Statistical Models and Experimental Design

A.P. Holmes   K.J. Friston

Contents

1   Introduction

In the absence of prior anatomical hypotheses regarding the physical location of a particular function, the statistical analysis of functional mapping experiments must proceed by assessing the acquired data for evidence of an experimentally induced effect at every intracerebral voxel individually and simultaneously.

After reconstruction, realignment, stereotactic normalisation and (possibly) smoothing, the data are ready for statistical analysis. This involves two steps: Firstly, statistics indicating evidence against a null hypothesis of no effect at each voxel are computed. This results in an "image" of statistics. Secondly, this statistic image must be assessed, reliably locating voxels where an effect is exhibited whilst limiting the possibility of false positives. In this chapter we shall address the former topic, the formation of an appropriate statistic image with known properties.

Current methods for assessing the data at each voxel are predominantly parametric: Specific forms of probability distribution are assumed for the data, and hypotheses specified in terms of models assumed for the (unknown) parameters of these distributions. The parameters are estimated, and a statistic reflecting evidence against the null hypothesis formed. Statistics with known null distribution are used, such that the probability (under the null hypothesis) of obtaining a statistic as or more extreme than that observed can be computed. This is hypothesis testing in the classical parametric sense. The majority of the statistical models used are special cases of the General Linear Model.

SPM has become an acronym in common use for the theoretical framework of "Fristonian" voxel based analysis of functional mapping experiments, for the software package implementing these processes, and for a statistic image (*Statistical Parametric Map*). Here we shall take SPM to refer to the software package. Statistical parametric maps are given the acronym SPM$\{t\}$, where the letter indicates the marginal null distribution, in this case indicating that voxels have values distributed as a Student's t-distribution under the null hypothesis. The long form *Statistical Parametric Mapping* will be used in reference to the conceptual and theoretical framework. This chapter is solely concerned with Statistical Parametric Mapping. A review of other approaches is beyond the scope of these notes.

2   The General Linear Model

Before turning to the specifics of PET and fMRI, consider the general linear model. In what follows, an understanding of basic matrix algebra and the statistical concepts of hypothesis testing are required. Healy (1986) presents a brief summary of matrix methods relevant to statistics. The statistically naive are directed towards Mould's excellent "Introductory Medical Statistics" (Mould, 1989), while the more mathematically experienced will find Chatfield's "Statistics for Technology" (Chatfield, 1983) useful. Draper & Smith (1981) give a good exposition of matrix methods for the general linear model, and go on to describe regression analysis in general. The definitive tome for practical statistical experimental design is Winer *et al.* (1991).

2.1 The General Linear Model — Introduction

Suppose we are to conduct an experiment during which we will measure a *response variable* (such as rCBF at a particular voxel) Y_j, where $j = 1, \ldots, J$ indexes the observation. In statistical parlance, Y_j is a random variable, conventionally denoted by a capital letter.[1] Suppose also that for each observation we have a set of L $(L < J)$ *explanatory variables* (each measured without error) denoted by x_{jl}, where $l = 1, \ldots, L$ indexes the explanatory variables. The explanatory variables may be continuous (or sometimes discrete) *covariates*, functions of covariates, or they may be *dummy* variables indicating the *levels* of an experimental *factor*.

A *general linear model* explains the variation in Y_j in terms of a linear combination of the explanatory variables, plus an error term:

$$Y_j = x_{j1}\beta_1 + \ldots + x_{jl}\beta_l + \ldots + x_{jL}\beta_L + \epsilon_j \tag{1}$$

Here the β_l are (unknown) parameters, corresponding to each of the L explanatory variables. The model is assumed to fit such that the errors ϵ_j are independent and identically distributed normal random variables with zero mean and variance σ^2, written $\epsilon_j \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Linear models with other error distributions are *Generalised Linear Models*, for which the acronym GLM is usually reserved. The use of the general linear model is also widely known as *regression analysis*.

2.1.1 Examples, dummy variables

Many classical parametric statistical procedures are special cases of the general linear model.

Linear regression

The simplest example is linear regression, where only one continuous explanatory variable x_j is measured (without error) for each observation $j = 1, \ldots, J$. The model is usually written as:

$$Y_j = \mu + x_j\beta + \epsilon_j \tag{2}$$

where the unknown parameter μ, the *constant term* in the model, is the Y "intercept", the expected value of Y at $x = 0$. β is the (unknown) regression slope. Here, $\epsilon_j \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. This can be re-written as a general linear model by the use of a dummy explanatory variable, an indicator variable x_{1j} whose values are all one:

$$Y_j = x_{1j}\mu + x_{2j}\beta_2 + \epsilon_j \tag{3}$$

which is of the form of Eqn.1 on replacing β_1 with μ.

---

[1] We talk of *random variables*, and of observations prior to their measurement, because classical (frequentist) statistics is concerned with what could have occurred in an experiment. Once the observations have been made, they are known, the residuals are known, and there is no randomness.

Two-sample t-test

Similarly the the two-sample t-test is a special case of a general linear model: Suppose Y_{1j} and Y_{2j} are two independent groups of random variables: The two-sample t-test assumes $Y_{qj} \stackrel{iid}{\sim} \mathcal{N}(\mu_q, \sigma^2)$, for $q = 1, 2$, and assesses the null hypothesis $\mathcal{H} : \mu_1 = \mu_2$. The standard statistical way of writing the model is:

$$Y_{qj} = \mu_q + \epsilon_{qj} \tag{4}$$

The q subscript on the μ_q indicates that there are two *levels* to the group *effect*, μ_1 and μ_2. Here, $\epsilon_{qj} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$. This can be re-written using two dummy variables x_{1qj} and x_{2qj} as:

$$Y_{qj} = x_{1qj}\mu_1 + x_{2qj}\mu_2 + \epsilon_{qj} \tag{5}$$

which is of the form of Eqn.1 on re-indexing for qj. Here the dummy variables indicate group membership, x_{1qj} indicates whether observation Y_{qj} is from the first group, taking the value 1 when $q = 1$, and 0 when $q = 2$. Similarly, $x_{2qj} = \begin{cases} 0 & \text{if } q = 1 \\ 1 & \text{if } q = 2 \end{cases}$

2.2   Matrix formulation

The general linear model can be succinctly expressed using matrix notation. Consider writing out Eqn.1 in full, for each observation j, giving a set of simultaneous equations:

$$
\begin{aligned}
Y_1 &= x_{11}\beta_1 + \ldots + x_{1l}\beta_l + \ldots + x_{1L}\beta_L + \epsilon_1 \\
\vdots &= \vdots \\
Y_j &= x_{j1}\beta_1 + \ldots + x_{jl}\beta_l + \ldots + x_{jL}\beta_L + \epsilon_j \\
\vdots &= \vdots \\
Y_J &= x_{J1}\beta_1 + \ldots + x_{Jl}\beta_l + \ldots + x_{JL}\beta_L + \epsilon_J
\end{aligned}
$$

Which are equivalent to:

$$
\begin{pmatrix} Y_1 \\ \vdots \\ Y_j \\ \vdots \\ Y_J \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1l} & \cdots & x_{1L} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{j1} & \cdots & x_{jl} & \cdots & x_{jL} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{J1} & \cdots & x_{Jl} & \cdots & x_{JL} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_l \\ \vdots \\ \beta_L \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_l \\ \vdots \\ \epsilon_L \end{pmatrix}
$$

of the form $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ for $\underline{Y}$ the column vector of observations, $\underline{\epsilon}$ the column vector of error terms, and $\underline{\beta}$ the column vector of parameters; $\underline{\beta} = [\beta_1 \cdots \beta_l \cdots \beta_J]^\top$. The $J \times L$ matrix X, with jl^{th} element x_{jl} is the *design matrix*. This has one row per observation, and one column per model parameter.

2.3   Estimation

Once an experiment has been completed, we have observations of the random variables Y_j, which we denote by y_j. Assuming that no two observations are equal *and* have

identical explanatory variables, the number of parameters L is (usually) less than the number of observations J, and the simultaneous equations implied by the general linear model (with $\underline{\epsilon} = \underline{0}$) cannot be solved. Therefore, some method of estimating parameters that "best fit" the data is required. This is achieved by *least squares*.

Denote a set of parameter estimates by $\underline{\tilde{\beta}} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_L)^{\top}$. These parameters lead to *fitted values* $\underline{\tilde{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_J)^{\top} = X\underline{\tilde{\beta}}$, giving residual errors $\underline{e} = (e_1, \ldots, e_J)^{\top} = \underline{Y} - \underline{\tilde{Y}} = \underline{Y} - X\underline{\tilde{\beta}}$. The *residual sum-of-squares* $S = \sum_{j=1}^{J} e_j^2 = \underline{e}^{\top}\underline{e}$ is the sum of the square differences between the actual and fitted values, and thus measures the fit of the model with these parameter estimates.[2] The *least squares* estimates are the parameter estimates which minimise the residual sum-of-squares. In full:

$$S = \sum_{j=1}^{J} \left(Y_j - x_{j1}\tilde{\beta}_1 - \ldots - x_{jL}\tilde{\beta}_L\right)^2$$

This is minimised when:

$$\frac{\partial S}{\partial \tilde{\beta}_l} = 2\sum_{j=1}^{J}(-x_{jl})\left(Y_j - x_{j1}\tilde{\beta}_1 - \ldots - x_{jL}\tilde{\beta}_L\right) = 0$$

This equation is simply the jth row of $X^{\top}\underline{Y} = (X^{\top}X)\underline{\tilde{\beta}}$. Thus, the least squares estimates, denoted by $\underline{\hat{\beta}}$, satisfy the *normal equations*:

$$X^{\top}\underline{Y} = (X^{\top}X)\underline{\hat{\beta}} \tag{6}$$

The least squares estimates are the *maximum likelihood estimates* (if the model is correct and the errors are normal), and are the *Best Linear Unbiased Estimates*[3]. That is, of all linear parameter estimates consisting of linear combinations of the observed data whose expectation is the true value of the parameters, the least squares estimates have the minimum variance.

If $(X^{\top}X)$ is invertible, which it is if and only if the design matrix X is of full rank, then the least squares estimates are:

$$\underline{\hat{\beta}} = \left(X^{\top}X\right)^{-1} X^{\top}\underline{Y} \tag{7}$$

2.4 Overdetermined models

If X has linearly dependent columns, it is *rank deficient*, $(X^{\top}X)$ is singular, and has no inverse. In this case the model is overparameterised: There are infinitely many parameter sets describing the same model. Correspondingly, there are infinitely many least squares estimates $\underline{\hat{\beta}}$ satisfying the normal equations.

2.4.1 One way ANOVA Example

The simplest example of such a model is the classic Q group one way analysis of variance (ANOVA) model:

$$Y_{qj} = \mu + \alpha_q + \epsilon_{qj} \tag{8}$$

---

[2] $\underline{e}^{\top}\underline{e}$ is the L_2 norm of $\underline{e}$
[3] Gauss-Markov theorem

where Y_{qj} is the jth observation in group $q = 1, \ldots, Q$. This model clearly doesn't uniquely specify the parameters: For any given μ and α_q, the parameters $\mu' = \mu + d$ and $\alpha'_q = \alpha_q \Leftrightarrow d$ give an equivalent model for any constant d. That is, the model is indeterminate up to the level of an additive constant between the constant term μ and the group effects α_q. Similarly for any set of least squares estimates $\hat{\mu}, \hat{\alpha}_q$. Here there is one degree of indeterminancy in the model, and the design matrix has rank Q, one less than the number of parameters (the number of columns of X). If the data vector $\underline{Y}$ has observations arranged by group, then for three groups ($Q = 3$), the design matrix and parameter vectors are

$$
X = \begin{bmatrix}
1 & 1 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
\vdots & \vdots & \vdots & \vdots \\
1 & 0 & 0 & 1
\end{bmatrix}
\qquad
\underline{\beta} = \begin{bmatrix}
\mu \\
\alpha_1 \\
\alpha_2 \\
\alpha_3
\end{bmatrix}
$$

Clearly this matrix is rank deficient: The first column, corresponding to the constant term μ, is the sum of the others, corresponding to the group effects α_q.

2.4.2 Parameter estimates

A set of least squares estimates may be found by imposing constraints on the estimates, or by using a pseudoinverse technique for $(X^\top X)$, which essentially implies a constraint. In either case it is important to remember that the actual estimates obtained depend on the particular constraint or pseudoinverse method chosen. This has implications for inference (§2.6): It is only meaningful to consider functions of the parameters that are uninfluenced by the particular constraint chosen.

2.4.3 Parameters for one way ANOVA example

(This section can be omitted without loss of continuity.)

Consider the one way ANOVA example. The simplest possible constraint is a *corner point* constraint, setting a redundant parameter to zero. Here a suitable constraint is $\hat{\mu} = 0$, effected by eliminating the first column of the design matrix, giving a design matrix of full rank. Then the least squares estimates are $\hat{\mu} = 0, \hat{\alpha}_q = \overline{Y}_{q\bullet}$, the sample mean of the group q observations.[4] This constraint interprets the group effects α_q as the group means, and the model could be succinctly written $Y_{qj} \sim \mathcal{N}(\alpha_q, \sigma^2)$.

Another possible constraint is a "sum-to-zero" constraint on the fitted group effects: $\sum_{q=1}^{Q} \hat{\alpha}_q = 0$. This constraint implies that any one group's fitted effect is minus the sum of the others, so the constraint can be effected by expressing the effect for (say) the last

---

[4]The standard statistical "bar and bullet" notation denotes that the mean has been taken over the values of the subscripts replaces by bullets.

group in terms of the others, and eliminating it from the design. (After fitting the model, the eliminated parameter can be obtained by substituting the parameter estimates into the constraint.) For three groups this gives $\hat{\alpha}_3 = -\hat{\alpha}_1 - \hat{\alpha}_2$, resulting in design matrix X and parameter vector $\underline{\beta}$ as:

$$
X = \begin{bmatrix}
1 & 1 & 0 \\
\vdots & \vdots & \vdots \\
1 & 1 & 0 \\
1 & 0 & 1 \\
\vdots & \vdots & \vdots \\
1 & 0 & 1 \\
1 & -1 & -1 \\
\vdots & \vdots & \vdots \\
1 & -1 & -1
\end{bmatrix}
\qquad
\underline{\beta} = \begin{bmatrix}
\mu \\
\alpha_1 \\
\alpha_2
\end{bmatrix}
$$

In this case $\hat{\mu} = \sum_{j=1}^{Q} (\overline{Y}_{q\bullet})/Q$, the mean of the group sample means (equal to the overall sample mean $\overline{Y}_{\bullet\bullet}$ if all groups have equal size). The group effects are estimated by $\hat{\alpha}_q = \overline{Y}_{q\bullet} - \hat{\mu}$, the mean of the group q observations after the overall effect has been subtracted, interpreted as the deviations for each group about the common response.

Alternatively a pseudoinverse method can be used, using $\mathrm{pinv}(X^{\top}X)$ in place of $(X^{\top}X)^{-1}$ in Eqn.7: A set of least squares estimates are then given by $\hat{\underline{\beta}} = \mathrm{pinv}(X^{\top}X)X^{\top}\underline{Y} = \mathrm{pinv}(X)\underline{Y}$. The pseudoinverse function implemented in MATLAB gives the Moore-Penrose pseudoinverse.[5] This results in the least squares parameter estimates with the minimum sum-of-squares (minimum L_2 norm $\|\hat{\underline{\beta}}\|_2$). For the one way ANOVA model, this can be shown to give parameter estimates $\hat{\mu} = \sum_{j=1}^{Q}(\overline{Y}_{q\bullet})/(1+Q)$ and $\hat{\alpha}_q = \overline{Y}_{q\bullet} - \hat{\mu}$, a rather strange constraint!

By cunning construction of the design matrix, the Moore-Penrose pseudoinverse can be utilised to impose particular constraints. For example, reparameterising the one way ANOVA model (Eqn.8) by expressing the group effects relative to their mean results in the imposition of a sum-to-zero constraint on the group effects. The model is then:

$$
Y_{qj} = \mu + (\alpha_q - \overline{\alpha}_q) + \epsilon_{qj} \tag{9}
$$

with design matrix (for three groups)

$$
X = \begin{bmatrix}
1 & 1-1/3 & -1/3 & -1/3 \\
\vdots & \vdots & \vdots & \vdots \\
1 & 1-1/3 & -1/3 & -1/3 \\
1 & -1/3 & 1-1/3 & -1/3 \\
\vdots & \vdots & \vdots & \vdots \\
1 & -1/3 & 1-1/3 & -1/3 \\
1 & -1/3 & -1/3 & 1-1/3 \\
\vdots & \vdots & \vdots & \vdots \\
1 & -1/3 & -1/3 & 1-1/3
\end{bmatrix}
\qquad
\underline{\beta} = \begin{bmatrix}
\mu \\
\alpha_1 \\
\alpha_2 \\
\alpha_3
\end{bmatrix}
$$

---

[5] If X is of full rank, then $\mathrm{pinv}(X^{\top}X)$ is an inefficient way of computing $(X^{\top}X)^{-1}$.

The least squares estimates using the Moore-Penrose pseudoinverse are then those of the sum-to-zero constraints considered above. The point to note here is that the design matrix still has a column for each effect, and has the same structure as the design matrix for the "plain" model. Thus, this implicit constraining is useful if one wishes to impose a particular constraint, for interpretation of the parameter estimates, but want to maintain the form of the design matrix for computational and visualisation reasons. Implicit sum-to-zero constraints are used in SPM, to allow plotting of the parameter estimates, and to give sensible adjusted values (§6.2).

2.5 Geometrical Perspective

For some, a geometrical perspective aids conceptualisation. (This section can be omitted without loss of continuity.)

The vector of observed values $\underline{Y}$ defines a single point in $\Re^J$, J-dimensional Euclidean space. $X\underline{\tilde{\beta}}$ is a linear combination of the columns of the design matrix X. The columns of X are J-vectors, so $X\tilde{\beta}$ for given $\tilde{\beta}$ defines a point in $\Re^J$. This point lies in the subspace of $\Re^J$ spanned by the columns of the design matrix, the X-space. The dimension of this subspace is $\text{rank}(X)$. (Recall that the space spanned by the columns of X is the set of points $X\underline{c}$ for all $\underline{c} \in \Re^L$.) The residual sum-of-squares for parameter estimates $\tilde{\beta}$ is the distance from $X\underline{\tilde{Y}}$ to $\underline{Y}$. Thus, the least squares estimates correspond to the point in the space spanned by the columns of X that is nearest to the data $\underline{Y}$. The perpendicular from $\underline{Y}$ to the X-space meets the X-space at $\underline{\hat{Y}} = X\hat{\beta}$. It is now clear why there are no unique least squares estimates if X is rank-deficient; for then any point in the X-space can be obtained by infinitely many linear combinations of the columns of X.

If X is of full rank, then define the *hat* matrix as $H = X\left(X^\top X\right)^{-1} X^\top$. Then $\underline{\hat{Y}} = H\underline{Y}$, and geometrically H is a projection onto the X-space. Similarly, the residual forming matrix is $R = (I \Leftrightarrow H)$ for I the identity matrix. Thus $R\underline{Y} = \underline{e}$, and R is a projection onto the space orthogonal to the X-plane.

As a concrete example, consider a simple linear regression with three observations. The observed data $\underline{Y} = (Y_1, Y_2, Y_3)^\top$ defines a point in three-dimensional Euclidean space ($\Re^3$). The model (Eqn.2) leads to a design matrix $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix}$ Provided the x_j's aren't all the same, the columns of X span a two dimensional subspace of $\Re^3$, a plane. (Fig.1).

2.6 Inference

2.6.1 Residual Sum of Squares

The residual variance σ^2 is estimated by the *residual mean square*, the residual sum-of-squares divided by the appropriate degrees of freedom: $\hat{\sigma}^2 = \frac{\underline{e}^\top \underline{e}}{J-p} \sim \sigma^2 \frac{\chi^2_{J-p}}{J-p}$ where $p = \text{rank}(X)$.
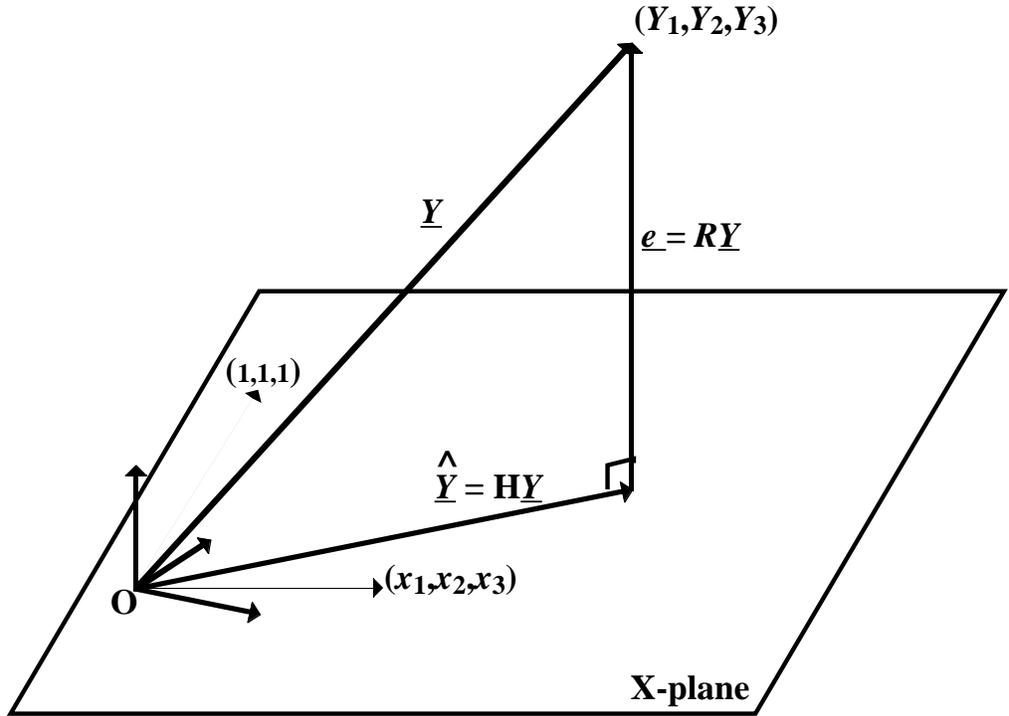
Figure 1: Geometrical perspective on 3 point linear regression

2.6.2 Linear Combinations of the Parameter Estimates

It is not too difficult to show that the parameter estimates are normally distributed: For X of full rank $\hat{\underline{\beta}} \sim \mathcal{N}\left(\underline{\beta}, \sigma^2 (X^\top X)^{-1}\right)$. From this it follows that for $\underline{c}$ a column vector of L weights:

$$\underline{c}^\top \hat{\underline{\beta}} \sim \mathcal{N}\left(\underline{c}^\top \underline{\beta}, \ \sigma^2 \underline{c}^\top (X^\top X)^{-1} \underline{c}\right) \tag{10}$$

Furthermore, $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are independent (Fisher's Law). Thus, prespecified hypotheses concerning linear compounds of the model parameters $\underline{c}^\top \underline{\beta}$ can be assessed using

$$\frac{\underline{c}^\top \hat{\underline{\beta}} \Leftrightarrow \underline{c}^\top \underline{\beta}}{\sqrt{\hat{\sigma}^2 \underline{c}^\top (X^\top X)^{-1} \underline{c}}} \sim t_{J-p} \tag{11}$$

That is, the hypothesis $\mathcal{H} : \underline{c}^\top \underline{\beta} = d$ can be assessed by comparing $T = \dfrac{\underline{c}^\top \hat{\underline{\beta}} - d}{\sqrt{\hat{\sigma}^2 \underline{c}^\top (X^\top X)^{-1} \underline{c}}}$ with a Student's t-distribution with $J \Leftrightarrow p$ degrees of freedom.

Example — Two-sample t-test

For example, consider the two-sample t-test (§2.1.1), The model (Eqn.4) leads to a design matrix X with two columns of dummy variables indicating group membership, and parameter vector $\underline{\beta} = (\mu_1, \mu_2)^\top$. Thus, the null hypothesis $\mathcal{H} : \mu_1 = \mu_2$ is equivalent to $\mathcal{H} : \underline{c}^\top \underline{\beta} = 0$ with $\underline{c} = (1, \Leftrightarrow 1)^\top$. The first column of the design matrix contains n_1 1's and

n_2 0's, indicating the group one measurements, the second column n_1 0's and n_2 1's. Thus $(X^\top X) = \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}$, $(X^\top X)^{-1} = \begin{pmatrix} 1/n_1 & 0 \\ 0 & 1/n_2 \end{pmatrix}$, and $\underline{c}^\top (X^\top X)^{-1}\underline{c} = 1/n_1 + 1/n_2$, giving t-statistic (by Eqn.11):

$$ T = \frac{\mu_1 \Leftrightarrow \mu_2}{\sqrt{\hat{\sigma}^2 \left(1/n_1 + 1/n_2\right)}} $$

...the standard formula for the two-sample t-statistic, with a Student's t-distribution of $n_1 + n_2 \Leftrightarrow 2$ degrees of freedom under the null hypothesis.

2.6.3 Estimable functions, contrasts

Recall (§2.4) that if the model is overparameterised (i.e. X is rank deficient), then there are infinitely many parameter sets describing the same model. Constraints or the use of a pseudoinverse technique pull out only one set of parameters from infinitely many. Therefore, when examining linear compounds $\underline{c}^\top \underline{\beta}$ of the parameters it is imperative to consider only compounds that are invariant over the space of possible parameters. Such linear compounds are called *contrasts*.

In detail (Scheffé, 1959), a linear function $\underline{c}^\top \underline{\beta}$ of the parameters is *estimable* if there is a linear unbiased estimate $\underline{c}'^\top \underline{Y}$ for some constant vector of weights $\underline{c}'$. That is $\underline{c}^\top \underline{\beta} = \mathsf{E}\left[\underline{c}'^\top \underline{Y}\right]$, so for an estimable function the natural estimate $\underline{c}^\top \hat{\underline{\beta}}$ is unique whatever solution $\hat{\underline{\beta}}$ of the normal equations is chosen (Gauss-Markov theorem). Further: $\underline{c}^\top \underline{\beta} = \mathsf{E}\left[\underline{c}'^\top \underline{Y}\right] = \underline{c}'^\top X \underline{\beta} \Rightarrow \underline{c} = \underline{c}'^\top X$, so $\underline{c}$ is a linear combination of the rows of X.

A *contrast* is an estimable function with the additional property $\underline{c}^\top \hat{\underline{\beta}} = \underline{c}'^\top \hat{\underline{Y}} = \underline{c}'^\top \underline{Y}$. Now $\underline{c}'^\top \hat{\underline{Y}} = \underline{c}'^\top \underline{Y} \Leftrightarrow \underline{c}'^\top H \underline{Y} = \underline{c}'^\top \underline{Y} \Leftrightarrow \underline{c}' = H\underline{c}'$ (since H is symmetric), so $\underline{c}'$ is in the X-space. Thus a contrast is an estimable function whose $\underline{c}'$ vector is a linear combination of the columns of X. Thus for a contrast it can be shown that $\underline{c}^\top \hat{\underline{\beta}} \sim \mathcal{N}\left(\underline{c}^\top \underline{\beta}, \sigma^2 \underline{c}'^\top \underline{c}'\right)$. Using a pseudoinverse technique, $H = X\mathrm{pinv}(X^\top X)X^\top$, so $\underline{c}' = H\underline{c}' \Rightarrow \underline{c}'^\top \underline{c}' = \underline{c}'^\top X\mathrm{pinv}(X^\top X)X^\top \underline{c}' = \underline{c}^\top \mathrm{pinv}(X^\top X)\underline{c}$ since $\underline{c} = \underline{c}'^\top X$ for an estimable function.

The above shows that the distributional results given above for unique designs (Eqn.10 & Eqn.11), apply for contrasts of the parameters of non-unique designs, where $(X^\top X)^{-1}$ is replaced by a pseudoinverse.

It remains to characterise which linear compounds of the parameters are contrasts. For most designs, contrasts have weights that sum to zero over the levels of each factor. For the one way ANOVA with parameter vector $\underline{\beta} = (\mu, \alpha_1, \ldots, \alpha_Q)^\top$, the linear compound $\underline{c}^\top \underline{\beta}$ with weights vector $\underline{c} = (c_0, c_1, \ldots, c_Q)$ is a contrast if $c_0 = 0$ and $\sum_{q=1}^{Q} c_q = 0$. Given that the indeterminacy in condition effects in this model is up to the level of an additive constant, the form of the contrast is intuitively correct. Contrasts of this form include no contribution from a constant added to all the condition effects. Other models and the form of appropriate contrasts are discussed in the next section (§3).

2.6.4 Extra Sum of Squares Principle

The *Extra sum-of-squares* principle provides a method of assessing general linear hypotheses, and for comparing models in a hierarchy.

Suppose we have a model with parameter vector $\underline{\beta}$ that can be partitioned into two, $\underline{\beta} = \left[\underline{\beta}_1^\mathsf{T} \vdots \underline{\beta}_2^\mathsf{T}\right]$, and suppose we wish to test $\mathcal{H} : \underline{\beta}_1 = \underline{0}$. The corresponding partitioning of the design matrix X is $X = \left[X_1 \vdots X_2\right]$, and the *full model* is:

$$\underline{Y} = \left[X_1 \vdots X_2\right] \begin{bmatrix} \underline{\beta}_1 \\ \cdots \\ \underline{\beta}_2 \end{bmatrix} + \underline{\epsilon}$$

which when $\mathcal{H}$ is true reduces to the *reduced model*: $\underline{Y} = X_2\underline{\beta}_2 + \underline{\epsilon}$. Denote the residual sum-of-squares for the full and reduced models by $S(\underline{\beta})$ and $S(\underline{\beta}_2)$ respectively. The *extra sum-of-squares* due to $\underline{\beta}_1$ after $\underline{\beta}_2$ is then defined as $S(\underline{\beta}_1|\underline{\beta}_2) = S(\underline{\beta}_2) \Leftrightarrow S(\underline{\beta})$. Under $\mathcal{H}$, $S(\underline{\beta}_1|\underline{\beta}_2) \sim \sigma^2 \chi_p^2$ independently of $S(\underline{\beta})$, where the degrees of freedom are $p = \text{rank}(X) \Leftrightarrow \text{rank}(X_2)$. (If $\mathcal{H}$ is not true, then $S(\underline{\beta}_1|\underline{\beta}_2)$ has a non-central chi-squared distribution, still independent of $S(\underline{\beta})$.) Therefore, the following F-statistic expresses evidence against $\mathcal{H}$:

$$F = \frac{\frac{S(\underline{\beta}_2) - S(\underline{\beta})}{p - p_2}}{\frac{S(\underline{\beta})}{J - p}} \sim F_{p - p_2, J - p} \tag{12}$$

where $p = \text{rank}(X)$ and $p_2 = \text{rank}(X_2)$. Significance can then be assessed by comparing this statistic with the appropriate F-distribution. Draper & Smith (1981) give derivations.

Example — One way ANOVA

For example, consider a one way ANOVA (§2.4.1, Eqn.8), where we wish to assess the omnibus null hypothesis that all the groups are identical: $\mathcal{H} : \alpha_1 = \alpha_2 = \ldots = \alpha_Q$. Under $\mathcal{H}$ the model reduces to $Y_{qj} = \mu + \epsilon_{qj}$. Since the ANOVA model contains a constant term, μ, *mathcal H* is equivalent to $\mathcal{H} : \alpha_1 = \alpha_2 = \ldots = \alpha_Q = 0$. Thus, let $\underline{\beta}_1 = (\alpha_1, \ldots, \alpha_Q)^\mathsf{T}$, and $\underline{\beta}_2 = \mu$. Eqn.12 then gives an F-statistic which is precisely the standard F-statistic for a one way ANOVA.

3   Modelling for PET

With the details of the General linear model covered, we turn our attention to some actual models used in functional brain mapping, discuss the practicalities of their application, and introduce the terminology of SPM. As the approach is massively univariate, we must consider a model for each and every voxel. Bear in mind that in the massively univariate approach of SPM, the same model form is used at every voxel simultaneously, with different parameters for each voxel. This fact leads to the notion of *image regression* and a multivariate perspective on the massively univariate approach, which is covered in section 5. Model selection issues are raised in section 6.5.

In this section, we shall concentrate on PET, with its mature family of standard statistical experimental designs. fMRI requires extensions to the general linear model theory, presented in the following section (§4), and special design considerations (discussed in a separate chapter).

Although most PET functional mapping experiments are on multiple subjects, many of the key concepts are readily demonstrated within the framework of a single subject experiment. Hence, for simplicity of exposition, the initial emphasis is on single subject experiments.

3.1   Global changes

Global cerebral blood flow (gCBF) varies, both between subjects, and over time in a single individual. If qualitative "count" measurements of relative activity (rA) are being used as an indicator of rCBF, then changes in the global activity gA reflect changes in the administered dose and head fraction, as well as changes in gCBF. Therefore, changes in regional cerebral blood flow (or rA) measurements across experimental conditions (measured at different times, and possibly on different subjects), are confounded by global changes. In the remainder of this chapter, we shall refer to regional cerebral blood flow (rCBF), regarding rA as an indicator of rCBF unless the discussion calls for a distinction.

Consider a simple single subject activation experiment, where a single subject is scanned repeatedly under both *baseline* (control) and *activation* (experimental) conditions. Inspection of rCBF alone at a single voxel may not indicate an experimentally induced effect, whereas the additional consideration of the gCBF for the respective scans may clearly differentiate between the two conditions. (Fig.2)

3.1.1   Measurement of gCBF

For normal subjects, the global cerebral blood flow is adequately measured as the mean rCBF over all intracerebral voxels. If Y_j^k is the rCBF at voxel $k = 1, \ldots, K$ of scan j, then denote the gCBF by $g_j = \overline{Y}_j^\bullet = \sum_{k=1}^K Y_j^k / K$.
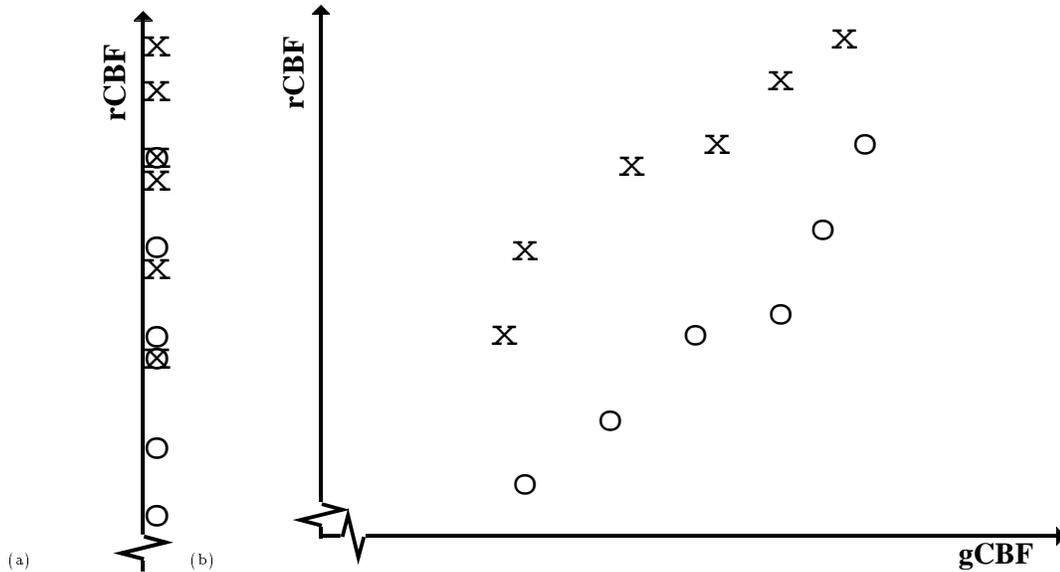
Figure 2: Single subject activation experiment, illustrative plots of rCBF at a single voxel: (a) Dot-plots of rCBF (b) Plot of rCBF vs. gCBF. Both plots indexed by condition: o for baseline, × for active.

3.1.2 Grand Mean Scaling

For qualitative counts data, it is common to scale the raw data into a physiological range, exploiting the near linearity of the quantification function to give parameters in an interpretable scale. This is usually achieved by scaling all images in a data set so that the *grand mean*, the mean global value, is a set value, usually chosen to be the canonical normal gCBF of 50ml/min/dl. The scaling factor is thus $\frac{50}{g_\bullet}$. Since the scaling factor is a constant (for each data set), this scaling doesn't change the relative values of the data, and therefore doesn't affect any subsequent statistical analysis. Henceforth, we shall assume that counts data have been scaled into a physiologically meaningful scale.

3.1.3 Proportional scaling gCBF normalisation

Conceptually the simplest way to account for global changes is to adjust the data by scaling all scans to have the some target gCBF, usually chosen to be the canonical normal gCBF of 50ml/min/dl. This gives *adjusted* rCBF at voxel k of $Y_j'^k = Y_j^k/(g_j/50)$. The adjusted data are then used as raw data for analysis.[6] On a graph of rCBF against gCBF, the adjusted values are the point where the line joining the observed rCBF to the origin intercepts gCBF $= 50$ (Fig.3a).

A simple two-condition single subject activation study would then be assessed using a

---

[6] Clearly grand mean scaling is redundant when followed by proportional scaling gCBF normalisation. So, within SPM, the target global value used is specified under "grand mean scaling", and defaults to 50.

two-sample t-statistic on the adjusted rCBF at each voxel. Let Y_{qj}^k denote the rCBF at voxel k of scan $j = 1, \ldots, M$ under condition $q = 1, 2$, and $Y_{qj}^{\prime k}$ the adjusted rCBF. The model is then (from Eqn.4): $Y_{qj}^{\prime k} = \alpha_q^k + \epsilon_{qj}^k$, where $\epsilon_{qj}^k \overset{iid}{\sim} \mathcal{N}(0, \sigma_k^2)$. Multiplying through by $g_j/50$ we arrive at:

$$Y_{qj}^k = \frac{g_j}{50} \times \alpha_q^k + \epsilon_{qj}^{\prime k} \tag{13}$$

This is a proportional regression (Fig.3b). Activation corresponds to a change in slope. However, this is a *weighted* regression, since the residual variance is weighted by the global flow: $\epsilon_{qj}^{\prime k} \overset{iid}{\sim} \mathcal{N}(0, (g_j/50)^2 \times \sigma_k^2)$. Proportional scaling isn't as simple as it first appears!

3.1.4 Modeling the effect of global changes

A more rigorous approach is to explicitly model the effect of global changes. Friston *et al.* (1990) proposed that for normal ranges of cerebral blood flow the relationship between regional and global flow would be well approximated by a straight line. For repeat scans of an individual under exactly the same conditions, the model is a simple regression at each voxel:

$$Y_j^k = \mu^k + \zeta^k (g_j \Leftrightarrow \overline{g}_\bullet) + \epsilon_j^k \tag{14}$$

where we assume normality of the errors, $\epsilon_q^k \overset{iid}{\sim} \mathcal{N}(0, \sigma_k^2)$.

The error variance σ_k^2 is allowed to vary between voxels. There is substantial evidence against an assumption of constant variance (homoscedasticity) at all points of the brain. This fact is perhaps to be expected, considering the different constituents and activities of grey and white matter, which is unfortunate, as the small sample sizes leave few degrees of freedom for variance estimation. If homoscedasticity can be assumed, variance estimates can legitimately be pooled across all voxels. Provided the image is much greater in extent than it's smoothness, this gives an estimate with sufficiently high (effective) degrees of freedom that it's variability is negligible. (Since the images are smooth, neighbouring voxels have correlated values, and hence the variance estimates at neighbouring voxels are correlated.) t-Statistics based on such a variance estimate are approximately normally distributed, the approximation failing only in the extreme tails of the distribution. This approach was initially endorsed by Worsley *et al.* (1992), using low resolution PET images which appeared homoscedastic, but is inappropriate for most applications.

If the (mean corrected) global values are considered as confounding covariates, adjusted values at voxel k could be constructed as $Y_j^{\prime k} = Y_j^k \Leftrightarrow \hat{\zeta}^k (g_j \Leftrightarrow \overline{g}_\bullet)$. However, analyses on such adjusted values don't take into account the degrees of freedom lost due to the adjustment, and the approach is not equivalent to the modeling approach which explicitly includes gCBF as a covariate.

3.1.5 ANCOVA

Consider the simple single subject activation experiment, and suppose an activation experimental condition induces a constant increase of rCBF over the baseline condition regardless of gCBF. Then, the activation condition can be added to the model as a simple
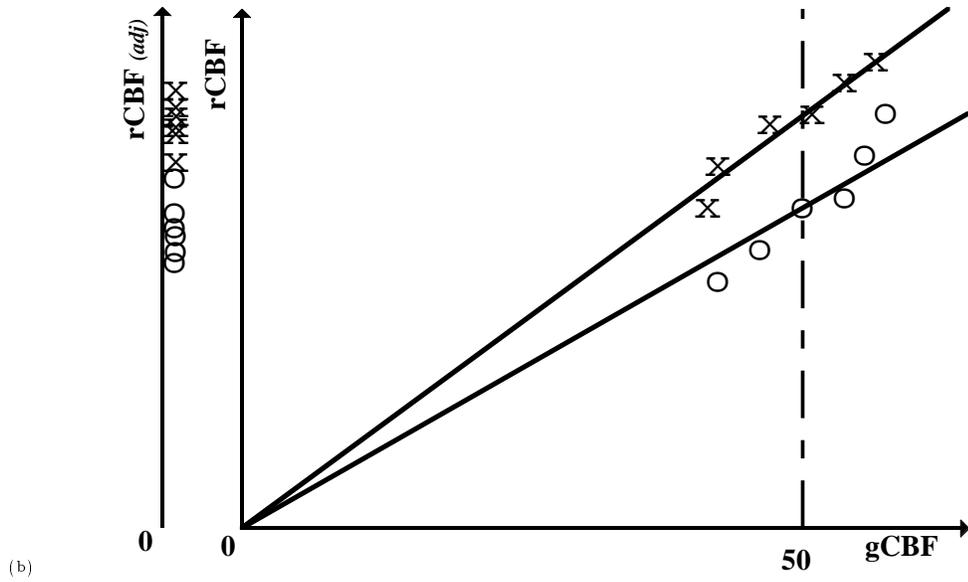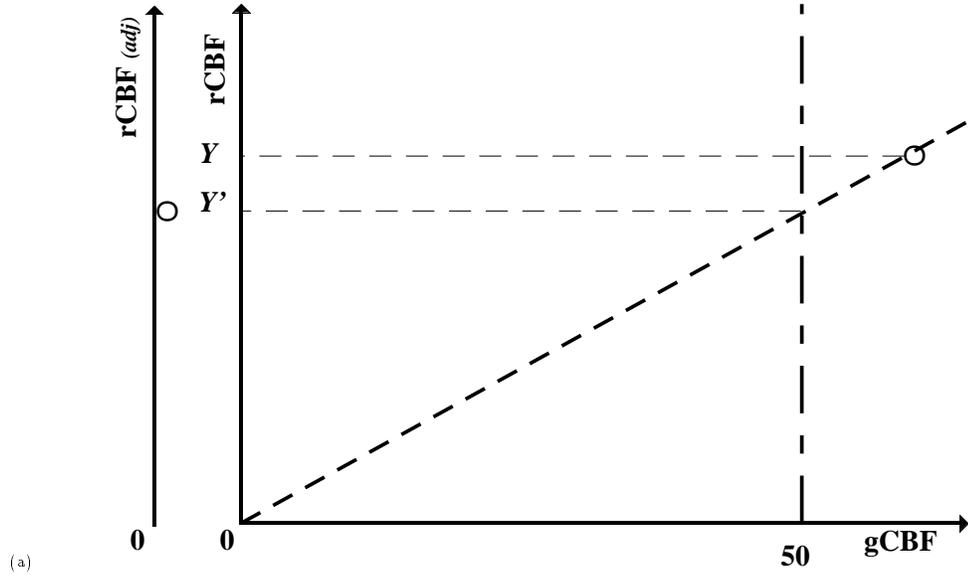
Figure 3: (a) Adjustment by proportional scaling (b) Simple single subject activation as a t-test on adjusted rCBF: Weighted proportional regression
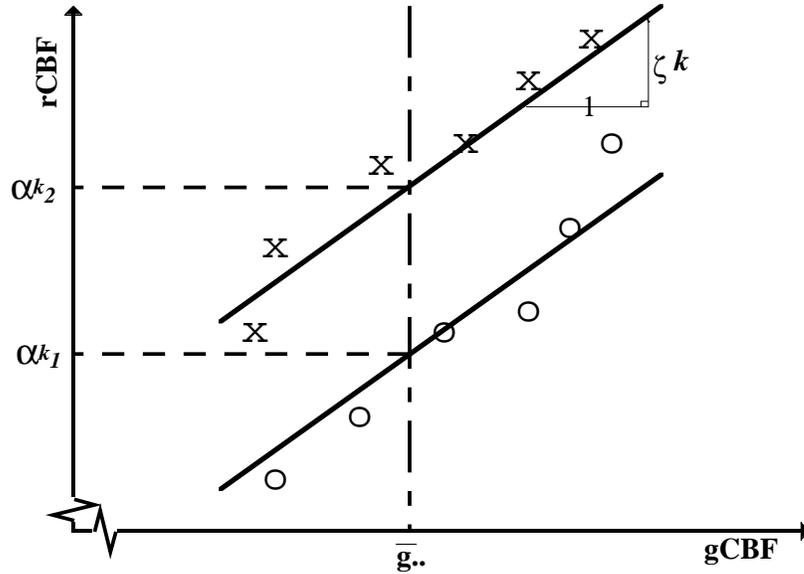
Figure 4: Single subject activation experiment, illustrative plot of AnCova model (Eqn.15) for rCBF vs gCBF.

additive effect, giving a classic *Analysis of Covariance*[7] (AnCova) model at each voxel:

$$Y_{qj}^k = \alpha_q^k + \zeta^k(g_{qj} \Leftrightarrow \overline{g}_\bullet) + \epsilon_{qj}^k \tag{15}$$

where $\epsilon_{qj}^k \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_k^2)$. Activation corresponds to a change in intercept, as depicted in figure 4. Note that the activation is assumed independent of the underlying gCBF; the regression lines for each condition have the same slope ζ^k. This is the AnCova parallel lines assumption.

3.1.6   Scaling verses modeling

Clearly a choice must be made; scaling or AnCova. Many authors have debated this issue, yet still no consensus exists.

Due to the weighting of the variance in the proportional regression model, the "extra sum-of-squares" F-test is inappropriate for comparing a proportional regression with a more general linear regression for repeat scans on a single individual under identical conditions (i.e. comparing the model of Eqn.13 for one condition with that of Eqn.14).

For normal subjects under normal conditions the range of gCBF exhibited is small, and located far from zero. Therefore for quantitative rCBF data on normal subjects under normal conditions, a simple regression model affords better modeling of the true relationship between regional and global flows than a proportional regression.

---

[7]AnCova is the standard way to account for an (independent) confounding covariate into a two sample problem.

If qualitative "count" measurements are being used as indicators of rCBF, then gA can vary considerably even if the actual gCBF is fairly stable. Possible reasons for this are differences in introduced dose, scanner sensitivity, subject head fraction, or similar. Clearly, for constant gCBF the relationship between regional and global activity is proportional. More importantly, the greater the introduced dose, the greater the variance in the regional activity, since this is Poisson in nature, again suggesting the appropriateness of the proportional model since in this case proportional scaling also stabilises the variance. The variance stabilising properties of the proportional scaling approach is the only substantial reason for not using ANCOVA type modeling approaches.

For a single subject activation study, provided the introduced dose is constant, the ANCOVA model is preferred over proportional scaling and two sample t-test, even for qualitative counts data. There is little evidence against the parallel lines assumption of an additive activation effect for cognitive paradigms within the normal range of gCBF. For multi-subject experiments on the same scanner using the same protocol and dosage, current experience is that the ANCOVA style approach is adequate. Recent authors have found little empirical difference between parallel analyses using proportional scaling and ANCOVA on the same data.

Special considerations apply if there are condition dependent changes in gCBF.

3.1.7   Global differences

Implicit in allowing for changes in gCBF (either by proportional scaling or ANCOVA) when assessing condition specific changes in rCBF, is the assumption that gCBF represents the underlying background flow, above which regional differences are assessed. That is, gCBF is independent of condition. Clearly, since gCBF is calculated as the mean intracerebral rCBF, an increase of rCBF in a particular brain region must cause an increase of gCBF unless there is a corresponding decrease of rCBF elsewhere in the brain. Similar problems can arise when comparing a group of subjects with a group of patients with brain atrophy, or when comparing pre and post-operative rCBF.

If gCBF actually varies considerably between conditions, as in pharmacological activation studies, then testing for a condition effect after allowing for global changes involves extrapolating the relationship between regional and global flow outside the range of the data. This extrapolation might not be valid, as illustrated in figure 5a.

If gCBF is increased by a large activation that is not associated with a corresponding deactivation, then comparison at a common gCBF will make non-activated regions (whose rCBF remained constant) appear falsely de-activated, and the magnitude of the activation will be similarly decreased. (Figure 5b illustrates the scenario for a simple single subject activation experiment using ANCOVA.) In such circumstances a better measure of the underlying background flow should be sought, for instance by examining the flow in brain regions known to be unaffected by the stimulus.
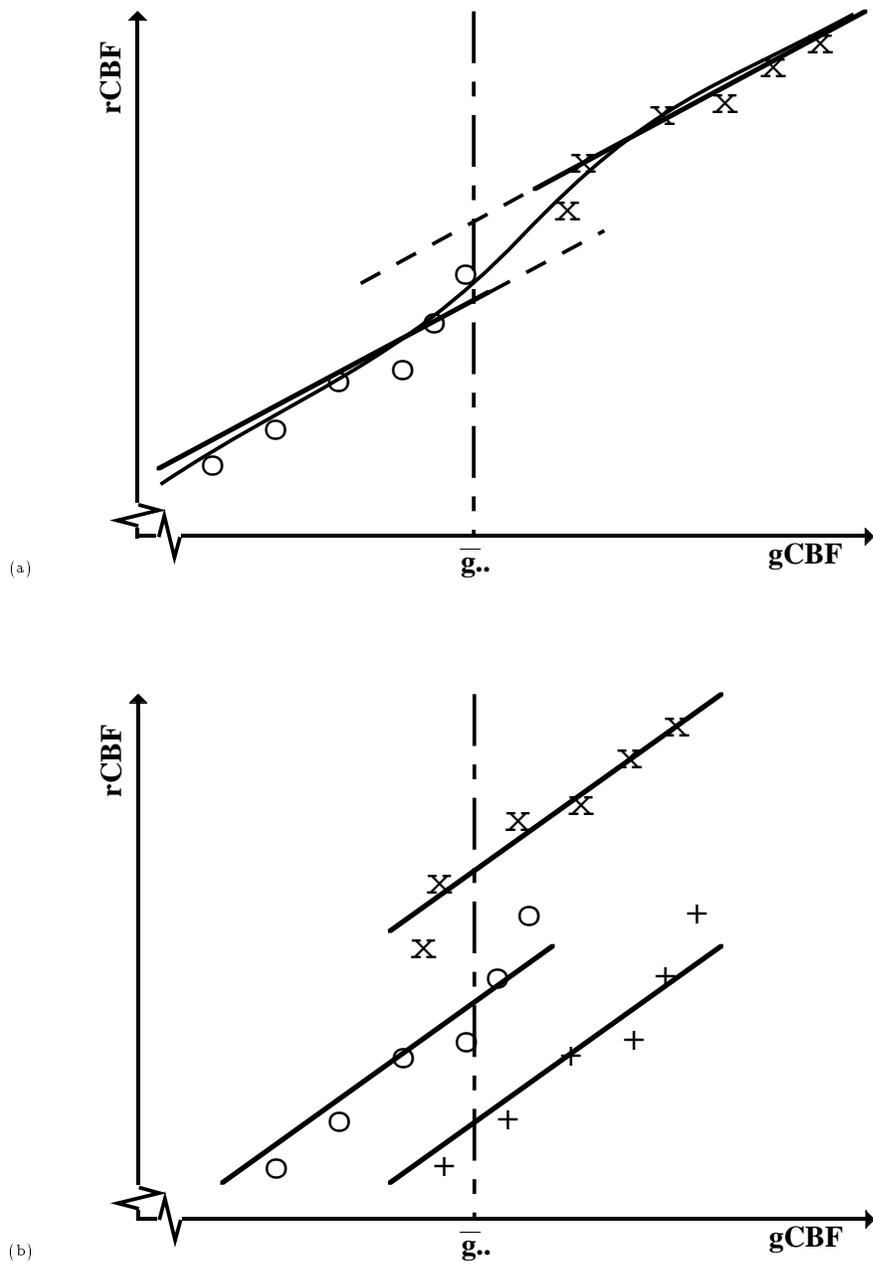
Figure 5: Single subject activation experiment, illustrative (ANCOVA) plots of rCBF vs. gCBF at a single voxel showing potential problems with global changes: (a) Large change in gCBF between conditions. The apparent activation relies on linear extrapolation of the baseline and active condition regressions (assumed to have the same slope) beyond the range of the data. The actual relationship between regional and global for no activation may be given by the curve, in which case there is no activation effect. (b) Large activation inducing increase in gCBF measured as brain mean rCBF. Symbol o denotes rest, × denotes active condition values if this is a truly activated voxel (in which case the activation is underestimated), while + denotes active condition values were this voxel not activated (in which case an apparent deactivation is seen).

18

3.2 Models

In the following subsections, the flexibility of the general linear model is demonstrated using models for various PET functional mapping experiments. For generality, ANCOVA style models are used, with gCBF included as a confounding covariate. The corresponding ANOVA models for data adjusted by proportional scaling can be obtained by omitting the global terms. Voxel level models are presented in the usual statistical notation, alongside the SPM description and "images" of example design matrices. The form of contrasts for each design are indicated, and some practicalities of the SPM interface are discussed.

Single subject models

3.2.1 Single subject activation design

The simplest experimental paradigm is the simple single subject activation experiment. The ANCOVA model for this design was discussed in the previous section on global normalisation. This model extends to more than two conditions in the obvious way. Suppose there are Q conditions, with M_q scans under condition Q. Let Y_{qj}^k denote the rCBF at voxel k in scan $j = 1, \ldots, M_q$ under condition $q = 1, \ldots, Q$. The model is:

$$Y_{qj}^k = \alpha_q^k + \mu^k + \zeta^k (g_{qj} \Leftrightarrow \overline{g}_{\bullet\bullet}) + \epsilon_{jq}^k \qquad (16)$$

There are $Q + 2$ parameters for the model at each voxel: The Q condition effects, the constant term μ^k, and the global regression effect, giving parameter vector $\underline{\beta}^k = (\alpha_1^k, \ldots, \alpha_Q^k, \mu^k, \zeta^k)^\top$ at each voxel. In this model, replications of the same condition are modelled with a single effect. The model is overparameterised, having only $Q \Leftrightarrow 1$ degrees of freedom, leaving $N \Leftrightarrow Q \Leftrightarrow 1$ residual degrees of freedom, where $N = \sum M_q$ is the total number of scans.

Allowable contrasts are linear compounds $\underline{c}^\top \underline{\beta}^k$ for which the weights sum to zero over the condition effects, and give zero weight to the constant term, i.e. $\sum_{q=1}^{Q} c_q = 0$. For example, to test $\mathcal{H}^k : \alpha_1^k = (\alpha_2^k + \alpha_3^k)/2$ against the one sided alternative $\overline{\mathcal{H}}^k : \alpha_1^k > (\alpha_2^k + \alpha_3^k)/2$, appropriate contrast weights would be $\underline{c} = (1, \Leftrightarrow \frac{1}{2}, \Leftrightarrow \frac{1}{2}, 0, \ldots, 0)^\top$. Large positive values of the t-statistic express evidence against the null hypothesis, in favour of the alternative hypothesis.[8] The global regression parameter is uniquely determined, and its value could be assessed (against a one-sided alternative) using contrast weights $\underline{c} = (0, \ldots, 0, 1)^\top$. However, the global confound is usually of no interest.

SPM partitions the model parameters (at each voxel) into effects of interest $(\underline{\beta}_1^k)$, and confounding effects, $\underline{\beta}_2^k$, with corresponding partitioning of the contrast weights vector $\underline{c}^\top = (\underline{c}_1^\top | \underline{c}_2^\top)$. Since the confounding effects are deemed of no interest, the weights for the confounding effects are assumed to be zero, and only the contrast weights for the effects of interest $(\underline{c}_1)$ are prompted for. In this case only the Q condition effects are designated of interest, so SPM requests a Q-vector of contrast weights. Henceforth, we shall give example contrast weights only for the effects of interest.

---

[8]In SPM all t-tests are one sided.

The constant term is of no interest, hence it's position with the confounding effects in the model (Eqn.16). The explicit inclusion of the constant term amongst the confounds enables the reduced model, with $\underline{\beta}_1^k = \underline{0}$, to fit the overall level. Thus the omnibus null hypothesis assessed by the extra sum-of-squares F-test for $\mathcal{H}^k : \underline{\beta}_1^k = \underline{0}$, assesses the omnibus hypothesis that there is no condition effect of any sort at voxel k, large values of the statistic expressing evidence against the null hypothesis in favour of the alternative hypothesis $\overline{\mathcal{H}}^k : \underline{\beta}_1^k \neq \underline{0}$

Design matrix "images"

Throughout SPM, linear models are represented by grayscale "images" of the design matrix used. Together with the list of scans, this "image" specifies the model at a glance. (Clearly the ordering of the scans is important, since each row of the design matrix corresponds to a particular scan. Reordering the rows gives the same model provided the scans are similarly reordered.)

A grey-scale "image" of an example design matrix for a single subject activation study with four scans under each of three conditions is shown in figure 6. The first three columns contain indicator variables (with values zero and one) indicating the condition. The last column contains the respective (mean corrected) gCBF values. The scans corresponding to this design matrix have been ordered such that all the scans for each condition appear together. Obviously the actual experiment would be carried out with a suitably randomised condition presentation order.

In the grey-scale design matrix images, ⇔1 is black, 0 mid-gray, and +1 white. Columns containing covariates are scaled by subtracting the mean (zero for centered covariates), and dividing the resulting values by their absolute maximum, giving values in [0,1]. (Design matrix blocks containing factor by covariate interactions (§3.2.5) are scaled such that the covariate values lie in (0,1], thus preserving representation of the "padding" zeros as mid-grey.)

3.2.2   Single subject parametric design

Consider the single subject parametric experiment where a single covariate of interest, or "score", is measured. For instance, the covariate may be a physiological variable, a task difficulty rating, or a performance score. It is desired to find regions where the rCBF values are highly correlated with the covariate, taking into account the effect of global changes. Figure 7a depicts the situation. If Y_j^k is the rCBF at voxel k of scan $j = 1, \ldots, J$ and s_j is the independent covariate, then a simple ANCOVA style model is a multiple regression with two covariates:

$$Y_j^k = \varrho^k (s_j \Leftrightarrow \overline{s}_\bullet) + \mu^k + \zeta^k (g_j \Leftrightarrow \overline{g}_\bullet) + \epsilon_j^k \tag{17}$$

Here, ϱ is the slope of the regression plane in the direction of increasing score, fitted separately for each voxel.

There are three model parameters, leaving $J \Leftrightarrow 3$ residual degrees of freedom. The design matrix (Fig.7b) has three columns, a column containing the (centered) score covariate,
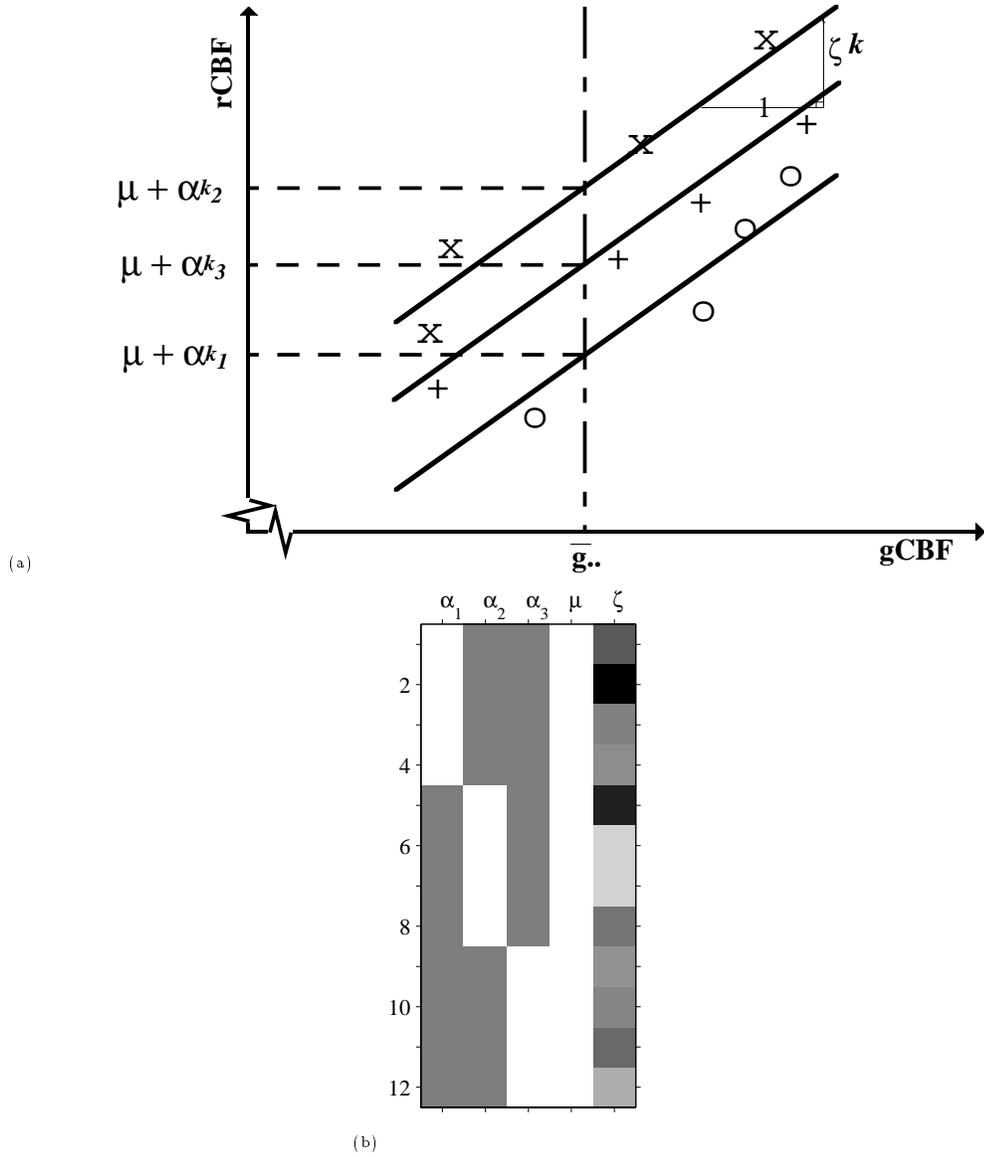
Figure 6: Single subject activation experiment, ANCOVA design)§3.2.1). Illustrations for a three-condition experiment with four scans in each of three conditions, ANCOVA design. (a) Illustrative plot of rCBF vs. gCBF. (b) Design matrix "image", with columns labelled by their respective parameters. The scans are ordered by condition.

a column of dummy 1's corresponding to μ^k, and a column containing the (centered) global values.

In SPM this is a "Single subject: Covariates only" design. The design is uniquely specified, so any linear combination of the three parameters is a contrast. SPM designates ϱ as the only effect of interest, so only a single contrast weight is requested. The null hypothesis of no score effect at voxel k, $\mathcal{H}^k : \varrho^k = 0$, can be assessed against against the one sided alternative hypotheses $\overline{\mathcal{H}}^k : \varrho^k > 0$ (rCBF increasing with score) with contrast weights for the effects of interest $\underline{c}_1 = +1$, and against $\overline{\mathcal{H}}^k : \varrho^k < 0$ (rCBF decreasing as score increases) with contrast weights $\underline{c}_1 = \Leftrightarrow 1$.

This simple model assumes a linear relationship between rCBF and the covariate (and other explanatory variables). More general relationships may be modeled by including other functions of the covariate. These functions of the covariate are essentially new explanatory variables, which if linearly combined still fit in the framework of the general linear model. For instance, if an exponential relationship is expected, $\ln(s_j)$ would be used in place of s_j. Fitting powers of covariate as additional explanatory variables leads to *polynomial regression*. More generally, sets of *basis functions* can be added as a suite of functions of a covariate to allow flexible modelling. This theme will be developed later in this chapter (for fMRI), and in other chapters.

3.2.3 Simple single subject activation revisited

As discussed in the General Linear Modeling section (§2), it is often possible to reparameterise the same model in many ways. As an example, consider again a simple two condition ($Q = 2$) single subject activation experiment, discussed above (§3.2.1). The model (Eqn.16) is:

$$Y_{qj}^k = \alpha_q^k + \mu^k + \zeta^k(g_{qj} \Leftrightarrow \overline{g}_{\bullet\bullet}) + \epsilon_{jq}^k$$

The model is over-determined, so consider a sum-to-zero constraint on the condition effects. For two conditions this implies $\alpha_1^k = \Leftrightarrow\alpha_2^k$. Substituting for α_2^k the resulting design matrix has a column containing $+1$'s and $\Leftrightarrow 1$'s indicating the condition $q = 1$ or $q = 2$ respectively, a column of 1's for the overall mean, and a column containing the (centered) gCBF (Fig.8). The corresponding parameter vector is $\underline{\beta}^k = (\alpha_1^k, \mu^k, \zeta^k)^\top$. Clearly this is the same design matrix as that for a parametric design with (non-centered) "score" covariate indicating the condition as active or baseline with $+1$ or $\Leftrightarrow 1$ respectively.[9] The hypothesis of no activation at voxel k, $\mathcal{H}^k : \alpha_1^k = 0$ can be tested against the one sided alternatives $\overline{\mathcal{H}}^k : \alpha_1^k > 0$ (activation) and $\overline{\mathcal{H}}^k : \alpha_1^k < 0$ with contrast weights for the effects of interest $\underline{c}_1 = 1$ and $\underline{c}_1 = \Leftrightarrow 1$ respectively. This example illustrates how the SPM interface may be used to enter "hand-built" blocks of design matrix as non-centered covariates.

3.2.4 Single subject: conditions and covariates

Frequently there are other confounding covariates in addition to gCBF that can be added into the model. For example, a linear time component could be modeled simply by

---

[9] In SPM94 this was how simple single subject activation studies were routinely entered.

(a)

(b)
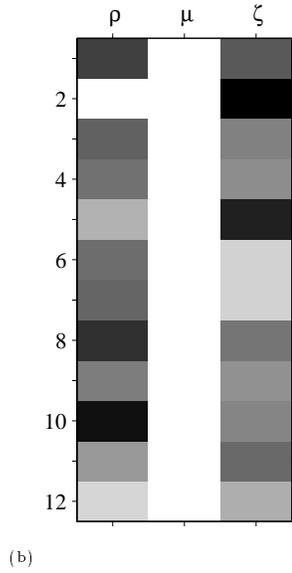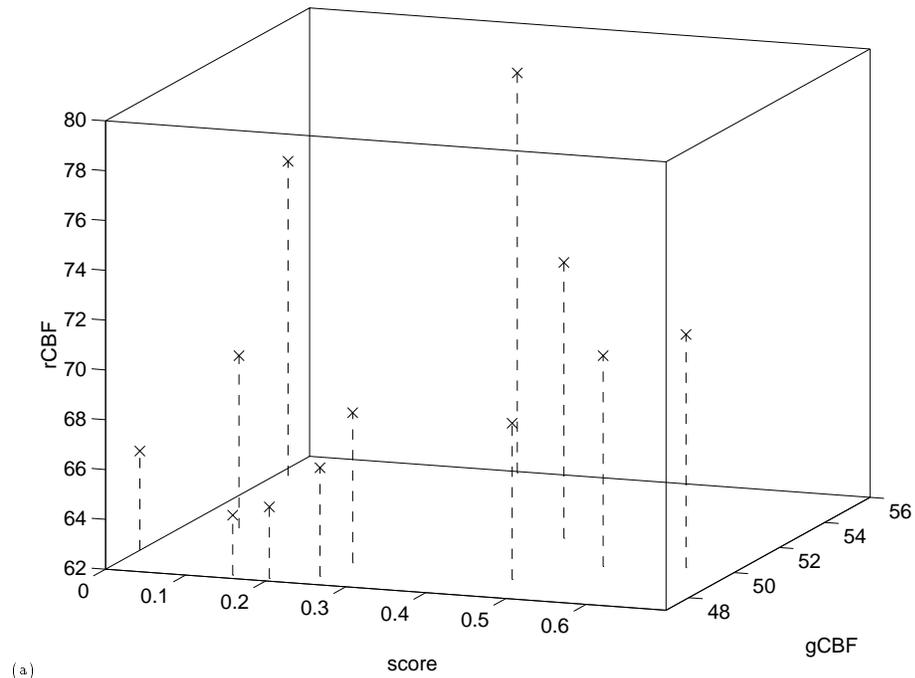
Figure 7: Single subject parametric experiment (§3.2.2): (a) Plot of rCBF vs. score and gCBF. (b) Design matrix "image" for Eqn.17, illustrated for a 12 scan experiment. Scans are ordered in the order of acquisition.
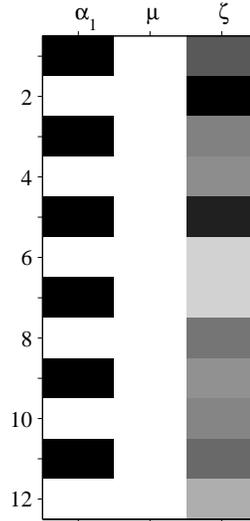
Figure 8: Example design matrix "image" for single subject simple activation experiment, with six scans in each of two conditions, formulated as a parametric design (§3.2.3). The twelve scans are ordered alternating between baseline and activation conditions, as might have been the order of acquisition.

entering the scan number as covariate. In SPM these appear in the design matrix as additional covariate columns adjacent to the global flow column. Care must be taken to enter the covariates in the same order that the scans were specified, since SPM asks for scans ordered by condition in many instances!

3.2.5 Factor by covariate interactions

A more interesting experimental scenario is when a parametric design is repeated under multiple conditions in the same subject(s). A specific example would be a PET language experiment in which, during each of twelve scans, lists of words are presented. Two types of word list (the two conditions) are presented at each of six rates (the parametric component). Interest may lie in locating regions where there is a difference in rCBF between conditions (accounting for changes in presentation rate), the *main* effect of condition; locating regions where rCBF increases with rate (accounting for condition), the main effect of rate; and possibly assessing evidence for condition specific responses in rCBF to changes in rate, an interaction effect.[10] Let Y_{qrj}^k denote the rCBF at voxel k for the j-th measurement under rate $r = 1, \ldots, R$ and condition $q = 1, \ldots, Q$, with s_r the rate covariate (some function of the rates). A suitable model is:

$$Y_{qrj}^k = \alpha_q^k + \varrho_q^k(s_{qr} \Leftrightarrow \overline{s}_{\bullet\bullet}) + \mu^k + \zeta^k(g_{qrj} \Leftrightarrow \overline{g}_{\bullet\bullet\bullet}) + \epsilon_{qrj}^k \qquad (18)$$

Note the q subscript on the voxel covariate effect parameter ϱ_q^k, indicating different slopes for each condition. Ignoring for the moment the global flow, the model describes two simple regressions with common error variance (Fig.9a). The SPM interface

---

[10]Two experimental factors *interact* if the level of one affects the expression of the other.

describes such factor by covariate interactions as "factor specific covariate fits". The interaction between condition and covariate effects is manifest as different regression slopes for each condition. There are $2Q + 2$ parameters for the model at each voxel, $\underline{\beta}^k = (\alpha_1^k, \ldots, \alpha_Q^k, \varrho_1^k, \ldots, \varrho_Q^k, \mu^k, \zeta^k)$, with $2Q + 1$ degrees of freedom. A design matrix "image" for the two condition example is shown in figure 9b. The factor by covariate interaction takes up the third and fourth columns, corresponding to the parameters ϱ_1^k and ϱ_2^k, the covariate being split between the columns according to condition, the remaining cells filled with zeros.

Only the constant term and global slope are designated confounding, giving $2Q$ effects of interest to specify contrast weights for, $\underline{\beta}_1^k = (\alpha_1^k, \ldots, \alpha_Q^k, \varrho_1^k, \ldots, \varrho_Q^k)$. As with the activation study model, contrasts have weights which sum to zero over the condition effects. For the 2 condition word presentation example, contrast weights $\underline{c}_1^\top = (0, 0, 1, 0)$ for the effects of interest express evidence against the null hypothesis that there is is no covariate effect in condition one, with large values indicating evidence of a positive covariate effect. Weights $\underline{c}_1^\top = (0, 0, \frac{1}{2}, \frac{1}{2})$ address the hypothesis that there is no average covariate effect across conditions, against the one sided alternative that the average covariate effect is positive. Weights $\underline{c}_1^\top = (0, 0, \Leftrightarrow 1, +1)$ address the hypothesis that there is no condition by covariate interaction, that is, that the regression slopes are the same, against the alternative that the condition 2 regression is steeper.

Conceptually, contrast weights $\underline{c}_1^\top = (\Leftrightarrow 1, +1, 0, 0)$ and $\underline{c}_1^\top = (+1, \Leftrightarrow 1, 0, 0)$ for the effects of interest assess the hypothesis of no condition effect against appropriate one-sided alternatives. However, the comparison of main effects is confounded in the presence of an interaction: In the above model, both gCBF and the rate covariate were centered, so the condition effects α_q^k are the relative heights of the respective regression lines (relative to μ^k) at the mean gCBF and mean rate covariate. Clearly if there is an interaction, then difference in the condition effects (the separation of the two regression lines) depends on where you look at them. Were the rate covariate not centered, the comparison would be at mean gCBF and zero rate, possibly yielding a different result.

Thus main effects of condition in such a design must be interpreted with caution. If there is little evidence for a condition dependent covariate effect then there is no problem. Otherwise, the relationship between rCBF and other design factors should be examined graphically to assess whether the perceived condition effect is sensitive to the level of the covariate.

Multi-subject designs

Frequently, experimentally induced changes of rCBF are subtle, such that analyses must be pooled across subjects to find statistically significant evidence of an experimentally induced effect. Further, (for appropriate models) the use of samples of subjects allows inference to be extended to the population from which they were sampled.

The simple single subject designs presented above must be extended to account for subject to subject differences. The simplest type of subject effect is an additive effect, a *block* effect in the parlance of experimental design. This implies that all subjects respond in the same way, save for an overall shift in rCBF (at each voxel). We extend our notation by adding subscript i for subjects, so Y_{iqj}^k is the rCBF at voxel k of scan j under condition

(a)

(b)
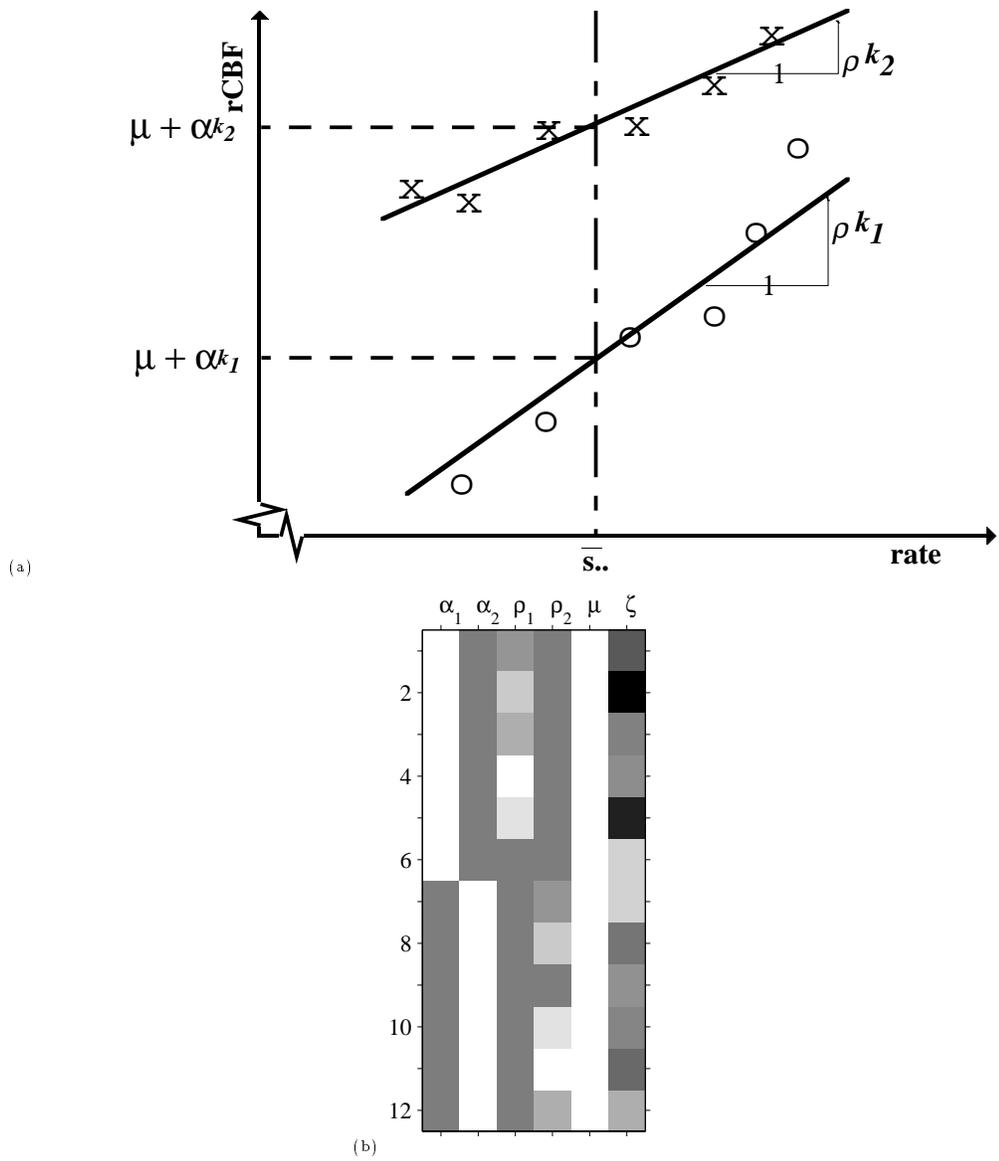
Figure 9: Single subject experiment with conditions, covariate, and condition by covariate interaction. (§3.2.5): (a) Illustrative plot of rCBF vs. rate. (b) Design matrix "image" for Eqn.18. Both illustrated for the two condition 12 scan experiment described in the text. The scans have been ordered by condition.

q on subject $i = 1, \ldots, N$.

3.2.6   Multi subject activation (replications)

For instance, the single subject activation model (Eqn.16) is extended by adding subject effects γ_i^k giving model:

$$Y_{iqj}^k = \alpha_q^k + \gamma_i^k + \zeta^k(g_{iqj} \Leftrightarrow \overline{g}_{\bullet\bullet\bullet}) + \epsilon_{iqj}^k \tag{19}$$

A schematic plot of rCBF vs. gCBF for this model is shown in figure 10a. In SPM terminology, this is a "multi-subject: replication of conditions" design. The parameter vector at voxel k is $\underline{\beta}^k = (\alpha_1^k, \ldots, \alpha_Q^k, \gamma_1^k, \ldots, \gamma_N^k, \zeta^k)^\top$. The design matrix (Fig.10b) has N columns of dummy variables corresponding to the subject effects. (Similarly a multi-subject parametric design could be derived from the single subject case (§3.2.2) by including appropriate additive subject effects.)

Again, the model is overparameterised, though this time we have omitted the explicit constant term from the confounds, since the subject effects can model an overall level. Adding a constant to each of the condition effects and subtracting it from each of the subject effects gives the same model. Bearing this in mind, it is clear that contrasts must have weights that sum to zero over both the subject effects and the condition effects. Subject effects are designated as confounding and not of interest, so SPM prompts only for contrast weights for the Q condition effects, and the same rules apply for these as with the single subject models.

3.2.7   Condition by replication interactions

The above model assumes that (accounting for global and subject effects) replications of the same condition give the same (expected) response. There are many reasons why this assumption may be inappropriate, such as time effects or learning effects. Time effects can be modeled by including appropriate functions of the scan number as confounding covariates. With multi-subject designs we have sufficient degrees of freedom available to enable the consideration of replication by condition interactions. Such interactions imply that the (expected) response to each condition is different between replications (having accounted for other effects in the model). Usually in statistical models, interaction terms are added to a model containing main effects. However, such a model is so overparameterised that the main effects may be omitted, leaving just the interaction terms. The model is:

$$Y_{iqj}^k = \alpha\vartheta_{(qj)}^k + \gamma_i^k + \zeta^k(g_{iqj} \Leftrightarrow \overline{g}_{\bullet\bullet\bullet}) + \epsilon_{iqj}^k \tag{20}$$

where $\alpha\vartheta_{(qj)}^k$ is the interaction effect for replication j of condition q, the condition-by-replication effect. As with the previous model, this model is overparameterised (by one degree of freedom), and contrasts must have weights which sum to zero over the condition-by-replication effects. There are as many of these condition-by-replication terms as there are scans per subject. (An identical model is arrived at by considering each replication of each experimental condition as a separate condition.) If the scans are reordered such that the j-th scan corresponds to the same replication of the same condition in each subject, then the condition-by-replication corresponds to the scan number. With this feature in

(a)

(b)
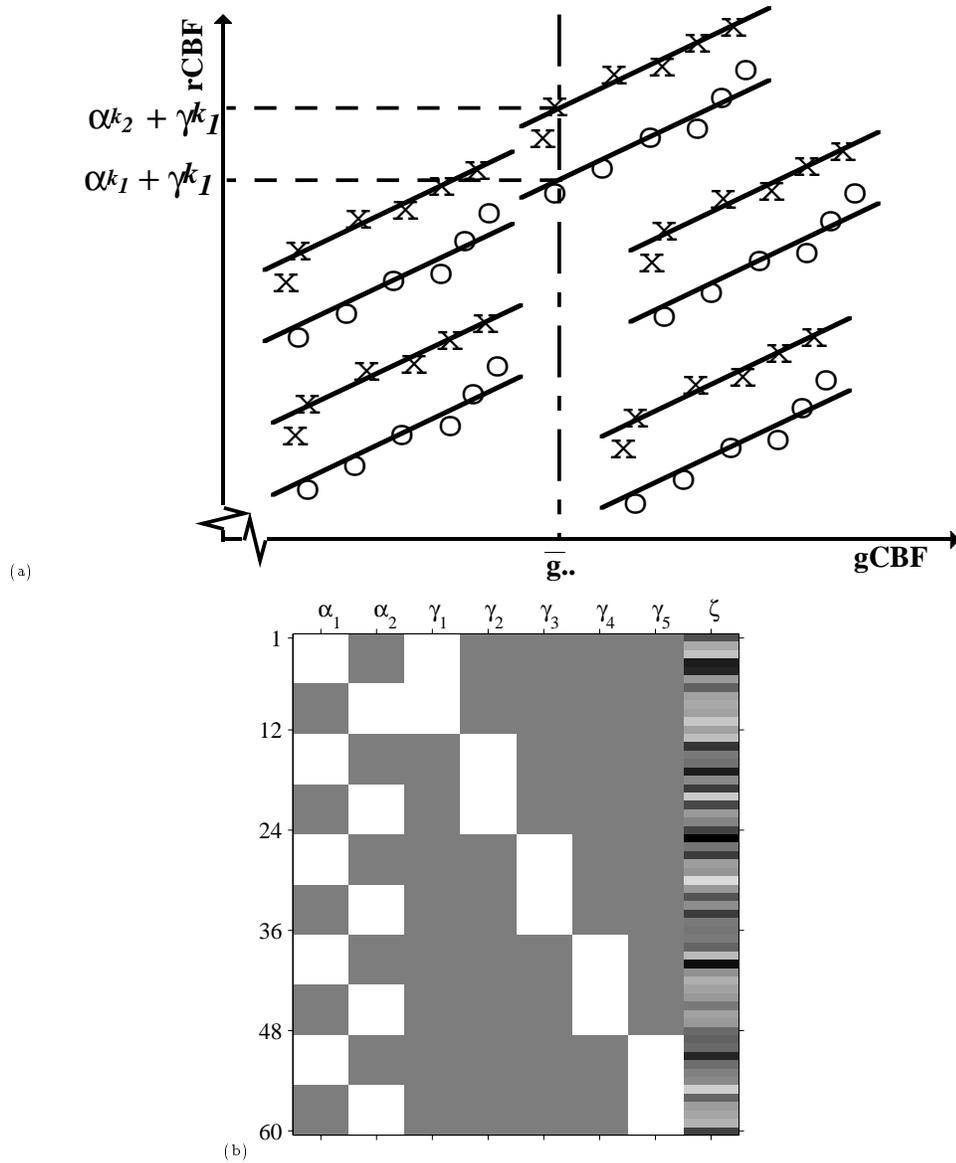
Figure 10: Multi-subject activation experiment, replication of conditions (§3.2.6), model Eqn.19. Illustrations for a 5 subject study, with six replications of each of two conditions per subject: (a) Illustrative plot of rCBF vs. gCBF. (b) Design matrix "image": The first two columns correspond to the condition effects, the next five to the subject effects, the last to the gCBF regression parameter. The design matrix corresponds to scans ordered by subject, and by condition within subjects.

28
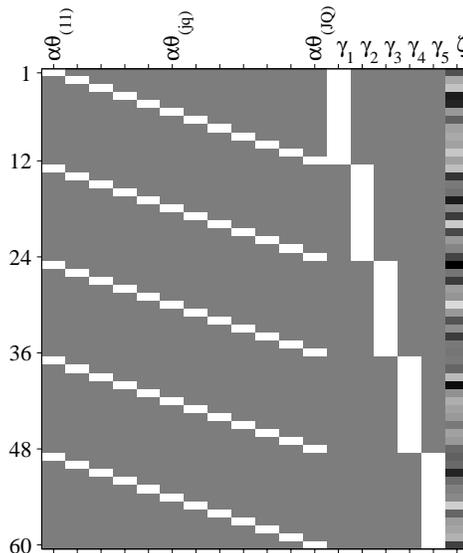
Figure 11: Multi-subject activation experiment, "classic" SPM design, where each replication of each experimental condition is considered as a separate condition (Eqn.20). Illustrative design matrix "image" for five subjects, each having 12 scans, the scans having been ordered by subject, and by condition and replication within subject. The columns are labelled with the corresponding parameter. The first twelve columns correspond to the "condition" effects, the next five to the subject effects, the last to the gCBF regression parameter.

mind it is usually simplest to reorder scans when using such a model. An example design matrix for 5 subjects scanned twelve times is shown in figure 11a, where the scans have been reordered. In SPM this is termed a "Multi-subject: conditions only" design.

This is the "classic" SPM ANCOVA described by Friston *et al.* (1990), and implemented in the original SPM software.[11] It offers great latitude for specification of contrasts. Appropriate contrasts can be used to assess main effects, specific forms of interaction, and even parametric effects. For instance, consider the verbal fluency data-set described by Friston *et al.* (1995b): Five subjects were scanned twelve times, six times under each of two conditions, word shadowing (condition A) and intrinsic word generation (condition B). The scans were reordered to ABABABABABAB for all subjects. (The actual condition presentation order was randomly allocated for each subject in blocks of two scans. That is, successive pairs of scans were chosen as AB or BA on the toss of a coin.) Then a contrast with weights (for the condition-by-replication effects) of $\underline{c}_1^\mathsf{T} = (\Leftrightarrow 1, 1, \Leftrightarrow 1, 1, \Leftrightarrow 1, 1, \Leftrightarrow 1, 1, \Leftrightarrow 1, 1, \Leftrightarrow 1, 1)$ assesses the hypothesis of no main effect of word generation (against the one-sided alternative of activation). A contrast with weights of $\underline{c}_1^\mathsf{T} = (5\frac{1}{2}, 4\frac{1}{2}, 3\frac{1}{2}, 2\frac{1}{2}, 1\frac{1}{2}, \frac{1}{2}, \Leftrightarrow\frac{1}{2}, \Leftrightarrow 1\frac{1}{2}, \Leftrightarrow 2\frac{1}{2}, \Leftrightarrow 3\frac{1}{2}, \Leftrightarrow 4\frac{1}{2}, \Leftrightarrow 5\frac{1}{2})$ is sensitive to linear decreases in rCBF over time, independent of condition, and accounting for subject effects and changes in gCBF. A contrast with weights of $\underline{c}_1^\mathsf{T} = (1, \Leftrightarrow 1, 1, \Leftrightarrow 1, 1, \Leftrightarrow 1, \Leftrightarrow 1, 1, \Leftrightarrow 1, 1, \Leftrightarrow 1, 1)$ assesses the interaction of time and condition, subtracting the activation in the first half of the experiment from that in the latter half.

---

[11]The original SPM software is now fondly remembered as SPM*classic*.

3.2.8   Interactions with subject

While it is (usually) reasonable to use ANCOVA style models to account for global flow, with regression parameter constant across conditions, the multi-subject models considered thus far assume additionally that this regression parameter is constant across subjects. It is quite possible that rCBF at the same location for different subjects will respond differentially to changes in gCBF – a subject by gCBF covariate interaction. The gCBF regression parameter can be allowed to vary from subject to subject. Extending the multi-subject activation (replication) model (Eqn.19) in this way gives:

$$Y_{iqj}^k = \alpha_q^k + \gamma_i^k + \zeta_i^k(g_{iqj} \Leftrightarrow \overline{g}_{\bullet\bullet\bullet}) + \epsilon_{iqj}^k \tag{21}$$

Note the i subscript on the global slope term, ζ_i^k, indicating a separate parameter for each subject. A schematic plot of rCBF vs. gCBF for this model and an example design matrix "image" are shown in figure 12. In the terminology of the SPM interface, this is an "ANCOVA by subject". The additional parameters are of no interest, and contrasts are as before.

Similarly, the SPM interface allows subject by covariate interactions, termed "subject specific fits". Currently SPM doesn't allow subject by condition interactions. If the parameters of interest are allowed to interact with subject effects, then account must be taken of variation between subjects. Essentially, the subject effects are *random effects*, since subjects are sampled from some population about which we wish to infer. In all the models considered thus far, the relevant variation is within subjects, and the variation of the subject effects can be ignored. A rigorous discussion of such *mixed effects* models is beyond the scope of this chapter. For details, see any intermediate level statistical treatment of the design of experiments.

3.3   Multi-study designs

The last class of SPM models for PET we consider are the "multi-study" models. In these models, subjects are grouped into two or more *studies*. The "multi-study" designs fit separate condition effects for each study. In statistical terms this is a *split plot* design. As an example consider two multi-subject activation studies, the first with five subjects scanned twelve times under two conditions (as described above in section 3.2.6), the second with three subjects scanned six times under three conditions. An example design matrix "image" for a model containing study specific condition effects, subject effects and study specific global regression (termed "ANCOVA by study" in SPM) is shown in figure 13. The first two columns of the design matrix correspond to the condition effects for the first study, the next two to the condition effects for the second study, the next eight to the subject effects, and the last to the gCBF regression parameter. (The corresponding scans are assumed to be ordered by study, by subject within study, and by condition within subject.)

Such multi-study designs are overparameterised by numerous degrees. An additional consideration for split-plot designs is the variation between subjects in each study. SPM only assesses the residual variability, the intra-subject variability, and therefore cannot be used to appropriately assess main effects of study, since these are differences between
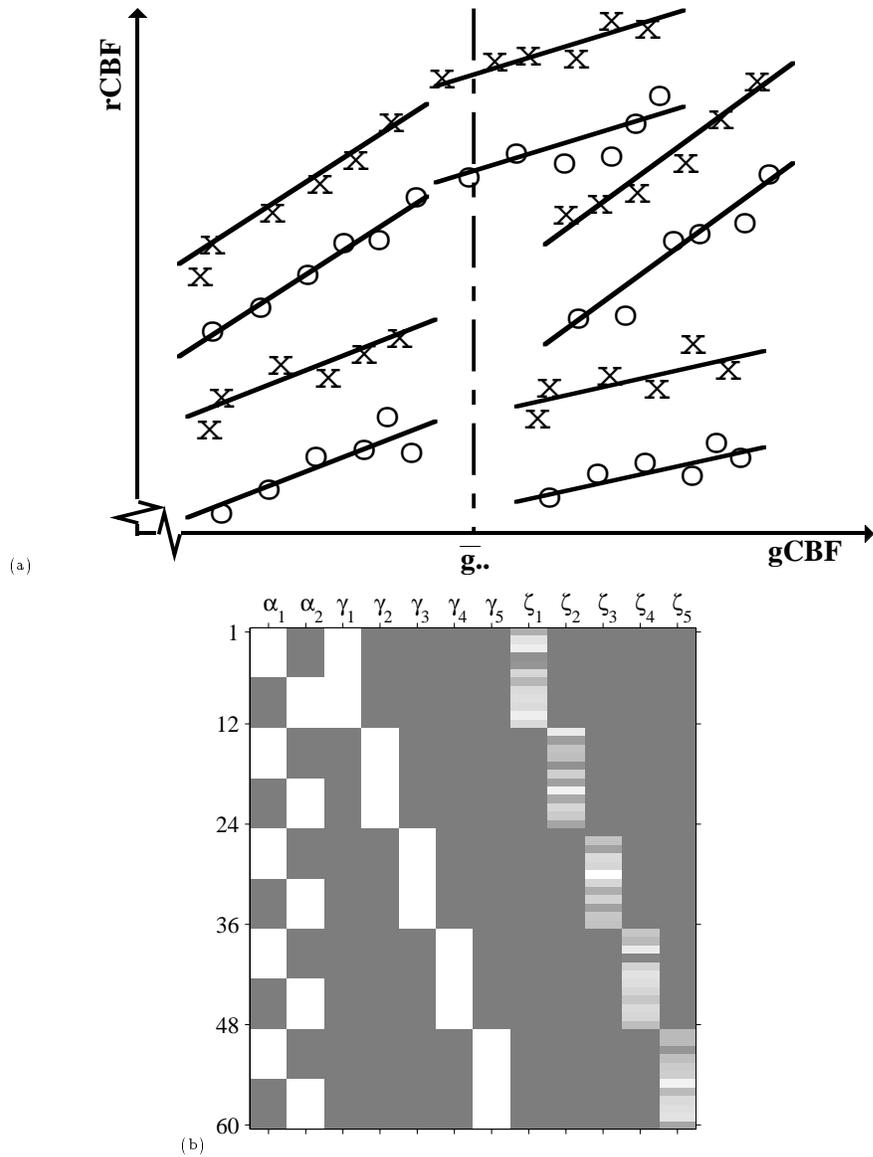
Figure 12: Multi-subject activation experiment, replication of conditions, ANCOVA by subject. Model Eqn.21. Illustrations for a 5 subject study, with six replications of each of two conditions per subject: (a) Illustrative plot of rCBF vs. gCBF. (b) Design matrix "image": The first two columns correspond to the condition effects, the next five to the subject effects, the last five the gCBF regression parameters for each subject. The design matrix corresponds to scans ordered by subject, and by condition within subjects.
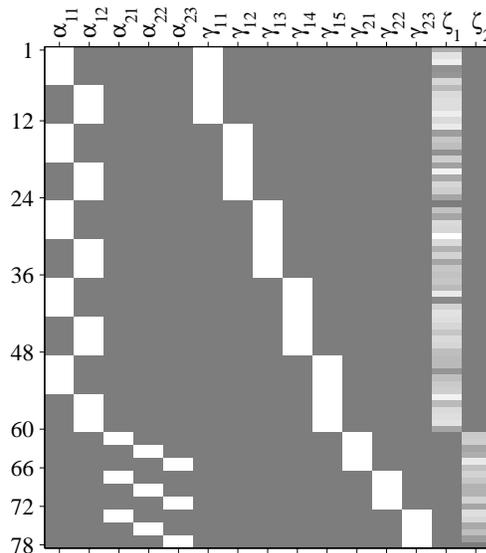
Figure 13: Design matrix "image" for the example multi-study activation experiment described in section 3.3.

subjects for which a measure of inter-subject variability is required. The only exception to this is the degenerate case when there is only one scan per subject, since in this case the residual variability is inter-subject. In this case the model reduces to that of a single subject activation experiment where groups are interpreted as conditions and subjects as replicate observations within the groups.

Valid contrasts for multi-study designs in SPM are contrasts whose weights, when considered for each of the studies individually, would define a contrast for the study. Thus, contrasts must have weights which sum to zero over the condition effects within each study. There remain three types of useful comparison available. The first is a comparison of condition effects within a single study, carried out in the context of a multi-study design; the contrast weights appropriate for the condition effects of the study of interest is entered, padded with zeros for the other study, e.g. $\underline{c}_1^\mathsf{T} = (1, \Leftrightarrow 1, 0, 0, 0)$ for the first study in our example. This will have additional power when compared to an analysis of this study in isolation, since the second study observations contribute to the variance estimates. The second is an average effect across studies; contrasts for a particular effect in each of the studies are concatenated, the combined contrast assessing a mean effect across studies. For example, if the second study in our example has the same conditions as the first, plus an additional condition, then such a contrast would have weights for the effects of interest $\underline{c}_1^\mathsf{T} = (\Leftrightarrow 1, 1, \Leftrightarrow 1, 1, 0)$. Lastly, differences of contrasts across studies can be assessed, such as differences in activation. The contrasts weights for the appropriate main effect in each study are concatenated, with some studies contrasts negated. In our example, $\underline{c}_1^\mathsf{T} = (\Leftrightarrow 1, 1, 1, \Leftrightarrow 1, 0)$ would be appropriate for locating regions where the first study activated more than the second, or where the second deactivated more than the first.

Assumption of model fit in this case includes the assumption that the error terms have equal variance (at each voxel) across studies. For very different study populations, or studies from different scanners or protocols (possibly showing large differences in the

measured gA between studies), this assumption may only be tenable after scaling the data for each study to have the same grand mean activity (grand mean scaling by study).

4 $f$unctional MRI

The analysis of functional Magnetic Resonance time series (fMRI) presents additional statistical problems. In this section some extensions to the basic general linear model theory for the analysis of fMRI time series are presented. A detailed discussion of design issues for fMRI is presented in another chapter, so only the simplest model will be considered here.

4.1 A time series linear model

Again, a linear model is used to model the fMRI time series of observations at each voxel, with the same model form for all voxels. Suppose we have a time series of N observations $Y_1, \ldots, Y_s, \ldots, Y_N$, acquired at times t_s, where $s = 1, \ldots, N$ is the *scan number*. The approach is to model the observed time series as a linear combination of explanatory functions, plus a residual error or "noise" term:

$$Y_s^k = \beta_1^k f^1(t_s) + \ldots + \beta_l^k f^l(t_s) + \ldots + \beta_L^k f^L(t_s) + \epsilon_s^k \tag{22}$$

Here the L functions $f^1(.), \ldots, f^L(.)$ are a suitable set of *basis functions*, designed such that linear combinations of them span the space of possible fMRI responses for this experiment, up to the level of residual noise. Clearly this is a general linear model, since the explanatory variables being the basis functions are evaluated at the time points of the fMRI time series. To emphasise this point, consider writing out the above (Eqn.22) for all time points t_s, to give the a set of simultaneous equations:

$$
\begin{aligned}
Y_1^k &= f^1(t_1)\beta_1^k + \ldots + \beta_l^k f^l(t_1) + \ldots + f^L(t_1)\beta_L^k + \epsilon_1^k \\
\vdots &= \vdots \\
Y_s^k &= f^1(t_s)\beta_1^k + \ldots + \beta_l^k f^l(t_s) + \ldots + f^L(t_s)\beta_L^k + \epsilon_s^k \\
\vdots &= \vdots \\
Y_N^k &= f^1(t_N)\beta_1^k + \ldots + \beta_l^k f^l(t_N) + \ldots + f^L(t_N)\beta_L^k + \epsilon_N^k
\end{aligned}
$$

which in matrix form is:

$$
\begin{pmatrix} Y_1^k \\ \vdots \\ Y_s^k \\ \vdots \\ Y_N^k \end{pmatrix}
=
\begin{pmatrix}
f^1(t_1) & \ldots & f^l(t_1) & \ldots & f^L(t_1) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
f^1(t_s) & \ldots & f^l(t_s) & \ldots & f^L(t_s) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
f^1(t_N) & \ldots & f^l(t_N) & \ldots & f^L(t_N)
\end{pmatrix}
\begin{pmatrix} \beta_1^k \\ \vdots \\ \beta_s^k \\ \vdots \\ \beta_N^k \end{pmatrix}
+
\begin{pmatrix} \epsilon_1^k \\ \vdots \\ \epsilon_s^k \\ \vdots \\ \epsilon_N^k \end{pmatrix}
$$

of the form of the general linear model in matrix form $\underline{Y}^k = X\underline{\beta}^k + \underline{\epsilon}^k$. Here each column of the design matrix X contains the values of one of the basis functions evaluated each

time point of the fMRI time series. That is, the columns of the design matrix are the discretised basis functions.

The basis functions must be chosen to span the space of all possible fMRI responses for the experiment in question, such that the residual "noise" terms ϵ_s^k are identically distributed with zero mean. Additionally, we assume that they are normally distributed with constant variance (at each voxel), $\epsilon_s^k \sim \mathcal{N}(0, \sigma_k^2)$, a reasonable assumption given the explicit and implicit smoothing of the required pre-processing steps.

Simplisticly, the fMRI time series can be regarded as the underlying neuronal activity expressed through the hæmodynamic response, with added noise. These noise sources include measurement noise, which is white, and physiological effects such cardiac and respiratory effects, aliased to low frequencies, and drifts. Noise sources and their impact on modelling and experimental design are discussed in detail in chapter 9. Experimental and low frequency confounding effects must be modelled, but for inter-scan intervals less than the length of the hæmodynamic response (TR < 7s), the residual "noise" time series will exhibit short term serial correlation. This *temporal autocorrelation* of the residual errors requires extensions to the standard theory of the General Linear Model developed earlier (§2).

4.2   Serially correlated regression — Temporal smoothing

The model should account for most of the temporal autocorrelation in an fMRI time series, leaving the residuals with temporal autocorrelation somewhat less than implied by the hæmodynamic response, and negligible for large inter-scan intervals. Were the form of the residual temporal autocorrelation known, results from the Statistical literature on serially correlated regression could be utilised. Further, by the matched filter theorem, since the fMRI signal for a point response has the form of the hæmodynamic response function, the signal is optimally recovered if the fMRI time series (and the basis functions of the design) are smoothed by convolution with the hæmodynamic response function. In the smoothed residual time series, the temporal autocorrelation of the raw residuals is swamped by that imposed by the temporal smoothing, which is known. Thus, temporal smoothing with an appropriate filter provides optimal parameter estimation, and swamps the residual autocorrelation of the raw residuals such that the temporal autocorrelation of the smoothed residual time series can be regarded as known.

4.2.1   A temporally smoothed model

Formally, the model at each voxel is convolved with an approximate hæmodynamic response function, leading to:
$$K\underline{Y} = KX\underline{\beta} + K\underline{\epsilon} \tag{23}$$
at each voxel. (For clarity, we shall drop the superscripting with the voxel index k.) Note that the parameter estimates remain unaffected by the temporal smoothing. Here the convolution is expressed by the Toeplitz matrix K, with the discretised smoothing kernel on the leading diagonal. Only a kernel approximating the hæmodynamic response is required, since the basis functions of the model fit the actual form of the hæmodynamic response to neuronal events. A Gaussian kernel would do, with standard

$$K =$$
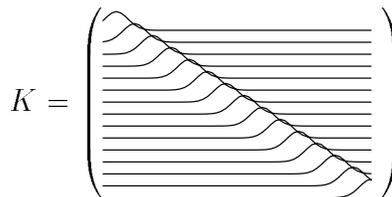
Figure 14: Illustrative toeplitz temporal smoothing matrix.

deviation parameter suitably chosen, say $\tau = \sqrt{8}$s.[12] In this case K would have elements $K_{ij} \propto \exp\left(-(i-j)^2/2\tau^2\right)$, as illustrated in figure 14.

The components of the error vector are then assumed independent $\epsilon_j \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, such that the autocovariance of the smoothed errors is completely specified by K and σ. Although the raw errors are slightly correlated, this autocorrelation is swamped by the smoothing making the assumption tenable.

4.2.2 Parameter estimates and distributional results

There is a large literature on the topic of serial correlation in regression, for example Watson (Watson, 1955) and Seber (Seber, 1977). We brought the approach to functional imaging and review it here (Friston *et al.*, 1995a; Worsley & Friston, 1995). Assuming X is of full-rank, the least-squares estimates of $\underline{\beta}$ are

$$\hat{\underline{\beta}} = (X^{*\top}X^*)^{-1}X^{*\top}K\underline{Y} \tag{24}$$

where $X^* = KX$. Although not fully optimal, $\hat{\underline{\beta}}$ is unbiased, with variance

$$\mathrm{var}\left[\hat{\underline{\beta}}\right] = \sigma^2 (X^{*\top}X^*)^{-1}X^{*\top}VX^*(X^{*\top}X^*)^{-1} \tag{25}$$

for $V = KK^\top$. This implies the following test statistic for a linear compound $\underline{c}^\top \hat{\underline{\beta}}$ of the parameters at voxel k:

$$T = \underline{c}^\top \hat{\underline{\beta}} \big/ (\underline{c}^\top \hat{\sigma}^2 (X^{*\top}X^*)^{-1}X^{*\top}VX^*(X^{*\top}X^*)^{-1}\underline{c})^{1/2} \tag{26}$$

with σ^2 estimated in the usual way by dividing the residual sum-of-squares by its expectation:

$$\hat{\sigma}^2 = \underline{e}^\top \underline{e} / \mathrm{trace}(RV) \tag{27}$$

where $\underline{e} = RKX$ is the vector of residuals and R the residual forming matrix. Here $\mathsf{E}\left[\hat{\sigma}^2\right] = \sigma^2$, and $\mathrm{var}\left[\hat{\sigma}^2\right] = 2\sigma^4 \frac{\mathrm{trace}(RVRV)}{\mathrm{trace}(RV)^2}$, so the effective degrees of freedom are then:

$$\nu = \frac{2\mathsf{E}\left[\hat{\sigma}^2\right]^2}{\mathrm{var}\left[\hat{\sigma}^2\right]} = \frac{\mathrm{trace}(RV)^2}{\mathrm{trace}(RVRV)} \tag{28}$$

By analogy with the χ^2 approximation for quadratic forms (Satterthwaite, 1946), the null distribution of T may be approximated by a t-distribution with ν degrees of freedom. Similarly, F-statistics for the extra sum-of-squares can be derived. Thus, hypotheses regarding contrasts of the parameter estimates, and general linear hypotheses can be assessed.

---

[12] A more elaborate approximation is utilised in the SPM software.
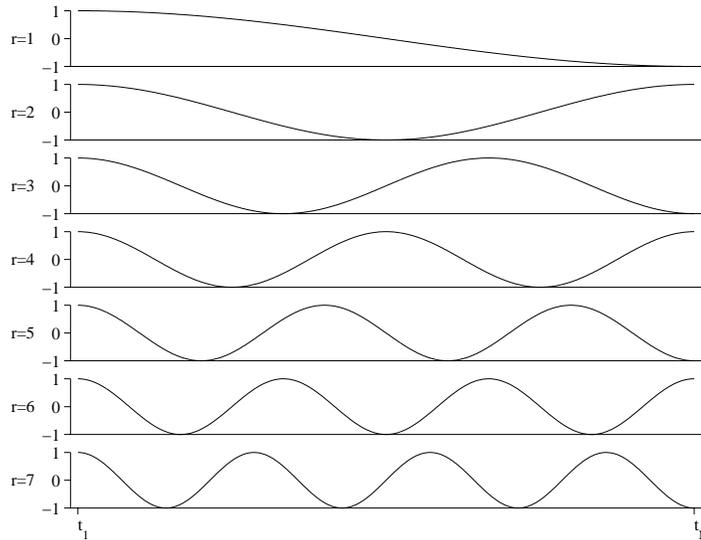
Figure 15: The "high-pass filter". Discrete cosine basis functions.

4.2.3 Smoothing fMRI data — Sacrilege?

Thus, temporal smoothing is the device permitting the use of standard parametric statistical methods. For short inter-scan intervals, less than about 7s say, temporal smoothing is necessary for the veracity of the statistical analysis. Many researchers, usually those with backgrounds in MRI physics, are strongly opposed to any form of smoothing for data which they have gone to great lengths to acquire at high spatial and temporal resolutions. However, such protestations against temporal smoothing are specious, since the temporal response to an experimental stimulus is governed by the hæmodynamic response. Nothing can be observed above this temporal resolution, and these responses are optimally recovered after convolution with a matched filter.

4.3 Low-frequency confounds — The "High pass" filter

As indicated above, the basis functions of the model must account for all components of the fMRI time series up to the level of residual noise with a small degree of temporal autocorrelation. In particular, in addition to the experimentally induced effects, low frequency confounds such as baseline drifts and aliased cardiac and respiratory effects must be modelled.

A suitable set of basis functions are discrete cosine functions: for a contiguous series of N scans acquired at times $t_1, \ldots, t_N$, the basis functions are of the form $f_r(t) = cos\left(r\pi \frac{t-t_1}{t_N-t_1}\right)$. The integer index r ranges from 1 (giving half a cosine cycle), to a user specified maximum, R, usually chosen such that the period of $f_r(t)$ is below that of the experimental paradigm. This basis set is similar to a Fourier basis set, but is superior for very low frequencies, such as long term drifts. The first seven such basis functions are illustrated in figure 15.

This proposed approach to modelling low-frequency confounds augments the design matrix with R columns, containing the discretised basis functions $f_r(t)$. Since the discrete cosine basis functions are designed to model low frequency effects, it has become known as the "*high-pass filter*". However, the other effects in the design matrix are and the basis functions for low frequency confounds are fitted simultaneously, rather than the low frequency effects being removed from the data prior to fitting the other effects, as the name "high-pass filter" might imply.

If the R columns of the high-pass filter are not orthogonal to the remainder of the design matrix, then some of the other effects can be modelled by the filter. For periodic designs, the respective partitions are (almost) orthogonal for a cut-off period of twice the experimental period.The extra sum-of-squares F-test assesses the additional variance modelled by the effects of interest beyond that modelled by the confounds, which includes the high-pass filter. Thus, as far as the F-test goes, the filter takes precedent over the effects of interest where there is non-orthogonality.

4.3.1 Example model

Experimental design and modelling issues for fMRI are dealt with in chapter 9, but it is useful to illustrate the concepts introduced above on a simple blocked activation experiment. 64-slice $3 \times 3 \times 3$ mm BOLD/EPI acquisitions were made every seven seconds on a single subject in a single session. 16 blocks of 6 scans (42s blocks) were used, the condition for successive blocks alternating between rest and auditory stimuli, starting with rest. During the auditory condition, the subject was presented binaurally with bi-syllabic words at a rate of 60 per minute. For various physical and psychological reasons the first few scans are atypical. Here we consider 7 blocks from the middle of the experiment, starting with a rest block.

For a single subject activation experiment of alternating epochs of baseline and activation conditions, the simplest model for the experimentally induced effects would be a mean for each condition, with conditions associated to scans reflecting the hæmodynamic lag. This would require 2 basis functions, a constant function, modelling the baseline measurement, and a box-car function of 0's (for rest epoch scans), and 1's (for active epoch scans), lagged 8s, modelling the additional signal due to the stimulus.[13] This minimalist model for the effects of interest must be augmented with appropriate basis functions for the low-frequency confounds and other confounds to give a suitable model for the fMRI time series.

For the example at hand, an example design matrix is shown in figure 16. The first column contains the lagged box car, padded at the top with a 1, indicating that the scan prior to the first rest scan considered came from a stimulus condition. The second column contains a vector of 1's, dummy variables for the baseline level. A high-pass filter cut-off of 168 s was chosen, this being twice the experimental period, giving a set of $K = 7$ discrete cosine basis functions with value $f_r(s) = cos\left(r\pi \frac{s-1}{N-1}\right)$ for scan $s = 1, \ldots, 84$, $r = 1, \ldots, K, K = 7$. These are columns 3 to 9 of the design matrix.

---

[13]A lagged box-car is an extremely crude approximation of the hæmodynamic response to a neuronal activation. Convolving a box-car with an estimated hæmodynamic response function would produce a better approximation. However, its simplicity and success have ensured its use, and it represents the
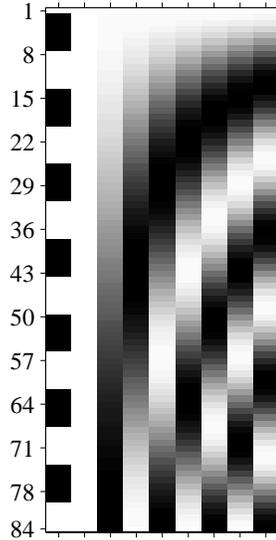
Figure 16: Design matrix X, for the example fMRI experiment described in §4.3.1.

The null hypothesis of no activation corresponds to zero amplitude of the box-car waveform. The amplitude of the box-car reference waveform is the only parameter of interest, so SPM prompts for contrast weight vectors of length one. Appropriate contrast weights are then $\underline{c}_1 = (+1)$ and $\underline{c}_1 = (-1)$ for alternative hypotheses of activation and deactivation respectively.

Voxels where the high-pass filter explains a significant amount of variation (F-test, $p < 0.001$, uncorrected, $(7, 72.87)$ d.f.) are shown in figure 17a. 8% of voxels exceed this threshold (0.1% expected), indicating significant evidence against the omnibus null hypothesis that the filter is redundant at all voxels. 0.1% of voxels are significant at $p = 0.05$ (corrected). Figure 17b shows the model fit for the voxel at $(30, -36, -14)$, where the F-statistic for the filter effects is highly significant ($p < 10^{-4}$ corrected). The upper plot shows the time series adjusted for global effects (dots), and the fitted condition effect (solid line). The lower plot shows the residuals remaining after fitting global and condition effects (dots), and the fitted low-pass filter (solid line).

5   A Multivariate perspective

The massively univariate approach outlined above involves fitting univariate linear models to the data at each voxel. Crucially, the same model form is used at each voxel, only the parameters are different. The model at voxel k is of the form:

$$\underline{Y}^k = X\underline{\beta}^k + \underline{\epsilon}$$

The values of the explanatory variables relate to each scan, the same explanatory variables are used for each voxel. Thus, the design matrix is the same for every voxel. Therefore,

---

minimalist model.
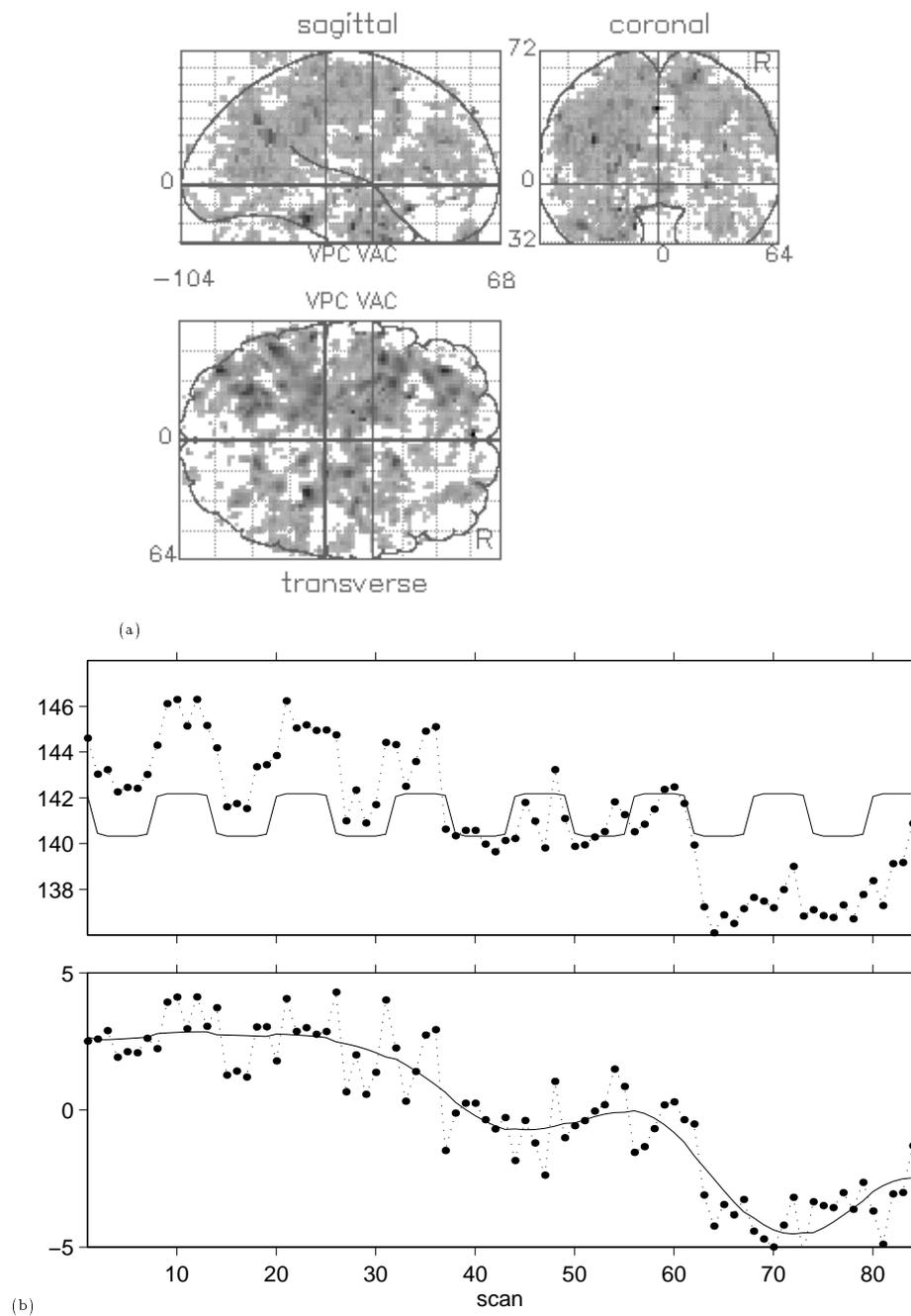
Figure 17: Illustration of the "high-pass filter". (a) Maximum intensity projection ($p < 0.001$, uncorrected) of the extra sum-of-squares F-statistic for the null hypothesis that the filter explains no additional variance beyond the box-car fit. (b) Timecourse plots for the voxel at Talairach co-ordinates $(30, \Leftrightarrow36, \Leftrightarrow14)$, illustrating the high-pass filter. See text (§4.3.1) for details.

the models for all intracerebral voxels $k = 1, \ldots, K$ can be expressed in a single matrix equation as:

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

where $Y = (\underline{Y}^1 \mid \ldots \mid \underline{Y}^k \mid \ldots \mid \underline{Y}^K)$ is the *data matrix*, arranged with scans in rows, one column per voxel. $\beta = (\underline{\beta}^1 \mid \ldots \mid \underline{\beta}^k \mid \ldots \mid \underline{\beta}^K)$ is the matrix of parameters, each column containing the parameters for a single voxel. Similarly the error matrix ϵ has ϵ^k as its k-th column. This equation therefore constitutes an *image regression*.

The least squares estimates for all the voxels may be simultaneously obtained as $\hat{\beta} = (X^\top X)^{-1} X^\top Y$.

This is the form of a multivariate regression. In multivariate regression it is usually assumed that the vectors $\underline{\epsilon}_j$ of errors associated with the j-th observation (the j-th row of the error matrix ϵ) have a k-variate multivariate normal distribution with zero mean and variance-covariance matrix Σ (Chatfield & Collins, 1980). The massively univariate approach of SPM only assumes that the marginal distributions are normal: $\epsilon_j^k \overset{iid}{\sim} \mathcal{N}(0, \sigma_k^2)$. Therefore we differentiate *image regression* from multivariate regression.

However, the random field approaches to assessing the significance of statistic images implicitly assume that the statistic images are discrete random fields. This in turn implies that the images of residual errors are Gaussian random fields, which in turn implies that the error vectors *are* multivariate normal.

Standard multivariate analyses of (raw) neuroimaging data are not possible. This is due to the fact that the number of variates (voxels) observed is much greater than the number of observations (scans). This fact results in a singular estimate of Σ, precluding most multivariate procedures, which utilise the determinant of the variance-covariance estimate. However, multivariate analyses are possible if the dimensionality of the data-set is reduced, a topic pursued in a separate chapter.

6   Miscellaneous Issues

6.1   Partitioning of the design, contrast specification

In general, SPM orders effects as (i) condition effects (PET); (ii) Other covariates of interest, either manually entered or constructed as basis functions for the effects of interest (*f*MRI); (iii) confounding subject/session (block) effects, or an explicit constant term μ; and (iv) other confounding covariates of no interest, including the global values if an ANCOVA style global normalisation is chosen, and (*f*MRI) the "high-pass filter". Effects in categories (i) & (ii) are classed as of interest, effects in categories (iii) & (iv) as of no interest.

Contrast weights are requested only for the parameters of interest, weights for other parameters are set to zero. For the designs implemented in SPM, the contrast weights corresponding to the condition effects must sum to zero within study. The package therefore ensures these weights specify a contrast by removing the mean (within study), dropping any contrasts whose weights are subsequently all zero. The exact design matrix

and contrasts used are printed by the package for inspection, and "bar charts" of contrasts are appended to design matrix "images" when presented with statistic images.

6.2   Adjusted data

Adjusted data are useful for illustrative plots of the nature of an effect, confounding effects having been removed. The partitioning of the model parameters (at each voxel) into effects of interest $(\underline{\beta}_1^k)$ and effects of no interest $(\underline{\beta}_2^k)$, corresponds to a partitioning of the design matrix $X = (X_1|X_2)$. Adjusted data are then computed for each voxel as $\underline{\tilde{Y}}^k = \underline{Y}^k \Leftrightarrow X_2\hat{\beta}_2^k$. In PET, the gCBF is entered *mean corrected* (a *centered* covariate), the adjustment is to the mean gCBF of $\overline{g}_\bullet$. For non-centered covariates, the adjustment is to a covariate value of zero, possibly implying an extrapolation of the fitted relationship way beyond the range of the covariate.

For overparameterised designs with indeterminancy between effects in both partitions, the exact form of the adjustment will depend on the constraints used (or implied) in obtaining parameter estimates. For instance, in the single subject PET activation model presented above (§3.2.1), the indeterminancy in the model corresponds to the addition / subtraction of a constant between the condition effects, designated of interest, and the overall level μ, designated as confounding. Various tactics have been employed in different versions of SPM. In SPM96, the overall mean is included in the adjusted data, alleviating the problem. This is achieved by explicitly removing the mean $\overline{Y}_\bullet^k$ from the data prior to model fitting, and subsequently adding it back to the adjusted data. In SPM97, the overall level, which isn't really interesting as has no meaning for qualitative data, is regarded as a confound. To ensure that the effects of interest do not absorb some of the overall level, the condition effects in PET are constrained to sum to zero, and the effects of interest in fMRI similarly carefully constructed. Rather than explicitly constrain the PET design matrix, leading to design matrix "images" that are potentially confusing, the condition effects are expressed relative to their mean (§2.4.3), and the Moore-Penrose pseudoinverse used to implicitly apply the constraint. This implicit constraining is exhibited in design matrix images, the effects of interest partition, X_1, having a darker background and grey patches where the indicator 1's would be.

6.3   Alternative hypotheses

Since the methods of assessing statistic images (described in the following chapter) only examine large values, the alternative hypotheses in SPM are implicitly one sided. A two-sided test can be obtained by specifying contrasts for increases and decreases (one is the negative of the other), and assessing the two resulting statistic images at half the significance level, easily carried out by doubling any p-values obtained.

6.4   Omnibus F-statistic — Any effects of interest?

An F-statistic using the "extra sum-of-squares" principle for the omnibus hypothesis that there are no effects of interest $(\mathcal{H}^k : \underline{\beta}_1^k = \underline{0})$ is computed by SPM for each voxel.

The resulting F-statistic image shows regions where there is evidence of some effect of interest. By careful partitioning of the design matrix, this F-statistic can be used to asses interesting hypotheses, particularly useful in fMRI for assessing the significance of subsets of basis functions. However, in SPM, the F-statistic for effects of interest is also used as a data reduction device.

F-statistic threshold filtering

In SPM, this F-statistic is thresholded (corresponding to a voxel level (*i.e.* uncorrected) upper tail probability of UFp under $\mathcal{H}^k$) to identify voxels for which adjusted data, model parameters and statistics are saved for subsequent display and inspection. This is purely a data reduction device. Smoothness estimators and the volume, parameters required for assessing the statistic image for significance, are computed over all intracerebral voxels.

The upper tail p-value threshold UFp is a user definable default, initially set to UF$p = 0.05$ for PET, UF$p = 0.001$ for fMRI. The implicit assumption in utilising this threshold as a data-reduction device is that there are no specific effects in voxels where there is no evidence of an overall effect. That is, for any voxel where a specific contrast is significantly different from zero at the chosen level, the extra sum-of-squares F-statistic exceeds the $100(1 \Leftrightarrow \text{UF}p)$ percentile of the appropriate F-distribution. Logically, a non-zero contrast of the actual parameters at voxel k implies $\mathcal{H}^k : \underline{\beta}_1^k = \underline{0}$ is false. However, significance tests only give "evidence against" null hypotheses at given levels. It is quite possible that a specific contrast will have t-statistic significant at $p = 0.05$ while the omnibus F-statistic is not quite significant at the same level, since the omnibus F-test attempts to account for all possible contrasts whilst maintaining test level. The more effects of interest, the more likely this will be true for some contrast.

Given that the specific contrasts are assessed at extremely high (uncorrected) significance levels, to account for the multiplicity of simultaneously assessing all voxels, it is highly unlikely that any voxels where the contrast is significant (corrected) has omnibus F-statistic failing to exceed the relatively low UFp threshold. However, low (uncorrected) significance levels may be appropriate if interest lies in specific regional hypotheses, if assessing the contrasts at a lenient level to generate hypotheses for future data-sets, or for identification of suprathreshold clusters of voxels. In these cases it is possible that the F-threshold will remove some relevant voxels, giving increased Type II errors, and a less stringent F-threshold (larger UFp) would be appropriate. UF$p = 1$ retains all voxels.

6.5 Model selection

In fitting a general linear model to a data set and then assessing questions in terms of hypotheses on the parameters of the model, it is implicitly assumed that the model fits. That is, that errors are independent and identically normally distributed with zero mean.

6.5.1 Problems of assessing model adequacy

The usual tools for assessing the adequacy of model fit for univariate linear models are predominantly graphical, for instance plots of residuals against fitted values to illustrate unmodelled structure or non-homogeneous residual variance; or, plots of residuals against the expected order statistics from a standard normal distribution (*normal scores*) to assess the validity of the normality assumption (sometimes called *Q-Q* plots). It is impractical to apply these methods routinely, given that they would have to be employed for each voxel separately. Although various diagnostic statistics can be used to summarise the information in such plots (for instance the Shapiro-Wilks test for normality, using the correlation of residuals and normal scores), the small number of observations (in PET) means such tests have extremely low power, particularly considering the multiple comparisons problem of testing at each voxel simultaneously.

6.5.2 Ill fitting models

If the model doesn't fit, for whatever reason, then the model assumptions do not hold and subsequent inference may not be valid. Consider an appropriate model, one that fits. If a reduced model is fitted (with some effects omitted), then the omitted effects contribute to the residual error, introducing structure into the residuals in defiance of their assumed independence. Usually, this results in increases in residual variance. However, the model has fewer parameters than the appropriate one, so the residual degrees of freedom are greater, implying increased confidence in the variance estimate. So, the variance (usually) goes up according to the omitted effects, but so do the degrees of freedom. Increasing variance and increasing degrees of freedom affect significance in opposite directions (for the same effect size), so in general all that can be concluded is that ensuing tests are invalid. However, the effect of increasing degrees of freedom is more pronounced for small degrees of freedom. For very low degrees of freedom the omission of terms may lead to more lenient tests despite an increase in residual variance.

If additional parameters are added to a model that is already adequate, then the (redundant) extra parameters reduce the degrees of freedom available for the estimation of residual variance. The reduced residual degrees of freedom result in reduced confidence in the variance estimates, manifest as larger p-values.

Further, it is quite possible that different models will be appropriate for data at different voxels. Thus, the pertinent approach is to chose a model for all voxels which includes all the effects relevant at any voxel.

6.5.3 Model selection

For a limited set of explanatory variables, (factors and covariates), the set of possible models for each voxel can be arranged into a hierarchy. Pairs of models at successive levels of the hierarchy differ by the inclusion/omission of a single term (factor, covariate, interaction effect, etc.). Related models in such a hierarchy can be compared using an extra sum-of-squares F-test (§2.6.4). Using this procedure various schemes for model selection can be developed. See Draper & Smith (1981) for details. The simplest are

forward selection, and *backward selection*. In the latter, the *saturated* model at the top of the hierarchy is chosen as a starting point. This model contains all plausible effects. This full model is then compared in turn with each of those in the next lower level of the model hierarchy (models derived from the current model by the omission of a single term), giving an F-statistic for the successive null hypotheses that the omitted effects are redundant in the model. If any of these comparisons is non-significant, then the terms corresponding to the lowest significance are dropped from the model. This continues until no non-significant terms are found, or until a minimal model is reached. The forward selection method is analogous.

In applying this (univariate) procedure to the problem of finding a simultaneous model for all voxels in a functional mapping experiment, there remains the problem of multiple comparisons. For each pair of models an F-statistic is obtained for each voxel, giving an F-statistic image that must be assessed. In this application, interest lies in assessing whether there is any evidence for any voxel that the larger model is required. It is not necessary to locate the particular voxels involved. Therefore, an *omnibus* test procedure with only weak control over experiment-wise Type I error is required, such as the "Sum-of-squares" or "exceedance proportion" tests described by Worsley *et al.* (1995). A backwards selection procedure using the latter test was proposed by Holmes (1994). Alternatively, the dimensionality of the data may be reduced by taking the first few principle components, and multivariate models selected on the reduced data using multivariate techniques.

6.5.4   Practical model selection

Currently, SPM has no features for formal model comparison. By careful construction of the design matrix, the default extra sum-of-squares F-statistic for the effects of interest can be used to produce an F-statistic image expressing evidence against the hypothesis that the parameters labelled of interest are superfluous. In the absence of a formal omnibus test for the whole statistic image, the F-statistic image should be interpreted leniently, using a low threshold, since it is Type II errors (false negatives) that are most important, not Type I errors (false positives).

Alternatively, a more pragmatic approach could be taken. An alternate heuristic perspective on a forwards selection procedure is this: Adding effects to a model improves the modeling of the data, removing components of the error at the expense of residual degrees of freedom. The best model provides an optimal trade off between modeling and "spending" degrees of freedom. A pragmatic approach from an investigators point of view is to start with a basic model, and successively add terms in a sensible order, while assessing the significance of a contrast of interest at each step. The "optimal" model will be the one that gives the most significant results, although the optimality is clearly from the investigators point of view! Clearly this approach is less than ideal, both statistically and practically. In particular, different models might indicate significant effects in different regions, a multiplicity which might lead to models being selected for their favourable results!

6.6 Experimental design

The most important consideration in statistical experimental design is the actual design of the experiment. This may seem circular, but frequently experiments are designed with little thought for what statistical analyses will be carried out. Statistical modelling is a component of experimental design, not a post-hoc panacea for badly thought out experiments!

A related but distinct topic is randomisation: In conducting designed experiments we wish to be able to attribute any observed effects to experimentally manipulated conditions. Usually this attribution can only be guaranteed if conditions are randomly allocated to a presentation order for each subject in a sensible manner (equivalently, experimental conditions are allocated to scan slots randomly). Further, this randomisation should be appropriately balanced, both across and within subjects. With such random allocation of conditions, any unexpected effects are randomly scattered between the conditions, and therefore only contribute to error.

6.7 Concluding remarks

The fixed effects general linear model provides a single framework for many statistical tests and models, giving great flexibility for experimental design and analysis. The SPM software package uses "image regression" to compute statistic images for pre-specified hypotheses regarding contrasts of the parameters of a general linear model fitted at each voxel. This facility presents a useful tool for functional neuroimage analysis.

References

Chatfield, C. 1983. *Statistics for Technology*. London: Chapman & Hall.

Chatfield, C., & Collins, A.J. 1980. *An Introduction to Multivariate Analysis*. London: Chapman & Hall.

Draper, N.R., & Smith, H. 1981. *Applied Regression Analysis*. 2nd edn. New York: John Wiley & Sons.

Friston, K.J., Frith, C.D., Liddle, P.F., Dolan, R.J., Lammertsma, A.A., & Frackowiak, R.S.J. 1990. The relationship between global and local changes in PET scans. *Journal of Cerebral Blood Flow and Metabolism*, **10**, 458–466.

Friston, K.J., Holmes, A.P., Poline, J-B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J., & Turner, R. 1995a. Analysis of fMRI time series revisited. *Neuroimage*, **2**, 45–53.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., & Frackowiak, R.S.J. 1995b. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, **2**, 189–210.

Healy, M.J.R. 1986. *Matrices for Statistics*. Oxford: Oxford University Press.

Holmes, A.P. 1994. *Statistical Issues in functional Neuroimaging*. Ph.D. thesis, University of Glasgow.

Mould, R.F. 1989. *Introductory Medical Statistics*. 2nd edn. London: Institute of Physics Publishing.

Satterthwaite, F.E. 1946. An approximate Distribution of Estimates of Variance Components. *Biometrics*, **2**, 110–114.

Scheffé, H. 1959. *The Analysis of Variance*. New York: Wiley.

Seber, G.A.F. 1977. *Linear Regression Analysis*. New York: Wiley.

Watson, G.S. 1955. Serial correlation in regression analysis: I. *Biometrika*, **42**, 327–341.

Winer, B.J., Brown, D.R., & Michels, K.M. 1991. *Statistical Principles in Experimental Design*. 3rd edn. New York: McGraw Hill.

Worsley, K.J., & Friston, K.J. 1995. Analysis of fMRI time-series revisited — again. *Neuroimage*, **2**, 173–181.

Worsley, K.J., Evans, A.C., Marrett, S., & Neelin, P. 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, **12**, 900–918.