SPM

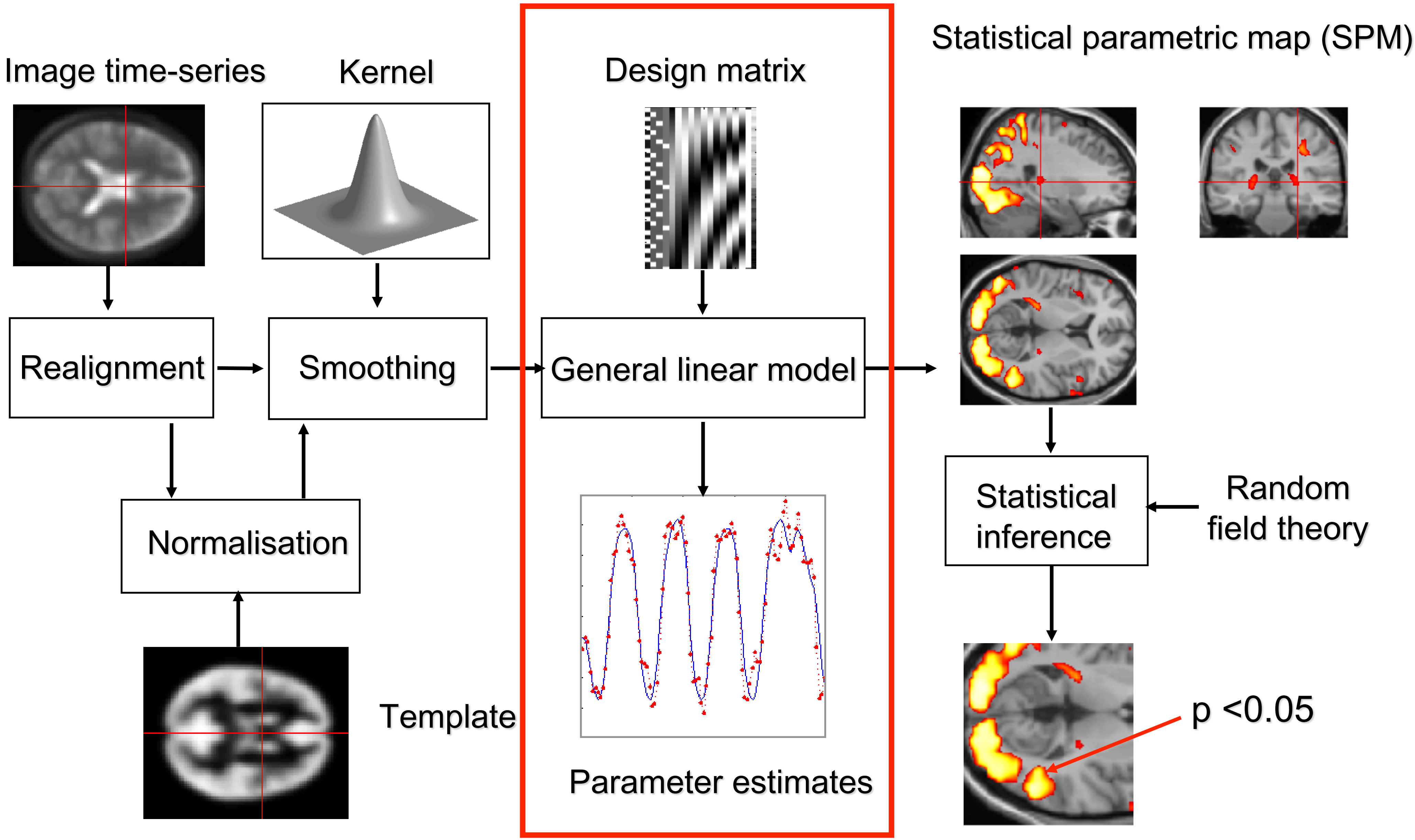
The General Linear Model (GLM)

Ged Ridgway
Wellcome Trust Centre for Neuroimaging
University College London

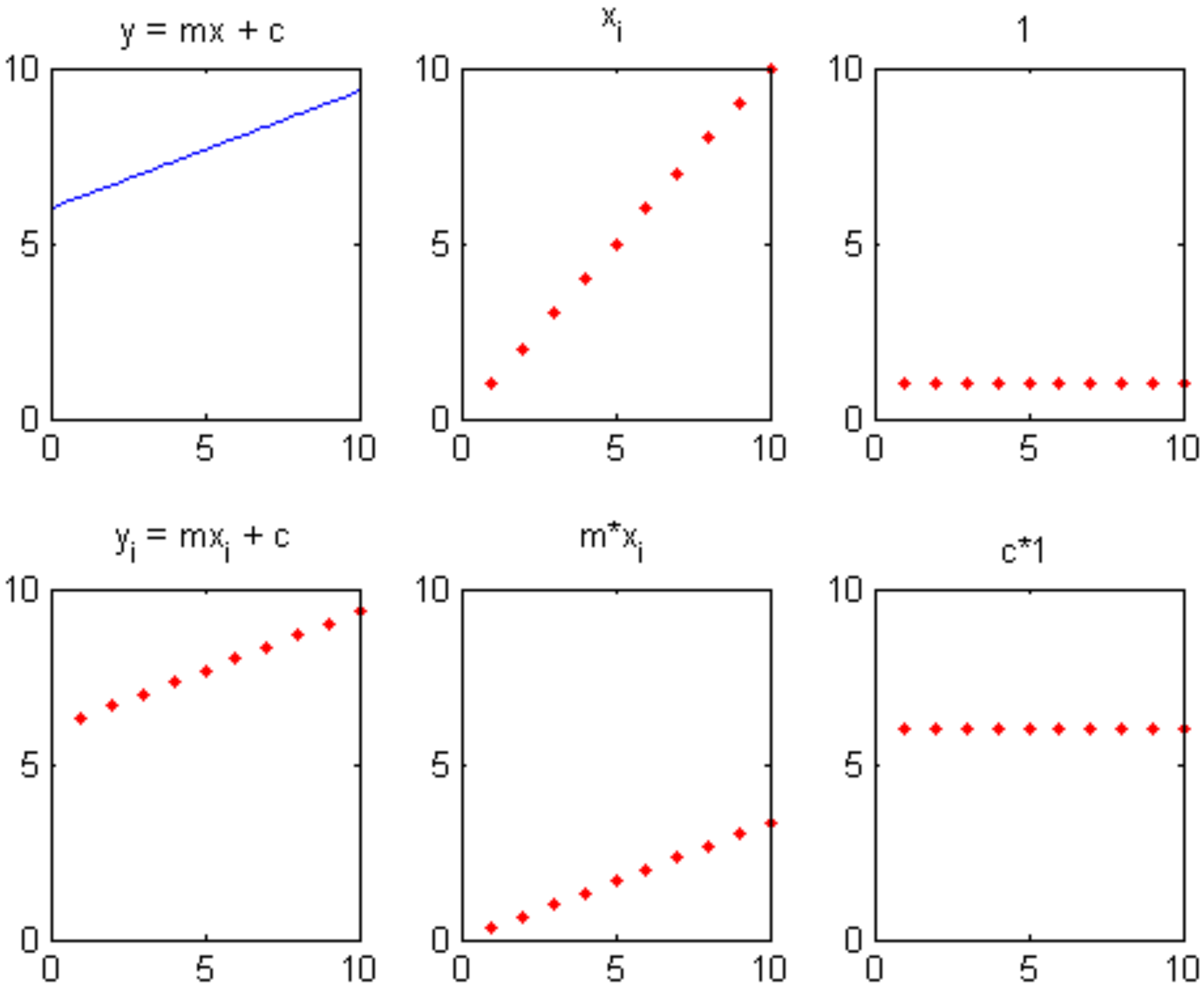
[slides from the FIL Methods group]

SPM Course
Vancouver, August 2010

Overview of SPM



Simple regression and the GLM



$$y_i = mx_i + c$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = m \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + c \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \mathbf{X\beta}$$

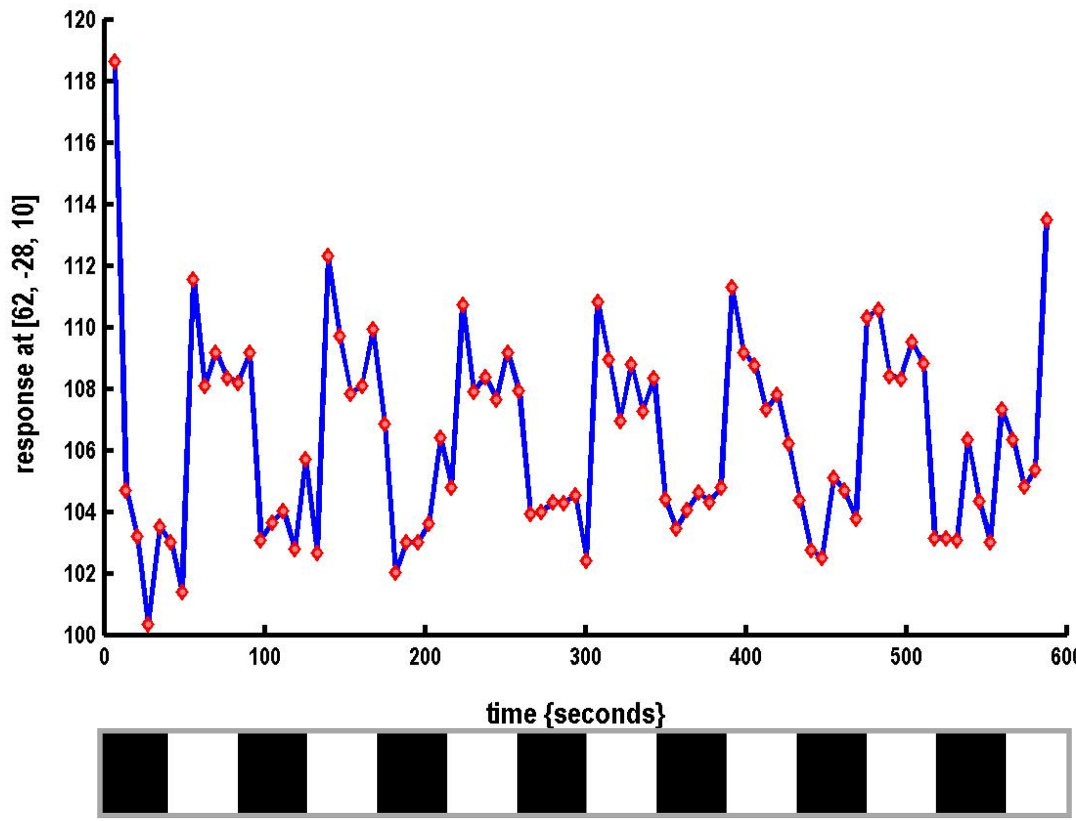
A very simple fMRI experiment

One session

Passive word listening versus rest

7 cycles of rest and listening

Blocks of 6 scans with 7 sec TR



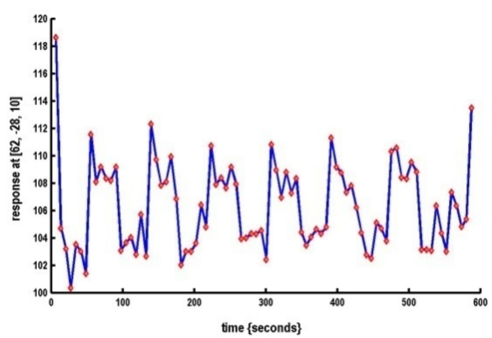
Stimulus function

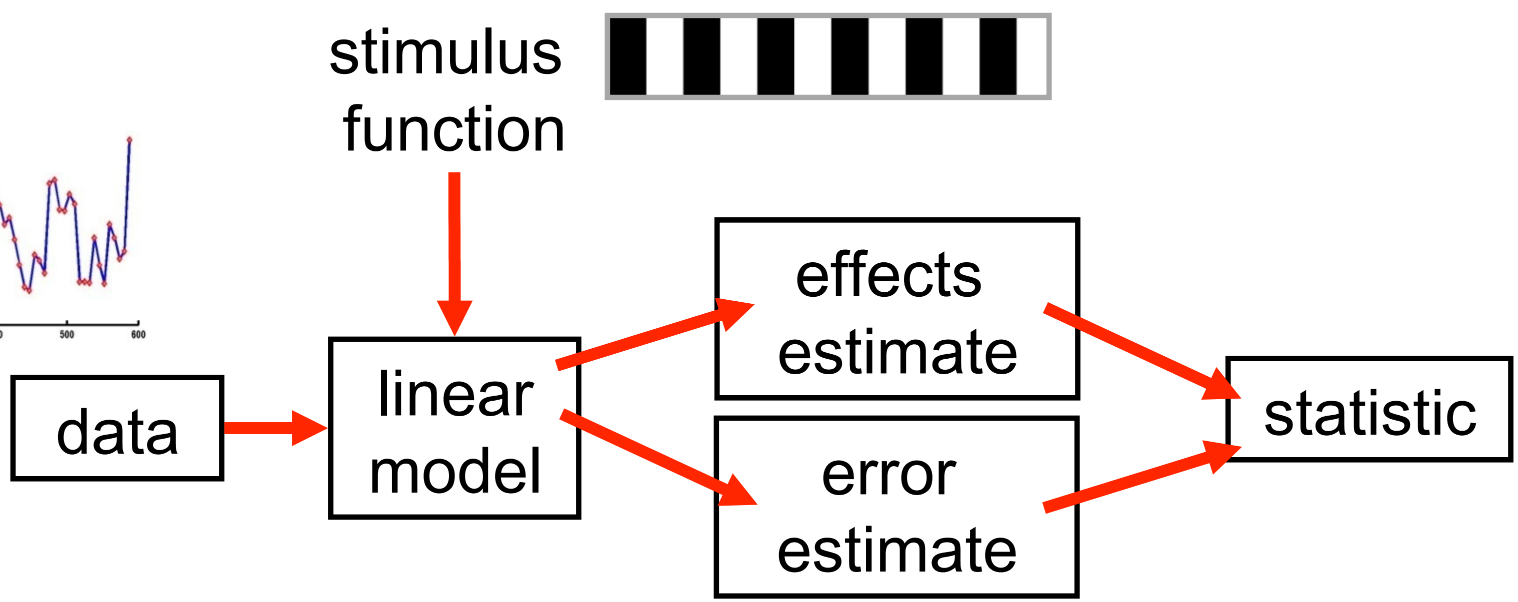
Question: Is there a change in the BOLD response between listening and rest?

Modelling the measured data

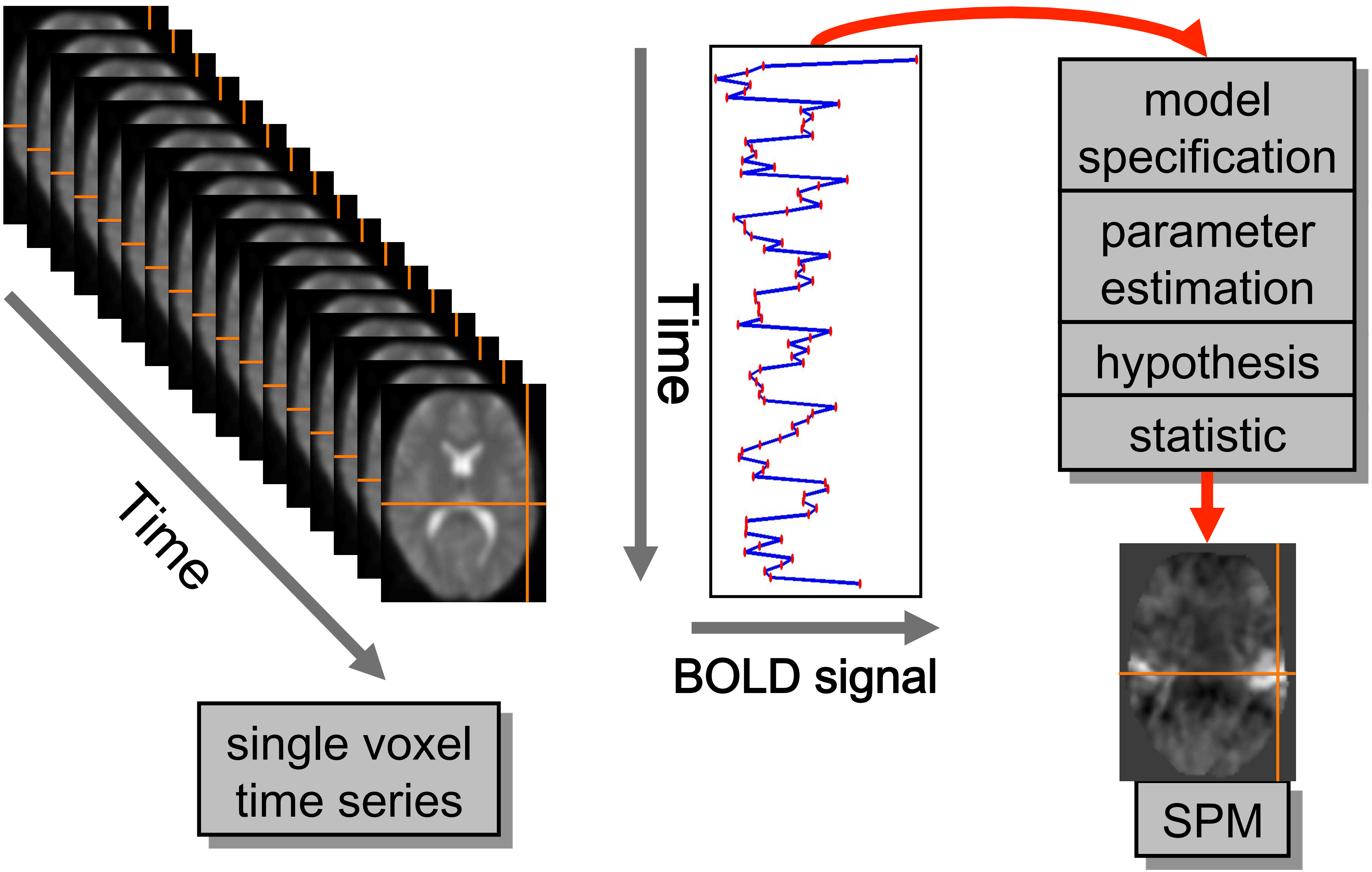
Why? Make inferences about effects of interest

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

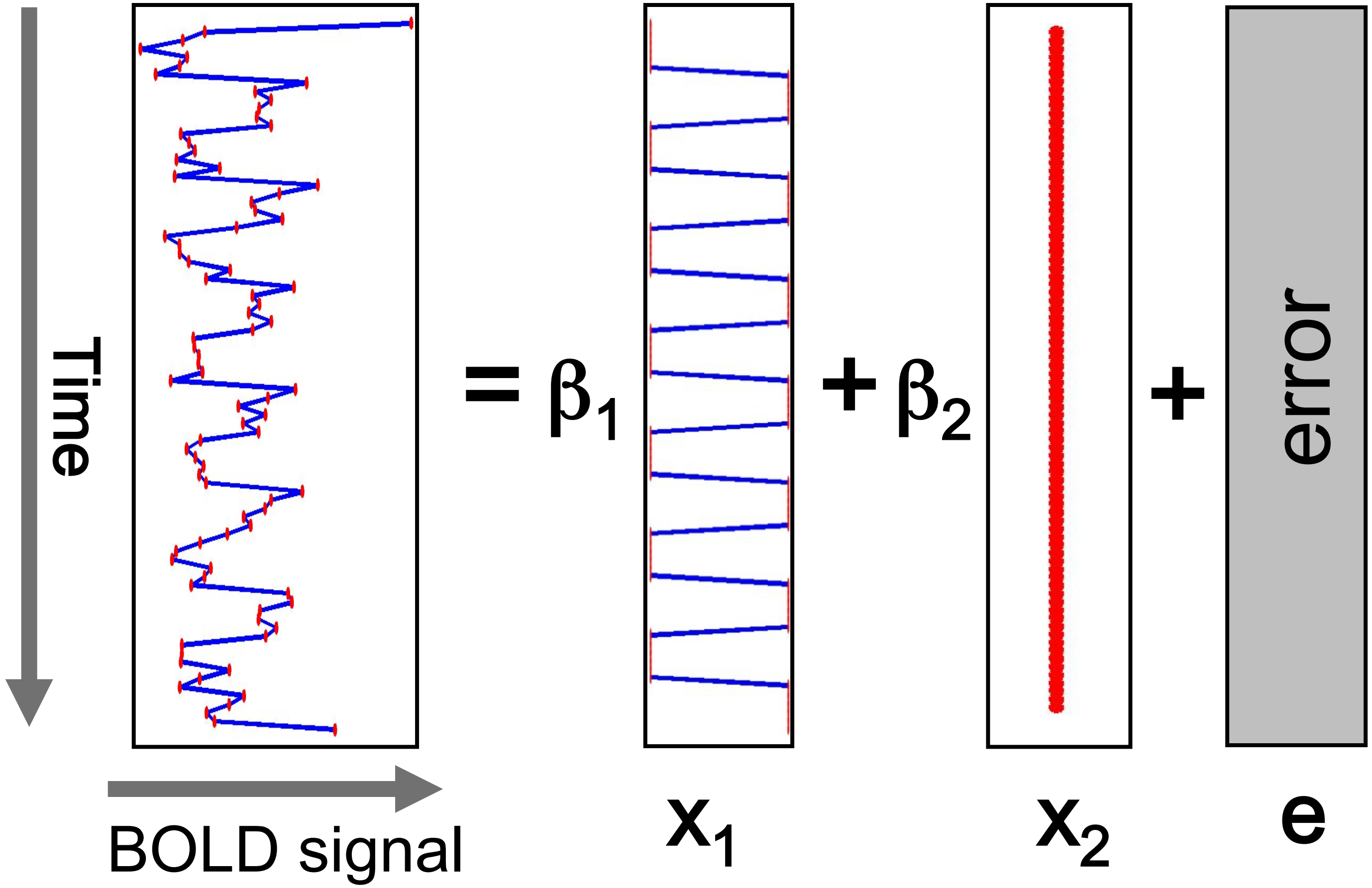




Voxel-wise time series analysis

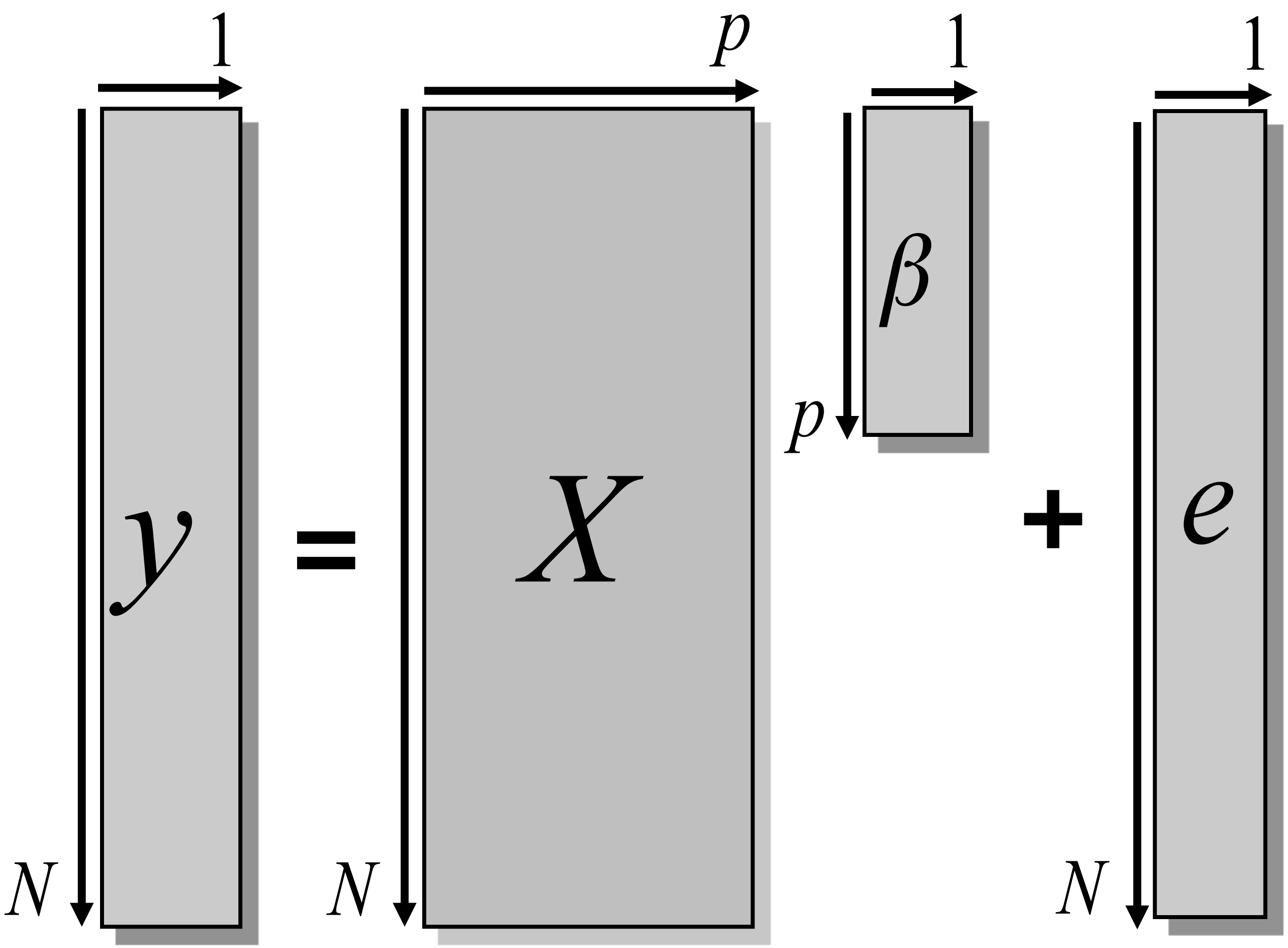


Single voxel regression model



$$y = x_1\beta_1 + x_2\beta_2 + e$$

Mass-univariate analysis: voxel-wise GLM



$$y = X\beta + e$$

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

Model is specified by

1. Design matrix X
2. Assumptions about e

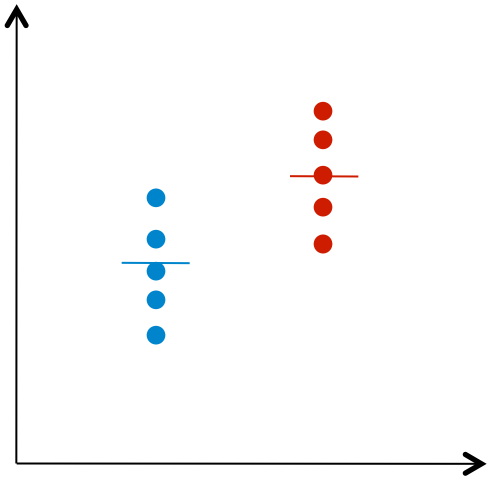
N: number of scans
p: number of regressors

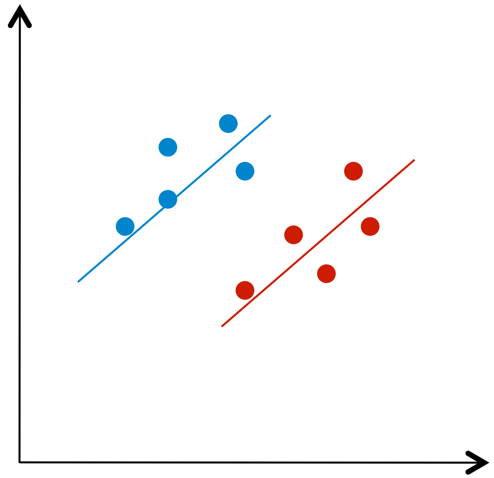
The design matrix embodies all available knowledge about experimentally controlled factors and potential confounds.

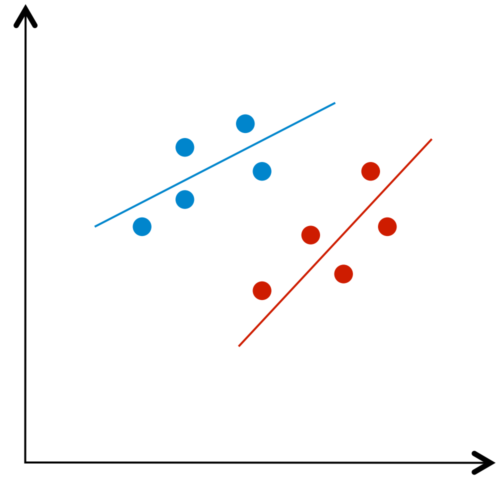
Example GLM “factorial design” models

- one sample *t*-test
- two sample *t*-test
- paired *t*-test
- Analysis of Variance (ANOVA)
- Factorial designs
- correlation
- linear regression
- multiple regression
- *F*-tests
- fMRI time series models
- etc…

Example GLM “factorial design” models





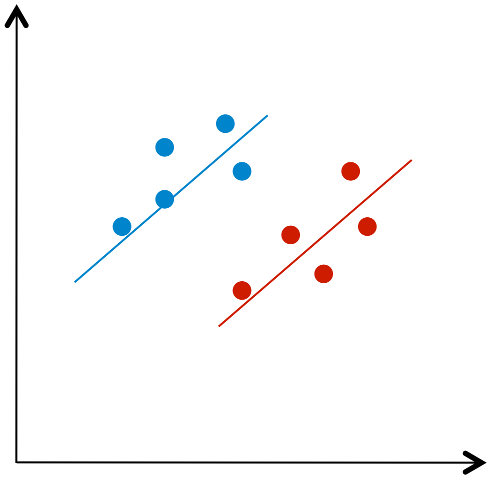


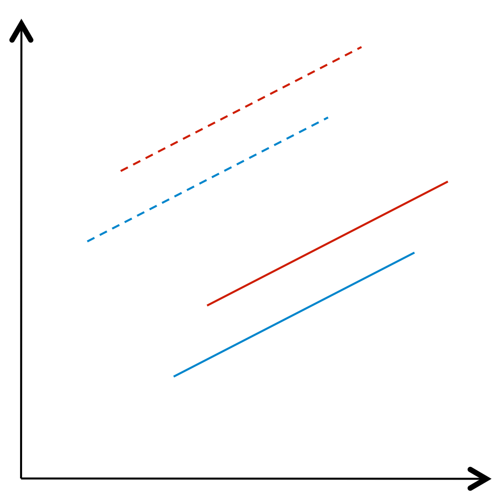
$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

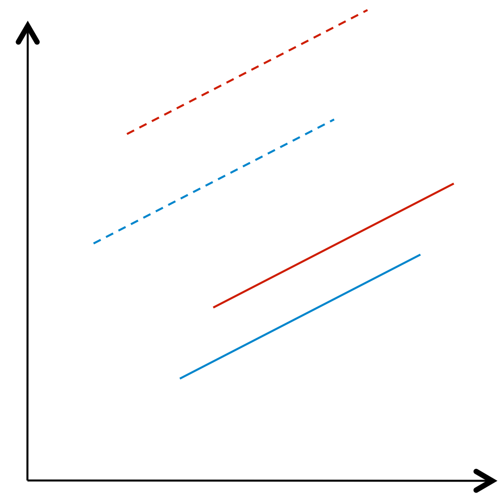
$$\begin{bmatrix} 1 & 0 & x_{1,1} \\ 1 & 0 & x_{1,2} \\ \vdots & \vdots & \vdots \\ 0 & 1 & x_{2,n-1} \\ 0 & 1 & x_{2,n} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & x_{1,1} & 0 \\ 1 & 0 & x_{1,2} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & x_{2,n-1} \\ 0 & 1 & 0 & x_{2,n} \end{bmatrix}$$

Example GLM "factorial design" models







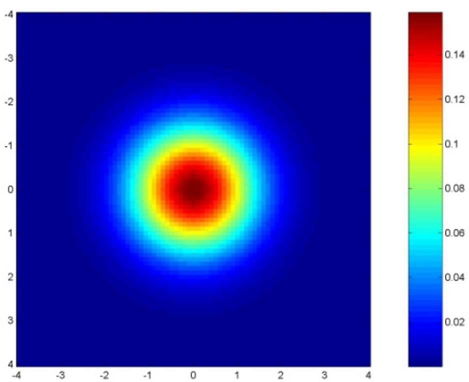
$$\begin{bmatrix} 1 & 0 & x_{1,1} \\ 1 & 0 & x_{1,2} \\ \vdots & \vdots & \vdots \\ 0 & 1 & x_{2,n-1} \\ 0 & 1 & x_{2,n} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & x_{1,1} \\ 1 & 0 & 1 & 0 & x_{1,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 1 & x_{2,n-1} \\ 0 & 1 & 1 & 0 & x_{2,n} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & x_{1,1} \\ 1 & 0 & 1 & 0 & x_{1,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & x_{2,n-1} \\ 0 & 1 & 1 & 1 & x_{2,n} \end{bmatrix}$$

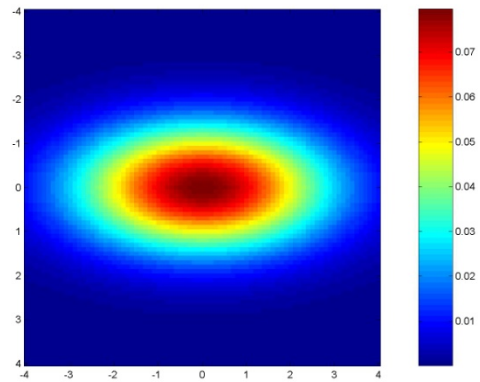
GLM assumes Gaussian “spherical” (i.i.d.) errors

sphericity = i.i.d.
error covariance is scalar multiple of identity matrix:
Cov(e) = $\sigma^2 I$



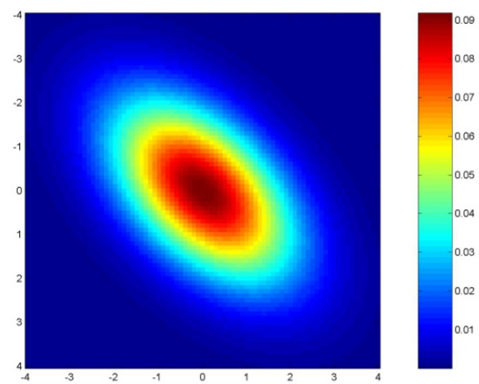
$$Cov(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Examples for non-sphericity:



$$Cov(e) = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

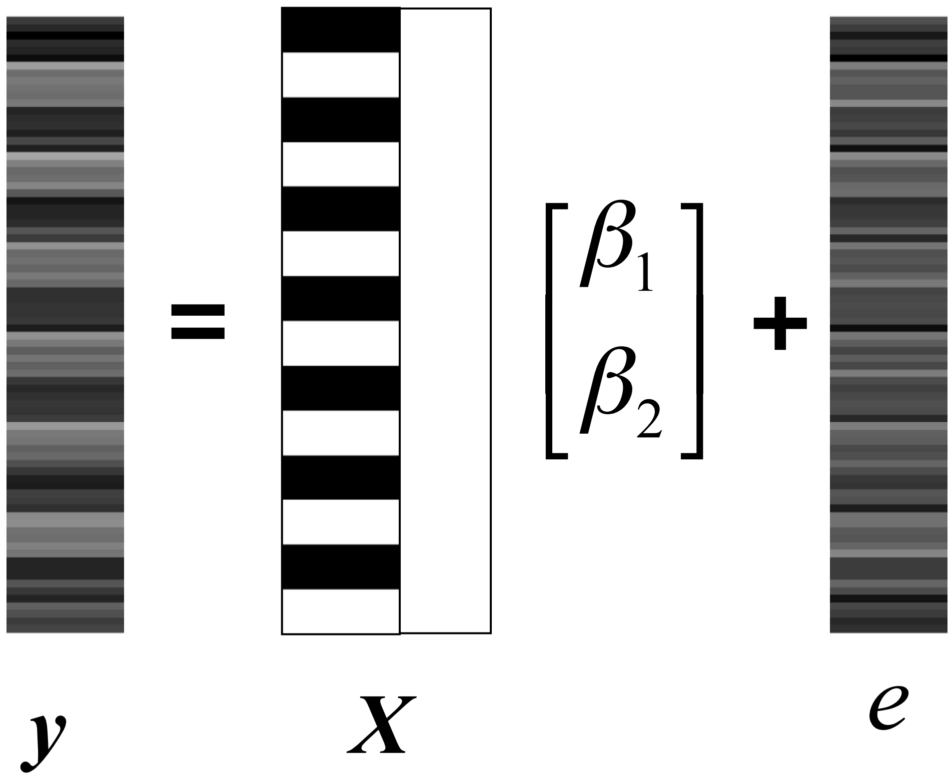
non-identity



$$Cov(e) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

non-independence

Parameter estimation



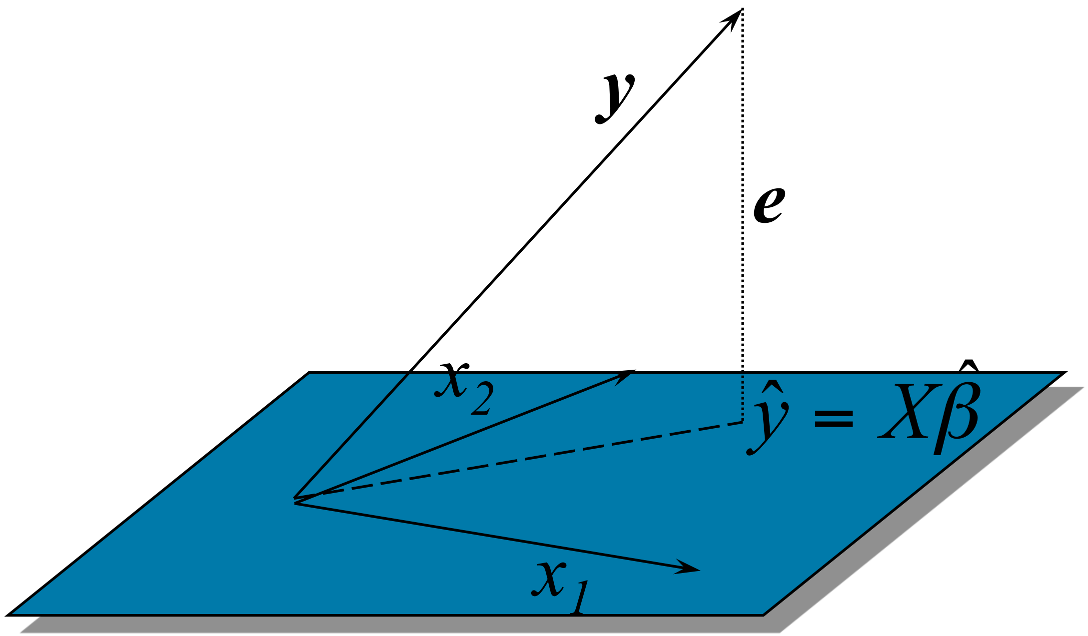
$$y = X\beta + e$$

Objective:
estimate parameters to minimize $\sum_{t=1}^{N} e_t^2$

Ordinary least squares
estimation (OLS)
(assuming i.i.d. error):

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

A geometric perspective on the GLM



Design space defined by X

Smallest errors (shortest error vector) when e is orthogonal to X

$$X^T e = 0$$

$$X^T (y - X\hat{\beta}) = 0$$

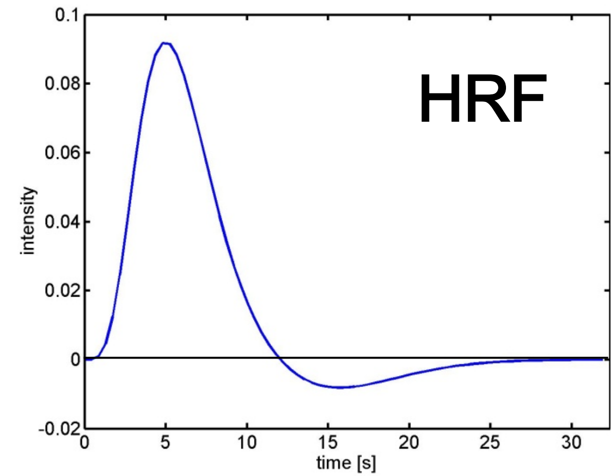
$$X^T y = X^T X\hat{\beta}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Ordinary Least Squares (OLS)

What are the problems of this model?

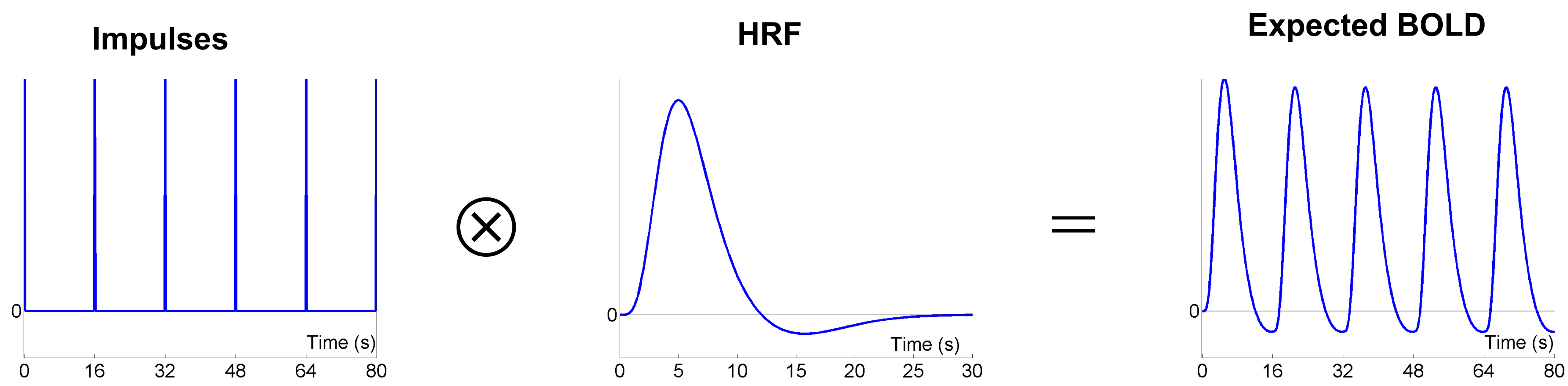
1. BOLD responses have a delayed and dispersed form.



2. The BOLD signal includes substantial amounts of low-frequency noise (eg due to scanner drift).

3. Due to breathing, heartbeat & unmodeled neuronal activity, the errors are serially correlated. This violates the assumptions of the noise model in the GLM

Problem 1: Shape of BOLD response
Solution: Convolution model



expected BOLD response

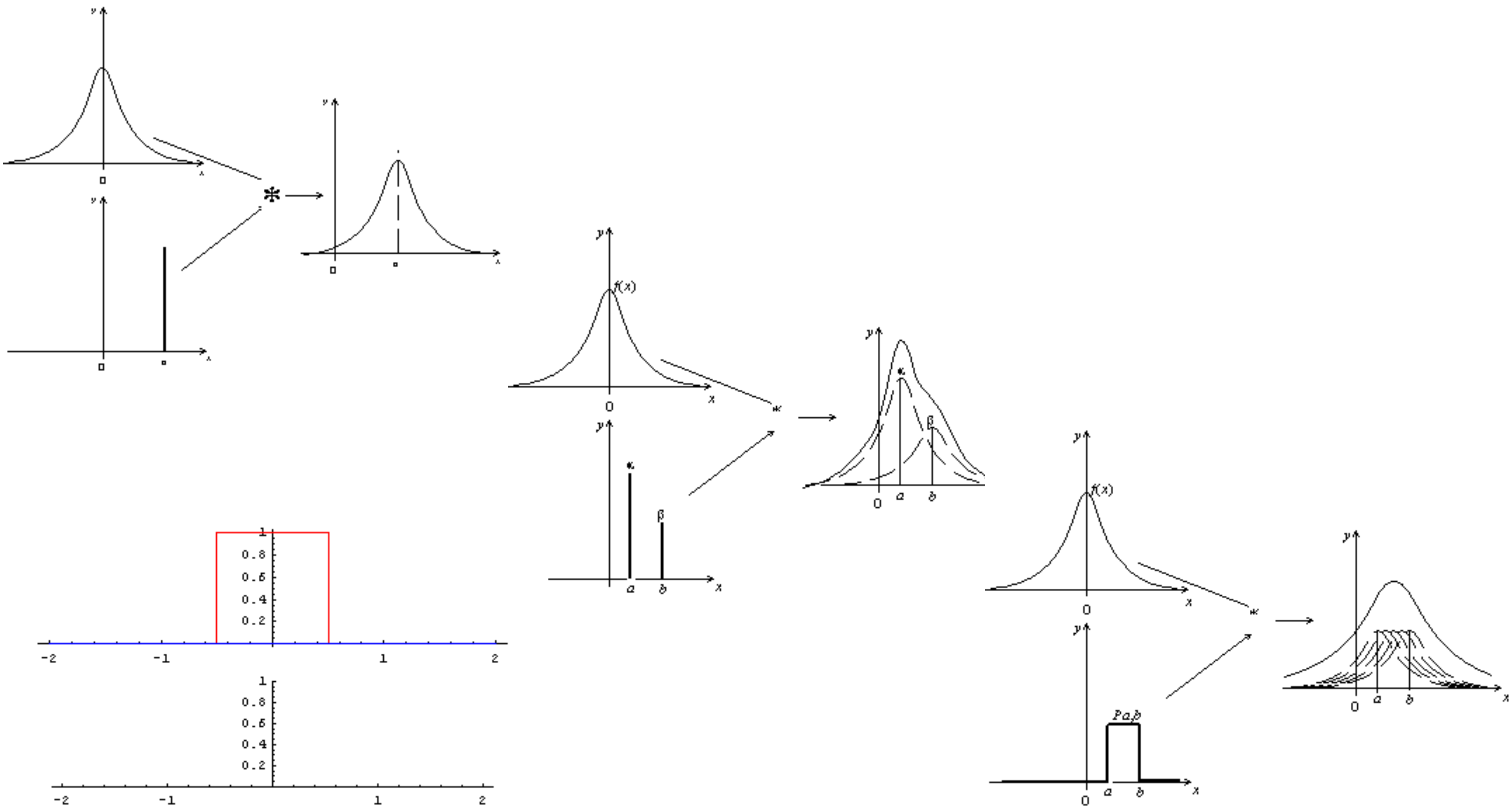
= input function ⊗ impulse response function (HRF)

$$f \otimes g(t) = \int_0^t f(\tau)g(t-\tau)d\tau$$

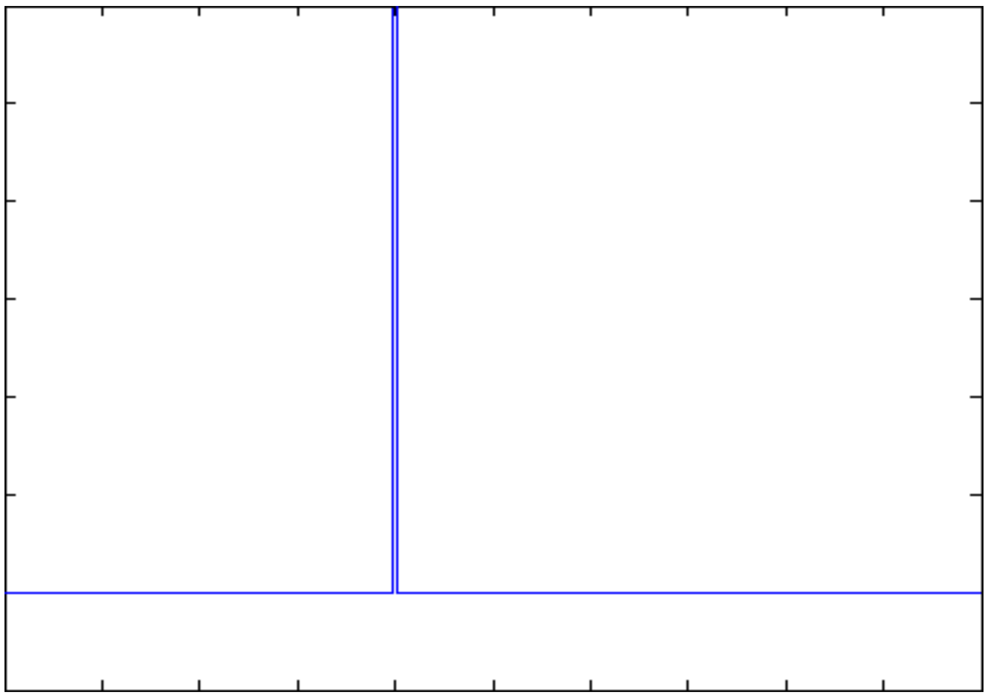
The response of a linear time-invariant (LTI) system is the convolution of the input with the system's response to an impulse (delta function).

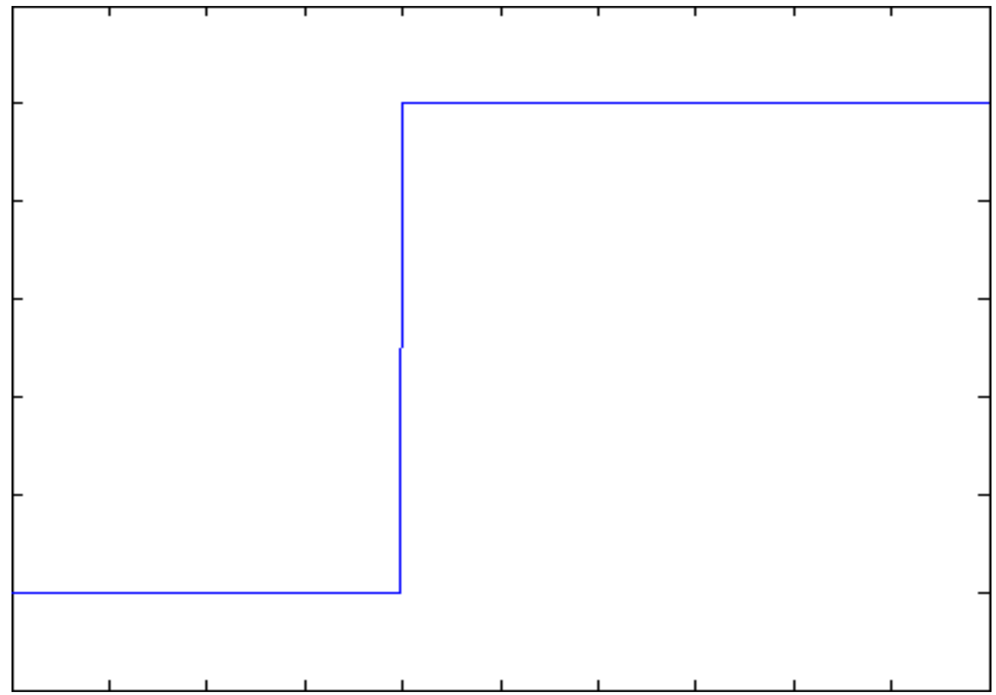
Convolution

Superposition principle



HRF convolution animations

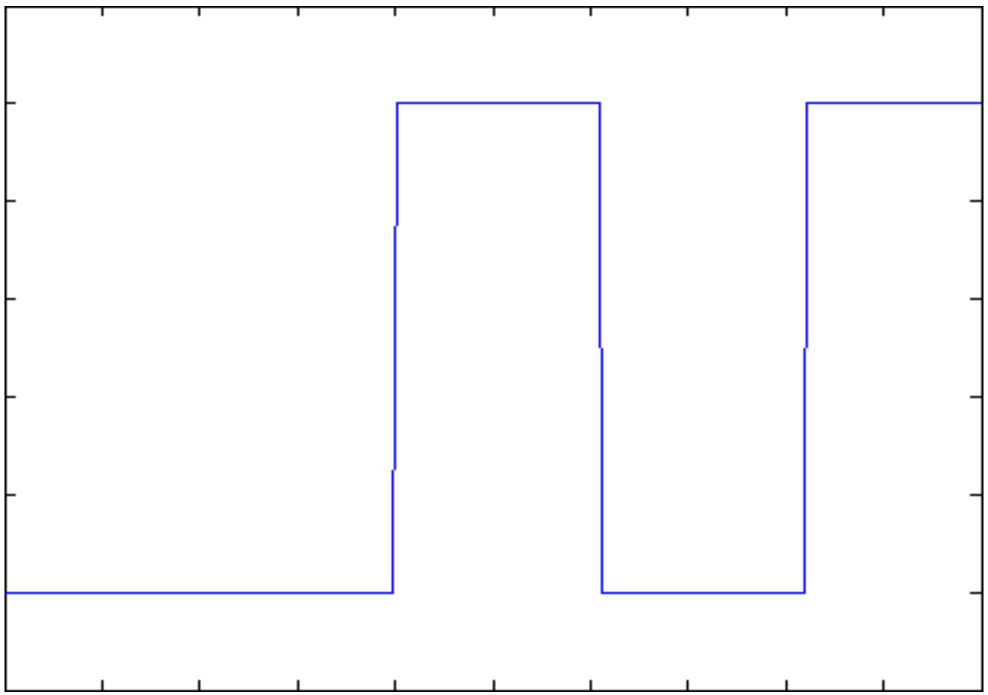


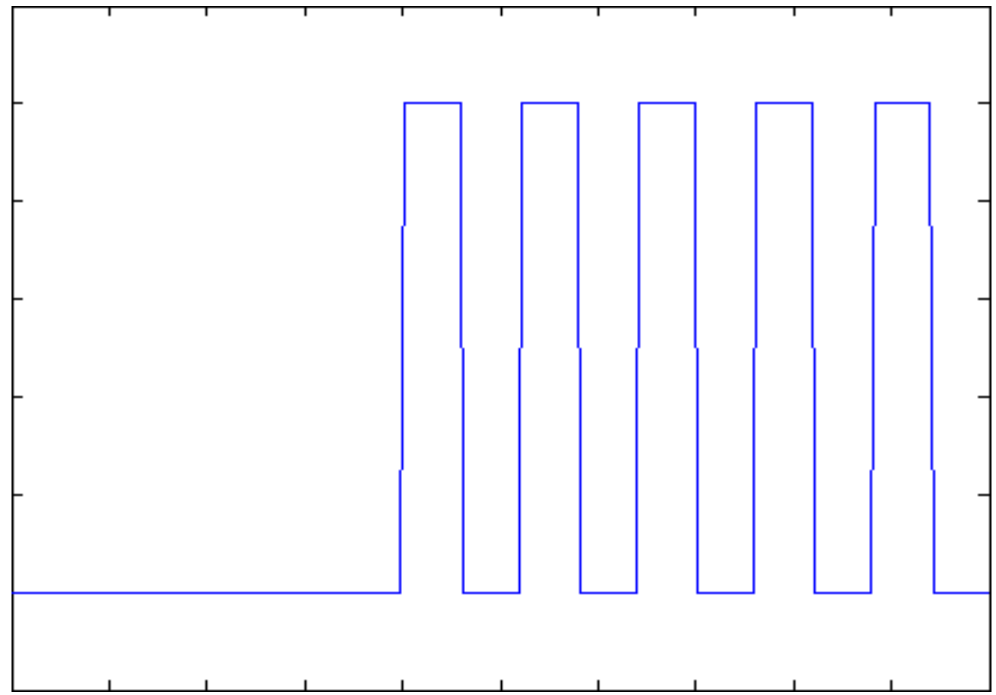


Sliding the reversed HRF past a unit-integral pulse and integrating the product recovers the HRF (hence the name impulse response)

The step response shows a delay and a slight overshoot, before reaching a steady-state. This gives some intuition for a square wave…

HRF convolution animations



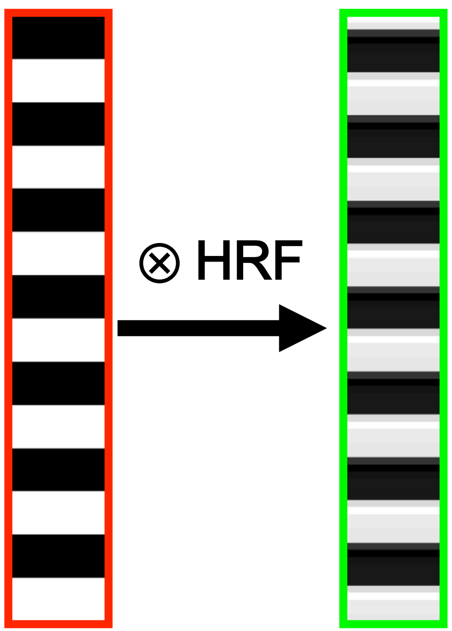


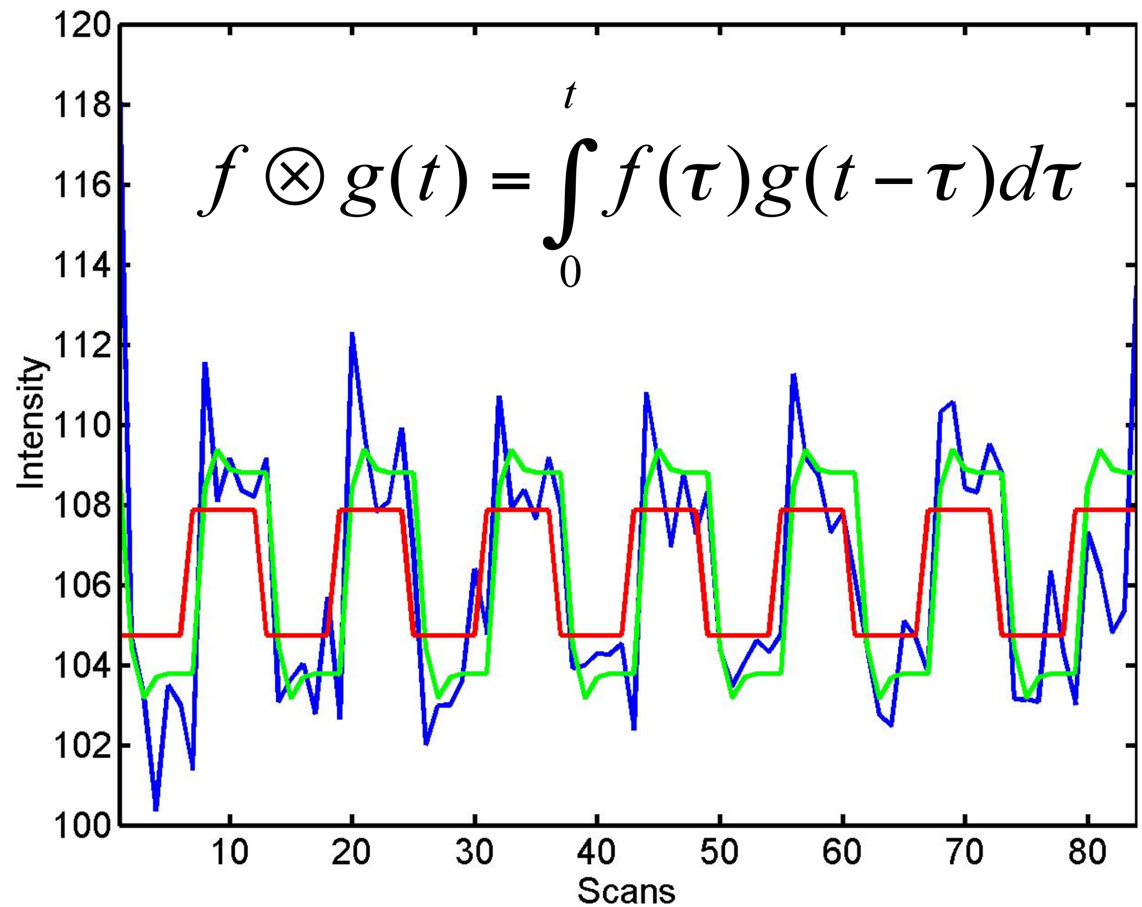
Slow square wave. Main (fundamental) frequency of output matches input, but with delay (phase-shift) and change in amplitude (and shape; sinusoids keep their shape)

Fast square wave. Amplitude is reduced, phase shift is more dramatic (output almost in anti-phase). Fourier transform of HRF yields magnitude and phase responses.

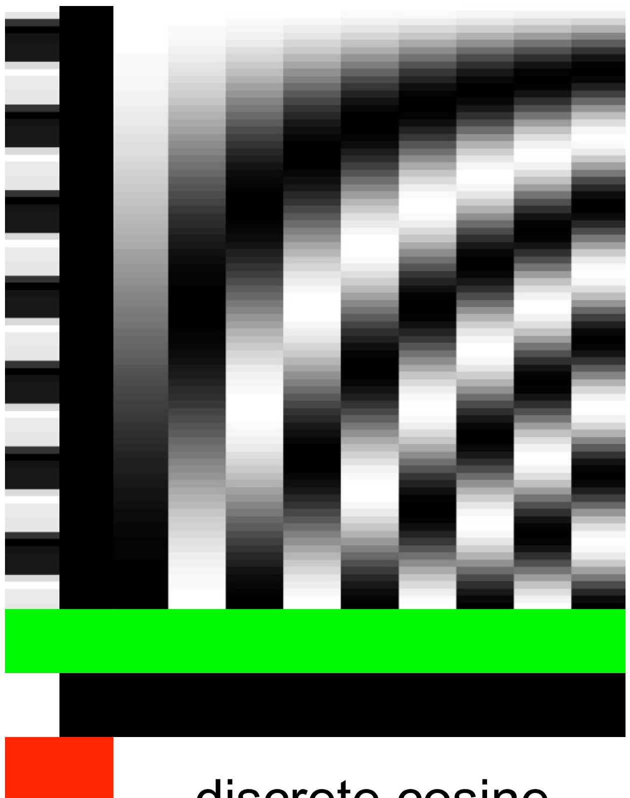
Convolution model of the BOLD response

Convolve stimulus function with a canonical hemodynamic response function (HRF):

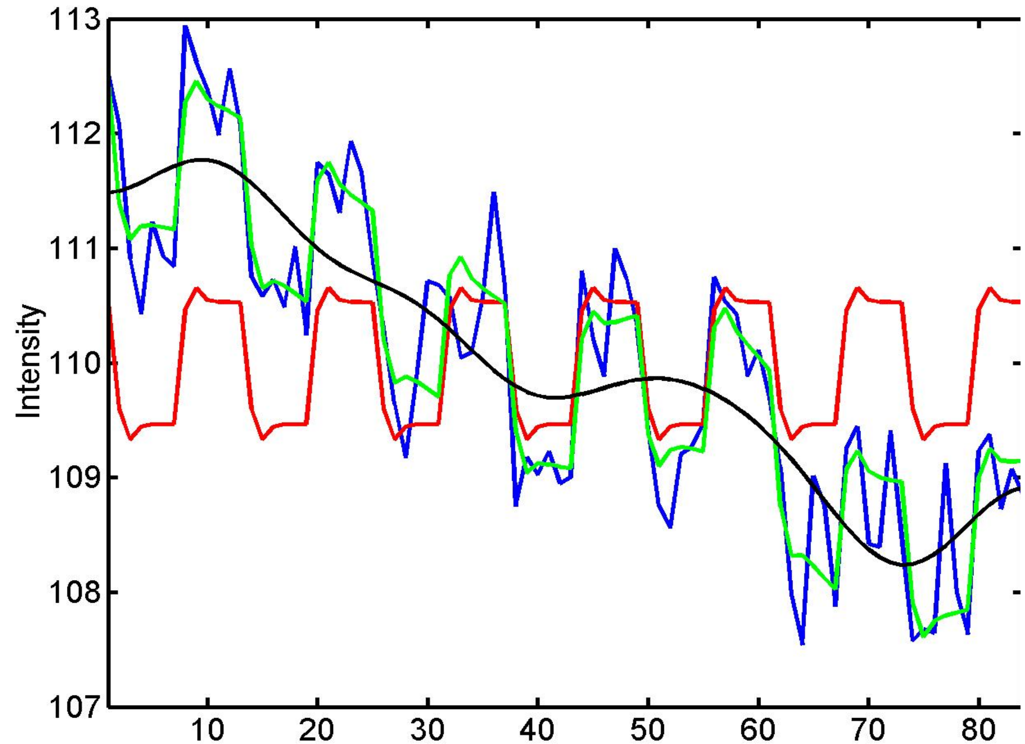




Problem 2: Low-frequency noise
Solution: High pass filtering



discrete cosine transform (DCT) set



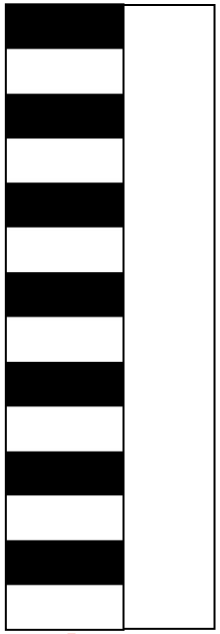
blue = data

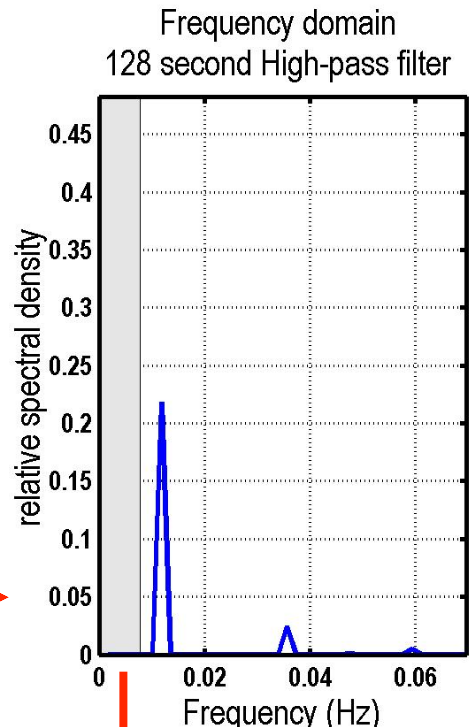
black = mean + low-frequency drift

green = predicted response, taking into account low-frequency drift

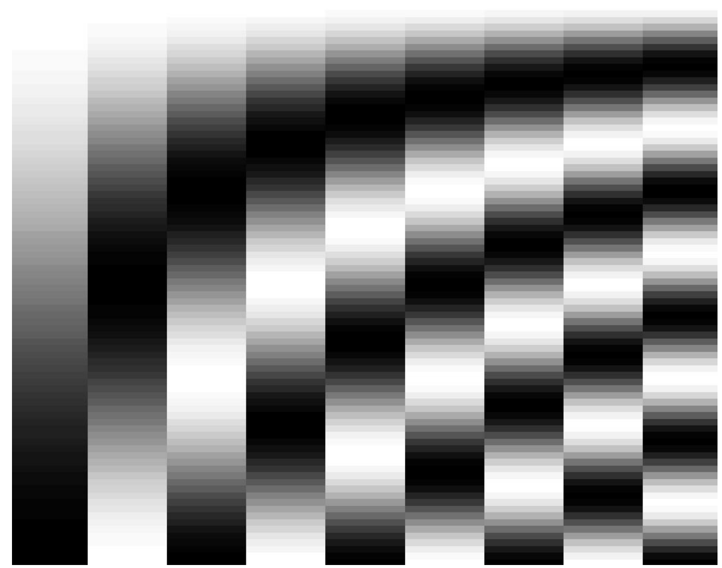
red = predicted response, NOT taking into account low-frequency drift

High pass filtering





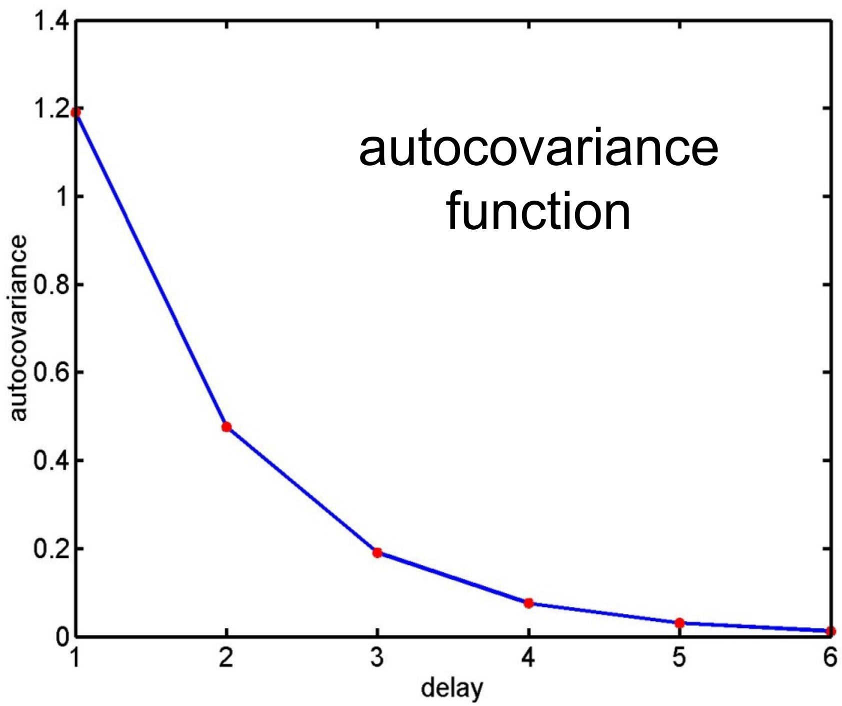
discrete cosine transform (DCT) set

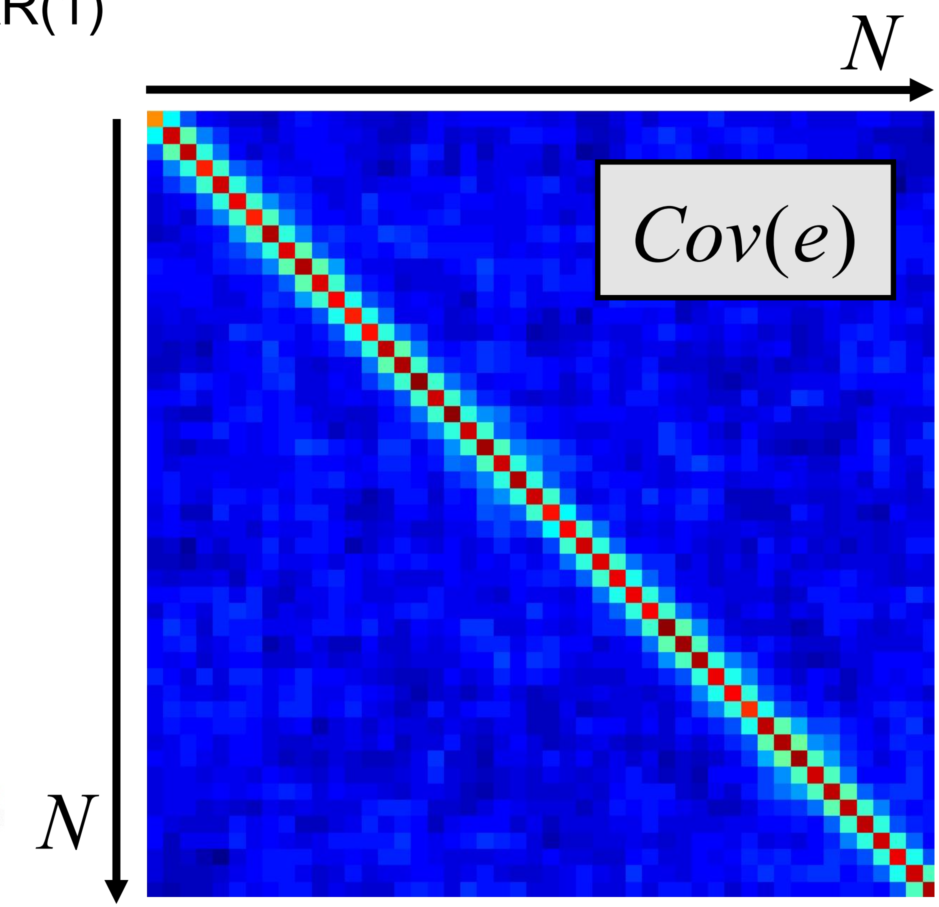


Problem 3: Serial correlations

$$e_t = ae_{t-1} + \varepsilon_t \text{ with } \varepsilon_t \sim N(0, \sigma^2)$$

1st order autoregressive process: AR(1)





Multiple covariance components

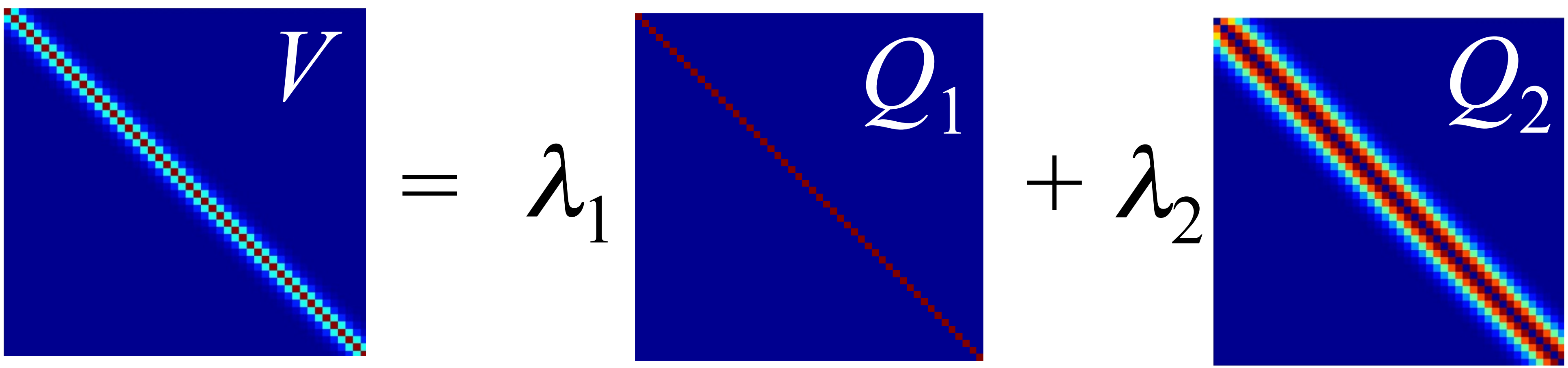
$$e_i \sim N(0, C_i)$$

enhanced noise model at voxel i

$$C_i = \sigma_i^2 V$$

$$V = \sum \lambda_j Q_j$$

error covariance components Q
and hyperparameters λ



Estimation of hyperparameters λ with ReML (Restricted Maximum Likelihood).

Parameters can then be estimated using Weighted Least Squares (WLS)

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

Let

$$W^T W = V^{-1}$$

Then

$$\hat{\beta} = (X^T W^T W X)^{-1} X^T W^T W y$$

$$\hat{\beta} = (X_s{}^T X_s)^{-1} X_s{}^T y_s$$

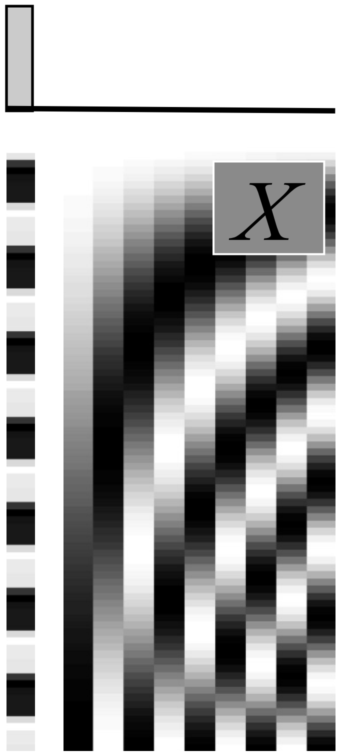
where

$$X_s = WX, y_s = Wy$$

WLS equivalent to OLS on whitened data and design

Contrasts & statistical parametric maps

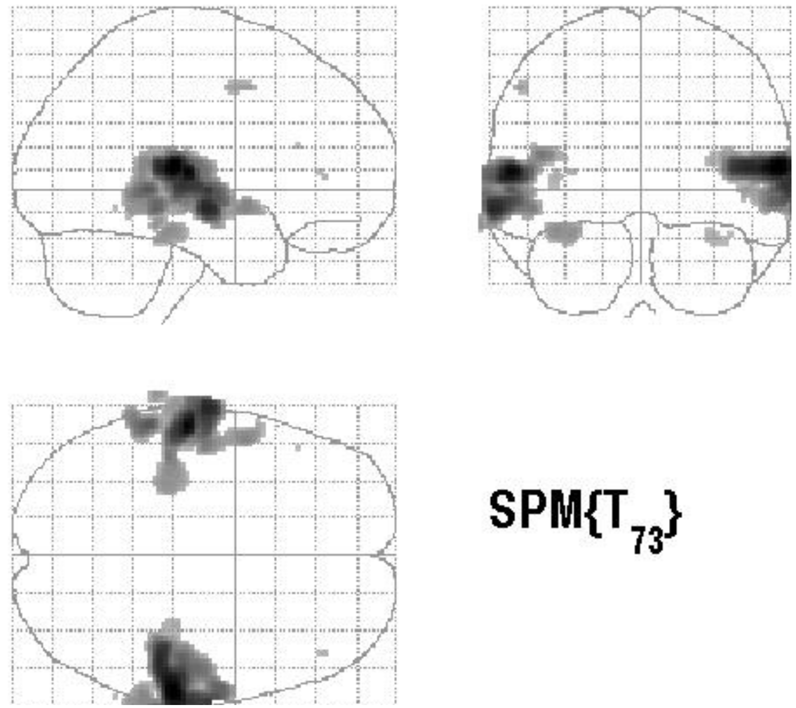
c = 1 0 0 0 0 0 0 0 0 0 0 0



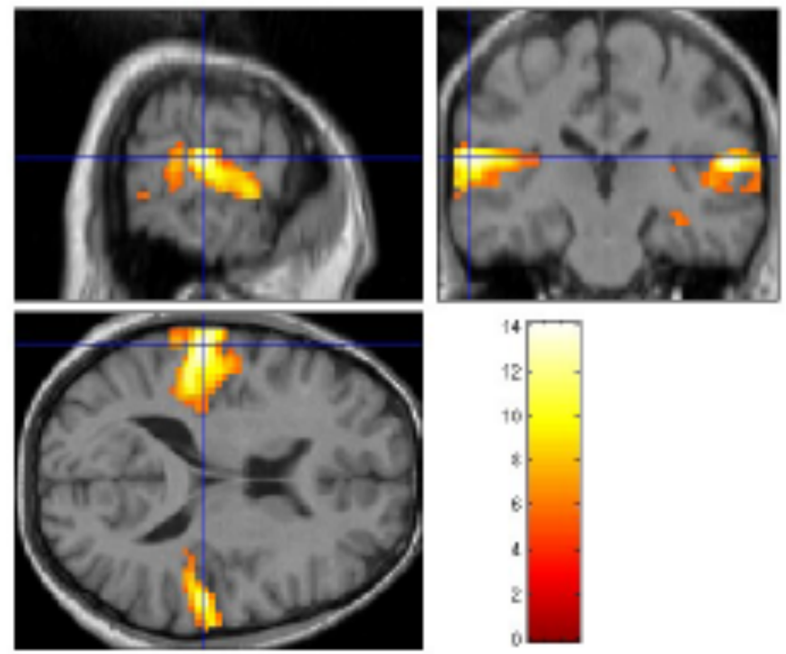
Q: activation during listening ?

Null hypothesis: $\beta_1 = 0$

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



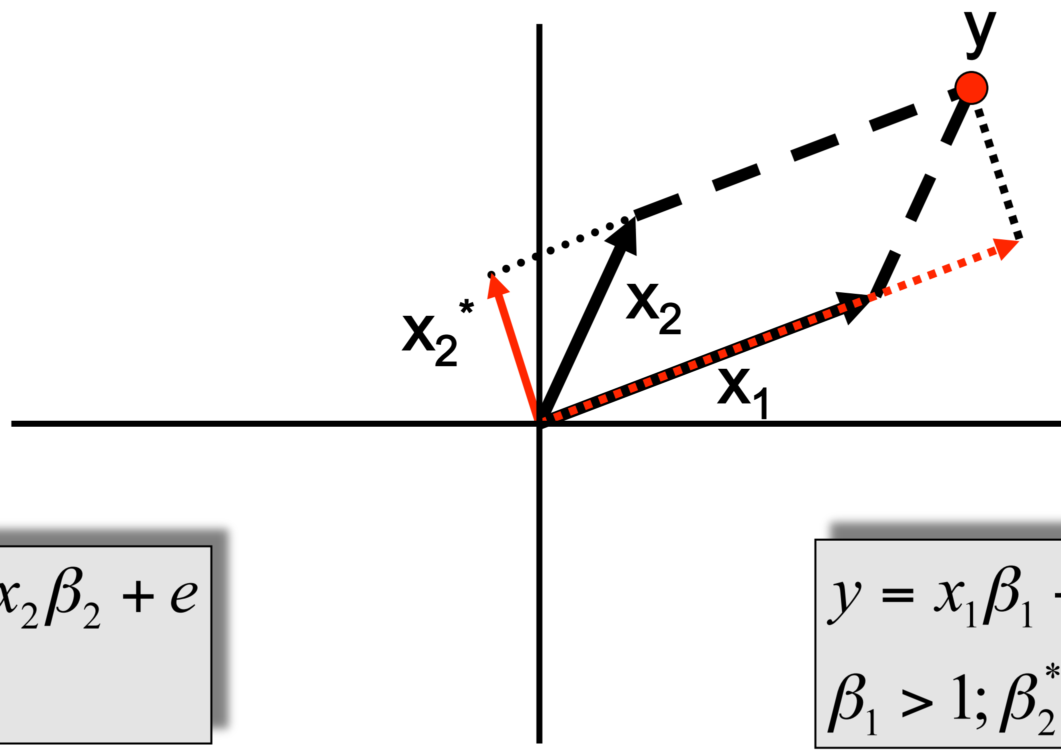
SPM{T_{73}}



Summary

- Mass univariate approach.
- Fit GLMs with design matrix, X, to data at different points in space to estimate local effect sizes, β
- GLM is a very general approach
- Hemodynamic Response Function
- High pass filtering
- Temporal autocorrelation

Correlated and orthogonal regressors



$$y = x_1\beta_1 + x_2\beta_2 + e$$
$$\beta_1 = \beta_2 = 1$$

$$y = x_1\beta_1 + x_2^*\beta_2^* + e$$
$$\beta_1 > 1; \beta_2^* = 1$$

Correlated regressors = explained variance is shared between regressors

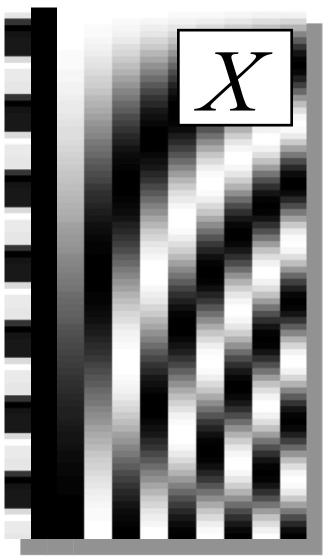
When x_2 is orthogonalized with regard to x_1, only the parameter estimate for x_1 changes, not that for x_2!

t-statistic based on ML estimates

$$Wy = WX\beta + We$$

$$\hat{\beta} = (WX)^{+}Wy$$

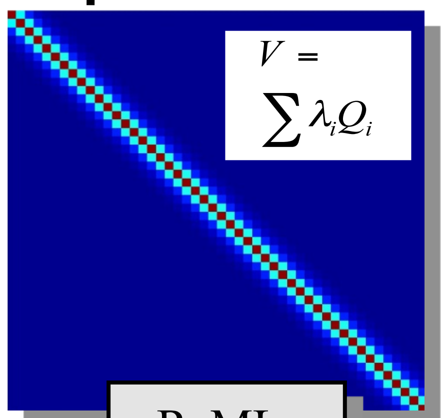
$$c = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0$$



$$t = \frac{c^T \hat{\beta}}{\hat{std}(c^T \hat{\beta})}$$

$$W = V^{-1/2}$$

$$\sigma^2 V = Cov(e)$$



ReML-estimates

$$\hat{std}(c^T \hat{\beta}) = \sqrt{\hat{\sigma}^2 c^T (WX)^{+} (WX)^{+T} c}$$

$$\hat{\sigma}^2 = \frac{\sum \left(Wy - WX\hat{\beta}\right)^2}{tr(R)}$$

$$R = I - WX(WX)^{+}$$

For brevity:

$$(WX)^{+} = (X^T W X)^{-1} X^T$$