ESTIMATING VARIANCE COMPONENTS FOR THE NESTED MODEL, MINQUE OR RESTRICTED MAXIMUM LIKELIHOOD

by

Francis Giesbrecht and Peter Burrows

Institute of Statistics
Mimeograph Series No. 1075
Raleigh - June 1976

Estimating Variance Components for the Nested Model, MINQUE or Restricted Maximum Likelihood

by

Francis Giesbrecht and Peter Burrows

ABSTRACT

This paper outlines an efficient method for computing MINQUE or restricted maximum likelihood estimates of variance components for the nested or hierarchical classification model. The method does not require storage of any large matrices and consequently does not impose limits on the size of the data set. Numerical examples indicate that the computations converge very rapidly.

Estimating Variance Components for the Nested Model, MINQUE or Restricted Maximum Likelihood

Francis Giesbrecht and Peter Burrows

I. Introduction

In a series of papers Rao (1970, 1971a, 1971b, 1972) has developed the MINQUE (MInimum Norm Quadratic Unbiased Estimation) theory of estimating variance components. These papers provide a complete solution to the problem in the sense that if normality holds then the method will give locally best (minimum variance) quadratic unbiased estimates of the variance components. Unfortunately, the form of the estimation equations given is such as to tax even a very large computer if one has a reasonably large data set since they involve inversion of an n x n matrix where n is the number of data points. Also the procedure requires prior information about the relative magnitudes of the variance components. The possibility of an iterative scheme seems reasonable, but places an even larger premium on efficient computations.

Hemmerle and Hartley (1973) develop a method for computing maximum likelihood estimates of the variance components. Corbeil and Searle (1976) modify the Hemmerle and Hartley algorithm to give restricted maximum likelihood estimates by partitioning the likelihood into two parts and then maximizing the part which is free of the fixed effects. However, both algorithms require the repeated inversion of a q x q matrix, where q is the total number of random levels in the model. Typically q will be much smaller than n, though for large data sets even that rapidly becomes unreasonable. Both algorithms are iterative and as such require only reasonable initial starting values for the unknown variance components rather than prior estimates.

The object of this paper is to show that for the hierarchal model the

MINQUE formulas can be re-written in a form that requires only that a p x p matrix be inverted, where p is the number of variance components or levels of nesting in the model. It is shown that all necessary totals can be obtained by one pass through the data set. It is also shown that if the MINQUE process is iterated (with some provision to replace negative components with small positive values) and converges, then the final answers will also satisfy the equations obtained by applying the restricted maximum likelihood method.

II. MINQUE With Invariance and Normality

Following Rao (1971a, 1971b) let Y be an n-vector of observations with structure

$$Y = X\beta + U_1\epsilon_1 + \dots + U_p\epsilon_p \tag{1}$$

where X is a known n x s matrix of rank s, β is an s-vector of unknown parameters, U_i is a known n x c_i matrix and ϵ_i is a c_i-vector of normal random variables with zero means and dispersion matrix I σ_i^2 of appropriate dimension. It is also assumed that ϵ_i and ϵ_j are uncorrelated for all $i \neq j$. Clearly

$$E[Y] = X\beta \qquad \text{and}$$

$$D[Y] = \sigma_1^2 V_1 + \dots + \sigma_p^2 V_p$$

where

$$V_i = U_i U_i' \qquad \text{for } i = 1, \dots, p.$$

The quadratic, unbiased and invariant (invariant in the sense that the estimates are unchanged if we replace the vector of observations Y by Y + Xb for any s-vector b) estimate of

$$p_1\sigma_1^2 + \dots + p_p\sigma_p^2$$

given by the MINQUE principle is

$$\lambda_1 Q_1 + \dots + \lambda_p Q_p$$

where

$$Q_i = Y'RV_iRY \ ,$$
$$R = V^{*-1}(I-P) \ ,$$
$$P = X(X'V^{*-1}X)^{-1}X'V^{*-1} \ ,$$
$$V^* = V_1 + \ldots + V_p$$

and the $\{\lambda_i\}$ are suitable (derivable) constants for arbitrary $\{p_i\}$. Replacing V^* by

$$V = \alpha_1 V_1 + \ldots + \alpha_p V_p$$

amounts to selecting a different norm and disturbs neither the unbiasedness nor invariance. Further, if the $\{\alpha_i\}$ happen to be proportional to the true but unknown $\{\sigma_i^2\}$ then the estimates obtained will also be the minimum variance quadratic unbiased invariant estimates at that particular point in the parameter space.

If all variance components are to be estimated simultaneously, rather than only a particular linear function, then the $\{\lambda_i\}$ need never be computed. The procedure is to compute the $\{Q_i\}$, equate to their expectations and solve the resulting system of equations. This can be written as

$$\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix} \begin{bmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \\ \vdots \\ \hat{\sigma}_p^2 \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_p \end{bmatrix} \qquad (2)$$

where $S_{ij} = tr(RV_iRV_j)$.

It seems reasonable that if prior knowledge of the relative magnitudes of the unknown variance components is available then this information should be used to define a set of $\{\alpha_i\}$ and the matrix V^* replaced by V in (2). Since (2) is a linear equation and the right hand sides are quadratic forms, the variance-covariance matrix of the estimated variance components follows immediately (in terms of the true variance components and $\{\alpha_i\}$).

Formally this provides a complete solution for the variance component estimation problem. However the computational aspects of obtaining the elements of (2) directly are formidable. They involve inverting an n x n matrix. Consequently the examination of special cases is in order. Of particular interest here is the nested analysis of variance model, since La Motte (1972) has obtained V^{-1} in a manageable form.

III. Connection with Restricted Maximum Likelihood

In order to obtain the maximum likelihood estimates of the variance components we proceed in two steps. The first is to assume $\sigma_i^2 = \alpha_i$ for $i = 1, \ldots, p$ and estimate β. This yields

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \ .$$

The second step is to compute residuals

$$Z = Y - X\hat{\beta}$$
$$= Q_v Y$$

and apply maximum likelihood to obtain estimates of $\{\sigma_i^2\}$.

Define

$$Z^* = T_v Y$$

where T_v consists of $n - s$ linearly independent rows of Q_v . The likelihood can now be written as

$$\mathcal{L} = \frac{K}{|T_v V T_v'|^{\frac{1}{2}}} \quad \exp\left\{-\tfrac{1}{2} Y'T_v' (T_v V T_v')^{-1} T_v Y\right\}.$$

Treating T_v as a constant (the $\{\sigma_i^2\}$ have been replaced by $\{\alpha_i\}$) gives the likelihood equations,

$$\frac{\partial \; \ell n \; \mathcal{L}}{\partial \; \sigma_i^2} = -\tfrac{1}{2}\mathrm{tr}[(T_v V T_v')^{-1} T_v V_i T_v']$$
$$+ \tfrac{1}{2} Y' T_v' (T_v V T_v')^{-1} T_v V_i T_v' (T_v V T_v')^{-1} T_v Y$$

for i = 1, 2, ..., p.

Equating the partial derivatives to zero and rearranging terms leads to

$$\mathrm{tr}[T_v'(T_v V T_v')^{-1} T_v V_i]$$
$$= Y' T_v' (T_v V T_v')^{-1} T_v V_i T_v' (T_v V T_v')^{-1} T_v Y \tag{3}$$

for i = 1, 2, ..., p.

Khatri (1966) has shown that for any symmetric positive definite matrix C,

$$T_c'(T_c C T_c')^{-1} T_c = C^{-1} Q_c .$$

Direct substitution into (3) gives

$$\mathrm{tr}(V^{-1} Q_v V_i) = Y' Q_v' V^{-1} V_i V^{-1} Q_v Y$$

for i = 1, 2, ..., p. Also we can replace V by $\hat{V}$ to obtain

$$\mathrm{tr}(\hat{V}^{-1} Q_{\hat{V}} V_i) = Y' Q_{\hat{V}} \hat{V}^{-1} V_i \hat{V}^{-1} Q_{\hat{V}} Y \tag{4}$$

for i = 1, 2, ..., p.

The quadratic forms in (4) are similar to those required by MINQUE, with the only difference being the nature of V or $\hat{V}$. In MINQUE theory V is a function of $\{\alpha_i\}$ while in maximum likelihood estimation theory $\hat{V}$ is a function of $\{\hat{\sigma}_i^2\}$. Also if the $\{\alpha_i\}$ are chosen correctly i.e. agree with $\{\sigma_i^2\}$ then the expected values of the quadratic forms obtained under the MINQUE principle can be written as

$$E[Y' Q_v' V^{-1} V_i V^{-1} Q_v Y] = \sum_j \mathrm{tr}(Q_v' V^{-1} V_i V^{-1} Q_v V_j) \sigma_j^2$$
$$= \mathrm{tr}(Q_v' V^{-1} V_i V^{-1} Q_v V)$$
$$= \mathrm{tr}(V^{-1} Q_v V_i) .$$

It follows immediately that if one uses an iterative scheme for solving the maximum likelihood equations in which the $\{\alpha_i\}$ are updated to agree with the $\{\hat{\sigma}_i^2\}$ (with some provision to replace negative values with some small positive constants) the solution will agree exactly with that obtained from the MINQUE equations using the same rules for iteration.

IV. Estimation in the Nested Model

The development in sections II and III has been in terms of a relatively general linear model. For the remainder of this paper we will specialize to the 3-level nested model with only the one fixed parameter μ. The 3-level nested model was selected because it is sufficiently complex to display the general pattern of terms that arise when more levels are introduced without the notation becoming too complex. The extension to the case where the random effects are nested within levels of a fixed effect is obvious and will not be discussed. The 3-level nested model is given by

$$y_{hijk} = \mu + a_h + b_{hi} + c_{hij} + e_{hijk} \qquad (5)$$

$$\text{for} \quad k = 1, 2, \ldots, n_{hij}$$
$$j = 1, 2, \ldots, m_{hi}$$
$$i = 1, 2, \ldots, m_h$$
$$h = 1, 2, \ldots, m$$

and where

$\{y_{hijk}\}$ are the observations,

μ is an unknown constant,

$\{a_h\}$ are independent $N(0, \sigma_1^2)$,

$\{b_{hi}\}$ are independent $N(0, \sigma_2^2)$,

$\{c_{hij}\}$ are independent $N(0, \sigma_3^2)$,

$\{e_{hijk}\}$ are independent $N(0, \sigma_4^2)$ and the latter 4 sets are also mutually independent.

Since we will have occasion to refer to the value obtained by summation over certain subscripts, we will use the convention of indicating such a sum by replacing the subscript with the + symbol. For example

$$n_{hi+} = \sum_j n_{hij}$$

and

$$n_{h++} = \sum_i n_{hi+} \qquad \text{etc.}$$

To establish the connection between (1) and (5), note that Y corresponds to the n_{+++} elements $\{y_{hijk}\}$ in dictionary order. X is a column of 1's and β the one element μ. U_1 corresponds to an n_{+++} x m matrix of 0's and 1's, U_2 to an n_{+++} x m_+ matrix of 0's and 1's, U_3 to an n_{+++} x m_{++} matrix and U_4 the n_{+++} x n_{+++} identity matrix. The covariance between y_{hijk} and $y_{h'i'j'k'}$ can be written as

$$\delta_{hh'}(\sigma_1^2 + \delta_{ii'}(\sigma_2^2 + \delta_{jj'}(\sigma_3^2 + \delta_{kk'}\sigma_4^2)))$$

where

$$\delta_{\ell m} = \begin{cases} 1 & \text{if } \ell = m \\ 0 & \text{if } \ell \neq m \end{cases} .$$

The corresponding element of V can be written as

$$v_{hijk,\, h'i'j'k'} = \delta_{hh'}(\alpha_1 + \delta_{ii'}(\alpha_2 + \delta_{jj'}(\alpha_3 + \delta_{kk'}\alpha_4)))$$

where α_i represents the prior estimate for σ_i^2 . Since only the relative magnitudes of the $\{\alpha_i\}$ are important they can be scaled so that one of them is 1. Also if one wishes the more primitive form of the MINQUE, Rao (1972), all can be set all equal to 1. The extension to more levels of nesting is immediate. The corresponding element of V^{-1} can be written as

$$\delta_{hh'}\delta_{ii'}\delta_{jj'}\delta_{kk'}\alpha_4^{-1} - \delta_{hh'}\delta_{ii'}\delta_{jj'}w_{hij}\alpha_4^{-1}$$

$$- \delta_{hh'}\delta_{ii'}w_{hij}w_{hij'}w_{hi}\alpha_3^{-1}$$

$$- \delta_{hh'}w_{hij}w_{hi'j'}w_{hi}w_{hi'}w_h\alpha_2^{-1}$$

where

$$w_{hij} = \alpha_3(\alpha_3\, n_{hij} + \alpha_4)^{-1} ,$$

$$w_{hi} = \alpha_2(\alpha_2\, n_{hi} + \alpha_3)^{-1} ,$$

$$w_h = \alpha_1(\alpha_1\, n_h + \alpha_2)^{-1} ,$$

$$w = (\sum_h w_h\, n_h)^{-1} ,$$

$$n_{hi} = \sum_j n_{hij}w_{hij} ,$$

and

$$n_h = \sum_i n_{hi}w_{hi} .$$

Summation over h′, i′, j′ and k′ yields

$$\alpha_1^{-1}w_{hij}w_{hi}w_h$$

as the element of the vector $V^{-1}X$ corresponding to y_{hijk} in Y. Also note that the scalar $X'V^{-1}X$ is $\alpha_1^{-1}w^{-1}$. The general element of R now follows as

$$\delta_{hh'}\delta_{ii'}\delta_{jj'}\delta_{kk'}\alpha_4^{-1} - \delta_{hh'}\delta_{ii'}\delta_{jj'}w_{hij}\alpha_4^{-1}$$

$$- \delta_{hh'}\delta_{ii'}w_{hij}w_{hi}w_{hij'}\alpha_3^{-1}$$

$$- \delta_{hh'}w_{hij}w_{hi}w_hw_{hi'}w_{hi'j'}\alpha_2^{-1}$$

$$- w_{hij}w_{hi}w_hww_{h'}w_{h'i'}w_{h'i'j'}\alpha_1^{-1} .$$

The pattern for the case with more levels is obvious.

The typical element of Z = RY is

$$z_{hijk} = y_{hijk}\alpha_4^{-1} - w_{hij}y_{hij+}\alpha_4^{-1}$$

$$- w_{hij}w_{hi}(\sum_{j'} w_{hij'}y_{hij'+})\alpha_3^{-1}$$

$$- w_{hij}w_{hi}w_h(\sum_{i'}\sum_{j'} w_{hi'}w_{hi'j'}y_{hi'j'+})\alpha_2^{-1}$$

$$- w_{hij}w_{hi}w_h w(\sum_{h'}\sum_{i'}\sum_{j'} w_{h'}w_{h'i'}w_{h'i'j'}y_{h'i'j'+})\alpha_1^{-1} ,$$

where $y_{hij+} = \sum_k y_{hijk}$.

The quadratic forms in (2) are now obtained as

$$\sum_h \sum_i \sum_j \sum_k z^2_{hijk} \quad , \quad \sum_h \sum_i \sum_j z^2_{hij+} \quad ,$$

$$\sum_h \sum_i z^2_{hi++} \quad \text{and} \quad \sum_h z^2_{h+++} \quad ,$$

where the + symbol indicates summation over the subscript before squaring. We note at this point that these quadratic forms are obtained by two passes through the data file and are not affected by size of data set.

In order to obtain expressions for the elements in RV_1 , RV_2 , RV_3 , RV_4 and the various products in a systematic pattern it is convenient to define

$$t_h = \alpha_2^{-1} + w_h w\, \alpha_1^{-1} \quad ,$$

$$t_{hi} = \alpha_3^{-1} + w_{hi}w_h t_h \quad \text{and}$$

$$t_{hij} = \alpha_4^{-1} + w_{hij}w_{hi}t_{hi} \quad .$$

The elements in RV_1 , RV_2 , RV_3 and RV_4 are given in Table 1 in a convenient form. Expressions for $tr(RV_iRV_j)$ for $4 \ge i \ge j \ge 1$ are given in Appendix A. These expressions can (and should) be condensed before being used for computations. However when this is done the pattern is not as clear. We also notice that if one arrives at a point where the estimates obtained agree with the prior values then the variance-covariance matrix of the estimates can be approximated by twice the inverse of the left side of 2.

Table I. Elements in the RV_1, RV_2, RV_3, and RV_4 Matrices

Coefficient	RV_4	RV_3	RV_2	RV_1
$\delta_{hh'}\delta_{ii'}\delta_{jj'}\delta_{kk'}$	$\alpha_4^{-1}-w_{hij}t_{hij}$	$w_{hij}(\alpha_3^{-1}-n_{hij}w_{hij}w_{hi}t_{hi})$	$w_{hij}w_{hi}(\alpha_2^{-1}-n_{hi}w_{hi}w_h t_h)$	$w_{hij}w_{hi}w_h(\alpha_1^{-1}-n_h w_h w\alpha_1^{-1})$
$\delta_{hh'}\delta_{ii'}\delta_{jj'}(1-\delta_{kk'})$	$-w_{hij}t_{hij}$	$w_{hij}(\alpha_3^{-1}-n_{hij}w_{hij}w_{hi}t_{hi})$	$w_{hij}w_{hi}(\alpha_2^{-1}-n_{hi}w_{hi}w_h t_h)$	$w_{hij}w_{hi}w_h(\alpha_1^{-1}-n_h w_h w\alpha_1^{-1})$
$\delta_{hh'}\delta_{ii'}(1-\delta_{jj'})$	$-w_{hij}w_{hi}w_{hij'}t_{hi}$	$-w_{hij}w_{hi}w_{hij'}n_{hij'}t_{hi}$	$w_{hij}w_{hi}(\alpha_2^{-1}-n_{hi}w_{hi}w_h t_h)$	$w_{hij}w_{hi}w_h(\alpha_1^{-1}-n_h w_h w\alpha_1^{-1})$
$\delta_{hh'}(1-\delta_{ii'})$	$-w_{hij}w_{hi}w_h w_{hi'}w_{hi'j}t_h$	$-w_{hij}w_{hi}w_h w_{hi'}w_{hi'j'}n_{hi'j'}t_n$	$-w_{hij}w_{hi}w_h w_{hi'}n_{hi'}t_h$	$w_{hij}w_{hi}w_h(\alpha_1^{-1}-n_h w_h w\alpha_1^{-1})$
$(1-\delta_{hh'})$	$-w_{hij}w_{hi}w_h ww_{h'}w_{h'i'}w_{h'i'j'}\alpha_1^{-1}$	$-w_{hij}w_{hi}w_h ww_{h'}w_{h'i'}w_{h'i'j'}n_{h'i'j'}\alpha_1^{-1}$	$-w_{hij}w_{hi}w_h ww_{h'}w_{h'i'}n_{h'i'}\alpha_1^{-1}$	$-w_{hij}w_{hi}w_h ww_{h'}n_{h'}\alpha_1^{-1}$

V. Computational Aspects

An examination of the formulas developed in the previous section shows that a very efficient computing routine is possible. If the data are sorted or arranged according to the nesting structure then all totals required in equation (2) can be obtained in one pass. Alternatively, if one wishes to minimize the programming effort, all the accumulations can be obtained rather easily by using any one of a number of standard programs for computing means and totals of subsets of observations in a data set. Summaries need to be obtained at each level of nesting. These summaries are then merged and a final set of accumulations obtained.

In order to examine the behavior of the suggested iterative scheme several analyses were performed on synthetic, unbalanced data sets. The results of three typical cases are given in Table II. The striking feature of these computations is the very rapid convergence. In fact it appears that 2 or at most 3 cycles are sufficient. The extent of the unbalance in the data can be judged from Table III.

VI. Acknowledgement

The authors wish to thank T. M. Gerig for his valuable comments and suggestions in the preparation of this paper.

TABLE II

Summary of Three Analyses

	σ_1^2	σ_2^2	σ_3^2	σ_4^2
A. True parameter values*	2.	.1	.5	1.
Starting values	1.	2.	1.	1.
Estimates first cycle	4.289	.006	.555	1.157
Estimates second cycle	4.146	-.015	.231	1.140
Estimates third cycle	4.133	-.011	.223	1.144
Estimates fourth cycle	4.134	-.012	.224	1.143
Estimates fifth cycle	4.134	-.012	.224	1.143
Estimates A.o.V.	4.398	-.061	.130	1.173
B. True parameter values*	2.	.1	.5	1.
Starting values	5.	1.	.1	1.
Estimates first cycle	4.292	.039	.211	1.128
Estimates second cycle	4.149	-.006	.220	1.142
Estimates third cycle	4.135	-.012	.224	1.143
Estimates fourth cycle	4.134	-.012	.224	1.143
C. True parameter values	2.	.1	.5	1.
Starting values	5.	1.	.1	1.
Estimates first cycle	3.187	.017	.354	1.135
Estimates second cycle	3.255	0.003	.330	1.158
Estimates third cycle	3.260	-.005	.332	1.158
Estimates fourth cycle	3.260	-.005	.332	1.158
Estimates fifth cycle	3.260	-.005	.332	1.158
Estimates A.o.V.	3.102	.014	.318	1.158

* These two used a common data set.

TABLE III Pattern of n_{hij}'s in Simulated Data Sets

		h									
	i	1	2	3	4	5	6	7	8	9	10
Sets A & B	1	2,1	2	1,1,2	3,1,1	2,1,3	3,3	2	3,1	3	3,2,1
	2	1,2	2,1	2	2,1,1	2,3,1	2,3	2	2,2,1		1,3,3
	3			2,1	2,3,1	1,3	1,1	2,3	1		1,1
	4			3	2,2	3	3,2				3
	5			3,1		2,3	2				
Set C	1	1,2,2	1,2	2	3	1,2,1	1,1	3,1	1	2	2
	2	3	3	2,1	1	1,1,3	2,2	1,2,3	2,3,1	1,1,1	1
	3		2	2	2,1,3	3	1,3,2	2,3,1	2,3,3		2
	4		1,3,1	3,3,3	2,2		1		3,2		
	5		1		1,3,3		1				

REFERENCES

Corbeil, R. R. and Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. Technometrics 18, 31-38.

Hemmerle, W. J. and Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. Technometrics 15, 819-832.

Khatri, C. G. (1966). A note on a manova model applied to problems in growth curve. Ann. Inst. Stat. Math. 18, 75-86.

LaMotte, L. R. (1972). Notes on the covariance matrix of a random nested anova model. Ann. Math. Statist. 43, 659-662.

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. J. Amer. Statist. Assoc. 65, 161-72.

Rao, C. R. (1971a). Estimation of variance and covariance components--MINQUE theory. J. Mult. Analysis 1, 257-75.

Rao, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. J. Mult. Analysis 1, 445-56.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. J. Amer. Statist. Assoc. 67, 112-15.

APPENDIX

$$\mathrm{tr}(RV_4RV_4) = \sum_h \sum_i \sum_j n_{hij}(\alpha_4^{-1} - w_{hij}t_{hij})^2 + \sum_h \sum_i \sum_j n_{hij}(n_{hij}-1)w_{hij}^2 t_{hij}^2$$

$$+ \sum_h \sum_i \sum_j ((\sum_j n_{hij}w_{hij}^2) - n_{hij}w_{hij}^2)n_{hij}w_{hij}^2 w_{hi}^2 t_{hi}^2$$

$$+ \sum_h \sum_i ((\sum_i \sum_j n_{hij}w_{hi\,j}^2 w_{ni}^2) - (\sum_j n_{hij} w_{nij}^2)w_{hi}^2)(\sum_j n_{hij} w_{hij}^2)w_{hi}^2 w_h^2 t_h^2$$

$$+ \sum_h ((\sum_h \sum_i \sum_j n_{h\,ij}w_{hij}^2 w_{hi}^2 w_h^2)$$

$$- (\sum_i \sum_j n_{hij}w_{hij}^2 w_{hi}^2)w_h^2)(\sum_i \sum_j n_{hij}w_{hij}^2 w_{hi}^2)w_h^2 w^2 \alpha_1^{-2} .$$

$$\mathrm{tr}(RV_4RV_3) = \sum_h \sum_i \sum_j n_{hij}w_{hij}(\alpha_4^{-1} - w_{nij}t_{nij})(\alpha_3^{-1} - n_{hij}w_{hij}w_{hi}t_{hi})$$

$$- \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1)w_{hij}^2 t_{hij}(\alpha_3^{-1} - n_{hij}w_{hij}w_{hi}t_{hi})$$

$$+ \sum_h \sum_i \sum_j ((\sum_j n_{hij} w_{hij}^2) - n_{hij}w_{hij}^2)n_{hij}^2 w_{hij}^2 w_{hi}^2 t_{hi}^2$$

$$+ \sum_h \sum_i ((\sum_i \sum_j n_{hi\,j}w_{ni\,j}^2 w_{hi}^2) - (\sum_j n_{hij}w_{hij}^2)w_{hi}^2)(\sum_j n_{hij}^2 w_{hij}^2)w_{hi}^2 w_h^2 t_h^2$$

$$+ \sum_h ((\sum_h \sum_i \sum_j n_{hij}w_{hij}^2 w_{hi}^2 w_h^2)$$

$$- (\sum_i \sum_j n_{hij}w_{hij}^2 w_{hi}^2 w_h^2))(\sum_i \sum_j n_{nij}^2 w_{hij}^2 w_{hi}^2)w_h^2 w^2 \alpha_1^{-2} .$$

$$\mathrm{tr}(RV_4RV_2) = \sum_h \sum_i \sum_j n_{hij} w_{hij} w_{hi} (\alpha_4^{-1} - w_{hij} t_{hij})(\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)$$

$$- \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1) w_{hij}^2 w_{hi} t_{hij} (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)$$

$$- \sum_h \sum_i \sum_j ((\sum_j n_{hij} w_{hij}^2) - n_{hij} w_{hij}^2) n_{hij} w_{hij} w_{hi}^2 t_{hi} (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)$$

$$+ \sum_h \sum_i ((\sum_i \sum_j n_{hij} w_{hij}^2 w_{hi}^2) - (\sum_j n_{hij} w_{hij}^2 w_{hi}^2)) n_{hi}^2 w_{hi}^2 w_h^2 t_h^2$$

$$+ \sum_h ((\sum_h \sum_i \sum_j n_{hij} w_{hij}^2 w_{hi}^2 w_h^2)$$

$$- (\sum_i \sum_j n_{hij} w_{hij}^2 w_{hi}^2 w_h^2))(\sum_i n_{hi}^2 w_{hi}^2) w_h^2 w^2 \alpha_1^{-2} .$$

$$\mathrm{tr}(RV_4RV_1) = \sum_h \sum_i \sum_j n_{hij} w_{hij} w_{hi} w_h (\alpha_4^{-1} - w_{hij} t_{hij})(\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$- \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1) w_{hij}^2 w_{hi} w_h t_{hij} (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$- \sum_h \sum_i \sum_j ((\sum_j n_{hij} w_{hij}^2)$$

$$- n_{hij} w_{hij}^2) n_{hij} w_{hij} w_{hi}^2 w_h t_{hi} (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$- \sum_h \sum_i ((\sum_i \sum_j n_{hij} w_{hij}^2 w_{hi}^2)$$

$$- (\sum_j n_{hij} w_{hij}^2 w_{hi}^2)) n_{hi} w_{hi} w_h^2 t_n (\alpha_1^{-1} - n_h w_n w \alpha_1^{-1})$$

$$+ \sum_h ((\sum_h \sum_i \sum_j n_{hij} w_{hij}^2 w_{hi}^2 w_h^2) - (\sum_i \sum_j n_{hij} w_{hij}^2 w_{hi}^2 w_h^2)) n_h^2 w_h^2 w^2 \alpha_1^{-2} .$$

$$\begin{aligned}
\mathrm{tr}(RV_3RV_3) &= \sum_h \sum_i \sum_j n_{hij} w_{hij}^2 (\alpha_3^{-1} - n_{hij} w_{hij} w_{hi} t_{hi})^2 \\
&+ \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1) w_{hij}^2 (\alpha_3^{-1} - n_{hij} w_{hij} w_{hi} t_{hi})^2 \\
&+ \sum_h \sum_i \sum_j ((\sum_j n_{hij}^2 w_{hij}^2) - n_{hij}^2 w_{hij}^2) n_{hij}^2 w_{hij}^2 w_{hi}^2 t_{hi}^2 \\
&+ \sum_h \sum_i ((\sum_i \sum_j n_{hij}^2 w_{hij}^2 w_{hi}^2) - (\sum_j n_{hij}^2 w_{hij}^2 w_{hi}^2))(\sum_j n_{hij}^2 w_{hij}^2) w_{hi}^2 w_h^2 t_h^2 \\
&+ \sum_h ((\sum_h \sum_i \sum_j n_{hij}^2 w_{hij}^2 w_{hi}^2 w_h^2) \\
&- (\sum_i \sum_j n_{hij}^2 w_{hij}^2 w_{hi}^2 w_h^2))(\sum_i \sum_j n_{hij}^2 w_{hij}^2 w_{hi}^2) w_h^2 w^2 \alpha_1^{-2} .
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}(RV_3RV_2) &= \sum_h \sum_i \sum_j n_{hij} w_{hij}^2 w_{hi} (\alpha_3^{-1} - n_{hij} w_{hij} w_{hi} t_{hi})(\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h) \\
&+ \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1) w_{hij}^2 w_{hi} (\alpha_3^{-1} - n_{hij} w_{hij} w_{hi} t_{hi})(\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h) \\
&- \sum_h \sum_i \sum_j ((\sum_j n_{hij}^2 w_{hij}^2) - n_{hij}^2 w_{hij}^2) n_{hij} w_{hij} w_{hi}^2 t_{hi} (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h) \\
&+ \sum_h \sum_i ((\sum_i \sum_j n_{hij}^2 w_{hij}^2 w_{hi}^2) - (\sum_j n_{hij}^2 w_{hij}^2) w_{hi}^2) n_{hi}^2 w_{hi}^2 w_h^2 t_h^2 \\
&+ \sum_h ((\sum_h \sum_i \sum_j n_{hij}^2 w_{hij}^2 w_{hi}^2 w_h^2) \\
&- (\sum_i \sum_j n_{hij}^2 w_{hij}^2 w_{hi}^2) w_h^2)(\sum_i n_{hi}^2 w_{hi}^2) w_h^2 w^2 \alpha_1^{-2} .
\end{aligned}$$

$$\mathrm{tr}(RV_3RV_1) = \sum_h \sum_i \sum_j n_{hij} w^2_{hij} w_{hi} w_h (\alpha_3^{-1} - n_{hij} w_{hij} w_{hi} t_{hi})(\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$+ \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1) w^2_{hij} w_{hi} w_h (\alpha_3^{-1} - n_{hij} w_{hij} w_{hi} t_{hi})(\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$- \sum_h \sum_i \sum_j ((\sum_j n^2_{hij} w^2_{hij}) - n^2_{hij} w^2_{hij}) n_{hij} w_{hij} w^2_{hi} w_h t_{hi} (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$- \sum_h \sum_i ((\sum_i \sum_j n^2_{hij} w^2_{hij} w^2_{hi}) - (\sum_j n^2_{hij} w^2_{hij}) w^2_{hi}) n_{hi} w_{hi} w^2_h t_h (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$+ \sum_h ((\sum_h \sum_i \sum_j n^2_{hij} w^2_{hij} w^2_{hi} w^2_h) - (\sum_i \sum_j n^2_{hij} w^2_{hij} w^2_{hi}) w^2_h) n^2_h w^2_h w^2 \alpha_1^{-2} .$$

$$\mathrm{tr}(RV_2RV_2) = \sum_h \sum_i \sum_j n_{hij} w^2_{hij} w^2_{hi} (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)^2$$

$$+ \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1) w^2_{hij} w^2_{hi} (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)^2$$

$$+ \sum_h \sum_i \sum_j (n_{hi} - n_{hij} w_{hij}) n_{hij} w_{hij} w^2_{hi} (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)^2$$

$$+ \sum_h \sum_i ((\sum_i n^2_{hi} w^2_{hi}) - n^2_{hi} w^2_{hi}) n^2_{hi} w^2_{hi} w^2_h t^2_h$$

$$+ \sum_h ((\sum_h \sum_i n^2_{hi} w^2_{hi} w^2_h) - (\sum_i n^2_{hi} w^2_{hi}) w^2_h)(\sum_i n^2_{hi} w^2_{hi}) w^2_h w^2 \alpha_1^{-2} .$$

$$\mathrm{tr}(RV_2RV_1) = \sum_h \sum_i \sum_j n_{hij} w_{hij}^2 w_{hi}^2 w_h (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)(\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$+ \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1) w_{hij}^2 w_{hi}^2 w_h (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)(\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$+ \sum_h \sum_i \sum_j (n_{hi} - n_{hij} w_{hij}) n_{hij} w_{hij} w_{hi}^2 w_h (\alpha_2^{-1} - n_{hi} w_{hi} w_h t_h)(\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$- \sum_h \sum_i ((\sum_i n_{hi}^2 w_{hi}^2) - n_{hi}^2 w_{hi}^2) n_{hi} w_{hi} w_h^2 t_h (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})$$

$$+ \sum_h ((\sum_h \sum_i n_{hi}^2 w_{hi}^2 w_h^2) - (\sum_i n_{hi}^2 w_{hi}^2) w_h^2) n_h^2 w_h^2 w^2 \alpha_1^{-2} .$$

$$\mathrm{tr}(RV_1RV_1) = \sum_h \sum_i \sum_j n_{hij} w_{hij}^2 w_{hi}^2 w_h^2 (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})^2$$

$$+ \sum_h \sum_i \sum_j n_{hij}(n_{hij} - 1) w_{hij}^2 w_{hi}^2 w_h^2 (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})^2$$

$$+ \sum_h \sum_i \sum_j (n_{hi} - n_{hij} w_{hij}) n_{hij} w_{hij} w_{hi}^2 w_h^2 (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})^2$$

$$+ \sum_h \sum_i (n_h - n_{hi} w_{hi}) n_{hi} w_{hi} w_h^2 (\alpha_1^{-1} - n_h w_h w \alpha_1^{-1})^2$$

$$+ \sum_h ((\sum_h n_h^2 w_h^2) - n_h^2 w_h^2) n_h^2 w_h^2 w^2 \alpha_1^{-2} .$$