The Annals of Statistics
2016, Vol. 44, No. 3, 1131–1164
DOI: 10.1214/15-AOS1400

RANDOMIZATION-BASED MODELS FOR MULTITIERED EXPERIMENTS: I. A CHAIN OF RANDOMIZATIONS

BY R. A. BAILEY[†] AND C. J. BRIEN[§]

University of St. Andrews[†], Queen Mary University of London[†], University of South Australia[§] and University of Adelaide[§]

We derive randomization-based models for experiments with a chain of randomizations. Estimation theory for these models leads to formulae for the estimators of treatment effects, their standard errors and expected mean squares in the analysis of variance. We discuss the practicalities in fitting these models and outline the difficulties that can occur, many of which do not arise in two-tiered experiments.

1. Introduction. Bailey [2, 3], following Grundy and Healy [22], outlines a method of deriving randomization-based models for experiments. These mixed models are randomization-based in the sense that the variance matrix on which the analysis is based is defined by the group of permutations for the randomization. The method applies to a general class of structures: those derived from a group of permutations which is stratifiable, in the sense defined in Section 2. This class includes all poset block structures and many other structures besides.

This approach applies to just a single randomization, as defined in [11] to be one that can be achieved using a single permutation of the set of observational units. Brien and Bailey [11–13] also describe experiments with multiple randomizations, which require multiple permutations.

EXAMPLE 1.1 (Simpler two-phase sensory experiment). Figure 1, equivalent to one presented in [11], Section 4.1, shows a two-phase sensory experiment. The first, or field, phase is a viticultural experiment and the second, or evaluation, phase involves the assessment of wine made from the produce of the first-phase vine-plots. The first phase uses a randomized complete block design, whilst the design in the second phase consists of a pair of 8×8 Latin squares on each occasion. The example involves two randomizations: one in which treatments are randomized to vine-plots, each comprised of several vines, and a second in which vine-plots are randomized to evaluations. This is called a chain of randomizations in [12].

The middle panel in Figure 1 shows that the proper randomization for the first phase is to randomly permute blocks and then randomly permute vine-plots inde-

Received August 2013; revised May 2015.
MSC2010 subject classifications. Primary 62B15, 62J05; secondary 62J10.
Key words and phrases. Analysis of variance, expected mean square, mixed model, multiphase experiments, multitiered experiments, randomization-based model, REML, structure, tier.

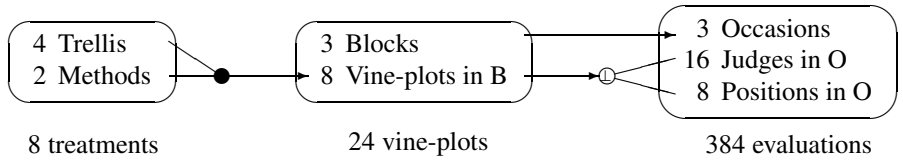

FIG. 1. *Randomization diagram for Example 1.1*: *treatments are randomized to vine-plots*, *which are in turn randomized to evaluations*; B *denotes Blocks*, O *denotes Occasions*.

pendently within each block. The right-hand panel shows the proper randomization for the second phase: randomly permute occasions; within each occasion, randomly permute all 16 judges and, independently, randomly permute all eight positions.

Brien and Bailey show how to assess the properties of experiments with multiple randomizations in [12, 13]. Curnow [20], in correcting the analysis of McIntyre [26], showed how to analyse the results of two-phase experiments by analysis of variance (anova). Wood, Williams and Speed [49] also discussed the analysis of two-phase experiments. Brien indicated how to use tiers to obtain the anova for multitiered experiments in [8] and derived expected mean squares under a mixed model in [9]. Brien and Payne [16] extended the sweep algorithm of Wilkinson [36, 47] to cover anova for multitiered experiments. Brien and Bailey [11] and Brien and Demétrio [14] describe how to analyse the data from such experiments by using mixed models. However, no one has so far given general formulae for the estimators of treatment effects and their standard errors for multitiered experiments, nor have formulae for the expected mean squares under randomization-based models been derived.

Section 2 formulates the randomization-based model for a two-tiered experiment and generalizes it to experiments with two randomizations in a chain. Section 3 describes families of expectation models that lead to a treatment decomposition; the assumption of structure balance is discussed. The properties of anova for randomization-based models are outlined in Section 4. Section 5 contains a set of examples. Sections 6–9 address the estimation of treatment effects, first for two-tiered experiments and then for various cases of three-tiered experiments. Section 10 generalizes this to an arbitrary number of randomizations in a chain. Section 11 covers the use of software in estimating model parameters, including a discussion of randomization-based models in the class of all mixed models. Statistical inference is discussed in Section 12. Section 13 briefly touches on models other than those described in Section 2. See Section 3 of [12] for definitions of some terms and notation specific to the approach we take.

2. Randomization-based models.

2.1. *The randomization-based model for a two-tiered experiment.* As in [11–13], we randomize the set of objects Γ to another set of objects Υ, so we

have a design function $h: \Upsilon \to \Gamma$. If the objects in Γ are treatments then $h(\upsilon)$ is the treatment assigned to unit υ in Υ. We associate a structure with each of Υ and Γ. If V_Υ is the space of all real vectors indexed by Υ, then a *structure* on Υ is an orthogonal decomposition of V_Υ. This is specified by a set of symmetric, idempotent, mutually orthogonal matrices projecting onto the subspaces of V_Υ in the decomposition. Similarly, structure on Γ is an orthogonal decomposition of the space V_Γ.

The usual initial assumption for the response Y_υ on unit υ in Υ is additive:

$$Y_\upsilon = w_\upsilon + \tau_{h(\upsilon)}. \tag{1}$$

In some approaches, w_υ is taken to be a constant, but here it is taken to be a random variable, as in [3, 22]. It depends only on the unit υ which is providing the response. On the other hand, τ_i, for i in Γ, is a constant. It depends only on the treatment i which is applied to υ. Permitting the w_υ to be random allows for measurement error, without the assumption of any particular form for it, and any random sampling of units that may occur.

Let G be a group of permutations of Υ. We usually take G to be the group of all permutations that preserve certain generalized factors on Υ, in the sense that if F is such a generalized factor, $g \in G$ and $F(\upsilon_1) = F(\upsilon_2)$ then we must have $F(g(\upsilon_1)) = F(g(\upsilon_2))$. Generalized factors are denoted by $F_1 \wedge \cdots \wedge F_n$ and are factors whose levels are the levels combinations of F_1, $F_2, \ldots$ and F_n, for $n \geq 1$. In [2, 3] it is argued that if we randomize by choosing g from G at random then it is appropriate to replace w_υ by W_υ, which is the mixture of the $w_{g(\upsilon)}$ over g in G. Hence, we get the randomization-based model

$$Y_\upsilon = W_\upsilon + \tau_{h(\upsilon)}, \tag{2}$$

where the W_υ are random variables which are exchangeable under G: in particular:

(P.a) if there is any g in G for which $g(\upsilon_1) = \upsilon_2$ then W_{υ_1} and W_{υ_2} have the same distribution, in particular, the same expectation;

(P.b) if there is any g in G for which $g(\upsilon_1) = \upsilon_2$ and $g(\upsilon_3) = \upsilon_4$ then the joint distribution of $(W_{\upsilon_1}, W_{\upsilon_3})$ is the same as the joint distribution of $(W_{\upsilon_2}, W_{\upsilon_4})$, in particular, $\mathrm{Cov}(W_{\upsilon_1}, W_{\upsilon_3}) = \mathrm{Cov}(W_{\upsilon_2}, W_{\upsilon_4})$.

If the group G is transitive on Υ then property (P.a) is true for all choices of υ_1 and υ_2, so we may incorporate the constant value of $\mathrm{E}(W_\upsilon)$ into each τ_i and so assume that $\mathrm{E}(W_\upsilon) = 0$ for all υ in Υ. We restrict attention to cases where G is transitive, which implies that every unrandomized factor on Υ is equi-replicate.

Let $\mathbf{Y}$ and $\mathbf{W}$ be the vectors of the random variables Y_υ and W_υ, respectively, and $\boldsymbol{\tau}$ the vector of treatment coefficients τ_i. Represent the design function h by the $\Upsilon \times \Gamma$ design matrix $\mathbf{X}_h$, with (υ, i)-entry equal to 1 if $h(\upsilon) = i$ and to 0 otherwise. Then equation (2) can be rewritten in vector form as $\mathbf{Y} = \mathbf{W} + \mathbf{X}_h\boldsymbol{\tau}$, and $\mathrm{E}(\mathbf{Y}) = \mathbf{X}_h\boldsymbol{\tau}$.

The pattern in the (co)variance matrix $\mathbf{C}$ of $\mathbf{W}$ is determined by property (P.b), which implies that $\mathbf{C}$ is a patterned matrix with the same entries, including multiplicities, in every row; only their order differs. The group G defines the set $\mathcal{B}$ of non-zero symmetric $\Upsilon \times \Upsilon$ adjacency matrices $\mathbf{B}$ with entries 0 and 1, whose sum is the all-1 matrix $\mathbf{J}$, as follows: if the (υ_1, υ_2)-entry of $\mathbf{B}$ is equal to 1 then the (υ_3, υ_4)-entry is equal to 1 if and only if there is some g in G for which either $g(\upsilon_1) = \upsilon_3$ and $g(\upsilon_2) = \upsilon_4$ or $g(\upsilon_1) = \upsilon_4$ and $g(\upsilon_2) = \upsilon_3$. Then the product of any two adjacency matrices is a linear combination of matrices in $\mathcal{B}$. Property (P.b) implies that there are (co)variances $\zeta_{\mathbf{B}}$ such that $\mathbf{C} = \sum_{\mathbf{B}\in\mathcal{B}} \zeta_{\mathbf{B}}\mathbf{B}$. For simple orthogonal block structures, this form of $\mathbf{C}$ is the same as the variance matrix for the null randomization distribution given by Nelder [29].

The group G is said to be *stratifiable* [1, 6] if the eigenvectors of the matrix $\mathbf{C}$ do not depend on the values of the entries $\zeta_{\mathbf{B}}$ but depend only on their pattern. Then the collection $\mathcal{V}$ of all possible variance matrices $\mathbf{C}$ has common eigenspaces, called *strata*, which form the structure on Υ, and $\mathcal{V}$ is said to have *orthogonal variance structure* (OVS). OVS is called "orthogonal block structure" by Houtman and Speed in [23]. Note that there is no linear dependence among the (co)variances $\zeta_{\mathbf{B}}$. Unless otherwise stated, we assume that G is stratifiable and so $\mathbf{C}$ has OVS. Then the number of strata is equal to the number of adjacency matrices.

Let $\mathcal{Q}$ be the collection of symmetric, mutually orthogonal, idempotent matrices projecting onto the strata. Then $\sum_{\mathbf{Q}\in\mathcal{Q}} \mathbf{Q}$ is the $\Upsilon \times \Upsilon$ identity matrix $\mathbf{I}_\Upsilon$, and the variance matrix can be re-expressed as

$$\mathbf{C} = \sum_{\mathbf{Q}\in\mathcal{Q}} \eta_{\mathbf{Q}}\mathbf{Q}, \tag{3}$$

with $\eta_{\mathbf{Q}} \geq 0$ for all $\mathbf{Q}$ in $\mathcal{Q}$. The values $\eta_{\mathbf{Q}}$ are the eigenvalues of $\mathbf{C}$ and are called *spectral components of variance*. The stratum corresponding to $\mathbf{Q}$ is the subspace within which all normalized contrasts have variance $\eta_{\mathbf{Q}}$ under randomization. Given $\mathcal{Q}$, any two matrices of the form (3) commute with each other. The matrices $\mathbf{Q}$ are linear combinations of the matrices $\mathbf{B}$, and vice versa, but in general there is no closed-form expression for the coefficients in these combinations.

2.2. *Application to poset block structures.* Most experiments conducted in practice, and all examples in this paper, have poset block structures on their units. These straightforward generalizations of simple orthogonal block structures are defined in [3, 42–44] and shown in [5] to have stratifiable permutation groups.

A poset block structure on Υ is defined by a set $\mathcal{H}$ of generalized factors on Υ satisfying certain conditions. Following [45], we write $F < H$ if F and H are in $\mathcal{H}$ and F is marginal to H. There are several ways in which to write $\mathbf{C}$, in terms of matrices and coefficients that depend on H in $\mathcal{H}$ [42, 44]:

$$\mathbf{C} = \sum_{H\in\mathcal{H}} \zeta_H\mathbf{B}_H = \sum_{H\in\mathcal{H}} \psi_H\mathbf{S}_H = \sum_{H\in\mathcal{H}} \eta_H\mathbf{Q}_H. \tag{4}$$

Here, ζ_H is the (co)variance under the randomization between elements of Υ with the same level of H but not the same level of any generalized factor F in $\mathcal{H}$ to which H is marginal; $\mathbf{B}_H$ is the $\Upsilon \times \Upsilon$ adjacency matrix with entry 1 for such pairs and entry 0 otherwise; ψ_H is a canonical component; $\mathbf{S}_H$ is the $\Upsilon \times \Upsilon$ relationship matrix [24] for H, with (υ_1, υ_2)-entry equal to 1 if υ_1 and υ_2 have the same level of H and to 0 otherwise. Thus, $\mathbf{S}_H = \sum_{F \geq H} \mathbf{B}_F$. If k_H is the common replication of all levels of H, then $k_H^{-1}\mathbf{S}_H = \sum_{F \leq H} \mathbf{Q}_F$. When H is the generalized factor consisting of all factors on Υ, the subscript H will sometimes be replaced by Υ, while the subscript for the generalized factor corresponding to the overall mean is denoted 0. Expressions in [42, 44] show how to convert one set of coefficients in equation (4) to another. In particular, $\zeta_H = \sum_{F \leq H} \psi_F$ and

$$\eta_H = \sum_{F \geq H} k_F \psi_F. \tag{5}$$

The natural interpretation of *canonical components* in this context is as components of excess covariance [32]. They are linear combinations of the covariances (ζ-parameters) [29, 32, 44]. They measure the difference between the (co)variance of the responses on the units in Υ that have the same level of a particular generalized factor and the combined covariances of all generalized factors marginal to it. Thus, ψ_H can be negative: for example, when $F < H$, $\psi_H = \zeta_H - \zeta_F$ and $\zeta_H < \zeta_F$. However, $\psi_\Upsilon = \eta_\Upsilon \geq 0$. Estimates of standard errors of treatment effects require estimates of the spectral components. On the other hand, scientifically interesting hypotheses about the canonical components are often formulated and tested [19, 32] (see also Section 11.3) and so estimates of them may also be required.

Randomization-based models for experiments with poset block structure are mixed models in which the only constraint on the canonical components is that the spectral components be non-negative. This is weaker than the assumption commonly made in variance components models, namely, that all variance components σ_H^2 are non-negative. These mixed models are discussed further in Section 11.

2.3. *The randomization-based model for an experiment with two randomizations in a chain.* For a chain of two randomizations, there are three sets: Υ is randomized to Ω, and Γ is randomized to Υ. Let the corresponding design maps be $f: \Omega \to \Upsilon$ and $h: \Upsilon \to \Gamma$, as in Figure 2. The elements of Γ will be referred to as treatments and Υ and Ω as unrandomized sets.

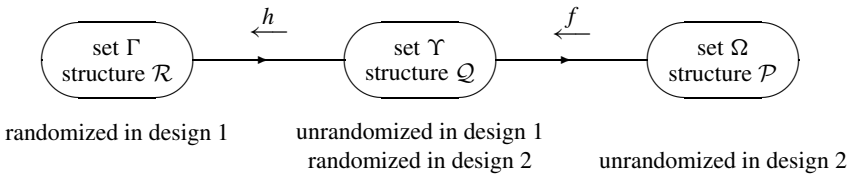

FIG. 2. *Diagram of an experiment with two randomizations in a chain.*

Suppose that f is randomized by choosing a random permutation from the group G_1 of permutations of Ω, and that G_1 is stratifiable with stratum projectors $\mathbf{P}$, for $\mathbf{P}$ in $\mathcal{P}$. Like the matrices $\mathbf{Q}$ in Section 2.1, the matrices $\mathbf{P}$ are known orthogonal idempotents summing to the $\Omega \times \Omega$ identity matrix $\mathbf{I}_\Omega$. In Section 2.1, the size of the idempotents is the size of Υ, while here it is the size of Ω.

Now let Y_ω be the response on observational unit ω in Ω. Applying the randomization argument from Section 2.1 to f gives $Y_\omega = Z_\omega + \tilde{Y}_{f(\omega)}$, where Z_ω is a random variable depending only on the unit ω and $\tilde{Y}_\upsilon$ is a notional effect associated with unit υ in Υ. Because G_1 is stratifiable, we can assume that the random variables Z_ω are identically distributed with mean zero, and that $\mathrm{Cov}(\mathbf{Z}) = \sum_{\mathbf{A}\in\mathcal{A}} \gamma_\mathbf{A}\mathbf{A}$, where $\mathcal{A}$ is the set of adjacency matrices arising from G_1 and the $\gamma_\mathbf{A}$ are the associated (co)variances. Following Section 2.1, we can also write $\mathrm{Cov}(\mathbf{Z}) = \sum_{\mathbf{P}\in\mathcal{P}} \xi_\mathbf{P}\mathbf{P}$ where, like the quantities $\eta_\mathbf{Q}$, the $\xi_\mathbf{P}$ are unknown nonnegative coefficients. Then the set of all possible matrices for $\mathrm{Cov}(\mathbf{Z})$ has OVS because G_1 is stratifiable.

Similarly, h is randomized by choosing a random permutation from the group G_2 of permutations of Υ, and G_2 is stratifiable with $\Upsilon \times \Upsilon$ stratum projectors $\mathbf{Q}$, for $\mathbf{Q}$ in $\mathcal{Q}$, as in Section 2.1. Rewriting equation (2) as $\tilde{Y}_\upsilon = W_\upsilon + \tau_{h(\upsilon)}$ gives

$$Y_\omega = Z_\omega + \tilde{Y}_{f(\omega)} = Z_\omega + W_{f(\omega)} + \tau_{h(f(\omega))}. \tag{6}$$

In turn, this randomization-based model can be rewritten in vector form as $\mathbf{Y} = \mathbf{Z} + \mathbf{X}_f\mathbf{W} + \mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}$, where $\mathbf{X}_f$ is the $\Omega \times \Upsilon$ design matrix for f. Hence, $\mathrm{E}(\mathbf{Y}) = \mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}$, and the variance matrix $\mathbf{V}$ of $\mathbf{Y}$ is given by

$$\mathbf{V} = \mathrm{Cov}(\mathbf{Z} + \mathbf{X}_f\mathbf{W}) = \sum_{\mathbf{A}\in\mathcal{A}} \gamma_\mathbf{A}\mathbf{A} + \sum_{\mathbf{B}\in\mathcal{B}} \zeta_\mathbf{B}\mathbf{X}_f\mathbf{B}\mathbf{X}_f' = \sum_{\mathbf{P}\in\mathcal{P}} \xi_\mathbf{P}\mathbf{P} + \sum_{\mathbf{Q}\in\mathcal{Q}} \eta_\mathbf{Q}\mathbf{X}_f\mathbf{Q}\mathbf{X}_f',$$

because $\mathbf{Z}$ and $\mathbf{W}$ are independent. The two sets $\mathcal{P}$ and $\mathcal{Q}$ of idempotents correspond to the eigenspaces of the variance matrices of $\mathbf{Z}$ and $\mathbf{W}$, respectively, but not necessarily to those of $\mathbf{V}$. Although the coefficients $\xi_\mathbf{P}$ and $\eta_\mathbf{Q}$ may not be eigenvalues of $\mathbf{V}$, we still call them spectral components of variance because they are the eigenvalues of $\mathrm{Cov}(\mathbf{Z})$ and $\mathrm{Cov}(\mathbf{W})$, respectively.

As noted in [12], the effect of the design function f is to embed a copy V_Υ^f of V_Υ inside the space V_Ω of real vectors indexed by Ω. Let $\mathbf{D}_f$ be the $\Upsilon \times \Upsilon$ diagonal matrix whose (ν, ν)-entry is the replication of unit ν. Then $\mathbf{X}_f'\mathbf{X}_f = \mathbf{D}_f$, and the matrix of orthogonal projection onto V_Υ^f is $\mathbf{X}_f\mathbf{D}_f^{-1}\mathbf{X}_f'$.

To further simplify $\mathbf{V}$, the design f be must be equi-replicate. If $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are in $\mathcal{Q}$ then $(\mathbf{X}_f\mathbf{Q}_1\mathbf{X}_f')(\mathbf{X}_f\mathbf{Q}_2\mathbf{X}_f') = \mathbf{X}_f\mathbf{Q}_1\mathbf{D}_f\mathbf{Q}_2\mathbf{X}_f'$. If the common replication is r then $\mathbf{D}_f = r\mathbf{I}_\Upsilon$, so if we put $\mathbf{Q}^f = r^{-1}\mathbf{X}_f\mathbf{Q}\mathbf{X}_f'$ then the $\mathbf{Q}^f$ are mutually orthogonal idempotents summing to $r^{-1}\mathbf{X}_f\mathbf{X}_f'$, which is the matrix of orthogonal

projection onto the subspace V_Υ^f. To simplify notation, as in [12], we shall write $\mathbf{Q}^f$ just as $\mathbf{Q}$, $\{\mathbf{Q}^f : \mathbf{Q} \in \mathcal{Q}\}$ as $\mathcal{Q}$, and $r^{-1}\mathbf{X}_f\mathbf{X}_f'$ as $\mathbf{I}_\mathcal{Q}$. Thus, we have

$$\mathbf{V} = \sum_{\mathbf{P}\in\mathcal{P}} \xi_\mathbf{P}\mathbf{P} + r\sum_{\mathbf{Q}\in\mathcal{Q}} \eta_\mathbf{Q}\mathbf{Q}. \tag{7}$$

This formula for $\mathbf{V}$ is similar to that in equation (2) of [49], but there are three differences. Here, the two collections of idempotents sum to $\mathbf{I}$ and $\mathbf{I}_\mathcal{Q}$, respectively, whereas those in [49] both sum to $\mathbf{I}$. Equation (7) is justified by the randomization; the formula in [49] is an assumed model. Finally, [49] does not require f to be equi-replicate, so the parameterization does not explicitly include the replication r.

2.4. *Pairs of poset block structures.* If the structure on Υ is a poset block structure with set $\mathcal{H}_2$ of generalized factors, then $r\sum_{H\in\mathcal{H}_2}\eta_H\mathbf{Q}_H^f = \sum_{H\in\mathcal{H}_2}\psi_H\mathbf{S}_H^f$, where $\mathbf{S}_H^f = \mathbf{X}_f\mathbf{S}_H\mathbf{X}_f'$ which is the $\Omega\times\Omega$ relationship matrix for H when it is regarded as a factor on Ω, in which case the common replication of its levels is rk_H. If we now write $\mathbf{S}_H^f$ just as $\mathbf{S}_H$, we have $r\sum_{H\in\mathcal{H}_2}\eta_H\mathbf{Q}_H = \sum_{H\in\mathcal{H}_2}\psi_H\mathbf{S}_H$.

Suppose that the structure on Ω is also a poset block structure, with set $\mathcal{H}_1$ of generalized factors. For H in $\mathcal{H}_1$, let the $\Omega\times\Omega$ relationship matrix for H be $\mathbf{T}_H$, with corresponding canonical component ϕ_H and common replication k_H. Then

$$\sum_{H\in\mathcal{H}_1}\xi_H\mathbf{P}_H = \sum_{H\in\mathcal{H}_1}\phi_H\mathbf{T}_H \quad\text{and}\quad \xi_H = \sum_{F\in\mathcal{H}_1, F\geq H} k_F\phi_F.$$

Thus, when both structures are poset block structures, equation (7) becomes

$$\mathbf{V} = \sum_{H\in\mathcal{H}_1}\phi_H\mathbf{T}_H + \sum_{H\in\mathcal{H}_2}\psi_H\mathbf{S}_H. \tag{8}$$

As noted in Section 2.2, even for poset block structures the randomization-based model for variance differs from a variance-components model. In equation (7), it is the coefficients $\xi_\mathbf{P}$ and $\eta_\mathbf{Q}$ which must be non-negative; the corresponding canonical components, except for ϕ_Ω and ψ_Υ, may well be negative.

3. Treatment decomposition and structure balance.

3.1. *Families of expectation models in a two-tiered experiment.* Consider the two-tiered set-up in Section 2.1. The design function h embeds a copy V_Γ^h of V_Γ inside V_Υ. Let $\mathbf{D}_h$ be the $\Gamma\times\Gamma$ diagonal matrix of replications of treatments. Then $\mathbf{X}_h'\mathbf{X}_h = \mathbf{D}_h$, and the matrix of orthogonal projection onto V_Γ^h is $\mathbf{X}_h\mathbf{D}_h^{-1}\mathbf{X}_h'$, which we write as $\mathbf{I}_\mathcal{R}$, because we always associate a structure $\mathcal{R}$ with Γ. The elements of $\mathcal{R}$ are derived from a family $\mathcal{M}$ of expectation models on Γ, as we now show.

With treatment effects fixed, data analysis usually proceeds by selecting a model from $\mathcal{M}$ and then estimating its parameters; see [4]. We assume that $\mathcal{M}$ defines

an orthogonal decomposition of V_Γ, in the following sense. There is a collection $\mathcal{R}$ of $\Gamma \times \Gamma$ symmetric, mutually orthogonal, idempotent matrices whose sum is the $\Gamma \times \Gamma$ identity matrix $\mathbf{I}_\Gamma$; each non-zero model in $\mathcal{M}$ is the subspace of V_Γ corresponding to a sum of one or more of the idempotents in $\mathcal{R}$; if M is such a model then there is at least one idempotent $\mathbf{R}$ in $\mathcal{R}$ such that $\mathrm{Im}(\mathbf{R}) \leq M$ and $M \cap (\mathrm{Im}(\mathbf{R}))^\perp$ is in $\mathcal{M}$; each $\mathbf{R}$ in $\mathcal{R}$ occurs at least once in this way, so that it corresponds to the extra sum of squares for fitting a larger model compared to a smaller model.

For $\mathbf{R}$ in $\mathcal{R}$, the subspace $\mathrm{Im}(\mathbf{R})$ of V_Γ is translated by h into a subspace of V_Γ^h whose $\Upsilon \times \Upsilon$ matrix $\mathbf{R}^h$ of orthogonal projection is $\mathbf{X}_h\mathbf{R}(\mathbf{R}\mathbf{D}_h\mathbf{R})^-\mathbf{R}\mathbf{X}_h'$. We require that h have the property that all such matrices commute with each other. When $\mathcal{M}$ is defined by a collection of orthogonal factors on Γ, this requirement is equivalent to the condition that the factors remain orthogonal when considered as factors on Υ. In the two-tiered context, we shall write $\mathbf{R}^h$ and $\{\mathbf{R}^h : \mathbf{R} \in \mathcal{R}\}$ simply as $\mathbf{R}$ and $\mathcal{R}$ from now on, so that $\sum_{\mathbf{R}\in\mathcal{R}} \mathbf{R} = \mathbf{I}_\mathcal{R}$.

There is no requirement for the design h to be equi-replicate. For example, suppose that Γ consists of the two levels of a treatment factor. If we parameterize the expectations as $\mu + \alpha$ and $\mu - \alpha$ then the estimators of μ and α are not orthogonal unless the levels are equally replicated. A model-focussed approach has one model M_1 in which we parameterize the expectations as α_1 and α_2, with a submodel M_2 in which they are both μ, and a further submodel M_3 in which both expectations are zero. Then $\mathcal{R} = \{\mathbf{R}_1, \mathbf{R}_0\}$, where $\mathbf{R}_0 = |\Upsilon|^{-1}\mathbf{J}$, which is the projector for the grand mean, and $\mathbf{R}_1 = \mathbf{I}_\mathcal{R} - \mathbf{R}_0$. Thus, we do have orthogonality, whether or not the replications are equal.

3.2. *Structure balance in a two-tiered experiment.* Until Section 9 inclusive, we insist that h be such that $\mathcal{R}$ is structure balanced in relation to $\mathcal{Q}$, in the sense defined in [12]. This means that there are scalars $\lambda_{\mathbf{QR}}$, for $\mathbf{Q}$ in $\mathcal{Q}$ and $\mathbf{R}$ in $\mathcal{R}$, such that $\mathbf{R}\mathbf{Q}\mathbf{I}_\mathcal{R} = \lambda_{\mathbf{QR}}\mathbf{R}$. Thus, (i) $\mathbf{R}\mathbf{Q}\mathbf{R} = \lambda_{\mathbf{QR}}\mathbf{R}$ and (ii) if $\mathbf{R}_1 \neq \mathbf{R}_2$ then $\mathbf{R}_1\mathbf{Q}\mathbf{R}_2 = \mathbf{0}$. The scalars $\lambda_{\mathbf{QR}}$ are called *efficiency factors*. It follows that each $\mathbf{Q}$ is the sum of the following mutually orthogonal idempotents: (i) $\mathbf{Q} \rhd \mathbf{R}$, for all $\mathbf{R}$ in $\mathcal{R}$ with $\lambda_{\mathbf{QR}} \neq 0$, and (ii) if it is non-zero, $\mathbf{Q} \vdash \mathcal{R}$. These idempotents are defined by $\mathbf{Q} \rhd \mathbf{R} = \lambda_{\mathbf{QR}}^{-1}\mathbf{Q}\mathbf{R}\mathbf{Q}$ and $\mathbf{Q} \vdash \mathcal{R} = \mathbf{Q} - \sum'_{\mathbf{R}\in\mathcal{R}} \mathbf{Q} \rhd \mathbf{R}$, where the summation is over those $\mathbf{R}$ for which $\lambda_{\mathbf{QR}} \neq 0$. This set of idempotents is denoted $\mathcal{Q} \rhd \mathcal{R}$ in [12].

For each $\mathbf{R}$ in $\mathcal{R}$, the efficiency factors $\lambda_{\mathbf{QR}}$ are non-negative and sum to 1. If each $\mathbf{R}$ has some $\mathbf{Q}$ in $\mathcal{Q}$ such that $\lambda_{\mathbf{QR}} = 1$ then the structure $\mathcal{R}$ is said to be *orthogonal* in relation to the structure $\mathcal{Q}$.

The matrices $\mathbf{Q}$ are determined by the group of permutations used for randomizing, and hence ultimately by the relevant information about Υ, such as blocks or managerial constraints. On the other hand, the matrices $\mathbf{R}$ depend on the family of expectation models chosen as appropriate. The former cannot be altered, but the latter may be refined, perhaps using pseudofactors, in order to achieve structure

balance [27, 50]. This is achieved by judicious replacement of some matrices $\mathbf{R}$ in $\mathcal{R}$ by sub-idempotents so that there is a refinement of the decomposition given by $\mathcal{R}$ into smaller subspaces; see [12], Section 4. Thus, we have a larger collection $\mathcal{R}^*$ of mutually orthogonal idempotents, such that each $\mathbf{R}$ in $\mathcal{R}$ is a sum of one or more of the idempotents in $\mathcal{R}^*$. For example, in a balanced lattice square design for k^2 treatments in $(k+1)/2$ squares, where k is odd, $\mathcal{R} = \{\mathbf{R}_0, \mathbf{R}_\mathrm{T}\}$ where $\mathbf{R}_0$ and $\mathbf{R}_\mathrm{T}$ are the idempotents for the Mean and Treatments, respectively. However, $\mathcal{R}$ is not structure balanced in relation to the structure $\mathcal{Q}$ defined by $(k+1)/2$ squares, each formed by k rows crossed with k columns. We form $\mathcal{R}^*$ by replacing $\mathbf{R}_\mathrm{T}$ by $\mathbf{R}_{\mathrm{T,R}}$ and $\mathbf{R}_{\mathrm{T,C}}$, which are the idempotents corresponding to the treatment subspaces partly confounded with rows and columns, respectively: then $\mathcal{R}^*$ is structure balanced in relation to $\mathcal{Q}$ (see [12], Example 5).

3.3. *Treatment structure and structure balance in a three-tiered experiment with two randomizations in a chain.* Now consider the three-tiered set-up in Section 2.3. As in Section 3.1, the effects $\boldsymbol{\tau}$ are taken to be fixed, and so we assume that the family of expectation models gives a set of mutually orthogonal idempotents $\mathbf{R}$ in $\mathcal{R}$ whose sum is the orthogonal projector onto V_Γ^h in V_Υ. Let $\mathbf{M}^h$ be the $\Upsilon \times \Upsilon$ idempotent for one of these expectation models. The corresponding $\Omega \times \Omega$ idempotent $(\mathbf{M}^h)^f$ is given by $(\mathbf{M}^h)^f = \mathbf{X}_f\mathbf{M}^h(\mathbf{M}^h\mathbf{D}_f\mathbf{M}^h)^-\mathbf{M}^h\mathbf{X}'_f = r^{-1}\mathbf{X}_f\mathbf{M}^h\mathbf{X}'_f$ since $\mathbf{D}_f = r\mathbf{I}_\Upsilon$. Therefore, putting $\mathbf{R}^f = r^{-1}\mathbf{X}_f\mathbf{R}\mathbf{X}'_f$ for $\mathbf{R}$ in $\mathcal{R}$, we see that the mutually orthogonal idempotents $\mathbf{R}$ in $\mathcal{R}$ translate to mutually orthogonal idempotents $\mathbf{R}^f$ on V_Ω. That is, because f is equi-replicate, the same formula is used to convert both the expectation idempotents and the variance idempotents from $\Upsilon \times \Upsilon$ matrices to $\Omega \times \Omega$ matrices. There is still no need for h to be equi-replicate. As in [12], we shall write $\mathbf{R}^f$ as $\mathbf{R}$ and $\{\mathbf{R}^f : \mathbf{R} \in \mathcal{R}\}$ as $\mathcal{R}$ in the three-tiered context. We continue to write $\sum_{\mathbf{R}\in\mathcal{R}} \mathbf{R}$ as $\mathbf{I}_\mathcal{R}$, which is now an $\Omega \times \Omega$ matrix.

In addition to the condition that f be equi-replicate, we also assume until Section 9 inclusive that $\mathcal{Q}$ is structure balanced in relation to $\mathcal{P}$, or can be made so, as described in Section 3.4. Then $\mathcal{R}$ is structure balanced in relation to $\mathcal{P} \rhd \mathcal{Q}$, $\mathcal{Q} \rhd \mathcal{R}$ is structure balanced in relation to $\mathcal{P}$, and $(\mathcal{P} \rhd \mathcal{Q}) \rhd \mathcal{R} = \mathcal{P} \rhd (\mathcal{Q} \rhd \mathcal{R})$ [12].

Let $\mathcal{Q}_1$ be the set of $\mathbf{Q}$ in $\mathcal{Q}$ for which there is an idempotent $\mathbf{P}$ in $\mathcal{P}$ with $\lambda_{\mathbf{PQ}} = 1$. Define the function c from $\mathcal{Q}_1$ to $\mathcal{P}$ such that $c(\mathbf{Q}) = \mathbf{P}$ for $\lambda_{\mathbf{PQ}} = 1$. If $\mathbf{Q} \in \mathcal{Q}_1$ then $\mathrm{Im}(\mathbf{Q}) \le \mathrm{Im}(c(\mathbf{Q}))$, and $\mathbf{QP} = \mathbf{PQ} = \mathbf{Q} = \mathbf{P} \rhd \mathbf{Q}$ if $\mathbf{P} = c(\mathbf{Q})$, while $\mathbf{QP} = \mathbf{PQ} = \mathbf{0}$ otherwise. Thus, $\mathcal{Q}$ is orthogonal in relation to $\mathcal{P}$ when $\mathcal{Q}_1 = \mathcal{Q}$.

For $\mathbf{P} \in \mathcal{P}$, equation (7) shows that $\mathbf{V}(\mathbf{P} \vdash \mathcal{Q}) = \xi_\mathbf{P}(\mathbf{P} \vdash \mathcal{Q})$, because $(\mathbf{P} \vdash \mathcal{Q})\mathbf{Q} = \mathbf{0}$ for all $\mathbf{Q}$ in $\mathcal{Q}$. Hence, $\mathrm{Im}(\mathbf{P} \vdash \mathcal{Q})$ is contained in an eigenspace of $\mathbf{V}$ with eigenvalue $\xi_\mathbf{P}$. Moreover, if $\mathbf{Q} \in \mathcal{Q}$ and $\lambda_{\mathbf{PQ}} \neq 0$, then

$$\mathbf{V}(\mathbf{P} \rhd \mathbf{Q}) = \mathbf{V}\frac{\mathbf{PQP}}{\lambda_{\mathbf{PQ}}} = \frac{\xi_\mathbf{P}}{\lambda_{\mathbf{PQ}}}\mathbf{PQP} + r\sum_{\mathbf{Q}^*}\frac{\eta_{\mathbf{Q}^*}}{\lambda_{\mathbf{PQ}}}\mathbf{Q}^*\mathbf{PQP} = \xi_\mathbf{P}(\mathbf{P} \rhd \mathbf{Q}) + r\eta_\mathbf{Q}\mathbf{QP}.$$

If $\mathbf{Q} \in \mathcal{Q}_1$ and $\mathbf{P} = c(\mathbf{Q})$ then $\mathrm{Im}(\mathbf{P} \rhd \mathbf{Q})$ is contained in an eigenspace of $\mathbf{V}$ with eigenvalue $\xi_{\mathbf{P}} + r\eta_{\mathbf{Q}}$. Otherwise, $\mathrm{Im}(\mathbf{P} \rhd \mathbf{Q})$ is not contained in any eigenspace of $\mathbf{V}$.

If $\mathcal{Q}$ is orthogonal in relation to $\mathcal{P}$, then

$$\mathbf{V} = \sum_{\mathbf{Q} \in \mathcal{Q}} (\xi_{c(\mathbf{Q})} + r\eta_{\mathbf{Q}})\mathbf{Q} + \sum_{\mathbf{P} \in \mathcal{P}} \xi_{\mathbf{P}}(\mathbf{P} \vdash \mathcal{Q}). \tag{9}$$

The idempotents in this expression are those in $\mathcal{P} \rhd \mathcal{Q}$, and the image of each is contained in an eigenspace of $\mathbf{V}$. Thus, the set of all positive semidefinite (p.s.d.) matrices of the form (9) commute with each other, and have common eigenspaces: we call this *commutative variance structure* (CVS). If, in addition, there is no linear dependence among the coefficients in (9), we have OVS.

3.4. *Choice of idempotents*. The matrices $\mathbf{P}$ are defined by the group G_1 of permutations used to randomize the design f. The matrices $\mathbf{Q}$ are first defined as matrices on V_Υ by the group G_2 of permutations used to randomize the design h, and then translated by f to matrices on V_Ω. The matrices $\mathbf{R}$ depend initially on the chosen family of expectation models, and are translated by h and then by f.

Strictly speaking, there is no freedom of choice over the $\mathbf{Q}$ matrices. However, as already outlined for design h in Section 3.2, it is sometimes possible to turn a design f without structure balance into one with structure balance by judicious replacement of some matrices $\mathbf{Q}$ in $\mathcal{Q}$ by sub-idempotents, yielding $\mathcal{Q}^*$. The variance matrix in equation (7) is defined by the original $\mathcal{Q}$: when it is rewritten in terms of $\mathcal{Q}^*$ it has the constraint that if $\mathbf{Q}$ in $\mathcal{Q}$ is the sum $\mathbf{Q}_1^* + \cdots + \mathbf{Q}_n^*$ with $\mathbf{Q}_i^*$ in $\mathcal{Q}^*$ then each of $\mathbf{Q}_1^*, \ldots, \mathbf{Q}_n^*$ has the *same* spectral component $\eta_{\mathbf{Q}}$.

There are two types of multiple randomization that form a chain; see [11, 12]. For *composed* randomizations, the randomizations may be done in either order, because neither needs knowledge of the outcome of the other. In contrast, *randomized-inclusive* randomizations have the complication that knowledge of the outcome of the randomization of Γ to Υ is needed before Υ can be randomized to Ω.

As explained in [11], Section 5.1 and [12], Section 6, this knowledge is needed in the second case partly because the structure $\mathcal{Q}$ on Υ defined by the randomization of design h is not structure balanced in relation to $\mathcal{P}$. Thus, $\mathcal{Q}$ needs to be refined into sub-idempotents, also called pseudosources, as described above. The second necessary ingredient for randomized-inclusive randomizations is that at least one source on Γ is confounded or partly confounded with one of the sources on Υ that needs to be split up. In order to work out the partial confounding of sources on Γ with those on Ω, it is necessary to keep track of the partial confounding of the former with the pseudosources on Υ. This may require pseudosources on Γ. Most importantly, the unrandomized version of f is constrained to ensure the correct partial confounding of (pseudo)sources on Γ with those on Ω.

Although this procedure is more complicated than that for composed randomizations, the randomization-based model is virtually the same. As above, we have to keep track of pseudosources. For those on Υ, it is important to remember that pseudosources of the same source have the *same* spectral component η. This complication can occur for experiments with two composed randomizations when the second randomization is not consonant, as defined in [11]. It always occurs for experiments with two randomized-inclusive randomizations; see Examples 2.1 and 2.2.

4. Analysis of variance.

4.1. *A two-tiered experiment.* Consider the two-tiered experiment in Sections 2.1 and 3.2. In [12], decomposition tables were used to display the decomposition of V_Υ appropriate for such an experiment. Such a table is a precursor to an anova table and consists of rows and columns. There is a set of columns for each tier: one column containing sources, one column containing degrees of freedom and, if the design is structure balanced but not orthogonal, a further column showing efficiency factors. The sources and pseudosources correspond to idempotents in $\mathcal{Q}$ or $\mathcal{R}$ which, when they are based on generalized (pseudo)factors, are labelled as described in Section 3 of [12]. Each row of the decomposition table corresponds to a subspace in the decomposition specified by $\mathcal{Q} \triangleright \mathcal{R}$. In this paper, we add a column for expected mean squares to decomposition tables in order to form skeleton anova tables.

The anova table for the analysis of a response variable when the variance matrix has the form (3), the τ_i are fixed effects and $\mathcal{R}$ is structure balanced in relation to $\mathcal{Q}$ is given in [2, 23, 30]. The data vector $\mathbf{y}$ is projected onto each stratum in turn and then $\mathbf{Qy}$, which is the projection into stratum $\mathrm{Im}(\mathbf{Q})$, is further decomposed according to the elements of $\mathcal{Q} \triangleright \mathcal{R}$ involving $\mathbf{Q}$. The following hold:

(A.a) The projections onto different strata are uncorrelated.

(A.b) Any orthonormal basis for $\mathrm{Im}(\mathbf{Q})$ gives uncorrelated random variables all with variance $\eta_{\mathbf{Q}}$.

(A.c) If $\lambda_{\mathbf{QR}} \neq 0$, then the expected mean square for $\mathbf{Q} \triangleright \mathbf{R}$ is equal to

$$\eta_{\mathbf{Q}} + \frac{\lambda_{\mathbf{QR}} \boldsymbol{\tau}' \mathbf{X}_h' \mathbf{R} \mathbf{X}_h \boldsymbol{\tau}}{\mathrm{rank}(\mathbf{R})}.$$

(A.d) If $\mathbf{Q} \vdash \mathcal{R}$ is non-zero, then the expected mean square for $\mathbf{Q} \vdash \mathcal{R}$ is equal to $\eta_{\mathbf{Q}}$.

For poset block structures, the expected mean squares in terms of the canonical components can obtained from those in (A.c) and (A.d) using equation (5).

The expression $\boldsymbol{\tau}' \mathbf{X}_h' \mathbf{R} \mathbf{X}_h \boldsymbol{\tau} / \mathrm{rank}(\mathbf{R})$ in (A.c) is a p.s.d. quadratic form in the parameters τ_i. If $\mathcal{R}$ is defined by a poset block structure on Γ then $\mathbf{R} = \mathbf{R}_F$ for a generalized factor F on Γ, just as $\mathbf{Q} = \mathbf{Q}_H$ in equation (4). In anova tables,

this expression is written as $q(F)$. In particular, $q_0 = \boldsymbol{\tau}'\mathbf{X}_h'\mathbf{R}_0\mathbf{X}_h\boldsymbol{\tau}$, where $\mathbf{R}_0 = |\Upsilon|^{-1}\mathbf{J}$.

4.2. *An experiment with two randomizations in a chain.* First, consider expectations. If $\mathbf{P} \in \mathcal{P}$, then $(\mathbf{P} \vdash \mathcal{Q})\mathbf{I}_\mathcal{Q} = \mathbf{0}$. If, further, $\mathbf{Q} \in \mathcal{Q}$ and $\lambda_{\mathbf{PQ}} \neq 0$, then $((\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R})\mathbf{I}_\mathcal{Q}\mathbf{I}_\mathcal{R} = \mathbf{0}$. Since $\mathrm{E}(\mathbf{Y}) = \mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau} = \mathbf{I}_\mathcal{Q}\mathbf{I}_\mathcal{R}\mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}$, it follows that $\mathrm{E}((\mathbf{P} \vdash \mathcal{Q})\mathbf{Y}) = \mathrm{E}(((\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R})\mathbf{Y}) = \mathbf{0}$. If, moreover, $\mathbf{R} \in \mathcal{R}$ and $\lambda_{\mathbf{QR}} \neq 0$, Section 5 of [12] shows that $(\mathbf{P} \rhd \mathbf{Q}) \rhd \mathbf{R} = \lambda_{\mathbf{PQ}}^{-1}\lambda_{\mathbf{QR}}^{-1}\mathbf{PQRQP}$. Therefore,

$$\mathrm{E}\big(((\mathbf{P} \rhd \mathbf{Q}) \rhd \mathbf{R})\mathbf{Y}\big) = \frac{1}{\lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}}}\mathbf{PQRQPI}_\mathcal{Q}\mathbf{I}_\mathcal{R}\mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau} = \mathbf{PQRX}_f\mathbf{X}_h\boldsymbol{\tau}.$$

Hence,

$$\begin{aligned}\big(\mathrm{E}(((\mathbf{P} \rhd \mathbf{Q}) \rhd \mathbf{R})\mathbf{Y})\big)'\mathrm{E}(((\mathbf{P} \rhd \mathbf{Q}) \rhd \mathbf{R})\mathbf{Y}) &= \boldsymbol{\tau}'\mathbf{X}_h'\mathbf{X}_f'\mathbf{RQPPQRX}_f\mathbf{X}_h\boldsymbol{\tau} \\ &= \lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}}\boldsymbol{\tau}'\mathbf{X}_h'\mathbf{X}_f'\mathbf{RX}_f\mathbf{X}_h\boldsymbol{\tau}.\end{aligned}$$

Consider a fixed $\mathbf{P}$ in $\mathcal{P}$. Equation (7) shows that

$$\begin{aligned}\mathrm{Cov}(\mathbf{PY}) = \mathbf{PVP} &= \xi_\mathbf{P}\mathbf{P} + r\sum_{\mathbf{Q}\in\mathcal{Q}}\eta_\mathbf{Q}\mathbf{PQP} = \xi_\mathbf{P}\mathbf{P} + r\sum_{\mathbf{Q}\in\mathcal{Q}}{}'\eta_\mathbf{Q}\lambda_{\mathbf{PQ}}\mathbf{P} \rhd \mathbf{Q} \\ &= \sum_{\mathbf{Q}\in\mathcal{Q}}{}'(\xi_\mathbf{P} + r\lambda_{\mathbf{PQ}}\eta_\mathbf{Q})(\mathbf{P} \rhd \mathbf{Q}) + \xi_\mathbf{P}(\mathbf{P} \vdash \mathcal{Q}).\end{aligned} \tag{10}$$

Here, $\sum'_{\mathbf{Q}\in\mathcal{Q}}$ denotes summation over $\mathbf{Q} \in \mathcal{Q}$ with $\lambda_{\mathbf{PQ}} \neq 0$. The matrices in equation (10) are mutually orthogonal idempotents which sum to $\mathbf{P}$ and have linearly independent coefficients. Hence, they are the projectors onto the eigenspaces of $\mathrm{Cov}(\mathbf{PY})$ with non-zero eigenvalues. Therefore, the results for $\mathbf{Y}$ in Section 4.1 carry over to $\mathbf{PY}$ as follows:

(A.e) The projections onto any two different subspaces of the form $\mathrm{Im}(\mathbf{P} \rhd \mathbf{Q})$ or $\mathrm{Im}(\mathbf{P} \vdash \mathcal{Q})$ are uncorrelated.

(A.f) If $\lambda_{\mathbf{PQ}} \neq 0$, any orthonormal basis for $\mathrm{Im}(\mathbf{P} \rhd \mathbf{Q})$ gives uncorrelated random variables all with variance $\xi_\mathbf{P} + r\lambda_{\mathbf{PQ}}\eta_\mathbf{Q}$.

(A.g) Any orthonormal basis for $\mathrm{Im}(\mathbf{P} \vdash \mathcal{Q})$ gives uncorrelated random variables all with variance $\xi_\mathbf{P}$.

(A.h) If $\lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}} \neq 0$, then the expected mean square for $(\mathbf{P} \rhd \mathbf{Q}) \rhd \mathbf{R}$ is

$$\xi_\mathbf{P} + r\lambda_{\mathbf{PQ}}\eta_\mathbf{Q} + \frac{\lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}}\boldsymbol{\tau}'\mathbf{X}_h'\mathbf{X}_f'\mathbf{RX}_f\mathbf{X}_h\boldsymbol{\tau}}{\mathrm{rank}(\mathbf{R})}.$$

(A.i) If $\lambda_{\mathbf{PQ}} \neq 0$ and $(\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R}$ is non-zero, then the expected mean square for $(\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R}$ is $\xi_\mathbf{P} + r\lambda_{\mathbf{PQ}}\eta_\mathbf{Q}$.

(A.j) If $\mathbf{P} \vdash \mathcal{Q}$ is non-zero, then the expected mean square for $\mathbf{P} \vdash \mathcal{Q}$ is $\xi_\mathbf{P}$.

For poset block structures, the spectral components $\xi_{\mathbf{P}}$ and $\eta_{\mathbf{Q}}$ can be expanded to express the expected mean squares in terms of the canonical components.

We write the expression $\boldsymbol{\tau}'\mathbf{X}_h'\mathbf{X}_f'\mathbf{R}\mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}/\operatorname{rank}(\mathbf{R})$ as $q(F)$ if $\mathcal{R}$ is defined by a poset block structure on Γ and $\mathbf{R}=\mathbf{R}_F$ for some generalized factor F on Γ.

Finally, consider the whole of $\mathcal{P}$. If $\mathcal{Q}$ is orthogonal in relation to $\mathcal{P}$ then we have CVS and so the projected data corresponding to any two different rows of the anova table are uncorrelated. Otherwise, we have the situation, such as that in [33], where some subspaces corresponding to idempotents of the form $\mathbf{P}\rhd\mathbf{Q}$ do not consist of eigenvectors of $\mathbf{V}$. Then it is still possible to do anova in the sense of decomposing the sum of squares of the responses according to the subspaces, and equating the observed values of the mean squares to their expectations, but this may not have all the properties of classical anova.

In particular, let $\mathbf{Q}$ be an idempotent in $\mathcal{Q}$ for which there are distinct $\mathbf{P}_1$ and $\mathbf{P}_2$ in $\mathcal{P}$ with $\lambda_{\mathbf{P}_1\mathbf{Q}}$ and $\lambda_{\mathbf{P}_2\mathbf{Q}}$ both non-zero. Then the projections of the data onto $\operatorname{Im}(\mathbf{P}_1)$ and $\operatorname{Im}(\mathbf{P}_2)$ are not independent, because

$$\begin{aligned}\operatorname{Cov}\bigl((\mathbf{P}_1\rhd\mathbf{Q})\mathbf{Y},(\mathbf{P}_2\rhd\mathbf{Q})\mathbf{Y}\bigr)&=(\mathbf{P}_1\rhd\mathbf{Q})'\Bigl(\sum_{\mathbf{P}\in\mathcal{P}}\xi_{\mathbf{P}}\mathbf{P}+r\sum_{\mathbf{Q}^*\in\mathcal{Q}}\eta_{\mathbf{Q}^*}\mathbf{Q}^*\Bigr)(\mathbf{P}_2\rhd\mathbf{Q})\\&=\frac{\mathbf{P}_1\mathbf{Q}\mathbf{P}_1}{\lambda_{\mathbf{P}_1\mathbf{Q}}}\Bigl(\sum_{\mathbf{P}\in\mathcal{P}}\xi_{\mathbf{P}}\mathbf{P}+r\sum_{\mathbf{Q}^*\in\mathcal{Q}}\eta_{\mathbf{Q}^*}\mathbf{Q}^*\Bigr)\frac{\mathbf{P}_2\mathbf{Q}\mathbf{P}_2}{\lambda_{\mathbf{P}_2\mathbf{Q}}}\\&=\frac{r\eta_{\mathbf{Q}}}{\lambda_{\mathbf{P}_1\mathbf{Q}}\lambda_{\mathbf{P}_2\mathbf{Q}}}\mathbf{P}_1\mathbf{Q}\mathbf{P}_1\mathbf{Q}\mathbf{P}_2\mathbf{Q}\mathbf{P}_2=r\eta_{\mathbf{Q}}\mathbf{P}_1\mathbf{Q}\mathbf{P}_2,\end{aligned}$$

which has the same rank as $\mathbf{Q}$. A similar calculation shows that $(\mathbf{P}_1\rhd\mathbf{Q}_1)\mathbf{Y}$ is not correlated with $(\mathbf{P}_2\rhd\mathbf{Q}_2)\mathbf{Y}$ if $\mathbf{Q}_1\neq\mathbf{Q}_2$.

5. Examples. Our first example shows how straightforward the anova table is when both designs are orthogonal, in the sense defined in Section 3.2, and the randomizations are composed. The second demonstrates that, even in orthogonal experiments with multiple randomizations, an η-component can occur with more than one ξ-component. Moreover, the fact that the numbers of units in the middle and right-hand tiers are equal in Example 2.1 but not in Example 1.1 leads to different difficulties, as shown in Section 8.2. Examples 1.2 and 2.2 are variants of Examples 1.1 and 2.1 that show the complications introduced by non-orthogonality. The first uses non-orthogonal designs in both phases and the second a non-orthogonal design only in the second phase; this leads to their $\mathbf{V}$ matrices having different properties.

EXAMPLE 1.1 (continued). The two randomizations involved in the example in Section 1 each use an orthogonal design. Table 1 gives the skeleton anova, including the expected mean squares under randomization; it is a revised and expanded version of Table 2 of [12] and exhibits the properties of the two-phase design. There is no need to show efficiency factors, because both designs are orthogonal.

TABLE 1
Skeleton analysis of variance for Example 1.1

evaluations tier		vine-plots tier		treatments tier		
source	**d.f.**	**source**	**d.f.**	**source**	**d.f.**	**E.M.S.**
Mean	1	Mean	1	Mean	1	$\xi_0 + 16\eta_0 + q_0$
Occasions	2	Blocks	2			$\xi_{\mathrm{O}} + 16\eta_{\mathrm{B}}$
Judges[O]	45					ξ_{OJ}
Positions[O]	21					ξ_{OP}
J # P[O]	315	Vine-plots[B]	21	Trellis	3	$\xi_{\mathrm{OJP}} + 16\eta_{\mathrm{BV}} + q(\mathrm{T})$
				Methods	1	$\xi_{\mathrm{OJP}} + 16\eta_{\mathrm{BV}} + q(\mathrm{M})$
				T # M	3	$\xi_{\mathrm{OJP}} + 16\eta_{\mathrm{BV}} + q(\mathrm{TM})$
				Residual	14	$\xi_{\mathrm{OJP}} + 16\eta_{\mathrm{BV}}$
		Residual	294			ξ_{OJP}

One consequence of this simple orthogonality, and the lack of pseudosources, is that each η-component appears in the final column in conjunction with exactly one ξ-component. Under randomization all of these components must be non-negative. However, canonical components such as ϕ_{OJ} and ϕ_{OP} can be negative. For example, $\phi_{\mathrm{OJ}} = \gamma_{\mathrm{OJ}} - \gamma_{\mathrm{O}}$, which is negative if the covariance between responses by the same judge (on an occasion) is less than the covariance between responses on the same occasion by different judges at different positions.

The appropriate "Residual" for each of the three treatments sources is the one with 14 degrees of freedom, which is $((\mathrm{J}\,\#\,\mathrm{P[O]}) \rhd (\text{Vine-plots[B]})) \vdash \mathcal{R}$, where $\mathcal{R}$ is the structure on the treatments tier.

EXAMPLE 2.1 (Wheat with Latin square in the laboratory phase). This experiment has a field phase followed by a laboratory phase. In the field phase, seven lines of wheat are grown on 49 plots; then a sample of grain from each plot is analysed in the laboratory. The two-phase design used is like one suggested in Example 4 of [26] and investigated in Web Appendix Example 1 of [15]: the first phase uses a randomized complete-block design with seven blocks; in the second phase, the plots from the first phase are randomized to 49 analyses, that occur at seven consecutive times in each of seven runs of the machine.

Figure 3, equivalent to Web Appendix Figure 1 of [15], shows the two randomizations. They are in a chain, but are randomized-inclusive rather than composed, because the allocation of lines to plots must be known before plots can be randomized to analyses. A pseudofactor P_1 is introduced, each of whose levels identifies the group of plots receiving a single line. In the second phase, Blocks are aliased

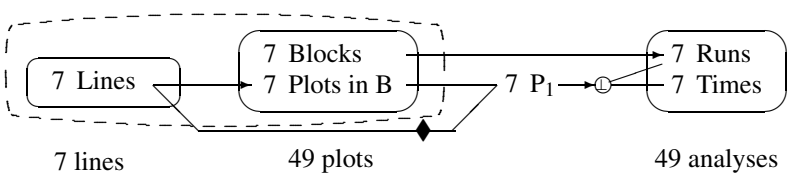

FIG. 3. *Randomized-inclusive randomizations in Example 2.1*: *lines are randomized to plots, and lines and plots are randomized to analyses*; B = *Blocks*; P_1 *is pseudofactor for Plots identifying plots assigned the same level of Lines*.

with Runs while levels of P_1 are allocated to make a Latin square whose rows and columns are Runs and Times. Runs and Times are then randomized independently.

The variance matrix under the randomizations is

$$\begin{aligned}\mathbf{V} &= \xi_0\mathbf{P}_0 + \xi_R\mathbf{P}_R + \xi_T\mathbf{P}_T + \xi_{RT}\mathbf{P}_{RT} + \eta_0\mathbf{Q}_0 + \eta_B\mathbf{Q}_B + \eta_{BP}\mathbf{Q}_{BP}\\ &= \phi_0\mathbf{T}_0 + \phi_R\mathbf{T}_R + \phi_T\mathbf{T}_T + \phi_{RT}\mathbf{T}_{RT} + \psi_0\mathbf{S}_0 + \psi_B\mathbf{S}_B + \psi_{BP}\mathbf{S}_{BP},\end{aligned}$$

where $\mathbf{T}_H$ and $\mathbf{S}_H$ are the relationship matrices and ϕ_H and ψ_H are the canonical components for a generalized factor H from the analyses or plots tiers, respectively.

However, $\mathcal{Q}$ cannot be structure-balanced in relation to $\mathcal{P}$, because no $\mathbf{P}$-matrix has rank as big as rank($\mathbf{Q}_{BP}$). As explained in Section 3.4, structure balance can be obtained by introducing a second pseudofactor P_2, each of whose levels identifies the group of plots allocated to a single time. Then $\mathbf{Q}_{BP}$ can be replaced by two $\mathbf{Q}^*$ matrices, one each for P_2 and the rest of Plots[Blocks]; both have coefficient η_{BP}.

The following expressions show how the canonical components in this example measure excess covariance:

$$\phi_{RT} = \gamma_{RT} - \gamma_R - \gamma_T + \gamma_0, \qquad \phi_R = \gamma_R - \gamma_0, \qquad \phi_T = \gamma_T - \gamma_0, \qquad \phi_0 = \gamma_0,$$

$$\psi_{BP} = \zeta_{BP} - \zeta_B, \qquad \psi_B = \zeta_B - \zeta_0, \qquad \psi_0 = \zeta_0.$$

Thus, ϕ_0, ϕ_R, ϕ_T and ϕ_{RT} measure, respectively, the basic covariance of "unrelated" analyses, the excess of the covariance of different analyses in the same run over that of "unrelated" analyses, the excess of the covariance of different analyses at the same time over that of "unrelated" analyses, and the excess of the variance of a single analysis over the appropriate linear combination of γ_0, γ_R and γ_T. The ψ-parameters from the plots tier can be similarly interpreted using the ζ-parameters.

The skeleton anova is in Table 2; both designs are orthogonal so efficiency factors are not shown. Now η_{BP} occurs with two different ξ-components because the source Plots[Blocks] has been split into two by the pseudofactor P_2. Table 2 shows that no spectral component is estimable and ϕ_T is the only estimable canonical component.

TABLE 2
Skeleton analysis of variance for Example 2.1 with expected mean squares in terms of spectral components and canonical components

analyses tier		plots tier[a]		lines tier		E.M.S.	
source	**d.f.**	**source**	**d.f.**	**source**	**d.f.**	**spectral components**	**canonical components**
Mean	1	Mean	1	Mean	1	$\xi_0 + \eta_0 + q_0$	$\phi_{RT} + 7\phi_R + 7\phi_T + 49\phi_0 + \psi_{BP} + 7\psi_B + 49\psi_0 + q_0$
Runs	6	Blocks	6			$\xi_R + \eta_B$	$\phi_{RT} + 7\phi_R + \psi_{BP} + 7\psi_B$
Times	6	P_2	6			$\xi_T + \eta_{BP}$	$\phi_{RT} + 7\phi_T + \psi_{BP}$
R # T	36	Plots[B]$_{\vdash}$	36	Lines	6	$\xi_{RT} + \eta_{BP} + q(\mathrm{L})$	$\phi_{RT} + \psi_{BP} + q(\mathrm{L})$
				Residual	30	$\xi_{RT} + \eta_{BP}$	$\phi_{RT} + \psi_{BP}$

[a] Plots[B]$_{\vdash}$ is the part of Plots[B] orthogonal to P_2.

EXAMPLE 1.2 (Elaborate two-phase sensory experiment). Section 3 of [16] gives a more elaborate version of Example 1.1. The randomization diagram for it, given in [11], is in Figure 4 and the decomposition table for it is derived in [12], Example 1. Here, the skeleton anova is in Table 3. Although there are pseudofactors for the Judges factor in the evaluations tier, they are ignored in doing the randomization as the six judges are permuted with no distinction. These pseudofactors are used only to obtain the systematic layout; they do not give rise to pseudosources.

In this example, neither design is orthogonal, and so efficiency factors are shown in the anova table. In the half-plots tier, only Columns[Squares] is not orthogonal to sources in the evaluations tier: thus, η_{QC} occurs in conjunction with two different

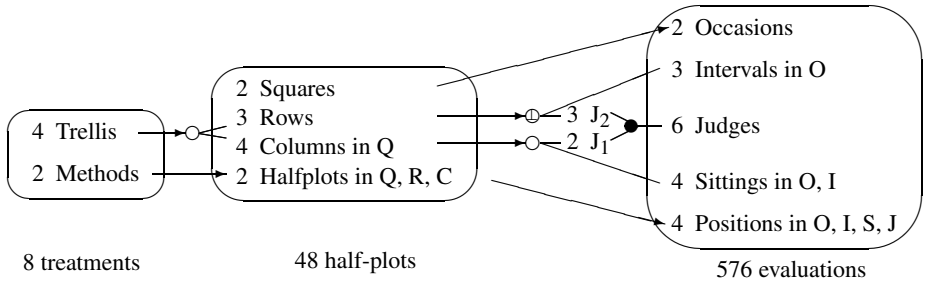

FIG. 4. *Randomization diagram for Example 1.2: treatments are randomized to half-plots, which are, in turn, randomized to evaluations*; Q, R, C, O, I, S, J *denote Squares, Rows, Columns, Occasions, Intervals, Sittings and Judges, respectively*; J_1 *and* J_2 *are pseudofactors for Judges.*

TABLE 3
Skeleton analysis of variance table for Example 1.2 (O = *Occasions*, I = *Intervals*, S = *Sittings*, J = *Judges*, P = *Positions*, Q = *Squares*, C = *Columns*, R = *Rows*, H = *Halfplots*, T = *Trellis*, M = *Methods*)

evaluations tier		**half-plots tier**			**treatments tier**			
source	**d.f.**	**eff.**	**source**	**d.f.**	**eff.**	**source**	**d.f.**	**E.M.S.**
Mean	1	1	Mean	1	1	Mean	1	$\xi_0 + 12\eta_0 + q_0$
O	1	1	Q	1				$\xi_O + 12\eta_Q$
I[O]	4							ξ_{OI}
S[O ∧ I]	18	$\frac{1}{3}$	C[Q]	6	$\frac{1}{27}$	T	3	$\xi_{OIS} + \frac{1}{3}12\eta_{QC} + \frac{1}{27}q(T)$
						Residual	3	$\xi_{OIS} + \frac{1}{3}12\eta_{QC}$
			Residual	12				ξ_{OIS}
J	5							ξ_J
O # J	5							ξ_{OJ}
I # J[O]	20	1	R	2				$\xi_{OIJ} + 12\eta_R$
		1	R # Q	2				$\xi_{OIJ} + 12\eta_{QR}$
			Residual	16				ξ_{OIJ}
S # J[O ∧ I]	90	$\frac{2}{3}$	C[Q]	6	$\frac{2}{27}$	T	3	$\xi_{OISJ} + \frac{2}{3}12\eta_{QC} + \frac{2}{27}q(T)$
						Residual	3	$\xi_{OISJ} + \frac{2}{3}12\eta_{QC}$
		1	R # C[Q]	12	$\frac{8}{9}$	T	3	$\xi_{OISJ} + 12\eta_{QRC} + \frac{8}{9}q(T)$
						Residual	9	$\xi_{OISJ} + 12\eta_{QRC}$
			Residual	72				ξ_{OISJ}
P[O ∧ I ∧ S ∧ J]	432	1	H[Q ∧ R ∧ C]	24	1	M	1	$\xi_{OISJP} + 12\eta_{QRCH} + q(M)$
					1	T # M	3	$\xi_{OISJP} + 12\eta_{QRCH} + q(TM)$
						Residual	20	$\xi_{OISJP} + 12\eta_{QRCH}$
			Residual	408				ξ_{OISJP}

ξ-components. Similarly, the treatment source Trellis is non-orthogonal to three idempotents in $\mathcal{P} \triangleright \mathcal{Q}$, and so information about Trellis differences is available in three different subspaces, as shown by the three occurrences of $q(T)$ in the table.

EXAMPLE 2.2 (Wheat with balanced lattice square in the laboratory phase). Example 9 of [11] extends Example 2.1 to include more lines so that a non-orthogonal design is needed in the second phase. Figure 5 gives the randomization diagram. In the field phase, 49 lines of wheat are investigated using a random-

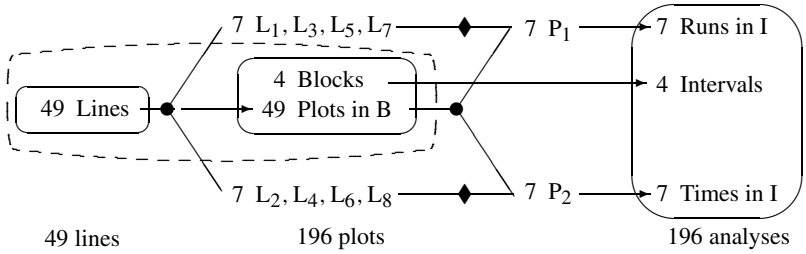

FIG. 5. *Randomized-inclusive randomizations in Example 2.2*: *lines are randomized to plots*, *then lines and plots are randomized to analyses*; B *denotes Blocks and* I *denotes Intervals*; $L_1, \ldots, L_8$ *are mutually orthogonal pseudofactors for Lines*; P_1 *and* P_2 *are pseudofactors for Plots*, *determined from different Lines pseudofactors in different blocks*.

ized complete-block design with four blocks. In the laboratory phase, a 7×7 balanced lattice square design with four replicates is used to assign the blocks, plots and lines to four intervals of seven runs by seven times. Pseudofactors are introduced for lines and plots to define the second-phase design. While expansion from $\mathcal{Q}$ to $\mathcal{Q}^*$ maintains orthogonality between the structures on plots and analyses, it induces non-orthogonality between the lines and plot structures, even though these were orthogonal in the first-phase design.

The variance matrix under the randomizations is

$$\begin{aligned}\mathbf{V} = {} & \xi_0\mathbf{P}_0 + \xi_{\mathrm{I}}\mathbf{P}_{\mathrm{I}} + \xi_{\mathrm{IR}}\mathbf{P}_{\mathrm{IR}} + \xi_{\mathrm{IT}}\mathbf{P}_{\mathrm{IT}} + \xi_{\mathrm{IRT}}\mathbf{P}_{\mathrm{IRT}} \\ & + \eta_0\mathbf{Q}_0 + \eta_{\mathrm{B}}\mathbf{Q}_{\mathrm{B}} + \eta_{\mathrm{BP}}\mathbf{Q}_{\mathrm{BP}}.\end{aligned}$$

Randomized-inclusive randomizations are used in this experiment, as the outcome of the randomization of lines to plots must be known before the plots can be randomized to analyses. The Plots pseudofactors P_1 and P_2 are used to ensure appropriate partial confounding of sources from the lines tier with sources in the analyses tier. These pseudofactors do not give idempotents in $\mathbf{V}$, because they do not contribute to the variance matrix; they are unknown before the randomization of lines to plots, and are not among the unrandomized factors, that give rise to covariance, in the randomization of plots to analyses. However, as in Example 2.1, $\mathbf{Q}_{\mathrm{BP}}$ is rewritten as the sum of three $\mathbf{Q}^*$-matrices each with coefficient η_{BP}, to make the design for the second phase structure balanced. This results in η_{BP} occurring with three different ξ-components in the skeleton anova in Table 4, which is an extended version of the decomposition table given for Example 5 in [12]. The expansion of $\mathcal{Q}$ to $\mathcal{Q}^*$ makes the structure on plots orthogonal to that on analyses, so there is no need to show efficiency factors for the plots tier.

6. Estimation in a two-tiered experiment. Estimation of treatment effects and variances is straightforward in a two-tiered experiment with structure balance.

TABLE 4
Skeleton analysis of variance for Example 2.2

analyses tier		**plots tier**[a]		**lines tier**			
source	**d.f.**	**source**	**d.f.**	**eff.**	**source**	**d.f.**	**E.M.S.**
Mean	1	Mean	1		Mean	1	$\xi_0 + \eta_0 + q_0$
Intervals	3	Blocks	3				$\xi_{\mathrm{I}} + \eta_{\mathrm{B}}$
Runs[I]	24	P_1[B]	24	$\frac{1}{4}$	$\mathrm{Lines}_{\mathrm{R}}$	24	$\xi_{\mathrm{IR}} + \eta_{\mathrm{BP}} + \frac{1}{4}q(\mathrm{L_R})$
Times[I]	24	P_2[B]	24	$\frac{1}{4}$	$\mathrm{Lines}_{\mathrm{T}}$	24	$\xi_{\mathrm{IT}} + \eta_{\mathrm{BP}} + \frac{1}{4}q(\mathrm{L_T})$
R # T[I]	144	$\mathrm{Plots[B]}_{\vdash}$	144	$\frac{3}{4}$	$\mathrm{Lines}_{\mathrm{R}}$	24	$\xi_{\mathrm{IRT}} + \eta_{\mathrm{BP}} + \frac{3}{4}q(\mathrm{L_R})$
				$\frac{3}{4}$	$\mathrm{Lines}_{\mathrm{T}}$	24	$\xi_{\mathrm{IRT}} + \eta_{\mathrm{BP}} + \frac{3}{4}q(\mathrm{L_T})$
					Residual	96	$\xi_{\mathrm{IRT}} + \eta_{\mathrm{BP}}$

[a] $\mathrm{Plots[B]}_{\vdash}$ is the part of Plots[B] orthogonal to P_1[B] and P_2[B].

6.1. *Estimating treatment effects and variances in one stratum.* For data satisfying the conditions in Section 4.1, the following are also shown in [2, 23, 30] for $\mathbf{Q}$ in $\mathcal{Q}$ and $\mathbf{R}$ in $\mathcal{R}$ with $\lambda_{\mathbf{QR}} \neq 0$.

(E.a) The best linear unbiased estimator of the treatment effects $\mathbf{R}\mathbf{X}_h\boldsymbol{\tau}$, using only the projected data $\mathbf{QY}$, is $\mathbf{R}(\mathbf{Q} \rhd \mathbf{R})\mathbf{Y}/\lambda_{\mathbf{QR}}$, which is equal to $\mathbf{RQY}/\lambda_{\mathbf{QR}}$.

(E.b) The variance matrix of the above estimator is $(\eta_{\mathbf{Q}}/\lambda_{\mathbf{QR}})\mathbf{R}$.

(E.c) From (A.d) in Section 4.1, an unbiased estimator of $\eta_{\mathbf{Q}}$ is given by the mean square for $\mathbf{Q} \vdash \mathcal{R}$, if $\mathbf{Q} \vdash \mathcal{R}$ is non-zero.

6.2. *Treatment structure orthogonal to variance structure.* If $\mathcal{R}$ is orthogonal in relation to $\mathcal{Q}$ then each $\mathbf{R}$ in $\mathcal{R}$ has some $\mathbf{Q}$ in $\mathcal{Q}$ such that $\lambda_{\mathbf{QR}} = 1$. Then all the information on $\mathbf{R}\mathbf{X}_h\boldsymbol{\tau}$ is in stratum $\mathrm{Im}(\mathbf{Q})$. Hence, result (E.a) in Section 6.1 gives $\mathbf{RQY}$ as the overall best linear unbiased estimator of $\mathbf{R}\mathbf{X}_h\boldsymbol{\tau}$. Result (E.b) shows that the variance matrix of this estimator is $\eta_{\mathbf{Q}}\mathbf{R}$, and result (E.c) that the mean square for $\mathbf{Q} \vdash \mathcal{R}$ is an unbiased estimator for $\eta_{\mathbf{Q}}$, if $\mathbf{Q} \vdash \mathcal{R}$ is non-zero.

6.3. *Estimating treatment effects from multiple strata when variances are known.* Suppose that $\mathcal{R}$ is not orthogonal in relation to $\mathcal{Q}$. As shown in [23, 31], if the coefficients $\eta_{\mathbf{Q}}$ are known then we can combine information on $\mathbf{R}\mathbf{X}_h\boldsymbol{\tau}$ from all strata for which $\lambda_{\mathbf{QR}} \neq 0$ to obtain its generalized least squares (GLS) estimator, which is the best linear unbiased estimator. In our notation, it is given by

$$\mathbf{R}\mathbf{X}_h\widehat{\boldsymbol{\tau}} = \theta_{\mathbf{R}}^{-1} \sum_{\mathbf{Q}\in\mathcal{Q}} \eta_{\mathbf{Q}}^{-1}\mathbf{RQY}, \tag{11}$$

where $\theta_{\mathbf{R}} = \sum_{\mathbf{Q}\in\mathcal{Q}} \lambda_{\mathbf{QR}}\eta_{\mathbf{Q}}^{-1}$. The variance matrix of this estimator is $\theta_{\mathbf{R}}^{-1}\mathbf{R}$.

6.4. *Estimating treatment effects and variances from multiple strata.* Usually the coefficients $\eta_{\mathbf{Q}}$ are unknown and must be estimated. One method uses the mean square for $\mathbf{Q} \vdash \mathcal{R}$ to estimate $\eta_{\mathbf{Q}}$. Nelder [31] argued that, especially for designs in which some strata have few Residual degrees of freedom, estimates should instead be obtained by equating the expected and observed values of the mean squares for what Houtman and Speed [23] called "actual residuals". These estimates are the same as those obtained with I-MINQUE and, even though normality is not assumed, they are the same as those obtained by REML [34], because as shown in [23], Section 4.5, [35] and [40], the same set of equations is solved for all. As will be discussed in Section 11.2, the assumptions on the variance parameters being estimated here are different from those for a variance-components model.

As noted in [23, 31], the estimation of the coefficients $\eta_{\mathbf{Q}}$ requires an iterative procedure, because their estimation needs the estimated value of $\boldsymbol{\tau}$ and vice versa. Given working estimates $\widehat{\eta}^*_{\mathbf{Q}}$ of $\eta_{\mathbf{Q}}$, a working estimate $\widehat{\boldsymbol{\tau}}^*$ of $\boldsymbol{\tau}$ can be obtained from equation (11): thus a revised estimate of each $\eta_{\mathbf{Q}}$ can be computed as

$$\frac{\mathbf{y}'(\mathbf{Q} \vdash \mathcal{R})\mathbf{y} + (\sum'_{\mathbf{R}\in\mathcal{R}}\{\mathbf{y}'(\mathbf{Q} \rhd \mathbf{R})\mathbf{y} - \lambda_{\mathbf{QR}}\widehat{\boldsymbol{\tau}}^{*\prime}\mathbf{X}_h'\mathbf{R}\mathbf{X}_h\widehat{\boldsymbol{\tau}}^*\})}{d'_{\mathbf{Q}}}, \tag{12}$$

where $\sum'_{\mathbf{R}\in\mathcal{R}}$ means summation over $\mathbf{R} \in \mathcal{R}$ for which $\lambda_{\mathbf{QR}} \neq 0$, and $d'_{\mathbf{Q}}$ are the *effective* degrees of freedom for this estimator, which are given by

$$d'_{\mathbf{Q}} = \operatorname{trace}(\mathbf{Q} \vdash \mathcal{R}) + \sum_{\mathbf{R}\in\mathcal{R}}{}' [1 - \theta_{\mathbf{R}}^{-1}(\widehat{\eta}^*_{\mathbf{Q}})^{-1}\lambda_{\mathbf{QR}}] \operatorname{trace}\mathbf{R}.$$

Since $\mathbf{Q} \vdash \mathcal{R}$ and $\mathbf{R}$ are both idempotent, their traces are equal to their ranks.

The numerator of expression (12) is the sum of two parts. The first is the Residual sum of squares in this stratum from the anova; the second is the difference between the sum of squares of the treatment estimates from just the data projected onto $\operatorname{Im}(\mathbf{Q})$ and the sum of squares of the combined estimates, summed over all $\mathbf{R}$ for which $\lambda_{\mathbf{QR}} \neq 0$. The former does not depend on $\eta_{\mathbf{Q}}$, but the latter does. The effective degrees of freedom make it clear that, even when $\mathbf{Q} \vdash \mathcal{R} = \mathbf{0}$, there can be information to estimate $\eta_{\mathbf{Q}}$.

If estimates of the canonical components are required, these can be obtained from the estimates of the spectral components.

7. Estimating treatment effects and variances in a single part of $\mathcal{P} \rhd \mathcal{Q}$. Suppose that $\lambda_{\mathbf{PQ}} \neq 0$, so that there is an idempotent $\mathbf{P} \rhd \mathbf{Q}$. Consider an idempotent $\mathbf{R}$ in $\mathcal{R}$ for which $\lambda_{\mathbf{P}\rhd\mathbf{Q},\mathbf{R}} \neq 0$. Theorem 5.1 of [12] shows that $\lambda_{\mathbf{P}\rhd\mathbf{Q},\mathbf{R}} = \lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}}$. Applying the results of Section 6.1 with $\mathbf{Y}$ and $\mathbf{Q}$ replaced by $\mathbf{PY}$ and $\mathbf{P} \rhd \mathbf{Q}$, respectively, and using equation (10) for $\operatorname{Cov}(\mathbf{PY})$, we find that the best linear unbiased estimator of the treatment effect $\mathbf{R}\mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}$, using only the projected data $(\mathbf{P} \rhd \mathbf{Q})\mathbf{Y}$, is $\mathbf{R}((\mathbf{P} \rhd \mathbf{Q}) \rhd \mathbf{R})\mathbf{Y}/\lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}}$, which is equal to $\mathbf{RQPY}/\lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}}$. Moreover, the variance matrix of this estimator is equal to

$\mathbf{R}(\xi_{\mathbf{P}} + r\lambda_{\mathbf{PQ}}\eta_{\mathbf{Q}})/\lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}}$. Result (A.i) in Section 4.2 shows that the mean square for $(\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R}$ is an unbiased estimator for $\xi_{\mathbf{P}} + r\lambda_{\mathbf{PQ}}\eta_{\mathbf{Q}}$, if $(\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R}$ is non-zero.

In Example 1.2, the effects for M and T # M are estimated in just the source P[O $\wedge$ I $\wedge$ S $\wedge$ J] $\rhd$ H[Q $\wedge$ R $\wedge$ C] and the Residual mean square for this source is an unbiased estimator of $\xi_{\mathrm{OISJP}} + 12\eta_{\mathrm{QRCH}}$.

8. Full estimation in a three-tiered experiment which is anova-applicable.

8.1. *Full or partial anova.* Call the triple $(\mathcal{P}, \mathcal{Q}, \mathcal{R})$ *anova-applicable* if it satisfies the following condition:

(13) for every $\mathbf{Q}$ in $\mathcal{Q}$, if $\mathbf{QI}_{\mathcal{R}} \neq \mathbf{0}$ then $\mathbf{Q} \in \mathcal{Q}_1$.

That is, if the source for an $\mathbf{R}$ in $\mathcal{R}$ is (partially) confounded with a source for some $\mathbf{Q}$, then the latter source must be confounded with the source corresponding to a single $\mathbf{P}$. Section 4.2 shows that when this condition is satisfied then the idempotents in $\mathcal{P} \rhd \mathcal{Q}$ whose subspaces are contained in eigenspaces of $\mathbf{V}$ include all those which have any part of $\mathcal{R}$ partially or totally confounded with them.

Condition (13) is satisfied when $\mathcal{Q}$ is orthogonal in relation to $\mathcal{P}$, so that $\mathcal{Q}_1 = \mathcal{Q}$. Then $\mathbf{V}$ is given by equation (9), possibly with OVS. Treatment effects and their variances can be estimated as in Section 6. Examples 1.1, 2.1 and 2.2 are like this. We call this *full anova*.

Under full anova, if no $(\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R} = \mathbf{0}$, we estimate linear combinations of spectral components from the anova, even if they are not needed for standard errors of treatment effects. Otherwise, formula (12) can be used, but with $\mathbf{P} \rhd \mathbf{Q}$ replacing $\mathbf{Q}$. In Example 1.1, all the information about each treatment source in $\mathcal{R} \setminus \mathbf{R}_0$ is contained in (J # P[O]) $\rhd$ V[B]. Also, the difference between the mean squares for ((J # P[O]) $\rhd$ (V[B])) $\vdash \mathcal{R}$ and (J # P[O]) $\vdash \mathcal{Q}$ estimates $16\eta_{\mathrm{BV}}$.

In general, put $\mathcal{P} * \mathcal{Q} = \mathcal{Q}_1 \cup \{\mathbf{P} \vdash \mathcal{Q} : \mathbf{P} \in \mathcal{P}\}$. Then the images of all the idempotents in $\mathcal{P} * \mathcal{Q}$ are contained in eigenspaces of $\mathbf{V}$. If $(\mathcal{P}, \mathcal{Q}, \mathcal{R})$ is anova-applicable but $\mathcal{Q}$ is not orthogonal in relation to $\mathcal{P}$ then we have *partial anova*, using only the information in $\mathcal{P} * \mathcal{Q}$. A treatment idempotent $\mathbf{R}$ in $\mathcal{R}$ may be non-orthogonal to more than one part of $\mathcal{P} \rhd \mathcal{Q}$, but these are all in $\mathcal{P} * \mathcal{Q}$. Section 4.2 shows that estimators of variances of treatment effects which are in different parts of $\mathcal{P} * \mathcal{Q}$ are uncorrelated, and so information can be combined as in Section 6.4.

However, the linear combinations of spectral components in the expected mean square for parts of $\mathcal{P} \rhd \mathcal{Q}$ outside $\mathcal{P} * \mathcal{Q}$ are not involved in this process, and their anova-estimators may not have good properties. A similar situation arises in two-tiered experiments if the group is not stratifiable but all treatment subspaces are contained in known eigenspaces of the variance matrix (see Example 16 in [3]).

In the special case that $\mathcal{R}$ is orthogonal in relation to $\mathcal{P} \rhd \mathcal{Q}$, each $\mathbf{R}$ in $\mathcal{R}$ has unique idempotents $\mathbf{Q}$ in $\mathcal{Q}$ and $\mathbf{P}$ in $\mathcal{P}$ such that $\lambda_{\mathbf{PQ}}\lambda_{\mathbf{QR}} = 1$, so that $\lambda_{\mathbf{PQ}} = \lambda_{\mathbf{QR}} = 1$. Hence, Condition (13) is satisfied. Then $\mathbf{P} \rhd \mathbf{Q} = \mathbf{Q}$, $\mathbf{RQP} = \mathbf{R}$ and the

effect $\mathbf{R}\mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}$ is estimated in just the one part $\mathbf{P} \rhd \mathbf{Q}$ of $\mathcal{P} * \mathcal{Q}$. The anova may be either full or partial. Tables 1 and 2 each specify full anova. The estimator for $\mathbf{R}\mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}$, and its variance matrix, obtained by simplifying the expressions given in Section 7, are $\mathbf{R}\mathbf{Y}$ and $(\xi_{\mathbf{P}} + r\eta_{\mathbf{Q}})\mathbf{R}$, respectively. If $\xi_{\mathbf{P}}$ and $\eta_{\mathbf{Q}}$ are not known, the mean square for $(\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R}$ provides an unbiased estimate of $\xi_{\mathbf{P}} + r\eta_{\mathbf{Q}}$, unless $(\mathbf{P} \rhd \mathbf{Q}) \vdash \mathcal{R} = \mathbf{0}$. Estimation of treatment effects and their standard errors proceeds exactly as in Section 6.2.

8.2. *Difficulties that do not arise in two-tiered experiments*. Even when $(\mathcal{P}, \mathcal{Q}, \mathcal{R})$ is anova-applicable, some phenomena can occur that are impossible in two-tiered experiments, even for the straightforward case where $\mathcal{P}$ and $\mathcal{Q}$ are both poset block structures, $\mathcal{Q}$ is orthogonal in relation to $\mathcal{P}$, and $\mathcal{R}$ is orthogonal in relation to $\mathcal{Q}$.

8.2.1. *Inestimability of some spectral and canonical components*. For a two-tiered experiment with OVS and $\mathcal{R}$ orthogonal in relation to $\mathcal{Q}$, the estimability of spectral components of variance is easily determined. If $\mathbf{Q} \vdash \mathcal{R}$ is non-zero then its mean square provides the best unbiased quadratic estimator of $\eta_{\mathbf{Q}}$; otherwise, there is no estimator for $\eta_{\mathbf{Q}}$. In particular, η_0 is never estimable.

In a three-tiered experiment, there are some sources for which the expected mean square is a linear combination of a ξ-parameter and an η-parameter. It may not be possible to estimate ξ and η separately. This affects the estimability of canonical components, although it appears that often more canonical, than spectral, components are estimable. The parameters ξ_0, η_0, ϕ_0 and ψ_0 are never estimable.

Otherwise, the simplest way in which two spectral components cannot be estimated separately occurs when a generalized factor F on Υ is randomly assigned to a generalized factor H on Ω with the same number of levels. Then only a linear combination of ξ_H and η_F can be estimated, and hence only a linear combination of ϕ_H and ψ_F. In Example 1.1, Blocks are assigned to Occasions, both of which have three levels. Then, only $\xi_{\mathrm{O}} + 16\eta_{\mathrm{B}}$ is estimable, as is shown in Table 1, where these two components only occur together. Correspondingly, only $\phi_{\mathrm{O}} + \psi_{\mathrm{B}}$ is estimable.

In the special case that $|\Upsilon| = |\Omega|$, Lemma 4.2 of [12] shows that $\mathcal{P} \rhd \mathcal{Q} = \mathcal{Q}$ and so there are no idempotents of the form $\mathbf{P} \vdash \mathcal{Q}$. Thus, every expected mean square contains one $\eta_{\mathbf{Q}}$ and one $\xi_{\mathbf{P}}$. If κ is any constant smaller in modulus than all the $\eta_{\mathbf{Q}}$ and all the $\xi_{\mathbf{P}}$, then $\kappa\mathbf{I}_\Omega$ can be added to $\sum \xi_{\mathbf{P}}\mathbf{P}$ and subtracted from $\sum \eta_{\mathbf{Q}}\mathbf{Q}$ without changing the variance matrix $\mathbf{V}$ in equation (7). Thus, none of the spectral components of variance can be estimated, although sums of the form $\xi_{c(\mathbf{Q})} + \eta_{\mathbf{Q}}$ can be. For estimates of standard errors, these sums are all that is needed, and so there is no problem. However, for comparing sources of variation, estimates of canonical components are required. Except for ϕ_Ω and ψ_Υ, each canonical component is a multiple of a difference between spectral components. This may well be estimable, even though the corresponding spectral component is not.

In Example 2.1, no spectral component is estimable. If relative magnitudes of sources of variation are to be investigated, then the canonical components may provide the information. Here, only ϕ_T is estimable; the sums $\phi_{RT} + \psi_{BP}$ and $\phi_R + \psi_B$ are estimable. Plans A and C for Example 4 of [11] are examples where none of the spectral components is estimable, yet most of the canonical components are.

8.2.2. *Negative estimates of spectral components.* As noted in Section 2.3, all spectral components of variance must be non-negative, and hence so must any linear combination with positive coefficients. However, if $\lambda_{\mathbf{PQ}} \neq 0$ and a mean square whose expectation is $\xi_{\mathbf{P}} + r\lambda_{\mathbf{PQ}}\eta_{\mathbf{Q}}$ is less than one whose expectation is $\xi_{\mathbf{P}}$, then the value of $\eta_{\mathbf{Q}}$ obtained by equating observed and expected means squares is negative; the anova-estimate of $\eta_{\mathbf{Q}}$ is then set to zero. This is the usual algorithm for constrained estimation, such as for non-negative variance components or simple linear regression with a non-negative slope.

In Example 1.1, Table 1 shows that the appropriate Residual source for all three treatment sources is the one with expected mean square equal to $\xi_{OJP} + 16\eta_{BV}$. If this mean square turns out to be smaller than the one whose expectation is ξ_{OJP} then we set η_{BV} to zero and combine the two Residual mean squares to obtain a better estimate of ξ_{OJP}. See Section 11.3 for further discussion.

8.2.3. *The effect of pseudosources.* If there are pseudosources for $\mathcal{Q}$ then some η-components occur with more than one ξ-component, even if $\mathcal{Q}$ is orthogonal in relation to $\mathcal{P}$. This can lead to what we call *linearly dependent commutative variance structure* (LDCVS), in which the eigenspaces of $\mathbf{V}$ are known but the eigenvalues satisfy some linear equations. This gives a set of sources whose expected mean squares are linearly dependent: simply equating them all to their data mean squares may give inconsistent results. Suppose that for i, j in $\{1, 2\}$ the idempotent $\mathbf{Q}^*_{ij}$ corresponds to a pseudosource for $\mathbf{Q}_j$ and is totally confounded with $\mathbf{P}_i$. Then the expected mean squares for the four idempotents $(\mathbf{P}_i \rhd \mathbf{Q}^*_{ij}) \vdash \mathcal{R}$ are $\xi_1 + r\eta_1$, $\xi_1 + r\eta_2$, $\xi_2 + r\eta_1$ and $\xi_2 + r\eta_2$. To estimate either the spectral or canonical components by equating expected and observed mean squares requires that the sum of the middle two observed mean squares is equal to the sum of the outer two.

Non-orthogonality between $\mathcal{Q}$ and $\mathcal{P}$ can produce a similar effect. In Table 3, the expected mean squares for $S[O \wedge I] \vdash \mathcal{Q}$, $(S[O \wedge I] \rhd C[Q]) \vdash \mathcal{R}$, $S \# J[O \wedge I] \vdash \mathcal{Q}$ and $(S \# J[O \wedge I] \rhd C[Q]) \vdash \mathcal{R}$ are linearly dependent.

9. Estimation in a three-tiered experiment which is not anova-applicable. If $\mathbf{V}$ is known then we can estimate treatment effects by GLS, which gives different results from ordinary least squares if $\mathcal{R}$ is not orthogonal in relation to $\mathcal{P} \rhd \mathcal{Q}$.

Equation (7) gives

$$\mathbf{V} = \mathbf{IVI} = \sum_{\mathbf{P}\in\mathcal{P}} \xi_{\mathbf{P}}\mathbf{P} + r \sum_{\mathbf{P}\in\mathcal{P}} \sum_{\mathbf{Q}\in\mathcal{Q}} \sum_{\mathbf{P}^*\in\mathcal{P}} \eta_{\mathbf{Q}}\mathbf{PQP}^*.$$

Put $\alpha_{\mathbf{Q}} = \sum_{\mathbf{P}}(\lambda_{\mathbf{PQ}}/\xi_{\mathbf{P}})$ for $\mathbf{Q}$ in $\mathcal{Q}$. Then direct calculation shows that

$$\mathbf{V}^{-1} = \sum_{\mathbf{P}\in\mathcal{P}} \frac{1}{\xi_{\mathbf{P}}}\mathbf{P} - \sum_{\mathbf{P}\in\mathcal{P}}\sum_{\mathbf{Q}\in\mathcal{Q}}\sum_{\mathbf{P}^*\in\mathcal{P}} \left(\frac{r\eta_{\mathbf{Q}}}{1+r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}}}\right)\frac{1}{\xi_{\mathbf{P}}\xi_{\mathbf{P}^*}}\mathbf{PQP}^*.$$

Consider $\mathbf{R}$ in $\mathcal{R}$. When the $\xi_{\mathbf{P}}$ and $\eta_{\mathbf{Q}}$ are known, the GLS estimator of the treatment effect $\mathbf{R}\mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}$ is $(\mathbf{RV}^{-1}\mathbf{R})^{-}\mathbf{RV}^{-1}\mathbf{Y}$, with variance matrix $(\mathbf{RV}^{-1}\mathbf{R})^{-}$. For a chain of randomizations, $\mathbf{R} = \mathbf{RI}_{\mathcal{Q}} = \sum_{\mathbf{Q}}\mathbf{RQ}$, so

$$\begin{aligned}
\mathbf{RV}^{-1} &= \sum_{\mathbf{P}}\sum_{\mathbf{Q}} \frac{1}{\xi_{\mathbf{P}}}\mathbf{RQP} - \sum_{\mathbf{P}}\sum_{\mathbf{Q}}\sum_{\mathbf{P}^*}\left(\frac{r\eta_{\mathbf{Q}}}{1+r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}}}\right)\frac{1}{\xi_{\mathbf{P}}\xi_{\mathbf{P}^*}}\mathbf{RQPQP}^* \\
&= \sum_{\mathbf{P}}\sum_{\mathbf{Q}} \frac{1}{\xi_{\mathbf{P}}}\mathbf{RQP} - \sum_{\mathbf{Q}}\sum_{\mathbf{P}^*}\left(\frac{r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}}}{1+r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}}}\right)\frac{1}{\xi_{\mathbf{P}^*}}\mathbf{RQP}^* \\
&= \sum_{\mathbf{P}}\sum_{\mathbf{Q}} \frac{1}{\xi_{\mathbf{P}}}\left(\frac{1}{1+r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}}}\right)\mathbf{RQP}.
\end{aligned}$$

Hence,

$$\mathbf{RV}^{-1}\mathbf{R} = \sum_{\mathbf{P}}\sum_{\mathbf{Q}} \frac{1}{\xi_{\mathbf{P}}}\left(\frac{1}{1+r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}}}\right)\mathbf{RQPQR} = \sum_{\mathbf{Q}}\left(\frac{\alpha_{\mathbf{Q}}\lambda_{\mathbf{QR}}}{1+r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}}}\right)\mathbf{R} = \theta_{\mathbf{R}}\mathbf{R},$$

with $\theta_{\mathbf{R}} = \sum_{\mathbf{Q}\in\mathcal{Q}}\alpha_{\mathbf{Q}}\lambda_{\mathbf{QR}}(1+r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}})^{-1}$. Thus, the GLS estimator of $\mathbf{R}\mathbf{X}_f\mathbf{X}_h\boldsymbol{\tau}$ is

$$\frac{1}{\theta_{\mathbf{R}}}\sum_{\mathbf{P}}\sum_{\mathbf{Q}} \frac{1}{\xi_{\mathbf{P}}}\left(\frac{1}{1+r\eta_{\mathbf{Q}}\alpha_{\mathbf{Q}}}\right)\mathbf{RQPY},$$

with variance matrix $\theta_{\mathbf{R}}^{-1}\mathbf{R}$.

This estimator is a linear combination of the $\mathbf{RQPY}$. There are no terms in $\mathbf{R}(\mathbf{P}\vdash\mathcal{Q})$, because all projectors of this form are zero for a chain of randomizations.

In the special case that $\mathcal{R}$ is orthogonal in relation to $\mathcal{Q}$ there is a unique $\mathbf{Q}$ such that $\mathbf{RQ} = \mathbf{R}$ while $\mathbf{RQ}^* = \mathbf{0}$ if $\mathbf{Q}^* \neq \mathbf{Q}$, so the estimator is a linear combination of the $\mathbf{RPY}$, as shown in [49]. The scalar $\theta_{\mathbf{R}}$ specializes to that given in [49].

In the anova-applicable case, we have $\theta_{\mathbf{R}} = \sum_{\mathbf{Q}\in\mathcal{Q}}\lambda_{\mathbf{QR}}(\xi_{c(\mathbf{Q})} + r\eta_{\mathbf{Q}})^{-1}$ and

$$\mathbf{R}\mathbf{X}_f\mathbf{X}_h\widehat{\boldsymbol{\tau}} = \theta_{\mathbf{R}}^{-1}\sum_{\mathbf{Q}\in\mathcal{Q}}(\xi_{c(\mathbf{Q})} + r\eta_{\mathbf{Q}})^{-1}\mathbf{RQY}.$$

If $\mathbf{V}$ is unknown, then spectral or canonical components need to be estimated by using a general estimation method, such as REML, and the treatment effects can be estimated using empirical GLS (EGLS). This can be done in Examples 1.2 and 2.2 when combined estimates of treatment effects are required and $\mathbf{V}$ is unknown.

10. Extension to more than two randomizations in a chain. We have seen that, even with structure balance, there can be difficulties with anova estimation for a three-tiered experiment. One solution can be to use software designed for fitting mixed models. This does not need to be restricted to designs with structure balance, or to experiments with three tiers, so we begin by generalizing Section 2.

As discussed in [11], Section 6, and [12], Section 7, more than two randomizations are possible. For example, a multiphase experiment can consist of p phases and involve p randomizations. Then there are p sets, Ω_i for $i = 1, \ldots, p$, and another set Γ for the first phase. There is a design function $h: \Omega_p \to \Gamma$; if the objects in Γ are treatments then $h(\omega)$ is the treatment assigned to unit ω in Ω_p. A stratifiable group G_p of permutations of Ω_p is used to randomize h. For $i = 1, \ldots, p-1$, there is a design function $f_i: \Omega_i \to \Omega_{i+1}$, so that $f_i(\omega)$ is the unit in Ω_{i+1} assigned to unit ω in Ω_i; there is also a stratifiable group G_i of permutations of Ω_i which is used to randomize f_i. It is assumed that f_i is equi-replicate, with replication r_{i+1}/r_i, where $r_1 = 1$, so that each element in Ω_i is assigned to r_i elements in Ω_1.

Let Y_ω be the response on unit ω in Ω_1. For ω in Ω_1, put $s_1(\omega) = \omega$, $s_{i+1}(\omega) = f_i(s_i(\omega))$ for $i = 1, \ldots, p-1$ and $t(\omega) = h(s_p(\omega))$. The randomization-based model in equation (6) can be generalized to

$$Y_\omega = \sum_{i=1}^{p} Z_{i,s_i(\omega)} + \tau_{t(\omega)},$$

where $Z_{i,s_i(\omega)}$ is the random effect, under randomization by G_i, for unit $s_i(\omega)$ in Ω_i.

For this model, $\mathrm{E}(\mathbf{Y}) = \mathbf{X}_s\mathbf{X}_h\boldsymbol{\tau}$, where $\mathbf{X}_s$ is the $\Omega_1 \times \Omega_p$ design matrix for s_p and $\mathbf{X}_h$ is the $\Omega_p \times \Gamma$ design matrix for h. Generalize $\mathbf{I}_{\mathcal{R}}$ to be the $\Omega_1 \times \Omega_1$ matrix of orthogonal projection onto $\mathrm{Im}(\mathbf{X}_s\mathbf{X}_h)$. Also, $\mathbf{V} = \sum_{i=1}^{p} \mathbf{V}_i$, where $\mathbf{V}_i = r_i \sum_{\mathbf{P}_{ij} \in \mathcal{P}_i} \xi_{ij}\mathbf{P}_{ij}$ and each $\mathbf{P}_{ij}$ is an idempotent of $\mathbf{V}_i$ with spectral component ξ_{ij}.

To this point, there is no need for structure balance, nor do any of the structures need to be defined by factors. However, if the randomization of f_i is based on a tier of factors $\mathcal{H}_i$ defining a poset block structure on Ω_i then $\mathbf{V}_i = \sum_{H \in \mathcal{H}_i} \phi_H \mathbf{S}_H$, where $\mathbf{S}_H$ is the $\Omega_1 \times \Omega_1$ relationship matrix for H considered as a factor on Ω_1.

If all of $f_1, \ldots, f_p$ and h are structure-balanced then the results of Sections 3, 4, 7 and 8 can be extended to more than two randomizations. In particular, generalize $\mathcal{Q}_1$ be the set of idempotents $\mathbf{Q}$ in $\mathcal{P}_p$ for which there is an idempotent $\mathbf{P}_{c_i(\mathbf{Q})}$ in $\mathcal{P}_i$ for $i = 1, \ldots, p$ such that $\mathbf{P}_{c_p(\mathbf{Q})} = \mathbf{Q}$ and $\mathbf{P}_{c_i(\mathbf{Q})}\mathbf{P}_{c_{i+1}(\mathbf{Q})} = \mathbf{P}_{c_{i+1}(\mathbf{Q})}$ for $i = 1, \ldots, p-1$. The condition for anova-applicability becomes

$$\text{for every } \mathbf{Q} \text{ in } \mathcal{P}_p, \text{ if } \mathbf{Q}\mathbf{I}_{\mathcal{R}} \neq \mathbf{0} \text{ then } \mathbf{Q} \in \mathcal{Q}_1.$$

11. Obtaining estimates from data for experiments with a chain of randomizations. How can standard software be used to obtain, from data, estimates

of treatment effects and their standard errors and/or estimates of canonical components, under randomization-based models? Assume that, for $i = 1, \ldots, p$, $\mathcal{P}_i$ is given by a poset block structure defined by a set $\mathcal{H}_i$ of generalized factors on Ω_i, which are then expressed as factors on Ω_1. Two possible procedures, based on mixed models, are anova and mixed-model fitting.

11.1. *Analysis of variance.* This is the method of choice for anova-applicable cases in which the structure $\mathcal{R}$ on Γ is also orthogonal in relation to $\mathcal{Q}_1$, and other cases in which it has been decided that each treatment effect is to be estimated from a single source, as might be done in Example 1.2. Other anova-applicable cases can be dealt with by anova followed by combination of information, as in Section 6.4.

Anova can also be used to estimate canonical components when there is CVS and $\mathcal{R}$ is orthogonal in relation to $\mathcal{Q}$. If there is LDCVS, then a generalized linear model (GLM) estimates the components. One fits a GLM to the observed mean squares involved in the estimation. The GLM has a gamma distribution, identity link, dispersion parameter equal to 2, weights equal to the degrees of freedom and an $\mathbf{X}$ matrix that contains, in each row, the coefficients of the canonical components for the expected mean square corresponding to the observed mean square.

The advantage of anova is that it is a non-iterative procedure in which all the quantities are well-defined. Further, non-negativity constraints are easily implemented as a manual procedure applied after the anova has been obtained, and the inestimability of some variance parameters is often inconsequential.

However, most anova software does not produce combined estimates of fixed effects, so that it is simpler to use mixed-model fitting. A further difficulty with anova for multitiered experiments is that software for it is not generally available, GenStat being the only package with specific facilities [17]. It may still be possible to produce the correct decomposition by omitting some sources. For example, the correct decomposition is obtained for Example 2.1 from an anova or a regression model with the sources Runs, Times, Lines and Residual. This is akin to fitting a mixed model of convenience [14], because it does not contain terms for all the potential sources of variation that have been identified for the experiment.

11.2. *Classes of mixed models for structures defined by factors.* Most mixed-model software uses a conditional model ([25], Appendix 1, and [40], Section 4.6):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\mathbf{U} + \mathbf{E},$$

with $\mathrm{E}(\mathbf{Y} \mid \mathbf{U}) = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\mathbf{U}$, $\mathrm{Cov}(\mathbf{U}) = \mathbf{G}$ and $\mathrm{Cov}(\mathbf{Y} \mid \mathbf{U}) = \mathrm{Cov}(\mathbf{E}) = \mathbf{R}$, where $\boldsymbol{\tau}$ is the vector of fixed-effects parameters, $\mathbf{X}$ is an indicator-variable matrix for fixed effects with one row for each observation and a column for each fixed effect, $\mathbf{Z}$ is an indicator-variable matrix with a row for each observation and a column for each random effect, $\mathbf{U}$ is the vector of random effects, $\mathbf{E}$ is the vector of random unit

effects, and $\mathbf{G}$ and $\mathbf{R}$ are symmetric matrix functions of the variance parameters. This usage of $\mathbf{R}$ and $\mathbf{Z}$ is unrelated to their usage elsewhere in the paper.

This conditional model can be re-expressed in the following marginal form:

$$\mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\tau} \quad \text{and} \quad \mathrm{Cov}(\mathbf{Y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}, \tag{14}$$

where $\mathbf{V}$ is p.s.d. We are concerned with models for the $\Omega_1 \times \Omega_1$ variance matrix $\mathbf{V}$ that are based on sets $\mathcal{H}_i$ of generalized factors on Ω_1. Put $\mathcal{H} = \bigcup_{i=1}^{p} \mathcal{H}_i$. Let $\mathcal{E}$ be the set of those generalized factors in $\mathcal{H}$ that uniquely index the units in Ω_1, and let $\mathcal{L} = \mathcal{H} \setminus \mathcal{E}$. Partitioning $\mathbf{Z}$, $\mathbf{U}$ and $\mathbf{G}$ conformably, according to the elements of $\mathcal{L}$, to form $\{\mathbf{Z}_L : L \in \mathcal{L}\}$, $\{\mathbf{U}_L : L \in \mathcal{L}\}$ and $\{\mathbf{G}_L : L \in \mathcal{L}\}$ and assuming that the different $\mathbf{U}_L$ are independent of each other and of $\mathbf{E}$, allows us to write $\mathbf{Z}\mathbf{G}\mathbf{Z}' = \sum_{L\in\mathcal{L}} \mathbf{Z}_L\mathbf{G}_L\mathbf{Z}'_L$ and $\mathbf{R} = \sum_{E\in\mathcal{E}} \mathbf{R}_E$.

All software allows the fitting of variance models based on variance components, for which $\mathbf{G}_L = \sigma_L^2\mathbf{I}_{m(L)}$, where $m(L)$ is the number of levels of L, and $\mathbf{R}_E = \sigma_E^2\mathbf{I}_{\Omega_1}$. The variance-components model for the matrix $\mathbf{V}$ in equation (14) is $\mathbf{V} = \sum_{L\in\mathcal{L}} \sigma_L^2\mathbf{S}_L + \sum_{E\in\mathcal{E}} \sigma_E^2\mathbf{I}_{\Omega_1}$, where $\mathbf{Z}_L\mathbf{Z}'_L = \mathbf{S}_L$. For such models, it is assumed that all variance components are non-negative, which implies that $\mathbf{V}$ is p.s.d.

Some software allows negative estimates of the variance components: this allows the fitting of a canonical-components model for the variance matrix, whose general form is obtained by replacing each σ^2-parameter with a ϕ-parameter:

$$\mathbf{V} = \sum_{L\in\mathcal{L}} \phi_L\mathbf{S}_L + \sum_{E\in\mathcal{E}} \phi_E\mathbf{I}_{\Omega_1}. \tag{15}$$

This differs from the variance-components model in that the ϕ_L, for L in $\mathcal{L}$, are not assumed to be non-negative, although $\mathbf{V}$ is required to be p.s.d.

Randomization-based models are inherently marginal linear mixed models. The expectation is as given in equation (14), with $\mathbf{X}$ replaced by $\mathbf{X}_s\mathbf{X}_h$. For the variance part of the model, equation (8) is generalized to

$$\mathbf{V} = \sum_{i=1}^{p} \mathbf{V}_i = \sum_{i=1}^{p} \sum_{H\in\mathcal{H}_i} \phi_H\mathbf{S}_H,$$

which is of the form given in equation (15).

The features of randomization-based models on poset block structures are:

(R.a) For $i = 1, \ldots, p$, all the factors initially defined on Ω_i are deemed random.

(R.b) For $i = 1, \ldots, p$, $\mathbf{V}_i$ is p.s.d., so that linear combinations of the canonical components corresponding to its spectral components must be non-negative. This implies that $\phi_{\Omega_i} \geq 0$ for $i = 1, \ldots, p$ but that other canonical components can be negative.

(R.c) The factors on Γ, the treatment factors, are usually regarded as fixed.

The set $\mathcal{V}$ of matrices assumed for $\mathbf{V}$ under any given model is a subset of the set $\mathcal{M}$ of $\Omega_1 \times \Omega_1$ real symmetric matrices. If $\mathcal{F}$ is any subset of $\mathcal{H}$, put $\mathcal{M}(\mathcal{F}) = \{\sum_{F\in\mathcal{F}} a_F \mathbf{S}_F : a_F \in \mathbb{R} \text{ for } F \text{ in } \mathcal{F}\}$. For the models described above, the sets of matrices are:

Unstructured: $\mathcal{V}_{\mathrm{US}} = \{\mathbf{M} \in \mathcal{M} : \mathbf{M} \text{ is p.s.d.}\}$;
Canonical-components: $\mathcal{V}_{\mathrm{CC}}(\mathcal{H}) = \mathcal{M}(\mathcal{H}) \cap \mathcal{V}_{\mathrm{US}}$;
Variance-components: $\mathcal{V}_{\mathrm{VC}}(\mathcal{H}) = \{\sum_{H\in\mathcal{H}} a_H \mathbf{S}_H : a_H \in \mathbb{R}_0^+ \text{ for } H \text{ in } \mathcal{H}\}$;
Randomization-based: $\mathcal{V}_{\mathrm{RB}}(\mathcal{H}_1, \ldots, \mathcal{H}_p) = \{\sum_{i=1}^{p} \mathbf{V}_i : \mathbf{V}_i \in \mathcal{V}_{\mathrm{CC}}(\mathcal{H}_i) \text{ for } i = 1, \ldots, p\}$.

Clearly, $\mathcal{V}_{\mathrm{VC}}(\mathcal{H}) \subset \mathcal{V}_{\mathrm{RB}}(\mathcal{H}_1, \ldots, \mathcal{H}_p) \subset \mathcal{V}_{\mathrm{CC}}(\mathcal{H}) \subset \mathcal{V}_{\mathrm{US}}$.

11.3. *Mixed-model fitting.* By mixed-model fitting, in the case where variances are unknown, we mean REML estimation of variance parameters followed by EGLS estimation of the fixed effects. It is preferred for estimation of effects in cases that are not anova-applicable, including all those without structure balance, and for estimation of canonical components when there is not OVS. It might also be deployed in anova-applicable cases because of software availability or because it is convenient to use a method that covers virtually all cases. Mixed-model fitting can also be used when $\mathbf{V}$ is known: the variance parameters are fixed at their known values. It cannot be used for those anova-applicable cases in which $\mathcal{R}$ is not orthogonal in relation to $\mathcal{Q}_1$ and separate analyses are required for different parts of $\mathcal{Q}_1$. Further advantages of mixed-model fitting are that pseudofactors are unnecessary and that combined estimates of treatment effects are available when $\mathcal{R}$ is not orthogonal to the other structures. A disadvantage of mixed-model fitting is that it is an iterative procedure that can have computational difficulties. Using anova estimates of canonical components as initial values helps to surmount these.

In obtaining the fitted values for a randomization-based model using data from an experiment, a problem is that mixed-model software usually fits only variance-components models and perhaps canonical-components models. The default for GenStat directives [46] is to fit canonical-components models, and it is an option in both ASReml-R [18], a commercial package for R [38], and in `PROC MIXED` in SAS [39]. The R packages `lme4` [7] and `nlme` [37] fit variance-components models only. Because $\mathcal{V}_{\mathrm{RB}}(\mathcal{H}_1, \ldots, \mathcal{H}_p) \nsubseteq \mathcal{V}_{\mathrm{VC}}(\mathcal{H})$, we recommend fitting canonical-components models. Even so, a number of difficulties arise: (i) all canonical components in the given model must be estimable, (ii) current software does not allow the separate specification of the factor sets $\mathcal{H}_i$ and so cannot impose the constraint that each $\mathbf{V}_i$ is p.s.d. and (iii) current software requires the fitted canonical components to be non-zero to avoid singularities in the matrices involved in the computations.

For some variance models, it is inherently impossible to estimate all variance parameters, as required in (i). This should be investigated when designing an

experiment so that any problems can be identified and rectified before the experiment is run. The skeleton anova is extremely useful for this, as well as in checking the properties of a design more generally. Software using eigenanalyses to obtain these tables for multitiered experiments, although without the expected mean squares, is available in GenStat and in the R package `dae` [10].

Inestimability of variance parameters can arise in two ways. First, a variance parameter is not estimable if it is completely confounded by one or more fixed effects. For example, if there is some H in $\mathcal{H}$ for which $\mathbf{S}_H\mathbf{I}_{\mathcal{R}} = \mathbf{S}_H$ then neither ϕ_H nor ξ_H is estimable. As noted in Section 8.2.1, components corresponding to the overall mean are never estimable and mixed-model-fitting software usually excludes them; if not, they must be dropped. However, if other canonical components are inestimable for this reason then this is usually a sign that the experiment suffers from some form of pseudo-replication; dropping such components results in incorrect estimates of standard errors and so is inadvisable. The avoidance of such design deficiencies is one reason to use a skeleton anova to check any proposed design.

Let $\tilde{\mathcal{H}}$ be the set of all factors in $\mathcal{H}$ which do not correspond to the overall mean. The other cause of inestimability is linear dependence among the matrices $\mathbf{S}_H$ for H in $\tilde{\mathcal{H}}$. Then canonical components need to be dropped so that those remaining correspond to a linearly independent set of $\mathbf{S}_H$ for H in $\mathcal{H}^*$, where $\mathcal{H}^* \subset \tilde{\mathcal{H}}$ and $\mathcal{V}_{\mathrm{CC}}(\mathcal{H}^*) = \mathcal{V}_{\mathrm{CC}}(\tilde{\mathcal{H}})$. The model based on $\mathcal{H}^*$ is a "model of convenience". There is a choice about which canonical components to drop in forming $\mathcal{H}^*$. When the design function $f_i : \Omega_i \to \Omega_{i+1}$ is structure balanced for $i = 1, \ldots, p$, then a skeleton anova can aid in detecting a cause of inestimability of the type outlined in Section 8.2.1 and so in determining which canonical components to drop.

Section 8.2.1 shows that in Example 1.1 the canonical components for Occasions and Blocks are inestimable. The term for one or other must be omitted from the mixed model. This should not be taken to imply that the analyser of the experiment is assuming that either does not contribute to the variability. Indeed, the estimated component should be viewed as estimating the sum of the two. Also, the spectral components for Occasions and Blocks are confounded and so it is not possible to check that each is non-negative. All the other spectral components except ξ_0 and η_0 are estimable and so their non-negativity can be verified.

For Example 2.1, the symbolic mixed model, derived using Step 1 in [14], is:

$$\text{Lines} \mid \text{Blocks} + \text{Blocks} \wedge \text{Plots} + \text{Runs} + \text{Times} + \text{Runs} \wedge \text{Times}.$$

As outlined in Section 8.2.1, none of the spectral components is estimable and so their non-negativity cannot be checked. Further, the only estimable canonical component is ϕ_{R}. One of Runs $\wedge$ Times and Blocks $\wedge$ Plots and one of Runs and Blocks needs to be omitted. Omitting Blocks $\wedge$ Plots and Blocks is, in effect, setting $\psi_{\mathrm{BP}} = \psi_{\mathrm{B}} = 0$, and hence $\eta_{\mathrm{BP}} = \eta_{\mathrm{B}} = 0$. Again, it is not assumed that this is the true value of the components. The constraint is imposed merely to obtain a solution, and the supposed estimate of ϕ_{RT} is actually an estimate of $\phi_{\mathrm{RT}} + \psi_{\mathrm{BP}}$.

For (ii), a check that the spectral components are non-negative is the only option to ensure that the constraints on them are met. The VSPECTRALCHECK procedure in GenStat [46] does this. Equation (5) is used to obtain the estimated spectral components from the estimated canonical components. If any spectral component is negative, then the linear combination of canonical components on the right-hand side of equation (5) must be constrained to zero in a refit of the model. If there are several negative components, they are constrained one at a time. Begin with the components based on the smallest number of factors and, if there are several of these, constrain the one whose estimate is furthest from zero. This is repeated until there are no negative spectral components remaining. If several spectral components are set to zero, this may force some canonical components to be zero also. Checking the data for Example 1.2, available from Statlib data sets (http://lib.stat.cmu.edu/datasets/sensory), reveals three negative spectral components, and sequentially constraining them results in just two zero estimates.

Difficulty (iii) occurs because the estimate of some canonical component happens to be zero. It must be addressed by removing this canonical component from the model. It cannot be anticipated ahead of having the data.

12. Statistical inference. In order to perform hypothesis tests or compute confidence intervals, one has to assume that the response follows a multivariate normal distribution whose expectation and variance are those described in Section 10 for the randomization-based model. Some justification for this approach is that, over all possible randomizations, the distribution of the data has this expectation and variance. The only further assumption that is required for inference is that of multivariate normality, although the guarantee for the associated expectation and covariance strictly applies only over multiple re-runs of the experiment. The role for randomization in an analysis based on this model is to ensure that the sources of variation taken into account by the designer have terms in the model; that is, it links the model to the design. Irrespective of the number of tiers, the randomization does not itself produce distributions whose third and higher-order moments are those of a multivariate normal distribution.

13. Other models. Steps 2 and 3 of the method in [14] suggest changes that could be made to the expectation and variance of randomization-based models. Here we concentrate on changing treatment factors from fixed to random and changing unrandomized factors from random to fixed. The first of these produces a randomization-based model, but the second does not; the latter does not preserve the variance matrix under randomization as part of the model.

13.1. *Treatment factors regarded as random.* The simplest modification to the model in equation (1) is to assume that the τ_i, for i in Γ, are random variables with common mean μ and variance matrix $\mathbf{C}_\Gamma$, which may be as simple as $\sigma^2_\Gamma \mathbf{I}_\Gamma$ or

may be based on a poset block structure on Γ. So long as h is equi-replicate, $\mathbf{C}_\Gamma$ translates easily to add an extra variance matrix to $\mathbf{V}$.

Varieties in early generation variety trials are often regarded as random; see [41]. If Lines are designated random in Example 2.2, then the variance matrix becomes

$$\begin{aligned}\mathbf{V} = {} & \xi_0\mathbf{P}_0 + \xi_{\mathrm{I}}\mathbf{P}_{\mathrm{I}} + \xi_{\mathrm{IR}}\mathbf{P}_{\mathrm{IR}} + \xi_{\mathrm{IT}}\mathbf{P}_{\mathrm{IT}} + \xi_{\mathrm{IRT}}\mathbf{P}_{\mathrm{IRT}} \\ & + \eta_0\mathbf{Q}_0 + \eta_{\mathrm{B}}\mathbf{Q}_{\mathrm{B}} + \eta_{\mathrm{BP}}\mathbf{Q}_{\mathrm{BP}} \\ & + 4\sigma^2_{\mathrm{L}}\mathbf{I}_{\mathcal{R}}.\end{aligned}$$

13.2. *Unrandomized factors regarded as fixed.* Sometimes it is appropriate to classify unrandomized factors such as Sites, Centres, Sex or Judges as fixed. It requires that there is no confounding between fixed sources. It results in the exclusion of the corresponding subspaces from the REML estimation of variance parameters, with canonical components effectively being set to zero and effects added to the expectation, so the variance matrix may have LDCVS. In the expected mean squares, $q(H)$ replaces $r_i k_H \phi_H$ if generalized factor H on Ω_i is designated as fixed.

Suppose that Judges in Example 1.2 is to be considered fixed. This removes ϕ_{J} from the expression for the variance matrix, and $\mathrm{Im}\,(\mathbf{P}_0 + \mathbf{P}_{\mathrm{J}})$ is excluded from the REML estimation of the canonical components. The effect on the expected mean squares in Table 3 is to replace ξ_{J} by $\xi_{\mathrm{OJ}} + q(\mathrm{J})$.

14. Discussion. This paper extends randomization-based models to multi-tiered experiments with two or more randomizations in a chain, and discusses the estimation of treatment effects and their standard errors, and canonical components, under the assumption of such a model. There are novel aspects to the estimability of spectral and canonical components in such experiments, including that the variance matrix can exhibit LDCVS.

We have emphasised the usefulness of a skeleton anova in checking the properties of a design and of anova in analysing anova-applicable experiments and for supplying initial estimates for mixed-model fitting. A limitation is software availability.

Otherwise, mixed-model fitting software is used to fit a randomization-based model. In this, one has to ensure that estimates of "variance components" can be negative and be vigilant that estimates of spectral components are non-negative.

While potentially negative canonical components are mandated for randomization-based models, they have the additional benefit of allowing for negative correlation, which is realistic in some circumstances; see [28]. Littell et al. [25], Section 4.7, recommend that unconstrained estimates be allowed in order to control Type I error, and show that they can achieve greater power; this agrees with the conclusions of Wolde-Tsadik and Afifi [48]. However, caution is required in

ascribing a negative estimate for a component to negative population correlation. As Searle, Casella and McCulloch [40], Section 3.5, show, for a variance component just above zero, there can be a high probability of a negative estimate if the number of treatments is less than 5 and the number of replicates less than 25. Gilmour and Goos [21] demonstrate that simply allowing negative variance components is not a panacea, especially in small experiments.

REFERENCES

[1] ALEJANDRO, P. P., BAILEY, R. A. and CAMERON, P. J. (2003). Association schemes and permutation groups. *Discrete Math*. **266** 47–67. MR1991706

[2] BAILEY, R. A. (1981). A unified approach to design of experiments. *J. Roy. Statist. Soc. Ser. A* **144** 214–223. MR0625801

[3] BAILEY, R. A. (1991). Strata for randomized experiments (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol*. **53** 27–78. MR1094275

[4] BAILEY, R. A. (2008). *Design of Comparative Experiments*. Cambridge Univ. Press, Cambridge. MR2422352

[5] BAILEY, R. A., PRAEGER, C. E., ROWLEY, C. A. and SPEED, T. P. (1983). Generalized wreath products of permutation groups. *Proc. Lond. Math. Soc.* (3) **47** 69–82. MR0698928

[6] BARDIN, A. and AZAÏS, J.-M. (1990). Une hypothèse minimale pour une théorie des plans d'expériences randomisés. *Rev. Stat. Appl*. **38** 21–41. MR1080504

[7] BATES, D., MAECHLER, M. and BOLKER, B. (2016). lme4: Linear mixed-effects models using S4 classes. Available at http://cran.at.r-project.org/web/packages/lme4/index.html (R package version 1.1-10, accessed January 8, 2016).

[8] BRIEN, C. J. (1983). Analysis of variance tables based on experimental structure. *Biometrics* **39** 51–59.

[9] BRIEN, C. J. (1992). Factorial linear model analysis. Ph.D. thesis, Department of Plant Science, The University of Adelaide, Adelaide, South Australia. Available at http://hdl.handle.net/2440/37701 (accessed January 8, 2016).

[10] BRIEN, C. J. (2015). dae: functions useful in the design and ANOVA of experiments. Available at http://cran.at.r-project.org/web/packages/dae/index.html (R package version 2.7-0, accessed April 26, 2015).

[11] BRIEN, C. J. and BAILEY, R. A. (2006). Multiple randomizations (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol*. **68** 571–609. MR2301010

[12] BRIEN, C. J. and BAILEY, R. A. (2009). Decomposition tables for experiments. I. A chain of randomizations. *Ann. Statist*. **37** 4184–4213. MR2572457

[13] BRIEN, C. J. and BAILEY, R. A. (2010). Decomposition tables for experiments. II. Two–one randomizations. *Ann. Statist*. **38** 3164–3190. MR2722467

[14] BRIEN, C. J. and DEMÉTRIO, C. G. B. (2009). Formulating mixed models for experiments, including longitudinal experiments. *J. Agric. Biol. Environ. Stat*. **14** 253–280. MR2750840

[15] BRIEN, C. J., HARCH, B. D., CORRELL, R. L. and BAILEY, R. A. (2011). Multiphase experiments with at least one later laboratory phase. I. Orthogonal designs. *J. Agric. Biol. Environ. Stat*. **16** 422–450. MR2843135

[16] BRIEN, C. J. and PAYNE, R. W. (1999). Tiers, structure formulae and the analysis of complicated experiments. *The Statistician* **48** 41–52.

[17] BRIEN, C. J. and PAYNE, R. W. (2012). AMTIER Procedure. In *GenStat Reference Manual Release* 15, *Part* 3 *Procedures* 106–108. VSN International, Hemel Hempstead,

UK. Available at http://www.vsni.co.uk/resources/documentation/genstat-reference-procedure-library/ (accessed January 8, 2016).

[18] BUTLER, D., CULLIS, B. R., GILMOUR, A. R. and GOGEL, B. J. (2009). *Mixed Models for S Language Environments*: *ASReml-R Reference Manual*. *Version* 3. DPI Publications, Brisbane. Available at http://www.vsni.co.uk/resources/documentation/ (accessed January 8, 2016).

[19] COX, D. R. (1958). The interpretation of the effects of non-additivity in the Latin square. *Biometrika* **45** 69–73.

[20] CURNOW, R. N. (1959). The analysis of a two phase experiment. *Biometrics* **15** 60–73.

[21] GILMOUR, S. G. and GOOS, P. (2009). Analysis of data from non-orthogonal multistratum designs in industrial experiments. *J. R. Stat. Soc. Ser. C Appl. Stat.* **58** 467–484. MR2750088

[22] GRUNDY, P. M. and HEALY, M. J. R. (1950). Restricted randomization and quasi-Latin squares. *J. R. Stat. Soc. Ser. B* **12** 286–291.

[23] HOUTMAN, A. M. and SPEED, T. P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist.* **11** 1069–1085. MR0720254

[24] JAMES, A. T. (1957). The relationship algebra of an experimental design. *Ann. Math. Stat.* **28** 993–1002. MR0097142

[25] LITTELL, R. C., MILLIKEN, G. A., STROUP, W. W., WOLFINGER, R. D. and SCHABENBERGER, O. (2006). *SAS for Mixed Models*, 2nd ed. SAS Press, Cary, NC.

[26] MCINTYRE, G. A. (1955). Design and analysis of two phase experiments. *Biometrics* **11** 324–334.

[27] MONOD, H. and BAILEY, R. A. (1992). Pseudofactors: Normal use to improve design and facilitate analysis. *J. R. Stat. Soc. Ser. C Appl. Stat.* **41** 317–336.

[28] NELDER, J. A. (1954). The interpretation of negative components of variance. *Biometrika* **41** 544–548. MR0065083

[29] NELDER, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc. Ser. A* **283** 147–162. MR0176576

[30] NELDER, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. *Proc. Roy. Soc. Ser. A* **283** 163–178. MR0174156

[31] NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **30** 303–311. MR0234582

[32] NELDER, J. A. (1977). A reformulation of linear models. *J. Roy. Statist. Soc. Ser. A* **140** 48–76. MR0458743

[33] OJIMA, Y. (1998). General formulae for expectations, variances and covariances of the mean squares for staggered nested designs. *J. Appl. Stat.* **25** 785–799. MR1652264

[34] PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58** 545–554. MR0319325

[35] PATTERSON, H. D. and THOMPSON, R. (1975). Maximum likelihood estimation of components of variance. In *Proceedings of the Eighth International Biometric Conference* (L. C. A. Corsten and T. Postelnicu, eds.) 197–207. Editura Academiei Republicii Socialiste România, Bucureşti.

[36] PAYNE, R. W. and WILKINSON, G. N. (1977). A general algorithm for analysis of variance. *J. R. Stat. Soc. Ser. C Appl. Stat.* **26** 251–260.

[37] PINHEIRO, J. C. and BATES, D. M. (2000). *Mixed Effects Models in S and S-PLUS*. Springer, New York.

[38] R DEVELOPMENT CORE TEAM (2016). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.r-project.org/ (accessed January 8, 2016).

[39] SAS INSTITUTE INC. (2010). *SAS/STAT® 9.22 User's Guide*. SAS Institute Inc., Cary, NC.
[40] SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York. MR1190470
[41] SMITH, A., CULLIS, B. and GILMOUR, A. (2001). The analysis of crop variety evaluation data in Australia. *Aust. N. Z. J. Stat.* **43** 129–145. MR1855705
[42] SPEED, T. P. (1987). What is an analysis of variance? (with discussion). *Ann. Statist.* **15** 885–941. MR0902237
[43] SPEED, T. P. and BAILEY, R. A. (1982). On a class of association schemes derived from lattices of equivalence relations. In *Algebraic Structures and Applications* (*University of Western Australia*, 1980) (P. Schultz, C. E. Praeger and R. P. Sullivan, eds.). *Lecture Notes in Pure and Applied Mathematics* **74** 55–74. Dekker, New York. MR0647166
[44] SPEED, T. P. and BAILEY, R. A. (1987). Factorial dispersion models. *Int. Stat. Rev.* **55** 261–277. MR0963143
[45] TJUR, T. (1984). Analysis of variance models in orthogonal designs (with discusion). *Int. Stat. Rev.* **52** 33–81. MR0967202
[46] VSN INTERNATIONAL (2015). *GenStat for Windows*, 18th ed. VSN International, Hemel Hempstead, UK. Available at GenStat.co.uk.
[47] WILKINSON, G. N. (1970). A general recursive procedure for analysis of variance. *Biometrika* **57** 19–46.
[48] WOLDE-TSADIK, G. and AFIFI, A. A. (1980). A comparison of the "sometimes pool", "sometimes switch" and "never pool" procedures in the two-way ANOVA random effects model. *Technometrics* **22** 367–373. MR0585634
[49] WOOD, J. T., WILLIAMS, E. R. and SPEED, T. P. (1988). Non-orthogonal block structure in two-phase designs. *Aust. J. Stat.* **30A** 225–237.
[50] YATES, F. (1936). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.* **26** 424–455. Reprinted with additional author's note in Yates, F. (1970) *Experimental Design: Selected Papers*, pages 147–180. Griffin, London.

SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF ST ANDREWS
ST ANDREWS, FIFE KY16 9SS
UNITED KINGDOM
AND
SCHOOL OF MATHEMATICAL SCIENCES
QUEEN MARY UNIVERSITY OF LONDON
MILE END ROAD
LONDON E1 4NS
UNITED KINGDOM
E-MAIL: rab24@st-andrews.ac.uk

PHENOMICS AND BIOINFORMATICS RESEARCH CENTRE
SCHOOL OF INFORMATION TECHNOLOGY
AND MATHEMATICAL SCIENCES
UNIVERSITY OF SOUTH AUSTRALIA
GPO BOX 2471
ADELAIDE, SA 5001
AUSTRALIA
AND
THE AUSTRALIAN CENTRE
FOR PLANT FUNCTIONAL GENOMICS
WAITE CAMPUS
UNIVERSITY OF ADELAIDE
URRBRAE, SA 5064
AUSTRALIA
E-MAIL: chris.brien@unisa.edu.au
URL: http://chris.brien.name