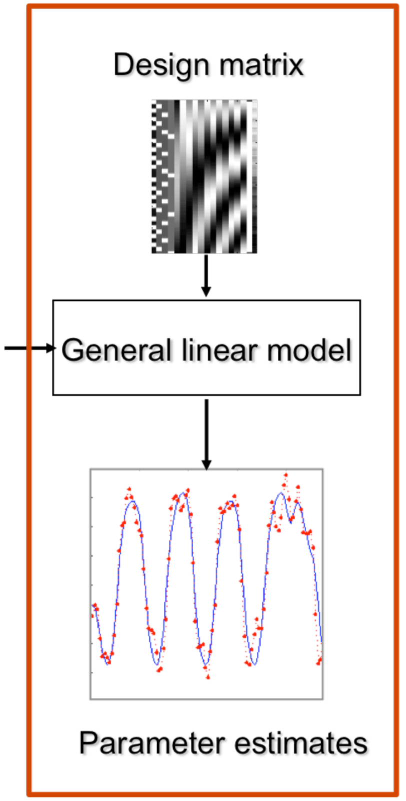
The General Linear Model (GLM)

Dr. Frederike Petzschner

petzschner@biomed.ee.ethz.ch

Translational Neuromodeling Unit (TNU)
Institute for Biomedical Engineering, University of Zurich & ETH Zurich



With many thanks for slides & images to:

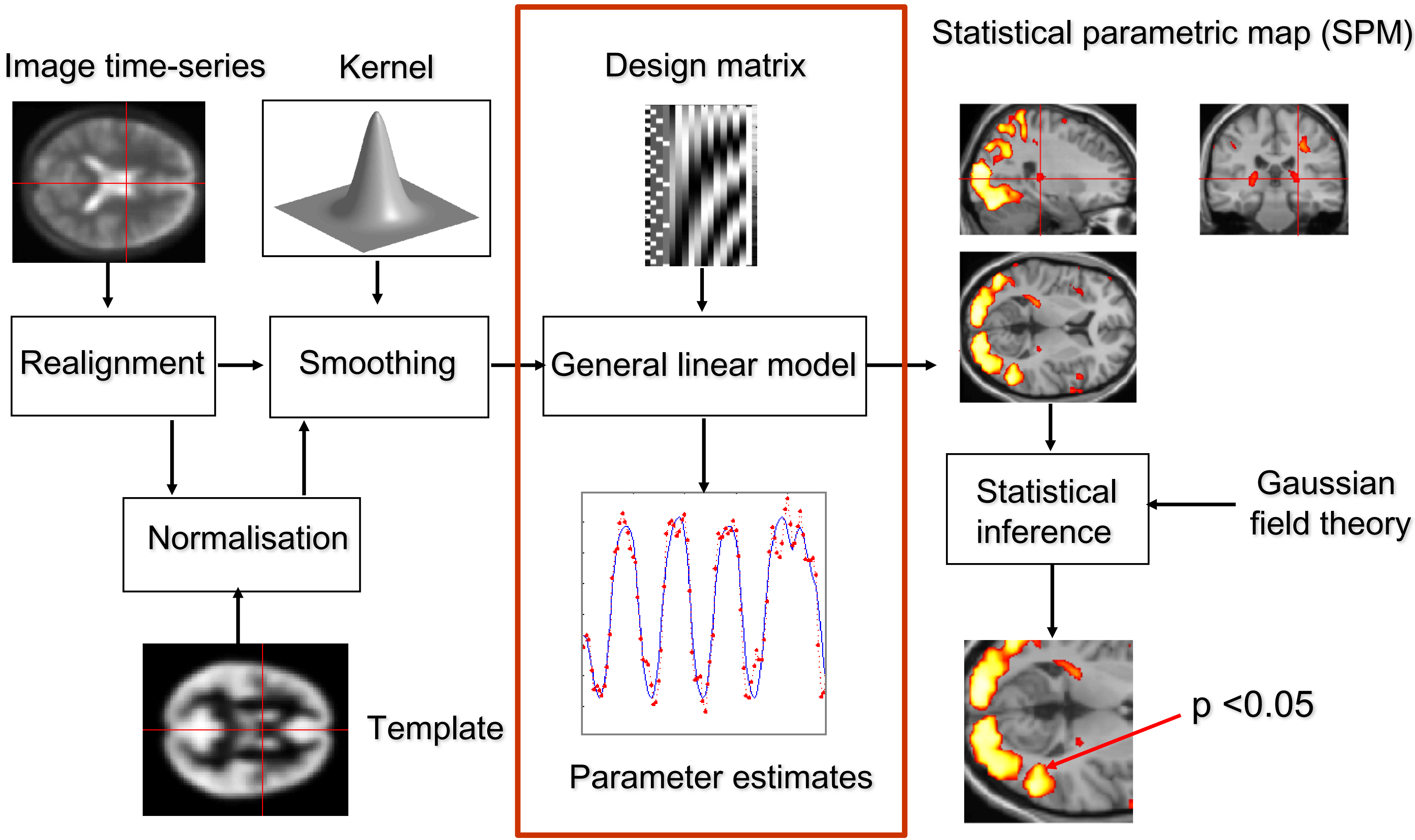
FIL Methods group, Virginia Flanagin and Klaas Enno Stephan

Translational Neuromodeling Unit

University of Zurich[UZH]

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

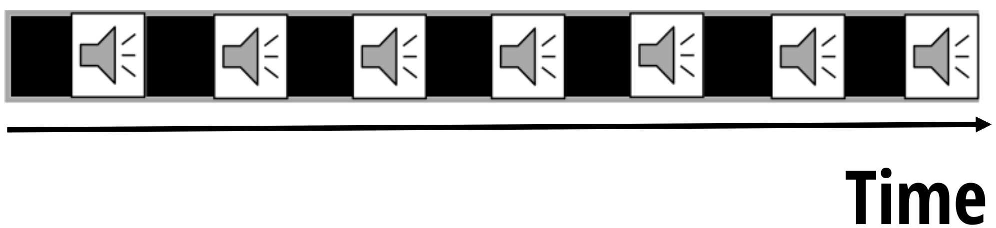
Overview of SPM

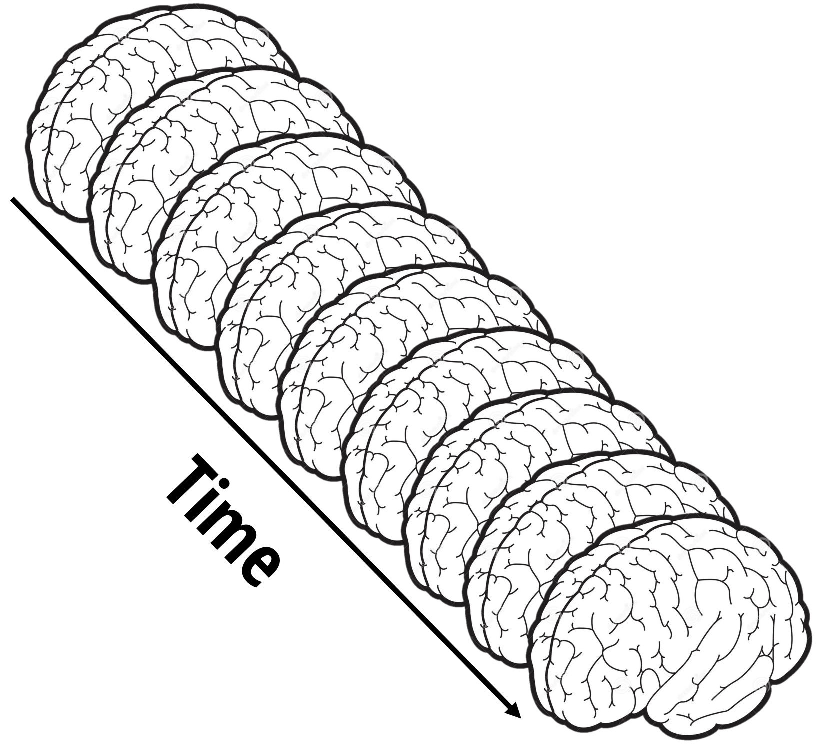


Research Question:

Where in the brain do we represent listening to sounds?

Image a very simple experiment…





SINGLE VOXEL TIME SERIES…

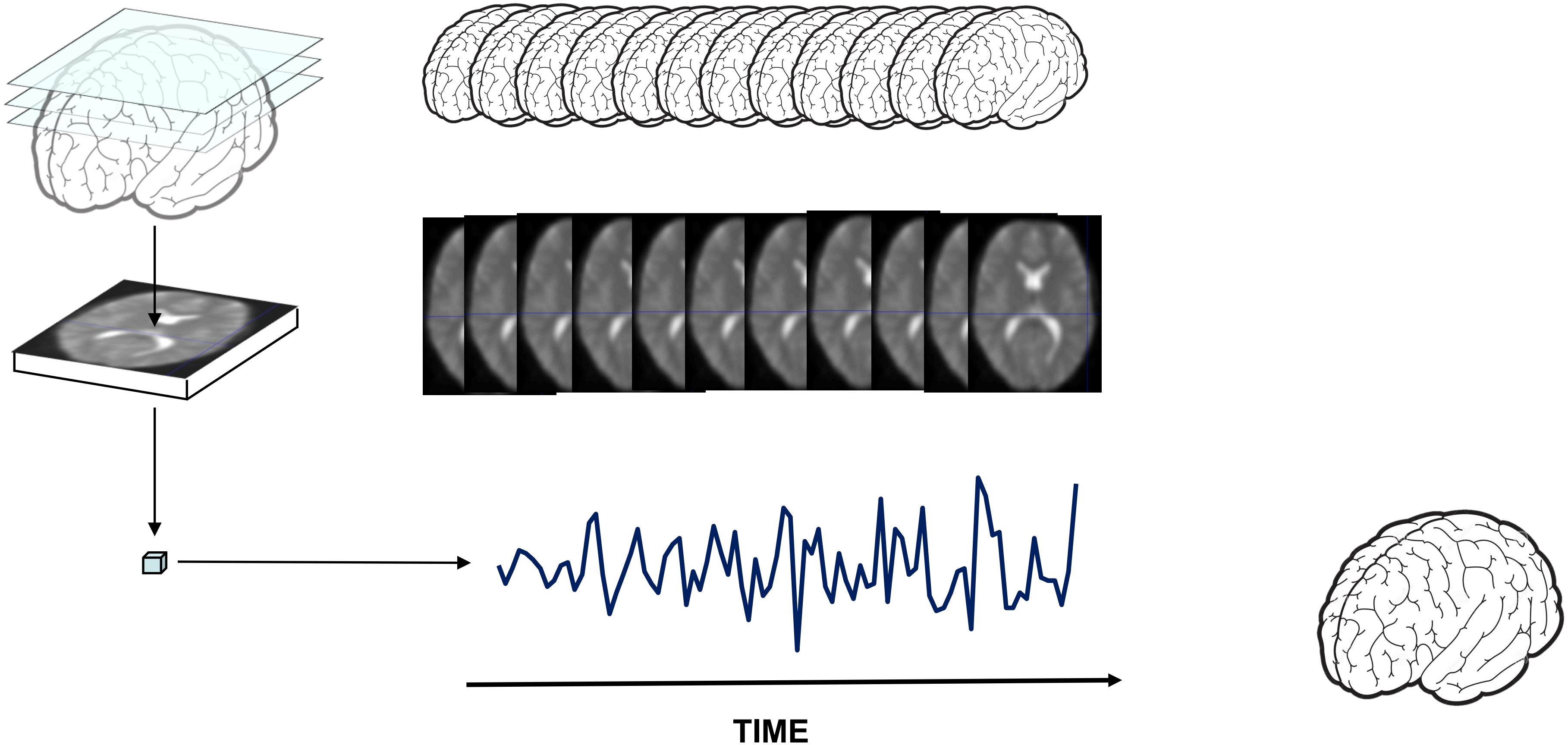


Image a very simple experiment…

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

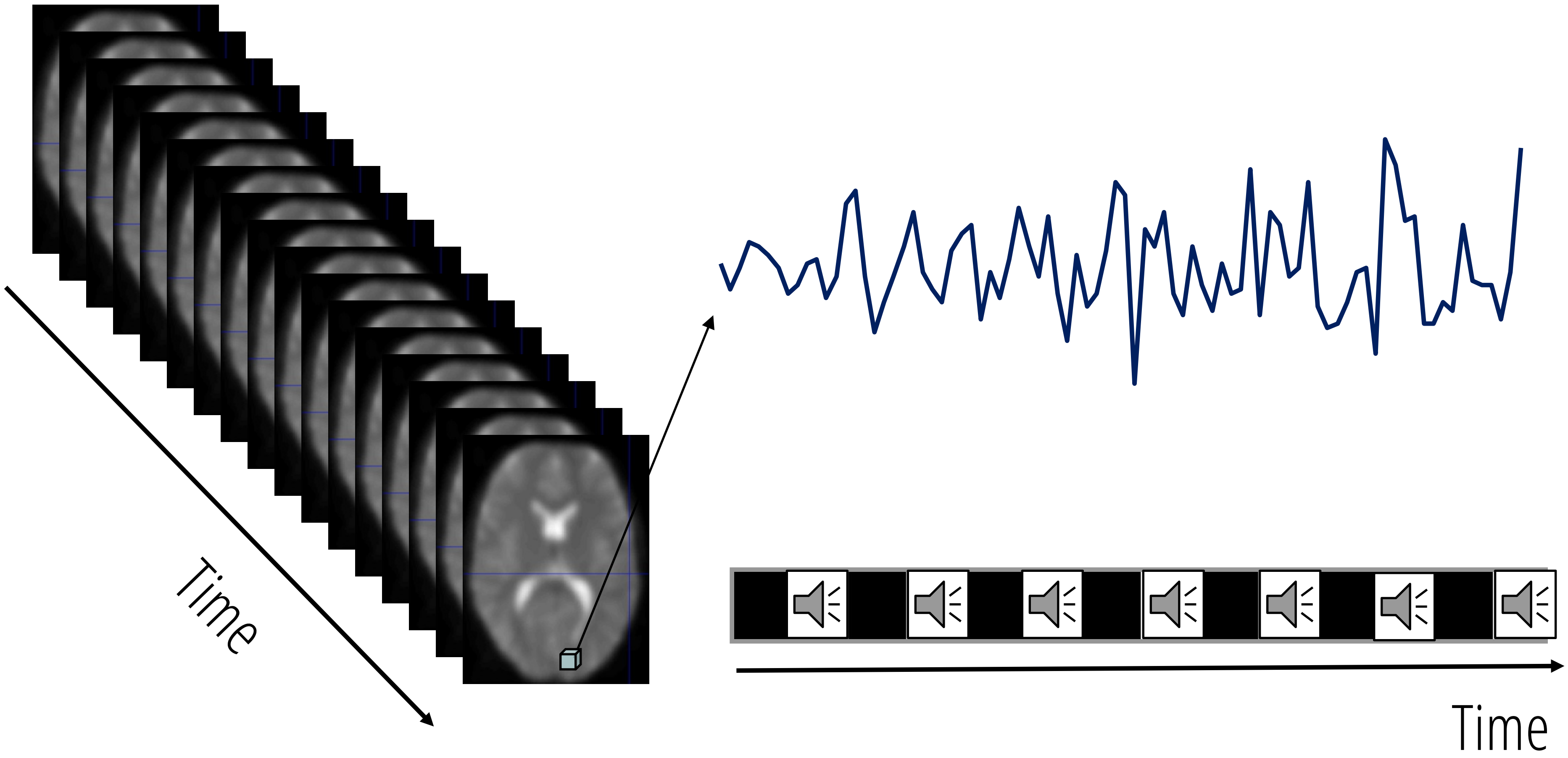
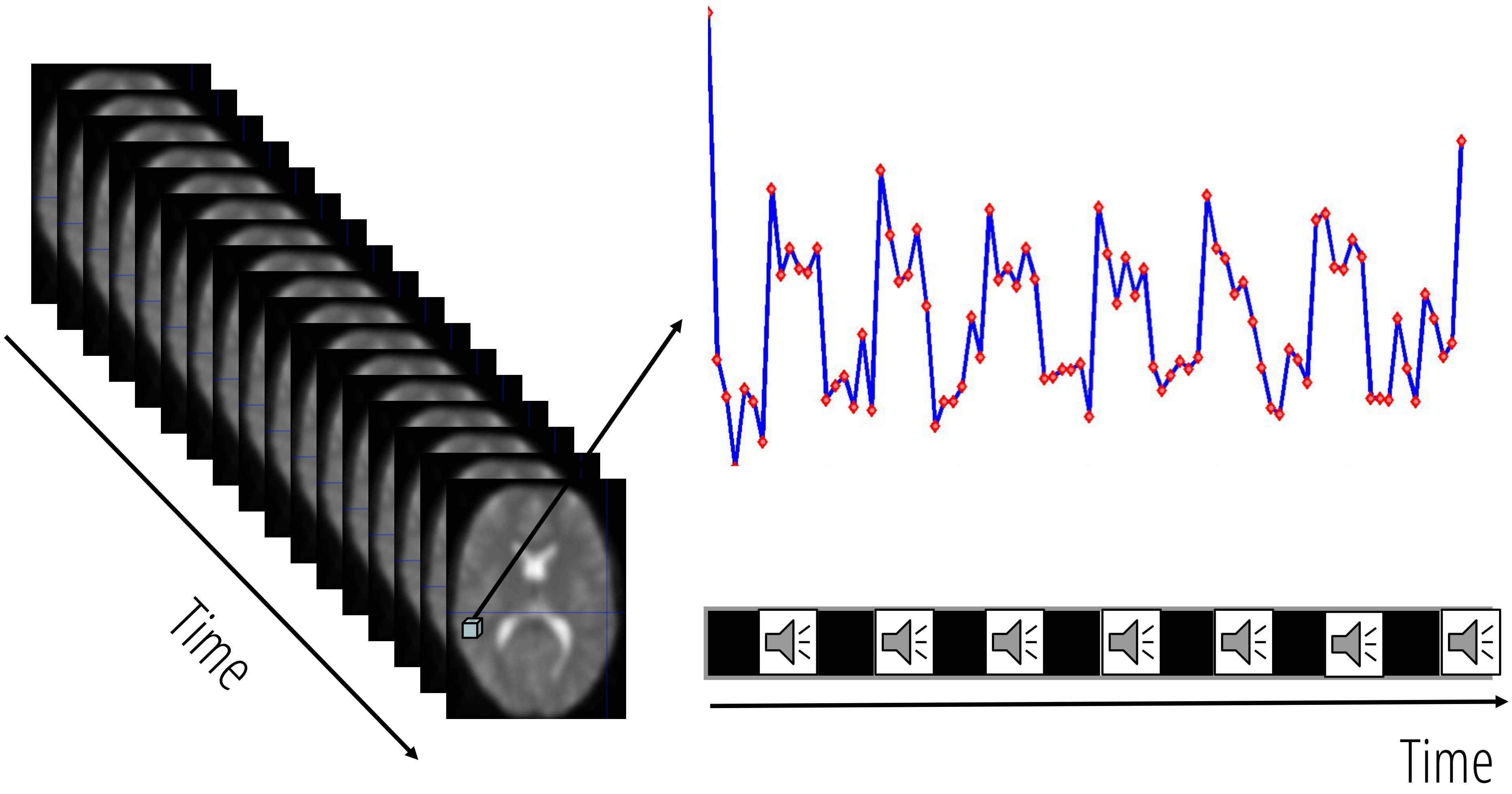
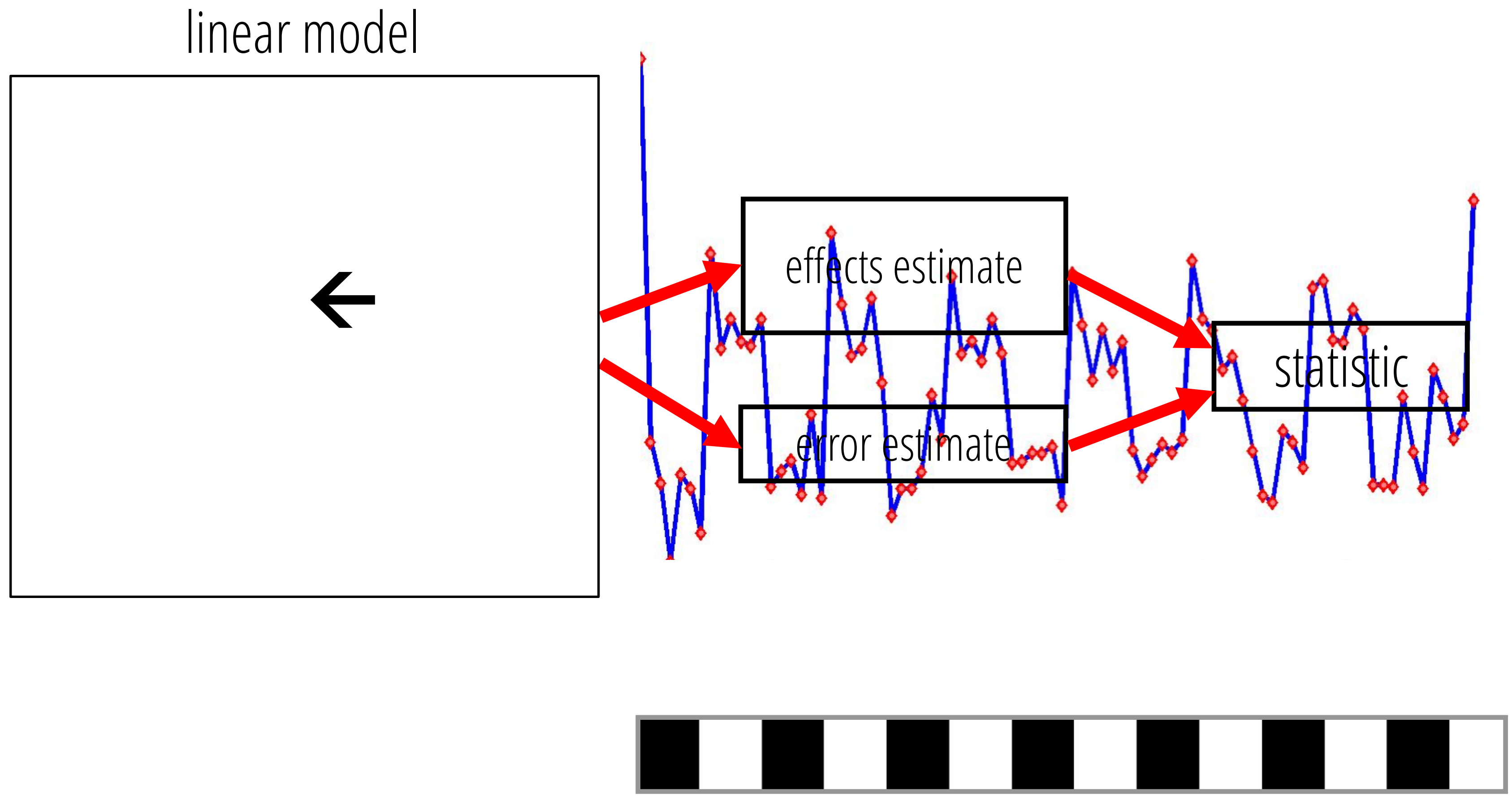


Image a very simple experiment…

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

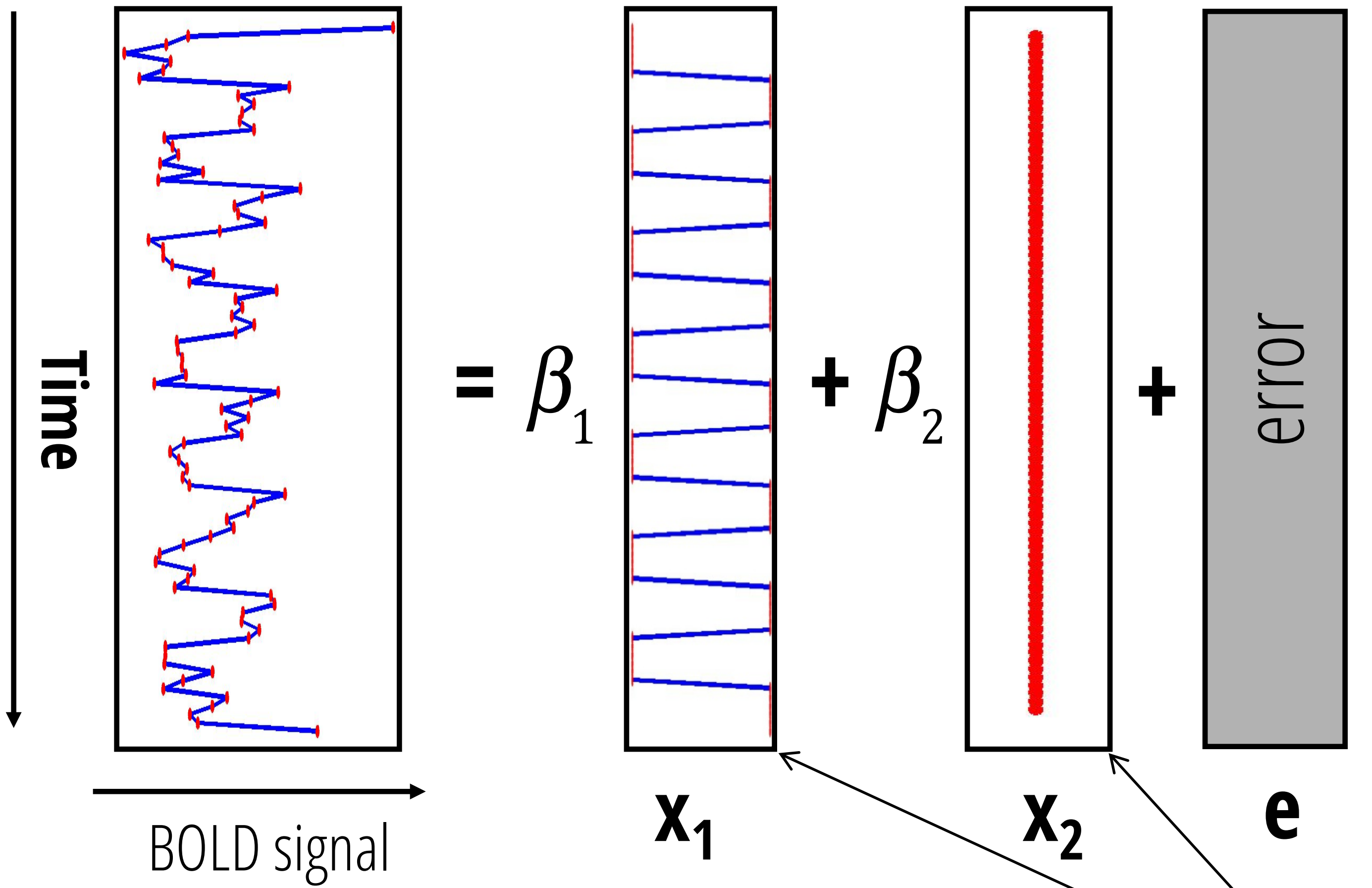


You need a model of your data...



Explain your data...

as a combination of experimental manipulation, confounds and errors

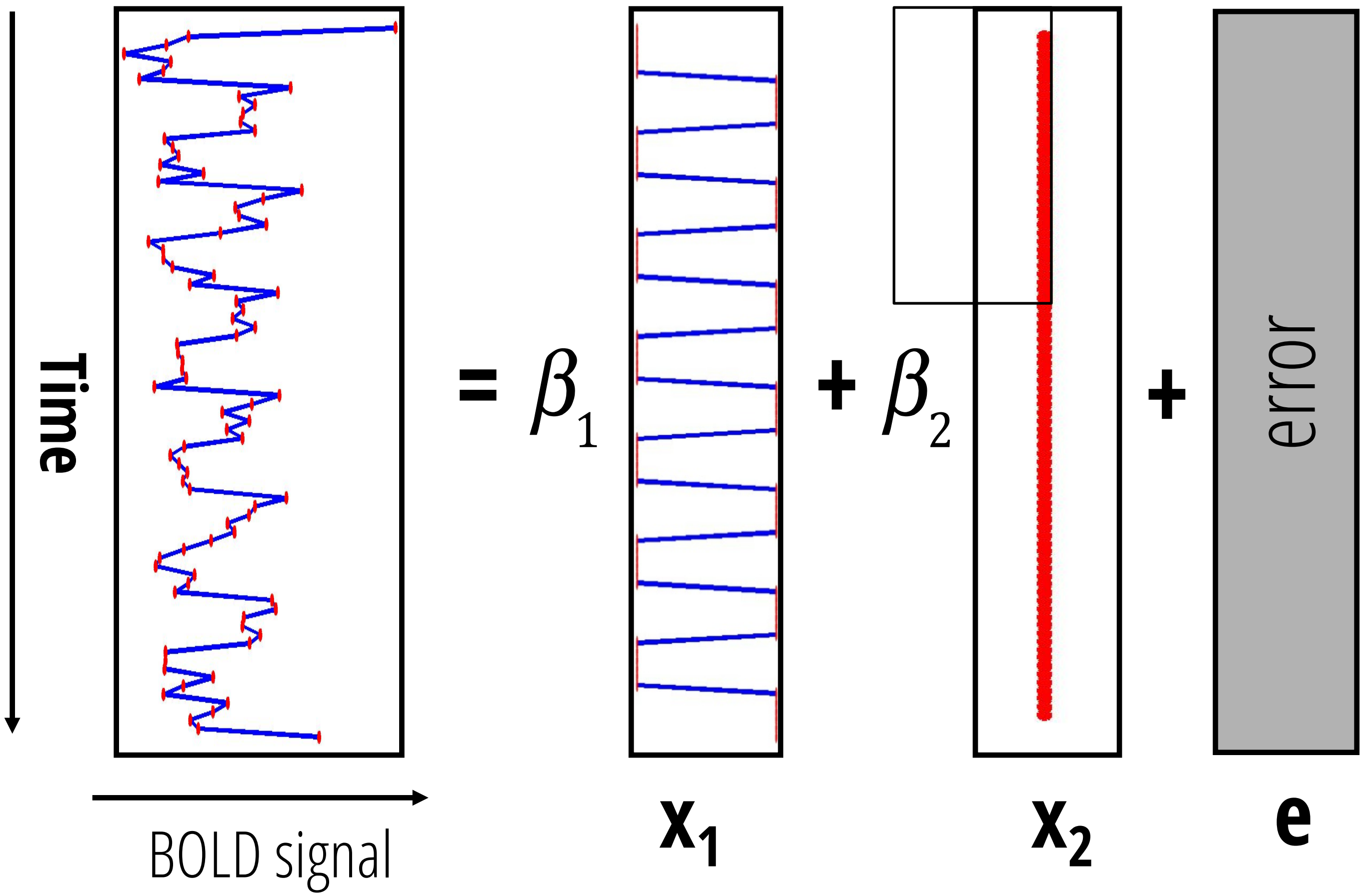


Single voxel regression model:

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

Explain your data...

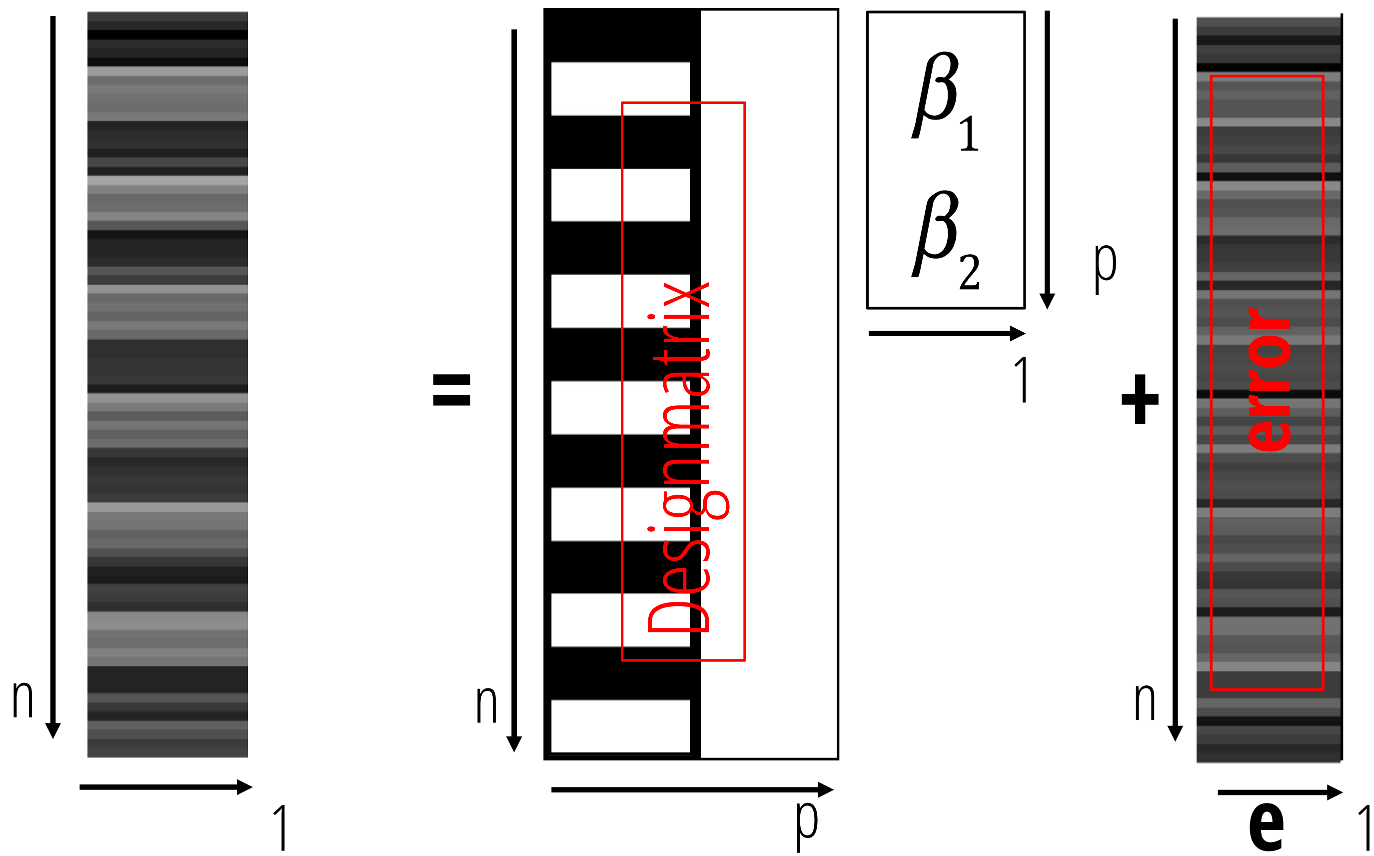
as a combination of experimental manipulation,confounds and errors



Single voxel regression model:

$$y : y = X\beta + e + e$$

The black and white version in SPM



$$y = X\beta + e$$

n: number of scans
p: number of regressors

GLM: mass-univariate parametric analysis

- 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…

Model assumptions

Designmatrix

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

error

You want to estimate your parameters such that you minimize:

$$\sum_{t=1}^{N} e_t^2$$

This can be done using an **Ordinary least squares** estimation (OLS) assuming an i.i.d. error

error

GLM assumes identical and independently distributed errors

i.i.d. = error covariance is a scalar multiple of the identity matrix $e \approx N(0, \sigma^2 I)$

$$Cov(e) = \begin{array}{cc} & \begin{array}{cc} t1 & t2 \end{array} \\ & \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \begin{array}{c} t1 \\ t2 \end{array} \end{array}$$

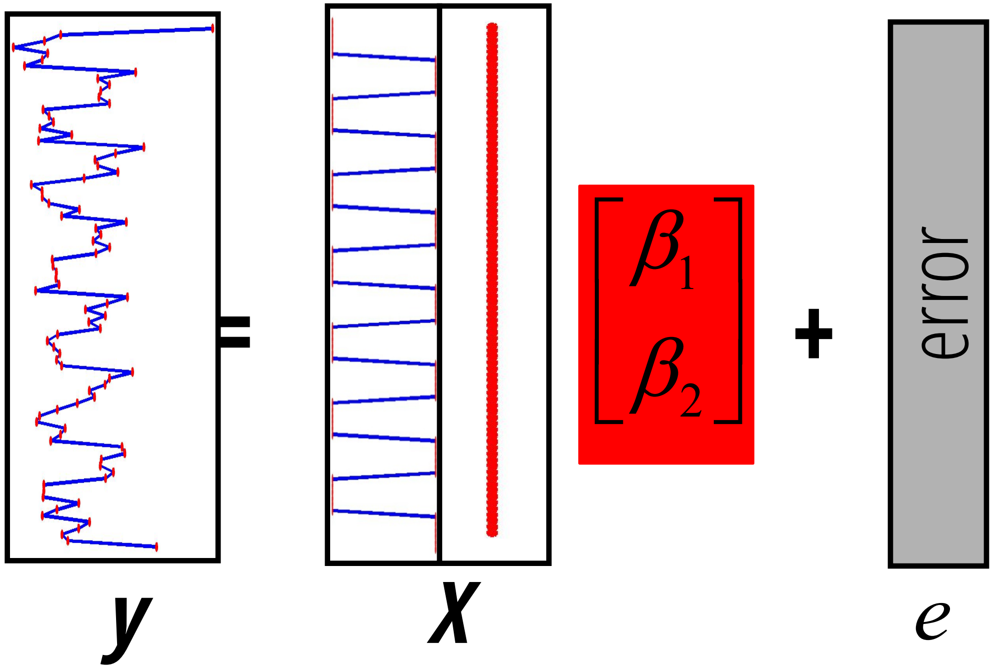
non-identity

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

non-independence

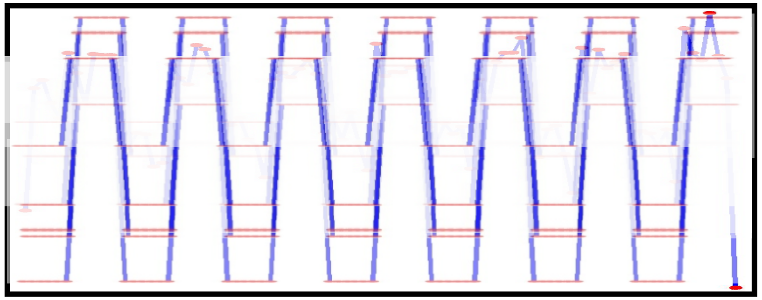
$$Cov(e) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

How to fit the model and estimate the parameters?

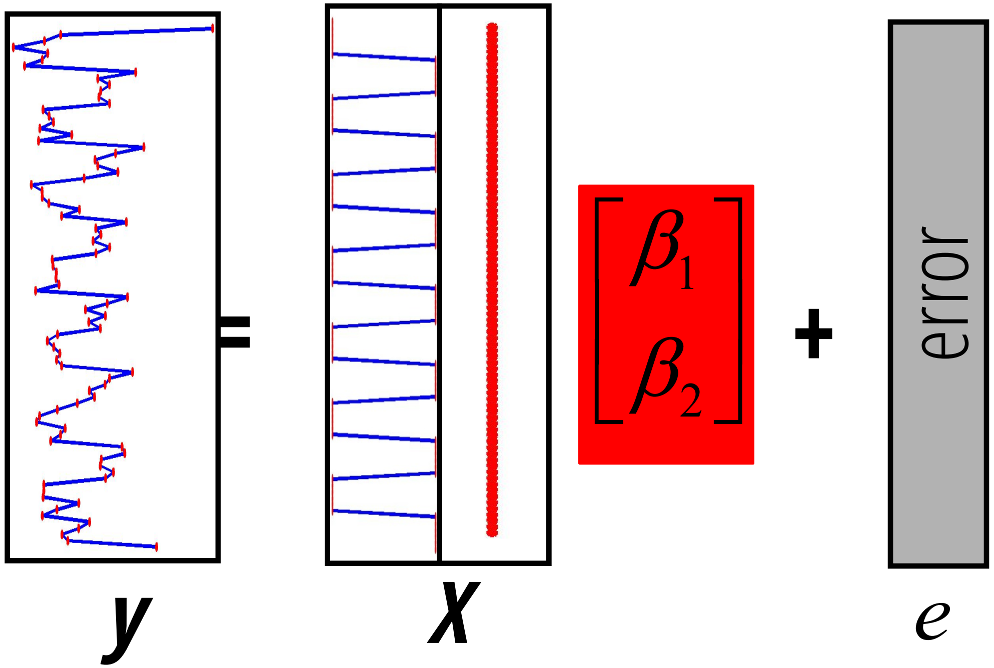


$$y = X \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + e$$

„Option 1": Per hand



How to fit the model and estimate the parameters?



$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

OLS (Ordinary Least Squares)

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

Data predicted by our model

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

Error between predicted and actual data

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

Goal is to determine the betas such that we minimize the quadratic error

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

OLS (Ordinary Least Squares)

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

The goal is to minimize the quadratic error between data and model

OLS (Ordinary Least Squares)

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

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

The goal is to minimize the quadratic error between data and model

OLS (Ordinary Least Squares)

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

The goal is to minimize the quadratic error between data and model

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

This is a scalar and the transpose of a scalar is a scalar ☺

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

OLS (Ordinary Least Squares)

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

The goal is to minimize the quadratic error between data and model

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

This is a scalar and the transpose of a scalar is a scalar ☺

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

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

OLS (Ordinary Least Squares)

The goal is to minimize the quadratic error between data and model

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

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

This is a scalar and the transpose of a scalar is a scalar ☺

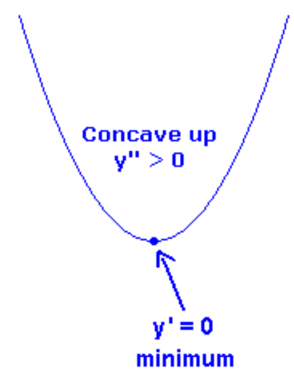
$$e^T e = y^T y \boxed{- y^T X\hat{\beta}} - \hat{\beta}^T X^T y + \hat{\beta}^T X^T X\hat{\beta}$$

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

You find the extremum of a function by taking its derivative and setting it to zero

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

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



OLS (Ordinary Least Squares)

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

The goal is to minimize the quadratic error between data and model

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

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

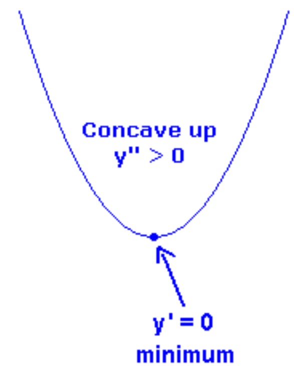
This is a scalar and the transpose of a scalar is a scalar ☺

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

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

You find the extremum of a function by taking its derivative and setting it to zero

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



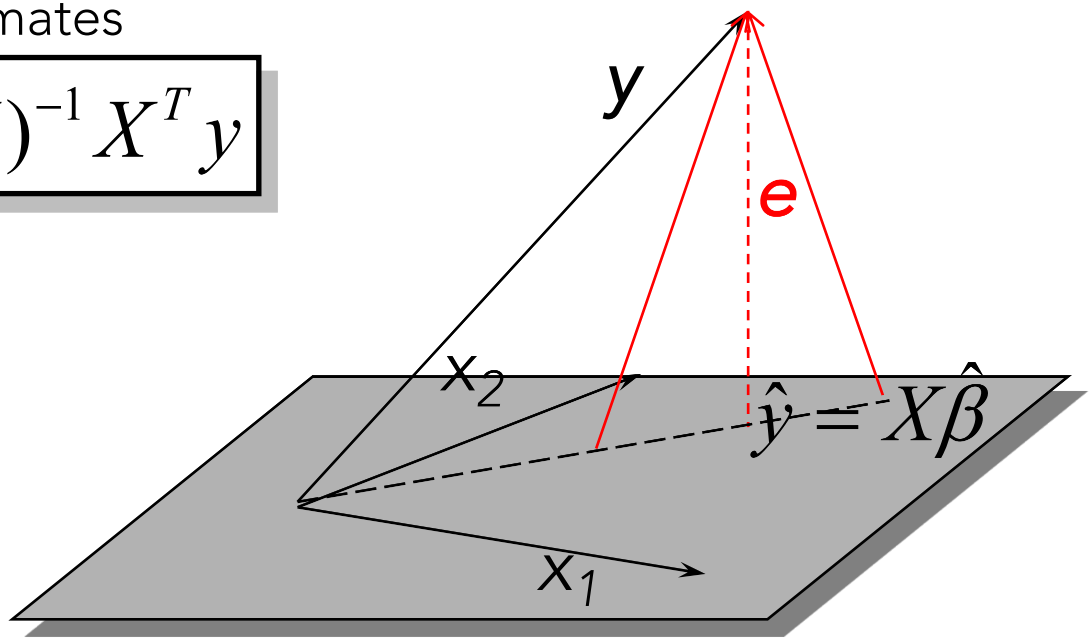
$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T y}$$

SOLUTION: OLS of the Parameters

A geometric perspective on the GLM

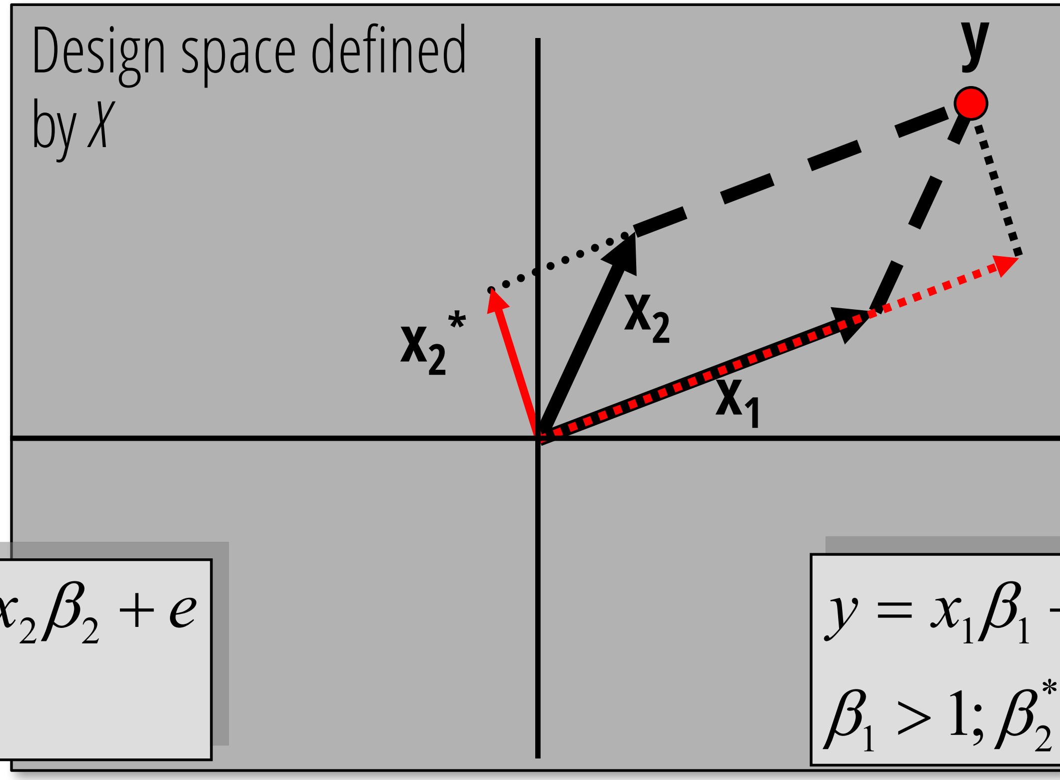
OLS estimates

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



Design space defined by *X*

Correlated and orthogonal regressors



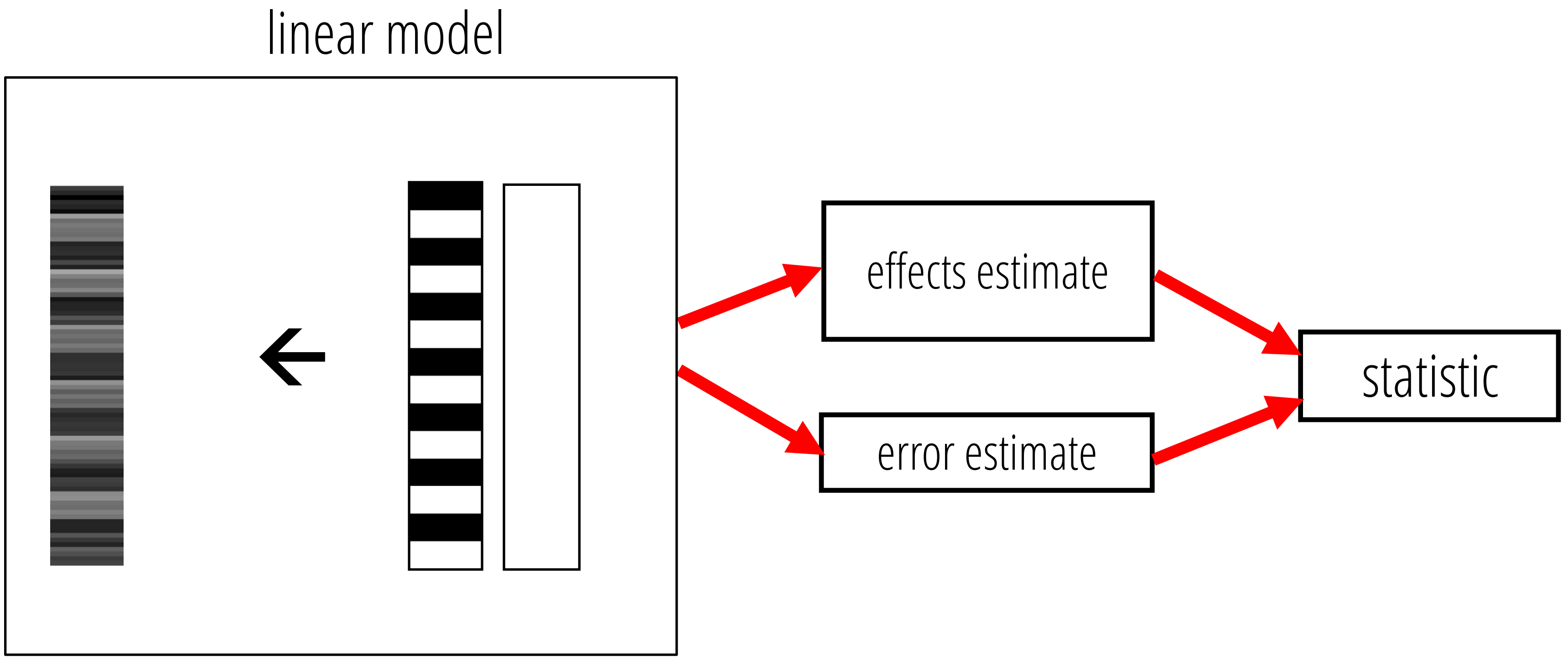
$$y = x_1\beta_1 + x_2\beta_2 + e$$
$$\beta_1 = \beta_2 = 1$$

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

Correlated regressors = explained variance is shared between regressors

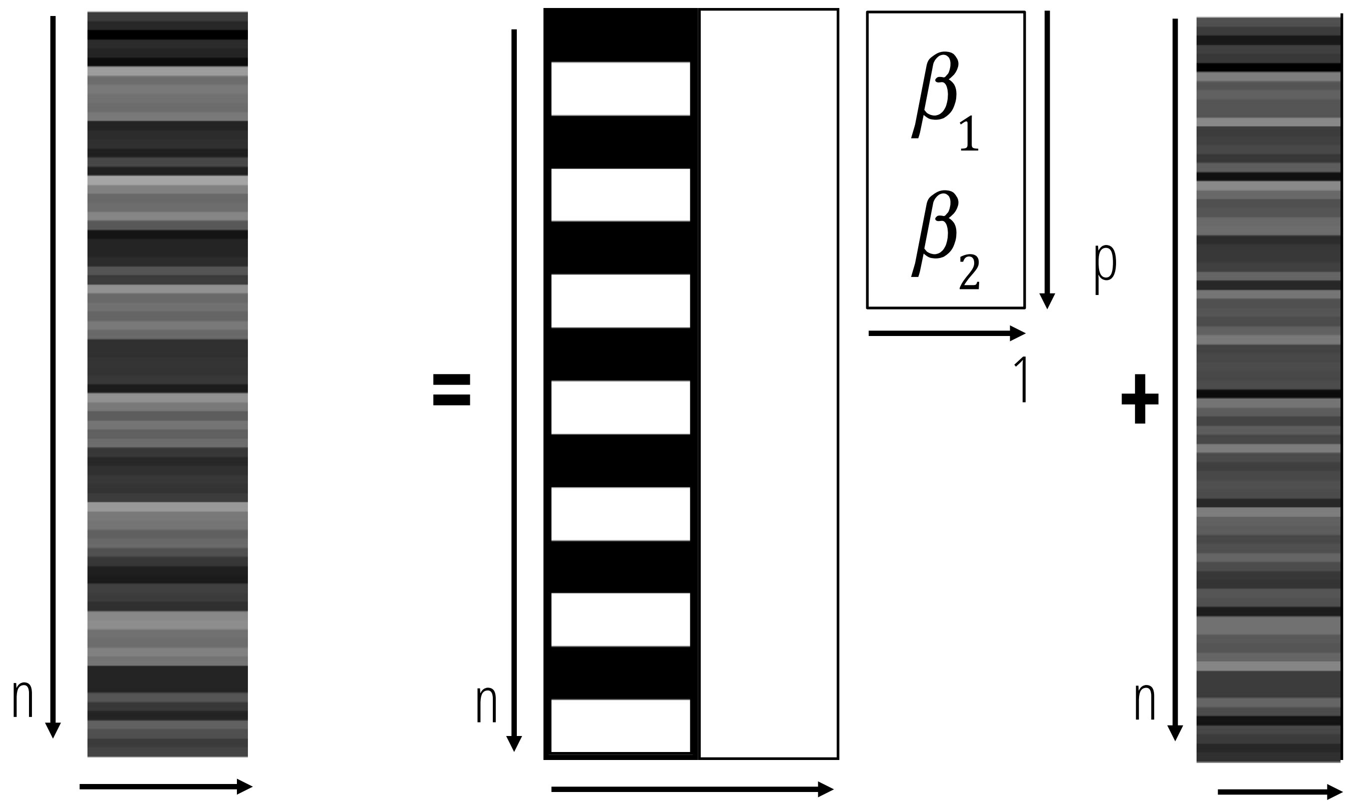
When x_2 is orthogonalized with regard to x_1, only the parameter estimate for x_1 changes, not that for x_2!

We are nearly there...

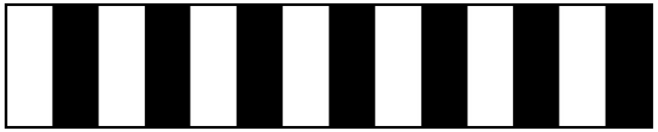


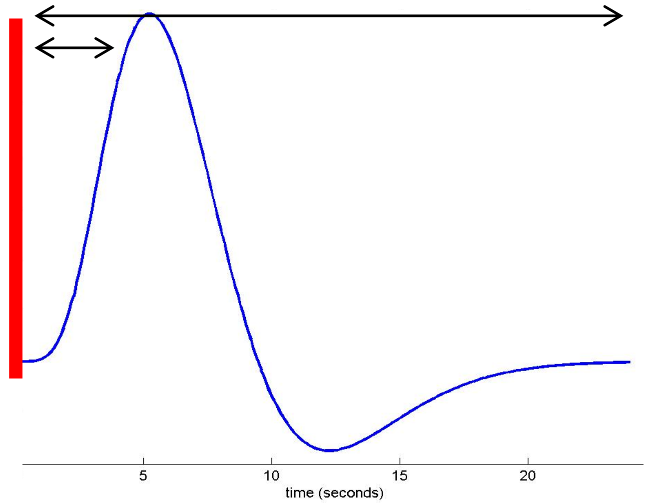
...but we are dealing with fMRI data

What are the problems?



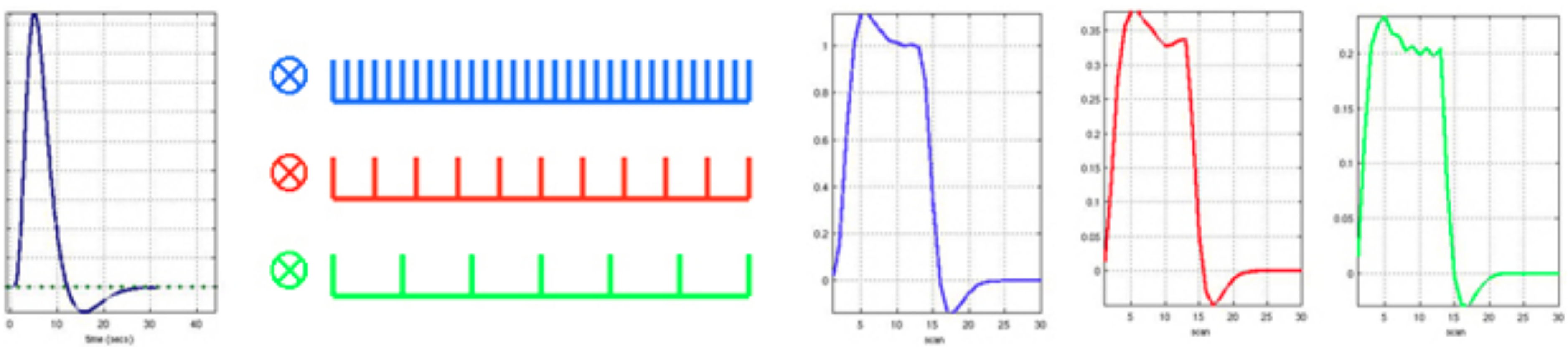
Problem 1: Shape of BOLD response





$$f \otimes g(t) = \int_{0} f(\tau)g(t-\tau)d\tau$$

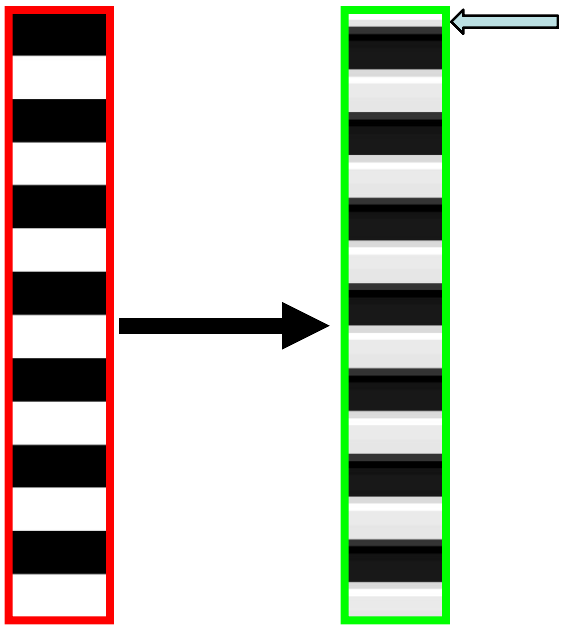
The response of a linear time-invariant (LTI) system is the convolution of the input with the system's response to an impulse (delta function).

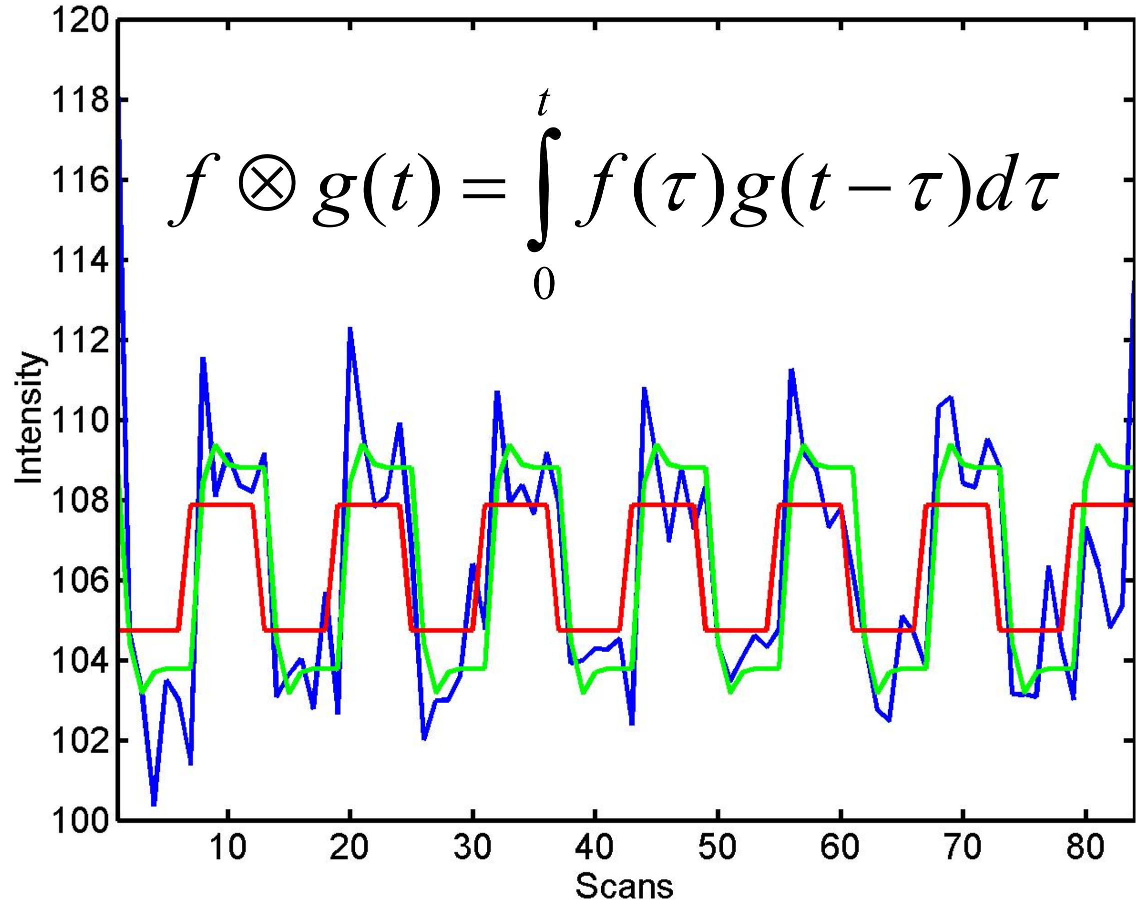


Solution: Convolution model of the BOLD response

expected BOLD response

= input function x impulse response function (HRF)



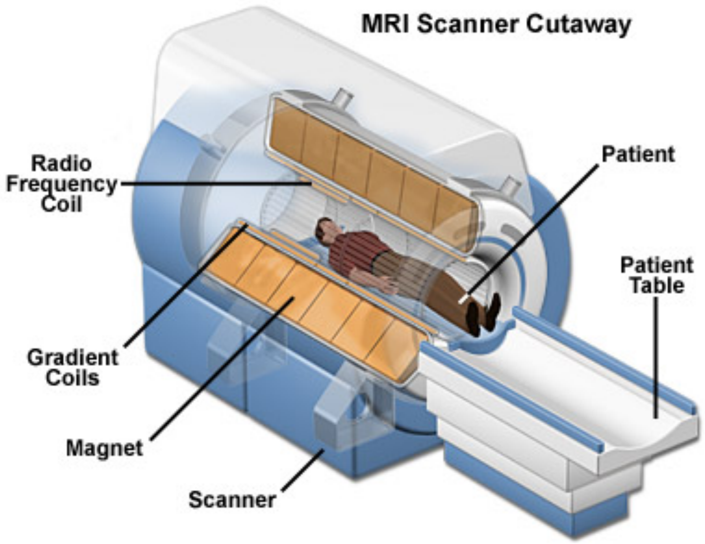


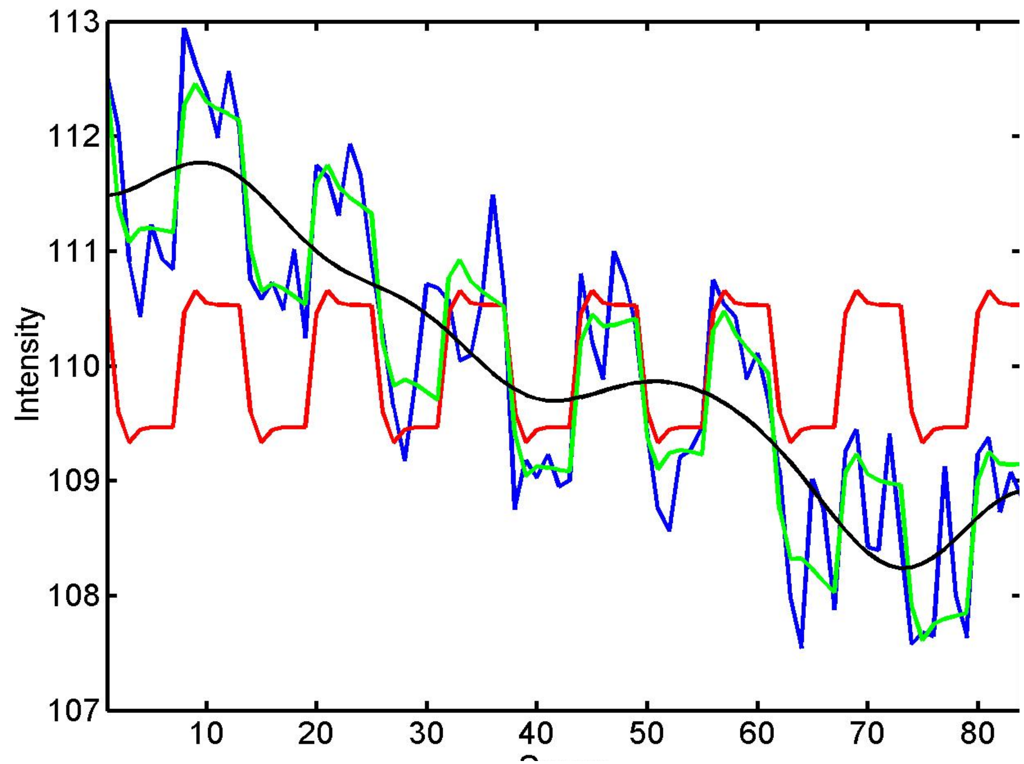
blue = data

green = predicted response, taking convolved with HRF

red = predicted response, NOT taking into account the HRF

Problem 2: Low frequency noise





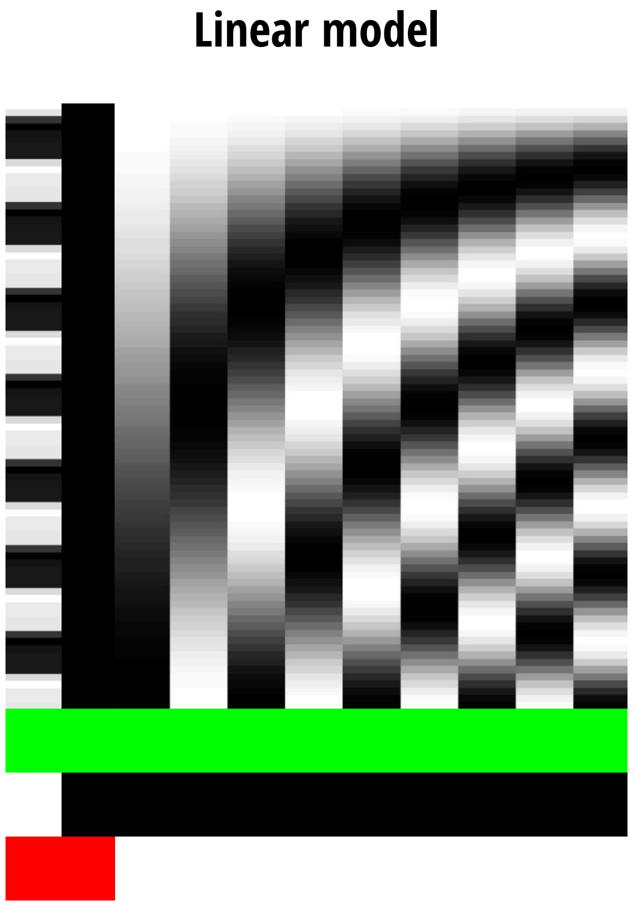
blue = data

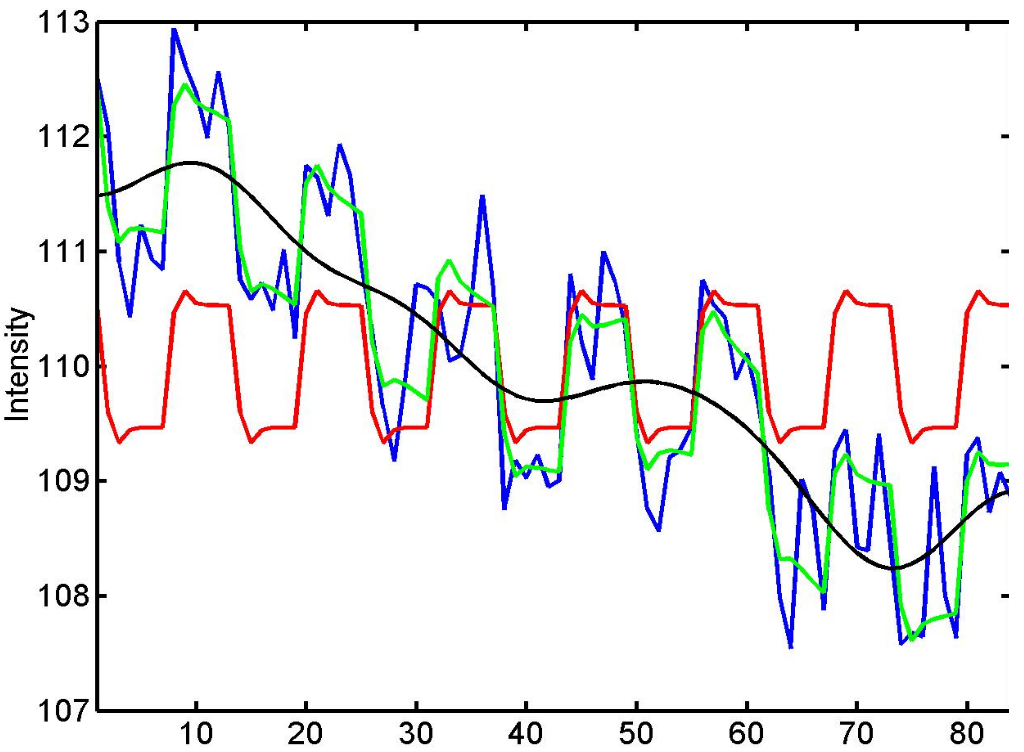
black = mean + low-frequency drift

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

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

Problem 2: Low frequency noise





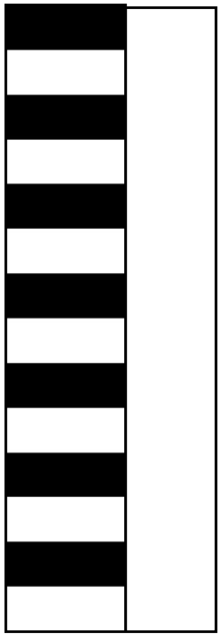
blue = data

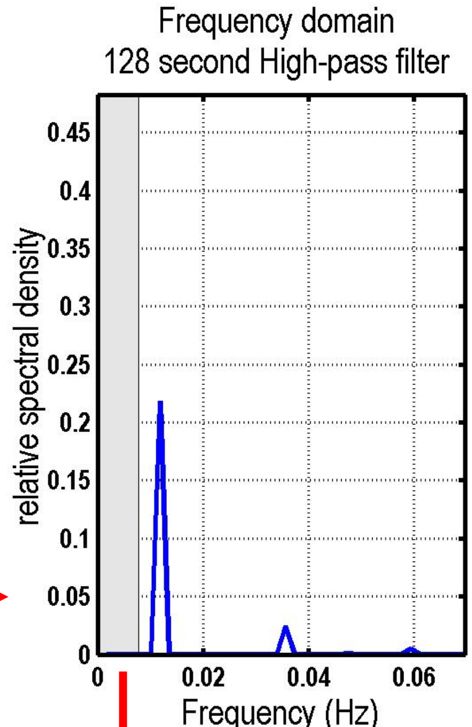
black = mean + low-frequency drift

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

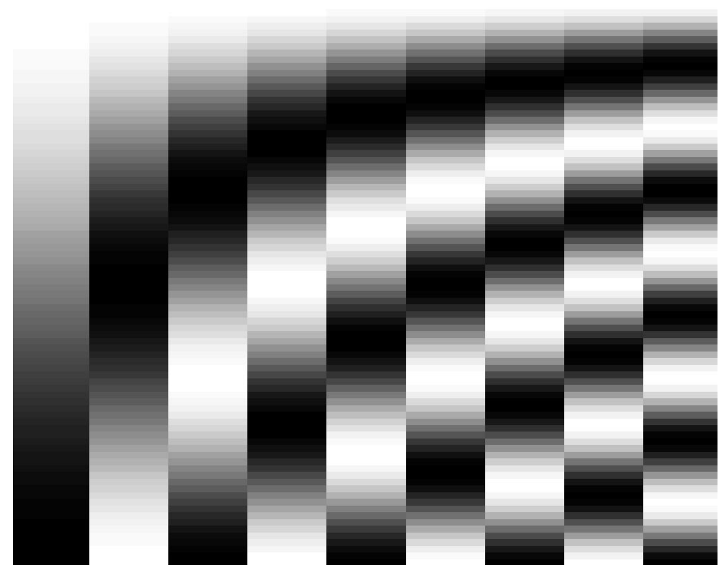
red = predicted response, NOT taking into account low-frequency drift

Solution 2: High pass filtering





discrete cosine transform (DCT) set



Problem 3: Serial correlations

i.i.d

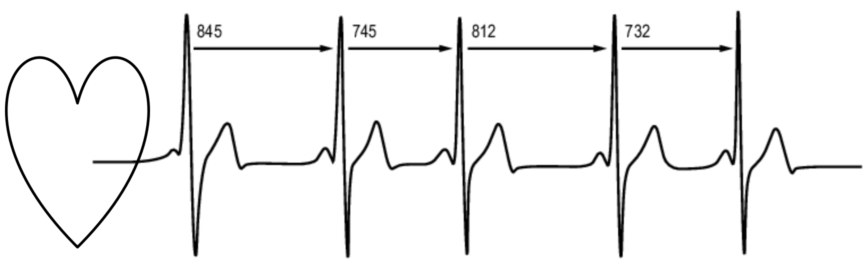
$$Cov(e) = \begin{array}{c c} & \begin{array}{cc} t1 & t2 \end{array} \\ & \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \begin{array}{c} t1 \\ t2 \end{array} \end{array}$$

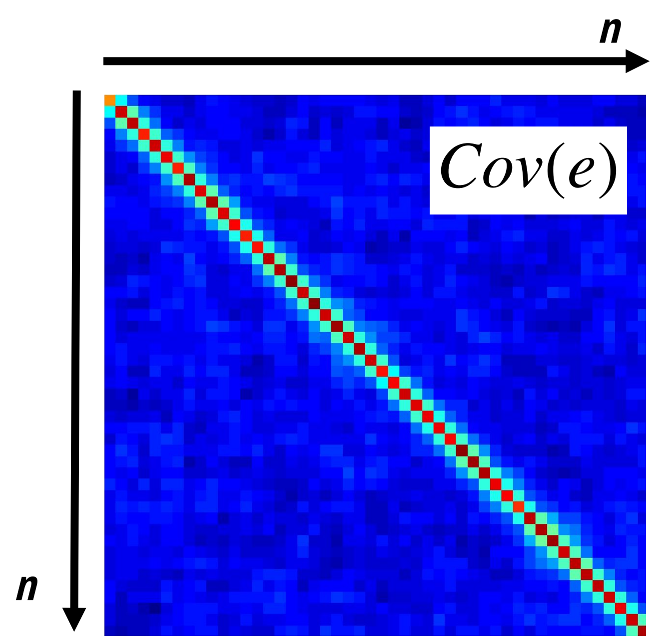
non-identity

$$Cov(e) = \left[\begin{array}{cc} 4 & 0 \\ 0 & 1 \end{array}\right]$$

non-independence

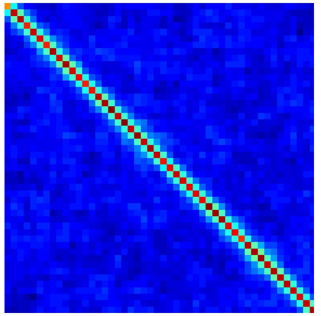
$$Cov(e) = \left[\begin{array}{cc} 2 & 1 \\ 1 & 2 \end{array}\right]$$





n: number of scans

Problem 3: Serial correlations



- Transform the signal into a space where the error is iid

$$\boxed{Wy = WX\beta + \underbrace{We}_{\text{This is i.i.d}}}$$

- *Pre-whitening*:

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

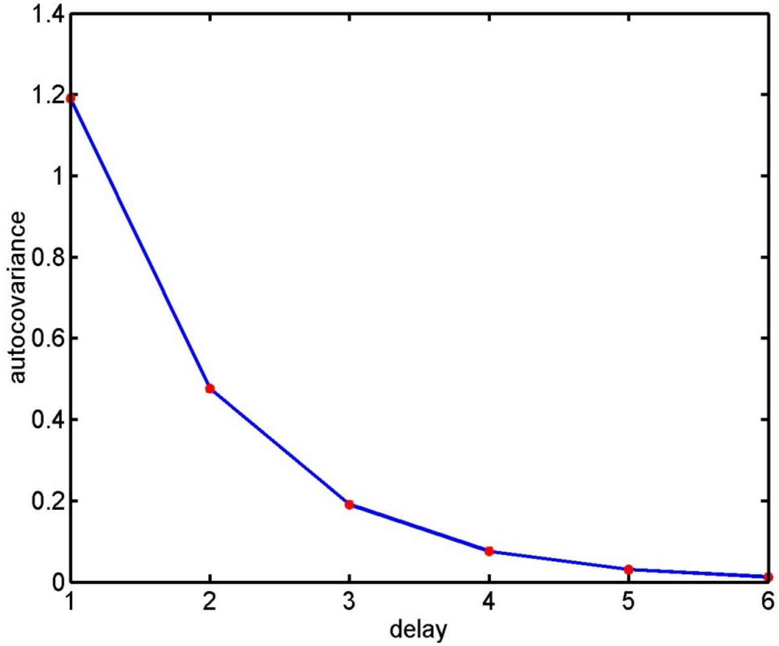
  2. Use estimated serial correlation to specify filter matrix W for whitening the data.

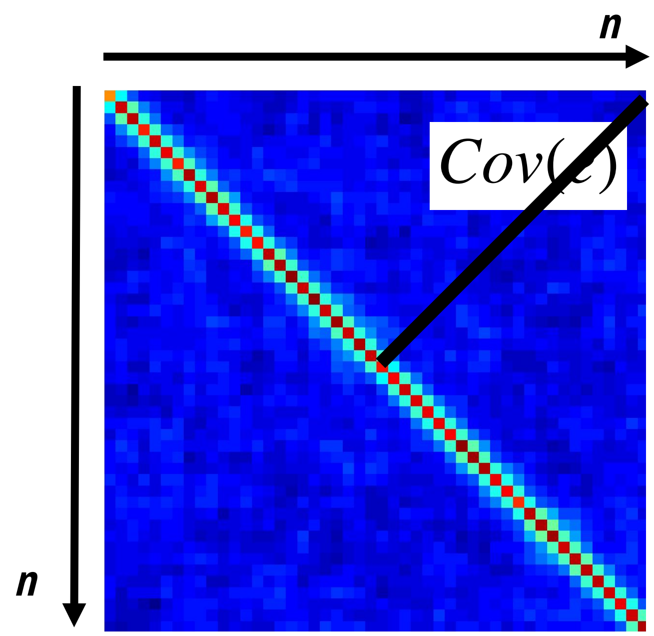
Problem 3: How to find W → Model the noise

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

1st order autoregressive process: AR(1)

autocovariance
function





n: number of scans

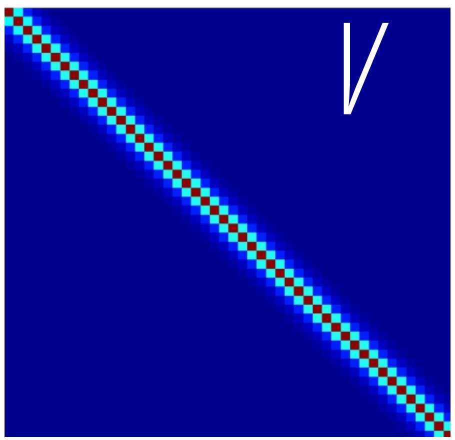
Model the noise: Multiple covariance components

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

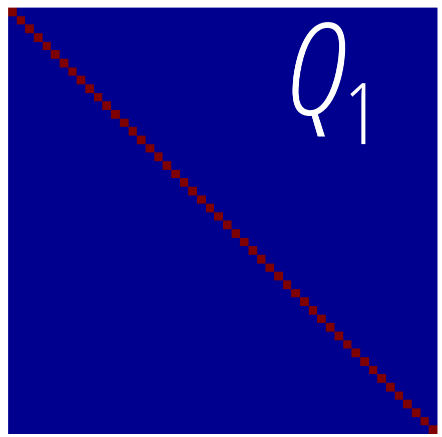
enhanced noise model

$$V \propto Cov(e)$$
$$V = \sum \lambda_i Q_i$$

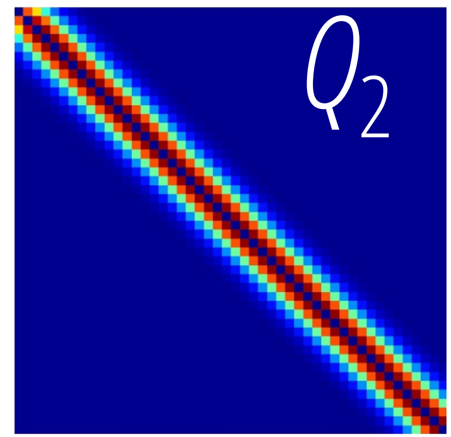
error covariance components *Q*
and hyperparameters



$$= \quad \lambda_1$$



$$+\lambda_2$$



Estimation of hyperparameters with EM (expectation maximisation) or ReML (restricted maximum likelihood).

How do we 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***

$$Wy = WX\beta + We \rightarrow y_s = X_s\beta + e_s$$

We are there...

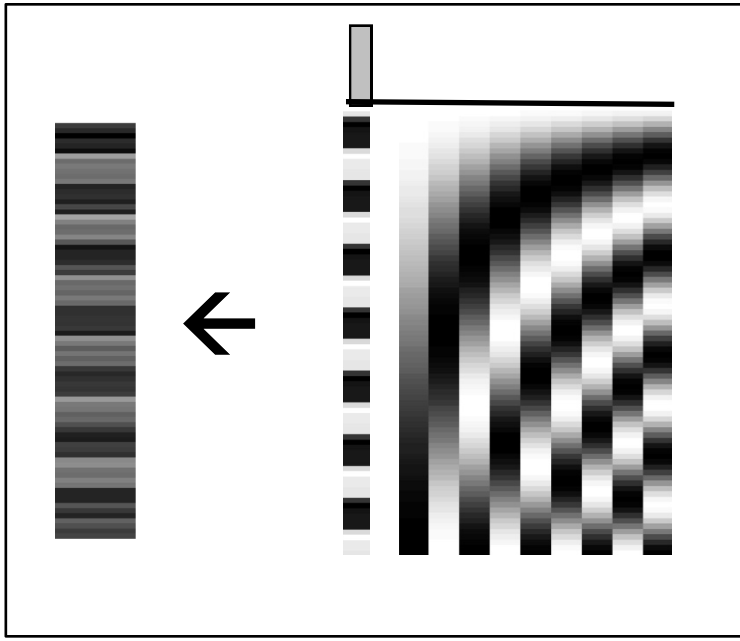
- the GLM models the effect of your experimental manipulation on the acquired data
- GLM includes all known experimental effects and confounds
- estimates effects an errors on a voxel-by-voxel basis

Because we are dealing with fMRI data there are a number of problems we need to take care of:

- Convolution with a canonical HRF
- High-pass filtering to account for low-frequency drifts
- Estimation of multiple variance components (e.g. to account for serial correlations)

We are there...

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



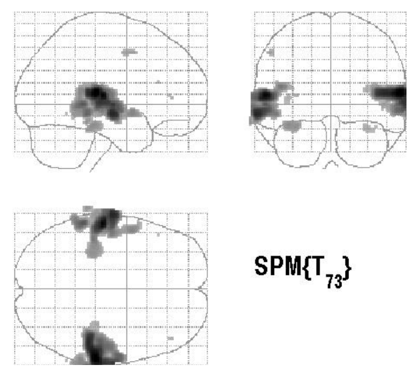
effects estimate

error estimate

Null hypothesis: $\beta_1 = 0$

statistic

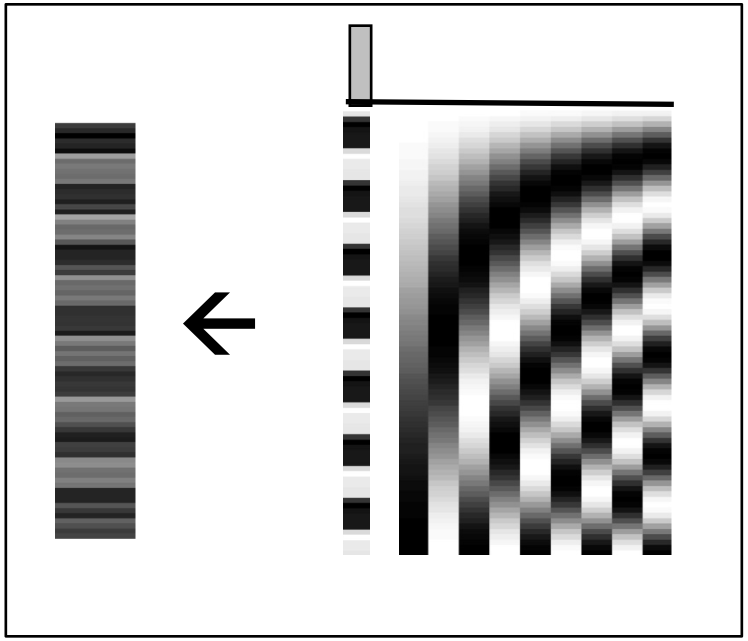
$$t = \frac{c^T \hat{\beta}}{Std(c^T \hat{\beta})}$$



→ Lecture: Classical (frequentist) inference

We are there...

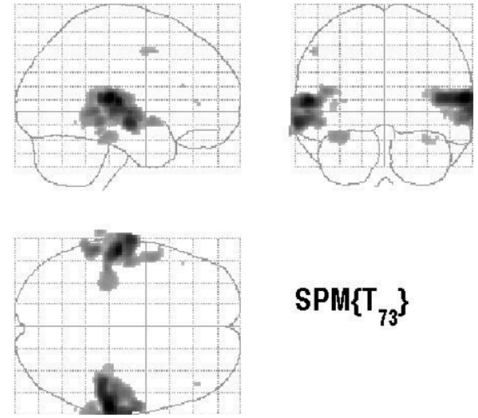
c = 1 0 0 0 0 0 0 0 0 0 0 0

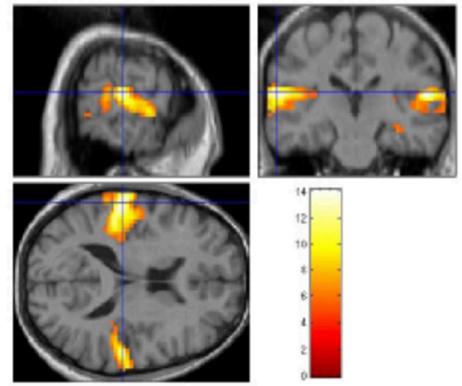


Q: activation during listening ?

Null hypothesis: $\beta_1 = 0$

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



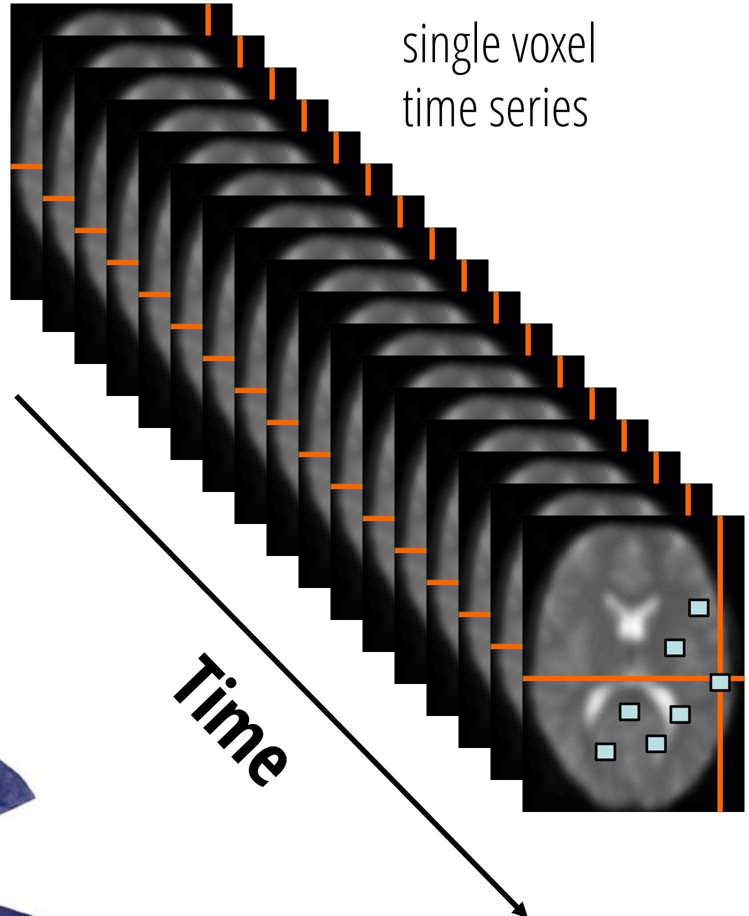


→ Lecture: Classical (frequentist) inference

So far we have looked at a single voxel...

- Mass-univariate approach:
  GLM applied to > 100,000 voxels

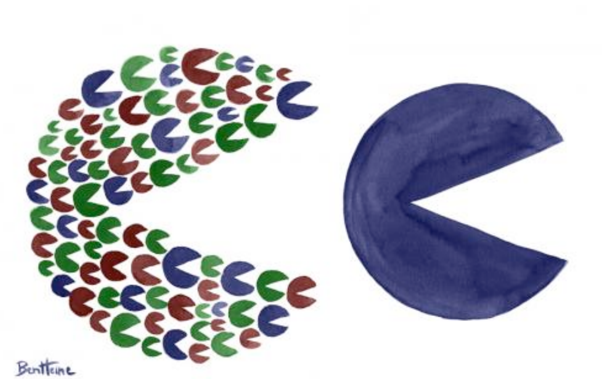
- Threshold of $p<0.05$ more than 5000 voxels significant by chance!



single voxel
time series

- Massive problem with multiple comparisons!

- Solution: Gaussian random field theory



How to build in physiological confounds?

- Head movements
- Aterial pulsations (particularly in the brain stem)
- Breathing
- Eye blinks
- Adaptation effects, fatigue, fluctuations in concentration etc.

- → Lecture: Noise models on fMRI and noise correction

Outlook: further challenges

- correction for multiple comparisons
- variability in the HRF across voxels
- slice timing
- limitations of frequentist statistics
    - → Bayesian analysis
- GLM ignores interactions among voxels

Thank you for listening!

- **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.