fMRI Modelling & Statistical Inference

Guillaume Flandin
Wellcome Trust Centre for Neuroimaging
University College London

SPM Course
Chicago, 22-23 Oct 2015

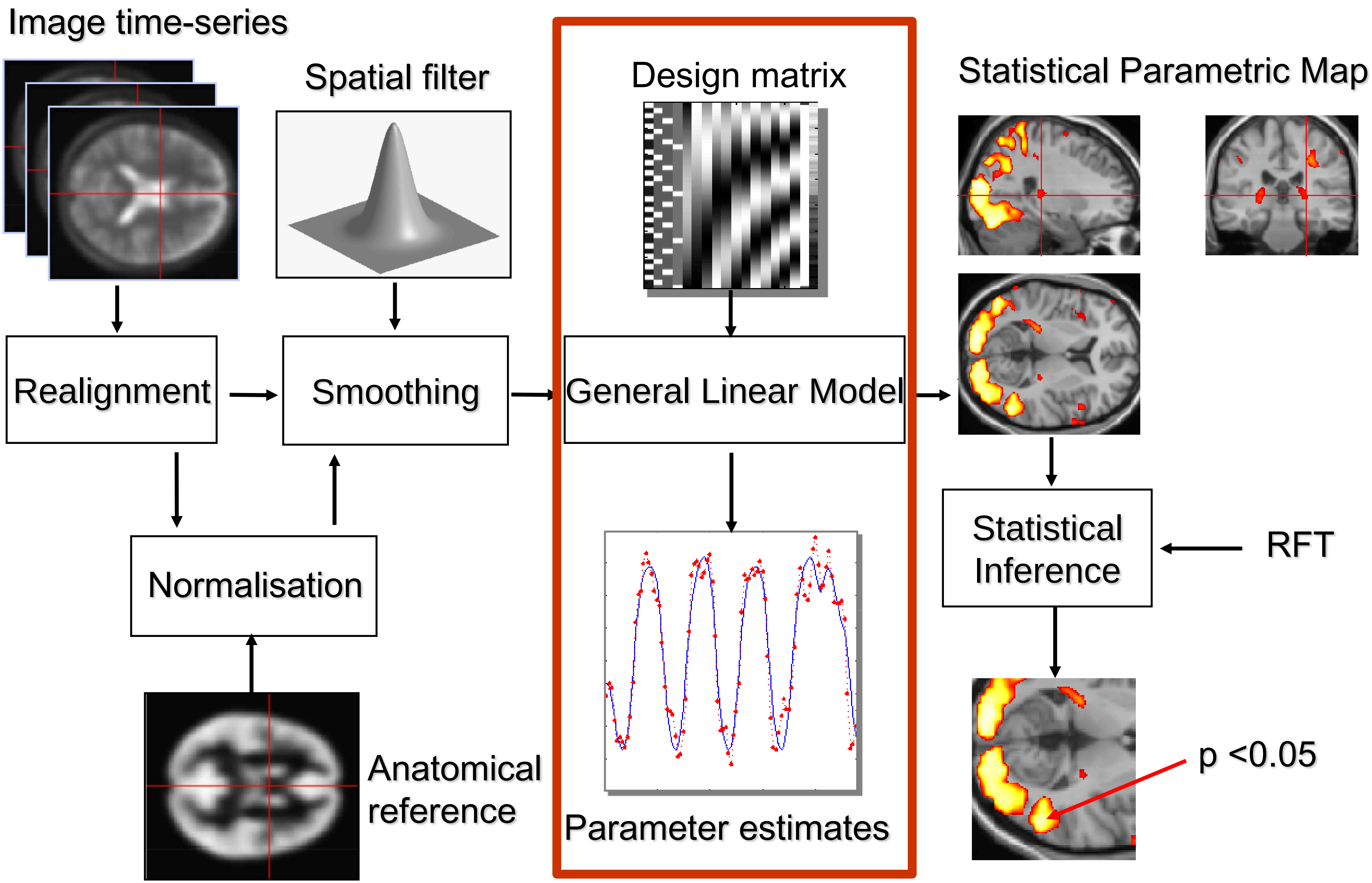


Image time-series

Spatial filter

Design matrix

Statistical Parametric Map

Realignment

Smoothing

General Linear Model

Statistical Inference

RFT

Normalisation

Anatomical reference

Parameter estimates

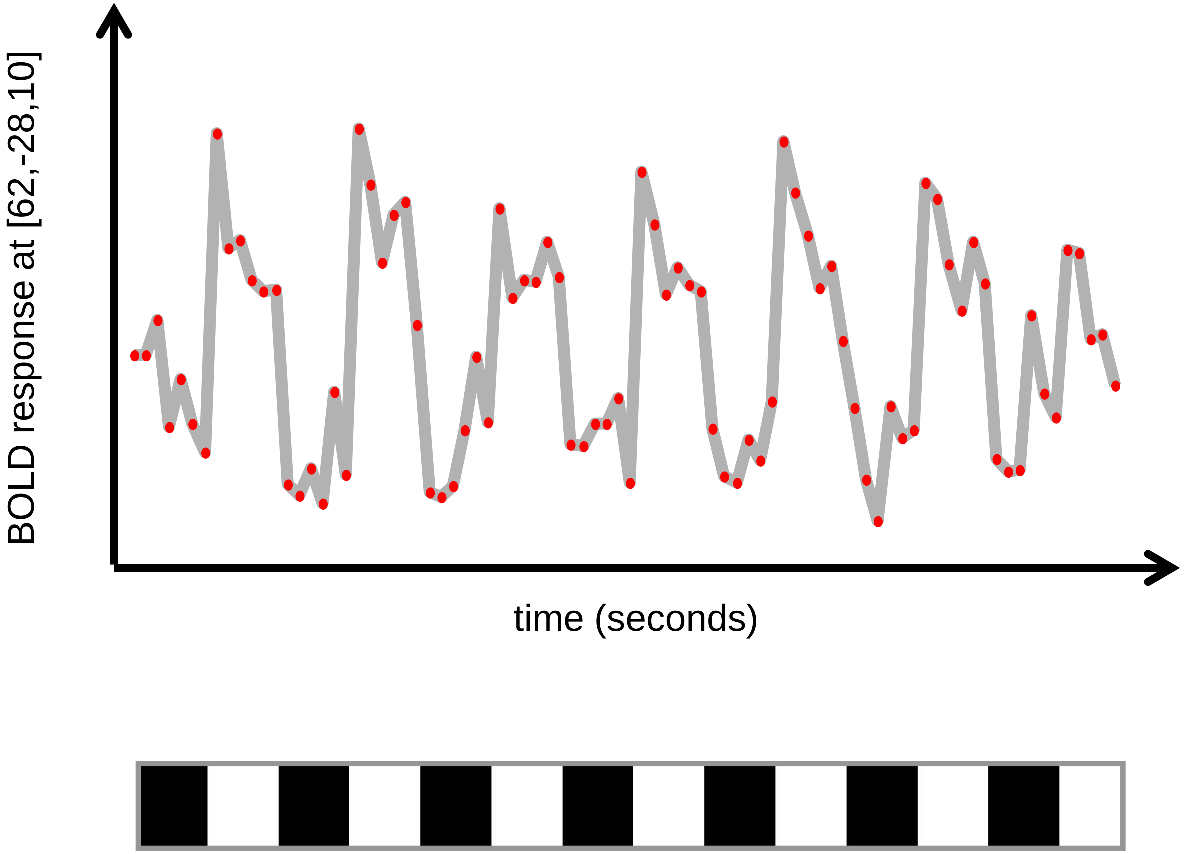
p <0.05

Example: Auditory block-design experiment

Passive word
listening
versus rest

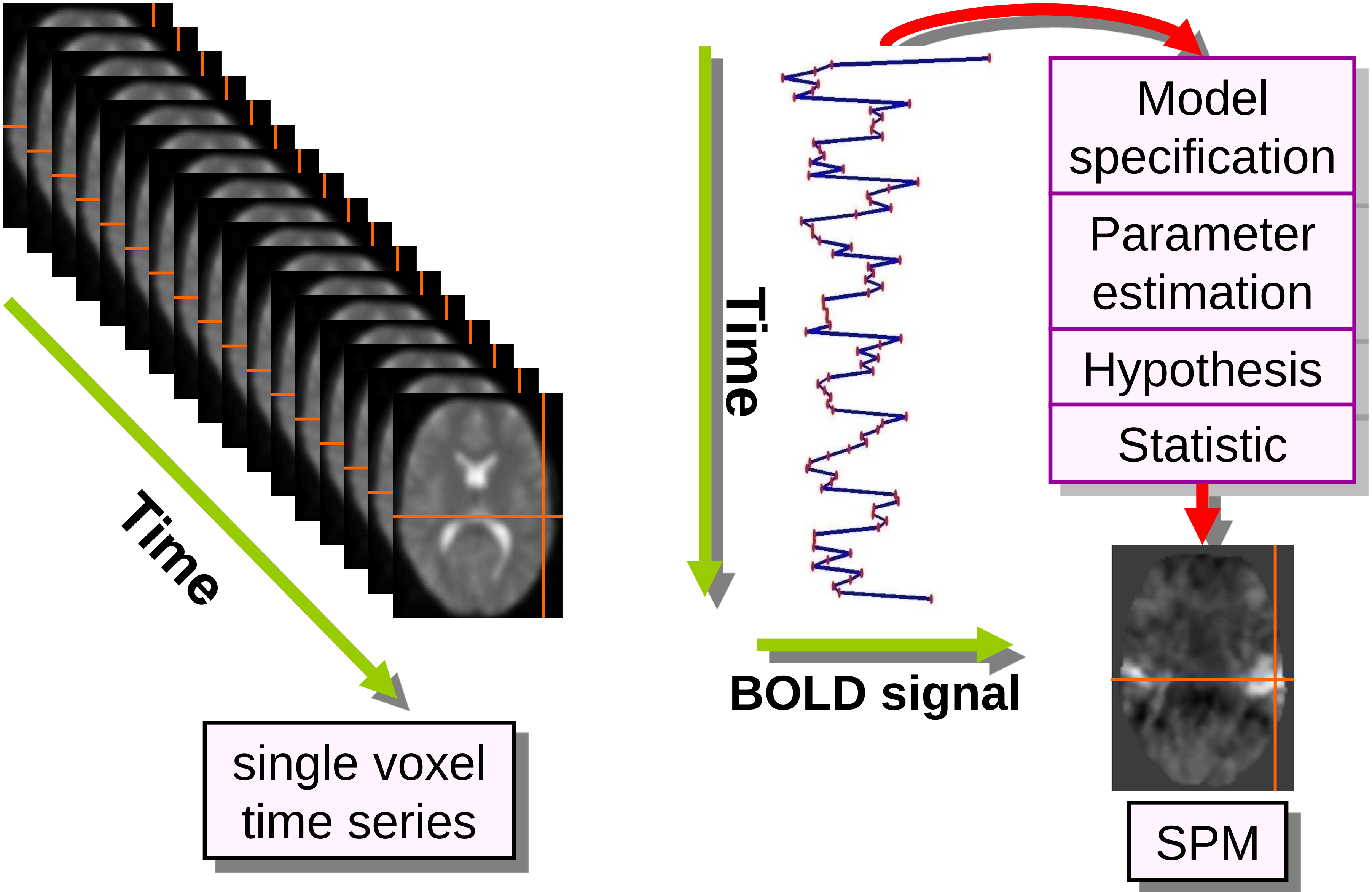
7 cycles of
rest and listening

Blocks of 6 scans
with 7 sec TR



Stimulus function

Voxel-wise time series analysis



Time

single voxel
time series

Time

BOLD signal

Model
specification

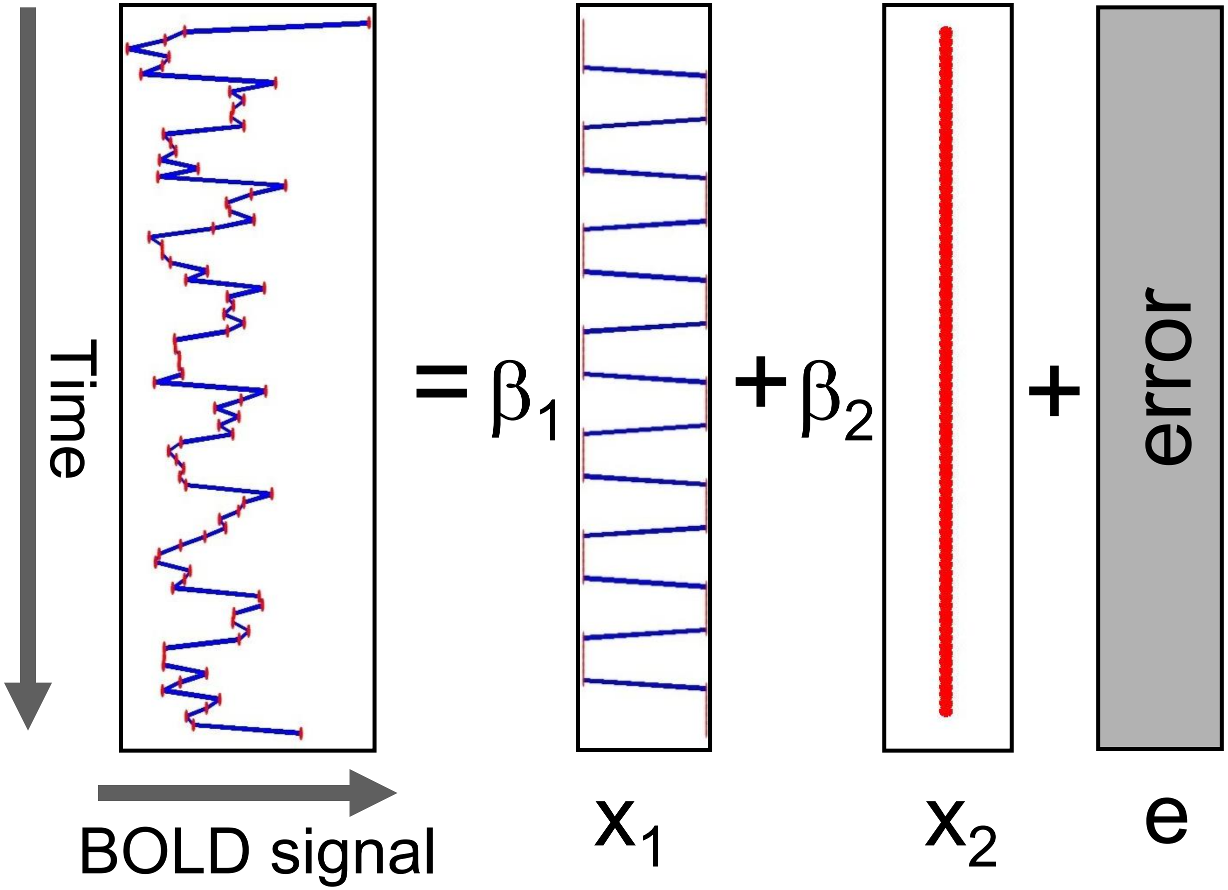
Parameter
estimation

Hypothesis

Statistic

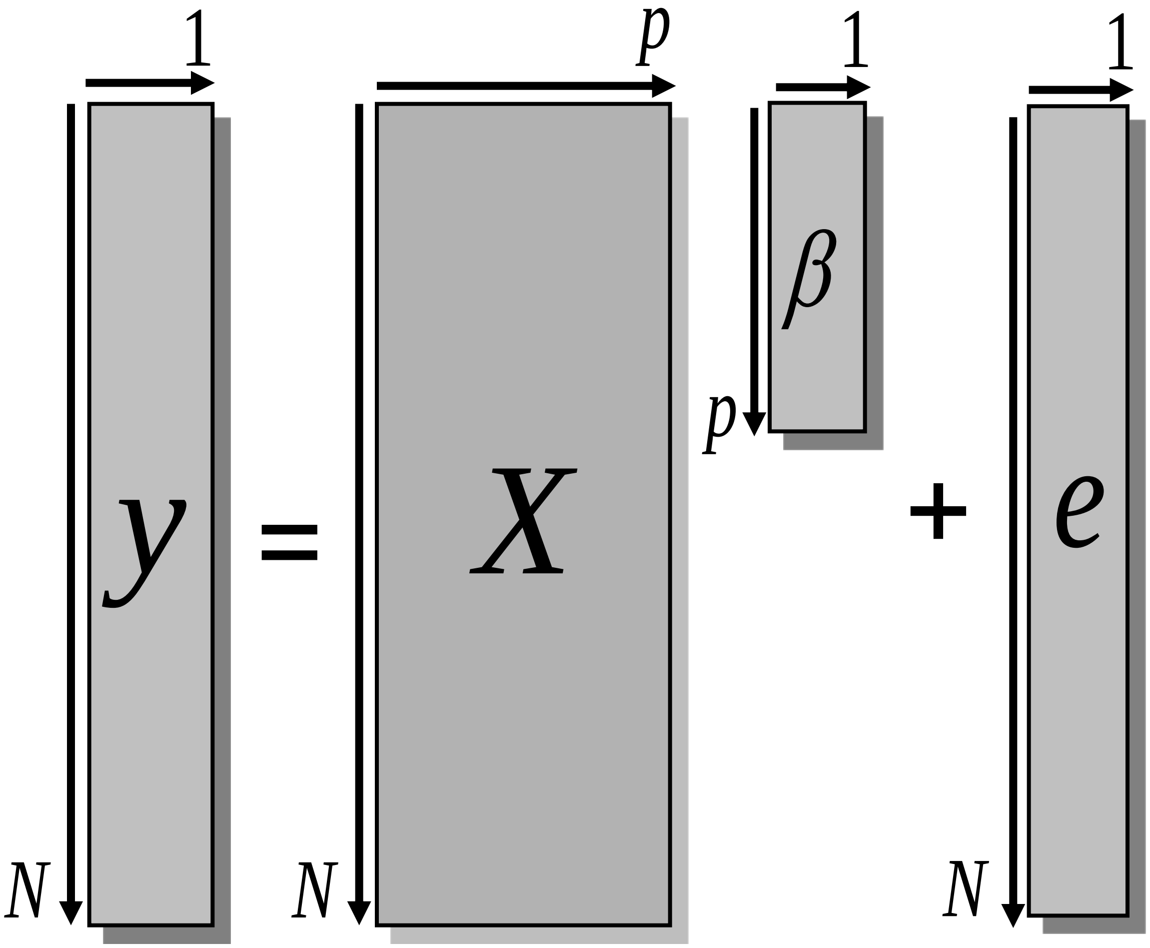
SPM

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

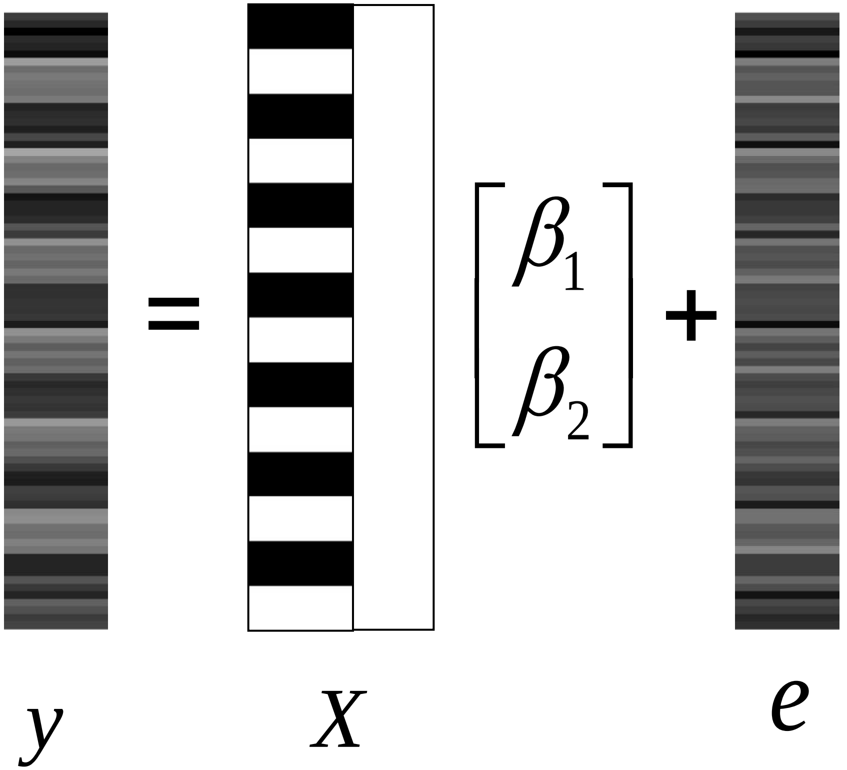
N: number of scans
p: number of regressors

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

GLM: a flexible framework for parametric analyses

- one sample *t*-test
- two sample *t*-test
- paired *t*-test
- Analysis of Variance (ANOVA)
- Analysis of Covariance (ANCoVA)
- correlation
- linear regression
- multiple regression

Parameter estimation



$$y = X\beta + e$$

Objective: estimate parameters to minimize

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

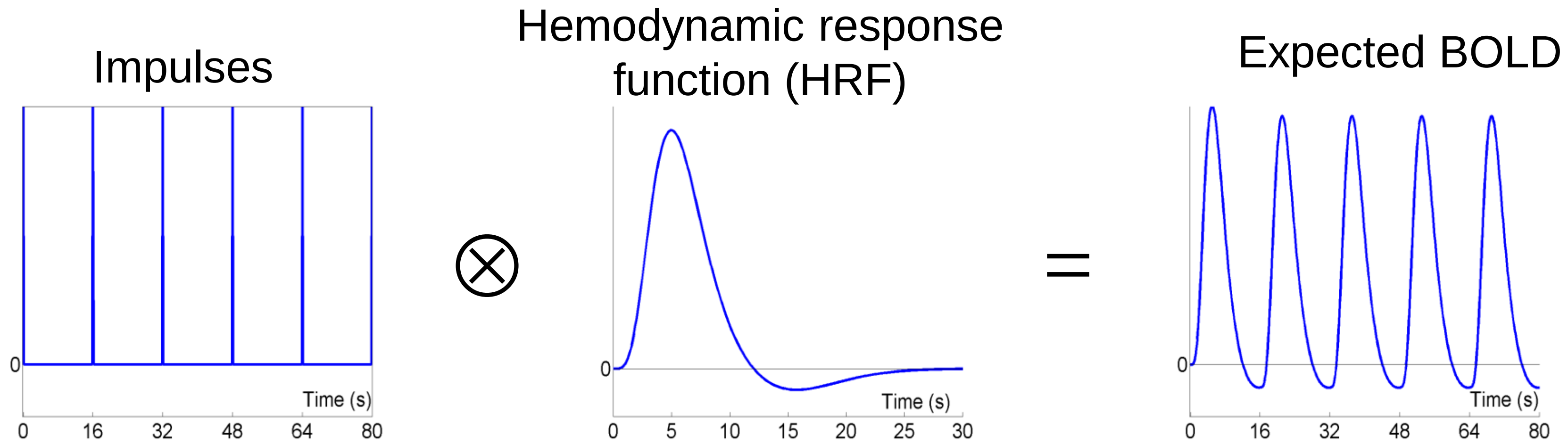
Ordinary least squares estimation (OLS) (assuming i.i.d. error):

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

Problems of this model with fMRI time series

1. The ***BOLD response*** has a delayed and dispersed shape.

BOLD response: Convolution model



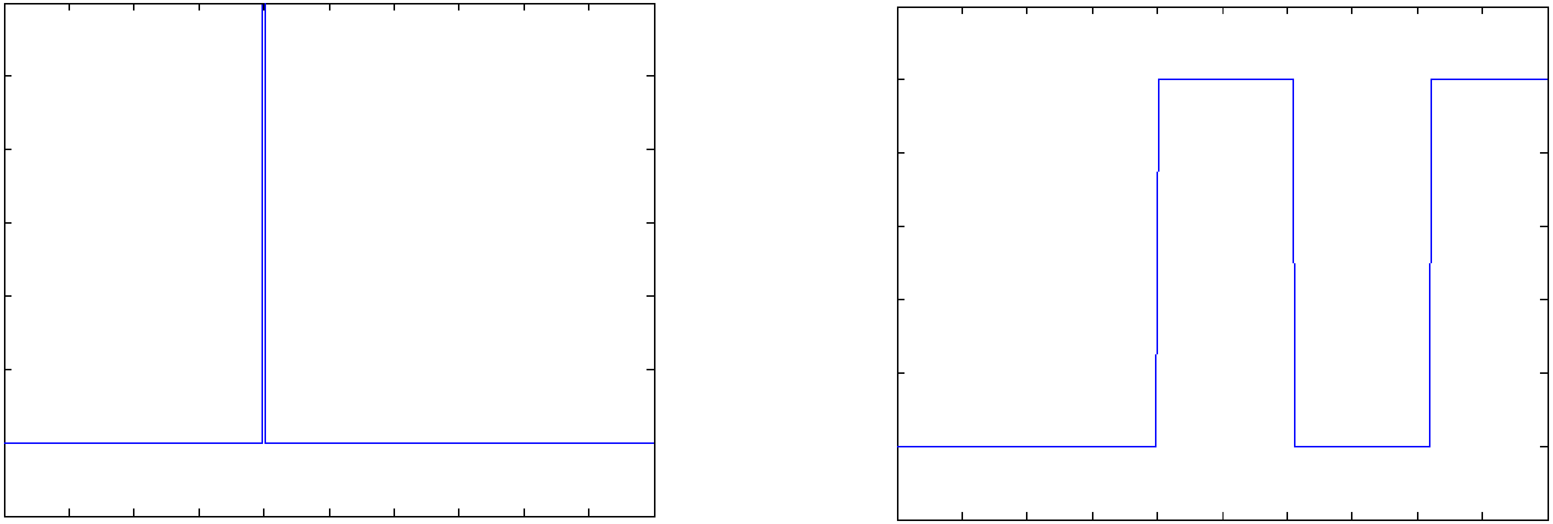
Linear time-invariant system: $f \otimes g(t) = \int_0^t f(\tau) g(t-\tau) d\tau$

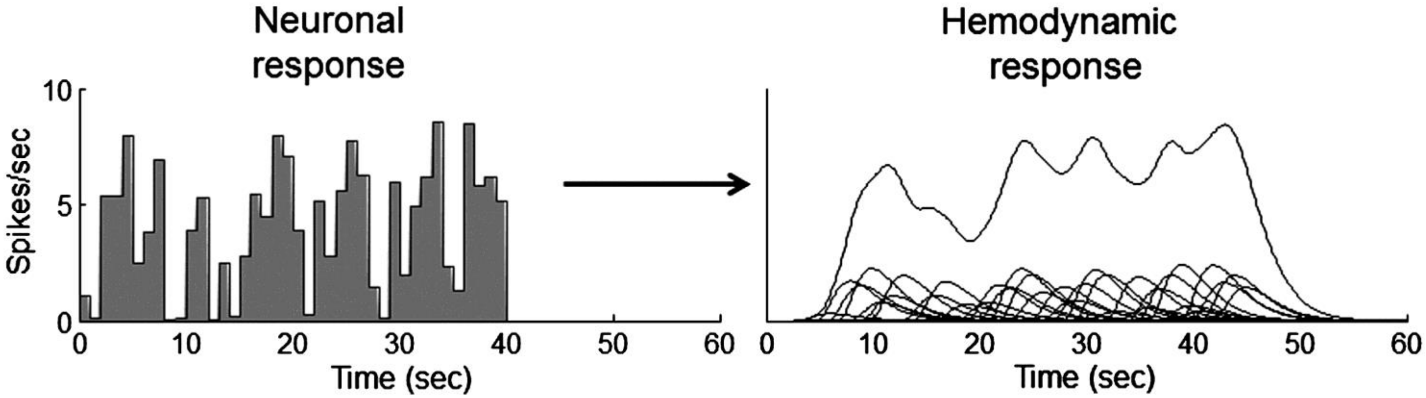
expected BOLD response

= input function ⊗ impulse response function (HRF)

Problem 1: BOLD response

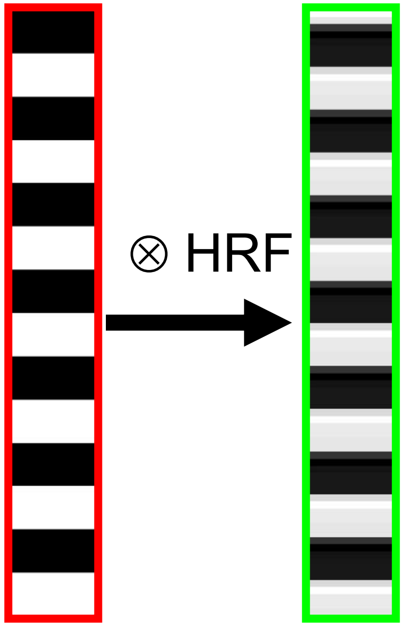
Solution: Convolution model

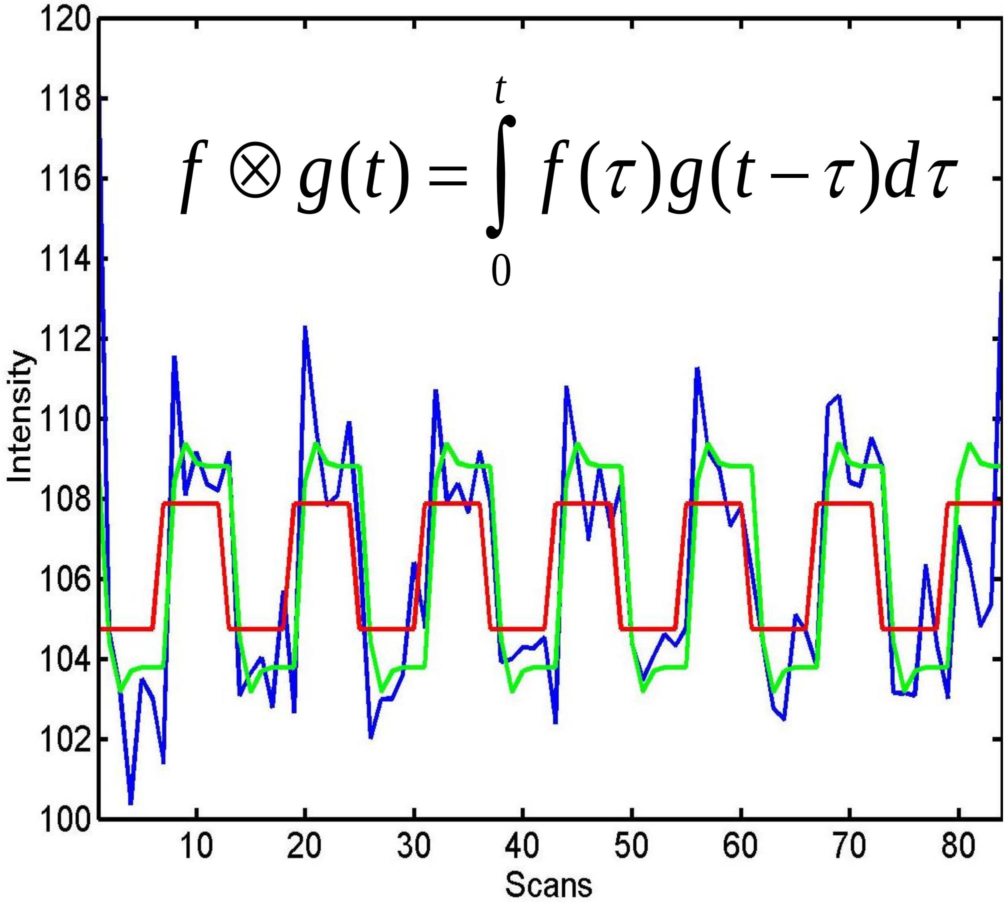




Convolution model of the BOLD response

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

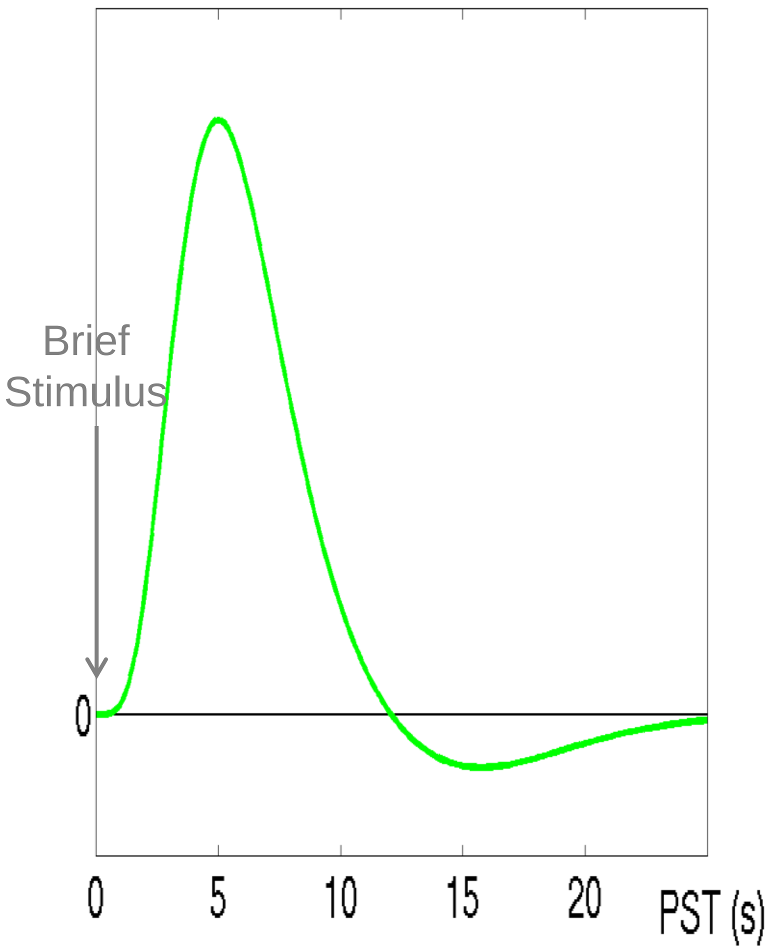




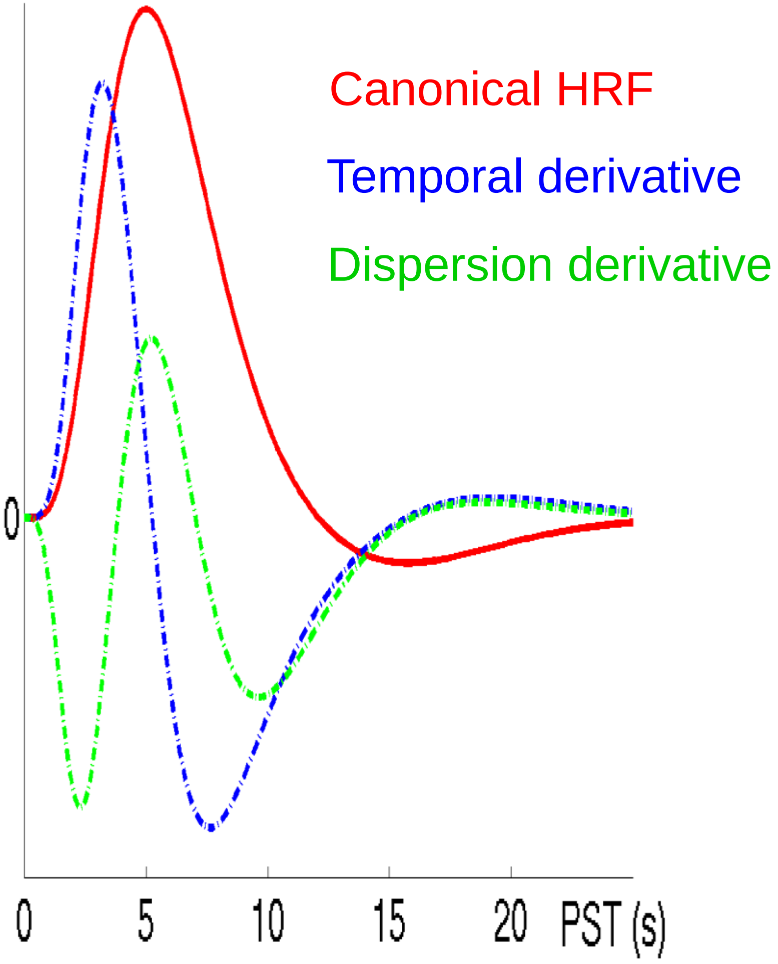
$$f \otimes g(t) = \int_0^t f(\tau)g(t-\tau)d\tau$$

Hemodynamic Response ⇨ Temporal Basis Set

Canonical HRF



Informed Basis Set



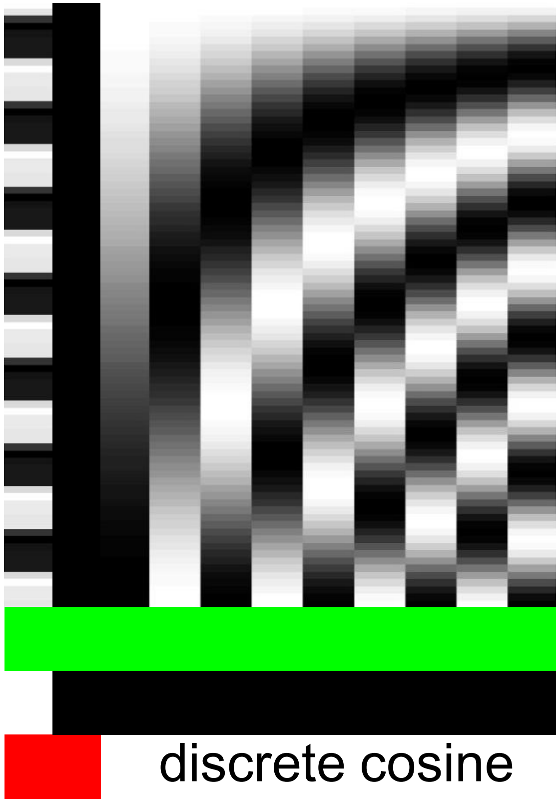
Problems of this model with fMRI time series

1. The *BOLD response* has a delayed and dispersed shape.

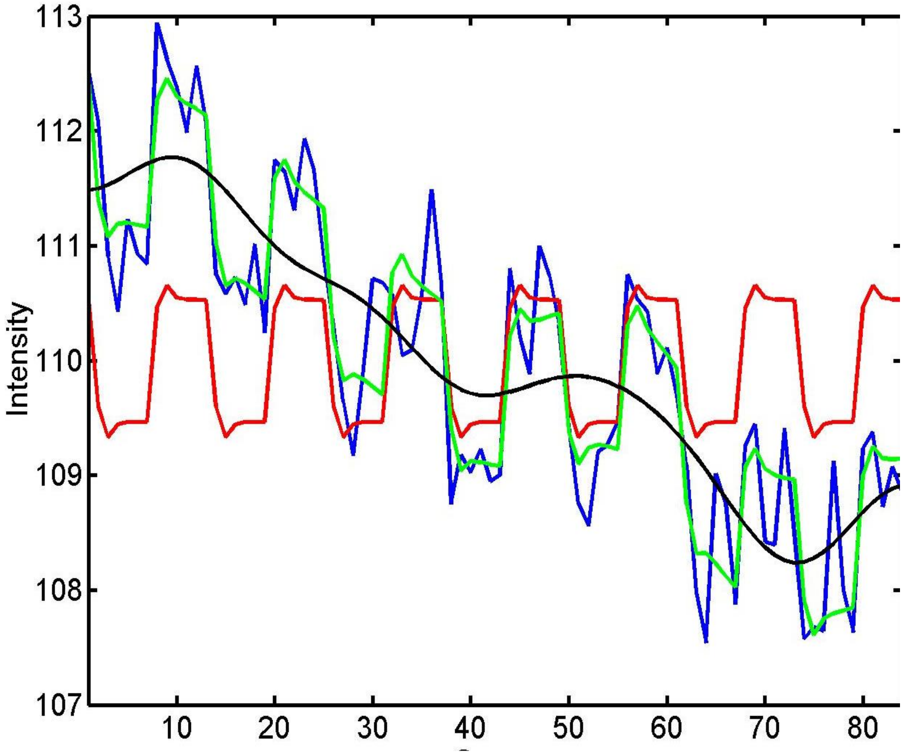
2. The BOLD signal includes substantial amounts of *low-frequency noise* (eg due to scanner drift).

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

Problems of this model with fMRI time series

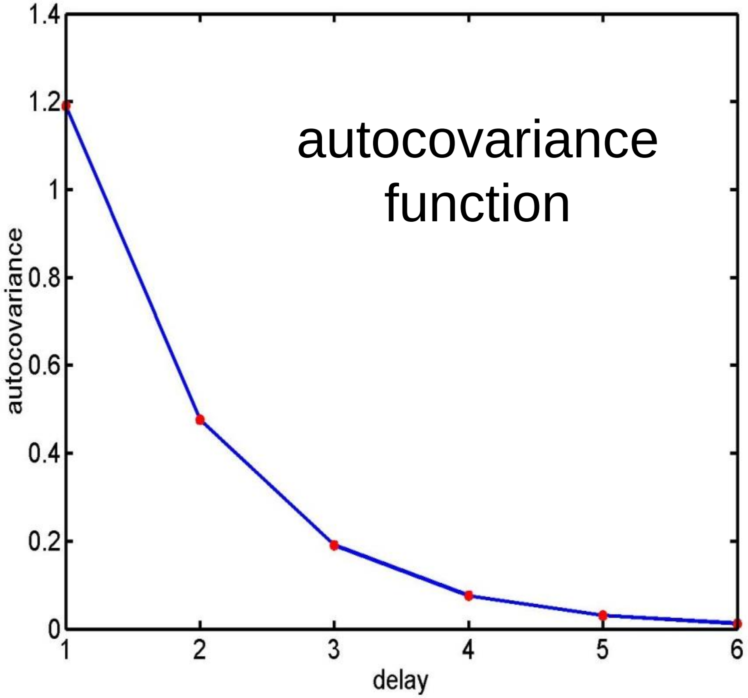
1. The *BOLD response* has a delayed and dispersed shape.

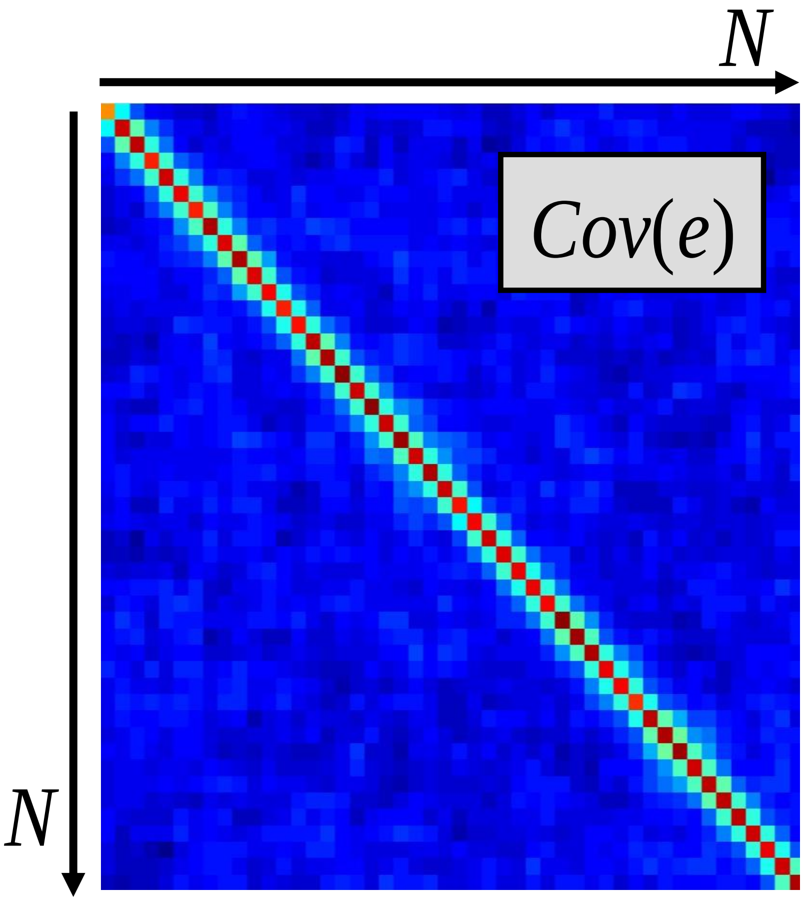
2. The BOLD signal includes substantial amounts of *low-frequency noise* (eg due to scanner drift).

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 3: Serial correlations

i.i.d: $\xcancel{e \sim N(0, \sigma^2 I)}$





Multiple covariance components

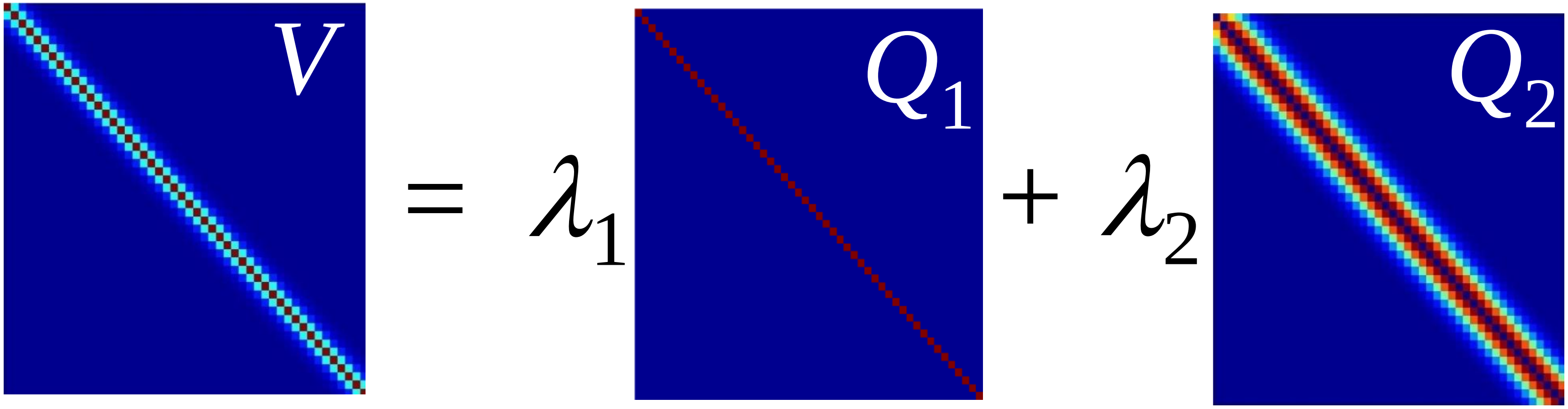
enhanced noise model at voxel *i*

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

$$C_i = \sigma_i^2 V$$

$$V = \sum \lambda_j Q_j$$

error covariance components Q
and hyperparameters λ

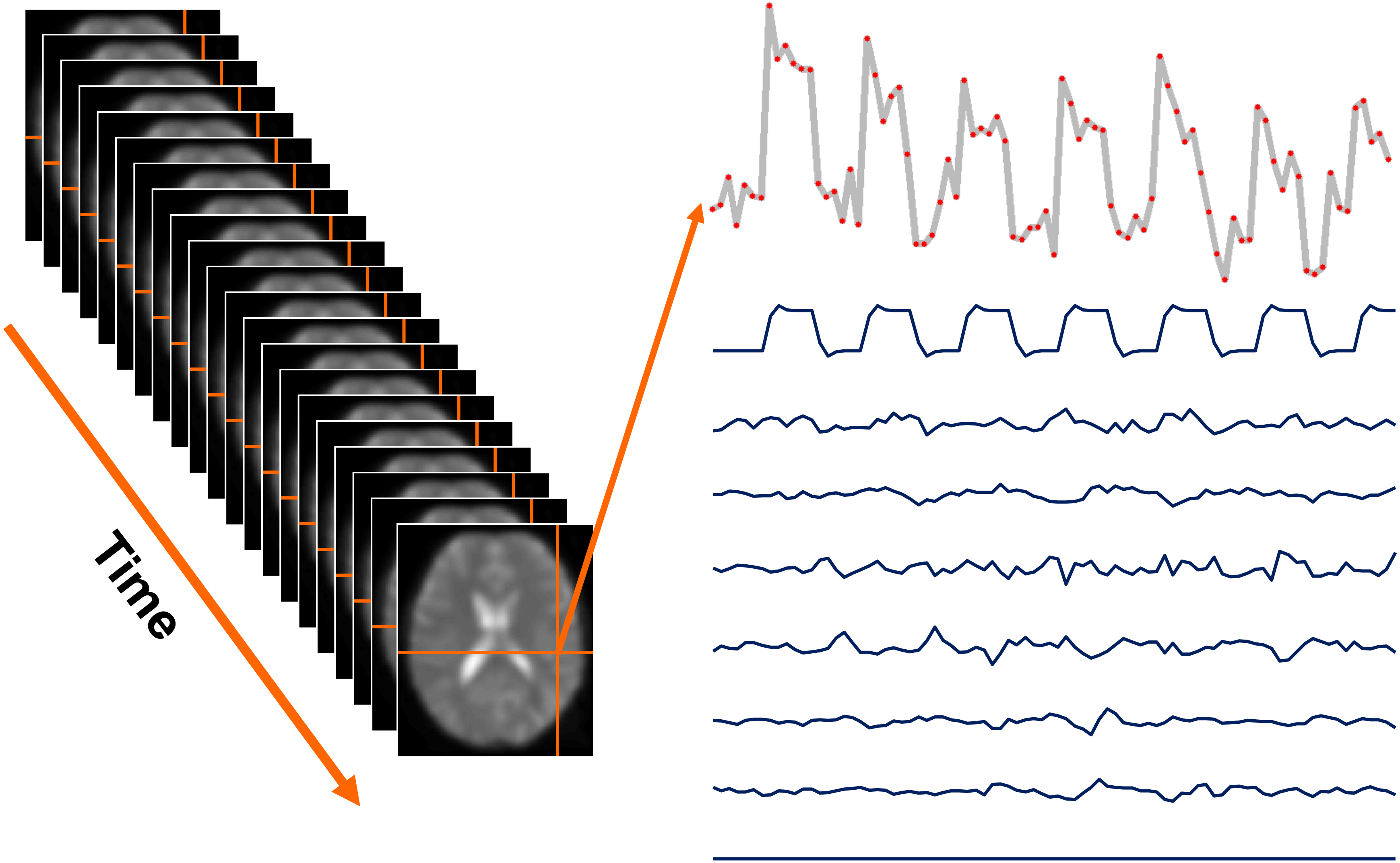


$$V = \lambda_1 Q_1 + \lambda_2 Q_2$$

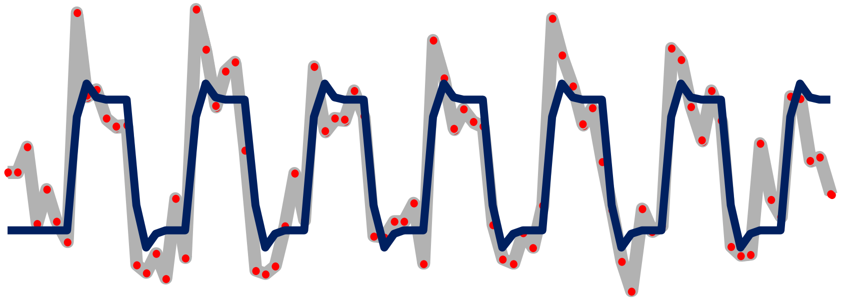
Estimation of hyperparameters λ with ReML (Restricted Maximum Likelihood).

Summary

A mass-univariate approach



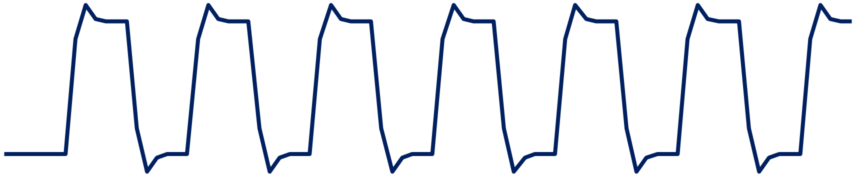
Estimation of the parameters



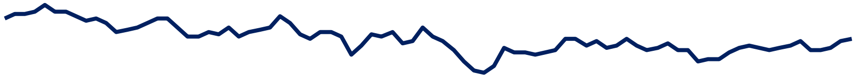
noise assumptions: $\varepsilon \sim N(0, \sigma^2 V)$

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

$\hat{\beta}_1 = 3.9831$

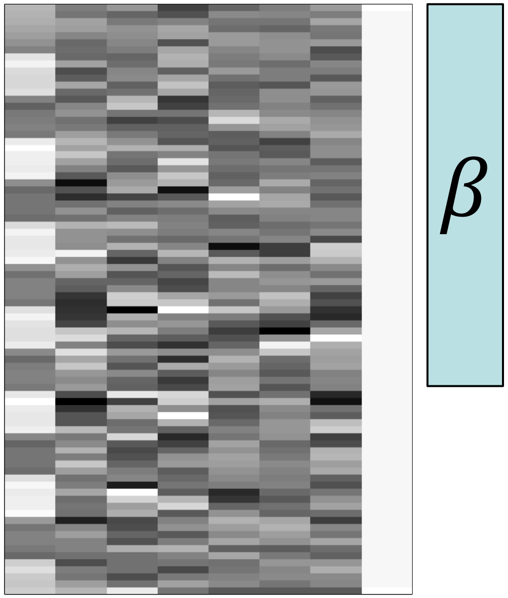


$\hat{\beta}_{2-7} = \{0.6871, 1.9598, 1.3902, 166.1007, 76.4770, -64.8189\}$



$\hat{\beta}_8 = 131.0040$

$y =$



$+\varepsilon$

$\hat{\varepsilon} =$

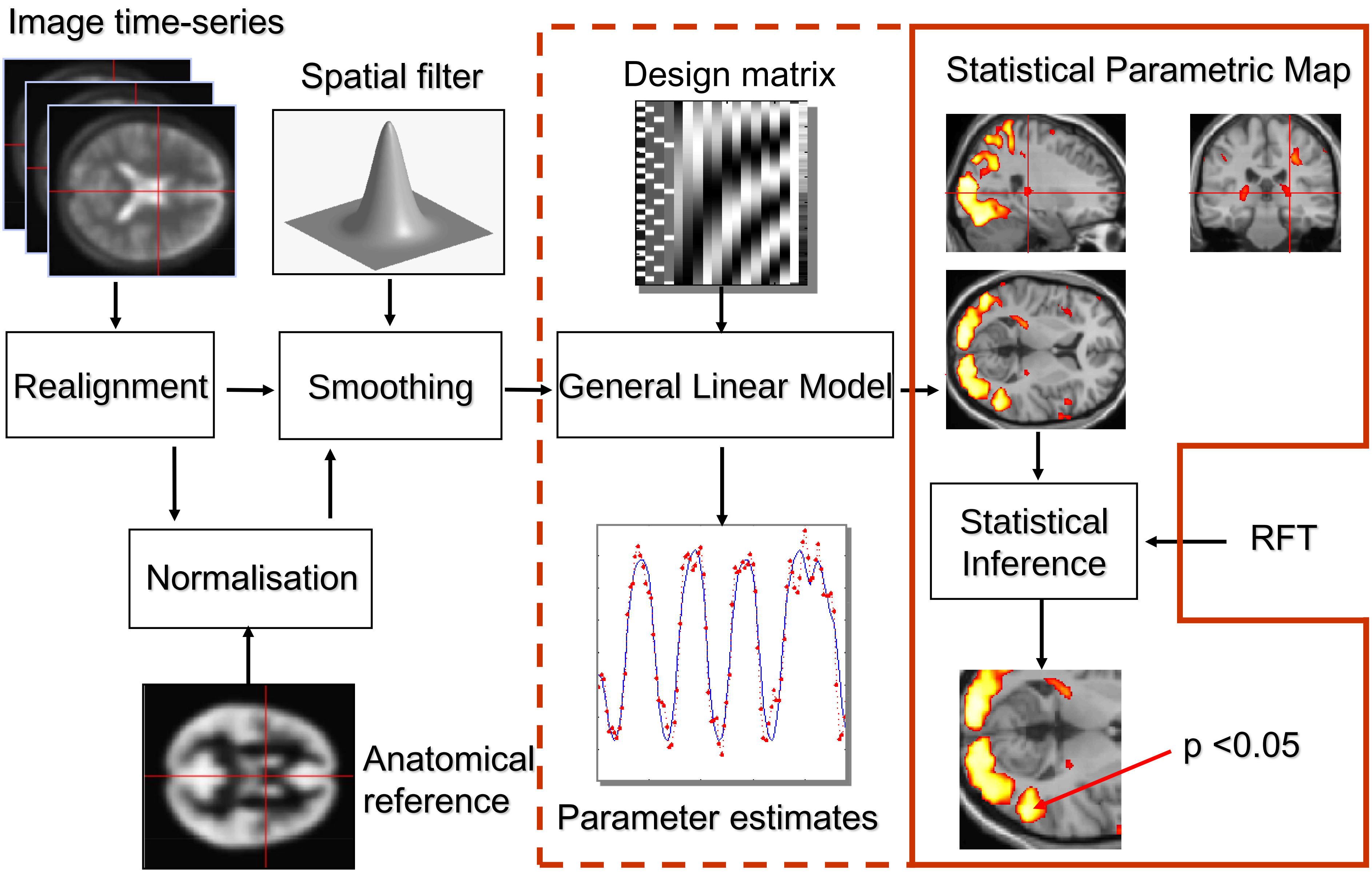


Image time-series

Spatial filter

Design matrix

Statistical Parametric Map

Realignment

Smoothing

General Linear Model

Statistical Inference

RFT

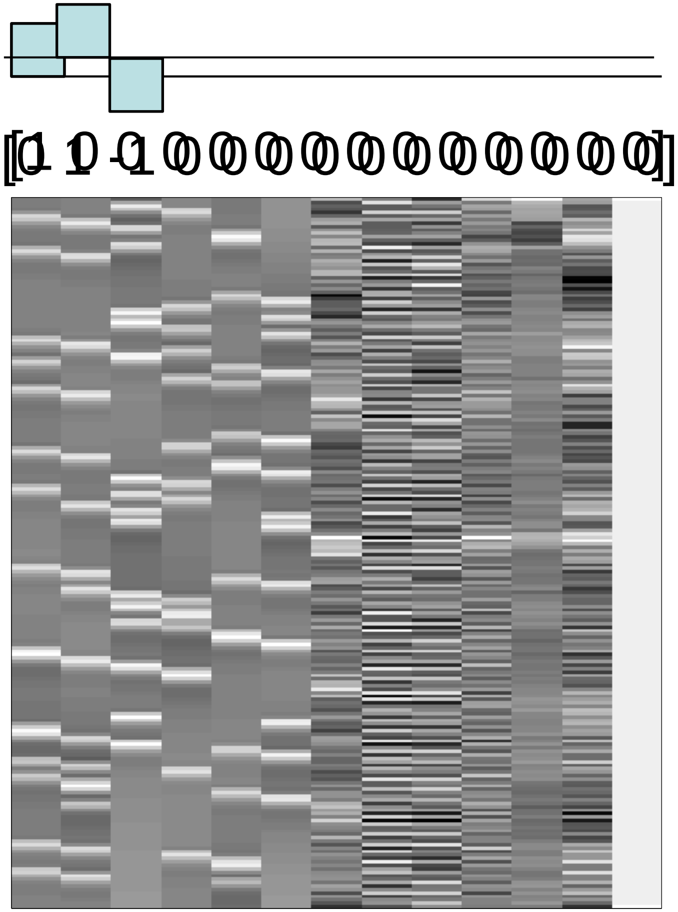
Normalisation

Anatomical reference

Parameter estimates

$p < 0.05$

Contrasts



- ☐ A contrast selects a specific effect of interest.
  - ⇨ A contrast c is a vector of length p.
  - ⇨ $c^T\beta$ is a linear combination of regression coefficients β.

$$c = [1\ 0\ 0\ 0\ \ \dots]^T$$

$$c^T\beta = \mathbf{1} \times \beta_1 + \mathbf{0} \times \beta_2 + \mathbf{0} \times \beta_3 + \mathbf{0} \times \beta_4 + \cdots$$

$$= \boldsymbol{\beta_1}$$

$$c = [0\ 1\ \ -1\ 0\ \ \dots]^T$$

$$c^T\beta = \mathbf{0} \times \beta_1 + \mathbf{1} \times \beta_2 + \mathbf{-1} \times \beta_3 + \mathbf{0} \times \beta_4 + \cdots$$

$$= \boldsymbol{\beta_2 - \beta_3}$$

Hypothesis Testing

To test an hypothesis, we construct “test statistics”.

- ☐ **Null Hypothesis H_0**

  Typically what we want to disprove (no effect).

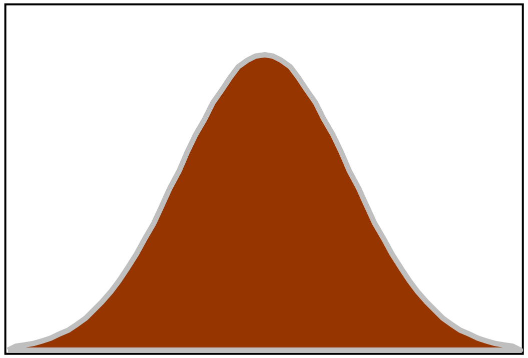
  ⇨ The Alternative Hypothesis H_A expresses outcome of interest.

- ☐ **Test Statistic T**

  The test statistic summarises evidence about H_0.

  Typically, test statistic is small in magnitude when the hypothesis H_0 is true and large when false.

  ⇨ We need to know the distribution of T under the null hypothesis.



Null Distribution of T

Hypothesis Testing

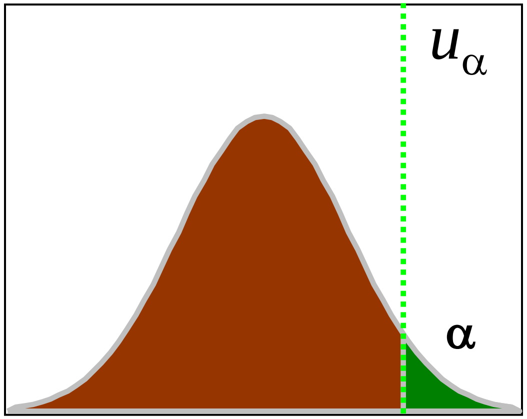
- ❑ ***Significance level α:***

  Acceptable *false positive rate α*.

  ⇨ threshold u_α

  Threshold u_α controls the false positive rate

  $$\alpha = p(T > u_\alpha \mid H_0)$$



Null Distribution of T

- ❑ **Conclusion about the hypothesis:**

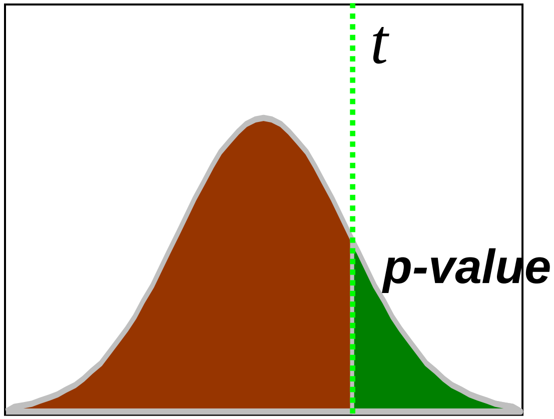
  We reject the null hypothesis in favour of the alternative hypothesis if $t > u_\alpha$

- ❑ ***p-value*:**

  A *p-value* summarises evidence against H_0.

  This is the chance of observing value more extreme than *t* under the null hypothesis.

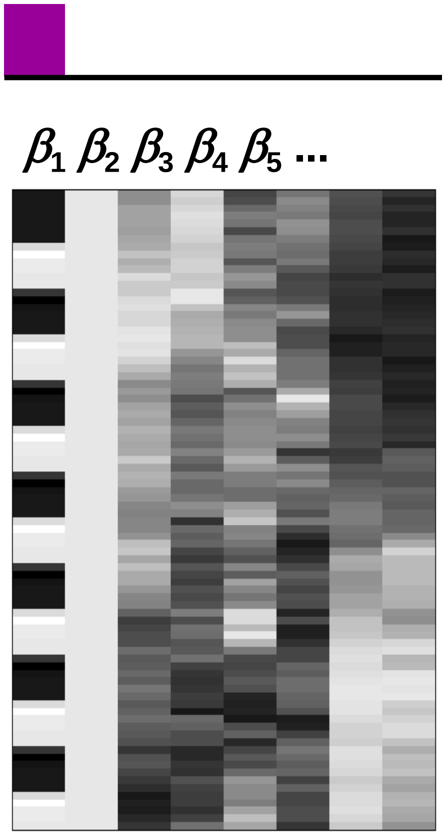
$$\boxed{p(T > t|H_0)}$$



Null Distribution of T

***T*-test** - one dimensional contrasts – SPM{*t*}

c^T = 1 0 0 0 0 0 0 0



Question: Amplitude of cond 1 > 0 ?

i.e.

$\beta_1 = c^T\beta > 0$?

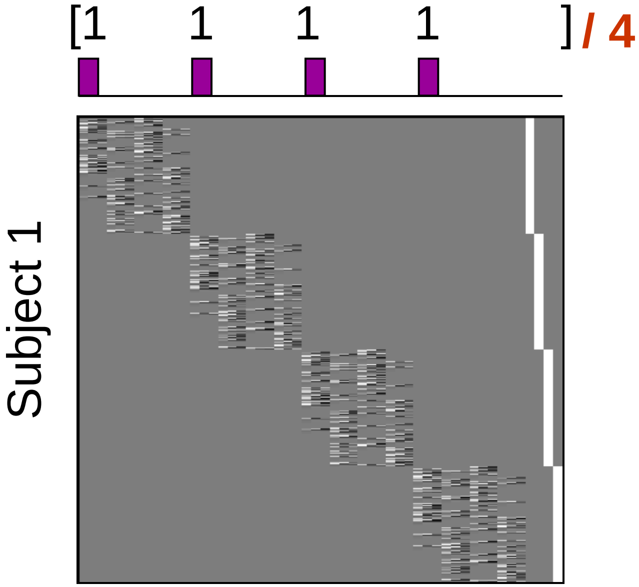
Null hypothesis: $\boxed{H_0: c^T\beta = 0}$ $H_A: c^T\beta > 0$

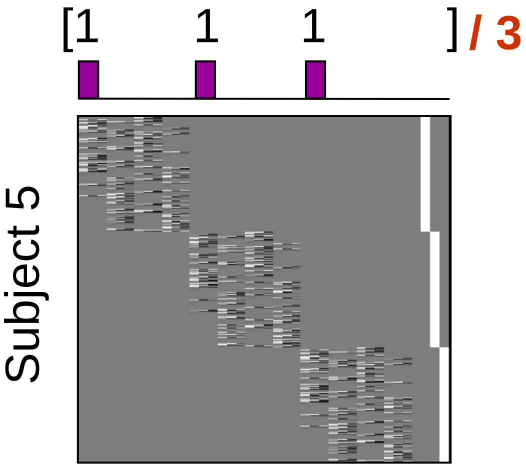
Test statistic:

$$T = \frac{\textbf{\textit{contrast} of estimated parameters}}{\sqrt{\textbf{variance estimate}}}$$

$$\boxed{T = \frac{c^T\hat{\beta}}{\sqrt{\mathrm{var}(c^T\hat{\beta})}} = \frac{c^T\hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T\left(X^TX\right)^{-1}c}} \sim t_{N-p}}$$

Scaling issue





$$T = \frac{c^T \hat{\beta}}{\sqrt{\mathrm{var}(c^T \hat{\beta})}} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T \left(X^T X\right)^{-1} c}}$$

- ❑ The *T*-statistic does not depend on the scaling of the regressors.
- ❑ The *T*-statistic does not depend on the scaling of the contrast.
- ❑ Contrast $c^T \hat{\beta}$ depends on scaling.

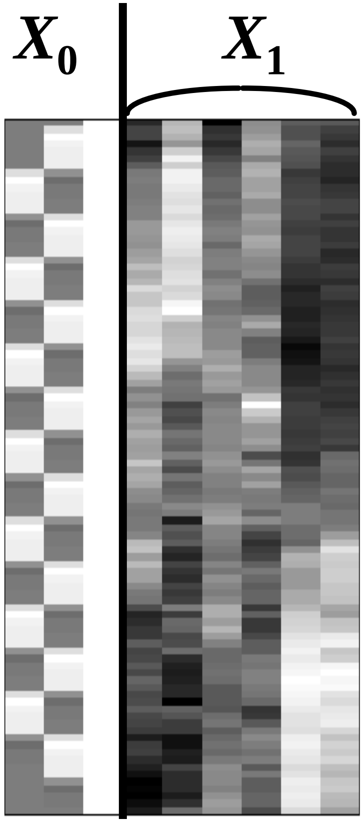
➢ Be careful of the interpretation of the contrasts $c^T \hat{\beta}$ themselves (eg, for a second level analysis):

sum ≠ average

***F*-test** - the extra-sum-of-squares principle

- Model comparison:

Null Hypothesis H_0: True model is X_0 (reduced model)



Full model ?

$\longrightarrow$ **RSS** $\sum \hat{\varepsilon}^2_{full}$

or Reduced model?

$\longrightarrow$ **RSS_0** $\sum \hat{\varepsilon}^2_{reduced}$

Test statistic: ratio of explained variability and unexplained variability (error)

$$F \propto \frac{RSS_0 - RSS}{RSS}$$

$$F \propto \frac{ESS}{RSS} \sim F_{\nu_1,\nu_2}$$

$\nu_1 = \mathrm{rank}(X) - \mathrm{rank}(X_0)$
$\nu_2 = N - \mathrm{rank}(X)$

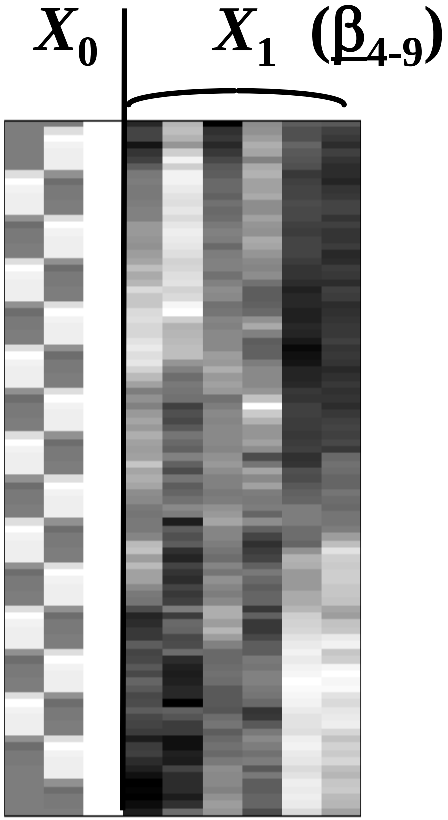
***F*-test** - multidimensional contrasts – SPM{*F*}

- Tests multiple linear hypotheses:

H_0: True model is X_0

H_0: $\beta_4 = \beta_5 = \ldots = \beta_9 = 0$

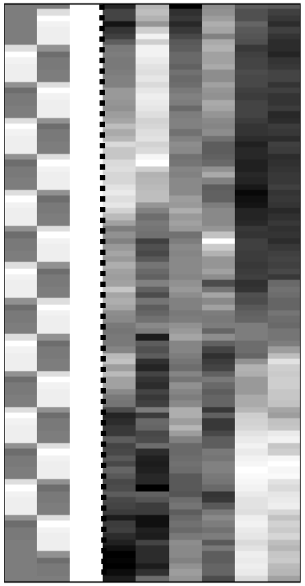
test H_0 : $c^T\beta = 0$?

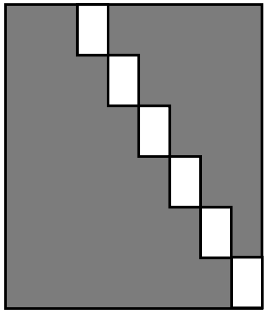


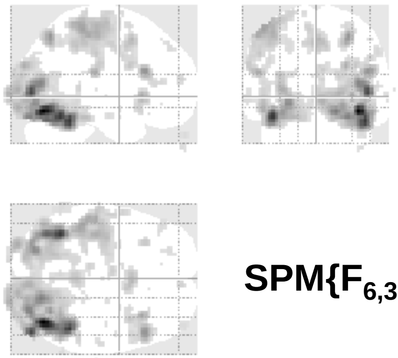
Full model?

Reduced model?

$$c^T = \begin{array}{ccc|cccccc}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}$$







SPM{$F_{6,322}$}

***F*-test:** summary

- F-tests can be viewed as testing for the additional variance explained by a larger model wrt a simpler (***nested***) model ⇨ ***model comparison***.

- Hypotheses:

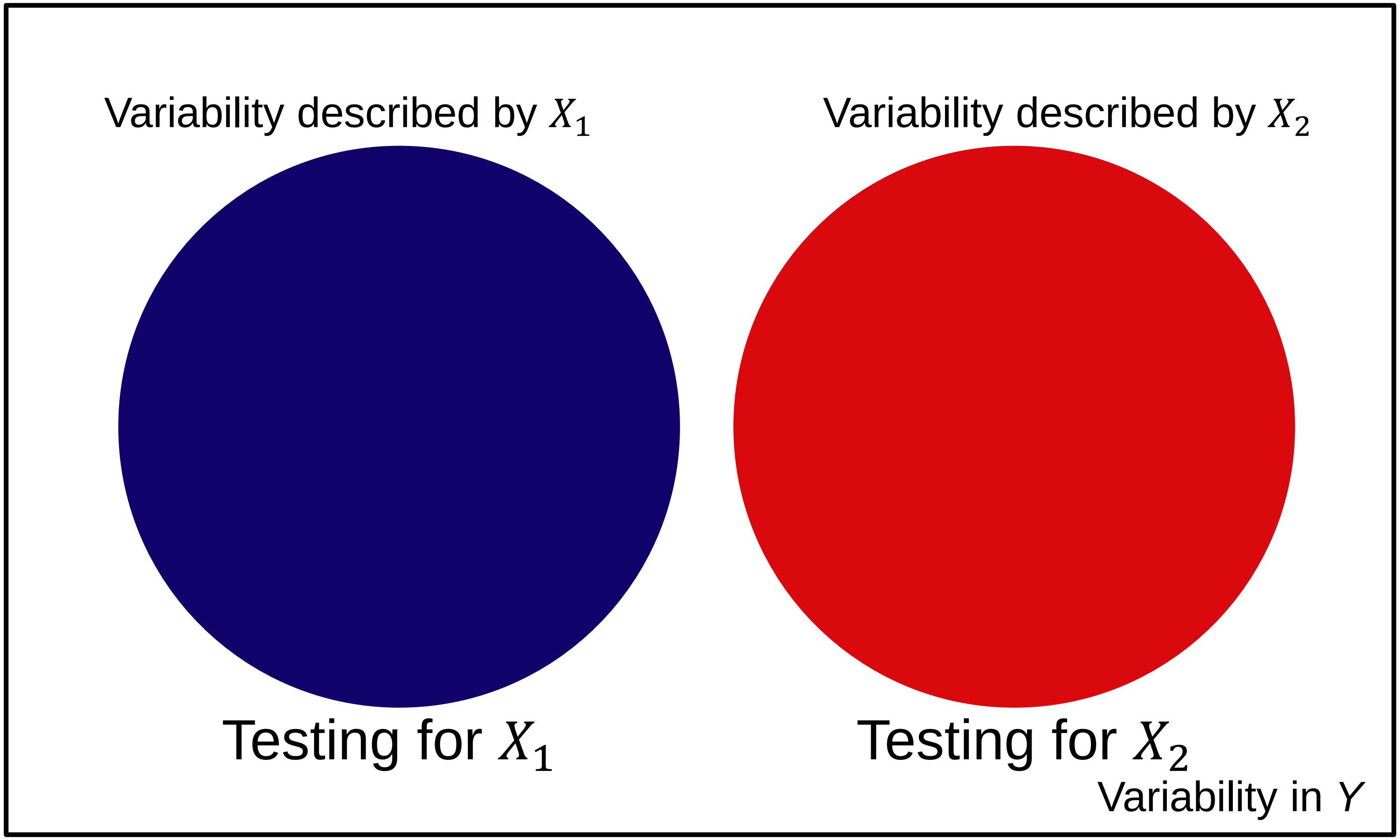
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{Null Hypothesis} H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$\text{Alternative Hypothesis } H_A : \text{at least one } \beta_k \neq 0$$

- In testing uni-dimensional contrast with an *F*-test, for example $\beta_1 - \beta_2$, the result will be the same as testing $\beta_2 - \beta_1$. It will be exactly the square of the *t*-test, testing for both positive and negative effects.

Orthogonal regressors



Variability described by X_1

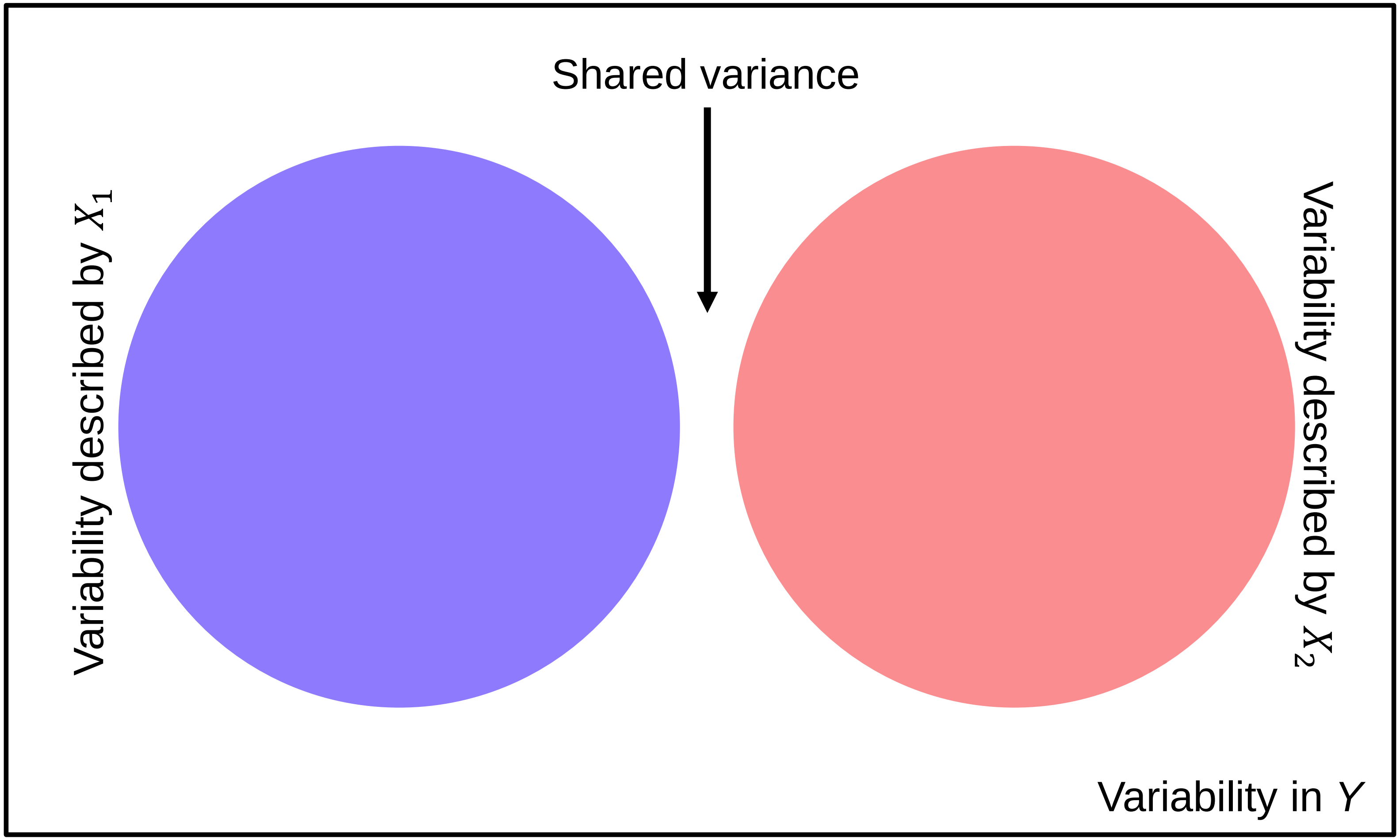
Variability described by X_2

Testing for X_1

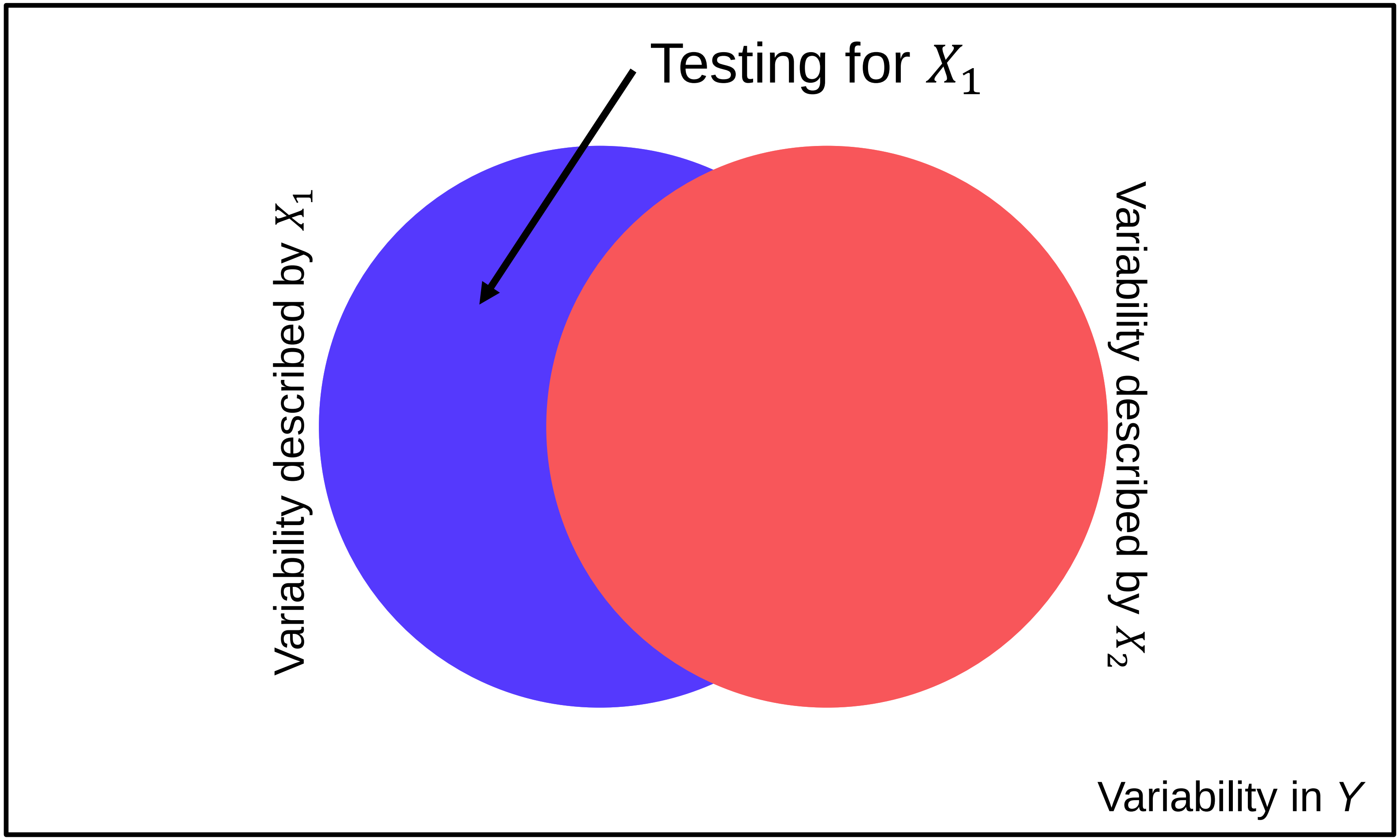
Testing for X_2

Variability in Y

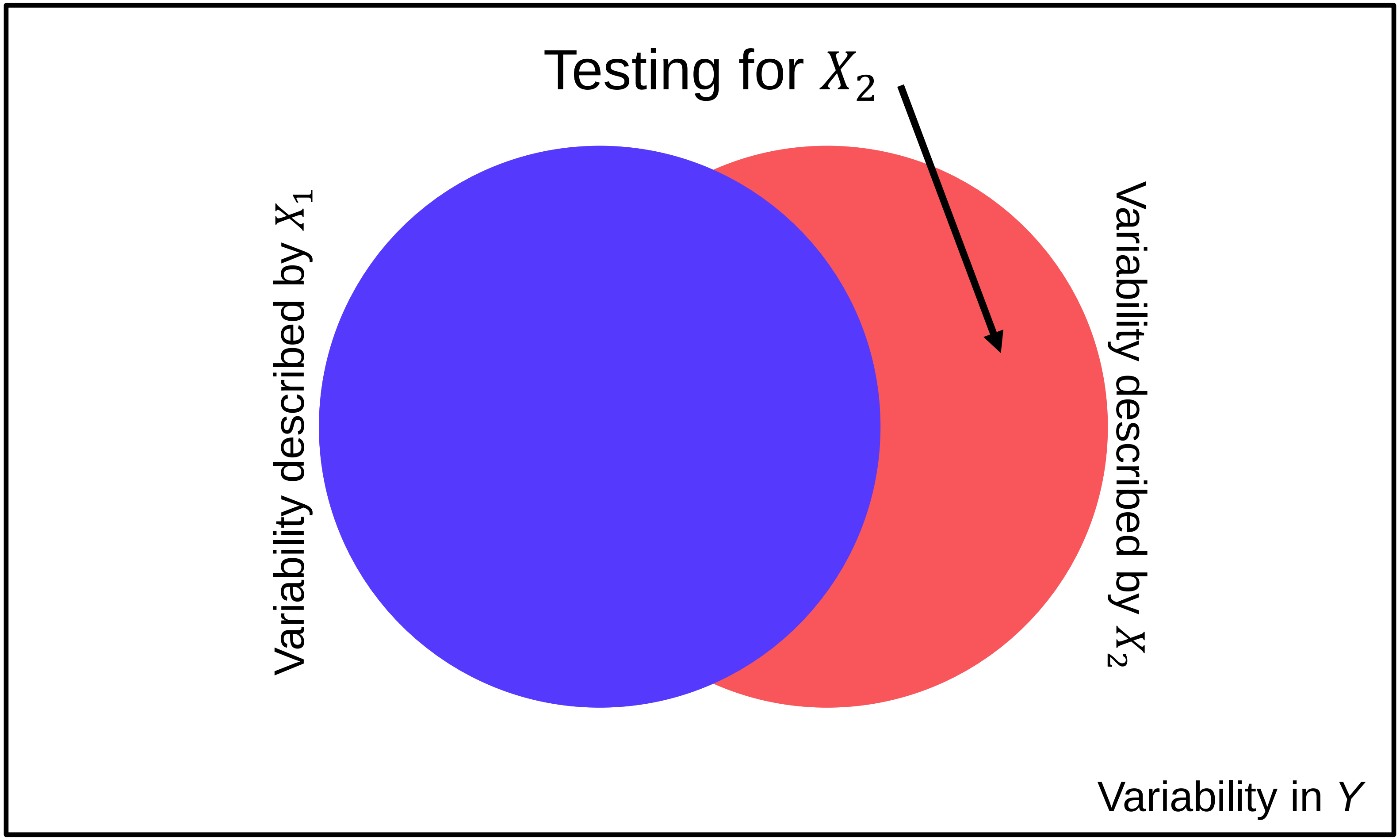
Correlated regressors



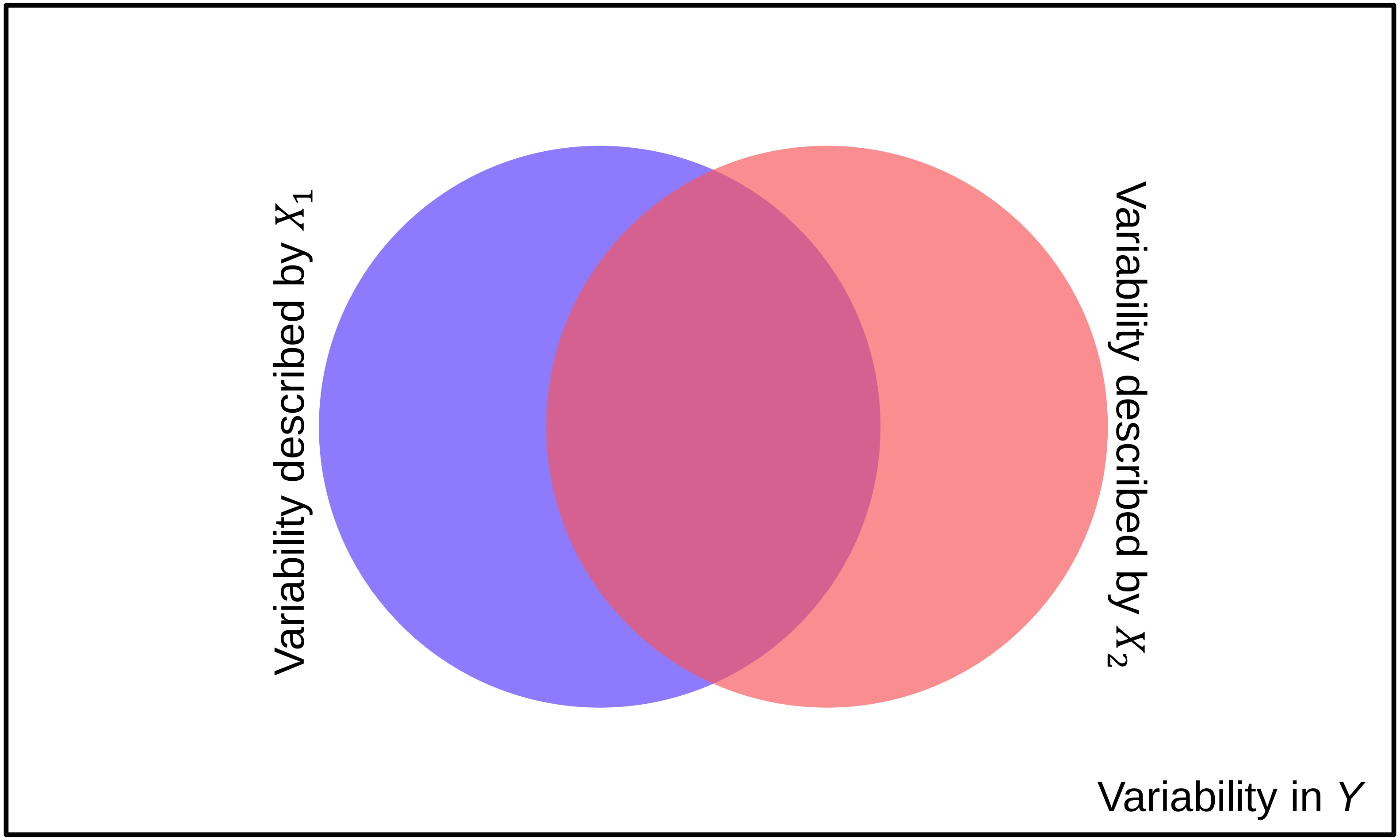
Correlated regressors



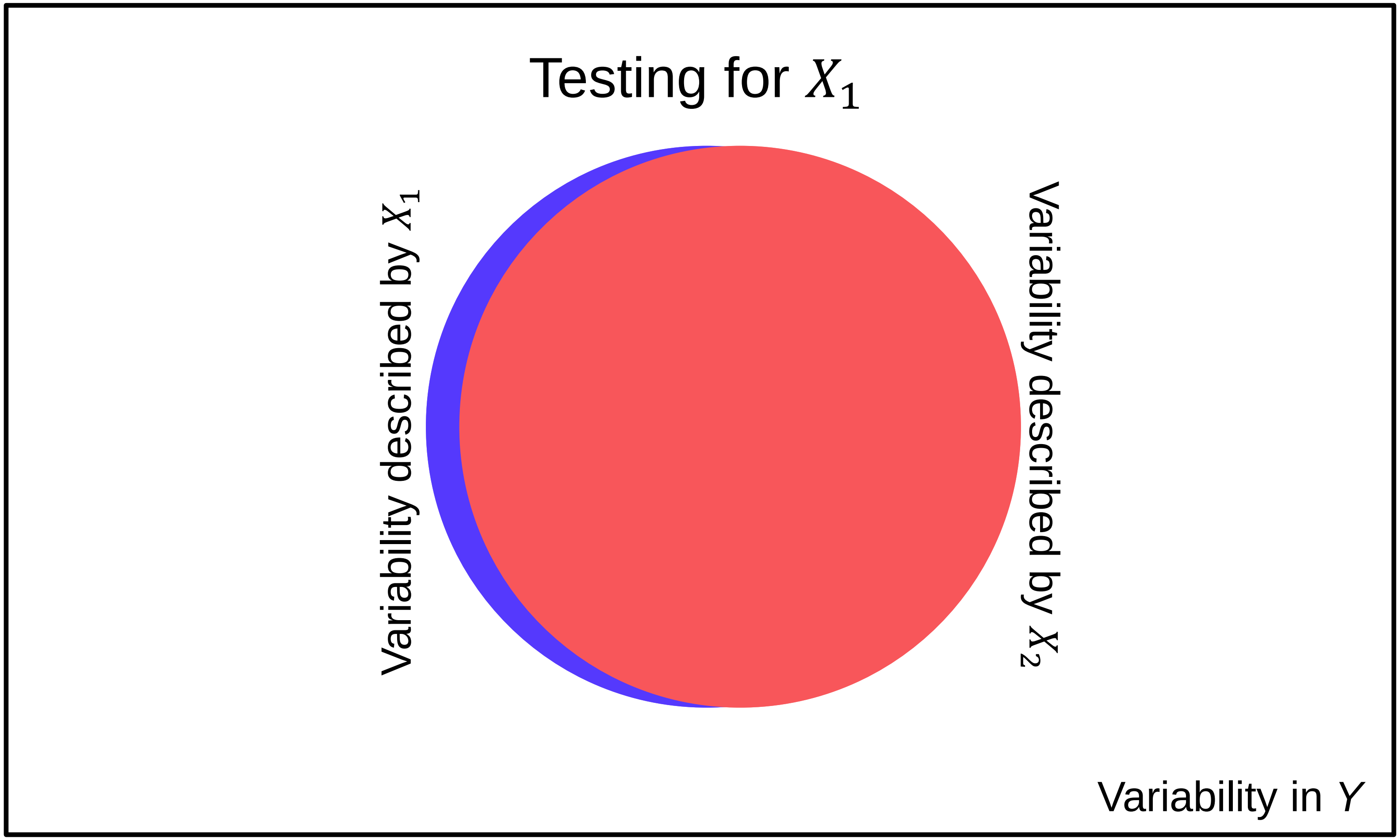
Correlated regressors



Correlated regressors



Correlated regressors



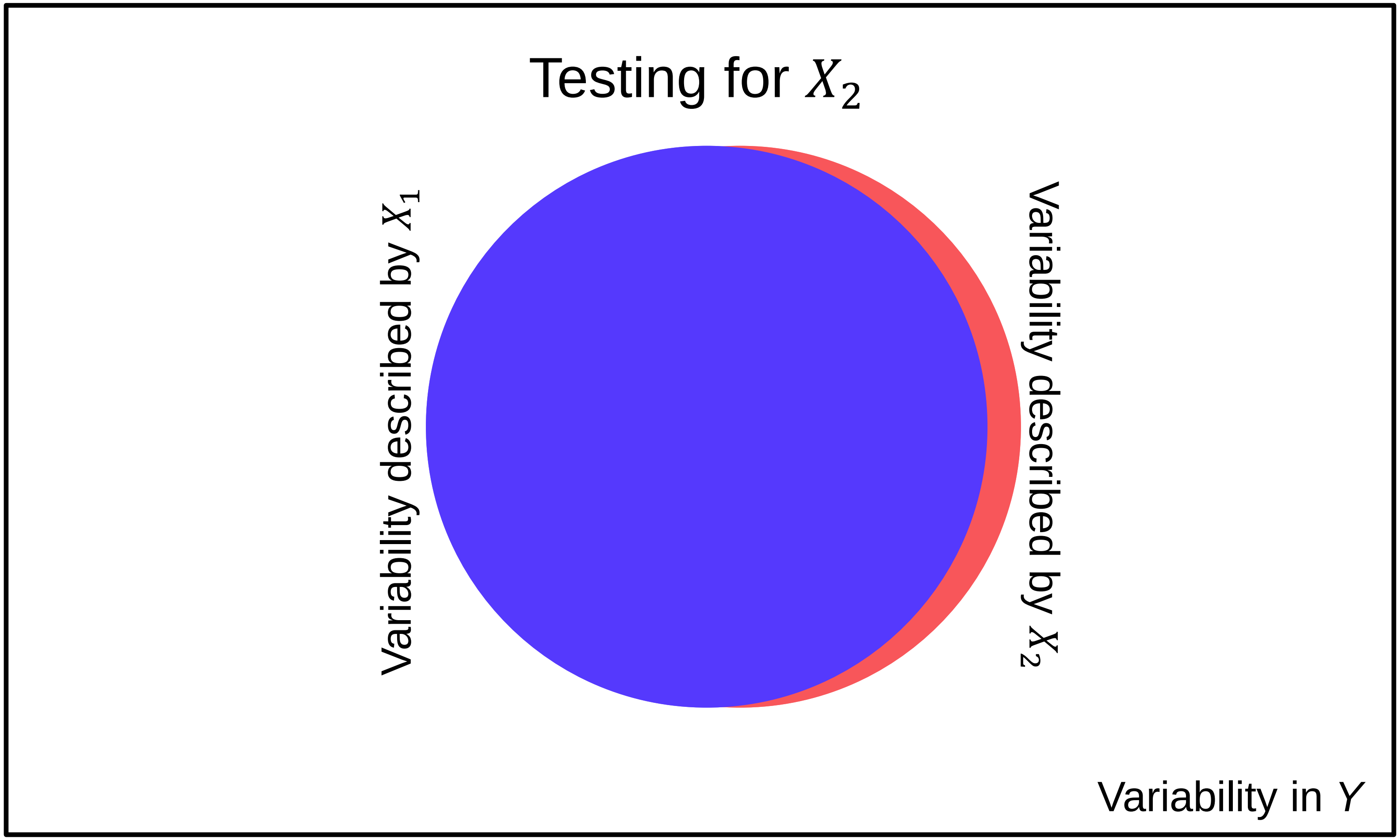
Testing for X_1

Variability described by X_1

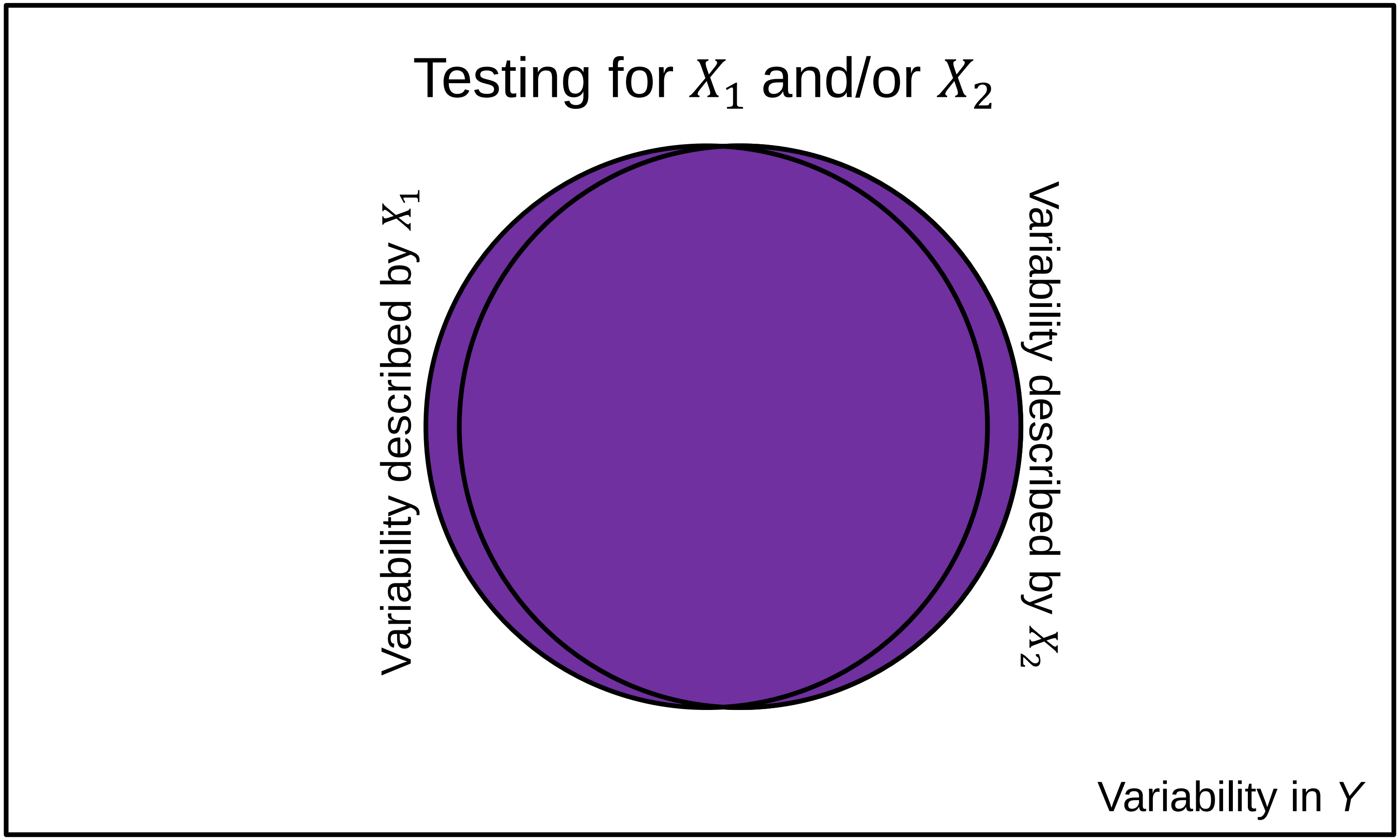
Variability described by X_2

Variability in Y

Correlated regressors

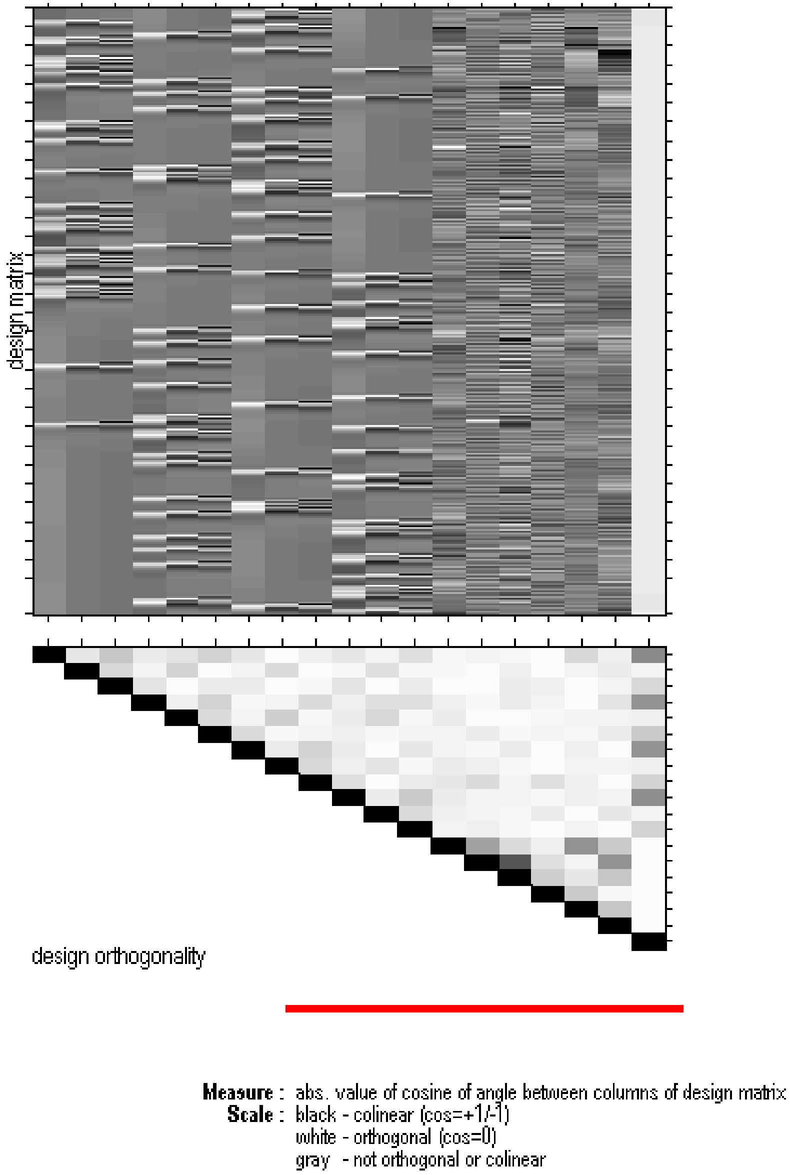


Correlated regressors



Orthogonalization of Regressors in fMRI Models, Mumford et al, PlosOne, 2015

Design orthogonality



Measure : abs. value of cosine of angle between columns of design matrix
Scale : black - colinear (cos=+1/-1)
white - orthogonal (cos=0)
gray - not orthogonal or colinear

- For each pair of columns of the design matrix, the orthogonality matrix depicts the magnitude of the **cosine of the angle** between them, with the range 0 to 1 mapped from white to black.

- If both vectors have **zero mean** then the cosine of the angle between the vectors is the same as the **correlation** between the two variates.

Design efficiency

$$T = \frac{c^T \hat{\beta}}{\sqrt{\mathrm{var}(c^T \hat{\beta})}}$$

- The aim is to minimize the standard error of a *t*-contrast (i.e. the denominator of a t-statistic).

$$\mathrm{var}(c^T \hat{\beta}) = \hat{\sigma}^2 c^T (X^T X)^{-1} c$$

- This is equivalent to maximizing the efficiency *e*:

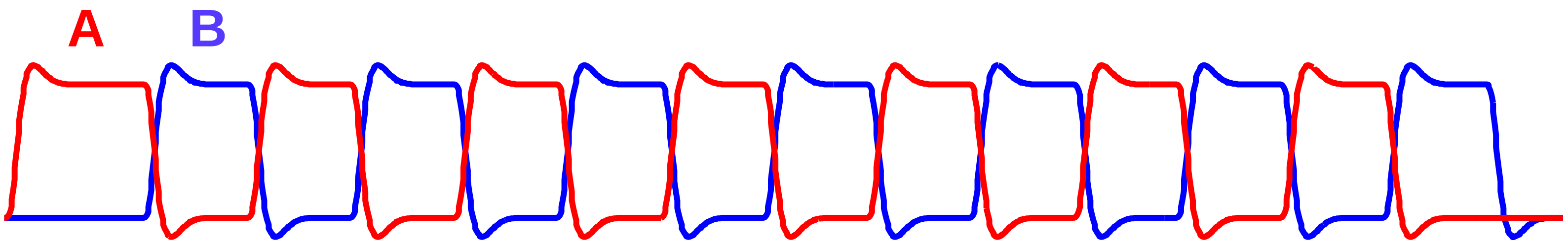
$$e(\hat{\sigma}^2, c, X) = (\underbrace{\boxed{\hat{\sigma}^2}}_{\text{Noise variance}} c^T \underbrace{\boxed{(X^T X)}}_{\text{Design variance}}{}^{-1} c)^{-1}$$

- If we assume that the noise variance is independent of the specific design:

$$e(c, X) = (c^T (X^T X)^{-1} c)^{-1}$$

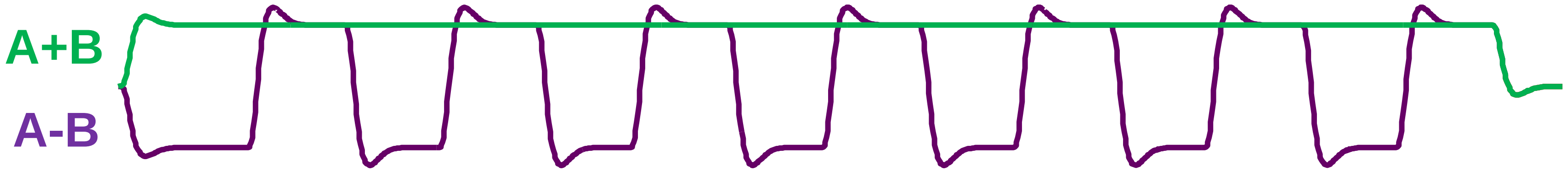
- This is a relative measure: all we can really say is that one design is more efficient than another (for a given contrast).

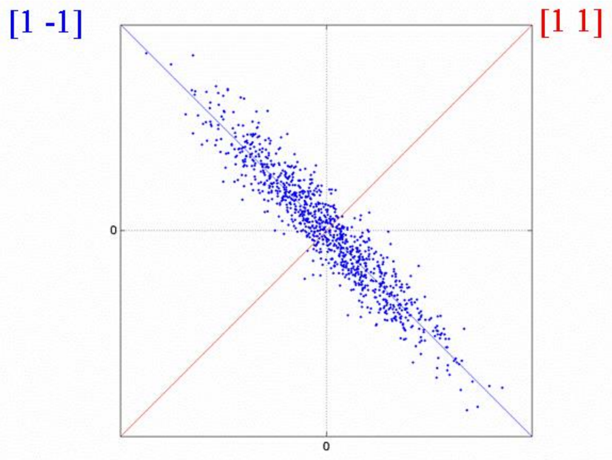
Design efficiency



$$X^T X = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$$

$$c = [1\ 0]^T: \quad e(c, X) = 18.1$$
$$c = [0.5\ 0.5]^T: \quad e(c, X) = 19.0$$
$$c = [1\ -1]^T: \quad e(c, X) = 95.2$$





- ❑ High correlation between regressors leads to low sensitivity to each regressor alone.
- ❑ We can still estimate efficiently the difference between them.

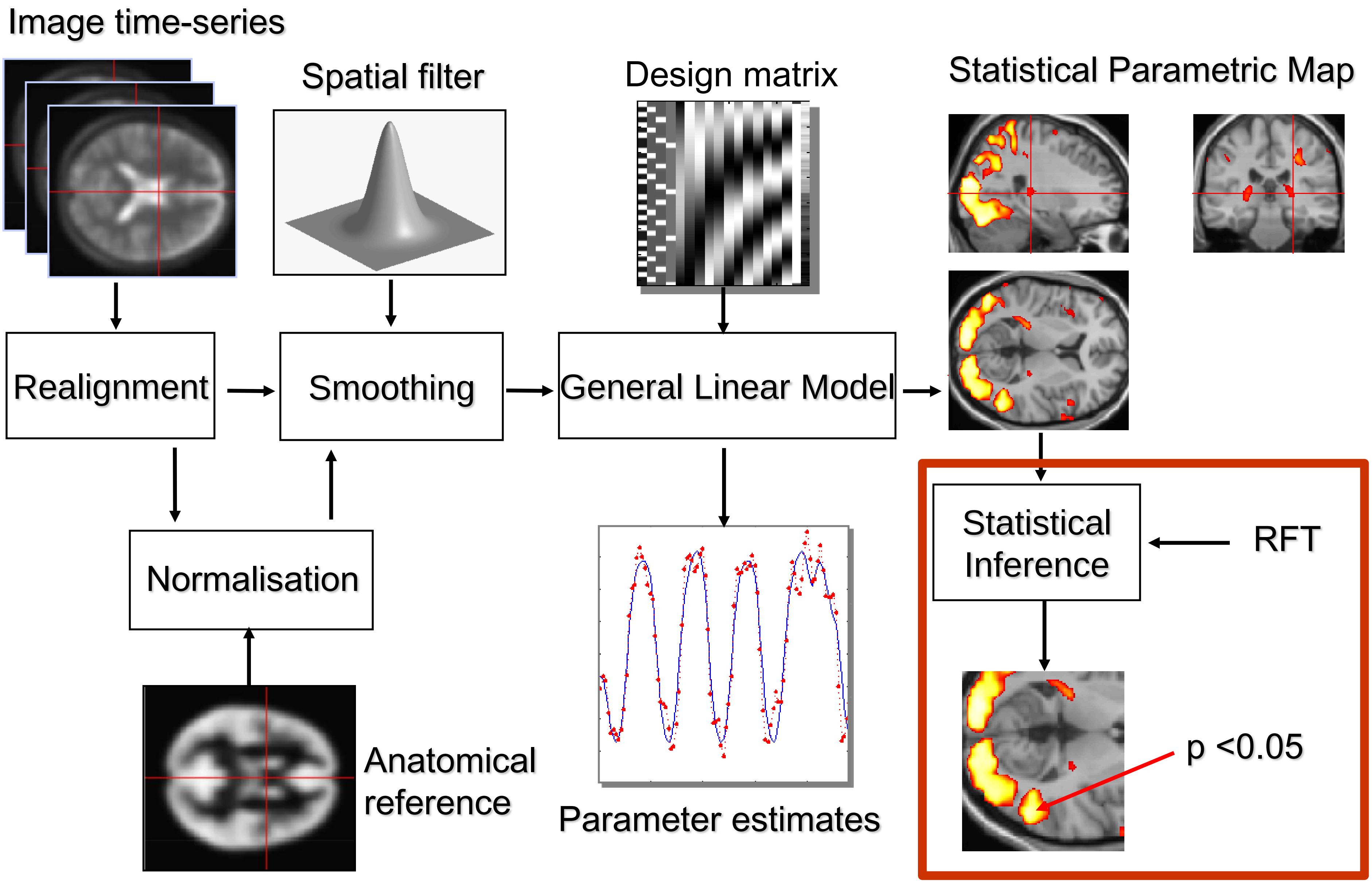


Image time-series

Spatial filter

Design matrix

Statistical Parametric Map

Realignment

Smoothing

General Linear Model

Normalisation

Anatomical reference

Parameter estimates

Statistical Inference

RFT

$p <0.05$