manuscript of Daley).

3.6 Which estimation procedure?

In terms of precision of estimates, it is unquestionably the case that any of the *other course score* procedures of sections 3.2 to 3.4 is preferable to the external reference measure procedure of section 3.5. This is a simple consequence of the fact that, if a parameter v_i is estimated by several measures, and a measurement error (or, errors in variables) model is appropriate, then information on the parameter is derived better from a reasonable combination of all the observations contributing more or less equally rather than relying on a single set of observations. Within this group of procedures, the criterion of unbiasedness of the scale parameters $\{\beta_j\}$, which is relevant in the tails of the distribution of $\{v_i\}$ though less critical than unbiasedness of the location parameters $\{\alpha_j\}$, means the method of moment procedure of section 3.4 is to be preferred.

It is possible in principle to investigate these methods via either or both of Monte Carlo methods and resampling procedures. The major practical problem associated with using the former is to construct a data set consistent with both the model and the pattern of courses $\mathscr{S}_i$ taken by students in relation to their general measures v_i. One solution is to use the estimates of both $\{v_i\}$ and $\{\sigma_j^2\}$ from a data set (e.g., as from the method of moment procedure), and replace the observed errors by simulated values $\{e'_{ij}\}$ which should then be reasonably independent. For the latter, jack-knife estimates of $\mathrm{Var}(v_i)$ for example may be appropriate through the use of a common set of subsamples for different estimation procedures.

The estimation procedures of sections 3.2 to 3.4 can also be used in conjunction with an external reference measure such as $\{\tilde{V}_i\}$ by regarding the latter as a set of scores from some course, as for example regarding it as the course s as under the constraint [1] above (3.10). Such a procedure was adopted in the analyses to which brief reference is made in Chapter 5 of *MATHEF* (1986).

The one-factor model and its associated estimation procedures can be used on subsets of courses when the latter are chosen by some external prescriptive criteria. For example, ad hoc analyses have been performed on classifying courses $j \in \mathscr{S}$ as lying in either a humanities (verbal) domain or a science and mathematics (quantitative) domain, and a procedure similar to that of section 3.5 followed within each of the two resulting subsets, whereas what has been sketched in section 3.4 would be much more appropriate. Again, all that is being reflected here is a lack of understanding of the logical need for any algorithm to be governed by a mathematical model that describes the context of the information being processed by the algorithm in such a way that, ideally, the principles underlying the algorithm and its application to the model are mutually consistent, optimal, and consistent with the data.

4 One-factor model procedures used on a two-factor model

It has been assumed so far that the one-factor model is a satisfactory description of the data in the sense that the sets of residuals $\{e_{ij}\}$ are mutually uncorrelated. (While we stated an assumption of independence at (2.2), all we have used, except for the maximum likelihood procedure which we have rejected, is this zero correlation property.) Since a *ranking* is a one-dimensional concept and the parameters $\{v_i\}$ correspond in a general sense to a first principal component of the multivariate set $\{Y_{ij}\}$, i.e., to the dominant component, it is arguable that at this stage it is enough to check that the resulting error terms are uncorrelated.

In practice, the data sets are such that a second component is always observable, and a third is also observable when certain external reference measures are used. It is therefore proper to consider the one-factor model estimation procedures in relation to these more detailed models. In this section we consider the following two-factor model which corresponds to the practical observation that many students tend to be relatively better in one of the two areas defined by a preponderance of verbal skills for one and quantitative skills for the other. (In more colloquial terms, students tend to be better in either the humanities area or the science and mathematics area.) Suppose then that we retain (1.1) but that instead of (2.1) we have

$$Y_{ij} = v_{i1} + \gamma_j v_{i2} + e'_{ij} \tag{4.1}$$

for some family of constants $\{\gamma_j\}$, general achievement measures $\{v_{i1}\}$, contrast measures $\{v_{i2}\}$, and residual variables $\{e'_{ij}\}$, such that over their common sub-candidatures, $\{v_{i1}\}$ and $\{v_{i2}\}$ are mutually uncorrelated and uncorrelated also with $\{e'_{ij}\}$. It is immediately recognizable that, in addition to the indeterminate parameters A, B as at (1.14) and (1.15) for the model at (2.1), there is another indeterminacy in the model at (4.1) in that the quantities $\{C\gamma_j\}$ and $\{C^{-1}v_{i2}\}$ yield the same description of any data set.

We content ourselves for the time being with observing that if (4.1) holds and we form the estimator of $Y_{i\cdot}$ at (1.10) by

$$\begin{aligned}\Sigma_{j\in\mathcal{S}_i} Y_{ij}/n_i &= \Sigma_{j\in\mathcal{S}_i} (v_{i1} + \gamma_j v_{i2} + e'_{ij})/n_i \\ &= v_{i1} + (\Sigma_{j\in\mathcal{S}_i} \gamma_j/n_i)v_{i2} + (\Sigma_{j\in\mathcal{S}_i} e'_{ij}/n_i) \\ &\equiv v_{i1} + \gamma'_{i\cdot}|v_{i2}| + e'_{i\cdot} \quad \text{say}, \qquad (4.2)\end{aligned}$$

then again the dominant component is v_{i1} but, typically, because a student if anything tends to have a majority of courses from the area of relative strength in terms of the contrast measure v_{i2}, this dominant component is increased by a fraction of the contrast measure. (It is tacitly being assumed here that the coefficients γ_j lie in the range $(-1, 1)$ or thereabouts, by appropriate choice of the arbitrary constant C.) The last statement means that, no matter what convention has been adopted with regard to the sign of v_{i2}, each student will tend to have a majority of courses for which γ_j has the same sign as v_{i2}, and thus, taking $\gamma'_{i\cdot} \equiv |\Sigma_{j\in\mathcal{S}_i} \gamma_j/n_i| = |\text{ave}'_i(\gamma_j)|$, the second term on the right-hand side of (4.2) is (usually) positive as implied.

The representation (4.2) makes little sense until we have some idea of the magnitude of the quantities involved. Our experience with data from three Australian sources indicates that $\text{var}_{\text{all}}(v_{i1}) : \text{var}_{\text{all}}(v_{i2}) \approx 4 : 1$ or larger, that $\gamma'_i \approx 0.2$ to 0.5, and that $\text{Var}(e'_{i\cdot}) \approx \text{var}_{\text{all}}(v_{i1})/10$ or less, so that the measures $Y_{i\cdot}$ can certainly be regarded as providing a classification of the population into several subgroups if that is required.

Questions of misclassification rates have been canvassed in Daley (1988).

A major benefit of having the representation (4.2) is that it explains observed covariances $\text{cov}_{jk}(Y_{ij}, Y_{ik})$ better than the one-factor model (2.1). To show this, we must make some assumptions that approximate the participation rates of students in various courses. To this end, assume that every student takes courses $j = 1$ and 2, these common courses being one in each of the major areas (e.g. every student takes English and Mathematics), that each student then takes three further courses in his area of strength, and that $\gamma'_{i.} \approx |1.0 - 1.0 + 3(0.5)|/5 = 0.3$. Suppose also that 60% of students are in the area of course 1 and 40% in the other. (We could equally use 50% in each: we choose otherwise in order to illustrate effects of imbalance, of which the first is that the representation at (4.2) must be modified by replacing $|v_{i2}|$ by the top 60% of v_{i2} for the area of course 1 , and the top 40% of $-v_{i2}$ for the area of course 2 .) We shall suppose that the raw scores X_{ij} have the representation at (4.2), much as we made the assumption about $\beta_j = 1$ at (3.6) in our study of biases of b_j in section 3. with $\gamma_1 = -\gamma_2 = 1$ for the sake of definiteness. We have

$$\Sigma_i v_{i1} = 0 = \Sigma_i v_{i2} = \text{ave}_1(X_{i1}) = \text{ave}_2(X_{i2}) ,$$

$$\text{var}_1(X_{i1}) = \text{var}_1(v_{i1}) + \text{var}_1(v_{i2}) + \text{Var}(e'_{i1}) ,$$

$$\text{var}_2(X_{i2}) = \text{var}_1(v_{i1}) + \text{var}_1(v_{i2}) + \text{Var}(e'_{i2}) ;$$

assuming for the sake of argument that the measures $\{v_{i2}\}$ have the distribution $N(0. s_2^2)$ and that the 60% group takes courses 3 , 5, and another, and that the 40% group takes courses 4 , 6 , and another, we have also

$$\text{ave}_3(X_{i3}) = \text{ave}_3(v_{i1}) + (0.5) \times \text{ave}\{\text{top } 60\% \text{ of } v_{i2}\} + 0$$

$$= \text{ave}_3(v_{i1}) + (0.5) \times (0.644 s_2) ,$$

$$\text{ave}_4(X_{i4}) = \text{ave}_4(v_{i1}) + (0.5) \times \text{ave}\{\text{top } 40\% \text{ of } - v_{i2}\} + 0$$

$$= \text{ave}_3(v_{i1}) + (0.5) \times (0.965 s_2) .$$

$$\text{var}_3(X_{i3}) = \text{var}_3(v_{i1}) + (0.5)^2 \times \text{var}\{\text{top } 60\ \% \text{ of } v_{i2}\} + \text{Var}(e'_{i3})$$

$$= \text{var}_3(v_{i1}) + (0.5)^2 \times (0.650 s_2)^2 + \text{Var}(e'_{i3}) ,$$

$$\text{var}_4(X_{i4}) = \text{var}_4(v_{i1}) + (0.5)^2 \times \text{var}(\text{top } 40\ \% \text{ of } - v_{i2}) + \text{Var}(e'_{i4})$$

$$= \text{var}_4(v_{i1}) + (0.5)^2 \times (0.560s_2)^2 + \text{Var}(e'_{i4})\,.$$

$$\text{cov}_{12}(X_{i1}, X_{i2}) = \text{cov}_{12}(v_{i1} + v_{i2} + e'_{i1}, v_{i1} - v_{i2} + e'_{i2})$$

$$= \text{var}_1(v_{i1}) - \text{var}_1(v_{i2})\,,$$

$$\text{cov}_{13}(X_{i1}, X_{i3}) = \text{cov}_{13}(v_{i1} + v_{i2} + e'_{i1}, \{v_{i1} + 0.5v_{i2} + e'_{i3}\text{: top 60\% of } v_{i2}\})$$

$$= \text{var}_3(v_{i1}) + (0.5) \times \text{var}_3(v_{i2})$$

$$= \text{var}_3(v_{i1}) + (0.5) \times (0.650s_2)^2\,,$$

$$\text{cov}_{24}(X_{i2}, X_{i4}) = \text{var}_4(v_{i1}) + (0.5) \times (0.560s_2)^2\,,$$

$$\text{cov}_{14}(X_{i1}, X_{i4}) = \text{cov}_{14}(v_{i1} + v_{i2} + e'_{i1}, \{v_{i1} - 0.5v_{i2} + e'_{i4}\text{: top 40\% of } -v_{i2}\})$$

$$= \text{var}_4(v_{i1}) - (0.5) \times \text{var}_4(v_{i2})$$

$$= \text{var}_4(v_{i1}) - (0.5) \times (0.560s_2)^2\,,$$

$$\text{cov}_{23}(X_{i2}, X_{i3}) = \text{var}_3(v_{i1}) - (0.5) \times (0.650s_2)^2\,,$$

$$\text{cov}_{35}(X_{i3}, X_{i5}) = \text{var}_3(v_{i1}) + (0.5)^2 \times (0.650s_2)^2\,,$$

$$\text{cov}_{46}(X_{i4}, X_{i6}) = \text{var}_4(v_{i1}) + (0.5)^2 \times (0.560s_2)^2\,.$$

To investigate the effect of using the method of moment estimation procedure on the data as though they conform to the one-factor model, consider the result of calculation after the first iterative step:

$$\text{ave}_1(X_{i\cdot}) = \text{ave}_1(v_{i1}) + (0.3) \times (0.773s_2)$$

$$= 0.232s_2 = \text{ave}_2(X_{i\cdot})\,,$$

$$\text{var}_1(X_{i\cdot}) = \text{var}_1(v_{i1}) + (0.3)^2 \times \text{var}\{\text{top 60\% of } v_{i2} \text{ and top 40\% of } -v_{i2}\} + \text{Var}(e'_{i\cdot})$$

$$= \mathrm{var}_1(v_{i1}) + (0.3)^2 \times (0.403s_2)^2 + \mathrm{Var}(e'_{i\cdot}) = \mathrm{var}_2(X_{i\cdot}) ,$$

$$\mathrm{ave}_3(X_{i\cdot}) = \mathrm{ave}_3(v_{i1}) + (0.3) \times (0.644s_2) ,$$

$$\mathrm{var}_3(X_{i\cdot}) = \mathrm{var}_3(v_{i1}) + (0.3)^2 \times (0.650s_2)^2 + \mathrm{Var}(e'_{i\cdot}) ,$$

$$\mathrm{ave}_4(X_{i\cdot}) = \mathrm{ave}_4(v_{i1}) + (0.3) \times (0.965s_2) ,$$

$$\mathrm{var}_4(X_{i\cdot}) = \mathrm{var}_4(v_{i1}) + (0.3)^2 \times (0.560s_2)^2 + \mathrm{Var}(e'_{i\cdot}) ,$$

$$\mathrm{cov}(X_{i\cdot}, X_{i1}) = \mathrm{var}_1(v_{i1}) + (0.3) \times \mathrm{cov}_1(\{\text{top } 60\% \ v_{i2} \text{ and top } 40\% \text{ of } -v_{i2}\}, v_{i2}) + (1/5)\mathrm{Var}(e'_{i1})$$

$$= \mathrm{var}_1(v_{i1}) + (0.3) \times (0.004s_2^2) + (0.2) \times \mathrm{Var}(e'_{i1}) ,$$

$$\mathrm{cov}(X_{i\cdot}, X_{i2}) = \mathrm{var}_1(v_{i1}) - (0.3) \times (0.004s_2^2) + (0.2) \times \mathrm{Var}(e'_{i2}) ,$$

$$\mathrm{cov}(X_{i\cdot}, X_{i3}) = \mathrm{var}_3(v_{i1}) + (0.15) \times \mathrm{cov}_3(v_{i2}, \{\text{top } 60\% \text{ of } v_{i2}\}) + (0.2) \times \mathrm{Var}(e'_{i3})$$

$$= \mathrm{var}_3(v_{i1}) + (0.15) \times (0.650s_2)^2 + (0.2) \times \mathrm{Var}(e'_{i3}) ,$$

$$\mathrm{cov}(X_{i\cdot}, X_{i4}) = \mathrm{var}_4(v_{i1}) + (0.15) \times (0.560s_2)^2 + (0.2) \times \mathrm{Var}(e'_{i4}) .$$

As the first iteration approximation to b_1 we have the ratio

$$[\mathrm{var}_1(v_{i1}) + 0.0361s_2^2 + \mathrm{Var}(e'_1)]/[\mathrm{var}_1(v_{i1}) + 0.001s_2^2 + (0.2)\mathrm{Var}(e'_{i1})] .$$

Assuming that $\mathrm{Var}(e'_1) \approx \mathrm{Var}(e'_{i1})/5 \approx s_2^2/5$ and that $s_2^2/\mathrm{var}_1(v_{i1}) \approx 1/5$, this ratio $\approx$ 1.007, indicating that the effect of assuming that (4.2) holds with X_{ij} rather than Y_{ij} introduces a bias that is smaller than any of the biases considered in connection with estimates of β_j in section 3. Making similar assumptions in connection with the other courses leads to ratios that are likewise within 1% of 1.00.

Within the simplified course choices and using the typical values of variances for the ratios just considered, the method of moment estimation procedure derived from a one-factor model produces estimators for the two-factor model that are somewhat smaller than the bias terms canvassed for other course score estimation procedures in section 3. Accordingly, on these theoretical grounds, the one-factor model constructed via method of moment estimation produces adequate estimators even for the two-factor model as above.

5 Reference test factor

It has long been known (Anastasi (1958) wrote of studies going back as far as 1929) that the relative performance of mid-teenage boys and girls on standardized tests such as SAT tests in USA differs from their relative performance under class-room assessment practices. Such differences would appear to be culturally based, or if not, a gender-linked interaction of the psyche with the mode of assessment in that multiple choice tests are predominantly used in standardized testing but not in most class-room based assessments. The presence of any such interaction is presumably not a gender trait per se but merely a gender-linked trait, in which case, if there exist at least two sets of pairs of assessments that can be regarded as being in similar areas, one from a standardized test or other multiple choice based test and the other from the classroom, then it should be possible to discern whether over all individuals it is feasible to postulate an analogue of (4.1) in the form

$$Y_{ij}(\delta_j) = v_{i1} + \gamma_j v_{i2} + \delta_j \Delta_i + e''_{ij} \tag{5.1}$$

where $\delta_j = +1$ or -1 and Δ_i denotes the relative performance of individual i as measured under two modes of assessment in course j.

Equation (5.1) has the consequence that if for example courses $j = 1$ and 2 have $\gamma_j = +1$ and -1 respectively, then the four sets of scores $\{Y_{i1}(1)\}$, $\{Y_{i1}(-1)\}$, $\{Y_{i2}(1)\}$, $\{Y_{i2}(-1)\}$, yield

$$\tfrac{1}{4}[Y_{i1}(1) + Y_{i2}(1) + Y_{i1}(-1) + Y_{i2}(-1)] = v_{i1} + e''_i(1), \tag{5.2}$$

$$\tfrac{1}{4}[Y_{i1}(1) - Y_{i2}(1) + Y_{i1}(-1) - Y_{i2}(-1)] = v_{i2} + e''_i(2), \tag{5.3}$$

$$\tfrac{1}{4}[Y_{i1}(1) + Y_{i2}(1) - Y_{i1}(-1) - Y_{i2}(-1)] = \Delta_i + e''_i(3), \tag{5.4}$$

$$\tfrac{1}{4}[Y_{i1}(1) - Y_{i2}(1) - Y_{i1}(-1) + Y_{i2}(-1)] = e''_i(4), \tag{5.5}$$

where on the assumption that the errors $e''_{i\cdot}$ are uncorrelated r.v.s with variances σ_1^2, ... σ_4^2 say, the sets of error terms $\{e''_i(r): i = 1, ..., N\}$ $(r = 1, ..., 4)$ are mutually uncorrelated with a common variance $\frac{1}{4}\sigma^2 \equiv (\sigma_1^2 + ... + \sigma_4^2)/16$. It is clear that some test of the model (5.1) is effected by forming the four linear contrasts (5.2)–(5.5) and finding their sums of squares, for which the expectations are

$$\text{var}_1(v_{i1}) + \tfrac{1}{4}\sigma^2,\ s_2^2 + \tfrac{1}{4}\sigma^2,\ \text{var}_1(\Delta_i) + \tfrac{1}{4}\sigma^2,\ \text{and}\ \tfrac{1}{4}\sigma^2\ \text{respectively}\,. \qquad (5.6)$$

Comparison of the observed mean squares with these expected mean squares, and in particular, that the last mean square is significantly smaller than any of the others, is evidence that (5.1) holds. Another test is effected by looking at the correlations of the sets of right-hand sides: near-zero correlations constitute additional evidence that (5.1) holds, being independent of the mean square evidence.

What is almost universally reported is that boys and girls differ in their relative abilities in the quantitative and verbal skill areas. In a report that admitted to having been written hastily, Masters and Beswick (1986) suggested and attempted to supply evidence that the gender-linked difference noted onwards from 1929 is attributable to an interaction of the relative participation rates of boys and girls in these two areas. This suggestion can be tested more thoroughly than in Masters and Beswick's analyses by using the model (5.1) in the following ways:

(1) Check the analyses based on (5.2)–(5.5) within each sex. If similar second-order properties are observed then it is evidence that the model (5.1) holds as a description of the scores of *individuals*, and that any systematic differences between subgroups formed on the basis of gender are merely gender-*linked* effects.

(2) Investigate the gender-difference of the averages of $\{Y_{ij}(1)\}$ and $\{Y_{ij}(-1)\}$ for each of $j = 1$ and 2 . If these gender-based differences are of similar sign and (better still) size for the two course areas, then it is evidence that the model-based averages of Δ_i within each sex are different. Moreover, they are not related to the verbal/quantitative contrast factors $\{v_{i2}\}$. (Such evidence was supplied to

the Committee that wrote *MATHEF* (1986) but not reported there.)

(3) Note in particular the correlations between the contrasts (5.3) and (5.4). If the gender-based differences observed as mode of assessment effects are attributable to verbal/quantitative contrast factors, then these correlations should differ from zero.

The presence of the factor $\{\Delta_i\}$ is of considerable concern for its effects, not only on the reference score equating procedure of section 3.5, but also when used in conjunction with any of the other course score estimation procedures of sections 3.2 to 3.4 in which scores such as $\{\tilde{V}_i\}$ at (3.38) are used as the scores of the particular course s for which $(\alpha_s, \beta_s) = (0,1)$ as at [1] above (3.10). This is particularly so whenever the mean squares $\sigma_\Delta^2(j) \equiv \text{var}_j(\Delta_i)$ differ considerably from the quantities σ_j^2 of (2.2) because the estimators b_j are affected by the ratio

$$[\text{var}_j(v_i) + \sigma_j^2]/[\text{var}_j(v_i) + \sigma_\Delta^2(j)] = (1 + 1/\Gamma_j)/(1 + 1/\Gamma_\Delta(j)) . \qquad (5.7)$$

So soon as $\Gamma_\Delta(j)$ is smaller than the general range 3 to 6 for Γ_j as at (3.21), distortion of the scale estimators b_j occurs and biases the contribution of the scores Y_{ij} from the courses j concerned. It is therefore appropriate to ensure that any sub-populations whose reference test scores are used in order to establish some form of comparability across groups which otherwise have vacuous common sub-candidatures $\mathscr{C}_{jk}$ $(j, k \neq s)$, have their ratios $\Gamma_\Delta(\cdot)$ (over the sub-population concerned) within the range 3 to 6. Put another way, the estimate of the mean square σ_Δ^2 within a sub-population can be considerably in excess of the purported measurement error associated with the reference test, and hence indicate a significant presence of mode of assessment differences $\{\Delta_i\}$; ***when this is so, it is essential to consider methods of reducing this observed mean square to the order of magnitude of the measurement error so as to comply with the fundamental assumption that (3.38) holds with measurement error only.***

References

ANASTASI, A. (1958). *Differential Psychology — Individual and Group Differences in Behaviour*, Third Ed. Macmillan, New York.

COONEY. G. H. (1975). Standardization procedures involving moderator variables — some theoretical considerations. *Aust. J. Educ.* **19**, 50–63.

COONEY, G. H. (1978). A critique of standardization by bivariate adjustment — a rejoinder. *Aust. J. Educ.* **22**, 323–325.

DALEY, D. J. (1985). How should NSW HSC examination marks be reported? *Independent Education* 15 (2) , 34–38.

DALEY, D. J. (1987). *Scaling NSW HSC Marks for School-leaver Admission. February 1987.* Report to the Canberra College of Advanced Education.

DALEY, D. J. (1988). Ranking in a one-factor model used to describe exam. marks. *Proc. Internat. Workshop/Seminar on Statistical Inference Procedures in Ranking and Selection*, Sydney, August 1987. American Sciences Press, Ohio (to appear).

DALEY, D. J. & EYLAND, E. A. (1987). The new and old HSC: Figures, facts and fantasies. *Independent Education* 17 (3) , 22–25.

DALEY, D. J. & SENETA, E. (1986). Modelling examination marks. *Aust. J. Statist.* **28**, 143–153.

HASOFER, A. M. (1978). A critique of standardization by bivariate adjustment. *Aust. J. Educ.* **22**, 319–322.

MASTERS, G. N. & BESWICK, D. G. (1986). The construction of tertiary entrance scores: principles and issues. Technical Report, Centre for the Study of Higher Education, University of Melbourne.

MATHEF (1986). *Making Admission to Higher Education Fairer.* Report of the Committee for the Review of Tertiary Entrance Score Calculations in the Australian Capital Territory. Australian Capital Territory Schools Authority, The Australian National University, and Canberra College of Advanced Education.

POTTHOFF, R. F. (1982). Some issues in test equating. In P. W. HOLLAND & D. B. RUBIN (Eds.), *Test Equating.* Academic Press, New York, 201–242.

SENETA, E. (1987). Report on the scaling of the 1986 New South Wales Higher School Certificate, University of Sydney.