Contrasts and Classical Inference

Jean-Baptiste Poline, Ferath Kherif and Will Penny

Contents

1 Introduction

The General Linear Model (GLM) characterises postulated relationships between our experimental manipulations and the observed data. These relations may consist of multiple effects all of which are contained within a specified design matrix. To test for a specific effect we use a 'contrast' which allows us to focus in on a particular characteristic of the data. The application of many different contrast vectors to the same design matrix allows us to test for multiple effects without having to re-fit the model. This is important in functional imaging because model fitting is computationally demanding.

There are often several ways to mathematically model an experimental paradigm. For example, in a functional imaging experiment the baseline condition can be modelled explicitly or not at all. This sort of issue generalises to more complex designs. Contrast specification and the interpretation of results are entirely dependent on the model specification which in turn depends on the design of the experiment. The most important step is clearly the specification of the experimental paradigm since if a design is clearly thought through, the questions asked of the data are generally easily specified.

In general, it is not very useful to know simply that a specific brain area was more active during one condition than another. We wish to know whether this difference is statistically significant. The definition of contrasts is therefore intimately related to statistical tests. We will therefore review the aspects of hypothesis testing that relate directly to the specification of contrasts.

This chapter is organised as follow. Section 2 reviews the basics beginning with some general comments about specification of contrasts and design matrices. In section 3, we review the raw material for the construction of contrasts namely the parameters of the linear model. In section 4, we describe some rules for constructing contrasts based on the t-test. In section 5 we discuss F-contrasts and in section 6 discuss the important issue of correlation between regressors.

2 Some general remarks

Thinking about which contrasts will be used should start before acquiring the data. Indeed, most of the problems concerning contrast specification come from poor design specification. Poor designs may be unclear about what the objective is or may try to answer too many questions in a single model. This often leads to a compromise in which it becomes difficult to provide clear answers to the questions of interest. This may seem obvious but still seems to be one of the main source of problems in functional imaging data analysis.

This previous remark does not completely preclude the use of a complex paradigm, in the sense that many conditions can and often should be included in the design. The process of recruiting subjects and acquiring the data is long and costly and therefore it is only natural that one would like to answer as many questions as possible with the same data. However, this requires careful

thinking about which contrasts will then be specified and whether they actually answer the question of interest.

Complex designs may also lead to testing many hypotheses at each voxel. The fact that this will increase the risk of false positives is an issue that is more often than not overlooked. Indeed, it is somewhat surprising that the problem of multiple comparisons across voxels has received a large amount of attention during the last ten years, while the problem of multiple comparisons across contrasts has not yet been addressed in the brain imaging literature. If there is no correction for the number of contrasts tested, results should be considered as exploratory (with an uncontrolled risk of error). Alternatively, Bonferroni correction can be applied to set a conservative bound on the risk of error.

One of the difficulties in brain imaging is that the form of the signal of interest is not precisely known because the haemodynamic response varies across subjects and brain regions. We therefore face a double task : estimation (what has happened ?) and detection (has anything happened ?).

3 Constructing models

3.1 What should be included in the model ?

Again, it is generally a good idea to think about how the experiment is going to be modelled and which comparison we wish to make *before* acquiring the data. We first review some of the questions often asked at the modelling step that have an impact on how the comparisons between conditions are going to be performed.

For a given experiment, different model parameterisations are usually possible and some of them may allow for an easier specification of contrasts. The specified model represents the a priori ideas about how the experimental paradigm influences the measured signal. The less prior information there is about the form of the induced signal, the larger the number of regressors required by the model, such that the combination of regressors can account for the possible signal variation.

To make this point clear, we take the example of an fMRI experiment investigating signal variation in the motor cortex when a subject is asked to press a device with four different forces (the "press" condition) interleaved with some "rest" periods.

The first question to be answered is whether the rest period should be modelled or not. In general, there is no difference whether the rest period is explicitly modelled or not, the only difference may arise because of edge effects during the convolution with the expected haemodynamic response function (for fMRI data). However, it has to be understood that the information contained in the data corresponds effectively to the difference between the conditions and the rest period. Therefore it is generally advisable not to model implicit conditions. One may think instead in terms of modelling the difference between two conditions. However, when the number of conditions is greater than 2, it

is often easier to model each condition separately and accept that there will be redundancy in the model and that only some comparisons will be valid (see section 4.2 on estimability below).

The second question that may arise is how to model the "press" condition. Here the prior information on neural activity and its metabolic consequence is essential. One may have very specific prior assumptions, for example, that the response should be linear with the force developed by the subject. In this case, one regressor representing this linear increase should be included in the model with, for instance, a value of one for the smallest force and a value of four for the greatest. In this design, we might then ask what value should this covariate take during the "rest" periods ? If zeros are assumed during this period, there is an explicit hypothesis in the model that the difference between the rest and the first force level is the same as the difference between the first and the second force level (or second and third, etc.). To relax this hypothesis, the difference between the "press" conditions and the rest period must be modelled explicitly.

3.2   Modelling the baseline

Generally the constant offset of the signal has to be modelled (as a column of ones) since the measured signal has an offset (it is not on average zero even without stimuli or task). Figure 1 shows this simple design, referred to as 'Model-1' which has three regressors[1] This is a 'linear parametric' model. Some questions that can then be put to the design would be

1. Is there a positive linear increase ? (testing if the parameter associated with the first regressor is significantly greater than zero);

2. Is there a difference between the rest and the first force level that is not the same as the one between other force levels ? In other words, is there an additive offset for the "press" condition not accounted for by the linear increase ? This would be tested using the second coefficient.

3. Is the average value of the signal during the rest period different from zero (positive or negative) ?

Note that in this first example the model could be re-parameterised, for instance by removing the mean of the first and second regressors to model only the difference around the mean (in fact, this is done automatically in SPM). In that case, the first and second questions would still be valid and the corresponding parameters be unchanged, but the interpretation of the third parameter would differ. It would take the value of the average of the measured data, therefore including a potential increase due to the first condition. On the other hand, if there is no information on whether the "press" condition might involve some negative response or not (this corresponds to the case where the

---

[1]Whilst it is usual to convolve regressors with an assumed haemodynamic response function, as described in Chapter 7, we have skipped this step so as to focus more keenly on design issues.

only a priori information is the difference between the levels of force) then it is more reasonable to remove the mean.

The above points are clarified by the following example. Suppose that the data y come from Model-1. We use parameters values [10 5 100] for the three regressors respectively. Analysis with a model in which the first two covariates have been mean centred leads to an estimation of [10 5 115] in the absence of noise. Clearly, the parameter estimates corresponding to the first two conditions of interest (now centred) are unchanged, but the parameter modelling the mean is changed (from 100 to 115). The reason is that it now includes the average of regressors one and two weighted by their respective parameter estimates.

3.3 Extending our first model

The hypothesis that the response increases linearly is a rather strong one. There are basically two ways to relax this assumption.

First, the linear increase modelled in the first covariate can be developed in a Taylor-like expansion, such that not only linear increases but also quadratic or cubic increases can be modelled. This solution leads to the inclusion of a new regressor that is simply constructed by the square of values of the linear increase regressor. This new model, Model-2, is shown in Figure 2. This is a 'quadratic-parametric model', a type of parametric modulation that is described further in Chapters 7 and 10.

The second solution is to have a non-parametric form, leaving the model completely free to capture any differences between the four force levels. This is achieved by representing each force level as a separate covariate. This example, Model-3, is shown in figure 3. This is a 'non-parametric model'. Clearly, modelling the difference between each level and the rest period renders the modelling of the average "press" versus rest condition redundant, since the average can be formed from the sum of the different force levels.

Note that what we want to model can be seen as two independent components; the differences between levels 1 and 2, levels 2 and 3, and levels 3 and 4 for one part, and the average activation over all levels of force (the main effect of force) for the other part. Note that the difference between levels 1 and 4 can be created with (1-2)+(2-3)+(3-4). Modelling differences between levels is similar to modelling interactions in factorial designs (see chapter on experimental design). We therefore have the choice here to model the difference of each level with the rest condition, or the main effect and the interaction. The questions that can be put to these two designs are exactly the same, they are just "rephrased". The two versions of this model, models 3 and 4, are shown in figures 3 (difference of each level with rest) and 4 (main effect and interactions).

The choice between parametric and non-parametric models often depends on the number of parameters that need to be modelled. If this number is large then parametric models might be preferred. A limited number of parameters (compared to the number of data points) with little prior information would generally lead to nonparametric models.

In each case, we may be unsure about how to test the effect of the force levels and which effects to test. Before answering these question more formally in the next sections, we briefly describe the issues involved. For the parametric models, we might be interested in the following questions

- Is there a linear increase or decrease in activation with force level (modelled by the first covariate) ?

- Is there a quadratic increase or decrease in *addition* to the linear variation (modelled by the second covariate) ?

- Is there anything that is either linear or quadratic in the response to the different levels of force (the joint use of the first and second covariate) ?

Should we in this instance centre the quadratic covariate or not ? The first answer to this is that it generally makes no difference. In general, only variations around the averaged signal over time are easily interpretable. Not centering this covariate would only make a difference if one were interested in the "mean" parameter. Likewise, the quadratic increase shown in Figure 2 can be decomposed into a linear increase and a "pure" quadratic increase (one decorrelated from a linear trend). Removing the component that can be explained by the linear increase from the quadratic increase makes the "linear increase" regressor parameter easier to interpret. But one has to be careful not to over-interpret a significant linear increase since even a signal equal to the second regressor in figure 2 may have a significant linear component [2].

For the nonparametric models, interesting questions might be

- Is there an overall difference between force levels and the rest condition (average difference between force levels and rest) ? This would involve the average of the first four regressors in model 3 and the first regressor in model 4.

- Are there differences between force conditions ? This is resolved by looking conjointly at all differences in force levels versus rest in model 3 and at regressors 2 to 4 in model 4.

- Would it be possible to test for a linear increase of the signal as a function of the force level ? Because any difference between condition levels has been modelled, it would not be easy to test for a *specific* linear increase. However, one can inspect the shape of the increase post-hoc by displaying the parameter estimates.

The re-parameterisation question is often framed in the following form. Should conditions A and B be modelled separately, or should the common part of A and B (A+B) be modelled as well as the difference (A-B) ? Note that if

---

[2]Since there is a common part to these two regressors, it is really a matter of interpretation whether this common part should be attributed to one or the other component. See section 6 and [1] for further discussion of this matter.

there is no third condition (or implicit condition as a null event or baseline) only (A-B) can be estimated from the data.

Another example often considered is the case where the data acquired during a period of time correspond to several factors in an experiment. For instance, consider an experiment comprising two factors, for instance, a modality factor where stimuli are presented either auditorily or visually, and a word factor where stimuli are either names, verbs or non-words. Rather than trying to model the factors (and possibly their interaction), it is often easier to model each level of the factor (here 2 by 3, yielding six conditions). If there is no further implicit or explicit rest (or baseline) then the questions of interest can be framed in terms of differences between these conditions. We return to this example in section 5.

4 Constructing and testing contrasts

4.1 Parameter estimation

We now turn to the issue of parameter estimation. As thoroughly reviewed in Chapter 7, the model considered is

$$Y = X\beta + \epsilon \tag{1}$$

simply expressing that the data Y (here Y is any time series of length n at a given location in the brain, see section A for notation), can be approximated with a linear combination of time series in X. The matrix X of dimension (n, p), therefore, contains all effects that may have an influence on the acquired signal. The quantity ϵ is additive noise and has a normal distribution with zero mean and covariance $\sigma^2 \Sigma_i$. The parameters β can then be estimated using least squares.

The most important thing to realise about model (1) is that it states that the expectation of the data Y is equal to $X\beta$. If this is not the case then the model is not appropriate and statistical results are likely to be invalid. This will occur if X does not contain all effects influencing the data, or contains too many regressors not related to the data.

A second important remark is that least squares estimation is a "good" estimation only under the assumption of normally distributed noise. This means that if there are outliers, the estimate $\hat{\beta}$ may be biased. The noise in functional imaging seems however close to normal and many aspects of the data processing stream, eg. spatial smoothing, have "normalising" properties [3]. The (true) parameters β are estimated from the data using

$$\hat{\beta} = (X^T X)^- X^T Y \tag{2}$$

where X^- denotes the (Moore-Penrose) pseudo inverse of X. The fitted data $\hat{Y}$ are defined as

$$\hat{Y} = X\hat{\beta} \tag{3}$$

---

[3] If the original noise properties are well known the most efficient way to analyse the data is the maximum likelihood procedure that would whiten the noise.

and represent what is predicted by the model. The estimated noise is

$$Y - \hat{Y} = RY = r \tag{4}$$

where

$$R = I_n - XX^- \tag{5}$$

The noise variance is estimated with

$$\hat{\sigma}^2 = Y^T R Y / \mathrm{tr}\left[R\Sigma_i\right] \tag{6}$$

Looking at formula (2) we realise that

- Parameters are dependent on the scaling chosen for the regressors in X. This scaling will not be important when the parameter estimate is compared to its standard deviation, but is important if regressors are entered "by hand" and then compared to each other in the same design. When entered through the dedicated interface in SPM , the regressors are appropriately scaled to yield sensible comparisons.

- Not all parameters may be estimable. This is the subject of the following subsection.

4.2   Estimability

One can appreciate that not all parameters may be estimable by taking the rather unlikely model that contains the same regressor twice, say x_1 and $x_2 = x_1$ having corresponding parameters β_1 and β_2. Clearly, there is no information in the data on which to base the choice of $\hat{\beta}_1$ compared to $\hat{\beta}_2$. In this specific case, any solution of the form $\hat{\beta}_1 + \hat{\beta}_2 = constant$ will provide the same fitted data, the same residuals, but an infinity of possible $\hat{\beta}_1$ and $\hat{\beta}_2$.

To generalise this argument we consider linear functions of the parameter estimates

$$\lambda_1 \hat{\beta}_1 + \lambda_2 \hat{\beta}_2 + ... = \lambda^T \hat{\beta} \tag{7}$$

The constants λ_i are the coefficients of a function that 'contrasts' the parameter estimates. The vector $\lambda^T = [\lambda_1 \lambda_2 ... \lambda_p]$, where p is the number of parameters in X, is referred to as the contrast vector. The word contrast is used for the result of the operation $\lambda^T \hat{\beta}$. A contrast is therefore a random variable, since $\hat{\beta}$ is estimated from noisy data.

This situation generalises each time a regressor can be constructed with a linear combination of the others. The matrix X is said to be rank deficient or degenerate if (some of) the parameter estimates are not unique and therefore do not convey any meaning by themselves. At first sight, this situation seems unlikely. However, especially for PET data, most design models are degenerate. This is because of the joint modelling of a constant term for the mean and of all the differences between any condition and the remaining scans.

A contrast is estimable if (and only if) the contrast vector can be written as a linear combination of the rows of X. This is because we get the information

about a contrast through combinations of the rows of Y. If no combination of of rows of X is equal to λ^T, then the contrast is not estimable[4].

In more technical terms, the contrast λ has to lie within the space of X^T, denoted by $\lambda \subset \mathcal{C}(X^T)$, or, equivalently, that λ is unchanged when projected orthogonally onto the rows of X (ie, that $P_{X^T}\lambda = \lambda$ with P_{X^T} being the 'projector' onto X^T (see appendix C).

The SPM interface ensures that any specified contrast is estimable, hence offering a protection against contrasts that would not make sense in degenerate designs. A further possible difficulty is that a contrast may be estimable but may be misinterpreted. One of the goals of this chapter is to improve the interpretation of contrasts.

4.3   Constructing and testing t-contrasts

If it is clear what the parameter estimates represent, then specification of contrasts is simple, especially in the case of t-contrasts. These contrasts are of the form described above ie. univariate linear combinations of parameter estimates. For instance, for model 1 we can ask if there is a linear increase by testing β_1 using the combination $1\beta_1 + 0\beta_2 + 0\beta_3$, that is, with the contrast vector $\lambda^T = [1\ 0\ 0]$. A linear decrease can be tested with $\lambda^T = [-1\ 0\ 0]$.

To test for the additive offset of the "press" condition, not accounted for by the linear increase, we use $\lambda^T = [0\ 1\ 0]$. Note here that the linear increase is starting with a value of one for the first force level, up to 4 for the fourth level (see figure 1).

When testing for the second regressor, *we are effectively removing that part of the signal that can be accounted for by the first regressor*. This means that the second regressor is not giving the average value of the difference between the "press" conditions and the rest condition. To obtain this, we would have to construct a re-parameterisation of model 1 and replace the first regressor so that it models *only* difference of "force levels" around an average difference between "press" and rest. This is achieved by orthogonalising the first regressor with respect to the second. This new model, model-5, is shown in figure 5. The parameter estimates of this new model are [10 30 100] as compared to [10 5 100] for model 1. This issue is detailed in [1] and the same effect can be seen in F-tests (see section 5). In other words, one should have clearly in mind not only what is but also what is *not* tested by the constructed statistics.

Another solution (useful in neuroimaging where estimating the parameters can be time consuming) is to work out the equivalent contrast (see section 6.2).

The contrast vector $\lambda^T = [1\ 1\ 0]$ is valid but difficult to interpret. For example, the individual effects may be strong but because they can have different signs the overall effect may be weak.

---

[4]Strictly, as we have described in Chapter 7, all contrasts are estimable by definition. If a linear combination of parameter estimates is not estimable then that linear combination is not a contrast. In this chapter, however, we often use the expression 'estimable contrast' for purposes of emphasis and because this term is used in the neuroimaging community.

For model 3 the average amplitude of the "press" condition compared to rest would be tested with $\lambda^T = [1\ 1\ 1\ 1\ 0]$. For model 4 the same effect can be tested with $\lambda^T = [1\ 0\ 0\ 0\ 0]$. The two contrasts give exactly the same t-maps. Note that in both cases, it is the average over levels that is tested, and this could be significant just because of the impact of one level.

An interesting question is whether we can easily test the linearity of the response to the four levels in those models. For model 3 the intuitive contrast to enter would be $\lambda^T = [1\ 2\ 3\ 4\ 0]$. This would indeed test for a linear increase of the force level, but in a very unspecific manner, in the sense that the test might be significant in a situation where only the fourth condition has a greater signal than in rest condition. This is because we are testing for the weighted sum of the related parameters. The test is therefore valid, but certainly does not ensure that the signal has a linear change with force levels. In other words, the model is flexible and we are testing a very restricted hypothesis, such that the shape of the predicted signal may be far away from the shape of the tested component.

4.4 Computing t-statistics

Whatever contrast is used, the contrast t-statistics are produced using [4, 6]

$$t_{df} = \lambda^T \hat{\beta}/\mathrm{SD}(\lambda^T \hat{\beta}) \tag{8}$$

where $\mathrm{SD}(z)$ denotes the standard deviation of z and is computed from the variance

$$\mathrm{var}\left[\lambda^T \hat{\beta}\right] = \hat{\sigma^2}\lambda^T (X^T X)^- X^T \Sigma_i X (X^T X)^- \lambda \tag{9}$$

For Gaussian errors t_{df} follows approximately a Student distribution with degrees of freedom given by $df = \mathrm{tr}\left[R\Sigma_i\right]^2/\mathrm{tr}\left[R\Sigma_i R\Sigma_i\right]$. At the voxel level, the value t_{df} is tested against the likeliness of this value under the null hypothesis.

The important point here is that the standard deviation of the contrast of parameter estimates depends on the matrix X. More specifically, when regressors are correlated, the variance of the corresponding parameter estimates increase. In other words, the stability of the estimation for one component is greater when other components included in the model are decorrelated. The dependence of the covariance of the estimated effects and the correlation within the model can be used, for instance, to optimise event related designs.

The test of t_{df} is one-tailed when testing for positivity only or negativity only and two-tailed when jointly testing for positive or negative effects (see section 5).

5 Constructing and testing F-contrasts

In this section, we will consider an experiment with two event-related conditions using the simple case of right and left motor responses. In this experiment, the

subject is asked to press a button with the right or left hand depending on a visual instruction (involving some attention). The events arrive pseudo-randomly but with a long inter-stimulus interval. We are interested in finding the brain regions that respond more to the right than to the left motor movement.

Our first model supposes that the shape of the haemodynamic response function (HRF) can be modelled by a 'canonical HRF' (see Chapter 10 for details). This model is shown in figure 6. To find the brain regions responding more to the left than to the right motor responses we can use $\lambda^T = [1 \ - \ 1 \ 0]$. Application of this contrast produces the SPM-t map shown in figure 7. This clearly shows activation of contralateral motor cortex plus other expected regions such as ipsilateral cerebellum.

Because there is an implicit baseline, the parameters are also interpretable on their own and when tested (SPM t-maps not shown), they show responses not only for the motor regions but also the expected visual regions [5]. Instead of having the two regressors being one for the left response and one for the right response, an equivalent model would have two regressors, the first modelling the response common to right and left and the second modelling the difference between these responses.

The fact that the HRF varies across brain regions and subjects can be accommodated as follows. A simple extension of the model of figure 6 is presented in figure 9, for which each response is modelled with three basis functions. These functions are able to model small variations in the delay and dispersion of the HRF, as described in Chapter 10. They are mean centred, so the mean parameter will represent the overall average of the data.

For this new model, how do we test for the effects of, for instance, the right motor response ? The most reasonable approach in the first instance is to test for all regressors modelling this response. This does not mean the sum (or average) of the parameter estimates since the sign of those parameter estimates is not interpretable, but rather the (weighted) sum of squares of those parameter estimates. The appropriate F-contrast is shown in figure 10.

One interpretation of the F-contrast is of the specification of a series of one dimensional contrasts, each of them testing against the null hypothesis that the parameter is zero. Because parameters are tested against zero one would have to reconstruct the fitted data and check the positive or negative aspects of the response.

To test for the *overall* difference between right and the left responses we use the contrast shown in figure 11. Note that multiplying the F-contrast coefficients by -1 does not change the value of the test. To see if the tested difference is "positive" or "negative" (if this makes sense since the modelled difference could be partly positive and partly negative) one has to look at the fitted signal corresponding to the extra sum of squares tested. The F-test image corresponding to this contrast is shown in figure 12. This image is very similar to the corresponding image for the simpler model (figure 8). Finally, figure 13 shows

---

[5]Interestingly, there is some ipsilateral activation in the motor cortex such that the "left-right" contrast is slightly less significant in the motor regions than the "left" [1 0 0] contrast.

that the more complex model provides a better fit to the data.

To conclude this section, we give a few more examples using the design described at the end of section 3. We suppose a 2 by 3 factorial design consisting of words presented either visually (V) or aurally (A) and belonging to 3 different categories (C1, C2, C3). The way the design is constructed is to model all the 6 event types in the following order in the design matrix; V-C1 (presented visually and in category one), V-C2, V-C3, A-C1, A-C2, A-C3. We can then test for the interaction between the modality and category factors. We suppose that the experiment is a rapid event related design with no implicit baseline, such that only comparisons between different kinds of event are meaningful (and not events in themselves). In a first example the events are modelled using only one basis function. A test for the main effect of modality would be the one presented in figure 14(a). Figure 14(b) shows the test for the main effect of categories. Note that because there is no implicit baseline here, the main effects of factors are differences between the levels. Finally, the interaction term would be tested for as in figure 14(c).

The number of rows in an interaction contrast (without implicit baseline) is given by

$$N_{rows} = \prod_{i=1}^{N}(l_i - 1) \tag{10}$$

where N is the number of factors and l_i the number of levels of factor i.

5.1 Interpretations of F-contrasts

There are two equivalent ways of thinking about F-contrasts. For example, we can think about the F-contrast in figure 10 as fitting a reduced model that does not contain the "right motor response" regressors. This reduced model would have a design matrix X_0 with zero entries where the "right motor response" regressors were in the 'full' design matrix X. The test then looks at the variance of the residuals (see section 4.1) as compared to that of the full model X. The F-test simply computes the extra sum of squares that can be accounted for by the inclusion in the model of the three "right hand" regressors. Following any statistical textbook (e.g. [2]) and the work of [6, 4], this is expressed by testing the following quantity :

$$F_{df_1,df_2} = \frac{(Y^T(I - P_{X_0})Y - Y^T(I - P_X)Y)/\nu_1}{Y^T(I - P_X)Y/\nu_2} \tag{11}$$

with

$$\nu_1 = \text{tr}\left((R_0 - R)\Sigma_i\right) \tag{12}$$
$$\nu_2 = \text{tr}\left(R\Sigma_i\right)$$

and

$$df_1 = \text{tr}\left((R_0 - R)\Sigma_i(R_0 - R)\Sigma_i\right)/tr((R_0 - R)\Sigma_i)^2 \tag{13}$$
$$df_2 = tr(R\Sigma_i R\Sigma_i)/\text{tr}\left(R\Sigma_i\right)^2 \tag{14}$$

where R_0 is the projector onto the residual space of X_0 and P_X is the orthogonal projector onto X (see appendix for a definition of a 'projector'.) We also have [6]

$$F_{df_1, df_2} \sim F(df_1, df_2) \tag{15}$$

Such a test can be implemented by specifying the columns of the design matrix that should be kept for the reduced model.

The second interpretation of the F-test is of the specification of a series of one dimensional contrasts, each of them testing against the null hypothesis that the parameter is zero. Note that in this case, parameters are tested against zero and therefore, to interpret this test one would have to reconstruct the fitted data and check the positive or negative aspects of the response.

We now formally show how the two interpretations of the F-test are linked. The model in equation (1), $Y = X\beta + \epsilon$, is restricted by the test $c^T\beta = 0$ where c is now a 'contrast matrix'. If c yields an estimable function then we can define a matrix H such that $c = H^T X$. Therefore, $H^T X\beta = 0$ which, together with equation (1), is equivalent to $Y \subset \mathcal{C}(X)$ and $Y \subset \mathcal{C}(H^\perp)$, the space orthogonal to H. It can be shown that the reduced model corresponding to this test can be chosen to be $X_0 = P_X - P_H$. This is a valid choice if and only if the space spanned by X_0 is the space defined by $\mathcal{C}(H)^\perp \bigcap \mathcal{C}(X)$ and it is easy to show that it is indeed the case.

If $\mathcal{C}(H) \subset \mathcal{C}(X)$, the numerator of equation 11 can be rewritten as

$$Y^T(R_0 - R)Y = Y^T(X_0 - R)Y = Y^T(P_X - X_0)Y = Y^T(P_H)Y \tag{16}$$

We choose H such that it satisfies the condition above with $H = (X^T)^- c$, which yields

$$
\begin{aligned}
Y^T(P_H)Y &= Y^T X (X^T X)^- X^T H (H^T H)^- H^T X (X^T X)^- X^T Y \tag{17} \\
&= \hat{\beta}^T c (H^T H)^- c^T \hat{\beta}
\end{aligned}
$$

The above rewriting of the F-test is important for several reasons. First it makes the specification and computation of F-tests feasible in the context of large data sets. Specifying a reduced model and computing the extra sum of squares using equation 11 would be too computationally demanding. Second, it helps to make the link between a t-test and the test of a reduced model and therefore helps to recall that what is tested is only the "extra" variability that cannot be explained by the reduced model. Third, it makes the test of complex interactions using F-tests more intuitive.

The F-contrast that looks at the total contribution of all the 'right regressors' is, however, quite a nonspecific test. One may have a specific hypothesis about the magnitude or the delay of the response and would like to test specifically for this. In the first instance, it can be thought that a reasonable test would be to use a t-test with contrast [0 0 0 1 0 0 0 0], testing for a positive parameter on the regressor modelling the standard HRF. This is perfectly valid, but it has to

---

[6]This is an approximate result with good properties when the number of points is not too small or if Σ_i is close to the identity matrix.

be remembered that what is tested here is the amount of adequacy of the BOLD response with this regressor, not the magnitude of the response. This means, for instance, that if the response has the shape of the one supposed by the standard model but is significantly delayed, the test might produce poor results even if the delay is appropriately taken into account by the other regressors. This might be quite important when comparing the magnitude of responses between two conditions : if this magnitude is the same but the delays are different across conditions, the test comparing simply the standard response regressors might be misinterpreted : a difference of delay might appear as a difference of magnitude *even if the basis function is decorrelated or orthogonal*. A related problem, the estimation of the delay of the HRF has been considered in an earlier chapter.

Note that the simplest F-contrasts are uni-dimensional, in which case the F-statistic is simply the square of the corresponding t-statistic. To visually differentiate between unidimensional F-contrasts and t-contrasts in the SPM interface, the former are displayed in terms of images and the latter as bars.

An important remark is that generally speaking, if we are confident about the shape of the expected response, F-tests are often less sensitive than t-tests. The reason is that the greater the flexibility of the tested space, the greater is the possibility of finding a signal that can have an important part of its variability explained by the model, and there is an implicit correction for this in the numerator of equation 11.

6   Correlation between regressors and other issues

Correlation between model regressors makes the tests more difficult to interpret. Unfortunately, such correlation is often imposed by the brain's dynamics, experimental design or external measurements. The risks of mis-interpretation have been extensively discussed in [1, 5]. To summarise, one could miss activations when testing for a given contrast of parameters if there is significant correlation with the rest of the design. An important example of this situation is when the response to a stimulus is highly correlated with a motor response.

If one believes that a region's activity will not be influenced by the motor response, then it is advisable to test this specific region by first removing from the motor response regressor all that can be explained by the stimulus. This is of course a "dangerous" action since if in fact the motor response does influence the signal in this region, then the test result will be overly significant. That is, experimental variance will be wrongly attributed to the stimulus.

6.1   Moving the variance across correlated regressors

If one decides that a regressor, or indeed several regressors or combination of those, should be orthogonalised with respect to some part of the design matrix before performing a test, it is not necessary to reparametrise and fit the model again. Once the model has been fitted, all the information needed to test some

effects can be found in the fitted parameter estimates. For instance, instead of testing the additional variance explained by a regressor, one may wish to test for all the variance that can be explained by this regressor. Using the definitions given in section 5.1, if c is the contrast testing for the extra sum of squares, it is easy to show that the contrast matrix

$$c_{Full\_space} = X^T X c \qquad (18)$$

tests for all the variance explained by the subspace of X defined by Xc since we then have $H = Xc$.

6.2 Contrasts and reparametrised models

The above principle can be generalised as follows. If the design matrix contains three subspaces say (S_1, S_2, S_3), one may wish to test for what is in S_1, having removed what could be explained by S_2 (but not by S_3). Other examples are conjunction analyses in which a series of contrasts can be modified such that the effects they test are orthogonal. This involves orthogonalising the subsequent subspaces tested. Therefore, results may differ depending on the order in which these contrasts are entered.

The principle to compute the contrast testing for various subspaces from parametrised versions of the model is simple. If X and X_p are two differently parametrised versions of the same model then we can define a matrix T such that $X_p = XT$. If c_p is a test expressed in X_p while the data have been fitted using X, the equivalent of c_p using the parameter estimates of X is

$$c = c_p(T^T X^T X T)^- T^T X^T X \qquad (19)$$

One should be careful using this sort of transformation, for instance, not putting variance of "no interest" in the space tested. Such transformations are often very useful, however, as the models do not require re-fitting. See [1] for further examples.

6.3 The estimation-detection dilemma

A crucial point when analysing functional imaging data (PET and fMRI alike) is that in general the model is not well known. The larger the model, the better in general would be the estimation of the signal, as long as the model is not starting to capture noise components. This often leads to less specific questions and to less sensitive tests compared to situations where the difference is known with better precision. There are two extreme choices:

- the use of a simple model with the danger of not having modelled some effects properly, a situation that may lead to biased results

- the use of a very flexible model with less sensitive tests and difficulties in the interpretation of the results

In other words, it is difficult to estimate the signal and at the same time test for this signal. A possible strategy would consist in using part of the data in an estimation phase that is separate from a testing phase. This will, however, involve "losing" some data. For an instance of such a strategy see [3].

6.4   FIR and random effects analyses

A typical example of a flexible model is Finite Impulse Response (FIR) modelling of event-related fMRI data, as described in Chapter 10. The model in this case is as flexible as possible since the haemodynamic response function is allowed to take any shape. A classic difficulty with this approach, however, is how to implement a random effects analysis (see Chapter 12). The difficulty arises because, usually, one takes a single contrast image per subject up to the second-level. With an FIR characterisation, however, one has multiple parameter estimates per subject and one must therefore take into account the covariance structure between parameters. Whilst this was prohibited in SPM99 it is now possible in SPM2, by making use of the 'non-sphericity' options as alluded to in Chapters 9, 12 and 17.

7   Summary

In a functional imaging experiment it is often the case that one is interested in many sorts of effects eg. the main effects of various conditions and the possible interactions between them. To investigate each of these effects one could fit several different GLMs and test hypotheses by looking at individual parameter estimates. Because functional imaging data sets are so large, however, this approach is impractical. A more expedient approach is to fit larger models and test for specific effects using specific contrasts.

In this chapter we have shown how specification of the design matrix is intimately related to the specification of contrasts. For example, it is often the case that main effects and interactions can be set up using parametric or nonparametric designs. These different designs lead to the use of different contrasts. Parametric approaches are favoured for factorial designs with many levels per factor. For contrasts to be interpretable they must be estimable and we have described the conditions for estimability.

In fMRI one can model haemodynamic responses using the canonical HRF. This allows one to test for activations using t-contrasts. To account for the variability in haemodynamic response across subjects and brain regions one can model the HRF using the canonical HRF plus its derivatives with respect to time and dispersion. Inferences about differences in activation can then be made using F-contrasts. We have shown that there are two equivalent ways of interpreting F-contrasts, one employing the extra-sum-of-squares principle to compare the model and a reduced model and one specifying a series of one-dimensional contrasts. Designs with correlation between regressors are less efficient and correlation can be removed by orthogonalising one effect with respect

to other effects. Finally, we have shown how such orthogonalisation can be applied retrospectively, ie. without having to refit the models.

A   Notation

Y	:	Data	The $(n, 1)$ time series, where n is the number of time points or scans.
c or λ	:	Contrast	weights of the parameter estimates used to form the (numerator) of the statistics
X	:	Design matrix or design model	the (n, p) matrix of regressors .
β	:	Model parameters	The true (unobservable) coefficients such that the weighted sum of the regressors is the expectation of our data (if X is correct)
$\hat{\beta}$	:	Parameter estimates	The computed estimation of the β using the data Y : $\hat{\beta} = (X^T X)X \, Y$
R	:	Residual forming matrix	Given a model X, the residual forming matrix $R = I_n - XX^-$ transforms the data Y into the residuals $r = RY$.
$\sigma^2 \Sigma_i$	:	scan (time) covariance	This (n, n) matrix describes the (noise) covariance between scans

B   Subspaces

Let us consider a set of p vectors x_i of dimension $(n, 1)$ (with $p < n$), such as regressors in fMRI. The space spanned by this set of vectors is formed of all possible vectors (say u) that can be expressed as a linear combination of the x_i : $u = \alpha_1 x_1 + \alpha_2 x_2 + ...\alpha_p x_p$. If the matrix X is formed with the x_i : $X = [x_1 x_2 ... x_p]$, we note this space as $\mathcal{C}(X)$.

Not all the x_i may be necessary to form $\mathcal{C}(X)$. The minimal number needed is called the rank of the matrix X. If only a subset of the x_i are selected, say that they form the smaller matrix X_0, the space spanned by X_0, $\mathcal{C}(X_0)$ is called a subspace of X. A contrast defines two subspaces of the design matrix X : one that is tested and one of "no interest", corresponding to the reduced model.

C   Orthogonal projection

The orthogonal projection of a vector x onto the space of a matrix A is the vector (for instance a time series) that is the closest to what can be predicted by linear combinations of the columns of A. The closest here is in the sense of a minimal sum of square errors. The projector onto A, denoted P_A, is unique and can be computed with $P_A = AA^-$, with A^- denoting the Moore-Penrose pseudo inverse[7] of A. For instance, in section 4.1, the fitted data $\hat{Y}$ can be computed with

$$\hat{Y} = P_X Y = XX^- Y = X(X^T X)^- XY = X\hat{\beta} \qquad (20)$$

---

[7]Any generalised inverse could be used.

Most of the operations needed when working with linear models only involve computations in the parameter space, as is shown in equation 17. For a further gain, if the design is degenerate, one can work in an orthonormal basis of the space of X. This is how the SPM code is implemented.

References

[1] A. Andrade, A.-L. Paradis, S. Rouquette, and J.-B. Poline. Ambiguous results in functional neuroimaging data analysis due to covariate correlation. *Neuroimage*, 10:483–486, 1999.

[2] Ronald Christensen. *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer, 1996.

[3] Kherif F., J.-B. Poline, F. Flandin, Benali H., S. Dehaene, and K.J. Worsley. Multivariate model specification for fMRI data. *Neuroimage*, 2002.

[4] K.J. Friston, A.P. Holmes, K.J. Worsley, J.-B. Poline, C.D. Frith, and R.S.J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

[5] Ashish Sen and Muni Srivastava. *Regression Analysis - Theory, Methods, and Applications*. Springer-Verlag, 1990.

[6] K. J. Worsley and K. J. Friston. Analysis of fMRI time-series revisited - again. *NeuroImage*, 2:173–181, 1995.
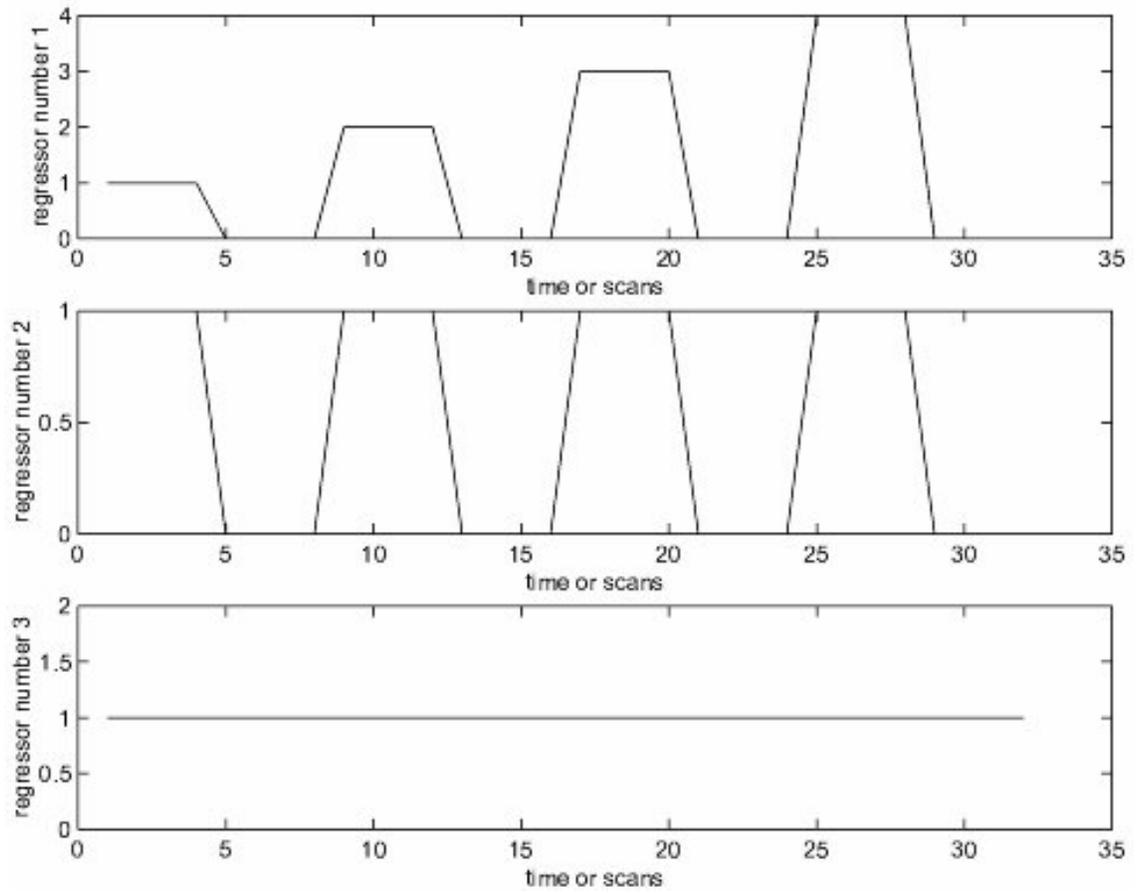
Figure 1: **Model-1**: Simple linear increase design showing the three regressors constituting the design matrix for the experiment described in section 3.2. The regressors, from top to bottom, model (i) the effects of a linear increase in force, (ii) the effect of force itself and (iii) the baseline response.
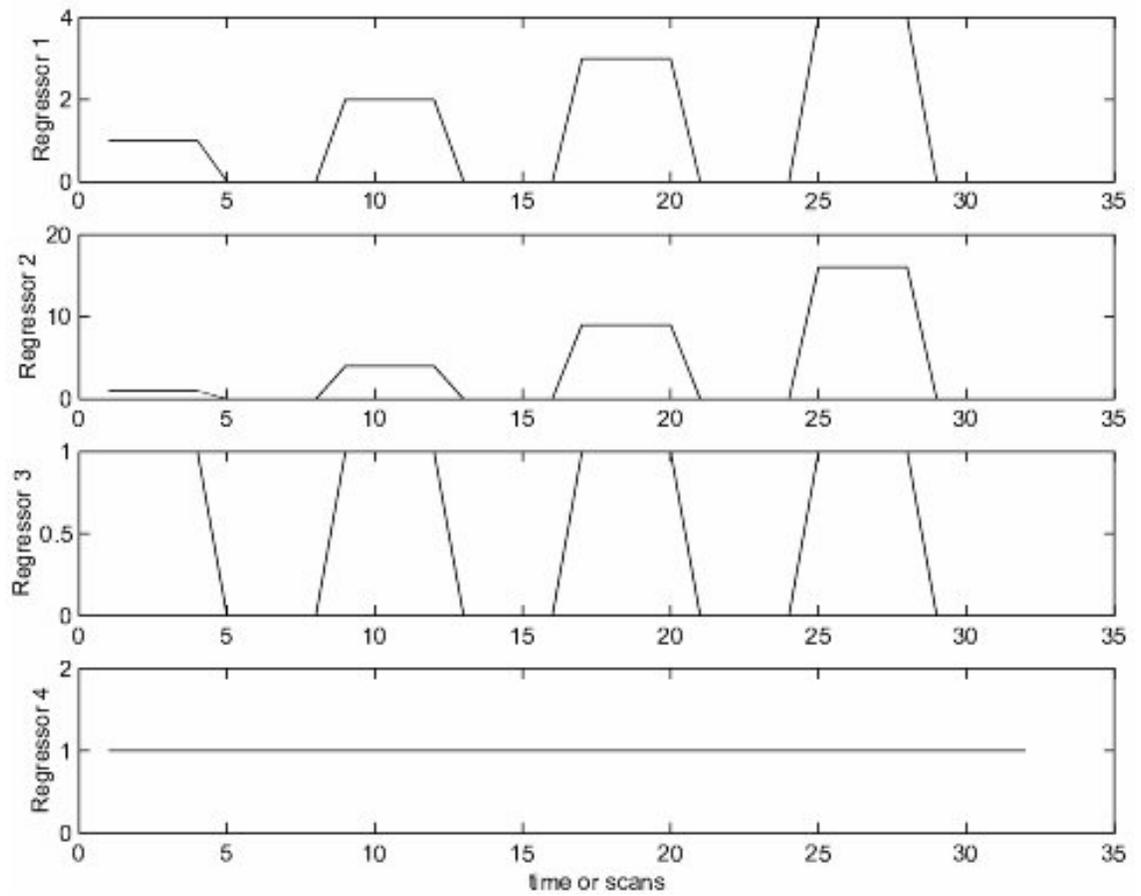
Figure 2: **Model-2**: Linear and quadratic increase covariates. Note the scale of the second covariate.
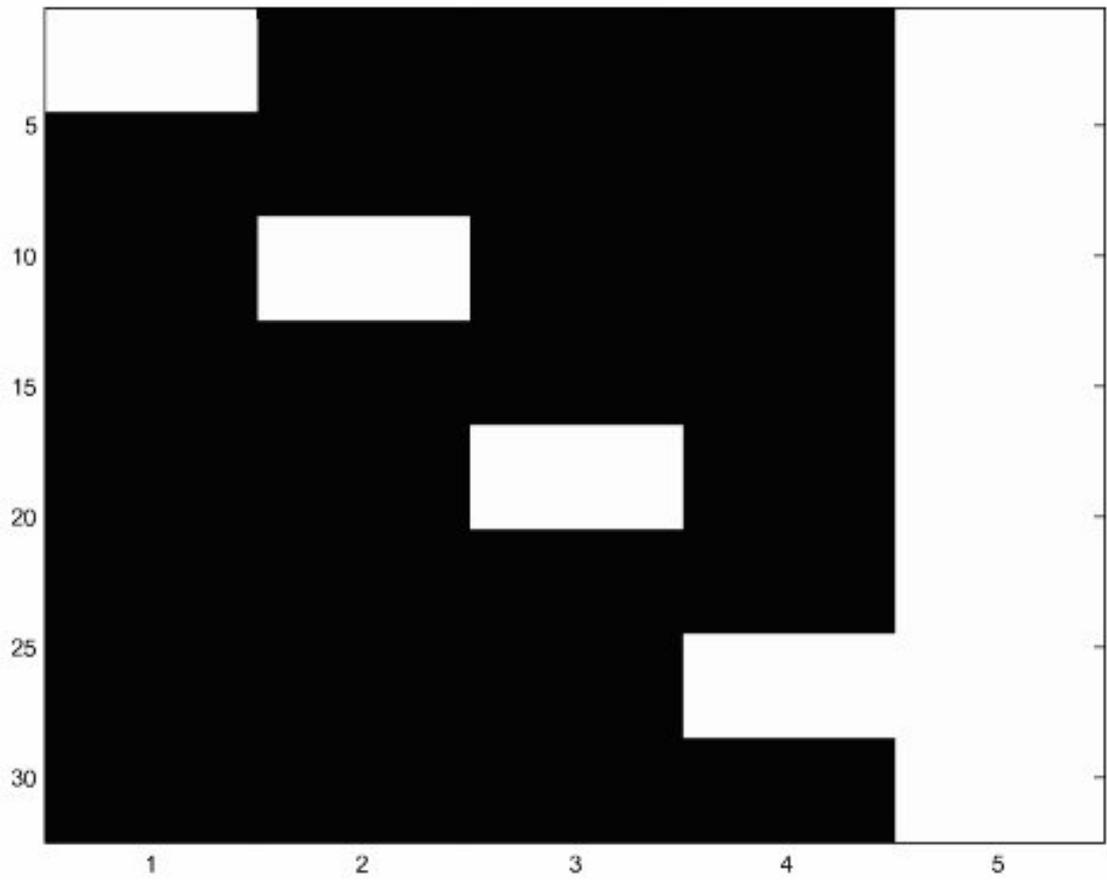
Figure 3: **Model-3** Different force levels are modelled using separate covariates. Black is 0 and white is 1 on this panel.
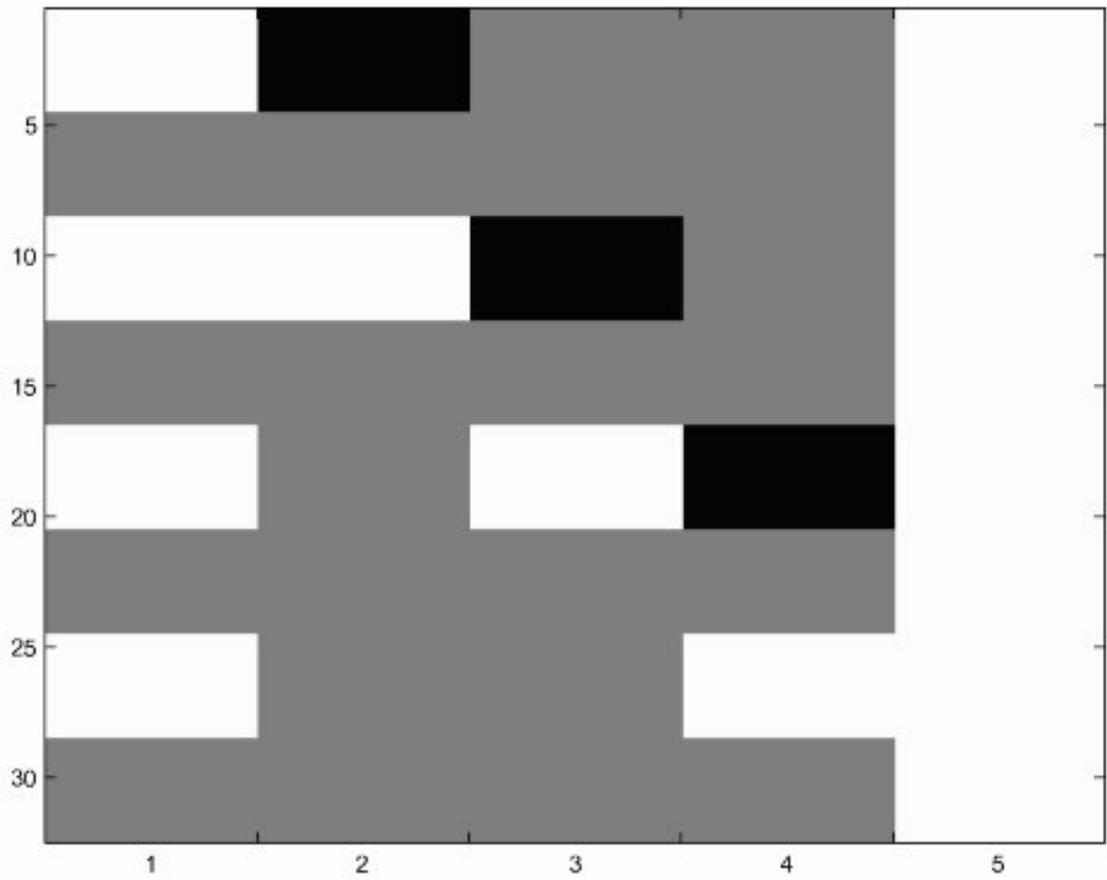
Figure 4: **Model-4** The main effect of force is modelled with the first regressor and the interactions are modelled with regressors 2 to 4
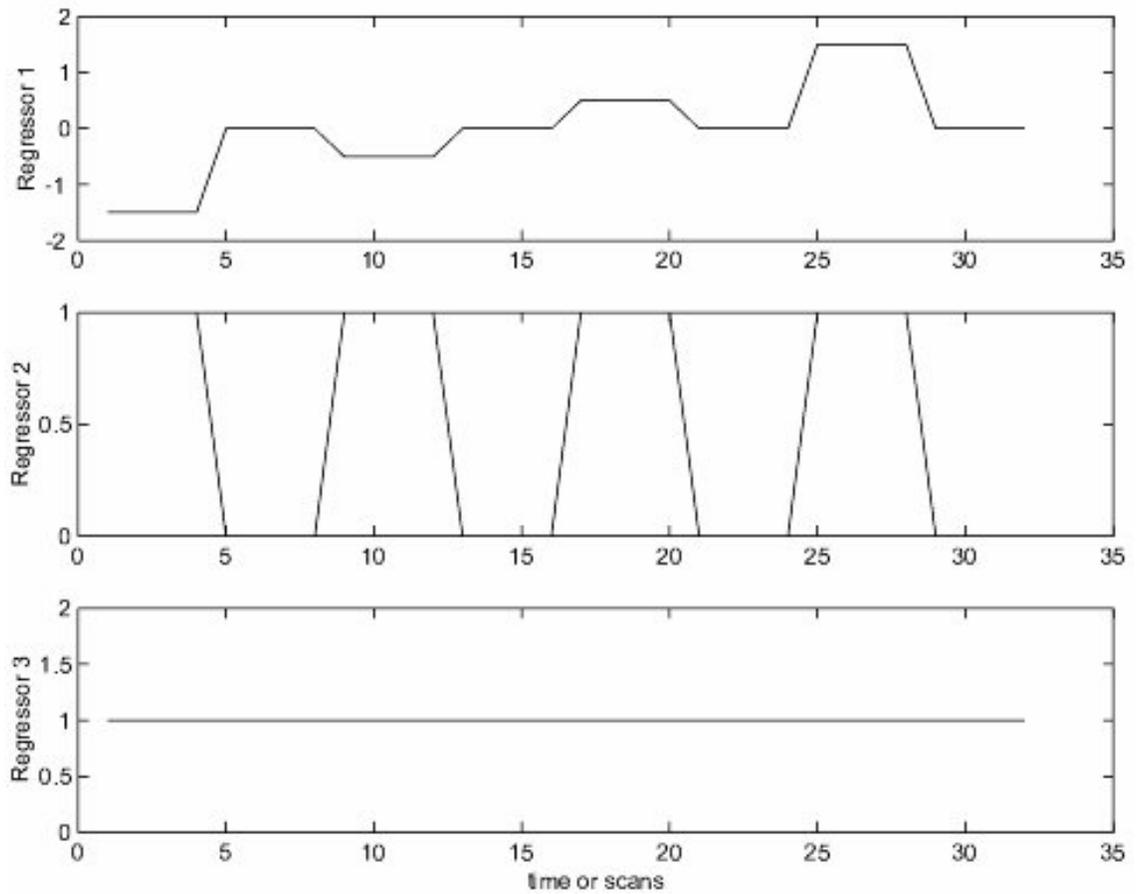
Figure 5: **Model-5** This is the same as model 1 but the main effect of force has been removed from the first regressor. This changes the interpretation of the second regressor.

23
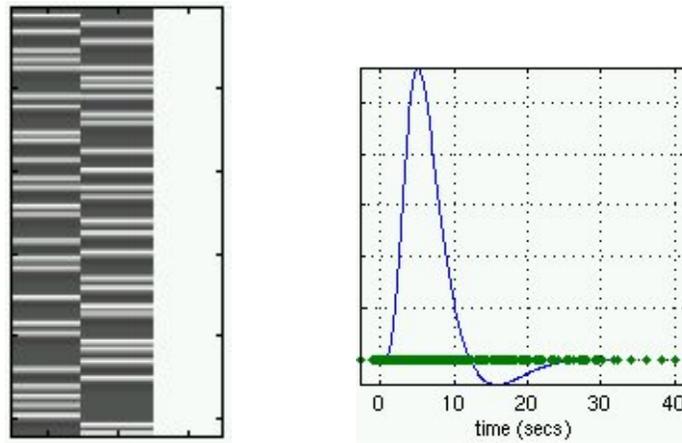
Figure 6: The left panel shows the design matrix for analysing two event related conditions (left or right motor responses) versus an implicit baseline in the case where the shape of the HRF is assumed known up to a scaling factor. This 'canonical' HRF is shown on the right panel. This HRF has been convolved with the series of Dirac functions, which occur at event onset times, to form the two leftmost regressors in the design matrix on the left.
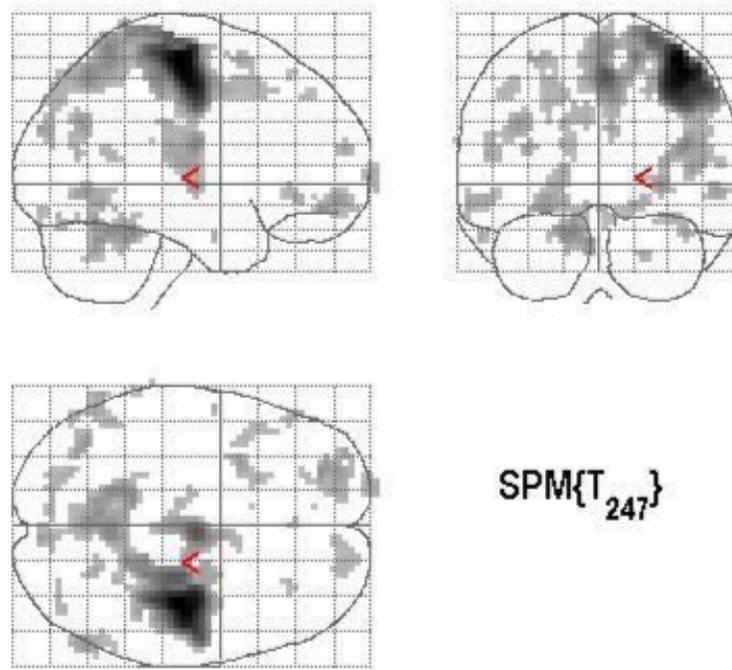
Figure 7: SPM-t image corresponding to the overall difference between the left and right responses. This map was produced using the $[1 \ -1 \ 0]$ contrast on the design matrix shown in figure 6.

SPM{F$_{[1,247]}$}
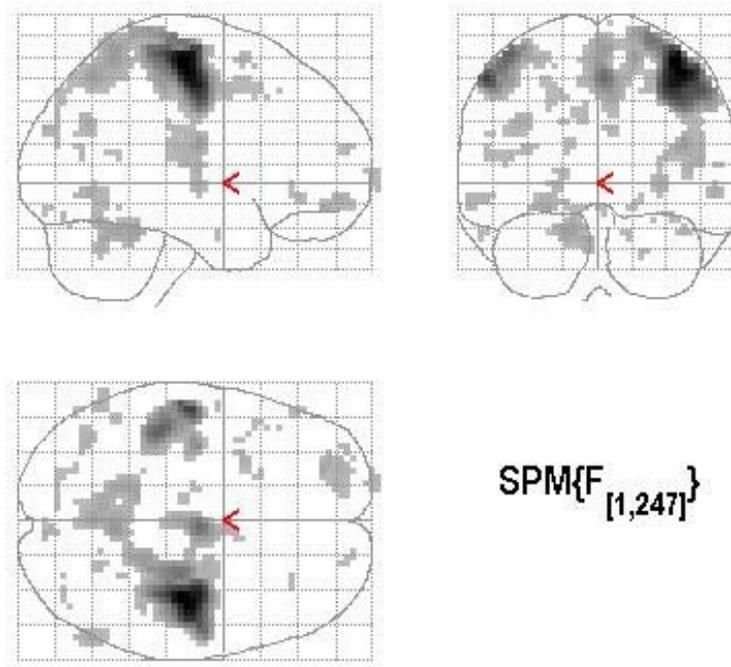
Figure 8: SPM-F image corresponding to the overall difference (positive or negative) from the left and right responses. This map was produced using the F-contrast [1 0 0; 0 1 0] on the design matrix shown in figure 6.
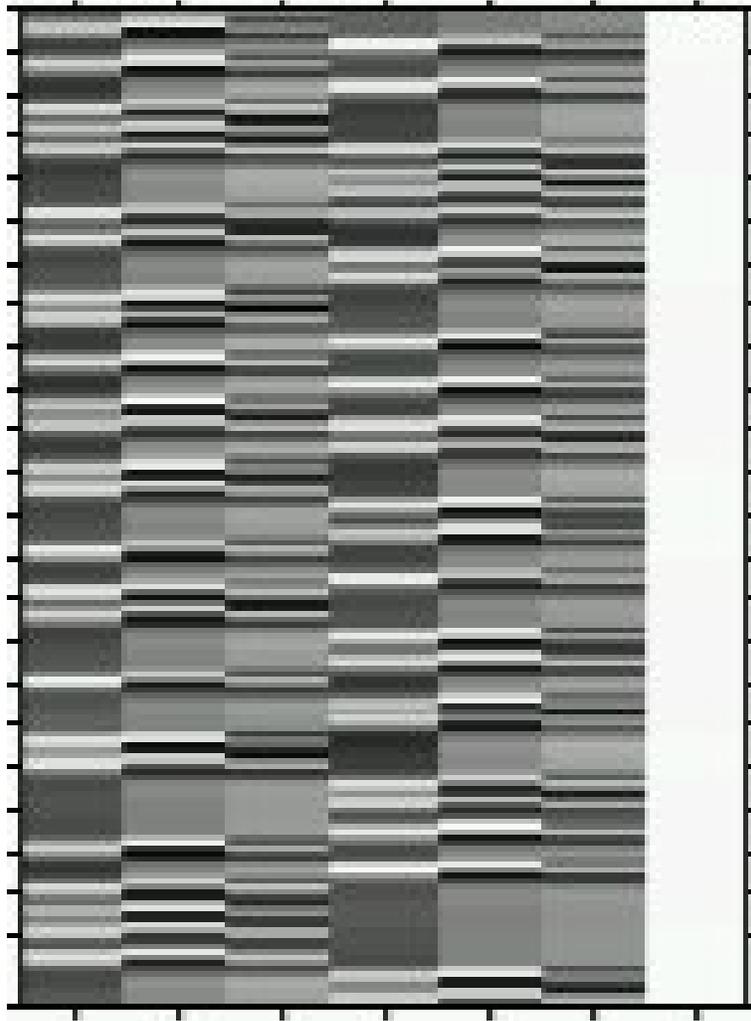
Figure 9: This is the same model as in figure 6, but we use 3 regressors to model each condition. The first three columns model the first condition (left motor response) while columns 4 to 6 model the second condition (right motor response). The three basis functions are the canonical HRF and its derivatives with respect to time and dispersion.
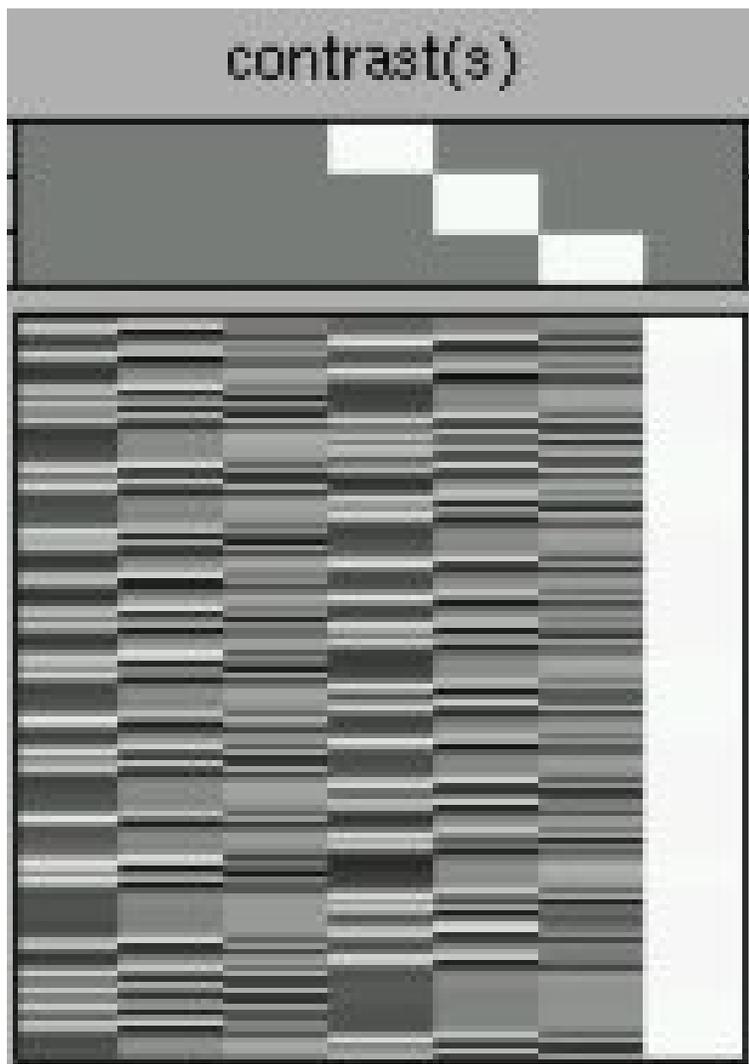
Figure 10: This figure shows a simple 'F-contrast' testing for the regressors modelling the right motor response. This corresponds to constructing the reduced model that does not contain the regressors that are "marked" with the F-contrast.
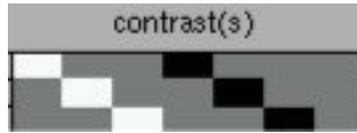
Figure 11: This figure shows the F-contrast used to test the overall difference (across basis functions) between the left and right responses.
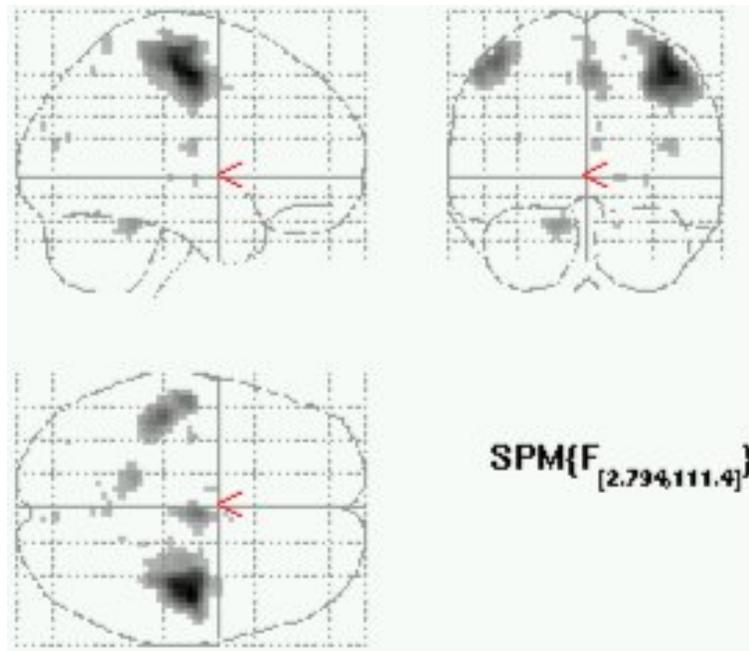

Figure 12: SPM-F image corresponding to the overall difference between the left and right responses. This map was produced using the F-contrast in figure 11 applied to the design matrix in figure 9.
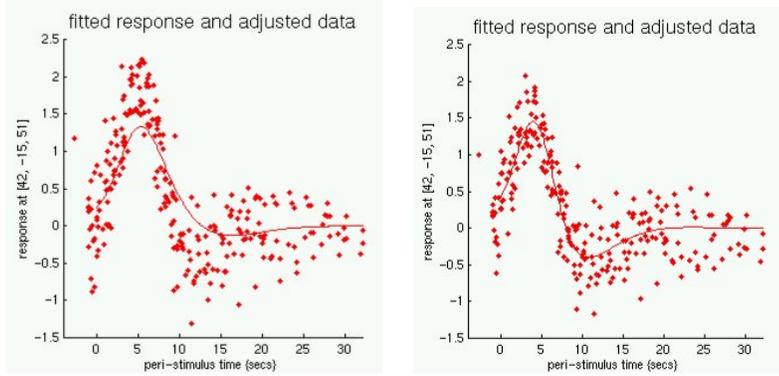
Figure 13: These plots show the haemodynamic response at a single voxel (the maxima of the SPM-F map in figure 12). The left plot shows the HRF as estimated using the simple model in figure 6 and demonstrates a certain lack of fit. This lack of fit is corrected (right panel) by using the more flexible model of figure 9.
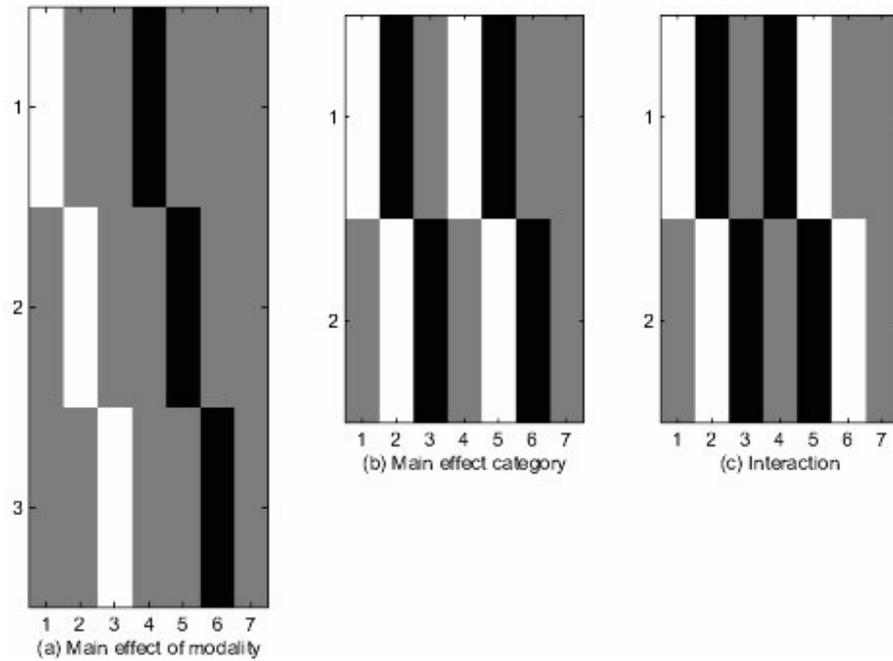

Figure 14: F-contrasts testing respectively for (a) the main effect of modality (b) the main effect of categories and (c) the interaction modality × category.