The general linear model and Statistical Parametric Mapping
I: Introduction to the GLM

Alexa Morcom

and Stefan Kiebel, Rik Henson, Andrew Holmes & J-B Poline

Overview

- Introduction
- Essential concepts
  - Modelling
  - Design matrix
  - Parameter estimates
  - Simple contrasts
- Summary

Some terminology

- SPM is based on a **mass univariate approach** that fits a model at each voxel
  - Is there an effect at location X? Investigate localisation of function or **functional specialisation**
  - How does region X interact with Y and Z? Investigate behaviour of networks or **functional integration**
- A **General(ised) Linear Model**
  - Effects are linear and additive
  - If errors are normal (Gaussian), **General** (SPM99)
  - If errors are not normal, **Generalised** (SPM2)

Classical statistics...

- **Parametric**
  - one sample *t*-test
  - two sample *t*-test
  - paired *t*-test
  - ANOVA
  - ANCOVA
  - correlation
  - linear regression
  - multiple regression
  - *F*-tests
  - etc…

all cases of the (**univariate**)
General Linear Model
Or, with non-normal errors, the
Generalised Linear Model

Classical statistics...

- **Parametric**
  - one sample *t*-test
  - two sample *t*-test
  - paired *t*-test
  - ANOVA
  - ANCOVA
  - correlation
  - linear regression
  - **multiple regression**
  - *F*-tests
  - etc…

all cases of the (**univariate**) **General Linear Model**

Or, with non-normal errors, the **Generalised Linear Model**

Classical statistics...

- **Parametric**
  - one sample *t*-test
  - two sample *t*-test
  - paired *t*-test
  - ANOVA
  - ANCOVA
  - correlation
  - linear regression
  - **multiple regression**
  - *F*-tests
  - etc…

all cases of the (**univariate**) **General Linear Model**

Or, with non-normal errors, the **Generalised Linear Model**

- **Multivariate?**

® **PCA/ SVD, MLM**

Classical statistics...

- **Parametric**
  - one sample *t*-test
  - two sample *t*-test
  - paired *t*-test
  - ANOVA
  - ANCOVA
  - correlation
  - linear regression
  - **multiple regression**
  - *F*-tests
  - etc…

all cases of the (**univariate**) **General Linear Model**

Or, with non-normal errors, the **Generalised Linear Model**

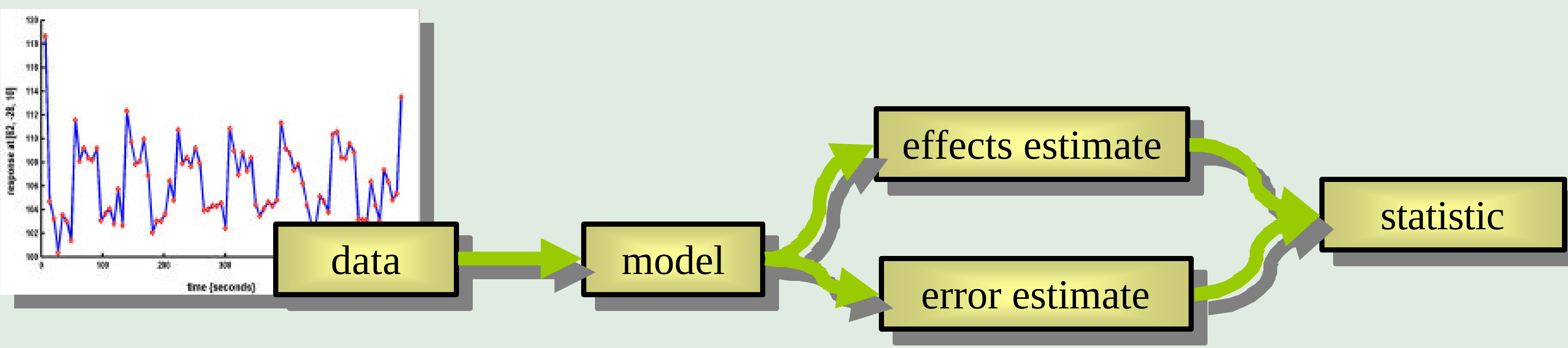
- **Multivariate?** ® **PCA/ SVD, MLM**
- **Non-parametric?** ® **SnPM**

Why modelling?

Why? — Make inferences about effects of interest

How?
1. Decompose data into effects and error
2. Form statistic using estimates of effects and error

Model? — Use any available knowledge



Choose your model

"All models are wrong, but some are useful"
George E.P.Box

A very general model

Variance

Bias

default SPM

No normalisation

No smoothing

Massive model

Captures signal

but ... not much sensitivity

Lots of normalisation

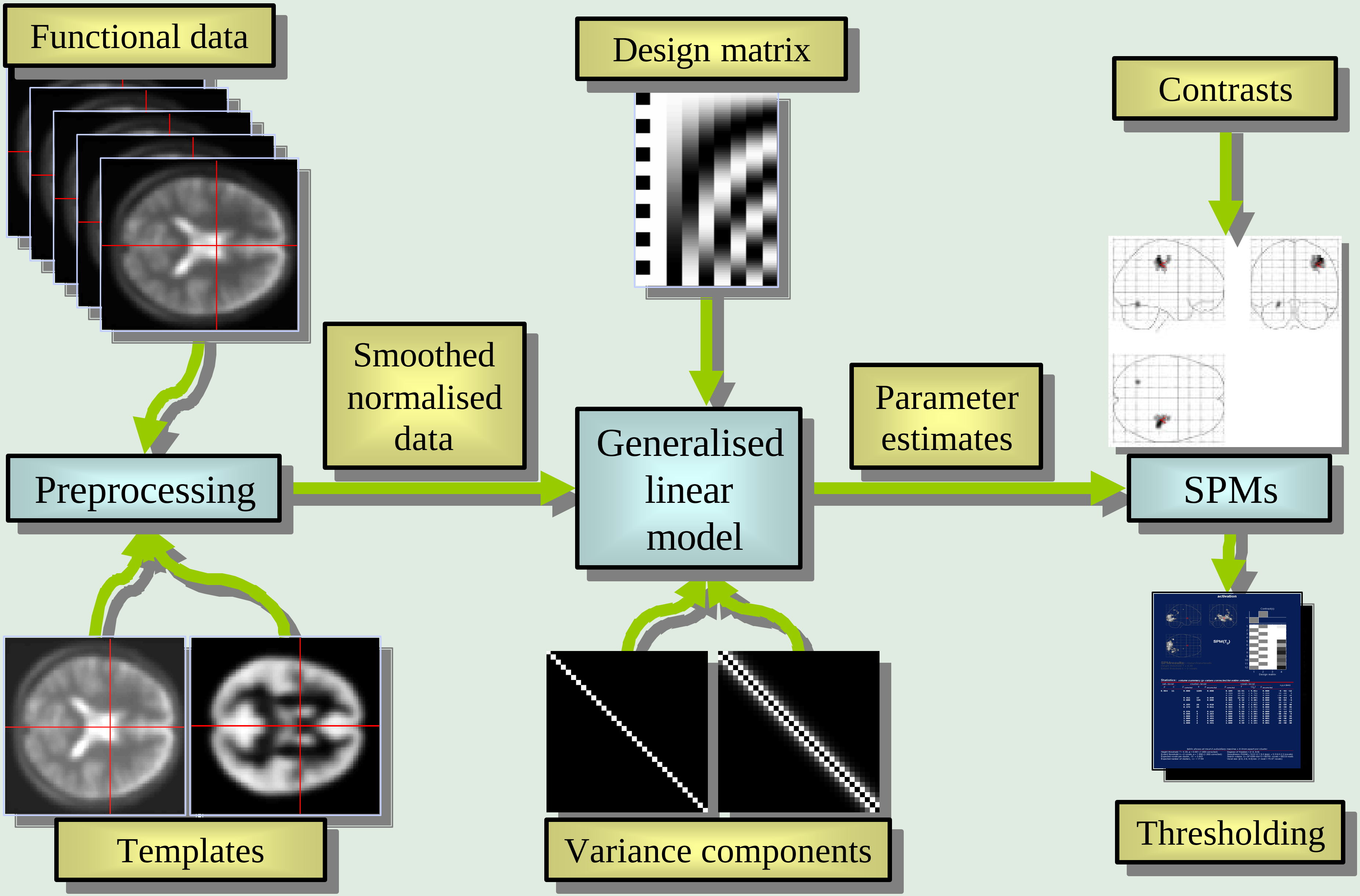
Lots of smoothing

Sparse model

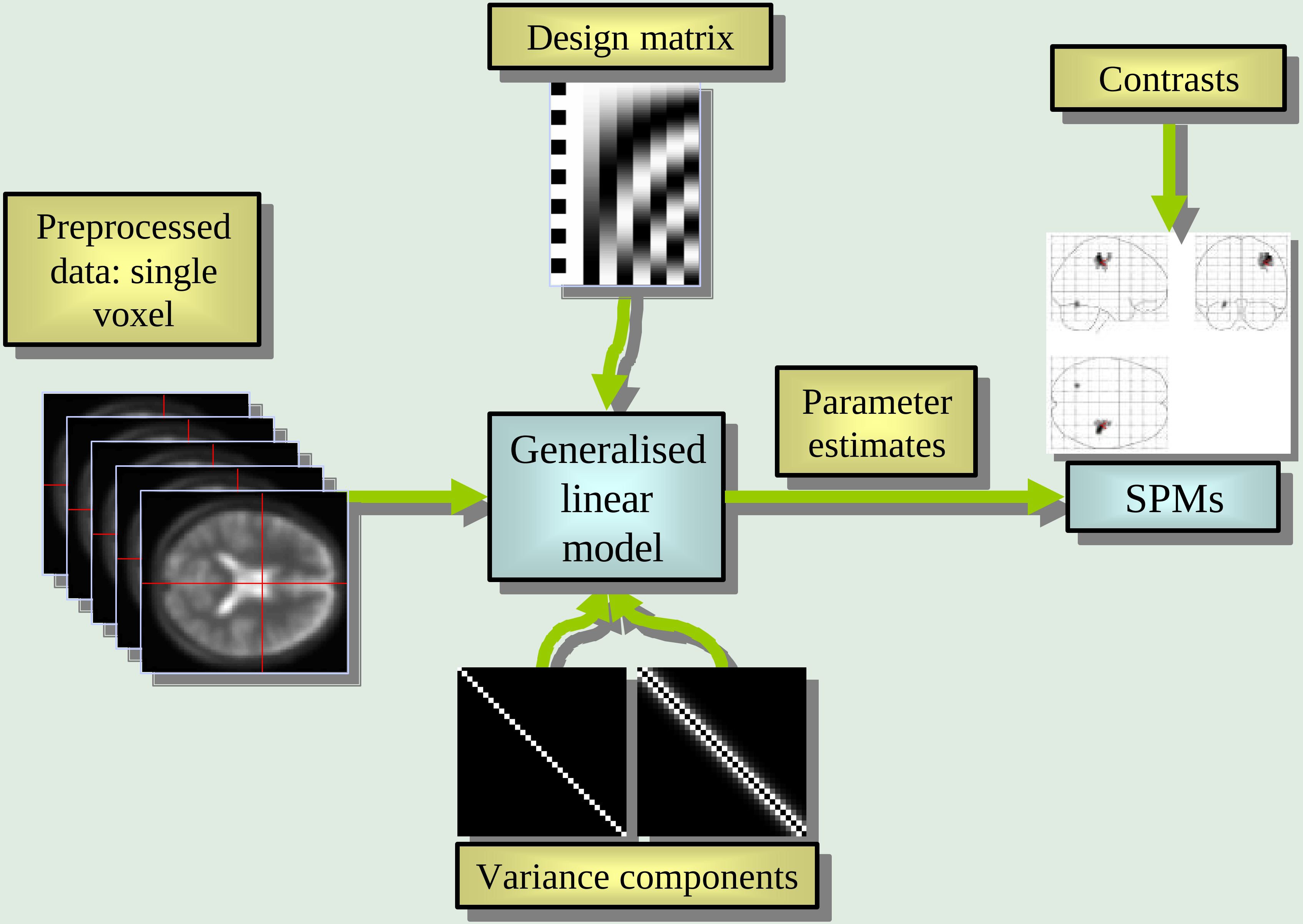
High sensitivity

but ... may miss signal

Modelling with SPM



Modelling with SPM



fMRI example

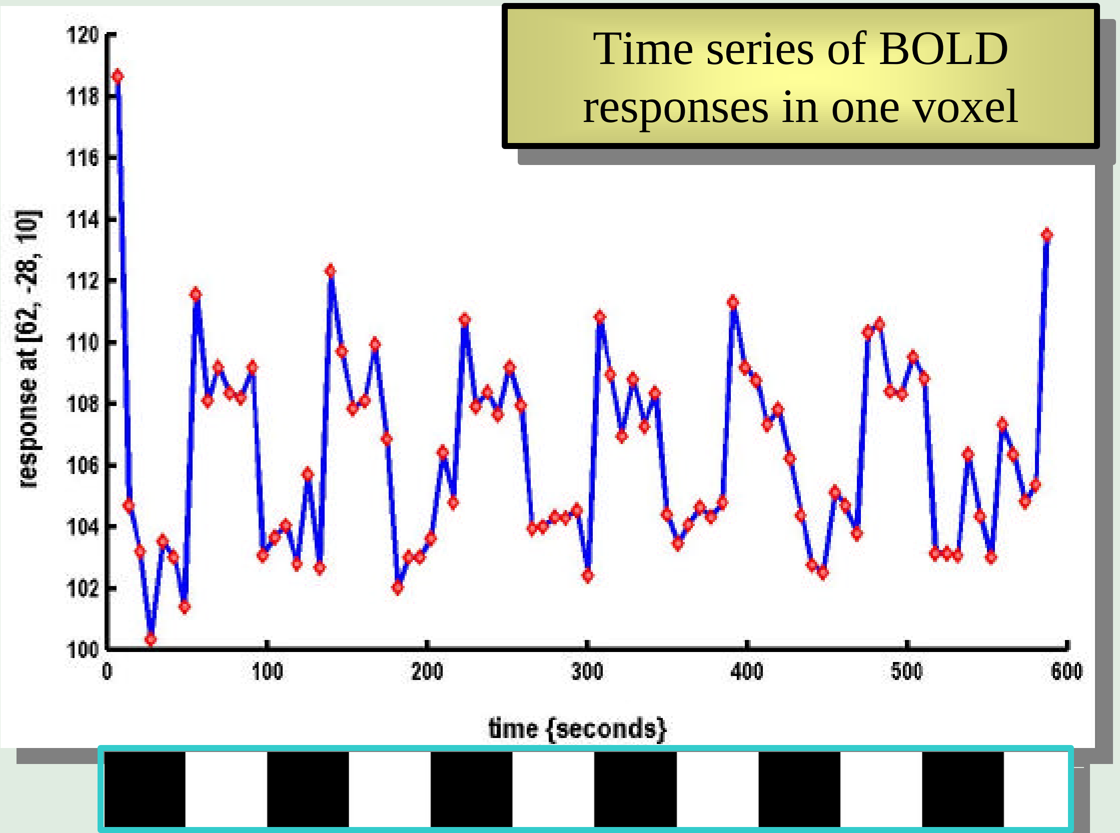
One session

Passive word listening versus rest

7 cycles of rest and listening

Each epoch 6 scans with 7 sec TR

Time series of BOLD responses in one voxel



Stimulus function

Question: Is there a change in the BOLD response between listening and rest?

GLM essentials

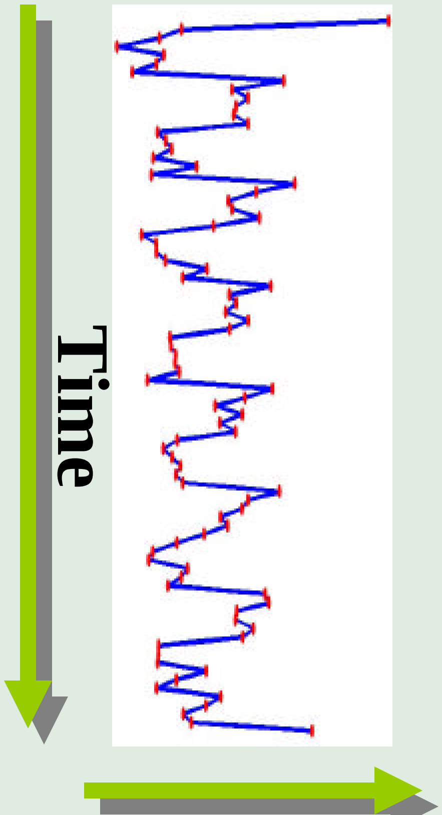
- The model
  - Design matrix: Effects of interest
  - Design matrix: Confounds (aka effects of no interest)
  - Residuals (error measures of the whole model)
- Estimate effects and error for data
  - Parameter estimates (aka betas)
  - Quantify specific effects using contrasts of parameter estimates
- Statistic
  - Compare estimated effects – the contrasts – with appropriate error measures
  - Are the effects surprisingly large?

GLM essentials

- **The model**
  - **Design matrix: Effects of interest**
  - **Design matrix: Confounds (aka effects of no interest)**
  - **Residuals (error measures of the whole model)**
- Estimate effects and error for data
  - Parameter estimates (aka betas)
  - Quantify specific effects using contrasts of parameter estimates
- Statistic
  - Compare estimated effects – the contrasts – with appropriate error measures
  - Are the effects surprisingly large?

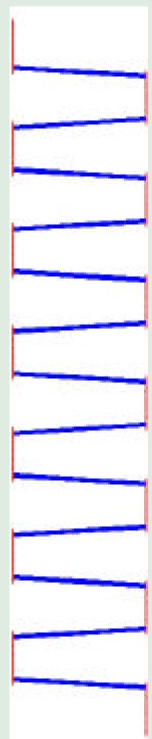
Regression model

General case

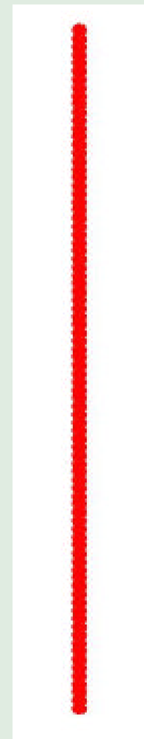


Intensity

$=\ b_1$



$+\ b_2$



$+$

x_1 x_2 e

$$e \sim N(0,\ s^2 I)$$

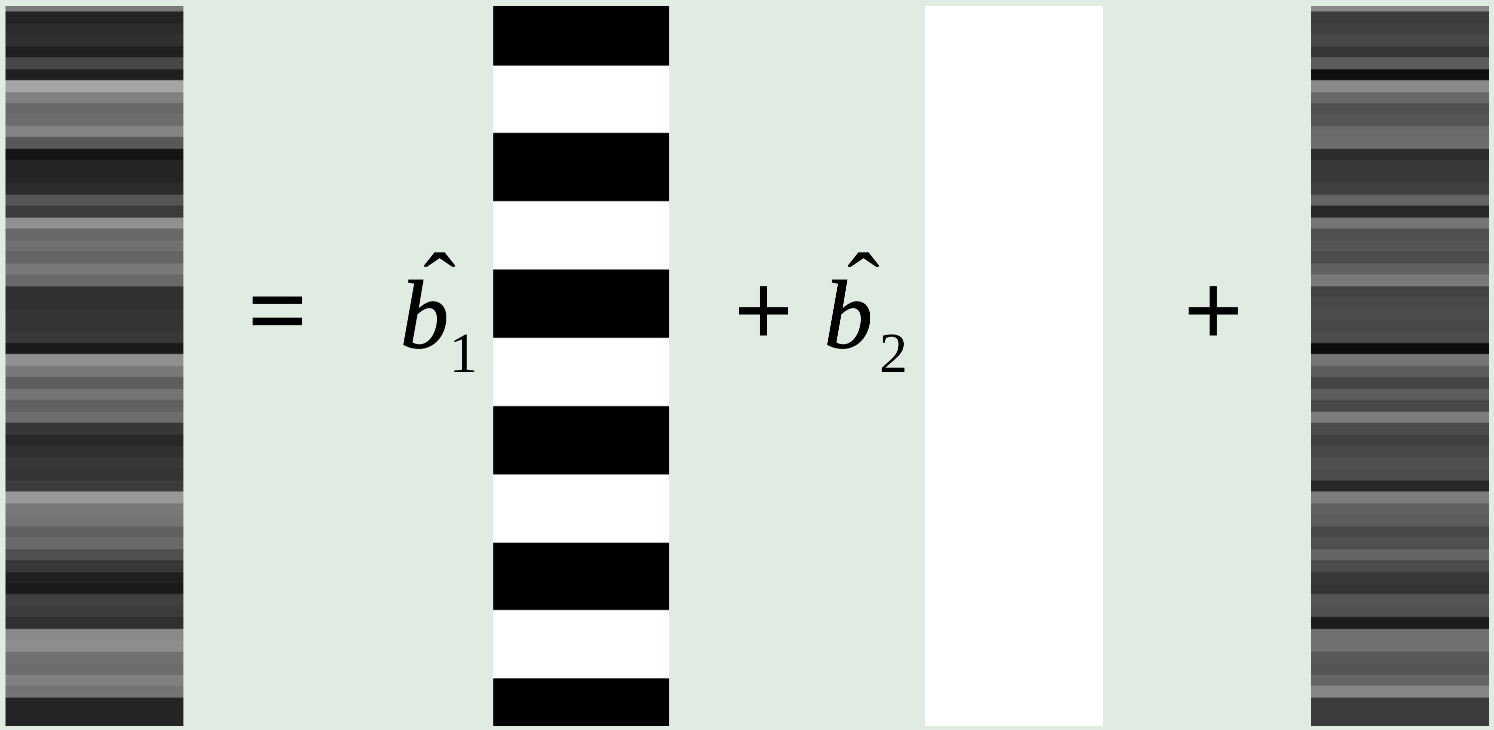
(error is normal and independently and identically distributed)

Question: Is there a change in the BOLD response between listening and rest?

Hypothesis test: $\beta_1 = 0$? (using t-statistic)

Regression model

General case



$$Y = \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{e}$$

Model is specified by

1. Design matrix X
2. Assumptions about ε

Matrix formulation

$$Y_i = b_1\, x_i + b_2 + e_i \quad i = 1,2,3$$

$$Y_1 = b_1\, x_1 + b_2 \times 1 + e_1$$

$$Y_2 = b_1\, x_2 + b_2 \times 1 + e_2$$

$$Y_3 = b_1\, x_3 + b_2 \times \underset{\uparrow\ \text{dummy variables}}{1} + e_3$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ \\ b_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

$$\underline{Y} = \underline{X}\ \underline{b} + \underline{e}$$

Matrix formulation

Hats = estimates

Linear regression

$$Y_i = b_1\, x_i + b_2 + e_i \qquad i = 1,2,3$$

$$Y_1 = b_1\, x_1 + b_2 \times 1 + e_1$$

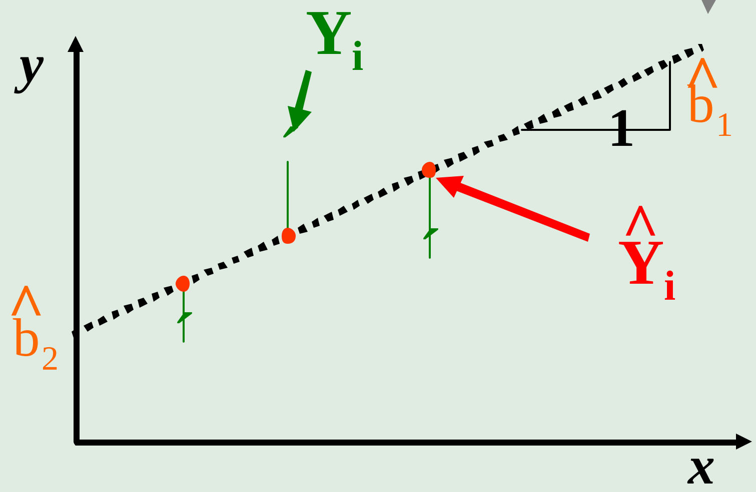
$$Y_2 = b_1\, x_2 + b_2 \times 1 + e_2$$

$$Y_3 = b_1\, x_3 + b_2 \times 1 + e_3$$

dummy variables

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ \\ b_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

$$\underline{Y} = \underline{X}\ \underline{b} + \underline{e}$$

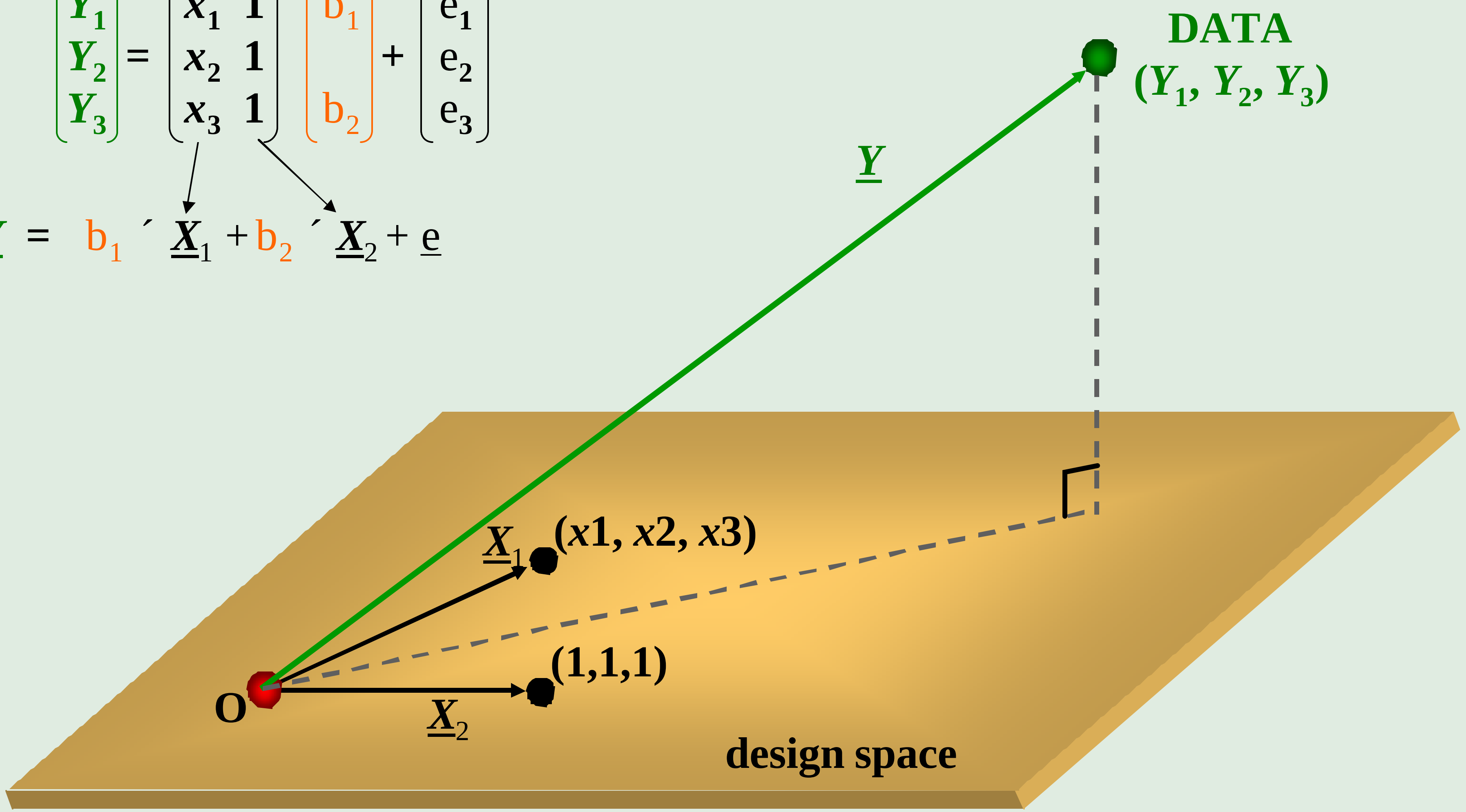


Parameter estimates	$\hat{b}_1$ & $\hat{b}_2$
Fitted values	$\hat{Y}_1$ & $\hat{Y}_2$
Residuals	$\hat{e}_1$ & $\hat{e}_2$

Geometrical perspective

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ \\ b_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

$$\underline{Y} = b_1 \times \underline{X}_1 + b_2 \times \underline{X}_2 + \underline{e}$$



GLM essentials

- The model
  - Design matrix: Effects of interest
  - Design matrix: Confounds (aka effects of no interest)
  - Residuals (error measures of the whole model)
- **Estimate effects and error for data**
  - **Parameter estimates (aka betas)**
  - **Quantify specific effects using contrasts of parameter estimates**
- Statistic
  - Compare estimated effects – the contrasts – with appropriate error measures
  - Are the effects surprisingly large?

Parameter estimation

Ordinary least squares

$$Y = Xb + e$$

$$\hat{b} = (X^T X)^{-1} X^T Y$$

Parameter estimates

$$\hat{e} = Y - X\hat{b}$$

residuals

Estimate parameters such that $\sum_{t=1}^{N} \hat{e}_t^2$ minimal

Least squares parameter estimate

Parameter estimation

Ordinary least squares

$$Y = Xb + e$$

$$\hat{b} = (X^T X)^{-1} X^T Y$$

$$\hat{e} = Y - X\hat{b}$$

Parameter estimates

residuals = r

Error variance s^2 = (sum of) squared residuals standardised for df

= $r^T r$ / df (sum of squares)

…where **degrees of freedom df** (assuming iid):

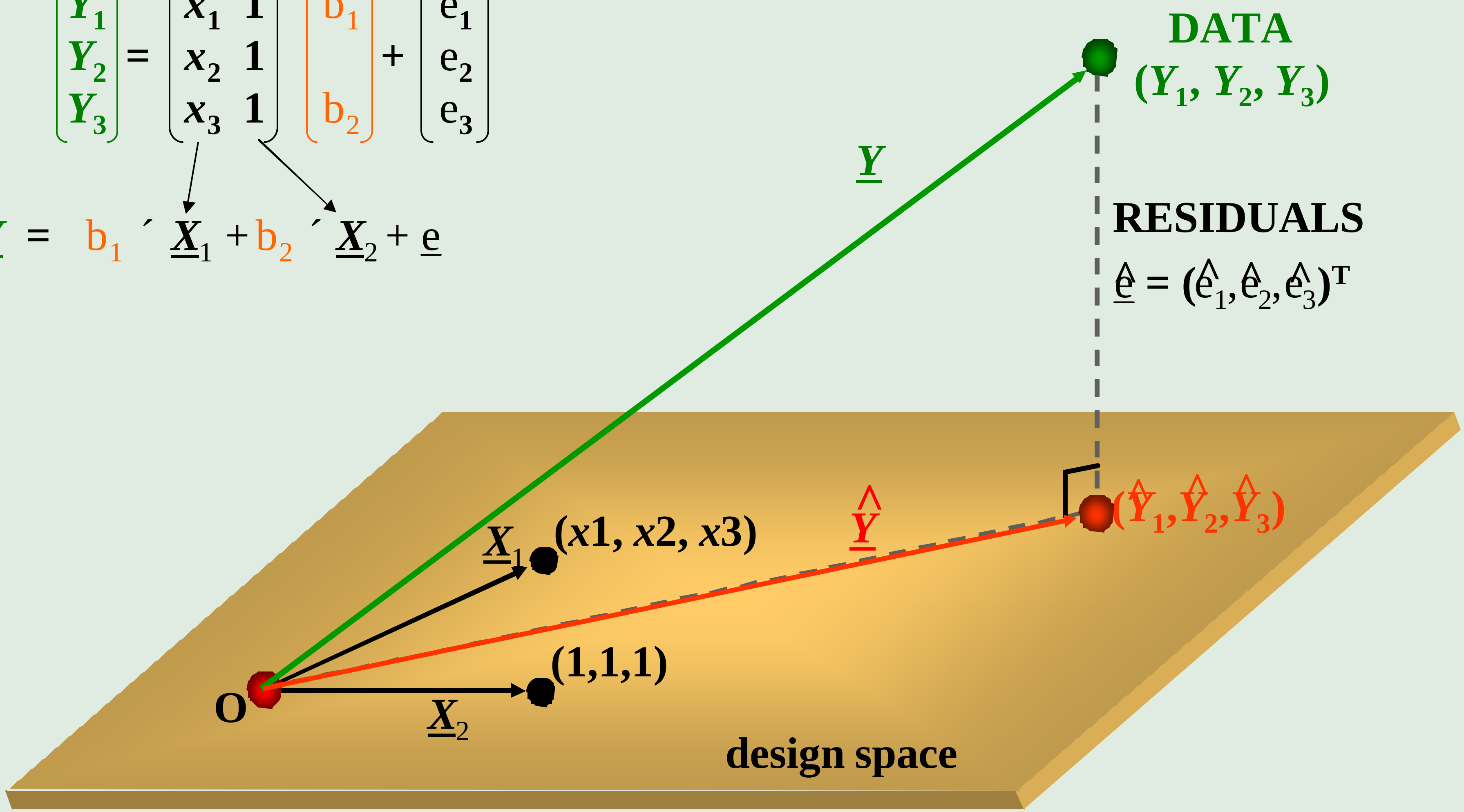
= N - rank(X)

(=N-P if X full rank)

Estimation (geometrical)

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

$$\underline{Y} = b_1 \times \underline{X}_1 + b_2 \times \underline{X}_2 + \underline{e}$$



GLM essentials

- The model
  - Design matrix: Effects of interest
  - Design matrix: Confounds (aka effects of no interest)
  - Residuals (error measures of the whole model)
- Estimate effects and error for data
  - Parameter estimates (aka betas)
  - Quantify specific effects using contrasts of parameter estimates
- **Statistic**
  - **Compare estimated effects – the contrasts – with appropriate error measures**
  - **Are the effects surprisingly large?**

Inference - contrasts

$$Y = Xb + e$$

A contrast = a linear combination of parameters: *c´* x b → **spm_con*img**

t-test: one-dimensional contrast, difference in means/ difference from zero

boxcar parameter > 0 ?

Null hypothesis: $b_1 = 0$

→ spmT_000*.img

SPM{t} map

***F*-test: tests multiple linear hypotheses – does subset of model account for significant variance**

Does boxcar parameter model anything?

Null hypothesis: variance of tested effects = error variance

→ ess_000*.img

SPM*{F}* map

t-statistic - example

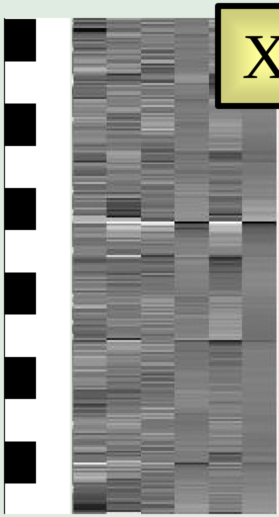
$$Y = Xb + e$$

$$t = \frac{c^T \hat{b}}{\hat{Std}(c^T \hat{b})}$$

Contrast of parameter estimates

Variance estimate

c = 1 0 0 0 0 0 0 0 0 0 0



X

Standard error of contrast depends on the **design**, and is larger with greater **residual error** and ‚greater‘ **covariance/ autocorrelation**

Degrees of freedom *d.f.* then = *n*-*p*, where *n* observations, *p* parameters

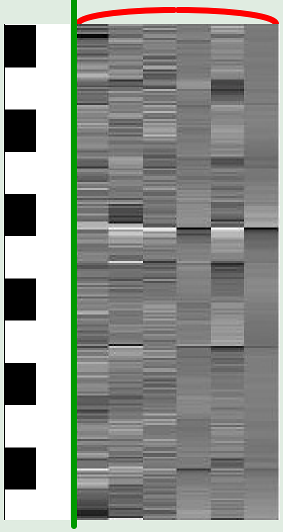
Tests for a **directional** difference in means

F-statistic - example

Do movement parameters (or other confounds) account for anything?

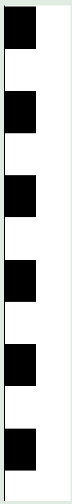
H_0: True model is X_0 **H_0: $\mathrm{b}_{3\text{-}9} = (0\ 0\ 0\ 0\ ...)$** **test H_0 : $c' \times \mathrm{b} = 0$?**

X_0 X_1 ($\underline{\mathrm{b}}_{3\text{-}9}$)



This model ?

X_0



Or this one ?

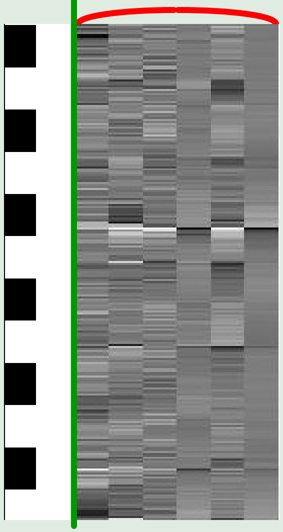
Null hypothesis H_0:
That all these betas $\mathrm{b}_{3\text{-}9}$ are zero, i.e. that no linear combination of the effects accounts for significant variance
This is a **non-directional** test

F-statistic - example

Do movement parameters (or other confounds) account for anything?

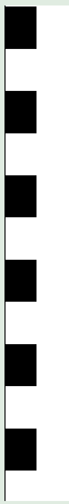
H_0: True model is X_0 **H_0: $b_{3\text{-}9} = (0\ 0\ 0\ 0\ ...)$** **test H_0 : $c' \times b = 0$?**

X_0 X_1 ($\underline{b}_{3\text{-}9}$)



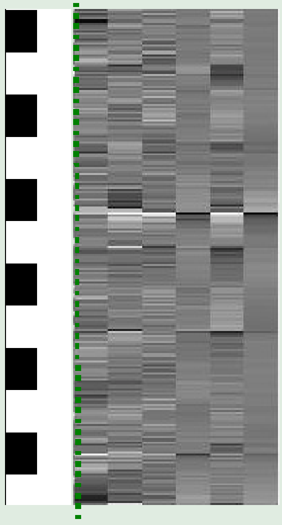
This model ?

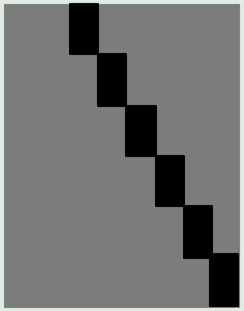
X_0



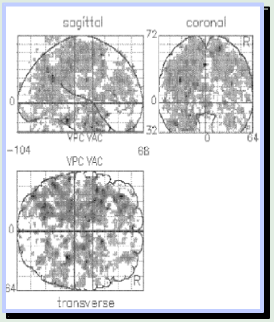
Or this one ?

$$c' = \begin{matrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix}$$





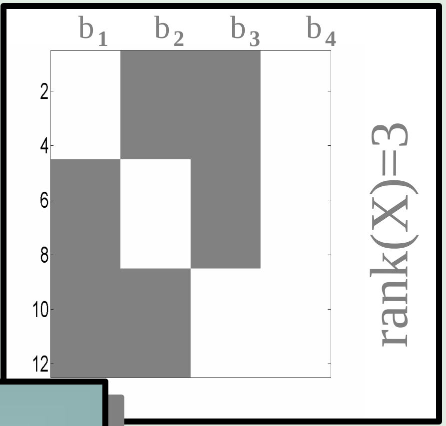
SPM{*F*}



Summary so far

- The essential model contains
  - Effects of interest
- A better model?
  - A better model (within reason) means smaller residual variance and more significant statistics
  - Capturing the signal – later
  - Add confounds/ effects of no interest
  - Example of movement parameters in fMRI
  - A further example (mainly relevant to PET)…

Example PET experiment



12 scans, 3 conditions (1-way ANOVA)

$$y_j = x_{1j}\, b_1 + x_{2j}\, b_2 + x_{3j}\, b_3 + x_{4j}\, b_4 + e_j$$

where (dummy) variables:

x_{1j} = *[0,1] = condition A (first 4 scans)*

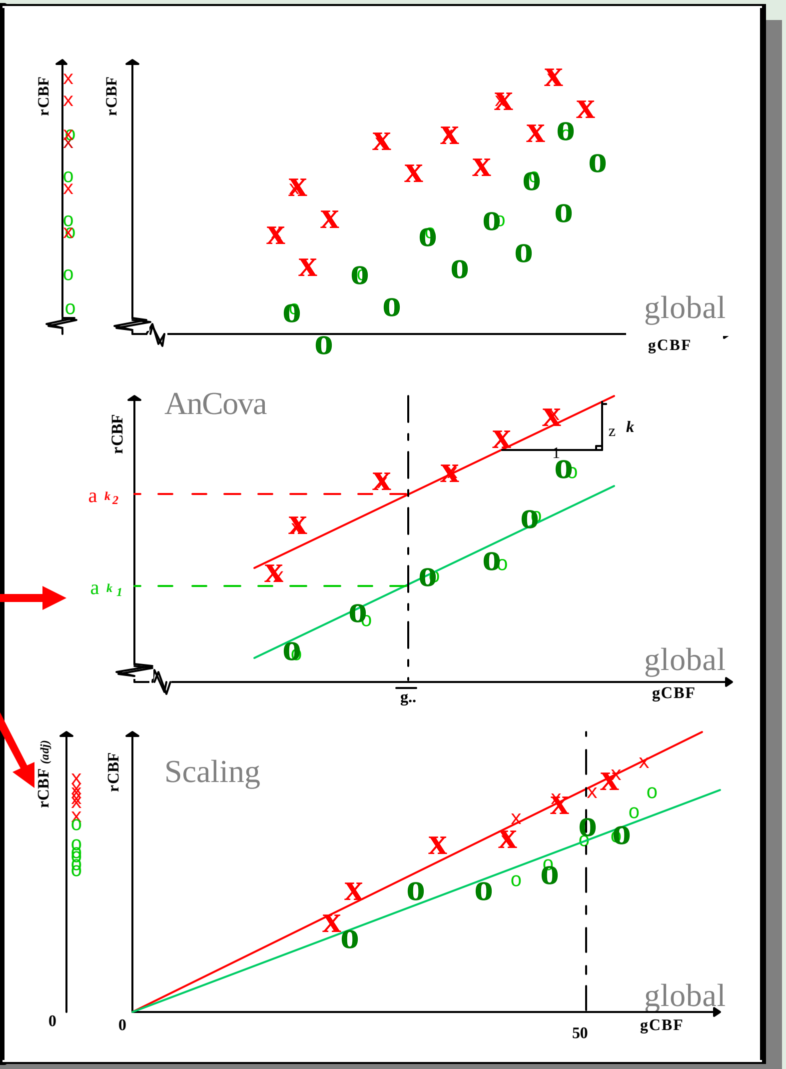
x_{2j} = *[0,1] = condition B (second 4 scans)*

x_{3j} = *[0,1] = condition C (third 4 scans)*

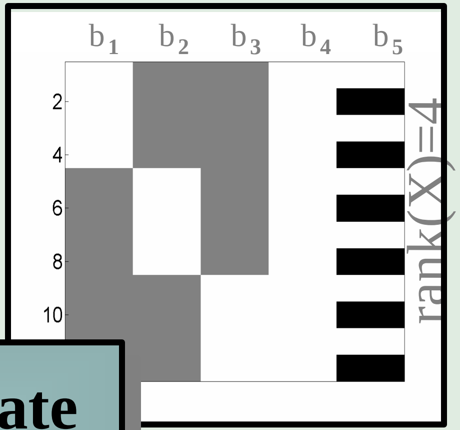
x_{4j} = *[1] = grand mean (session constant)*

Global effects

- May be variation in PET tracer dose from scan to scan
- Such "global" changes in image intensity (gCBF) confound local / regional (rCBF) changes of experiment
- Adjust for global effects by:
  - AnCova (Additive Model) - *PET?*
  - Proportional Scaling - *fMRI?*
- Can improve statistics when orthogonal to effects of interest…
- …but can also worsen when effects of interest correlated with global



Global effects (AnCova)



12 scans, 3 conditions, 1 confounding covariate

$$y_j = x_{1j}\, b_1 + x_{2j}\, b_2 + x_{3j}\, b_3 + x_{4j}\, b_4 + \mathbf{x_{5j}\, b_5} + e_j$$

where (dummy) variables:

x_{1j} = *[0,1] = condition A (first 4 scans)*

x_{2j} = *[0,1] = condition B (second 4 scans)*

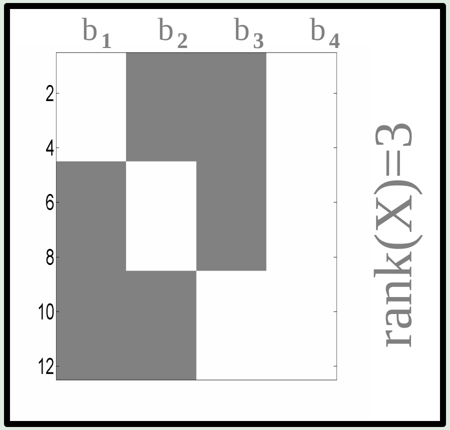
x_{3j} = *[0,1] = condition C (third 4 scans)*

x_{4j} = *[1] = grand mean (session constant)*

x_{5j} = global signal (mean over all voxels)

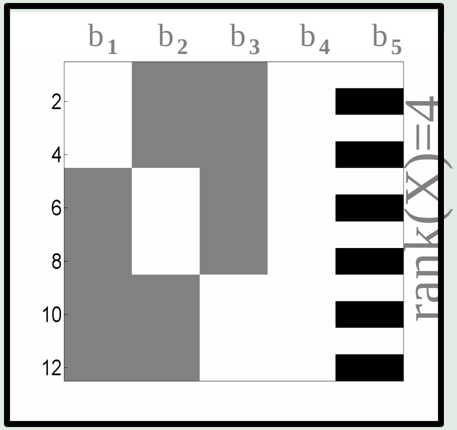
Global effects (AnCova)

No Global



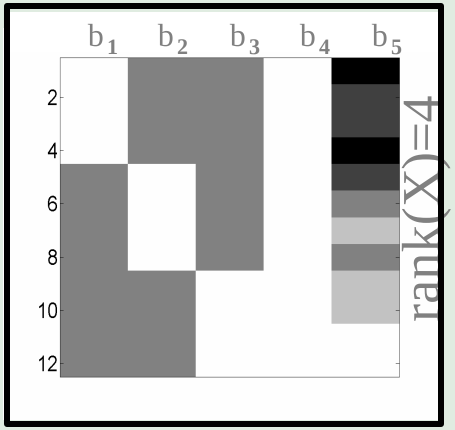
- Global effects not accounted for
- Maximum degrees of freedom (global uses one)

Orthogonal global



- Global effects independent of effects of interest
- Smaller residual variance
- Larger T statistic
- More significant

Correlated global



- Global effects correlated with effects of interest
- Smaller effect &/or larger residuals
- Smaller T statistic
- Less significant

Global effects (scaling)

- **Two types of scaling: Grand Mean scaling and Global scaling**
  - Grand Mean scaling is automatic, global scaling is optional
  - Grand Mean scales by 100/mean over all voxels and ALL scans (i.e, single number per session)
  - Global scaling scales by 100/mean over all voxels for EACH scan (i.e, a different scaling factor every scan)
- **Problem with global scaling is that TRUE global is not (normally) known… … only estimated by the mean over voxels**
  - So if there is a large signal change over many voxels, the global **estimate** will be confounded by local changes
  - This can produce artifactual deactivations in other regions after global scaling
- **Since most sources of global variability in fMRI are low frequency (drift), high-pass filtering may be sufficient, and many people do not use global scaling**

Summary

- General(ised) linear model partitions data into
  - Effects of interest & confounds/ effects of no interest
  - Error
- Least squares estimation
  - Minimises difference between model & data
  - To do this, assumptions made about errors – more later
- Inference at every voxel
  - Test hypothesis using contrast – more later
  - Inference can be Bayesian as well as classical
- Next: Applying the GLM to fMRI