Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

NORMAL MULTIVARIATE ANALYSIS OF FAMILIES OF REGRESSION COEFFICIENT VECTORS[†]

BY ALAN T. JAMES

University of Adelaide, South Australia

An exposition is given of results derived in James and Venables, Matrix Weighting of Several Regression Coefficient Vectors (1993). The results show that for small sample random effects models, an estimated random effects variance matrix may be used in weight matrices without causing undue error in the weighted mean. Exact error variances are quoted for a mean with estimated weights for the two sample case in one and two dimensions. Simulation is used to determine errors for a practical example of six 5-variate samples. A curious range anomaly is illustrated which arises if random effects are ignored when present.

1. Introduction. The random effects model of Henderson *et al.* (1959) can combine the results of $p+1$ similar regressions by specifying that the regression parameter vectors, $\boldsymbol{\beta}_i \in R^n$, are random and multinormally distributed, $\boldsymbol{\beta}_i \in \mathrm{N}(\boldsymbol{\beta}_0, \boldsymbol{\Delta})$. A sample regression vector, $\boldsymbol{b}_i, i = 1, \cdots, p+1$, then has a conditional distribution, $\boldsymbol{b}_i | \boldsymbol{\beta}_i \sim \mathrm{N}(\boldsymbol{\beta}_i, \boldsymbol{\Gamma}_i)$, and a marginal distribution, $\boldsymbol{b}_i \sim \mathrm{N}(\boldsymbol{\beta}, \boldsymbol{\Gamma}_i + \boldsymbol{\Delta})$. The Maximum Likelihood, ML, estimate of $\boldsymbol{\beta}_0$ is then a matrix weighted mean of the $\boldsymbol{b}_i$ with weights, $(\boldsymbol{\Gamma}_i + \boldsymbol{\Delta})^{-1}$,

$$\hat{\boldsymbol{\beta}}_0 = \left(\sum_{i=1}^{m} (\boldsymbol{\Gamma}_i + \boldsymbol{\Delta})^{-1} \right)^{-1} \sum_{i=1}^{m} (\boldsymbol{\Gamma}_i + \boldsymbol{\Delta})^{-1} \boldsymbol{b}_i$$

If the variance matrices, $\boldsymbol{\Gamma}_i$, of the $\boldsymbol{b}_i$ are all equal, that is if the data is *balanced*, then the weights are all equal and the ML estimate of $\boldsymbol{\beta}_0$ is simply the average of the $\boldsymbol{b}_i$,

$$\hat{\boldsymbol{\beta}}_0 = \bar{\boldsymbol{b}}$$

For unbalanced data, the ML estimate, $\hat{\boldsymbol{\beta}}_0$, will depend upon the between regressions variance matrix, $\boldsymbol{\Delta}$, to the extent of the imbalance. Since there is

---

[†] Research supported by Australian Research Council Grant A6 8931380
AMS 1980 Subject Classifications: Primary 62H12, Secondary 62J10.
Key words and phrases: Conditional bias, cutoff function, efficiency, estimated generalized least squares, matrix weighting, moment estimator, random effects model, range anomaly, residual maximum likelihood, small sample random effects model, simulation, unbalanced data.

usually no prior information about Δ, an estimate, $\hat{\Delta}$, has to be substituted for Δ, in computing a weighted mean,

$$\tilde{\beta}_0 = \left(\sum_{i=1}^{m} (\Gamma_i + \widehat{\Delta})^{-1} \right)^{-1} \sum_{i=1}^{m} (\Gamma_i + \widehat{\Delta})^{-1} \boldsymbol{b}_i$$

The sampling error of $\widehat{\Delta}$ naturally increases the error variance of the weighted mean, $\tilde{\beta}_0$, over var($\hat{\beta}_0$), but it is generally very difficult to evaluate by how much. When the number of regressions, $p+1$, is small, the sampling error of $\widehat{\Delta}$ will be so large that authors such as Gumpertz and Pantula (1989) have concluded that a weighted mean based upon it would be unreliable and advocated the sample average, $\bar{\boldsymbol{b}}$, or in some cases, on the assumption that $\Delta = 0$, the use of Γ_i^{-1} as weights.

The present paper presents theory and simulations that demonstrate in some cases, and suggest in general, that the prejudice against the use of $\widehat{\Delta}$ in weights, however few the regressions, $p+1$, is unfounded. The paper also illustrates a grave range anomaly that can arise in a weighted mean ignoring Δ. This method, which weights by Γ_i^{-1} is equivalent to an Ordinary Least Squares, OLS, analysis of the pooled data vector, $\boldsymbol{y}$, of the $p+1$ sets of original observations. The method will be subsequently referred to as OLS.

A reason why errors in the weights are not as serious as one would suppose at first sight, is that errors in the $\widehat{\Delta}$ and $\boldsymbol{b}_i$ are to some extent mutually compensating. If for some contrast, $\boldsymbol{\lambda}'\boldsymbol{\beta}$, the between regressions variance component, $\boldsymbol{\lambda}'\Delta\boldsymbol{\lambda}$, is fortuitously underestimated by $\boldsymbol{\lambda}'\widehat{\Delta}\boldsymbol{\lambda}$, then this implies that the $\boldsymbol{\lambda}'\boldsymbol{b}_i$ do not vary much and consequently the weighting is inconsequential. On the other hand, if $\boldsymbol{\lambda}'\widehat{\Delta}\boldsymbol{\lambda}$ overestimates $\boldsymbol{\lambda}'\Delta\boldsymbol{\lambda}$, then the weighting is moving towards the equality which yields $\boldsymbol{\lambda}'\bar{\boldsymbol{b}}$. As a result of the compensation, the error of $\tilde{\boldsymbol{\beta}}$ is not as great as if the errors in $\widehat{\Delta}$ and the $\boldsymbol{b}_i$ were independent.

Models for comparison of two or more mean vectors, or more general linear models can be built up, but the inferential issues with which we are concerned are the same as for the estimation of a single mean vector, $\boldsymbol{\beta}$, and between regressions variance component matrix, Δ. Such models were expressed in an algebraically equivalent form in terms of the original observations, $\boldsymbol{y}$, by Henderson *et al.* (1959), namely,

$$\boldsymbol{y} = X\boldsymbol{\beta}_0 + Z\boldsymbol{b} + \varepsilon,$$

where $\boldsymbol{b} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{D})$ and $\varepsilon \sim \mathrm{N}(\mathbf{0}, \Gamma)$. The weighted mean vector, $\hat{\beta}_0$, is then obtained by Generalized Least Squares, GLS, and if an estimate $\widehat{\Delta}_0$ is substituted for Δ to yield, $\tilde{\boldsymbol{\beta}}$, the method is called Estimated GLS, EGLS.

The relative merit of $\bar{\boldsymbol{b}}$ and $\tilde{\boldsymbol{\beta}}$ can be judged by comparing their error variance matrices. For more than two samples, $p > 1$, the error variance matrix, var($\tilde{\boldsymbol{\beta}}$), of $\tilde{\boldsymbol{\beta}}$ seems too complicated to compute mathematically, when one allows for the effect of the error of $\widehat{\boldsymbol{\Delta}}$, because this Residual Maximum Likelihood, REML, estimator has to be computed by iteration. By simulation, however, one can estimate var($\tilde{\boldsymbol{\beta}}$), for any particular value of $\boldsymbol{\Delta}$, and compare it with var($\bar{\boldsymbol{b}}$) as follows.

One can take $\boldsymbol{\beta} = \boldsymbol{0}$ and compute a large number, N, of independent estimates, $\tilde{\boldsymbol{\beta}}_i$, of it from random numbers. Their crude mean square, $\sum_{i=1}^{N} \tilde{\boldsymbol{\beta}}_i \tilde{\boldsymbol{\beta}}_i'/N$, is an estimate of var($\tilde{\boldsymbol{\beta}}$) on N degrees of freedom.

For speed of simulation, the moment estimator, $\boldsymbol{\Delta}^{(m)}$ of $\boldsymbol{\Delta}$ was used instead of the REML estimator, $\widehat{\boldsymbol{\Delta}}$. Since the coefficient of variation of an estimated variance is $\sqrt{2/N}$, three figure accuracy requires up to $N = 2$ million simulations and two figure accuracy 20,000. And this is only for each possible value of $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p+1}$. Consequently, the simulation needs to be supplemented, if possible, by theory which gives a broad picture of the relative accuracies of $\bar{\boldsymbol{b}}$ and $\tilde{\boldsymbol{\beta}}$ over all the range of $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p+1}$.

At present the mathematical and computing problems seem tractable only for the case of two samples, $p = 1$, and have been worked out only for the scalar case, $n = 1$, and the two dimensional vector case, $n = 2$. The explicit theory for the two cases, nevertheless, sheds much light on the general situation.

Firstly, since the sampling error of $\widehat{\boldsymbol{\Delta}}$ increases with decreasing p, the case $p = 1$ is the most unfavourable. We shall see that var($\tilde{\boldsymbol{\beta}}$) $<$ var($\bar{\boldsymbol{b}}$) for small $\boldsymbol{\Delta}$, in the sense of positive definite matrices, and only negligibly greater for large $\boldsymbol{\Delta}$. This result in the least favourable case, $p = 1$, gives one confidence in the estimate $\tilde{\boldsymbol{\beta}}$ for larger p.

Secondly, although even in the scalar case, var($\tilde{\beta}$) depends upon three parameters, Γ_1, Γ_2 and Δ, one can compare the magnitude of var($\tilde{\beta}$) with var($\bar{b}$) by means of a single function, $\Delta/(\Gamma_1 + \Gamma_2)$, of the three parameters.

In the two dimensional vector case, the theory points to two extreme forms of imbalance:

a. scalar imbalance in which $\boldsymbol{\Gamma}_1$ is proportional to $\boldsymbol{\Gamma}_2$ but the eigenvalues of $\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2$ are either both large or small.
b. eccentric imbalance in which one eigenvalue of $\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2$ is large and the other is small. In the case when this is due to numerically large correlations of opposite sign, the OLS method can lead to a serious range anomaly, as we shall illustrate.

2. Two Vectors.

For two vectors, $\boldsymbol{b}_1, \boldsymbol{b}_2$, with $p = 1$, the REML estimator, $\widehat{\boldsymbol{\Delta}}$, is the moment estimator, $\boldsymbol{\Delta}^{(m)}$, given explicitly by

$$\hat{\boldsymbol{\Delta}} = \begin{cases} 0 & \text{if } \chi^2 \leq 1, \\ (\chi^2 - 1)\boldsymbol{d}\,\boldsymbol{d}'/(2\chi^2) & \text{if } \chi^2 \geq 1, \end{cases} \tag{2.1}$$

where $\boldsymbol{d} = \boldsymbol{b}_2 - \boldsymbol{b}_1, \chi^2 = \boldsymbol{d}'\boldsymbol{\Gamma}^{-1}\boldsymbol{d}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2$.

The mean with estimated weights is

$$\tilde{\boldsymbol{\beta}} = \bar{\boldsymbol{b}} - \boldsymbol{\Phi}\boldsymbol{c}(\boldsymbol{d}), \tag{2.2}$$

where, what we shall call, the imbalance factor is

$$\boldsymbol{\Phi} = (\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_1)\boldsymbol{\Gamma}^{-1} \tag{2.3}$$

and cutoff function is

$$\boldsymbol{c}(\boldsymbol{d}) = \begin{cases} \boldsymbol{d}/2 & \text{if } \chi^2 \leq 1, \\ \boldsymbol{d}/(2\chi^2) & \text{if } \chi^2 \geq 1. \end{cases} \tag{2.4}$$

The variance of $\tilde{\boldsymbol{\beta}}$ is

$$\operatorname{var}(\tilde{\boldsymbol{\beta}}) = \operatorname{var}(\bar{\boldsymbol{b}}) + \boldsymbol{\Phi}\, \boldsymbol{V}_k \boldsymbol{\Phi}', \tag{2.5}$$

where $\boldsymbol{V}_k$ is the variance kernel, given by

$$\boldsymbol{V}_k = \operatorname{var}(\boldsymbol{c}) - \frac{1}{2}\boldsymbol{\Lambda}^{-1}\mathrm{E}[\boldsymbol{d}\boldsymbol{c}'] - \frac{1}{2}\mathrm{E}[\boldsymbol{c}\boldsymbol{d}']\boldsymbol{\Lambda}^{-1}, \tag{2.6}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Gamma}_d\boldsymbol{\Gamma}^{-1}$ with $\boldsymbol{\Gamma}_d = \operatorname{var}(\boldsymbol{d}) = \boldsymbol{\Gamma} + 2\boldsymbol{\Delta}$, and $\operatorname{var}(\bar{\boldsymbol{b}}) = \boldsymbol{\Gamma}_d/4$.

The variance kernel, $\boldsymbol{V}_k$, expresses the difference of the variances of $\tilde{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{b}}$ in the most extreme cases of imbalance. Its positive or negative definiteness and magnitude therefore express their relative accuracies.

2.1. *The scalar case*

When $n = 1$, we have

$$\tilde{b} = \bar{b} - \Phi c(d),$$

where the imbalance factor is $\Phi = (\Gamma_2 - \Gamma_1)/\Gamma$ and the cutoff function is

$$c(d) = \begin{cases} d/2 & \text{if } |d| \leq \sqrt{\Gamma}, \\ \Gamma/(2d) & \text{if } |d| \geq \sqrt{\Gamma}. \end{cases}$$

Its graph is given in Figure 1.

Since

$$\left(\operatorname{var}(\tilde{b}) - \operatorname{var}(\bar{b})\right)/\operatorname{var}(\bar{b}) = \Phi^2 V_k/\operatorname{var}(\bar{b}),$$

where the variance kernel is

$$V_k = \operatorname{var}(c) - \mathrm{E}[dc]/\Lambda$$

the variances of $\tilde{b}$ and $\bar{b}$ can be compared by studying the function

$$f(\Delta/\Gamma) = V_k/\operatorname{var}(\bar{b}).$$

The function f can be explicitly evaluated in terms of exponential functions and normal integrals. Its graph is shown in Figure 2.

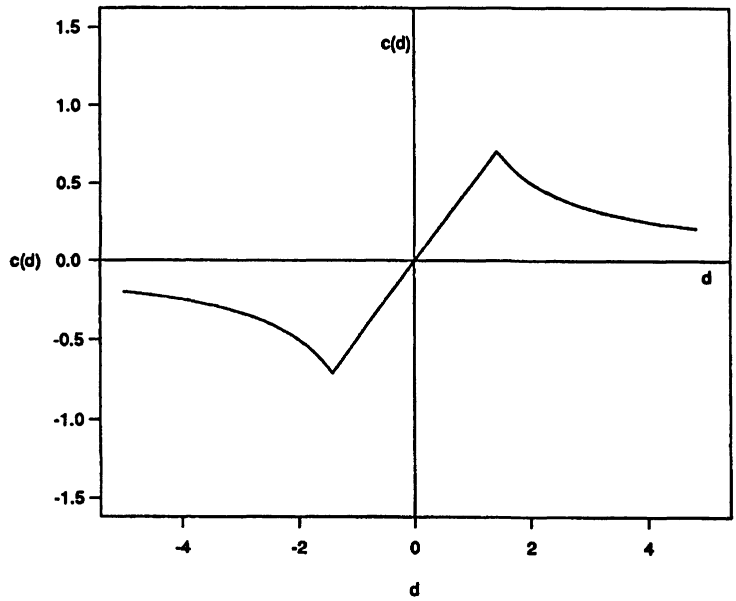

Figure 1. The cutoff function, $c(d)$. For two-sample scalar random effects models c multiplied by the imbalance factor, $(\Gamma_2 - \Gamma_1)/(\Gamma_2 + \Gamma_1)$, gives the correction substracted from the average, $\bar{b}$, to give the empirically weighted mean, $\tilde{\beta}$

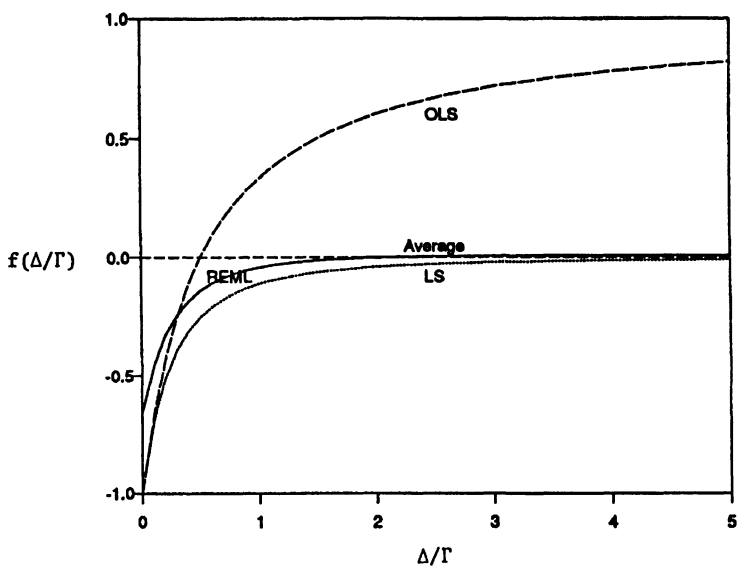

Figure 2. Difference of variance estimators as a percentage of the variance of the average

For comparison with other estimators, m, such as the least squares estimator, $m = \hat{\beta}$, and the estimator, $m = \beta^+ = (\Gamma_1^{-1}\beta_1 + \Gamma_1^{-1}\beta_2)/(\Gamma_1^{-1} + \Gamma_2^{-1})$, based on the assumption $\Delta = 0$, we also plot the functions corresponding to f, namely,

$$\left(\left(\frac{\operatorname{var}(m) - \operatorname{var}(\bar{b})}{\operatorname{var}(\bar{b})}\right) \Big/ \Phi^2\right) \times 100\,\%.$$

Not knowing Δ, we cannot compute the least squares estimator, $\hat{\beta}$, but its variance constitutes a lower bound for the variances of all estimators. The abscissa in Figure 2 constitutes the curve for $m = \bar{b}$.

For $\Delta/\Gamma \leq 2$, the empirically weighted mean, $\tilde{\beta}$, has lower variance than the equally weighted mean, $\bar{b}$ and the difference is substantial for $\Delta/\Gamma < 1$. For $\Delta/\Gamma > 2$, $\tilde{\beta}$ has a slightly higher variance than $\bar{b}$ but from a practical point of view, the difference is negligible. Hence estimated weighting is to be recommended.

For Δ/Γ from 0 to 0.3, one would do best by weighting without random effects. Unless one has definite information that $\Delta/\Gamma \leq 0.3$, however, one cannot be sure of this and estimated weighting should be used.

At about $\Delta/\Gamma = 1.5$, there is 50:50 chance that random effects will be significant, inferring that $\Delta > 0$. Hence above 1.5, it is clear that random effects must be specified, but in this range, $\bar{b}$ has about the same accuracy as $\tilde{\beta}$.

For $0.5 < \Delta/\Gamma < 1.5$, there is low power in the significance test, but the weighted mean without random effects, β^+ is highly inefficient. This is the dangerous region if one ignores Δ in the hopes that there are no random effects. Such a procedure is highly nonrobust. If the data is completely unbalanced with Γ_2/Γ_1 large, then if one is certain of no random effect, that is that $\Delta = 0$, one can ignore b_2 and use b_1 as the estimator of β. It has a negligible variance, and hence is 100% below the average, $\bar{b} = (b_1 + b_2)/2$. In the absence of certain knowledge of no random effect, it is dangerous to ignore its possibility. If one allows for a possible random effect when in fact $\Delta = 0$, one pays a penalty that the variance of $\tilde{\beta}$ is only decreased by 70% of $\bar{b}$ instead of 100%.

Within the range $0 < \Delta/\Gamma < 1.5$, the empirically weighted mean, $\tilde{\beta}$, has less error variance than the average, $\bar{b}$, and substantially less in the lowest part of this range. The results confirm the use of estimated weighting.

2.2. *The two dimensional case*

In the case of two vectors $\boldsymbol{b}_1, \boldsymbol{b}_2 \in R^2$, $p = 1$, $n = 2$, we can compare the error variance matrices of the mean $\tilde{\boldsymbol{\beta}}$ obtained from estimated weights with the average, $\bar{\boldsymbol{b}}$, by making a nonsingular linear transformation

$$\boldsymbol{b}_i \to L\boldsymbol{b}_i, \qquad \tilde{\boldsymbol{\beta}} \to L\tilde{\boldsymbol{\beta}}, \qquad \bar{\boldsymbol{b}} \to L\bar{\boldsymbol{b}}$$

such that the induced congruence transformations map $\mathbf{\Gamma}$ to I and diagonalize $\mathbf{\Delta}$

$$\mathbf{\Gamma} \to L\mathbf{\Gamma}L' = I, \qquad \mathbf{\Delta} \to L\mathbf{\Delta}L' = \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix}.$$

The variance kernel, V_k, given by equation (2.6) undergoes the same congruence transformation and becomes a diagonal matrix

$$\begin{bmatrix} z_1(\delta_1, \delta_2) & 0 \\ 0 & z_2(\delta_1, \delta_2) \end{bmatrix}$$

which is a function of δ_1, δ_2 with $z_2(\delta_1, \delta_2) = z_1(\delta_2, \delta_1)$.

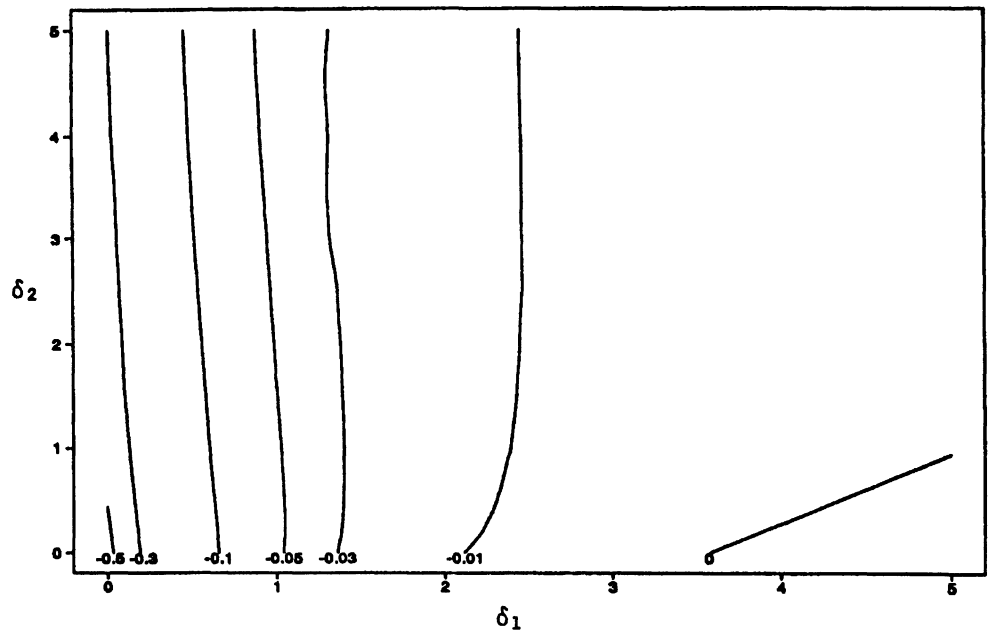

Figure 3. Bivariate case (a). Contours for the function $f_1(\delta_1, \delta_2)$ given by (2.7)

There are two extreme cases of imbalance:

a) Scalar imbalance due to different but proportionate variance matrices.

$$\mathbf{\Gamma}_1 = 0, \qquad \mathbf{\Gamma}_2 = I, \qquad \mathbf{\Gamma}_2 - \mathbf{\Gamma}_1 = I.$$

This is analogous with the scalar case. Then

$$f_1(\delta_1, \delta_2) = \left(\operatorname{var}(\tilde{\beta}_1) - \operatorname{var}(\bar{b}_{.1})\right)/\operatorname{var}(\bar{b}_{.1}) = 4z_1(\delta_1, \delta_2)/(1 + 2\delta_1). \qquad (2.7)$$

b) Eccentric imbalance which can be produced by difference of correlation.

$$\mathbf{\Gamma}_1 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}, \qquad \mathbf{\Gamma}_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \qquad \mathbf{\Gamma}_2 - \mathbf{\Gamma}_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

This is the case where OLS, which ignores $\boldsymbol{\Delta}$, leads to range anomaly. We have

$$f_2(\delta_1, \delta_2) = \frac{\operatorname{var}(\tilde{\beta}_1) - \operatorname{var}(\bar{b}_{.1})}{\operatorname{var}(\bar{b}_{.1})} = \frac{4z_2(\delta_1, \delta_2)}{1 + 2\delta_1} = \frac{4z_1(\delta_2, \delta_1)}{1 + 2\delta_1}. \qquad (2.8)$$

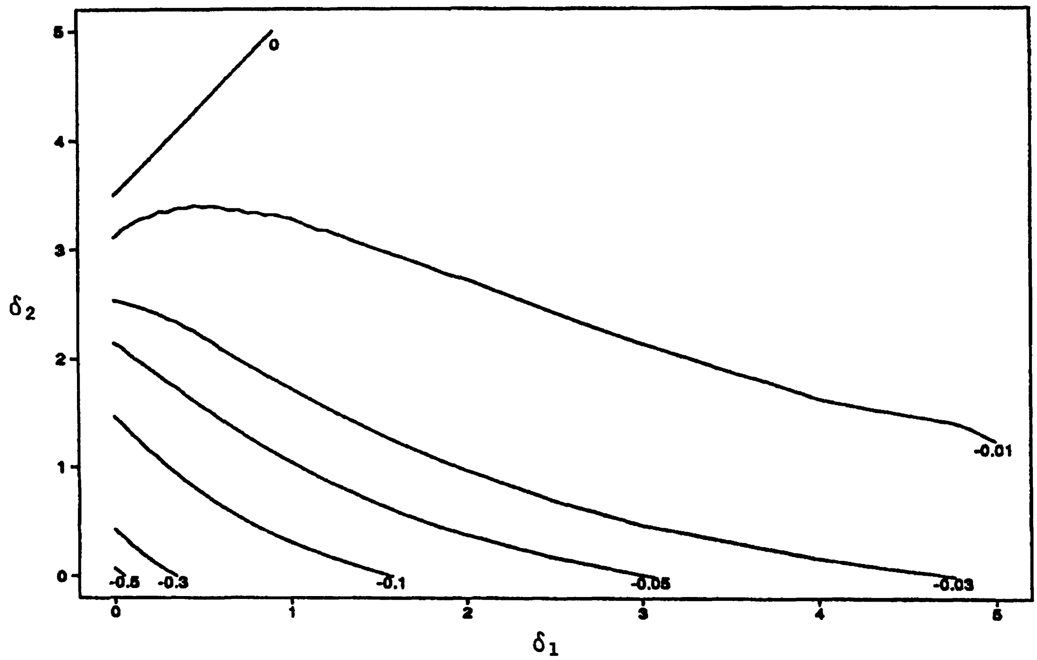

Figure 4. Bivariate case (b). Contours for the function $f_2(\delta_1, \delta_2)$ given by (2.8)

The function z_1 was evaluated using *Maple*. Contour diagrams of f_1 and f_2 are shown in Figures 3 and 4. For small δ_1, δ_2, f_1 and f_2 are highly negative, showing that in these regions $\tilde{\boldsymbol{\beta}}$ is substantially more accurate than $\bar{\boldsymbol{b}}$. For large δ_1, δ_2 the functions f_1 and f_2 become positive but not large.

3. An Example from Mitochondrial Experiments. Mitochondria are numerous organelles within the cells of plants and animals which generate the aerobic power. James, Wiskich and Conyers (1989, 1993) performed six experiments to test a model. The results were fitted by nonlinear regression giving rise to 6 sample regression vectors, five of whose components pertaining to the mitochondria, are given as the row vectors, $\boldsymbol{b}_i'$, $i = 1, 2, \cdots, 6$, of Table

1 with accurately estimated standard errors. The imbalance of the data lies in the large differences between the correlation matrices, and between the standard errors for the 6 experiments.

Table 1. Six independent row vectors of five regression coefficients with their standard errors, and four types of mean vectors. e_1, e_2, e_3 are potentials and r_1, r_2 are resistances measuring mitochondrial performance

Rat number, i	e_1 ln units	e_2 ln units	e_3 ln units	$100r_1$ ln units$\%^{-1}$	$100r_2$ ln units$\%^{-1}$
Hexokinase					
1	4.249±0.055	4.786±0.065	5.538±0.080	0.775±0.144	1.078±0.056
2	4.380±0.048	4.985±0.057	5.401±0.064	1.408±0.158	1.072±0.078
3	4.556±0.090	5.145±0.095	5.913±0.112	1.943±0.303	1.176±0.107
4	4.424±0.066	4.976±0.073	5.567±0.079	0.864±0.172	1.023±0.068
5	4.537±0.047	5.016±0.057	5.700±0.069	1.292±0.118	0.794±0.044
6	4.391±0.038	4.899±0.043	5.370±0.048	0.861±0.118	1.076±0.046
Mean Vector (Weight)					
Average (I)	4.423±0.046	4.968±0.049	5.582±0.082	1.191±0.184	1 .037±0.053
Simulated Error	0.042	0.043	0.071	0.164	0.057
Matrix (Γ_i^{-1})	4.283±0.019	4.780±0.022	5.348±0.025	0.842±0.056	0.945±0.023
REML $((\Gamma_i + \hat{\Delta})^{-1})$	4.407±0.041	4.946±0.041	5.557±0.070	1.144±0.161	1.027±0.055
Moment $((\Gamma_i + \Delta^{(m)})^{-1})$	4.415±0.046	4.954±0.049	5.567±0.082	1.164±0.182	1.031±0.052
Simulated Error	0.042	0.043	0.071	0.161	0.056

A biological interpretation of these results requires the 6 regression coefficient vectors to be summarized by a single vector. For a coherent multivariate analysis, matrix weights given by the inverses of the variance matrices should be used. When this was done using matrix weights Γ_i^{-1}, because the estimate of Δ was so much subject to error, the matrix weighted means shown in Table 1 were obtained whose second and third components lay outside the range of the 6 values of which they are supposed to form a summary! This is what we refer to as the potential *range anomaly* of OLS.

In Section 4, we illustrate by an artificial example just how it comes about, and how it does not appear to occur under the more realistic random effects model.

4. Matrix-Weighted Mean Displacement. Suppose two independent bivariate sample vectors had the following values with known variance matrices given by

$$\boldsymbol{b}_1 = \begin{bmatrix} 0 \\ a \end{bmatrix}, \quad \boldsymbol{\Gamma}_1 = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}, \qquad \boldsymbol{b}_2 = \begin{bmatrix} 0 \\ -a \end{bmatrix}, \quad \boldsymbol{\Gamma}_2 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

It looks highly artificial to take both observed abscissae as zero, but it makes the figure described below easier to comprehend. One could rotate it a little and displace it from zero without altering the essential argument.

We begin by treating this 'data' by OLS which ignores random effects, to illustrate the consequences.

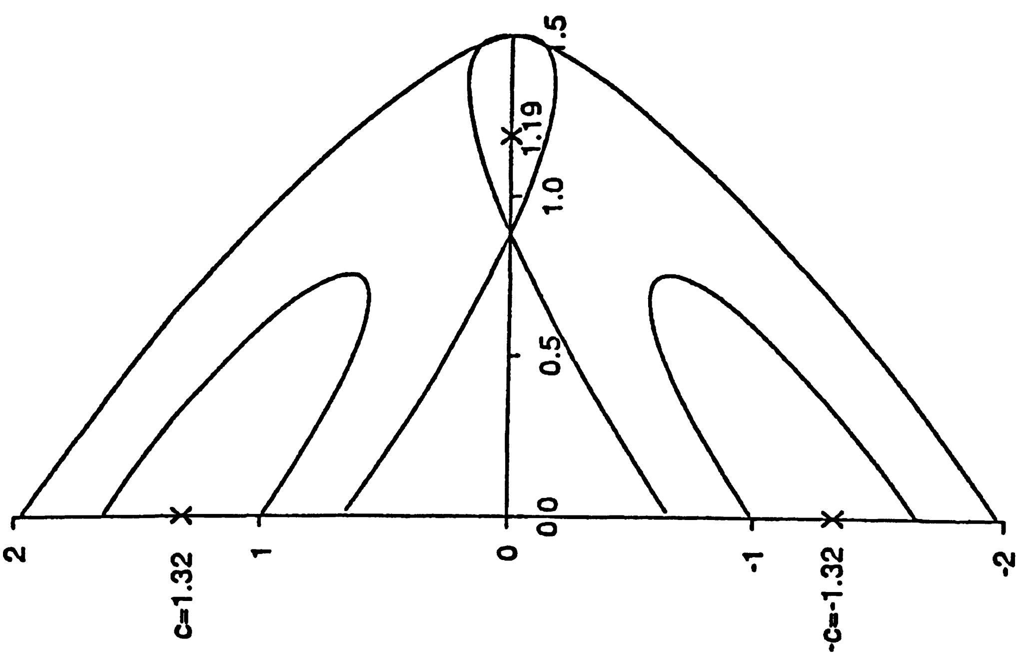

Figure 5. Likelihood contours for two separate bivariate likelihoods

Figure 5 shows the contours of the separate likelihood functions for the expectation vector, obtained from each observed vector. One sees how two ridges of high likelihood extend out from the observed vectors when the correlations are numerically large, and how, when the correlations are very unequal, the ridges will intersect in a region of high product of their likelihoods. If it can be assumed that the two observed vectors have a common expectation vector, its likelihood function is the product of their likelihoods. This function is illustrated graphically by the surface shown in Figure 6. Hence one can see how a matrix weighted mean can be well away from a scalar weighted mean, due to the first of the two variates having a strong covariance on the second.

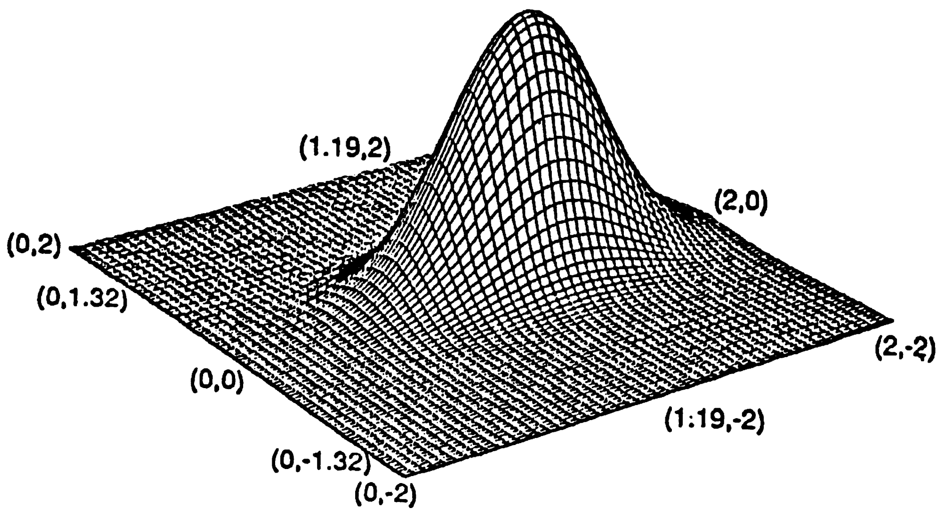

Figure 6. Likelihood surface given by the product of the likelihoods in Figure 5

If it is reasonable to assume that the expectations of the two vectors are equal, and if the region of the intersection of the two ridges of high likelihood is well within the two confidence intervals of the expectation obtained from the individual observed vectors, then the matrix weighted mean is acceptable. The considerable displacement and high accuracy are due to the extra information from the strong covariance.

On the other hand, if the observed vectors differ significantly as tested by

$$\chi_2^2 = \boldsymbol{d}' \, (\boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2)^{-1} \boldsymbol{d} = 2a^2 ,$$

where $\boldsymbol{d} = \boldsymbol{b}_2 - \boldsymbol{b}_1$, then the assumption underlying OLS is significantly rejected. A random effects model, however, will fit the 'data'. A variance component matrix must be estimated even though it is only on one degree of freedom. It is given for the n dimensional case by equation (2.1).

We now come to a common situation in which random effects may possibly be zero or negligible but this cannot be assumed. Suppose there is no *à priori* certainty that the expectations are equal, but χ_2^2 is below significance, as for example, $\chi_2^2 = 3.5$ when $a = \sqrt{3.5/2} = 1.32$. If, in an attempt to use OLS, we persist in specifying a model of equal expectations on the basis that the estimates do not differ significantly at the 5% level, then the ML estimate of the mean vector, with standard errors, is

$$\boldsymbol{\beta}^{+} = \begin{bmatrix} 1.19 \pm 0.31 \\ 0 \ \pm 0.31 \end{bmatrix}.$$

On this specification, the first element is highly significantly different from zero. But this inference depends entirely upon the assumption of no random effect. If the assumption is doubtful, the inference is correspondingly dubious.

If on the other hand one specifies a random effects model, then the estimated variance component matrix is

$$\widehat{\boldsymbol{\Delta}} = \frac{1}{2}\left(1 - \frac{1}{2a^2}\right)\begin{bmatrix} 0 & 0 \\ 0 & 4a^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 5/2 \end{bmatrix}$$

giving a weighted mean of $\hat{\boldsymbol{\beta}} = \left[\begin{smallmatrix} 0.34 \pm 0.62 \\ 0 \ \pm 1.16 \end{smallmatrix}\right]$ if we ignore the sampling error of $\widehat{\boldsymbol{\Delta}}$. If we allow for the sampling error of $\widehat{\boldsymbol{\Delta}}$ according to theory developed in Section 2.2 but evaluate the error at $\boldsymbol{\Delta} = \widehat{\boldsymbol{\Delta}}$, we obtain

$$\tilde{\boldsymbol{\beta}} = \begin{bmatrix} 0.34 \pm 0.67 \\ 0 \ \pm 1.25 \end{bmatrix}.$$

This seems a far more reasonable inference. This weighted mean is the value at which the likelihood surface for the random coefficients model, shown in Figure 7, has its maximum. The surface also shows high likelihood at $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$. At $\boldsymbol{\Delta} = \widehat{\boldsymbol{\Delta}}$, this estimate, $\tilde{\boldsymbol{\beta}}$, has a smaller error than the average

$$\bar{\boldsymbol{b}} = \begin{bmatrix} 0 \pm 0.71 \\ 0 \pm 1.32 \end{bmatrix}.$$

The range anomaly comes about by treating data with a nonzero $\boldsymbol{\Delta}$ as if it were zero. But $\boldsymbol{\Delta}$ is not the direct cause of the range anomaly, it is an indirect cause. $\boldsymbol{\Delta}$ operates only through a large difference, $\boldsymbol{d} = \boldsymbol{b}_2 - \boldsymbol{b}_1$, which is reflected in $\widehat{\boldsymbol{\Delta}}$ as a function of $\boldsymbol{d}$. If $\boldsymbol{d}$ is fortuitously small, then there is no range anomaly when $\boldsymbol{\Delta}$ is ignored. Likewise, when the use of weights ignoring $\boldsymbol{\Delta}$ produces a range anomaly, it is $\widehat{\boldsymbol{\Delta}}$ as a function $\boldsymbol{d}$ which rectifies it. In dimensions in which $\widehat{\boldsymbol{\Delta}}$ is singular, there are no differences to create a range anomaly.

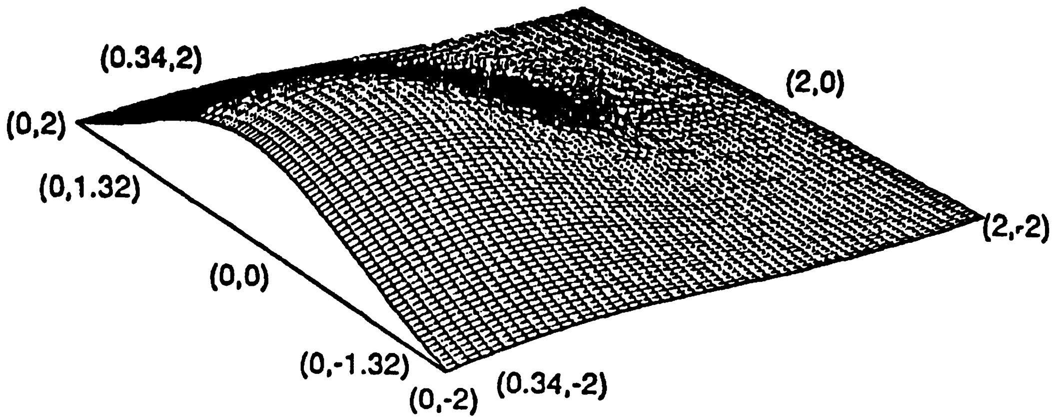

Figure 7. Likelihood surface for the parameter vector in the random coefficients model

On the other hand, if $\hat{\boldsymbol{\Delta}}$ fortuitously overestimates $\boldsymbol{\Delta}$, then the weights move conservatively towards the equality which gives the average, $\bar{\boldsymbol{b}}$.

These considerations led us to question a widely held belief that large errors in $\widehat{\boldsymbol{\Delta}}$ would produce excessive errors if it was used in weighting of means, and hence to investigate mathematically what would otherwise be regarded as the hopelessly inaccurate situation of two samples, which fortunately was mathematically tractable.

5. Simulation. For the mitochondrial data, $N = 600000$ simulations were done, at $\boldsymbol{\Delta} = \widehat{\boldsymbol{\Delta}}$,

The variances are given in Table 2.

One sees that the variances of the components of the average vector, $\bar{\boldsymbol{b}}$, are about 2-3% higher than for the moment estimator, $\boldsymbol{\beta}^{(m)}$, which in turn is about the same amount higher than for the least squares estimator, $\hat{\boldsymbol{\beta}}$.

The variance matrices, $\mathrm{var}(\bar{\boldsymbol{b}})$ and $\mathrm{var}(\boldsymbol{\beta}^{(m)})$ can be compared by the eigenvalues of the inverse of the first times the second, because if $\boldsymbol{a}$ is an eigenvector corresponding to the eigenvalue, λ, that is $\mathrm{var}(\bar{\boldsymbol{b}})\boldsymbol{a} = \mathrm{var}(\boldsymbol{\beta}^{(m)})\boldsymbol{a}\lambda$, then $\lambda = (\boldsymbol{a}'\mathrm{var}(\boldsymbol{\beta}^{(m)})\boldsymbol{a})/(\boldsymbol{a}'\mathrm{var}(\bar{\boldsymbol{b}})\boldsymbol{a})$ is the ratio of the variances of the estimates of the contrast, $\boldsymbol{a}'\boldsymbol{\beta}$, that is the efficiency of $\boldsymbol{a}'\bar{\boldsymbol{b}}$ relative to $\boldsymbol{a}'\boldsymbol{\beta}^{(m)}$. The absolute efficiencies are found by eigenvalues relative to the least squares estimator, $\hat{\boldsymbol{\beta}}$, as in Table 3.

Table 2. Variances of estimators × 100

Theoretical Least Squares	0.1672	0.1709	0.4942	2.5988	0.3040
Average	0.1777	0.1817	0.5089	2.6778	0.3241
Simulated Least Squares	0.1677	0.1714	0.4945	2.6057	0.3050
Moment	0.1734	0.1776	0.5031	2.6498	0.3151
Average	0.1784	0.1825	0.5097	2.6871	0.3254

Table 3. Eigenvalues measuring efficiencies and relative effici encies

$\text{var}(\bar{\boldsymbol{b}})^{-1}\text{var}(\boldsymbol{\beta}^{(m)})$	0.869	0.965	0.985	1.000	0.999
$\text{var}(\boldsymbol{\beta}^{(m)})^{-1}\text{var}(\hat{\boldsymbol{\beta}})$	0.853	0.950	0.980	0.998	0.999
$\text{var}(\bar{\boldsymbol{b}})^{-1}\text{var}(\hat{\boldsymbol{\beta}})$	0.741	0.916	0.965	0.997	0.999

One sees that all the eigenvalues are close to 1 except for the first, for which the efficiency of $\bar{\boldsymbol{b}}$ relative to $\boldsymbol{\beta}^{(m)}$ is 0.869.

6. Conclusions. Estimated weights can always be used in unbalanced small sample random effects models.

There is never any appreciable loss of accuracy compared with a simple average, but there can be a considerable gain if the random effects are small.

OLS should not be used unless one has definite prior knowledge that there are no random effects.

Simulation can be used to find the error variance of a moment estimator of a mean.

Acknowledgements. We gratefully acknowledge considerable help from Dr. Arunas P. Verbyla and also the work of Clarence Tiong in the mathematics, statistics, computing and preparation of this paper.

REFERENCES

GUMPERTZ, M. and PANTULA, S. G. (1989). A simple approach to inference in random coefficient models, *The American Statistician*, **43**, 203–210.

HENDERSON, C. A., KEMPTHORNE, O., SEARLE, S. R. and VON KROSIGK, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling, *Biometrics*, **15**, 192–218.

JAMES, A. T., WISKICH, J. T. and CONYERS, R. A. J. (1989). The statistical modelling of mitochondrial power supply, *Journal of Mathematical Biology*, **27**, 29–48.

JAMES, A. T., WISKICH, J. T. and CONYERS, R. A. J. (1993). t-REML for robust heteroscedastic regression analysis of mitochondrial power. *Biometrics*, **49**, 339–356.

JAMES, A. T. and VENABLES, W. N. (1993). Matrix weighting of several regression coefficient vectors, *The Annals of Statistics*, **21**, 1093–1114.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ADELAIDE
ADELAIDE, SOUTH AUSTRALIA 5005
AUSTRALIA