# General Multi-Level Linear Modelling for Group Analysis in FMRI

FMRIB Technical Report TR01CB1

A related paper has been accepted for publication in NeuroImage

**Christian F. Beckmann***, **Mark Jenkinson*** **and Stephen M. Smith**

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB),
Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital,
Headley Way, Headington, Oxford, UK

Corresponding author: `beckmann@fmrib.ox.ac.uk`
*These authors contributed equally to the work.

**Abstract**

This paper discusses general modelling of multi-subject and/or multi-session FMRI data. In particular, we show that a two-level mixed-effects model (where parameters of interest at the group level are estimated from parameter and variance estimates from the single-session level) can be made equivalent to a single complete mixed-effects model (where parameters of interest at the group level are estimated directly from all of the original single-sessions' time-series data) if the (co-)variance at the second level is set equal to the sum of the (co-)variances in the single-level form, using the BLUE with known covariances. This result has significant implications for group studies in FMRI, since it shows that the group analysis only requires values of the parameter estimates and their (co-)variance from the first level, generalising the well established 'summary statistics' approach in FMRI. The simple and generalised framework allows for different pre-whitening and different first-level regressors to be used for each subject. The framework incorporates multiple levels and cases such as repeated measures, paired or unpaired $t$-tests and $F$-tests at the group level; explicit examples of such models are given in the paper. Using numerical simulations based on typical first level covariance structures from real FMRI data we demonstrate that by taking into account lower-level covariances and heterogeneity a substantial increase in higher-level $Z$-score is possible.

## 1 Introduction

Functional magnetic resonance imaging studies are typically used to address questions about activation effects in populations of subjects. This generally involves a multi-subject and/or multi-session approach where data are analysed in such a way as to allow for hypothesis tests at the group level [9, 19], e.g. in order to assess whether the observed effects are common and stable across or between groups of interest.

Calculating the level and probability of brain activation for a single subject is typically achieved using a linear model of the signal together with a Gaussian noise model for the residuals. This model is commonly referred to as the General Linear Model (GLM) and much attention to date has been focussed on ways of modelling and fitting the (time-series) signal and residual noise at the individual single-session level [2, 17, 18].

In order to carry out higher-level anlayses it is straightforward to formulate a complete single-level GLM that relates various parameters of interest at the group-level to the full set of (time-series) data available [3, 13]. In FMRI, where the human and computational costs involved in data analysis are relatively high, however, it is desirable to be able to make group-level inferences using the *results* of separate first-level analyses without the need to re-analyse any of the individual subject data. Within such a two-level approach, group parameters of interest can then easily be refined as more data become available. The natural question to address is if and how a two-level approach to group analysis restricts the type of hypotheses we are able to infer at the group level.

In order to be able to generate results that extend to the population, we also need to account for the fact that the individual subjects themselves are sampled from the population and thus are random quantities with associated variances. It is exactly this step that marks the transition from a simple *fixed-effects model* to a *mixed-effects model*[1] and it is imperative to formulate a model at the group level that allows for the explicit modelling of these additional variance terms [9, 5].

In this paper we revisit the single-level mixed-effect GLM for FMRI group analysis which relates the effect of interest at the group level (e.g. difference of patients and controls) to the individual FMRI time-series. This model is well known and encompasses both the familiar single-subject models as well as a general group model [9, 3, 4, 13]. By using the full GLM at the group level, one obtains a very general and flexible framework which allows us to examine more complex relationships beyond the simple tests such as mean group activation. An example of a more general group analysis would be to test whether activation is correlated with some other variable such as drug dosage or disability score.

The main result shows that this single-level GLM can be decomposed into an equivalent two-level version so that group analyses can be performed using only the lower-level parameter *estimates* and their (co-)variances, from the individual subject analyses. This result is formally stated in terms of a model equivalence theorem. The practical consequence is that it is possible to perform valid group analyses in two stages: firstly the individual subject analyses, and secondly, a single group analysis performed on the output of the combined estimates from the individual subject analyses. Furthermore, there are very few restrictions placed upon the model and so it can be used in very general conditions, such as when the estimates at the individual subject level were obtained using different pre-whitening for each subject or indeed different regressors. We finally give examples of how, in this general framework, group tests of interest, like paired or unpaired $t$-tests, can easily be formulated. The model equivalence result applies not only to FMRI studies but to any mixed-effects GLM that can be split into two levels in the same way as presented here.

All models used here are instances of the univariate GLM. This means treating each single, registered voxel separately and assuming Gaussian noise. Parameter estimation is being performed (at all levels) using pre-whitening approach [2, 17] which is known to be the best linear unbiased estimator (BLUE) for these models [16].

In this paper we make a distinction between the problems involved in modelling of FMRI data at the group level and estimation of the relevant (co-)variances. While estimation of variances is an important issue when implementing such models, we consider the problems involved as beyond the scope of this paper. A companion paper on practical issues of estimation, using Bayesian methods, is currently being prepared and these techniques form the basis of what has been implemented in the latest software release of FSL [7].

# 2 Models

To begin with we consider the familiar two-level univariate GLM for FMRI. That is, the model that in the first level deals with individual subjects, relating time-series to activation, and in the second level deals with a group of subjects or sessions (or both), relating the combined individual activation estimates to some group parameter, such as mean activation level.

## 2.1 Two-level GLM

Consider an experiment where there are $N$ subjects and that for each subject, $k$, the preprocessed FMRI data is $Y_k$ (a vector of $T$ time points), the design matrix is $X_k$ and the parameter estimates are $\beta_k$ (for $k = 1, \ldots, N$). The two-level model for this experiment is

$$Y_k = X_k\beta_k + \epsilon_k \quad (1)$$
$$\beta = X_{\scriptscriptstyle G}\beta_{\scriptscriptstyle G} + \eta \quad (2)$$

where $\epsilon_k$ specifies the single-subject residuals, $\eta$ specifies the residuals of the group activation (parameter)

[1] Note that in the FMRI literature this has previously been referred to as a *random-effects model*. Within this paper, however, the separate fixed- and random-effects contributions to the mixed-effects variance are considered, thus making a clear distinction between "random-" and "mixed-effects" important.

scores, and where

$$\mathrm{E}(\epsilon_k) = 0\ , \qquad \mathrm{Cov}(\epsilon_k) = V_k\ , \qquad \mathrm{E}(\eta) = 0\ , \qquad \mathrm{Cov}(\eta) = V_G, \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix}$$

denotes the combined first-level parameter estimates of the whole group, assembled into a single vector, $X_G$ is the group-level design matrix (e.g. separating controls from normals) and $\beta_G$ is the final vector of group-level parameters.

## 2.2 Single-level GLM

The two-level model written above can be re-written as a single-level model by substituting equation 2 into a block form of equation 1, giving

$$Y = XX_G\beta_G + X\eta + \epsilon \tag{3}$$

where now all the first-level GLMs have been combined such that

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & X_N \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix},$$

$$\mathrm{E}(\eta) = 0\ , \qquad \mathrm{E}(\epsilon) = 0\ , \qquad \mathrm{Cov}(\eta) = V_G\ , \qquad \text{and} \qquad \mathrm{Cov}(\epsilon) = V = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & V_N \end{bmatrix}.$$

The two error terms in equation 3 can simply be combined such that

$$Y = XX_G\beta_G + \gamma, \tag{4}$$

where

$$\mathrm{E}(\gamma) = 0 \quad \text{and} \quad \mathrm{Cov}(\gamma) = W = XV_GX^T + V.$$

It is easy to see that this *model* is equivalent to the two-level version presented as equations 1 and 2.

## 2.3 Estimation

The BLUEs for both the two-level GLM and the single-level GLM can be calculated using the General Least Squares approach [16].

Initially consider the two-level GLM. The parameter estimates at the first level are

$$\begin{aligned} \widehat{\beta} &= (X^TV^{-1}X)^{-1}X^TV^{-1}Y, \\ \mathrm{Cov}(\widehat{\beta}) &= (X^TV^{-1}X)^{-1}. \end{aligned} \tag{5}$$

Similarly, the estimates of the (second-level) group parameters are given by

$$\begin{aligned} \widehat{\beta}_G &= (X_G^TV_G^{-1}X_G)^{-1}X_G^TV_G^{-1}\beta, \\ \mathrm{Cov}(\widehat{\beta}_G) &= (X_G^TV_G^{-1}X_G)^{-1}. \end{aligned}$$

In practice, however, the second-level model uses the *estimates* from the first level as input and not the true (but unobservable) parameters. That is, equation 2 is modified, becoming

$$\widehat{\beta} = X_G\beta_G + \eta'. \tag{6}$$

Therefore the two-level model, as used in practice, is specified by equations 1 and 6. This has significant implications, as the two-level version is no longer precisely equivalent to the single-level model in terms of estimation. In particular, the estimation of the group parameters in the two-level model now is

$$\begin{aligned} \widehat{\beta}_G &= (X_G^T V_{G2}^{-1} X_G)^{-1} X_G^T V_{G2}^{-1} \widehat{\beta}, \\ \mathrm{Cov}(\widehat{\beta}_G) &= (X_G^T V_{G2}^{-1} X_G)^{-1}, \end{aligned} \tag{7}$$

where $V_{G2} = \mathrm{Cov}(\eta')$ represents the potentially different covariance in this new two-level model.
Now consider the single-level GLM (equation 4), where the BLUE is

$$\begin{aligned} \widehat{\beta}_G &= (X_G^T X^T W^{-1} X X_G)^{-1} X_G^T X^T W^{-1} Y, \\ \mathrm{Cov}(\widehat{\beta}_G) &= (X_G^T X^T W^{-1} X X_G)^{-1}. \end{aligned} \tag{8}$$

This equation directly relates the group parameter estimates of interest, $\widehat{\beta}_G$, to the full data vector $Y$ and so requires the GLM to be solved for matrices of greatly increased size.
Thus, instead of solving the single-level model all at once, we wish to use the two-level approach. However, substituting equation 5 into equation 7 gives the two-level group parameter estimates as:

$$\begin{aligned} \widehat{\beta}_G &= (X_G^T V_{G2}^{-1} X_G)^{-1} X_G^T V_{G2}^{-1} (X^T V^{-1} X)^{-1} X^T V^{-1} Y, \\ \mathrm{Cov}(\widehat{\beta}_G) &= (X_G^T V_{G2}^{-1} X_G)^{-1}. \end{aligned} \tag{9}$$

If this estimation (equation 9) can be made equivalent to the single-level estimation (equation 8) by accounting for the covariances of the first-level estimates within the second-level (i.e. setting $V_{G2}$ appropriately), then the two approaches become exactly equivalent. This turns out to be possible and the general equivalence result is presented in the next section.

## 3 Model Equivalence

**Theorem:**
The two-level model specified by equations 1 and 6 is fully equivalent to the single-level model specified by equation 4 in terms of both modelling *and estimation* when

$$V_{G2} = V_G + (X^T V^{-1} X)^{-1}. \tag{10}$$

That is, the second-level covariance is set as the sum of the group-covariance from the single-level model and the first-level parameter covariance from the two-level model.
**Proof:**
We employ the *Sherman-Morrison-Woodbury* formula (equation 17 from the appendix)
to write $W^{-1}$ as

$$W^{-1} = V^{-1} - V^{-1} X \left( V_G^{-1} + X^T V^{-1} X \right)^{-1} X^T V^{-1}, \tag{11}$$

and let

$$Q = (X^T V^{-1} X)^{-1},$$

which is the covariance estimate for the $\widehat{\beta}$s, from equation 5.
Then

$$\begin{aligned} X^T W^{-1} X &= \left( \mathbf{I} - \left( V_G^{-1} Q + \mathbf{I} \right)^{-1} \right) Q^{-1} \\ &= \left( V_G + Q \right)^{-1}, \end{aligned}$$

by using equation 17 again, and

$$X^T W^{-1} = (V_G + Q)^{-1} Q X^T V^{-1}.$$

Inserting this into equation 8 gives

$$\begin{aligned} \widehat{\beta}_G &= \left( X_G^T (V_G + Q)^{-1} X_G \right)^{-1} X_G^T (V_G + Q)^{-1} \left( X^T V^{-1} X \right)^{-1} X^T V^{-1} Y, \\ \mathrm{Cov}(\widehat{\beta}_G) &= \left( X_G^T (V_G + Q)^{-1} X_G \right)^{-1}, \end{aligned}$$

which becomes equivalent to equation 9 if

$$V_{G2} = V_G + Q. \qquad \square$$

For this choice of $V_{G2}$, the group-level parameter estimates can be written as

$$\widehat{\beta}_G = \left(X_G^T \left(V_G + Q\right)^{-1} X_G\right)^{-1} X_G^T \left(V_G + Q\right)^{-1} \widehat{\beta},$$

that is, they become a function of the first-level parameter estimates $\widehat{\beta}$ and their associated covariances $Q = \left(X^T V^{-1} X\right)^{-1}$ only.

Note that by simply applying this theorem multiple times, these results extend to any multi-level GLM. For example, one can calculate parameter estimates for groups of groups using a multi-level approach by only keeping track of parameter estimates and associated covariances at each level. Note also that parts of this proof can simply be obtained by characterising the necessary conditions on sufficient statistics for $\beta_G$ [16].

# 4 Confounds

The previous section assumed that all the first-level parameter estimates would be used in the second-level model. This is often not true, as parameters of no interest (confounds) are often present in the first-level to remove unwanted signals (e.g. motion estimates as regressors to remove motion-related artefacts). This section shows that, when confounds are present in the first level, the model equivalence theorem still holds. That is, the second-level (group) analysis does *not* need to know about the confounds defined in the first level and only requires the estimates and covariance estimates of the parameters of interest. The proof of this result follows.

Consider a first-level GLM including confounds. This model can be written as

$$Y_k = X_k \beta_k + Z_k \alpha_k + \epsilon_k,$$

where $\alpha_k$ are the parameters of no interest (confounds) and $Z_k$ contains the regressors of no interest. Combining across subjects but grouping the $\beta_k$ and $\alpha_k$ separately gives the model

$$Y = X' \beta' + \epsilon,$$

where

$$X' = [X \ Z] = \begin{bmatrix} X_1 & 0 & \cdots & 0 & Z_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 & 0 & Z_2 & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & X_N & 0 & 0 & \cdots & Z_N \end{bmatrix}, \quad \text{and} \quad \beta' = \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}.$$

The parameter estimates in this case are

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\alpha} \end{bmatrix} = \begin{bmatrix} (X^T V^{-1} X) & (X^T V^{-1} Z) \\ (Z^T V^{-1} X) & (Z^T V^{-1} Z) \end{bmatrix}^{-1} \begin{bmatrix} X^T V^{-1} \\ Z^T V^{-1} \end{bmatrix} Y.$$

It is always possible to re-write such a model so that the confounds are orthogonal to the signals of interest in the pre-whitened space, without affecting the estimates of the signals of interest. This result is formally stated and proven in the appendix as Theorem A. Therefore, we can impose

$$Z^T V^{-1} X = 0 \qquad (12)$$

which gives

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\alpha} \end{bmatrix} = \begin{bmatrix} (X^T V^{-1} X)^{-1} X^T V^{-1} Y \\ (Z^T V^{-1} Z)^{-1} Z^T V^{-1} Y \end{bmatrix}. \qquad (13)$$

In the second-level, only the parameters of interest, $\widehat{\beta}$, appear. That is

$$\widehat{\beta} = X_G \beta_G + \eta',$$

which is identical to equation 6, so that the parameter estimates are exactly the same as those given in equation 7.

On the other hand, the single-level model with confounds is written as

$$Y = [(X X_G) \;\; Z] \begin{bmatrix} \beta_G \\ \alpha \end{bmatrix} + X\eta + \epsilon,$$

which has the same noise terms and covariance structure as before. The parameter estimates are given by

$$\begin{bmatrix} \widehat{\beta}_G \\ \widehat{\alpha} \end{bmatrix} = \begin{bmatrix} (X_G^T X^T W^{-1} X X_G) & (X_G^T X^T W^{-1} Z) \\ (Z^T W^{-1} X X_G) & (Z^T W^{-1} Z) \end{bmatrix}^{-1} \begin{bmatrix} X_G^T X^T W^{-1} \\ Z^T W^{-1} \end{bmatrix} Y.$$

Using equation 11 from the previous section, together with the orthogonality condition in equation 12, it is easy to show that:

$$Z^T W^{-1} = Z^T V^{-1} \quad \text{and hence} \quad Z^T W^{-1} X = 0.$$

This gives the parameter estimates as:

$$\widehat{\beta}_G = (X_G^T X^T W^{-1} X X_G)^{-1} X_G^T X^T W^{-1} Y \tag{14}$$

$$\widehat{\alpha} = (Z^T V^{-1} Z)^{-1} Z^T V^{-1} Y \tag{15}$$

where equation 14 is the same as equation 8, and consequently, using the previous theorem, equivalent to the two-level estimation when $V_{G2} = V_G + (X^T V^{-1} X)^{-1}$, while equation 15 is precisely the same as the two-level version (equation 13) independent of the choice of $V_{G2}$. The variances of the parameter estimates are also equal under these conditions, which can be verified by straightforward calculations.

Therefore the model equivalence result still holds when confounds are present.

# 5 Contrasts

This section demonstrates that the model equivalence result also holds when contrasts are used. That is, when contrasts are specified at the first level, only the contrasts of parameter estimates (copes), and their covariances, need to be used in the second level.

Contrasts are defined in the linear model in order to take linear combinations of the given parameters that form more meaningful or useful quantities. For instance, the difference between the responses to two different stimuli is often of greater interest than the individual responses, although it is easier to specify the design matrices, $X_k$, in terms of the individual stimuli.

Mathematically, each contrast is a column vector that represents how the old parameters, $\beta$, are combined to form new parameters, $b$. A collection of contrasts can be represented by a matrix $C_1$, such that

$$b = C_1^T \beta,$$

where, for the time being, we assume $C_1$ only contains linearly independent contrasts.

When there are fewer contrasts specified than original parameters this implicitly defines a set of confounds. These confounds represent the remaining amount of signal modelled by the original parameters but not by the copes. It is shown in Theorem B (in the appendix) that a model with contrasts can be written as a model involving confounds but no contrasts. Using this theorem the problem can be reformulated into a confound problem for which model equivalence has already been established.

In terms of the second-level analysis, it is only the desired contrasts of parameter estimates (copes), $\widehat{b}$, and their covariances, $\mathrm{Cov}(\widehat{b})$, which are required. That is

$$\widehat{b} = X_G \beta_G + \eta'.$$

Hence the introduction of contrasts does not affect the form of the model, but just substitutes $\widehat{b}$ and $\mathrm{Cov}(\widehat{b})$ for $\widehat{\beta}$ and $\mathrm{Cov}(\widehat{\beta})$ respectively at the second-level.

In the above derivation, the only restriction is that the set of copes passed into the second level must not contain linear dependence. This is for precisely the same reasons that the regressors in any GLM must not be linearly dependent. Consequently, the set of contrasts specifying these copes at the first level must be linearly independent. This, however, does not prevent linearly dependent copes being estimated in the first level, but just that they cannot be passed into a second-level analysis all at once — they would need to be combined into a linearly independent set first or tested in separate second-level analyses, if desired.

Consequently, all the previous results still hold when using contrasts.

# 6 Estimation of Variance Components

In the previous sections it was assumed that all variance terms are known a-priori. In practice, these quantities are unknown and will need to be estimated as part of the model fitting. Variance component estimation is a challenging task in itself, having generated a variety of approaches. Any approach to variance estimation (or combination of approaches) can easily be combined with the multi-level GLM to provide a practical multi-level method; this section discusses some of the more popular approaches.

There are no differences between the first level and any other level from the modelling perspective; however, in practice, estimating the variances is substantially different. At the first-level there typically exists considerable serial auto-correlation (in FMRI time series data) but with a large number of observations. A considerable amount of literature is devoted to specifying the form for the first-level covariance matrix $V$ and estimating its parameters in the single session case [17, 14, 19]. In contrast, higher level variance component estimation is typically troubled by having very few observations, while serial auto-correlation between these normally is, and often can be, ignored.

When the number of observations is very low, this imposes restrictions in the types of model which are practically estimable. For instance, while it is possible to formulate a model where the variance about the group mean is different for each session/subject, such a model is not estimable because there is only a single measurement per session/subject.

Several approaches to estimation within the multi-level GLM currently exist, of which the parametric techniques can, most easily, be split into Bayesian and non-Bayesian approaches.

Classically, variance components tend to be estimated separately using iterative estimation schemes employing Ordinary Least Squares (OLS), Expectation Maximisation (EM) or Restricted Maximum Likelihood (ReML), see [16] for details.

The specific choice of the variance component estimator will also determine the 'effective' degrees of freedom of the variance estimate. For ordinary least squares estimates, the degrees of freedom simply is $n - 1$. When more sophisticated estimation techniques and/or variance structures are being used, this changes significantly [19, 10]; e.g. when enforcing mixed- effects variance to be greater than the fixed-effects variance, this can increase the effective degrees of freedom, a quantity typically estimated using a Satterthwaite approximation [15].

As an example of a non-Bayesian approach, [19] estimates variance components at each level of the split-level model separately. First-level estimation incorporates autoregressive AR($p$) noise estimated from the lag-$l$-autocovariance matrices and utilises the pre-whitening approach to generate BLUEs for the first-level parameters. At higher levels, they propose EM for estimation of the random effects variance contribution, in order to reduce bias in the variance estimation - a potential problem in higher-level analyses if simple OLS were used (note that [12] shows that this is not necessary at first-level). Positivity of the random- effects variance, avoiding what is known as the 'negative variance problem' [11] (where mixed-effects variance estimates are lower than fixed-effects variances implying negative random-effects variance), is partially addressed but not strictly enforced. As a separate stage, in order to boost effective degrees of freedom for later mixed-effects inference, the authors propose to post-hoc spatially regularise the random-effects variance via smoothing a random variance ratio image. As a consequence, the resulting analysis does not produce a mixed-effects analysis. It should be noted that this specific form of spatial regularisation (or indeed any) is not a necessary ingredient of an EM approach to variance estimation.

In contradistinction [6] have proposed an empirical-Bayesian approach [6] for estimation of the full single-level model. Unlike the previous case and the techniques advocated in this paper, this relates the parameters of interest to the full set of original data, i.e. does not utilise the 'summary statistics' approach. Parameter and variance component estimation is no longer separated. Conditional posterior point estimates are generated using EM which give rise to posterior probability maps.

More recently, [1] have placed the higher-level GLM estimation in a *fully* Bayesian framework. Using appropriate reference priors, the method is based on Markov-Chain-Monte-Carlo sampling from the full posterior distribution of $p(\beta_G|\widehat{\beta})$. Under the parametric model assumptions, the posterior has a non-central $t$-distribution which the method fits to the MCMC samples. This is done in order to both estimate the appropriate degrees of freedom for the mixed-effects inference, and in order to avoid having to sample densely far into the tail of the posterior. As in the empirical-Bayesian case, parameters and their relevant variance components are estimated together. Also, as the number of degrees of freedom is estimated as part of the $t$-distribution fit there is no need to separately approximate this quantity e.g. via the Satterthwaite approximation [16]. This technique provides for an unbiased and efficient estimation of multi- level GLMs within the 'summary statistics' approach, also strictly enforcing positivity of variance components at all levels. This technique, combined with a pre-whitening approach to first-level estimation [17] has been implemented as part of FSL [1, 7].

# 7 Limitations

In this section we discuss the limitations of the model framework presented above. The first and most obvious limitation is that the framework is purely linear, as it is an instance of the GLM. Linear modelling, however, is a very common and flexible model for FMRI data. Secondly, in order for the model to be decomposed into two levels, and hence allow for efficient two-level computation, it is necessary that the stacked first-level covariance matrix, $V$, does not contain any non-zero off-diagonal blocks. This is equivalent to assuming that the first-level residuals are not correlated between sessions or subjects.

Efficient computation at the second-level requires full access to the first-level parameter estimates and associated covariances. This involves both the variances of the parameter estimates and the covariances *between* different parameters. It is not sufficient to only use the first-level statistical parametric maps (i.e. $t$-scores, $z$-scores or $F$-scores). Finally, the estimation of covariance parameters at the second-level often imposes restrictions in the types of model which are practically estimable. This is often more problematic at the second level because there is usually only a small number of subjects/sessions that are being modelled. For instance, while it is possible to formulate a model where the variance about the group mean is different for each session/subject, such a model is not estimable because there is only a single measurement per session/subject.

# 8 Examples

In this section we show how various group level parameters of interest can easily be calculated within the GLM framework. This amounts to specifying a suitable group design matrix $X_G$, a covariance structure $V_G$ and possibly a contrast vector $C_G$. Note that unlike the case of first-level designs, the mean parameter value is often of interest and hence the design matrix, $X_G$, must always model the group mean activation, that is, the unit vector must always be included in the span of $X_G$.

For several of the examples it is easy to show the added benefit of the proposed framework using numerical simulations. These principally contrast the heteroscedastic model (allowing for different individual first-level variances) with the homoscedastic model (where first-level variances are assumed to be identical). These comparisons show substantial increases in $Z$-statistics over a wide range of realistic scenarios.

## 8.1 Average Group Activation

In order to calculate the average group activation, we model the individual subject activation as being normally distributed according to

$$\beta_k \sim \mathcal{N}(\beta_G, \sigma_s^2)$$

where $\beta_G$ represents the average group activation and is usually estimated as

$$\widehat{\beta}_G = \frac{1}{N}\sum_k \widehat{\beta_k}$$

and where $\sigma_s^2$ denotes the between-subject variance.

We will model the first-level within-subject covariances to be subject-specific and model the between-subject variances (from the group mean) as equal across the group. That is

$$V = \begin{bmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_N \end{bmatrix}, \qquad V_G = \sigma_s^2 \mathbf{I}, \qquad \text{and} \qquad X_G = (1 \cdots 1)^T.$$

Then the adjusted second-level covariance matrix is

$$\begin{aligned} V_{G2} &= V_G + Q = V_G + \left(X^T V^{-1} X\right)^{-1} \\ &= \begin{bmatrix} \left(X_1^T V_1^{-1} X_1\right)^{-1} & & 0 \\ & \ddots & \\ 0 & & \left(X_N^T V_N^{-1} X_N\right)^{-1} \end{bmatrix} + \sigma_s^2 \mathbf{I}. \end{aligned}$$

Define $u_k = \left(X_k^T V_k^{-1} X_k\right)^{-1} + \sigma_s^2$, which is the sum of the within- and between-subject covariances. Then the estimate of the group parameter writes as

$$\widehat{\beta}_G = \left(X_G^T V_{G2}^{-1} X_G\right)^{-1} X_G^T V_{G2}^{-1} \widehat{\beta} = \left(\sum_i u_i^{-1}\right)^{-1} \left(u_1^{-1} \cdots u_N^{-1}\right) \widehat{\beta}, \tag{16}$$

where the inverse sum over $u_k^{-1}$s is the associated variance.

Hence we see that in the general framework, the mean group activation parameter is a weighted average of the combined subject- specific activations, where the weights are inversely proportional to the subject-specific variances. This adjustment is advantageous in the case where the individual time-series model does not fit well for a particular subject, $j$, generating an unusual value for $\widehat{\beta}_j$ (an outlier) but also a large $\text{Var}(\widehat{\beta}_j)$. If no correction for the first-level variance is done, then this outlier can significantly affect the estimation of the group (between-subject) variance. If, however, this first-level correction is performed, the increased variance in this parameter will effectively de-weight the contribution of this outlier to the group variance estimate, since we use General Least Squares estimation.

In the much simpler case, where the within-subject covariances are $V_k = \sigma_w^2 \mathbf{I}$, and the $X_k$ are normalised, such that $X_k^T X_k = 1$ for all $k$, then

$$u_k = \sigma_w^2 + \sigma_s^2 \ \text{ and } \ \widehat{\beta}_G = \frac{1}{N} \sum_k \widehat{\beta_k} \ \text{ with associated variance } \ \text{Var}(\widehat{\beta}_G) = \frac{\sigma_w^2 + \sigma_s^2}{N}. \tag{17}$$

The test for significance is then carried out in a $t$-test where

$$T = \widehat{\beta}_G / \sqrt{\text{Var}(\widehat{\beta}_G)}$$

has an approximate $t$-distribution.

#### 8.1.1 Numerical simulation

In this section we illustrate the potential benefit of a heteroscedastic variance model (i.e. by allowing for separate first-level variances) compared to the homoscedastic variance model (otherwise known as OLS for random-effects analysis of FMRI data [9]).

Both the heteroscedastic model (equation 16) and the homoscedastic model (equation 17) provide an unbiased estimate of the group level parameter of interest $\beta_G$. They differ, however, in their associated variance, $\text{Var}(\widehat{\beta}_G)$, and therefore will give different $Z-$ or $T-$statistic. The associated variance for the heteroscedastic model is always less than or equal to that of the homoscedastic model, as can be shown using Jensen's inequality. This will result in an increase in the expected statistics values for the model in equation 16:

$$\frac{\langle \Delta Z \rangle}{\langle Z_{\text{homo}} \rangle} = \sqrt{\frac{\langle \text{Var}(\widehat{\beta}_G) \rangle_{\text{homo}}}{\langle \text{Var}(\widehat{\beta}_G) \rangle_{\text{hetero}}}} - 1$$

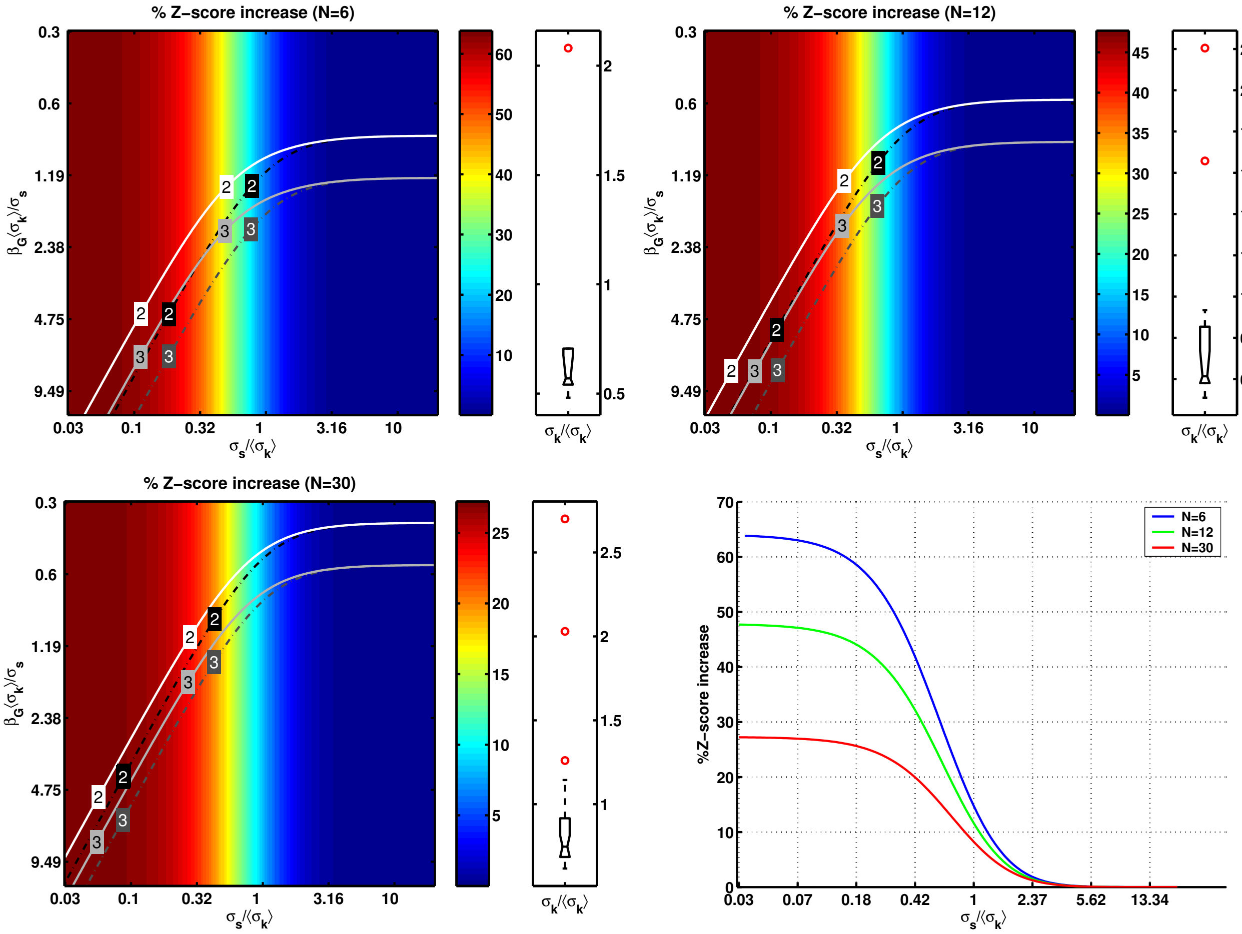


Figure 1: Expected % $Z$-score increase for the mean group activation for the case of typical first level covariance containing a few large outliers with different group sizes (N=6, 12 and 30). The first three images shows the increase as a function of the ratio of group level and mean first-level standard deviations ($\sigma_s/\langle\sigma_k\rangle$ on the $x$-axis) vs. relative group level activation ($y$-axis). Contour lines are shown for group-level $Z = 2.0$ and $Z = 3.0$ in the case of the heteroscedastic model (equation 16; solid lines) and the homoscedastic model (OLS; equation 17; dash-dotted). The difference between the solid and dash-dotted lines corresponds exactly to the color-coded percentage increase. The boxplots show the individual first level standard deviation compared to their mean ($\sigma_k/\langle\sigma_k\rangle$) as an indicator of heteroscedacity. The percentage changes, which are independent of $\beta_G$, are summarised in the fourth plot for all three variance configurations.

As a quantitative example of this increase we have generated a simulated group FMRI study, with a known set of first-level variances $V_k = \sigma_k^2\mathbf{I}$, together with a second-level variance $\sigma_s^2$. From these known values we calculate the expected percentage increase in $Z$- statistic (see above) and show this versus changes in the ratio of the second-level variance to the mean first-level variance (the assumed variance in the homo- scedastic model). First-level variances were taken from a real FMRI group study (simple motor paradigm) estimated using the GLS pre-whitening approach as implemented in FILM (part of FSL [7, 17]). These estimates of first-level variances are realistic representatives of typical first-level variance structures and are defined to be ground-truth in the following simulation.

Figure 1 shows the expected percentage increase in $Z$-scores for three different sets of first-level variances, each of them containing one or more 'outliers' (see boxplots). As can be seen from the first three images, the expected percentage increase, for a given first-level variance structure, only depends on the ratio of second-level variance to mean first-level variance ($\sigma_s/\langle\sigma_k\rangle$) and is independent of the group-level effect size $\beta_G$. As the second- level variance becomes considerably larger than the mean first- level variance, the expected increase in $Z$-score tends to zero. When these variances are approximately equal, the heteroscedastic model allows for a $\sim 7-14\%$ increase in $Z$-score, increasing to much larger values as the second-level variance decreases relative to the mean first-level variance. The group-level effect size determines the actual $Z$-level, as shown by the

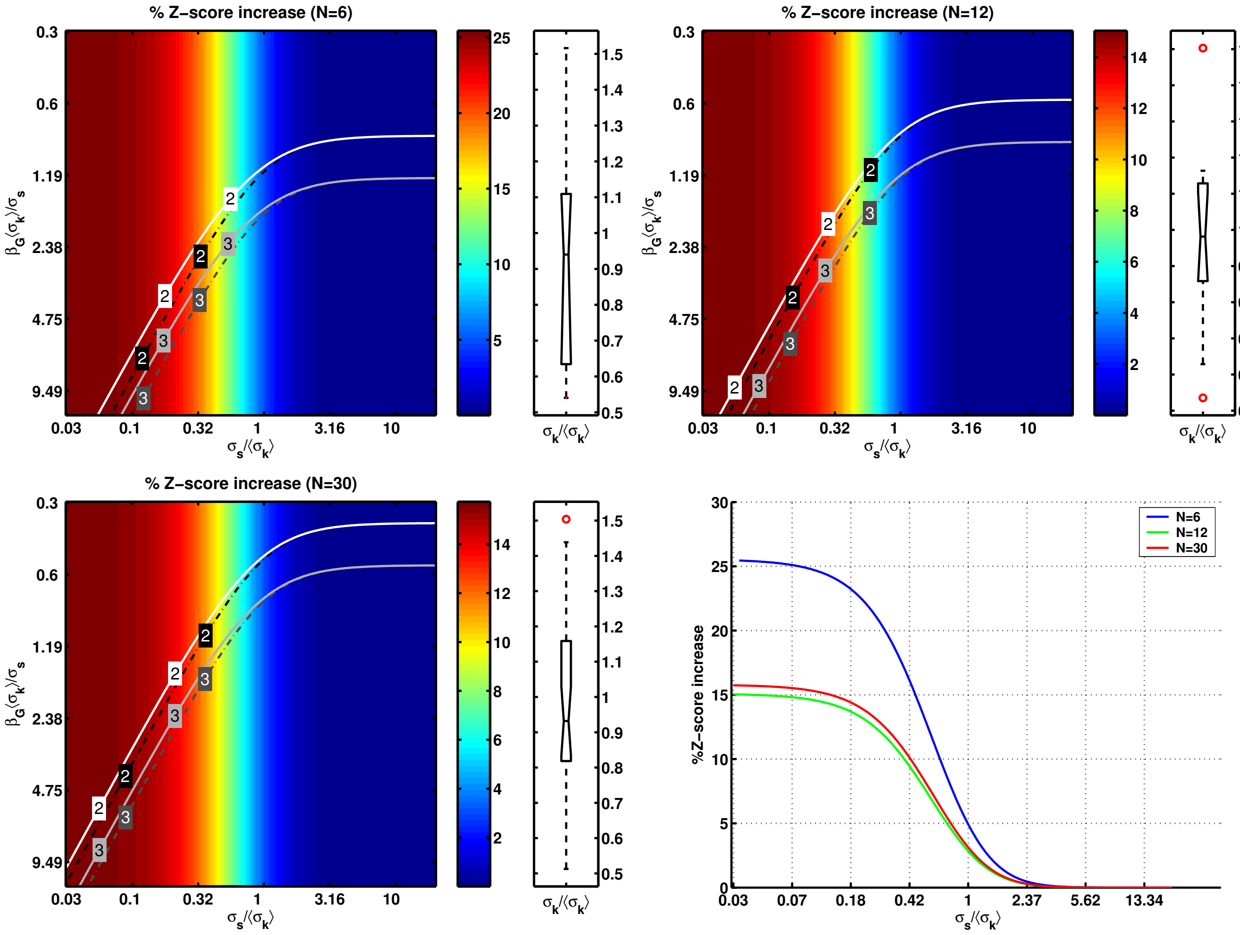


Figure 2: Expected % $Z$-score increase for the mean group activation for the case of typical first level covariance (no large outliers) with different group sizes (N=6, 12 and 30). See figure 1 for details.

overlaid contour plots (solid lines show $Z$- score levels for the heteroscedastic model while dash-dotted lines show the $Z$-score levels for the homoscedastic model). Over a large range of variance ratios the improvements gained by the heteroscedastic model have substantial implications on post- thresholded results, e.g. typical sub-threshold values of $2.0$ increase to super-threshold values of $\sim 3.0$.

Figure 2 shows similar plots for cases where the first-level variances do not contain large outliers, that is, they are close to the homoscedastic model. In this case, the expected percentage $Z$-score increase is smaller, but still potentially important.

It is clear from these figures, that the exact configuration of first-level variances has a major impact on the improvements gained by using the heteroscedastic model. This dependency is further investigated in figure 3, where we show the expected (log-) increase in $Z$-score for a set of variance configurations estimated from $50200$ voxels in a real group study, assuming equal second-level and mean first-level variances. This histogram shows that approximately $1000$ voxels have a $> 5\%$ increase in expected $Z$-score.

## 8.2 Unpaired Group Difference

We assume that the individual subjects are grouped in two groups and that within each group the first-level parameters are normally distributed around a group-specific mean. That is

$$\beta_k \sim \mathcal{N}(\mu_1, \sigma_{s_1}^2), k \in G_1 \quad \text{and} \quad \beta_k \sim \mathcal{N}(\mu_2, \sigma_{s_2}^2), k \in G_2.$$

In order to simplify further notation and without loss of generality we assume that the subjects $1, \ldots, r$ belong to the first group and subjects $r+1, \ldots, N$ belong to the second group. We do not make any assumption about

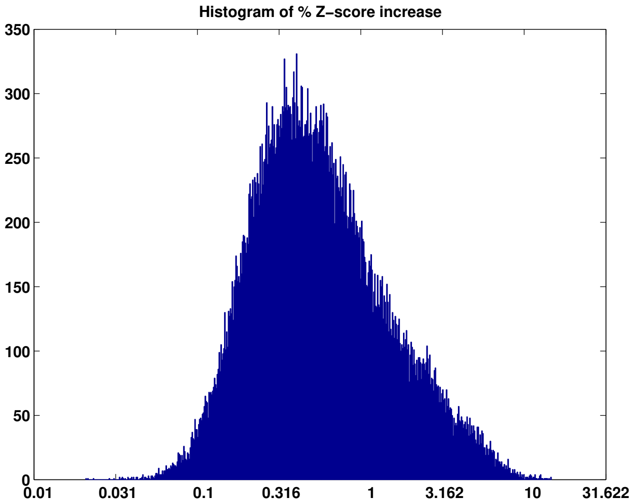


Figure 3: Histogram of log- % $Z$-score increase (50200 voxels, N=12 subjects, $\sigma_s/\langle\sigma_k\rangle = 1$): $\sim$ 2% of voxels (about 1000) have an increase of $> 5\%$ in $Z$-score.

the first-level covariance structure and simply set

$$V = \begin{bmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_N \end{bmatrix}, \; V_G = \begin{bmatrix} \sigma^2_{s_1}\mathbf{I}_r & 0 \\ 0 & \sigma^2_{s_2}\mathbf{I}_{N-r} \end{bmatrix}, \; X_G = \begin{pmatrix} 1\cdots 1 & -1\cdots -1 \\ 1 & \cdots \quad 1 \end{pmatrix}^T \quad \text{and} \quad C_G = (2 \;\; 0)^T.$$

Then $V_{G2}$ is block diagonal with elements $u_k = \left(X_k^T V_k^{-1} X_k\right)^{-1} + \sigma^2_{s_k}$. If we define $s_1 = \sum_{G1} u_k^{-1}$ and $s_2 = \sum_{G2} u_k^{-1}$, the group parameter estimate writes as

$$\begin{aligned} \widehat{b_G} &= C_G^T \left[ \begin{pmatrix} \frac{1}{u_1}\cdots\frac{1}{u_r} & \frac{-1}{u_{r+1}}\cdots\frac{-1}{u_N} \\ \frac{1}{u_1}\cdots\frac{1}{u_r} & \frac{1}{u_{r+1}}\cdots\frac{1}{u_N} \end{pmatrix} X_G \right]^{-1} X_G^T V_{G2}^{-1} \widehat{\beta} \\ &= \frac{C_G^T}{4 s_1 s_2} \begin{bmatrix} s_1+s_2 & s_2-s_1 \\ s_2-s_1 & s_1+s_2 \end{bmatrix} \begin{bmatrix} \sum_{G1}\frac{\widehat{\beta_k}}{u_k} - \sum_{G2}\frac{\widehat{\beta_k}}{u_k} \\ \sum_{G1}\frac{\widehat{\beta_k}}{u_k} + \sum_{G2}\frac{\widehat{\beta_k}}{u_k} \end{bmatrix} \\ &= \left( \sum_{G1} \frac{\widehat{\beta_k}}{s_1 u_k} - \sum_{G2} \frac{\widehat{\beta_k}}{s_2 u_k} \right), \end{aligned}$$

where the variance, as usual, is calculated from the first term as

$$\mathrm{Var}(\widehat{b_G}) = C_G^T \left( X_G^T V_{G2}^{-1} X_G \right)^{-1} C_G = \frac{1}{s_1} + \frac{1}{s_2}.$$

Under the same assumptions as before, of equal covariance at the first level and normalised designs (i.e. homoscedastic model), these equation simplify to

$$u_k = \sigma_w^2 + \sigma^2_{s_k}, \quad s_1 = \frac{r}{\sigma_w^2 + \sigma^2_{s_1}}, \quad s_2 = \frac{N-r}{\sigma_w^2 + \sigma^2_{s_2}}, \quad \text{and thus} \quad \widehat{b_G} = \frac{1}{r}\sum_{G1}\widehat{\beta_k} - \frac{1}{N-r}\sum_{G2}\widehat{\beta_k}$$

with

$$\mathrm{Var}(\widehat{b_G}) = \frac{\sigma^2_{s_1}}{r} + \frac{\sigma^2_{s_2}}{N-r} + \frac{N\sigma_w^2}{r(N-r)}.$$

Note that the second level contrast includes an appropriate scaling constant. This factor becomes irrelevant once the group parameter of interest is combined with its variance to form a test statistic.

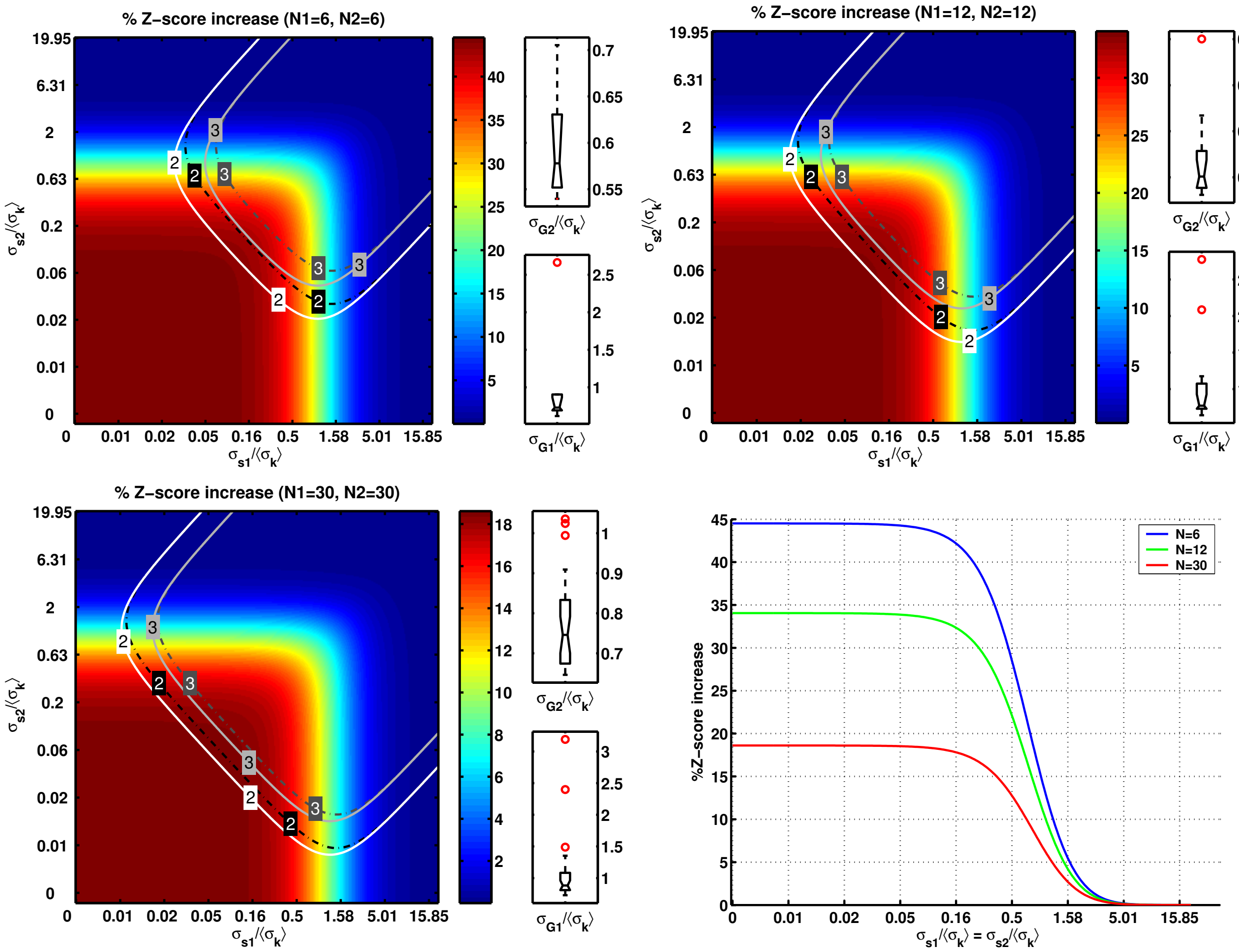


Figure 4: Expected % $Z$-score increase for the unpaired group difference at different group sizes (N=6, 12 and 30). The first three images show the increase as a function of the standard deviation of group 1 and mean first-level standard deviation ($\sigma_{s1}/\langle\sigma_k\rangle$ on the$x$-axis) vs. the standard deviation of group 2 relative to the mean first level standard deviation ($\sigma_{s2}/\langle\sigma_k\rangle$ on the $y$-axis). Contour lines are shown for the $Z = 2.0$ and $Z = 3.0$ group level thresholds in the case of the heteroscedastic model (solid lines) and homoscedastic model (OLS; dash-dotted). The boxplots show the individual first level standard deviation compared to the mean first-level standard deviation ($\sigma_k/\langle\sigma_k\rangle$) as an indicator of heteroscedacity. The expected increase in $Z$- score (assuming $\sigma_{s1} = \sigma_{s2}$) are shown in the fourth plot.

### 8.2.1 Numerical simulation

Figure 4 shows numerical simulation of the expected $Z$-score increase for different first-level variance configurations. As before, the expected increase is independent of the second-level effect size but will depend on the first- level variance configuration for group 1 and group 2 as well as the different second-level variances $\sigma_{s1}$ and $\sigma_{s2}$. The added flexibility of the heteroscedastic model is important for a variety of real FMRI experiment where the two groups naturally will have different variance configurations, e.g. studies of patients vs. non-patients. Once again, significant changes in $Z$-score (e.g. $> 10\%$) can be seen over a large set of configurations.

## 8.3 Repeated Measures

In the case of repeated measures, where some or all $N$ subjects were scanned more than once, we might want to account for a possible difference in the between-subject variance and the between-session variance of the same subject. Thus, we simply extend the previous model to write

$$\beta_{kj} \sim \mathcal{N}(\bar{\beta}_k, \sigma_b^2) \quad \text{and} \quad \bar{\beta}_k \sim \mathcal{N}(\beta_G, \sigma_s^2),$$

where $\beta_{kj}$ denotes the first-level parameter for subject $k$ obtained at session $j$, $\beta_G$ is the population mean, $\bar{\beta}_k$ is the subject-specific mean, $\sigma_s^2$ is the between-subject variance and $\sigma_b^2$ is the between-session variance. Note that this can be extended to the case of repeated measures under different conditions by introducing more variance terms. The existence of these additional variance terms will be reflected in the covariance matrix $V_G$ simply by allowing for off-diagonal elements. As an example, consider the case of 3 subjects each measured twice, then with $\sigma_c^2 = \sigma_b^2 + \sigma_s^2$

$$V_G = \begin{bmatrix} \sigma_c^2 & \sigma_s^2 & & & & \\ \sigma_s^2 & \sigma_c^2 & & & & \\ & & \sigma_c^2 & \sigma_s^2 & & \\ & & \sigma_s^2 & \sigma_c^2 & & \\ & & & & \sigma_c^2 & \sigma_s^2 \\ & & & & \sigma_s^2 & \sigma_c^2 \end{bmatrix},$$

where, for simplicity, we order the subjects in pairs appropriately. Hence, we can simply adjust any covariance structure $V_G$ at the group level to take account of the existence of repeated measurements. Alternatively, note that because our previous results extend to multi-level GLMs, we can simply combine the different sessions per subject on a second level and calculate group parameters of interest in a third level.

## 8.4 Paired $t$-tests

Let us assume that for each of $N$ subject there exist two measurements obtained under different conditions $s_1$, $s_2$ and that we are interested in the significance of the mean group difference $b_G = \sum_k b_k/N = \sum_k(\beta_{ks_1} - \beta_{ks_2})/N$. We will assume that the between-subject variance and the between-session variance is equal across subjects and conditions. Note that this is a notational simplification within this framework which might or might not become a necessary condition once we try to estimate the associated group-level parameters. Similar to the previous sections, we model this as

$$\beta_{ks_i} \sim \mathcal{N}(\mu_i, \sigma_b^2 + \sigma_s^2), \quad \text{and} \quad \mathrm{Cov}(\beta_{ks_1}, \beta_{ks_2}) = \sigma_s^2.$$

Let

$$V = \begin{bmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_{2N} \end{bmatrix}, \quad V_G = \begin{bmatrix} U & & 0 \\ & \ddots & \\ 0 & & U \end{bmatrix} \text{ with } U = \begin{bmatrix} \sigma_c^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_c^2 \end{bmatrix} \text{ and } X_G = \begin{bmatrix} 1 & 1 & & & \\ -1 & 1 & & & \\ & & 1 & & \\ \vdots & & 1 & & \\ & & & \ddots & \\ 1 & & & & 1 \\ -1 & & & & 1 \end{bmatrix},$$

where, again, $\sigma_c^2 = \sigma_b^2 + \sigma_s^2$ and where the group design matrix, $X_G$, de-means the first level estimates for each subject. Assume for simplicity that $V_k = \sigma_w^2 \mathbf{I}$, $X_k^T X_k = 1$ and define $u^2 = \sigma_w^2 + \sigma_c^2$. Then $V_{G2}^{-1}$ will be block diagonal with blocks of

$$\widetilde{U}^{-1} = \frac{1}{u^4 - \sigma_s^4} \begin{bmatrix} u^2 & -\sigma_s^2 \\ -\sigma_s^2 & u^2 \end{bmatrix}, \quad \text{where} \quad \widetilde{U} = \begin{bmatrix} u^2 & \sigma_s^2 \\ \sigma_s^2 & u^2 \end{bmatrix}.$$

Furthermore, let $c_1 = u^2 + \sigma_s^2 = \sigma_w^2 + \sigma_b^2 + 2\sigma_s^2$ and $c_2 = u^2 - \sigma_s^2 = \sigma_w^2 + \sigma_b^2$. Then

$$X_G^T V_{G2}^{-1} = \begin{bmatrix} c_2^{-1} & -c_2^{-1} & \dots & c_2^{-1} & -c_2^{-1} \\ c_1^{-1} & c_1^{-1} & & & \\ & & \ddots & & \\ & & & c_1^{-1} & c_1^{-1} \end{bmatrix}, \quad \text{and} \quad \left(X_G^T V_{G2}^{-1} X_G\right)^{-1} = \frac{1}{2} \begin{bmatrix} c_2/N & & & \\ & c_1 & & \\ & & \ddots & \\ & & & c_1 \end{bmatrix},$$

so that

$$\widehat{\beta}_G = \frac{1}{2}\left(\sum_k \frac{\widehat{\beta}_{ks_1} - \widehat{\beta}_{ks_2}}{N}, \widehat{\beta}_{1s_1} + \widehat{\beta}_{1s_2}, \dots, \widehat{\beta}_{Ns_1} + \widehat{\beta}_{Ns_2}\right)^T.$$

Using $C_G = (1, 0, \dots, 0)^T$, the group parameter estimate writes as

$$\widehat{b}_G = C_G^T \widehat{\beta}_G = \sum_k \frac{\widehat{\beta}_{ks_1} - \widehat{\beta}_{ks_2}}{N} \quad \text{and} \quad \mathrm{Var}(\widehat{b}_G) = \frac{c_2}{2N}.$$

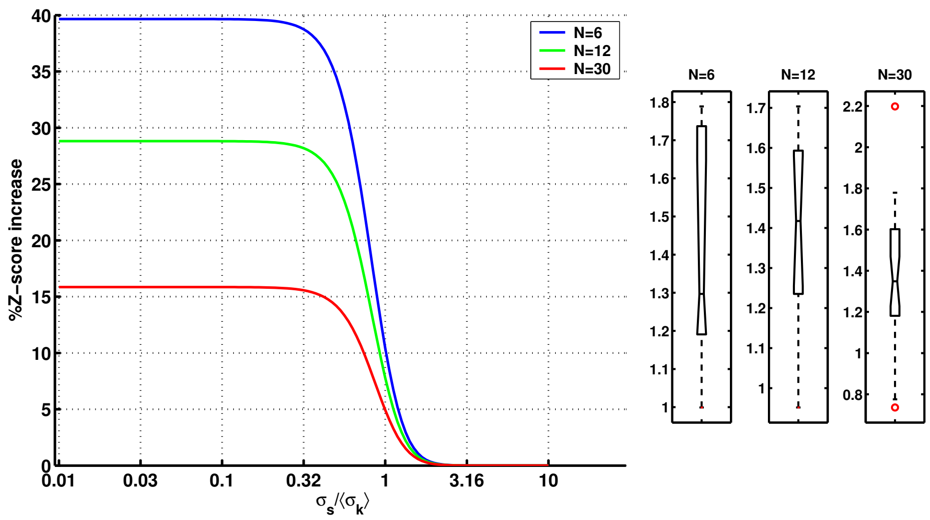


Figure 5: Expected % increase in $Z$-score for the paired $t$-test at different group sizes (N=6, 12 and 30). Boxplots show the standard deviation of the paired differences.

As expected, the variance of the group level estimate no longer depends on the between-subject variance $\sigma_s^2$. Note that this approach is equivalent to using a three level approach with an unpaired $t$-test of de-meaned repeated measures.

#### 8.4.1 Numerical simulation

Similar to before, figure 5 shows the expected percentage increase in $Z$-score for the heteroscedastic vs. the homoscedastic model. Again, depending on the explicit configurations of first-level variances, a substantial increase in $Z$-score can be observed.

### 8.5 $F$-tests

Assume that for a set of $N$ subjects we model their hæmodynamic response functions using the same set of $M$ basis functions for each subject and consider the case where we wish to test for the population mean activation. Thus, we implicitly assume that there exists a single hæmodynamic response function that is representative for the group activation. We will not restrict the choice of basis functions (i.e. we do not require the basis functions to be orthogonal) and therefore allow for general correlations between the individual basis funtion fits, but assume that the covariance structure is the same for each individual. That is, we model the subject-specific vector of fits as distributed according to a multivariate normal distribution

$$\boldsymbol{\beta}_k \sim \mathcal{N}(\boldsymbol{\beta}_G, V_B),$$

and let

$$V_G = \begin{bmatrix} V_B & & 0 \\ & \ddots & \\ 0 & & V_B \end{bmatrix},$$

where $V_B$ is the covariance matrix of the $M$ basis function fits. Then the group-level design matrix

$$X_G = [\mathbf{I}_M, \cdots, \mathbf{I}_M]^T$$

combines the $M$ individual basis functions across subjects such that

$$\widehat{\beta}_G = \left(\widehat{\beta_{G_1}}, \ldots, \widehat{\beta_{G_M}}\right)^T,$$

where the individual values are the $M$ mean basis coefficients. In order to assess the average population activation, we need to test if any of the basis function coefficients are significantly non-zero. This can be achieved by calculating

$$F = \frac{1}{M} \beta_G^T C_G \left( \widehat{\mathrm{Var}}(C_G \beta_G) \right)^{-1} C_G \beta_G, \quad \text{with} \quad C_G = \mathbf{I}_M,$$

which approximately follows an $F$-distribution. If, instead, we wish to assess if the final $p$ basis functions contribute significantly to the mean fit, we simply set

$$C_G = [0 \quad \mathbf{I}_p]$$

and change $M$ to $p$.

# 9 Conclusion

In this paper we have shown that it is possible to efficiently test, using only first-level parameter and (co-)variance estimates, general hypotheses for a mixed-effects[2] group analysis model within the framework of the multi-level GLM under the BLUE for known variances. In particular, we have demonstrated the equivalence between a multi- level GLM and a single-level version if (and only if) the (co-)variance structure is modified appropriately. The result has a natural interpretation in that the variance used in the second-level analysis must be the sum of the (co-)variances from the individual and group levels. This eqivalence allows the second-level analysis to be carried out efficiently, only using first-level results, without any need to revisit the FMRI time-series data.

A further consequence of the model equivalence is that we can switch freely between the two forms of the model, depending on which one is more convenient at any one time. Also, the equivalence relation naturally extends to multi-level GLMs, so this framework encompasses more complex analysis scenarios, e.g. where data is grouped first across sessions and later across subjects.

This paper deliberately restricts itself to the issues involved in modelling and does not directly address the problem of estimation. The main problem in estimation for group studies in FMRI is the generally small number of subjects/sessions and the associated difficulty in obtaining accurate (co-)variance estimates. This problem will occur irrespective of the specific model used. Naturally, any advanced estimation technique can be used for this multi-level model. These problems of estimation have been investigated separately using a fully Bayesian approach [1]. This is implemented within the current release version of FSL [7] and will be the topic of a forthcoming paper.

We have provided examples of how various group designs can be formulated within this general framework, including commonly used designs such as paired and unpaired $t$-tests or $F$-tests at the group level. The explicit equations in each of these cases can be derived easily from equations 7 and 10, such that the differences in the exact formulation of these tests and their assumptions become explicit. For example, all results in section 8 have been derived without the need for balanced designs, i.e. they allow for different first-level regressors. Balanced designs, therefore, are not a requirement of the multi-level GLM but potentially are a constraint introduced by a specific estimation technique. This becomes relevant when the first-level regressors need to be subject specific, e.g. are related to the subject's performance in a cognitive task.

To demonstrate some of the benefits of more flexible mixed-effects modelling some numerical simulations have been generated which contrasted the heteroscedastic model versus the standard homoscedastic (OLS) random-effects model. Using typical first-level variance structures from real FMRI data, these simulations demonstrate that by taking into account the relevant lower-level variances and heterogeneity amongst them, a substantial increase in higher-level $Z$-scores is possible.

# Acknowledgements

The authors wish to thank Mark Woolrich, Didier Leibovici, Timothy Behrens and Joe Devlin for the contribution they made by taking part in many helpful discussions and thinking through examples and counter-examples which have helped to greatly improve this manuscript. Funding from the UK MRC and EPSRC and GSK is gratefully acknowledged.

[2]This naturally includes fixed-effects tests simply by assuming that the group-level (co-)variance matrix $V_G$ is zero.

# Appendix

The general form of the *Sherman-Morrison-Woodbury* formula [8] is

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1}. \tag{17}$$

Also, the inverse for a matrix in block form is given by

$$\begin{bmatrix} A & B \\ B^T & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A - BD^{-1}B^T\right)^{-1} & -A^{-1}B\left(D - B^TA^{-1}B\right)^{-1} \\ -D^{-1}B^T\left(A - BD^{-1}B^T\right)^{-1} & \left(D - B^TA^{-1}B\right)^{-1} \end{bmatrix}.$$

**Theorem A:**
Any model of the form

$$Y = [X_1 \;\; Z_1] \begin{bmatrix} \beta_1 \\ \alpha_1 \end{bmatrix} + \epsilon$$

can be rewritten as

$$Y = [X_2 \;\; Z_2] \begin{bmatrix} \beta_2 \\ \alpha_2 \end{bmatrix} + \epsilon,$$

where $Z_2^T V^{-1} X_2 = 0$ whilst being completely equivalent in terms of the estimated parameters of interest ($\widehat{\beta_2} = \widehat{\beta_1}$ , $\text{Cov}(\widehat{\beta_2}) = \text{Cov}(\widehat{\beta_1})$) and the modelled signal space: Span$(X_2)$ ∪ Span$(Z_2)$ = Span$(X_1)$ ∪ Span$(Z_1)$ in the pre-whitened space. Note that Cov$(\epsilon) = V$ for both models.
That is, the signals of interest can be made orthogonal to the confounds without affecting the estimation of the parameters or the residuals.
**Proof:**
The proof is by construction, where we show that orthogonalising $X_1$ with respect to $Z_1$ gives the desired results. Let

$$X_2 = X_1 - P_{Z_1} X_1 \quad \text{and} \quad Z_2 = Z_1$$

where $P_{Z_1} = Z_1(Z_1^T V^{-1} Z_1)^{-1} Z_1^T V^{-1}$ is the projection matrix for $Z_1$ in the pre-whitened space.
These equations give $Z_2^T V^{-1} X_2 = Z_1^T V^{-1}(\mathbf{I} - P_{Z_1})X_1 = 0$. Also, the combined span of $X_2$ and $Z_2$ is clearly the same as that of $X_1$ and $Z_1$.
Now consider the covariances

$$\text{Cov}\left(\begin{bmatrix} \widehat{\beta_1} \\ \widehat{\alpha_1} \end{bmatrix}\right) = \begin{bmatrix} X_1^T V^{-1} X_1 & X_1^T V^{-1} Z_1 \\ Z_1^T V^{-1} X_1 & Z_1^T V^{-1} Z_1 \end{bmatrix}^{-1} =: \begin{bmatrix} A & B \\ B^T & D \end{bmatrix}^{-1}.$$

Using the block matrix inverse, this gives

$$\text{Cov}(\widehat{\beta_1}) \;\; = \;\; \left(A - BD^{-1}B^T\right)^{-1},$$

while, since the off-diagonal blocks are zero in the second case, the calculation simply gives

$$\begin{aligned} \text{Cov}(\widehat{\beta_2}) &= (X_2^T V^{-1} X_2)^{-1} \\ &= \left\{(X_1^T - BD^{-1}Z_1^T)V^{-1}(X_1 - Z_1 D^{-1} B^T)\right\}^{-1} \\ &= \left(A - BD^{-1}B^T\right)^{-1} \\ &= \text{Cov}(\widehat{\beta_1}). \end{aligned}$$

For the first model, the parameter estimates, given by equation 5, can be written using the matrix block inversion formula, giving

$$\widehat{\beta_1} = \left(A - BD^{-1}B^T\right)^{-1} X_1^T V^{-1} Y - A^{-1}B\left(D - B^T A^{-1} B\right)^{-1} Z_1^T V^{-1} Y \tag{12}$$

while for the second model, the block diagonal form yields the familiar form

$$\begin{aligned}\widehat{\beta_2} &= (X_2^T V^{-1} X_2)^{-1} X_2^T V^{-1} Y \\ &= \mathrm{Cov}(\widehat{\beta_2}) \left( X_1^T V^{-1} - B D^{-1} Z_1^T V^{-1} \right) Y.\end{aligned}$$

Applying the *Sherman-Morrison-Woodbury* formula to the second term in equation 12 gives

$$\begin{aligned}A^{-1} B \left( D - B^T A^{-1} B \right)^{-1} &= A^{-1} \left( \mathbf{I} + B D^{-1} B^T (A - B D^{-1} B^T)^{-1} \right) B D^{-1} \\ &= (A - B D^{-1} B^T)^{-1} B D^{-1} \\ &= \mathrm{Cov}(\widehat{\beta_1}) B D^{-1}.\end{aligned}$$

Substituting this into equation 12 gives

$$\begin{aligned}\widehat{\beta_1} &= \mathrm{Cov}(\widehat{\beta_1}) \left( X_1^T V^{-1} Y - B D^{-1} Z_1^T V^{-1} Y \right) \\ &= \widehat{\beta_2}\end{aligned}$$

□

**Theorem B:**
Given the standard GLM, $Y = X\beta + \epsilon$, and a set of linearly independent contrasts specified by $C_1$ such that $\widehat{b_1} = C_1^T \widehat{\beta}$, then an equivalent model without contrasts, but with confounds, exists in the form

$$Y = [X_2 \;\; Z_2] \begin{bmatrix} b \\ \alpha \end{bmatrix} + \epsilon.$$

That is, $\widehat{b} = C_1^T \widehat{\beta}$, $\mathrm{Cov}(\widehat{b}) = C_1^T \mathrm{Cov}(\widehat{\beta}) C_1$ and the modelled signal space: $\mathrm{Span}(X_2) \cup \mathrm{Span}(Z_2) = \mathrm{Span}(X)$ in the pre-whitened space. Note that $\mathrm{Cov}(\epsilon) = V$ for both models.
**Proof:**
The proof is, again, by construction. Firstly, let $C_2$ be a set of contrasts that when combined with $C_1$ form a complete linearly independent set of contrasts. That is, the matrix $C = [C_1 \quad C_2]$ will be full rank (and hence invertible). Then let

$$X_2 = XQC_1F_1 \quad \text{and} \quad Z_2 = XQC_3F_3$$

where

$$Q = (X^T V^{-1} X)^{-1}, \quad F_1 = (C_1^T Q C_1)^{-1}, \quad C_3 = C_2 - P_{C_1} C_2, P_{C_1} = C_1 (C_1^T Q C_1)^{-1} C_1^T Q, \quad \text{and} \quad F_3 = (C_3^T Q C_3)^{-1}.$$

From these definitions it is easy to see that $C_1^T Q C_3 = 0$, which represents an orthogonality condition. As before, it is straightforward to verify that the combined span of $X_2$ and $Z_2$ is equal to the span of $X$. Consequently,

$$Z_2^T V^{-1} X_2 = F_3 C_3^T Q X^T V^{-1} X Q C_1 F_1 = F_3 C_3^T Q C_1 F_1 = 0.$$

Therefore, $Z_2$ and $X_2$ are orthogonal as well.
The estimation equations for the model become

$$\begin{aligned}\mathrm{Cov}\left( \begin{bmatrix} \widehat{b} \\ \widehat{\alpha} \end{bmatrix} \right) &= \begin{bmatrix} (X_2^T V^{-1} X_2)^{-1} & 0 \\ 0 & (Z_2^T V^{-1} Z_2)^{-1} \end{bmatrix}, \\ \begin{bmatrix} \widehat{b} \\ \widehat{\alpha} \end{bmatrix} &= \begin{bmatrix} (X_2^T V^{-1} X_2)^{-1} X_2^T V^{-1} \\ (Z_2^T V^{-1} Z_2)^{-1} Z_2^T V^{-1} \end{bmatrix} Y.\end{aligned}$$

Thus

$$\begin{aligned}\mathrm{Cov}(\widehat{b}) &= \left( F_1 C_1^T Q X^T V^{-1} X Q C_1 F_1 \right)^{-1} \\ &= \left( F_1 C_1^T Q C_1 F_1 \right)^{-1} \\ &= F_1^{-1} = C_1^T (X^T V^{-1} X)^{-1} C_1 \\ &= C_1^T \mathrm{Cov}(\widehat{\beta}) C_1\end{aligned}$$

and

$$
\begin{aligned}
\widehat{b} &= \mathrm{Cov}(\widehat{b})(F_1 C_1^T Q X^T) V^{-1} Y \\
&= C_1^T (Q X^T V^{-1}) Y \\
&= C_1^T \widehat{\beta}
\end{aligned}
\qquad \square
$$

# References


[1] T. Behrens, M.W. Woolrich, and S.M. Smith. Multi-subject null hypothesis testing using a fully Bayesian framework: Theory. In *Ninth Int. Conf. on Functional Mapping of the Human Brain*, 2003.

[2] E. Bullmore, M. Brammer, S.C.R. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham. Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, 35(2):261–277, 1996.

[3] A. Cnaan, N.M. Laird, and P. Slasor. Tutorial in Biostatistics: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, 16:2349–2380, 1997.

[4] B.S. Everitt. The analysis of Repeated Measures: A Practical Review with Examples. *Statistician*, 44(1):113–135, 1995.

[5] L. Frison and S.J. Pocock. Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in Medicine*, 11:1685–1704, 1992.

[6] K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner. Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage*, 16:465–483, 2002.

[7] FSL. `http://www.fmrib.ox.ac.uk/fsl`.

[8] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, second edition, 1989.

[9] A.P Holmes and K.J. Friston. Generalisability, random effects & population inference. In *Fourth Int. Conf. on Functional Mapping of the Human Brain, NeuroImage*, volume 7, page S754, 1998.

[10] P. Jezzard, P.M. Matthews, and S.M. Smith, editors. *Functional MRI: An Introduction to Methods*. OUP, Oxford, 2001.

[11] D.G. Leibovici. Negative variance problem in fMRI. In *Seventh Int. Conf. on Functional Mapping of the Human Brain*, 2001.

[12] J.L. Marchini and S.M. Smith. On bias in the estimation of autocorrelations for fMRI voxel time-series analysis. *NeuroImage*, 18(1):83–90, 2003.

[13] J.N.S. Matthews, D.G. Altman, M.J. Campbell, and P. Royston. Analysis of serial measurements in medical research. *British Medical Journal*, 300:230–235, 1990.

[14] P.L. Purdon, V. Solo, E. Brown, R. Buckner, M. Rotte, and R.M. Weisskoff. fMRI noise variability across subject and trials: Insights for noise estimation methods. In *Fourth Int. Conf. on Functional Mapping of the Human Brain, NeuroImage*, volume 7, page S617, 1998.

[15] F.E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bull.*, 2:110–114, 1946.

[16] S.R. Searle, G. Casella, and C.E. McCulloch. *Variance Components*. Wiley, 1992.

[17] M.W. Woolrich, B.D. Ripley, J.M. Brady, and S.M. Smith. Temporal autocorrelation in univariate linear modelling of FMRI data. *NeuroImage*, 14(6):1370–1386, 2001.

[18] K.J. Worsley and K.J. Friston. Analysis of fMRI time series revisited - again. *NeuroImage*, 2:173–181, 1995.

[19] K.J. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans. A general statistical analysis for fMRI data. *NeuroImage*, 15(1):1–15, 2002.