The general linear model for fMRI

Methods and Models in fMRI, 17.10.2017

Jakob Heinzle
heinzle@biomed.ee.ethz.ch

Translational Neuromodeling Unit (TNU)
Institute for Biomedical Engineering (IBT)
University and ETH Zürich

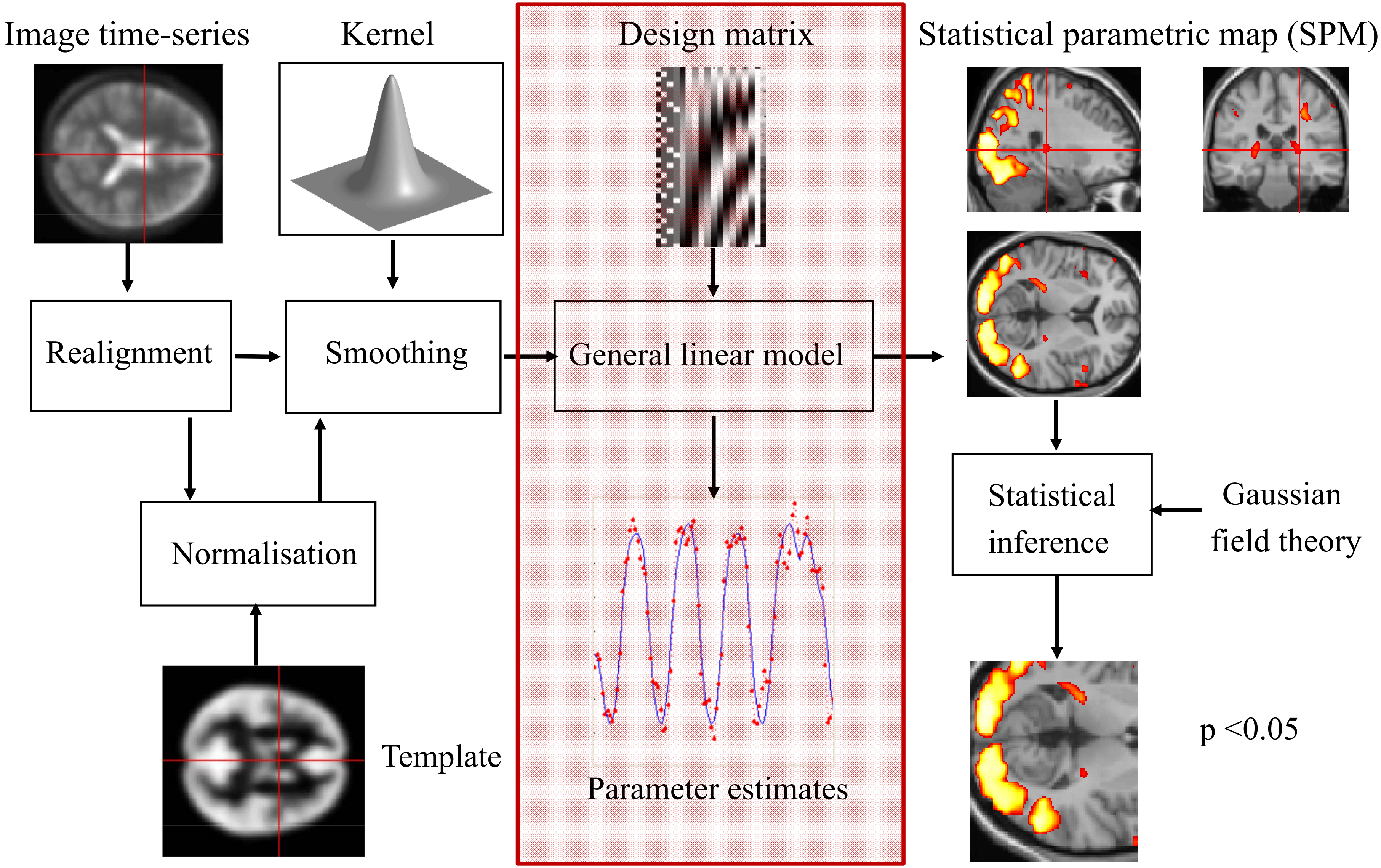
Many thanks to
K. E. Stephan and
F. Petzschner for material

Translational Neuromodeling Unit

University of Zurich[UZH]

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Overview of SPM



What is the problem we want to solve?

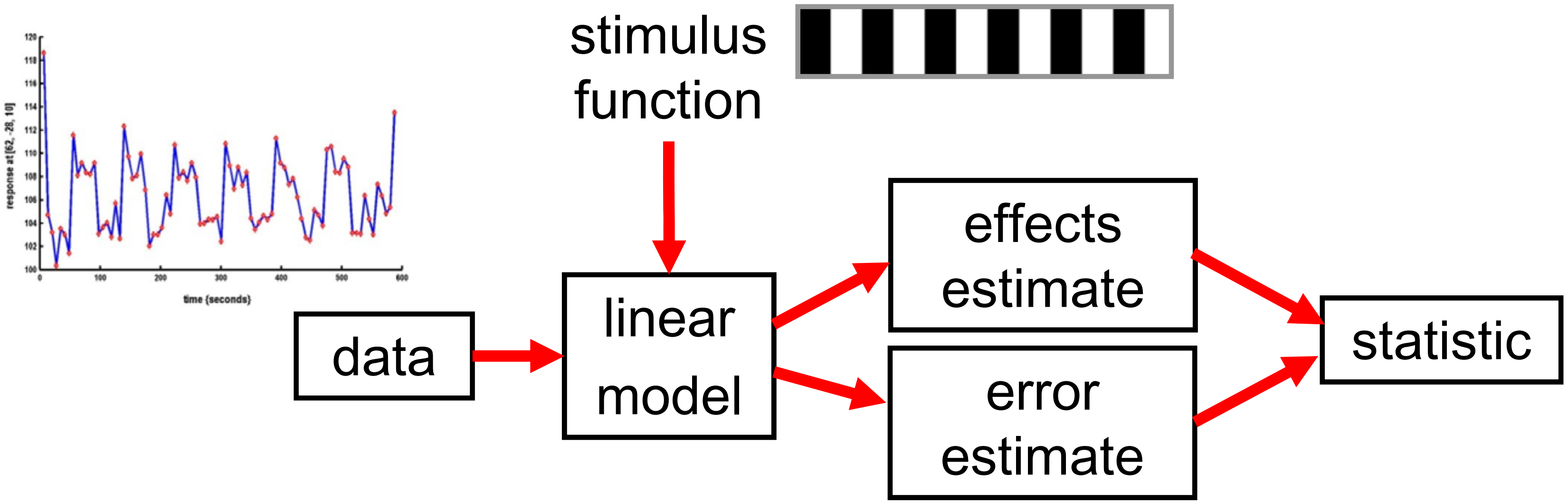
- We have an experimental paradigm and want to test whether brain activity is (linearly) related to the paradigm.
- We will try to solve the problem by modeling the data.

Modelling the measured data

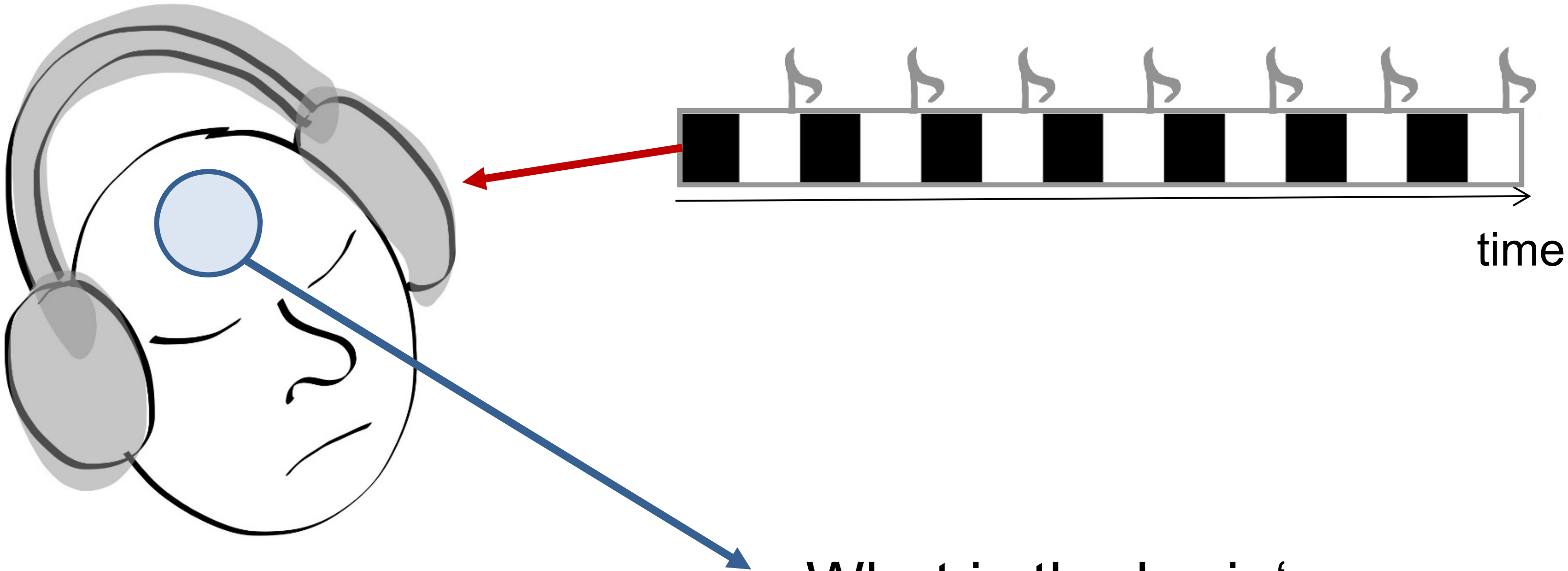
Why? Make inferences about effects of interest

How?

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



A very simple experiment

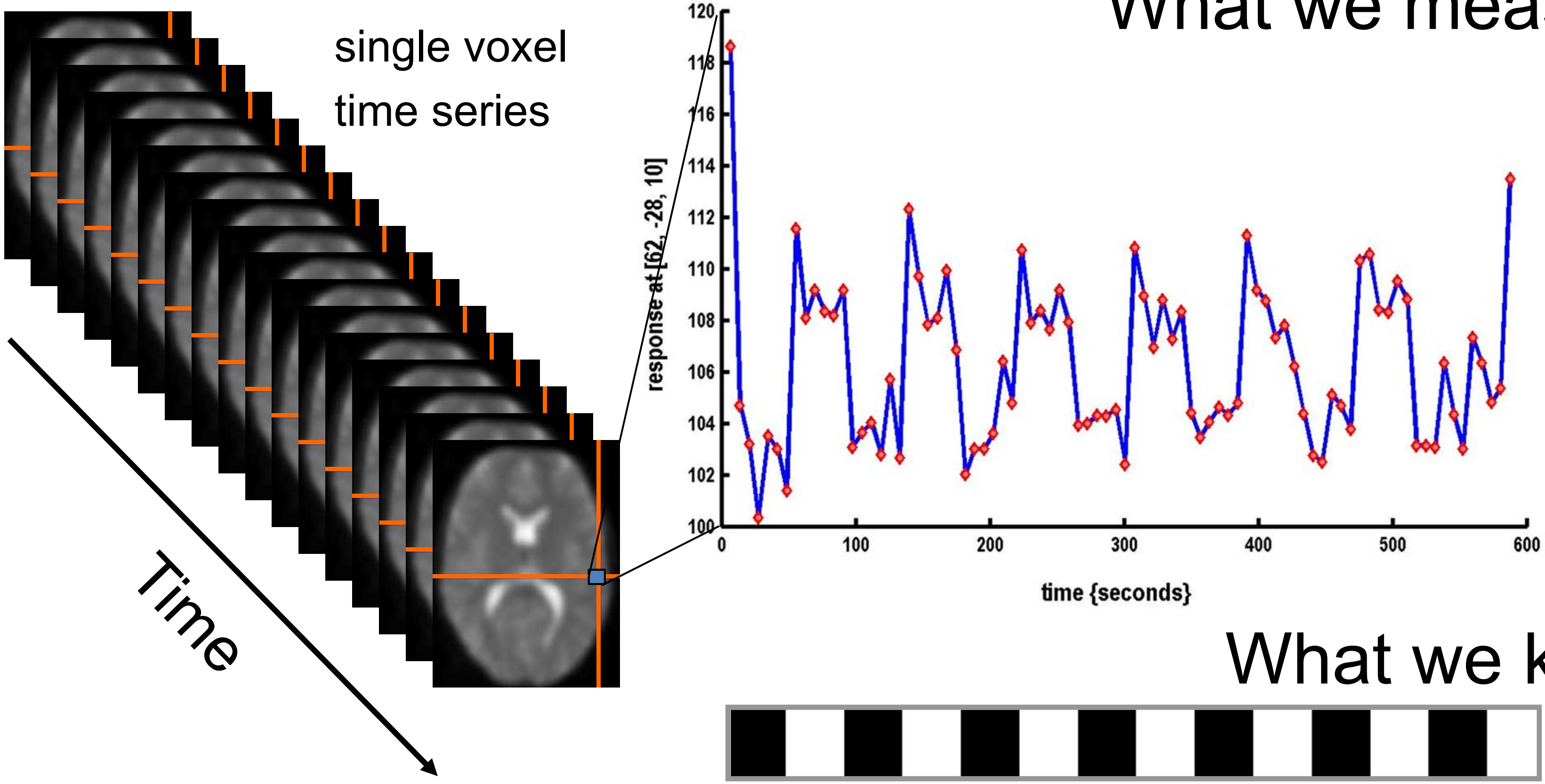


- One session
- 7 cycles of rest and listening
- Blocks of 6 scans with 7 sec TR

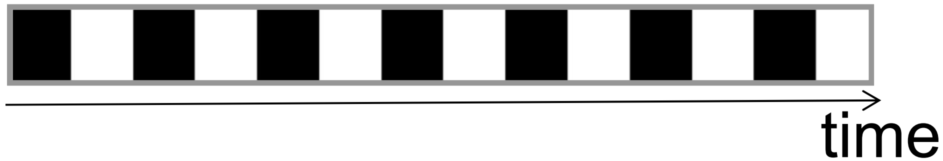
What is the brain‘s response to such a stimulation?

How is brain data related to the input?

What we measure.



What we know.

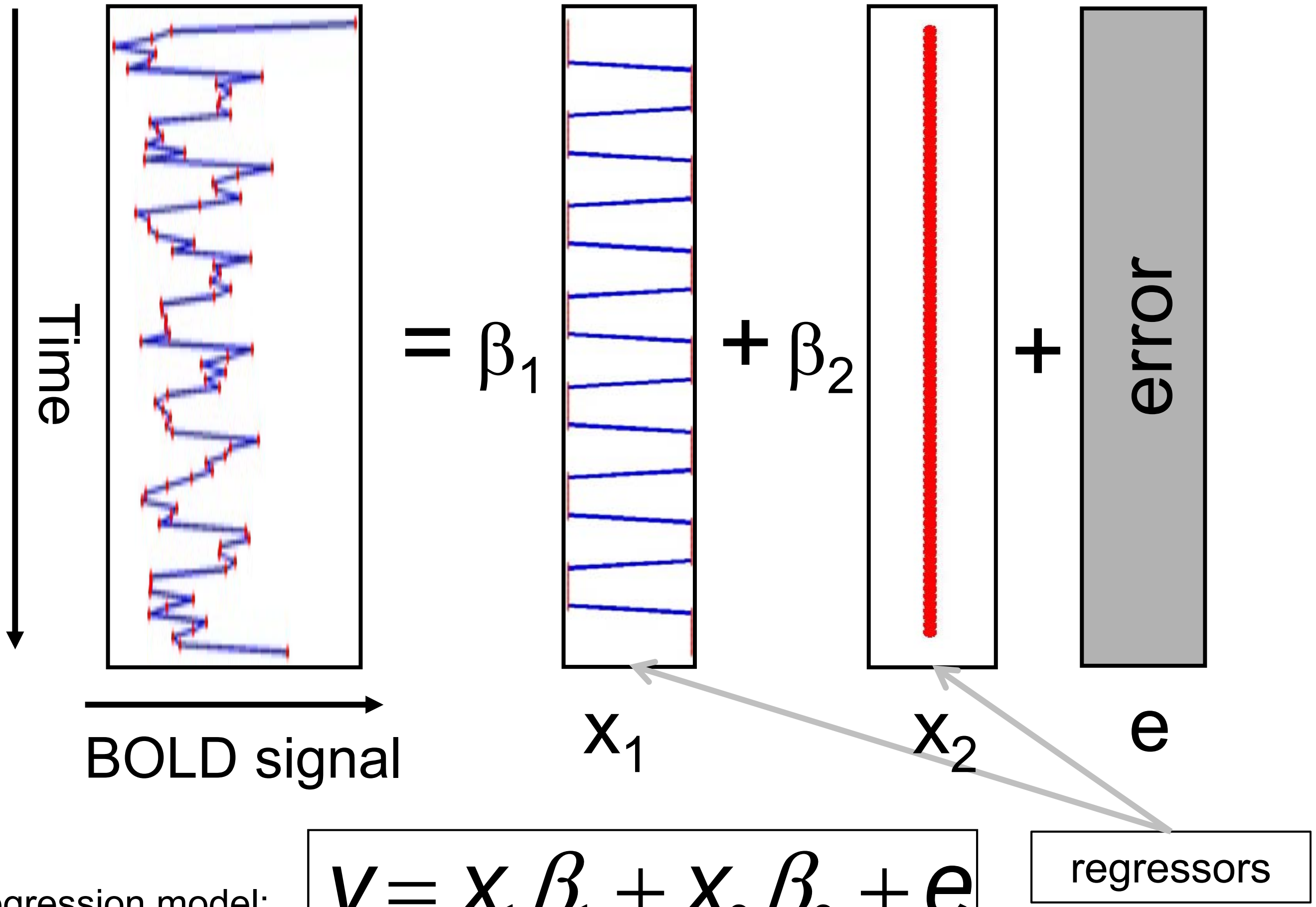


Question: Is there a change in the BOLD response between listening and rest?

A linear model of the data

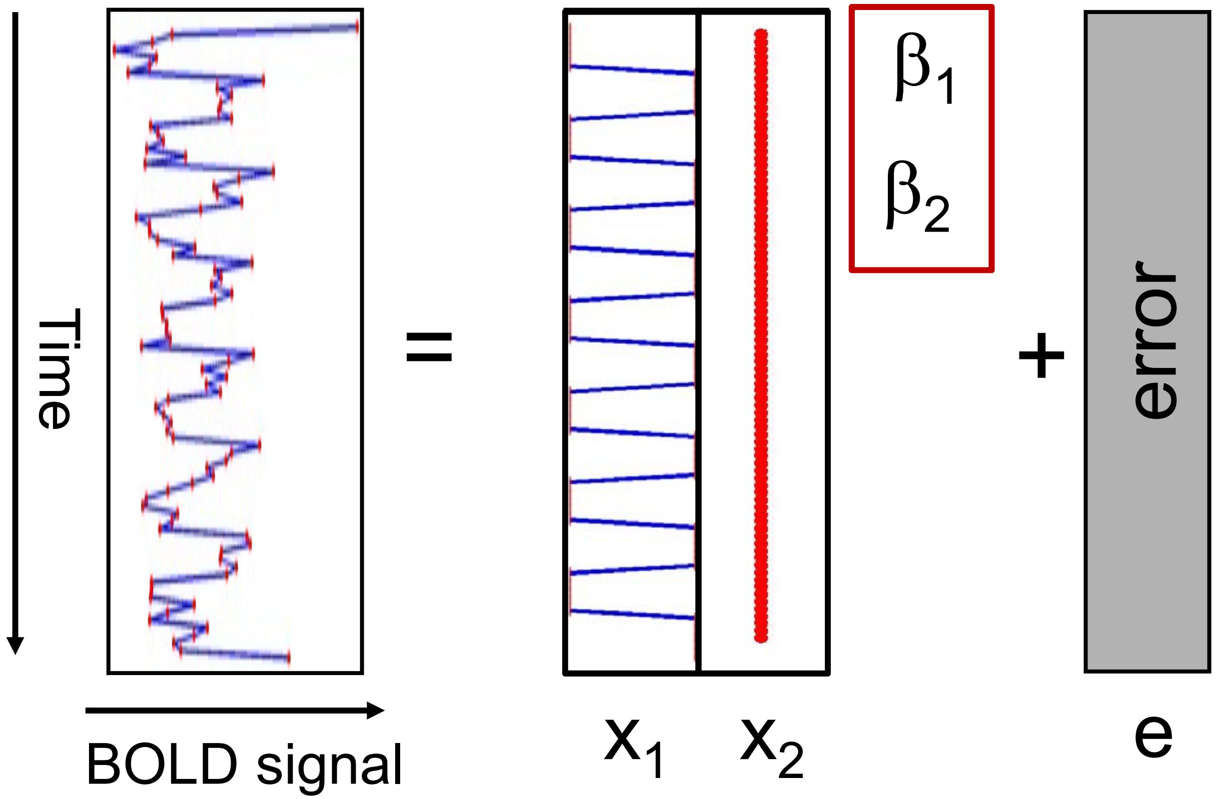
Explain your data…

as a combination of experimental manipulation,confounds and errors



Single voxel regression model: $$y = x_1\beta_1 + x_2\beta_2 + e$$

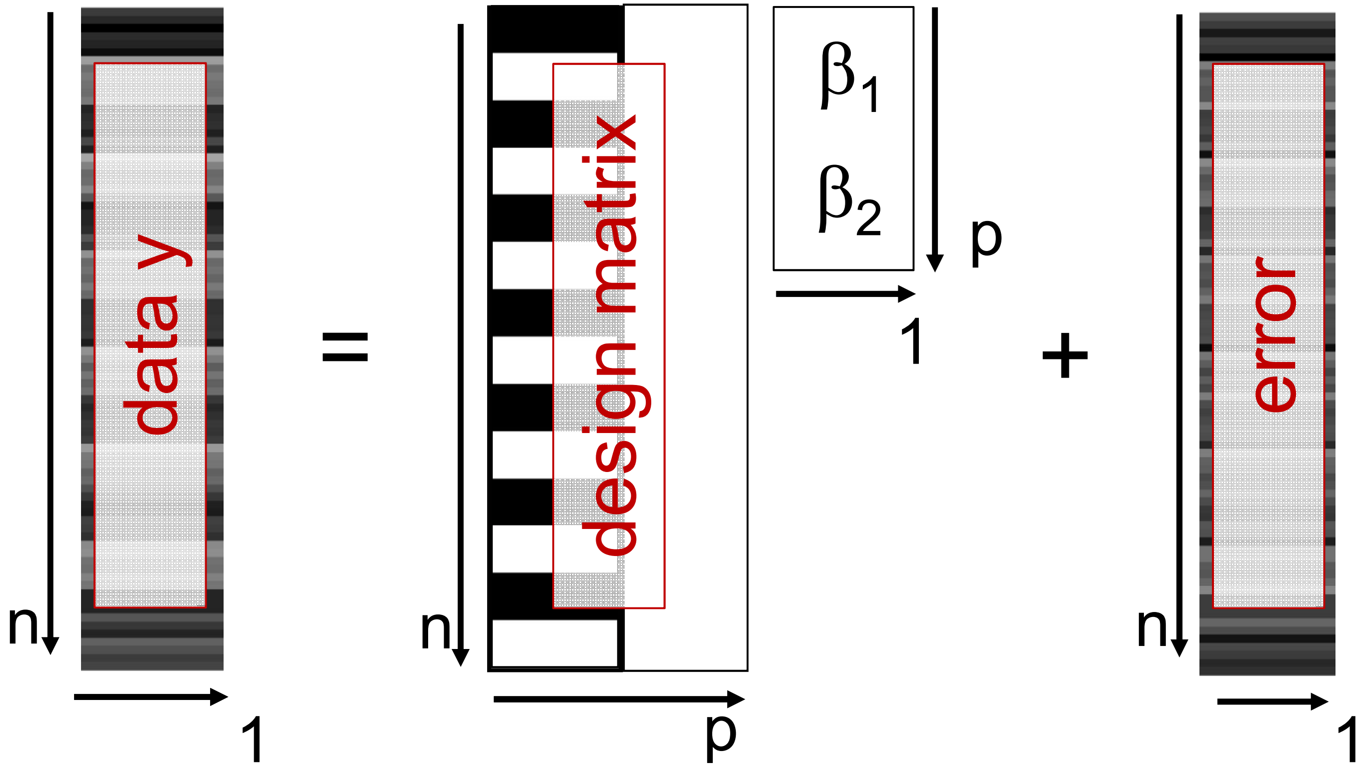
Writing everything in matrix notation



Single voxel regression model:

$$y = X\beta + e$$

The way it looks in SPM



n: number of scans

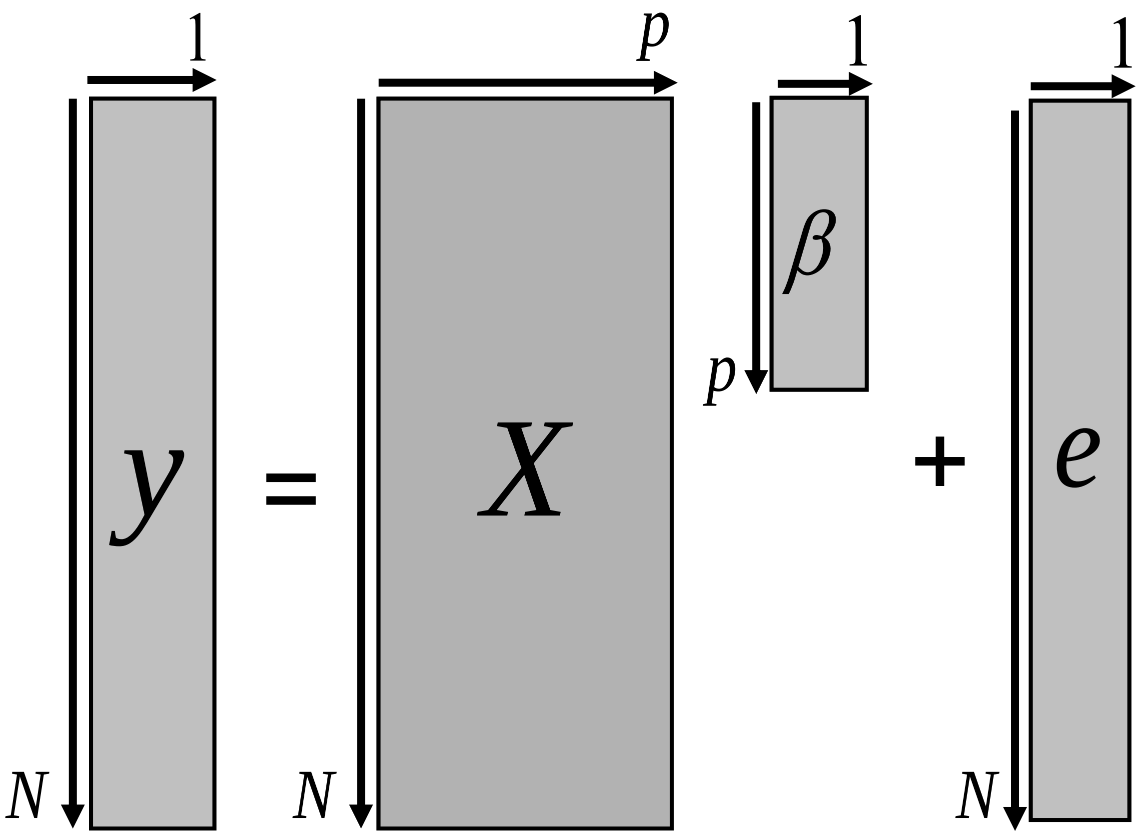
p: number of regressors

$$y = X\beta + e$$

We need …

- … to specify the design matrix.
- … specify a noise model, e.g. $e \sim N(0, \sigma^2 I)$
- … and then, estimate the parameters b that minimize the error $\sum_{t=1}^{N} e_t^2$
  - Minimization of the error depends on assumptions about the noise.

Summary: Mass-univariate GLM



$$y = X\beta + e$$

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

Model is specified by

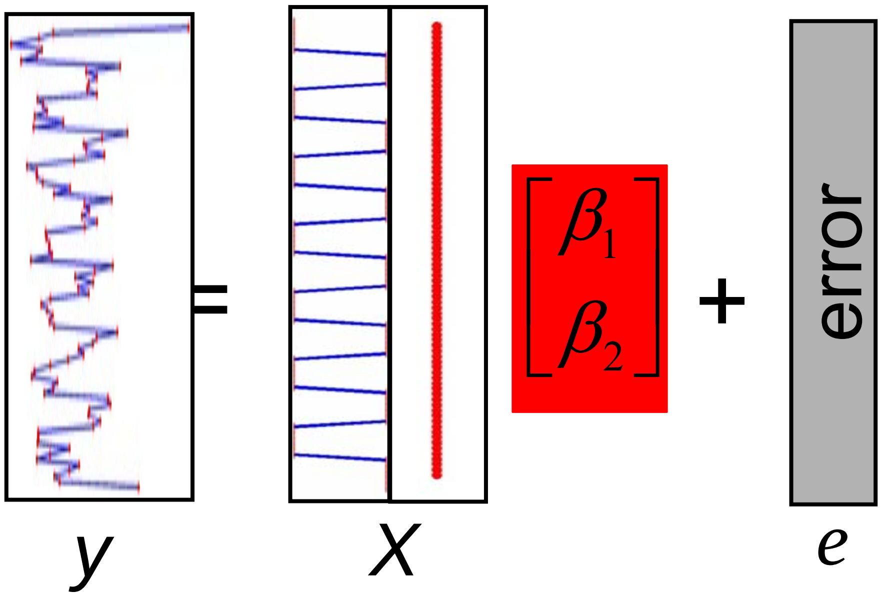
1. Design matrix X
2. Assumptions about e

N: number of scans

p: number of regressors

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

How to fit the model parameters.



OLS (Ordinary Least Squares)

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

$e = y - \hat{y}$

$e = y - X\hat{\beta}$

Data predicted by our model

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

e = error between predicted and actual data

Goal is to determine the betas that minimize the quadratic error

OLS – Ordinary least squares

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

We want to minimize the quadratic error between data and model

OLS – Ordinary least squares

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

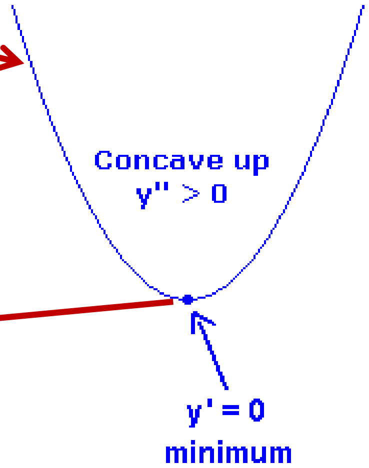
$$e^T e = (y^T - \hat{\beta}^T X^T)(y - X\hat{\beta})$$

$$e^T e = y^T y - y^T X\hat{\beta} - \hat{\beta}^T X^T y + \hat{\beta}^T X^T X\hat{\beta}$$

$$e^T e = y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X\hat{\beta}$$

$$\frac{\partial e^T e}{\partial \hat{\beta}} = -2X^T y + 2X^T X\hat{\beta}$$

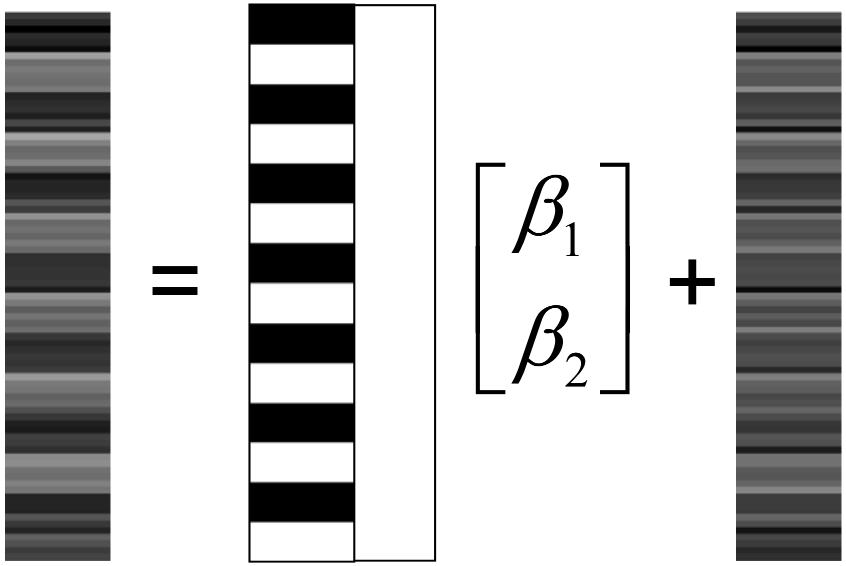
$$0 = -2X^T y + 2X^T X\hat{\beta}$$



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

OLS estimate for β

Summary: OLS solution



$$y \quad X \begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix} \quad e$$

$$y = X\beta + e$$

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

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

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

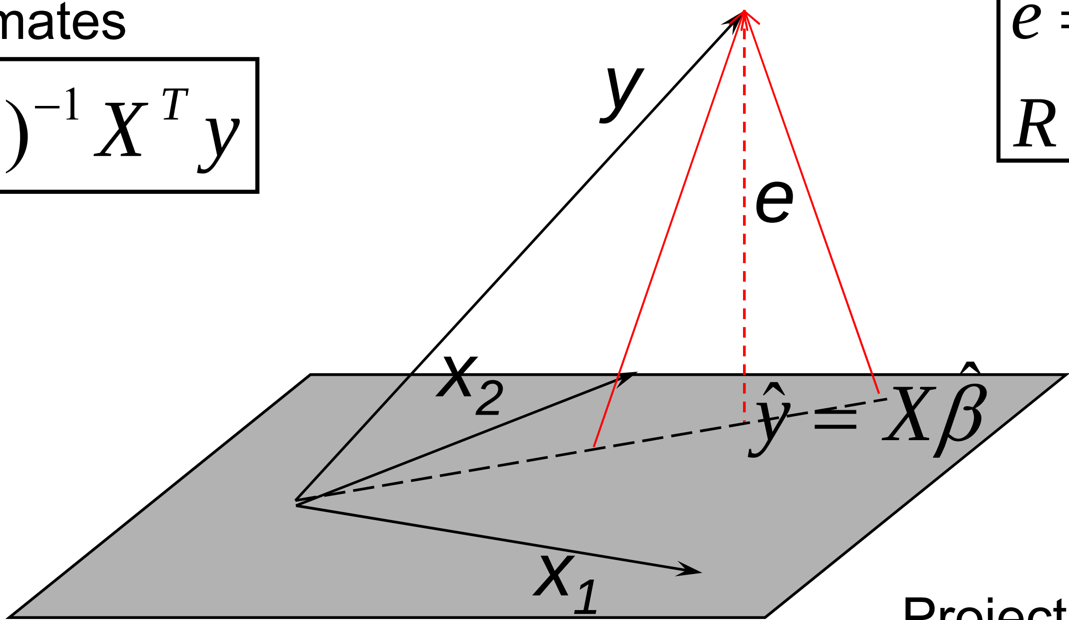
Geometric perspective

Residual
forming matrix R

$$e = Ry$$
$$R = I - P$$

OLS estimates

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



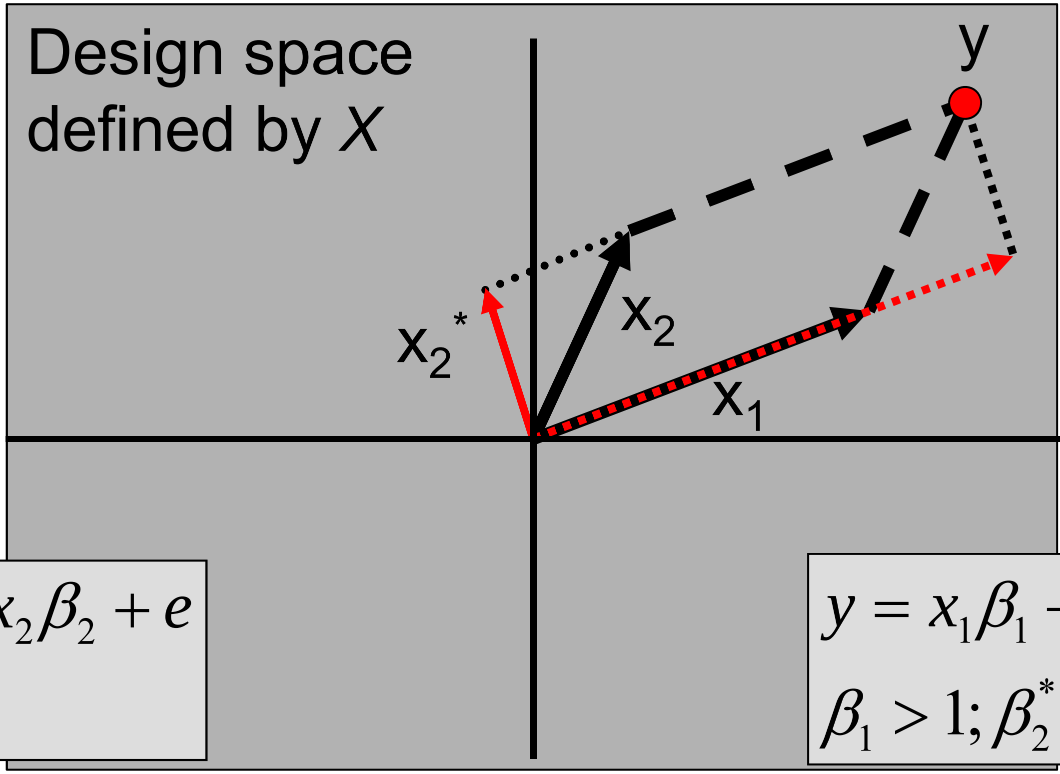
$$\hat{y} = X\hat{\beta}$$

Design space
defined by *X*

Projection matrix P

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

Correlated and orthogonalized regressors



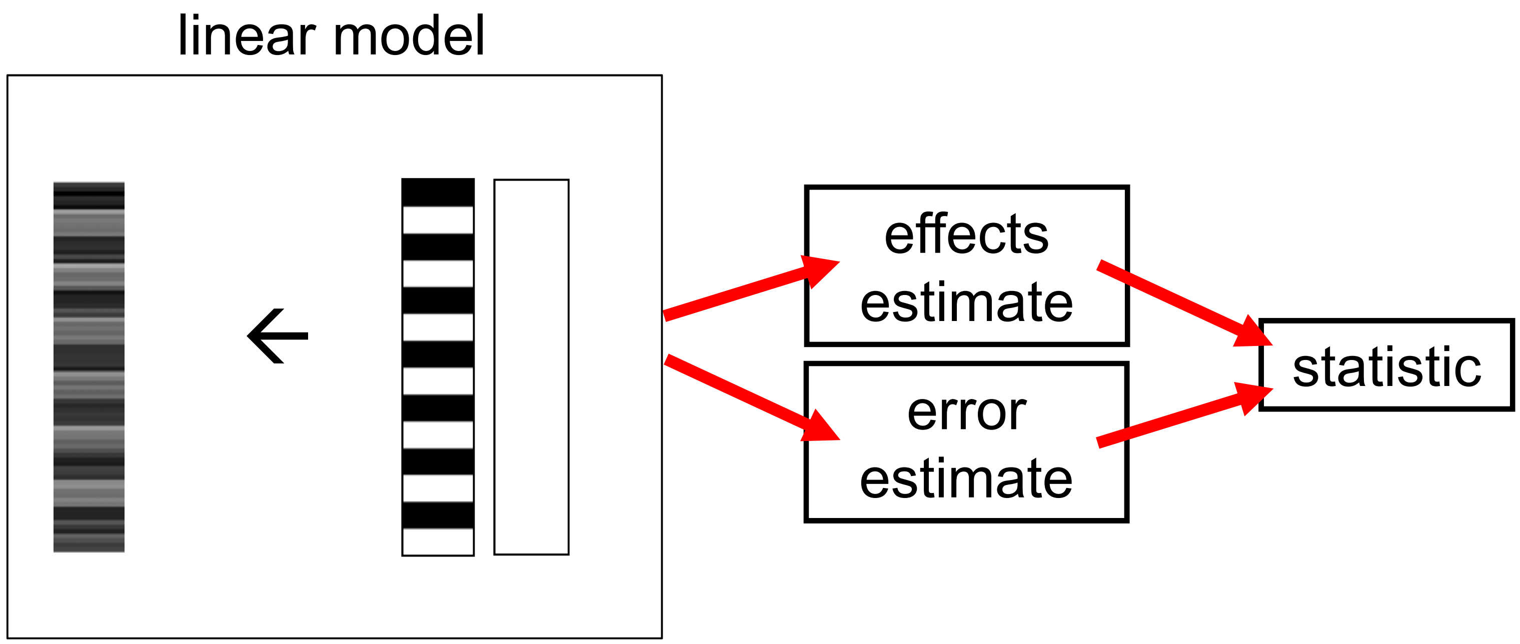
$$y = x_1\beta_1 + x_2\beta_2 + e$$
$$\beta_1 = \beta_2 = 1$$

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

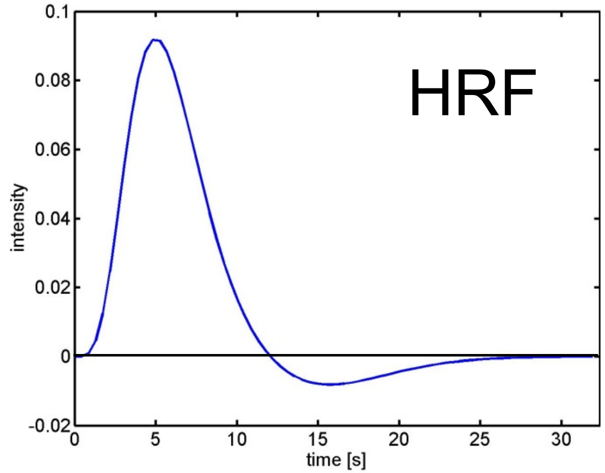
Correlated regressors = explained variance is shared between regressors

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

We are nearly there …



Problems of this model

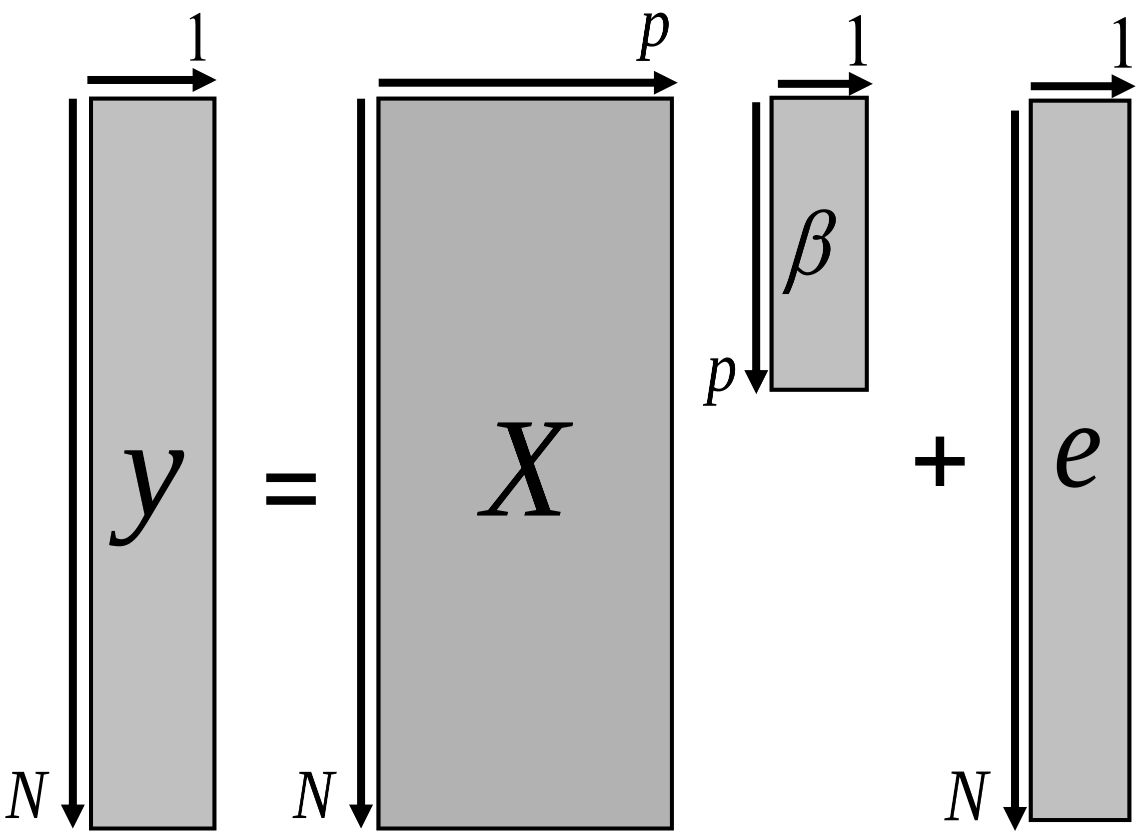


1. BOLD responses have a delayed and dispersed form (cf. Lecture 1).

2. The BOLD signal includes substantial amounts of low-frequency noise.

3. The data are serially correlated (temporally autocorrelated) $\rightarrow$ this violates the assumptions of the noise model in the GLM

Summary: Mass-univariate GLM



$$y = X\beta + e$$

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

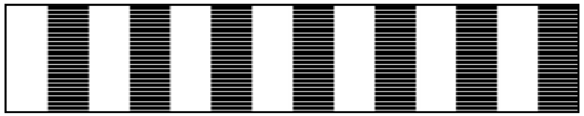
Model is specified by

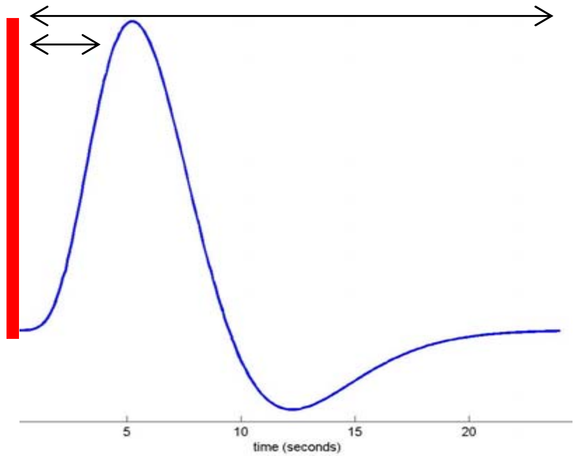
1. Design matrix X
2. Assumptions about e

N: number of scans
p: number of regressors

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

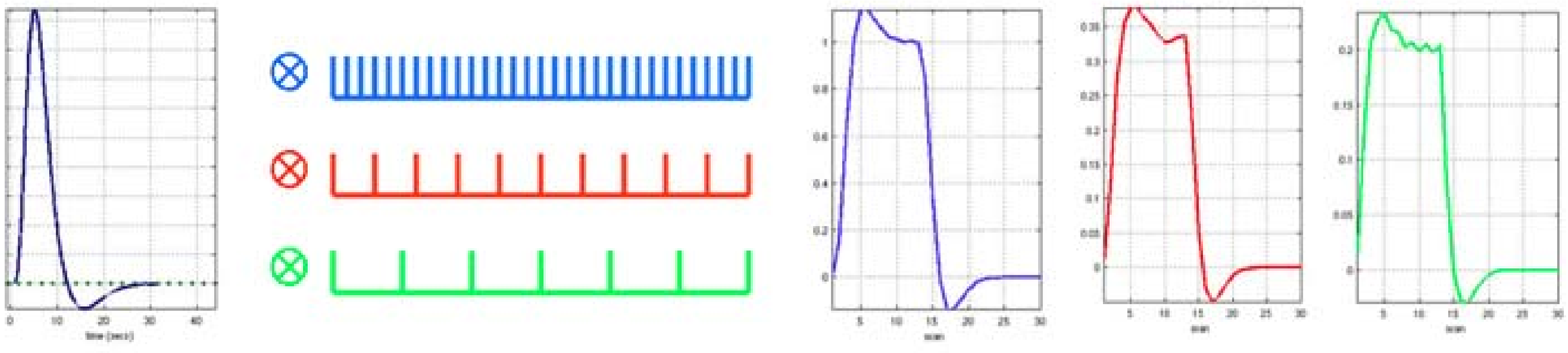
Problem 1: The BOLD response





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

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).

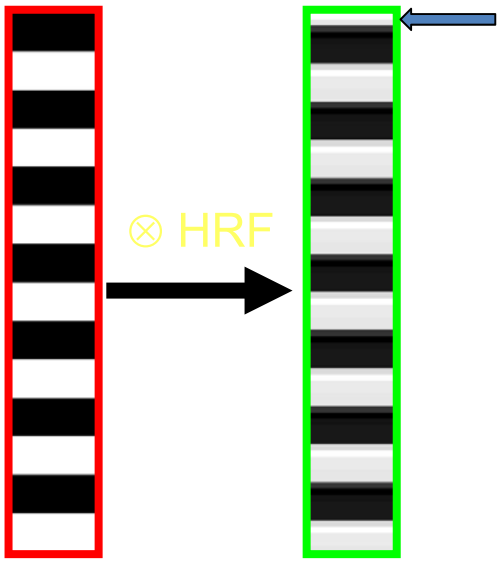


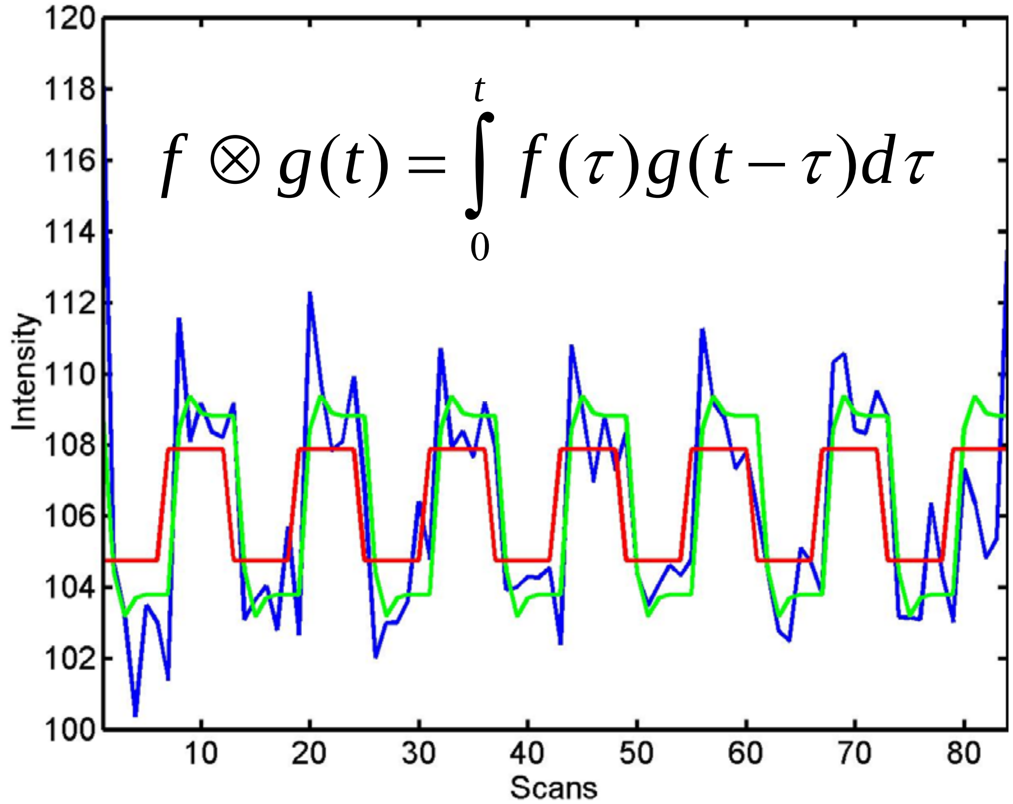
Basic math: What is a convolution?

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

Solution: Convolution with the HRF

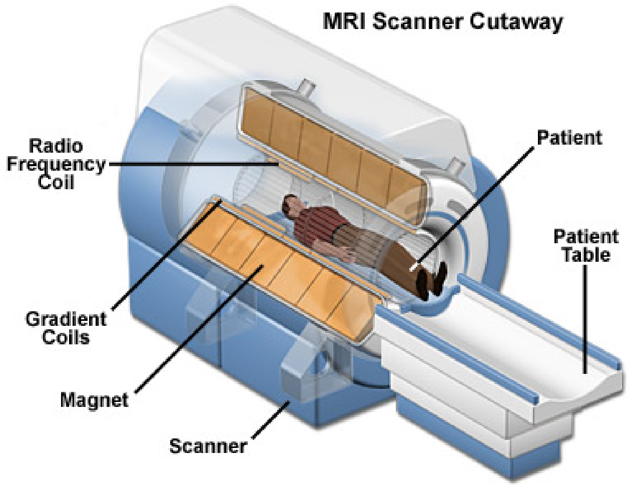
expected BOLD response
= input function ⊗ impulse response function (HRF)

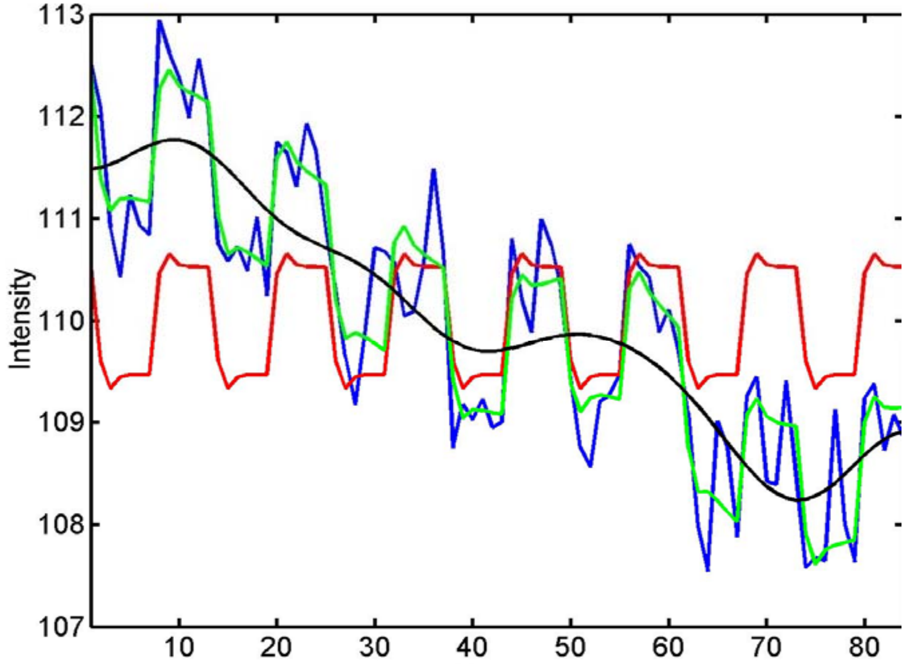




blue =	data
green =	predicted response, taking convolved with HRF
red =	predicted response, NOT taking into account the HRF

Problem 2: Low frequency noise





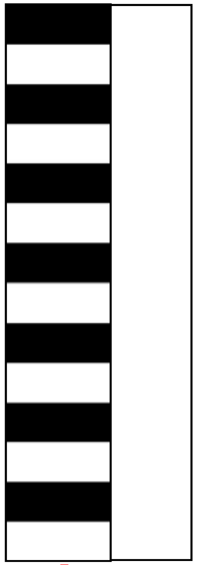
blue = data

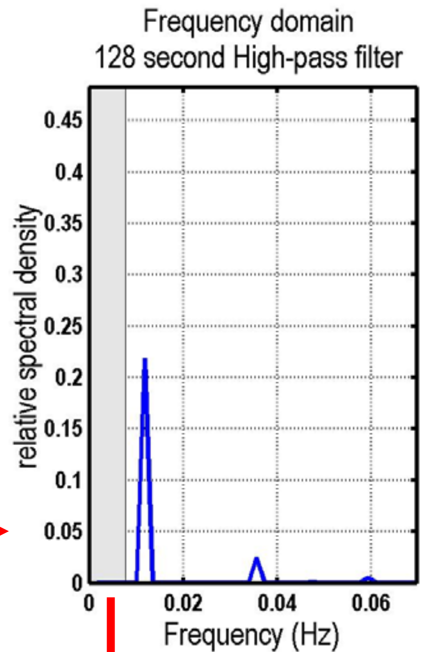
black = mean + low-frequency drift

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

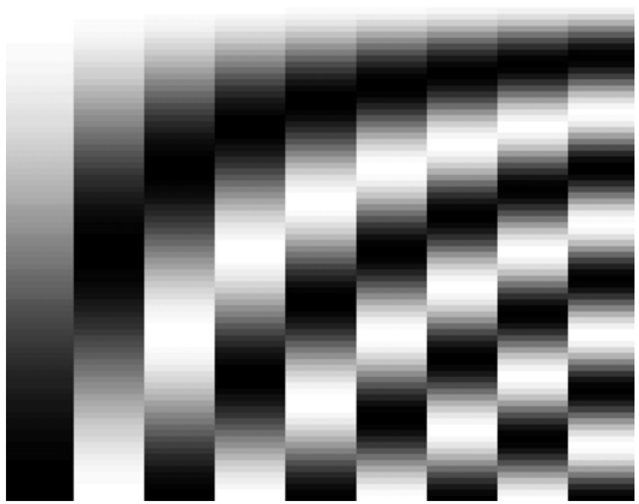
red = predicted response, NOT taking into account low-frequency drift

Solution 2: High-pass filtering



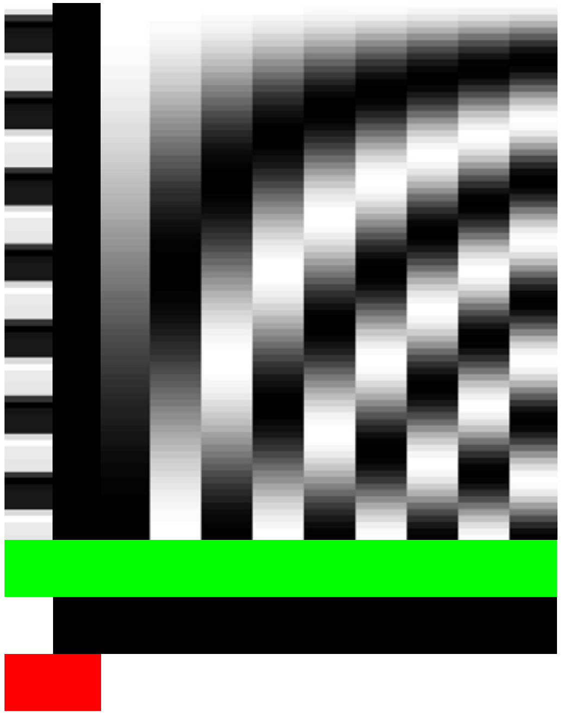


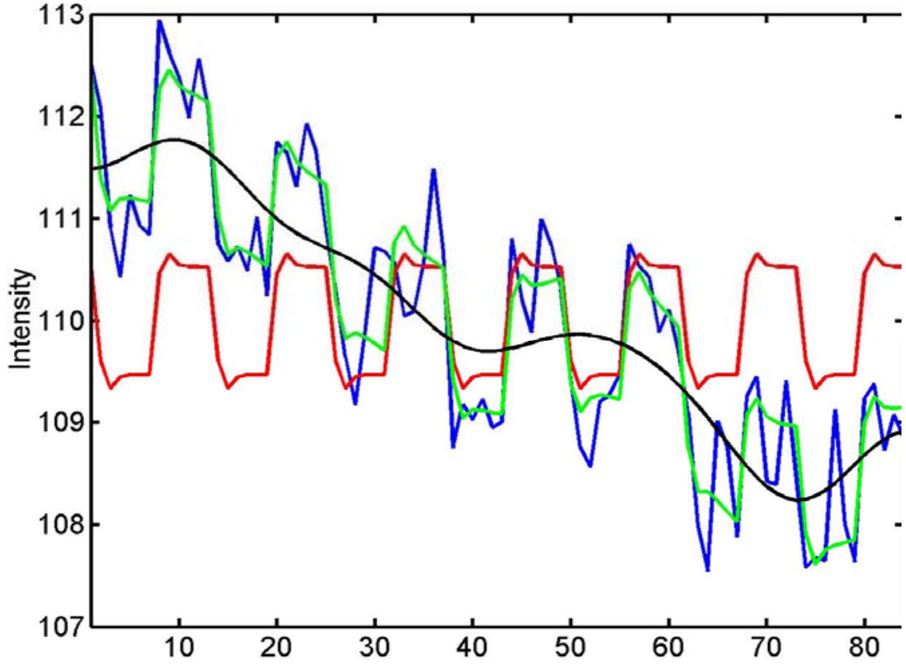
discrete cosine transform (DCT) set



Solution 2: High-pass filtering

Linear model





blue = data

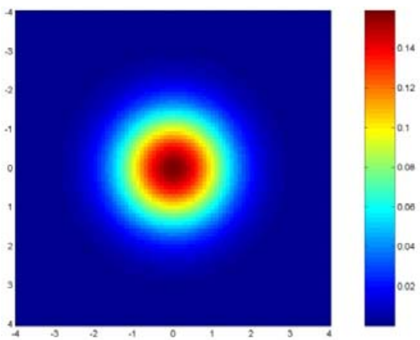
black = mean + low-frequency drift

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

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

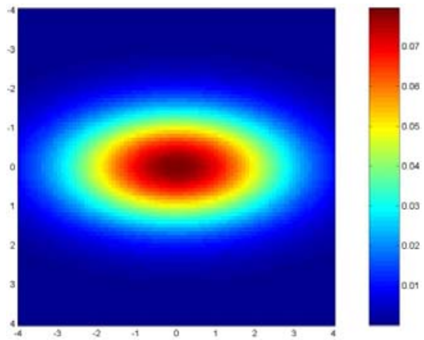
Problem 3: Serial correlations

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



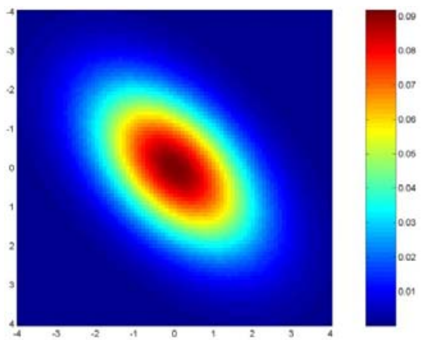
$$Cov(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Examples for non-sphericity:



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

non-identity



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

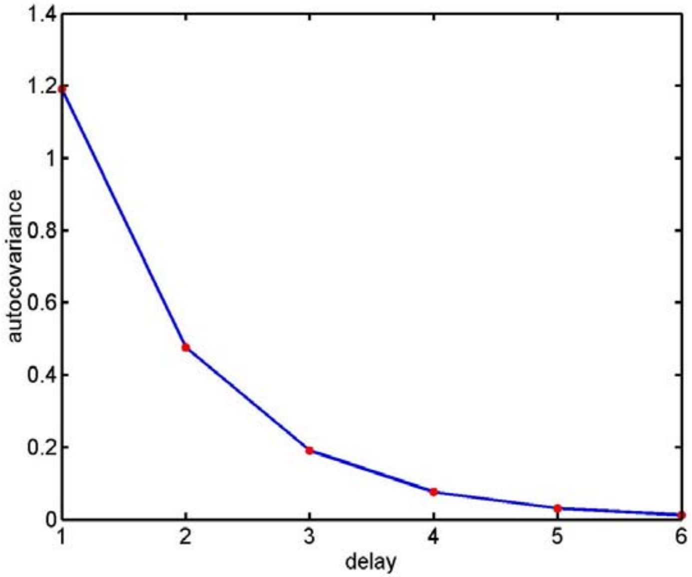
non-independence

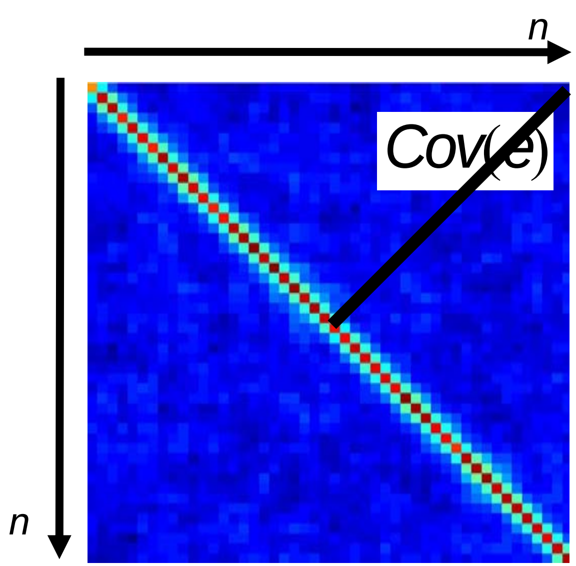
Problem 3: Serial correlations

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

1st order autoregressive process: AR(1)

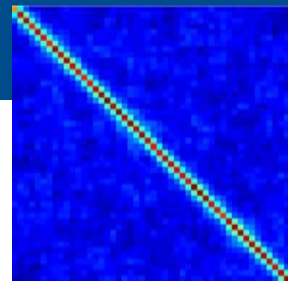
autocovariance function





n: number of scans

Solution 3: Pre-whitening



- *Pre-whitening*:

  1. Use an enhanced noise model with multiple error covariance components, i.e. e ~ $N(0, \sigma^2 V)$ instead of e ~ $N(0, \sigma^2 I)$.

  2. Use estimated serial correlation to specify filter matrix *W* for whitening the data.

This is i.i.d

$$Wy = WX\beta + \underbrace{We}$$

How to define W?

- Enhanced noise model

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

- Remember linear transform for Gaussians

$$x \sim N(\mu, \sigma^2), y = ax$$
$$\Rightarrow y \sim N(a\mu, a^2\sigma^2)$$

- Choose *W* such that error covariance becomes spherical

$$We \sim N(0, \sigma^2 W^2 V)$$
$$\Rightarrow W^2 V = I$$
$$\Rightarrow W = V^{-1/2}$$

- Conclusion: *W* is a simple function of *V* $\Rightarrow$ so how do we estimate *V* ?

$$Wy = WX\beta + We$$

Find W – multiple covariance components.

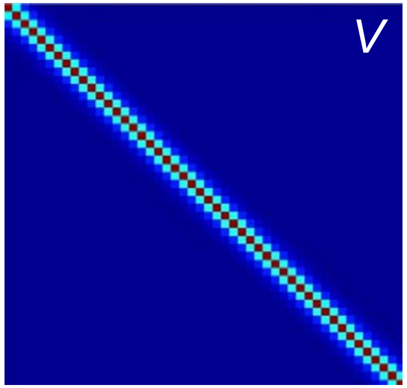
$$e \sim N(0, \sigma^2 V)$$

enhanced noise model

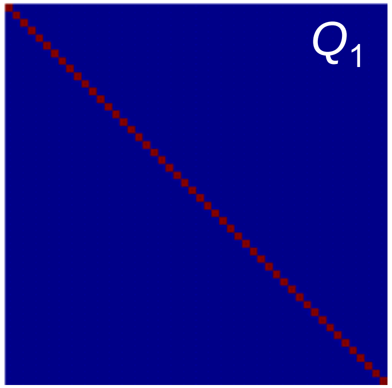
$$V \propto Cov(e)$$

$$V = \sum \lambda_i Q_i$$

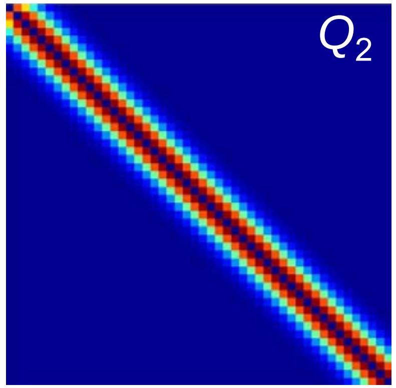
error covariance components Q
and hyperparameters λ



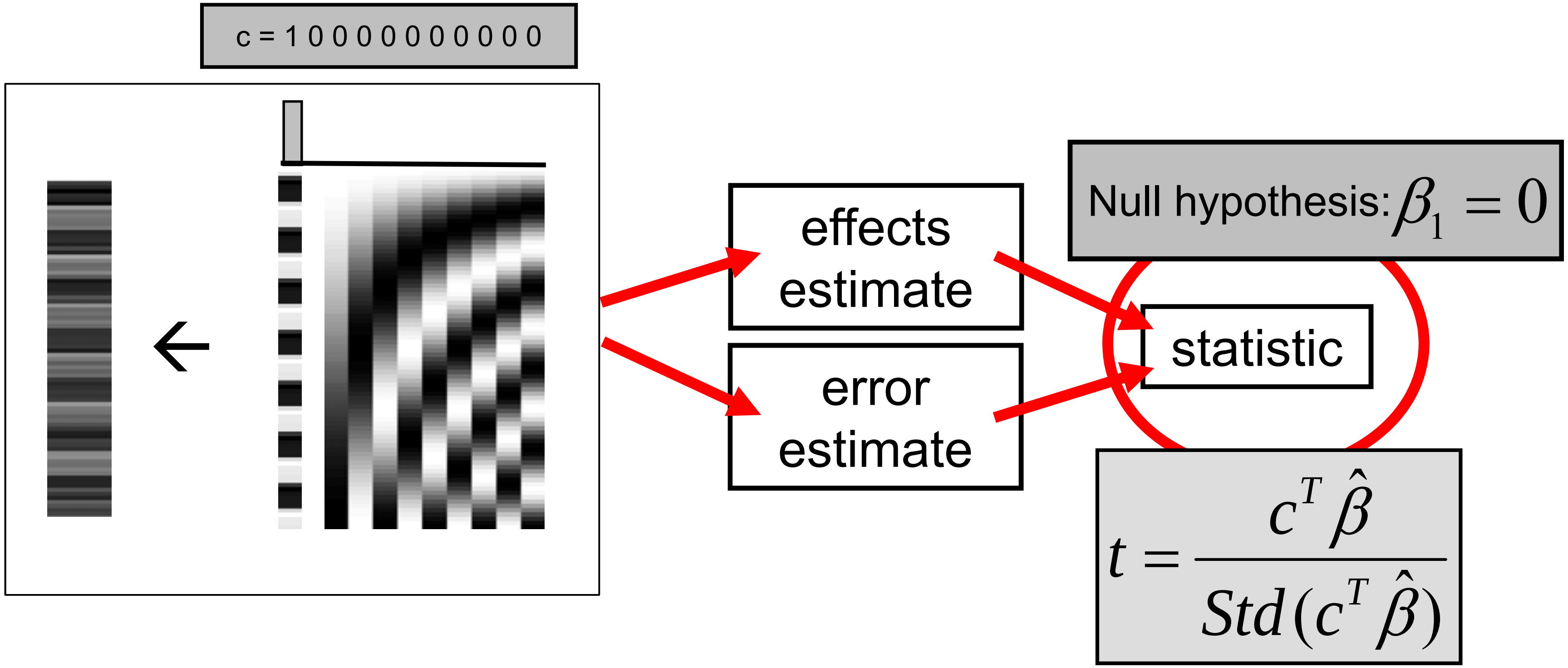
$= \lambda_1$



$+ \lambda_2$



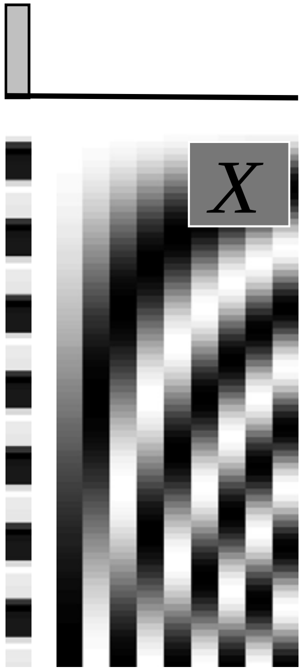
Estimation of hyperparameters λ with EM (expectation maximisation) or ReML (restricted maximum likelihood). For more details see (Friston et al, Neuroimage, 16:465; 2002)



→ Lecture: Classical (frequentist) inference

Outlook: Contrasts and statistical maps

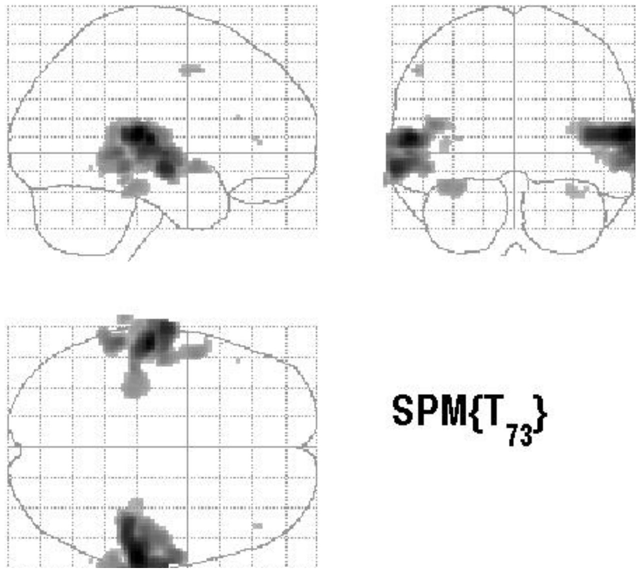
c = 1 0 0 0 0 0 0 0 0 0 0

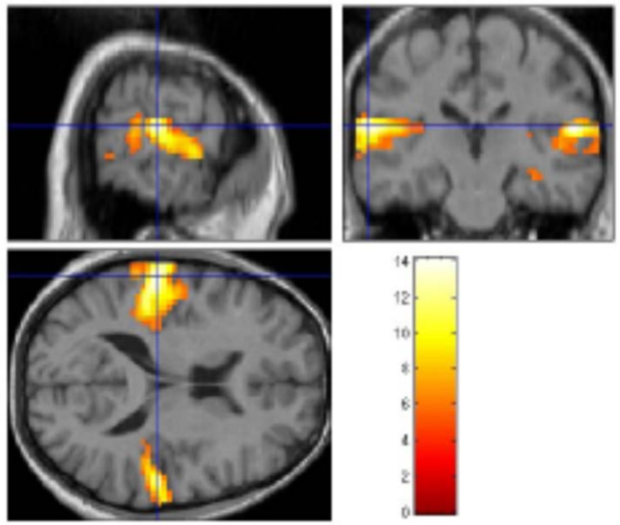


Q: activation during listening ?

Null hypothesis: $\beta_1 = 0$

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



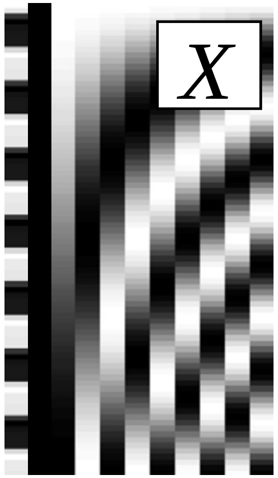


Summary of GLM

$$Wy = WX\beta + We$$

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

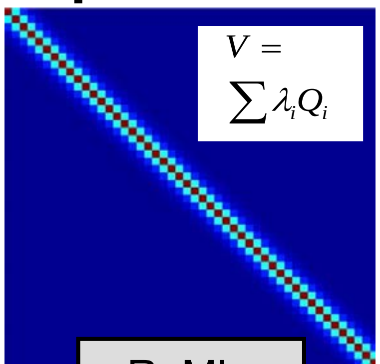
$$c = 1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0$$



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

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

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



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:

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

Physiological confounds

- head movements
- arterial pulsations (particularly bad in brain stem)
- breathing
- eye blinks (visual cortex)
- adaptation effects, fatigue, fluctuations in concentration, etc.

→ Lecture: Noise models in fMRI and noise correction

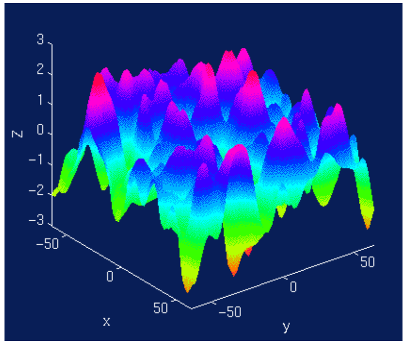
Outlook – further challenges

- correction for multiple comparisons
- variability in the HRF across voxels
- slice timing
- limitations of frequentist statistics
  $\rightarrow$ Bayesian analyses
- GLM ignores interactions among voxels
  $\rightarrow$ models of effective connectivity

These issues are discussed in future lectures.

Correction for multiple comparison

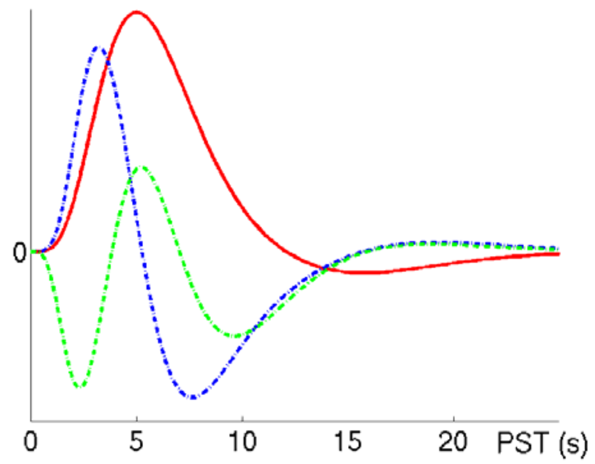
- Mass-univariate approach:
  We apply the GLM to each of a huge number of voxels (usually > 100,000).
- Threshold of $p<0.05 \rightarrow$ more than 5000 voxels significant by chance!
- Massive problem with multiple comparisons!
- Solution: Gaussian random field theory



→ Lecture: Multiple comparison correction

Variability in the BOLD response

- HRF varies substantially across voxels and subjects
- For example, latency can differ by ± 1 second
- Solution: use multiple basis functions
- See talk on event-related fMRI

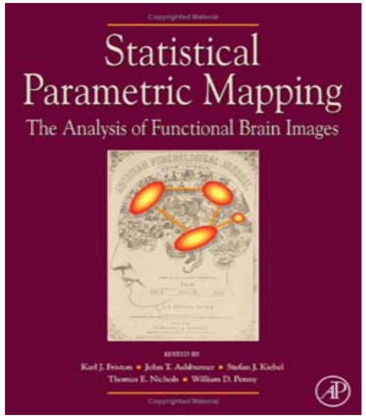


Summary

- Mass-univariate approach: same GLM for each voxel
- GLM includes all known experimental effects and confounds
- Convolution with a canonical HRF
- High-pass filtering to account for low-frequency drifts, implemented by a set of cosine functions.
- Estimation of multiple variance components (e.g. to account for serial correlations)

Bibliography

Friston, Ashburner, Kiebel, Nichols, Penny (2007) *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier.

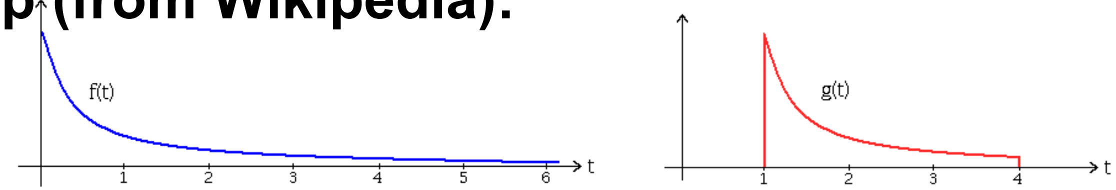


- Christensen R (1996) *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer.
- Friston KJ et al. (1995) Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* 2: 189-210.

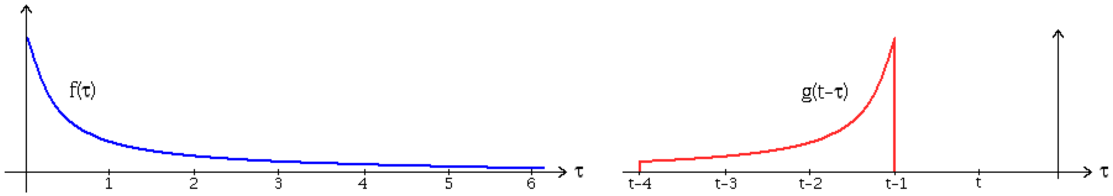
Supplementary slides

Convolution step-by-step (from Wikipedia):

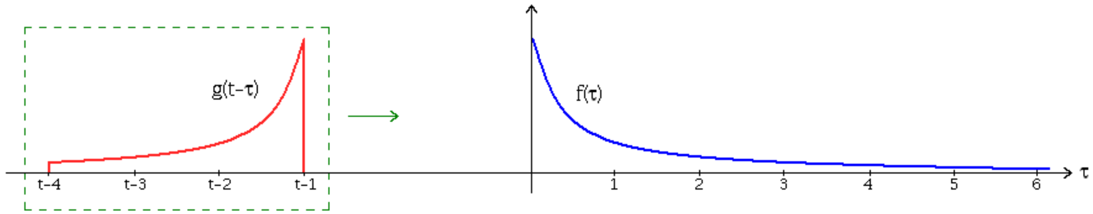
1. Express each function in terms of a dummy variable τ.



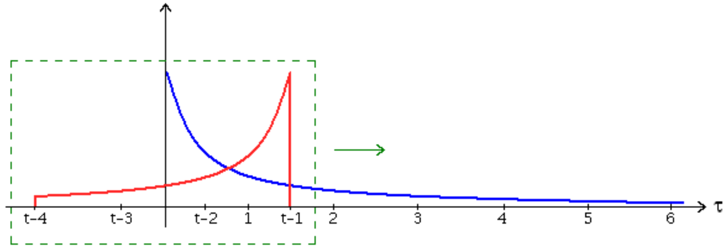
2. Reflect one of the functions: $g(\tau) \rightarrow g(-\tau)$.



3. Add a time-offset, t, which allows $g(t-\tau)$ to slide along the τ-axis.



4. Start t at -∞ and slide it all the way to +∞. Wherever the two functions intersect, find the integral of their product. In other words, compute a sliding, weighted-average of function $f(\tau)$, where the weighting function is $g(-\tau)$.



The resulting waveform (not shown here) is the convolution of functions f and g. If f(t) is a unit impulse, the result of this process is simply g(t), which is therefore called the impulse response.