Introduction à la statistique médicale

Statistical Parametric Mapping short course

Course 3:

General Linear Model,
Contrast & Inference

Christophe Phillips, Ir PhD
GIGA – CRC *In Vivo* Imaging &
GIGA – *In Silico* Medicine

LIÈGE université
GIGA
CRC In vivo Imaging

fnrs
LA LIBERTÉ DE CHERCHER

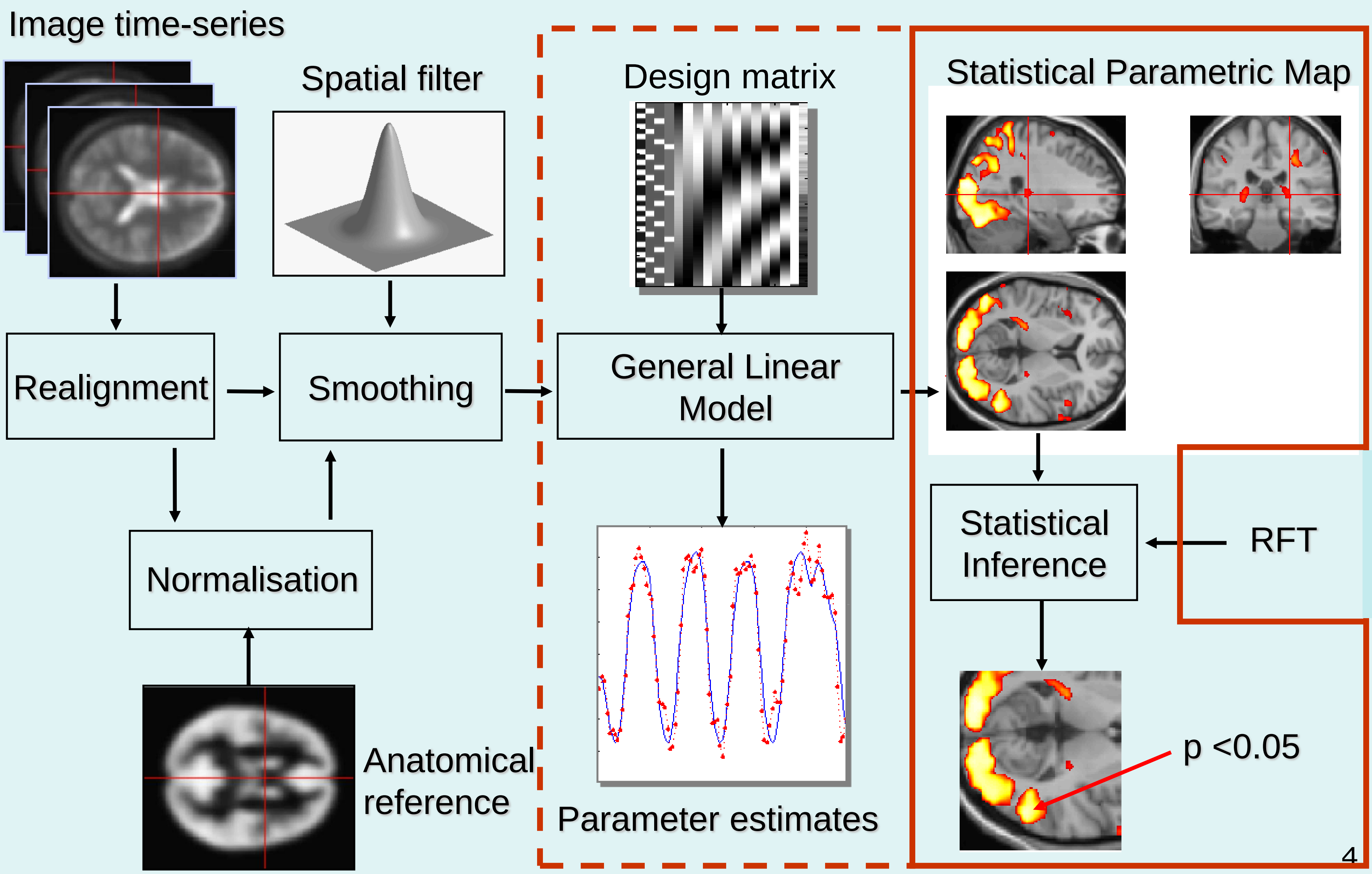
Content

- **Introduction**
- **Contrast & Inference**
- **Orthogonality issue**
- **Conclusion**

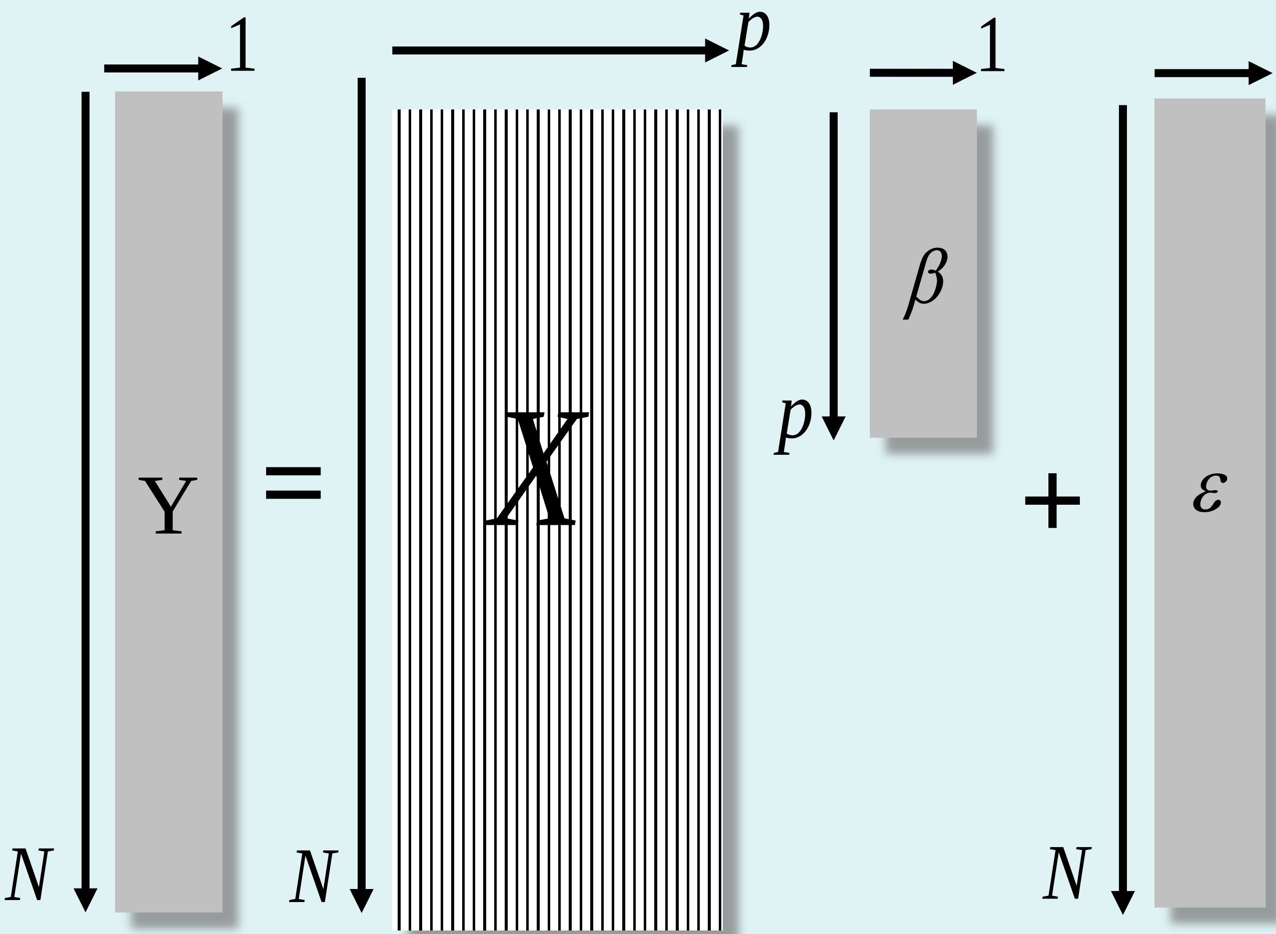
Content

- **Introduction**
  - **Generalized Linear Model**
  - **Estimated parameters**
- **Contrast & Inference**
- **Orthogonality issue**
- **Conclusion**

SPM work flow



General Linear Model



$$Y = X\beta + \varepsilon$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$$

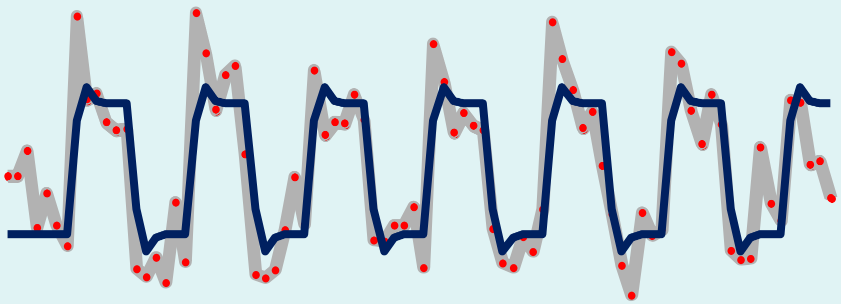
This is for a SINGLE voxel !

Design matrix **X** is the same for ALL voxels !

N: number of scans
p: number of regressors

Model is specified by
1. Design matrix **X**
2. Assumptions about **ε**

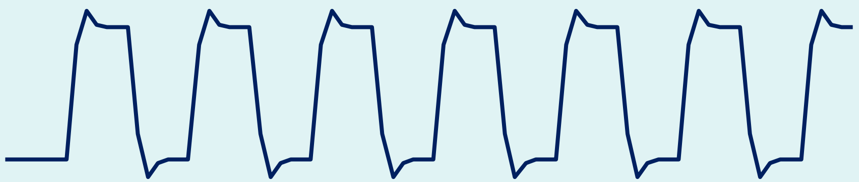
Estimation of the parameters



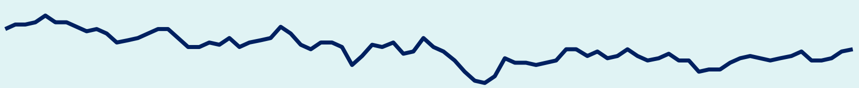
i.i.d. assumptions: $\varepsilon \sim N(0, \sigma^2 I)$

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

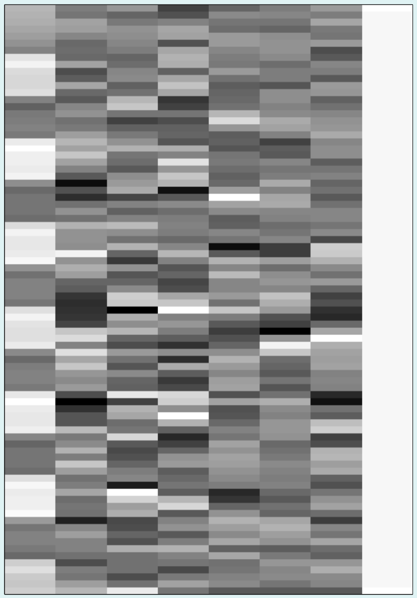
$\hat{\beta}_1 = 3.9831$

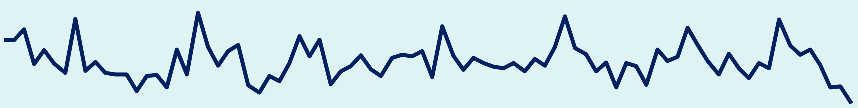


$\hat{\beta}_{2-7} = \{0.6871, 1.9598, 1.3902, 166.1007, 76.4770, -64.8189\}$



$\hat{\beta}_8 = 131.0040$

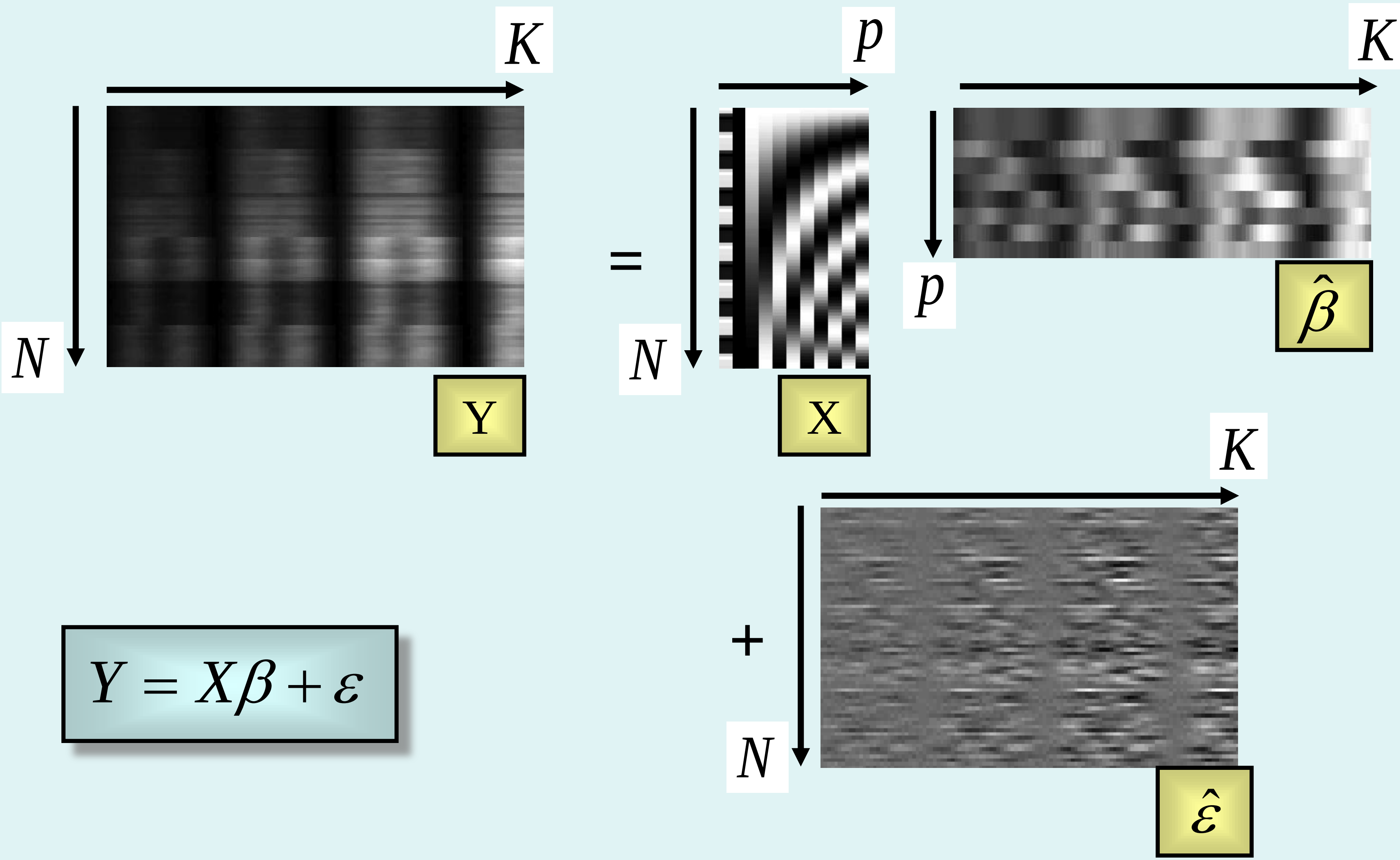
$y =$  β $+\varepsilon$

$\hat{\varepsilon} =$ 

$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$

$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N-p}$

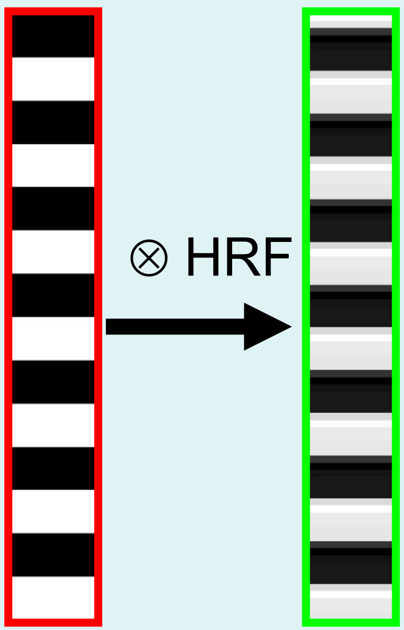
GLM & Mass univariate approach

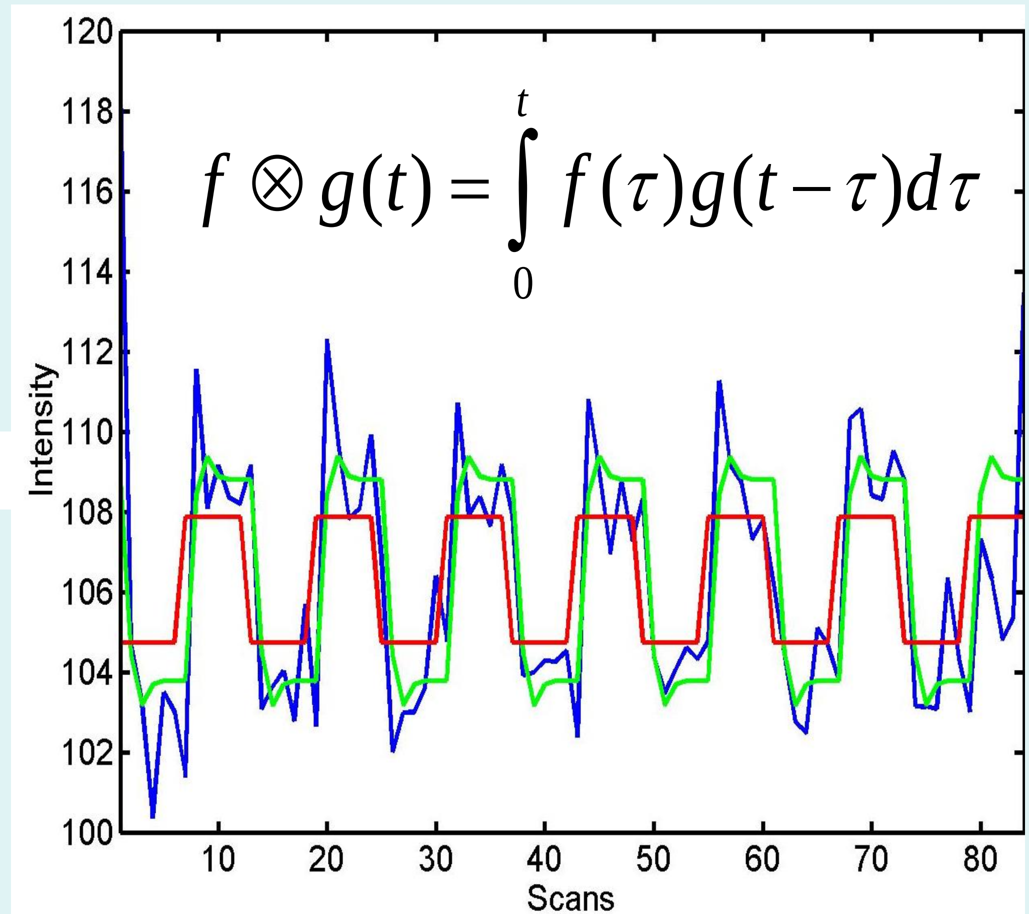


$$Y = X\beta + \varepsilon$$

Convolution model of the BOLD response

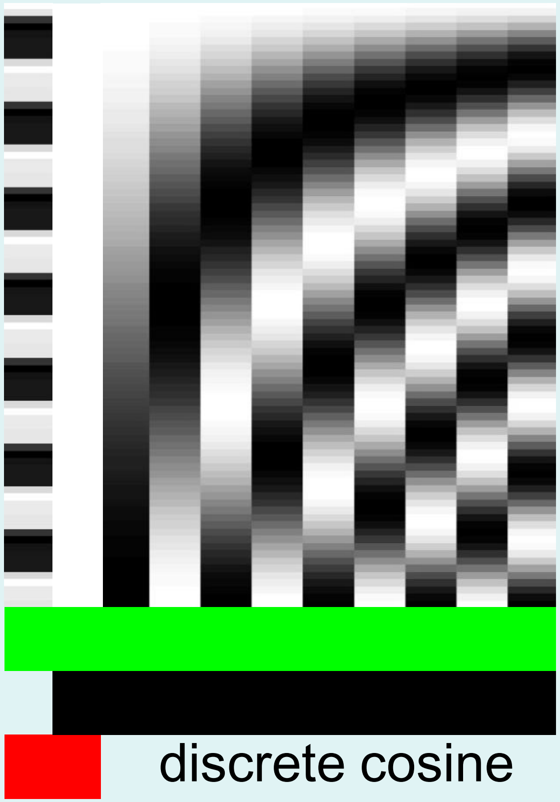
Convolve stimulus function with a canonical hemodynamic response function (HRF):



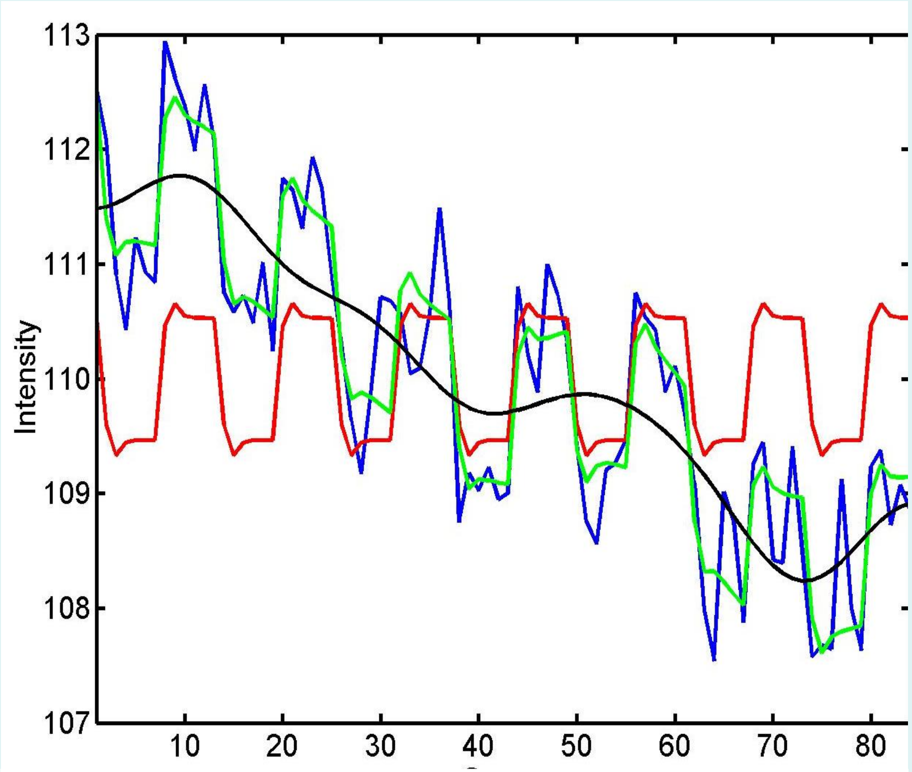


Low-frequency noise

Solution: High pass filtering



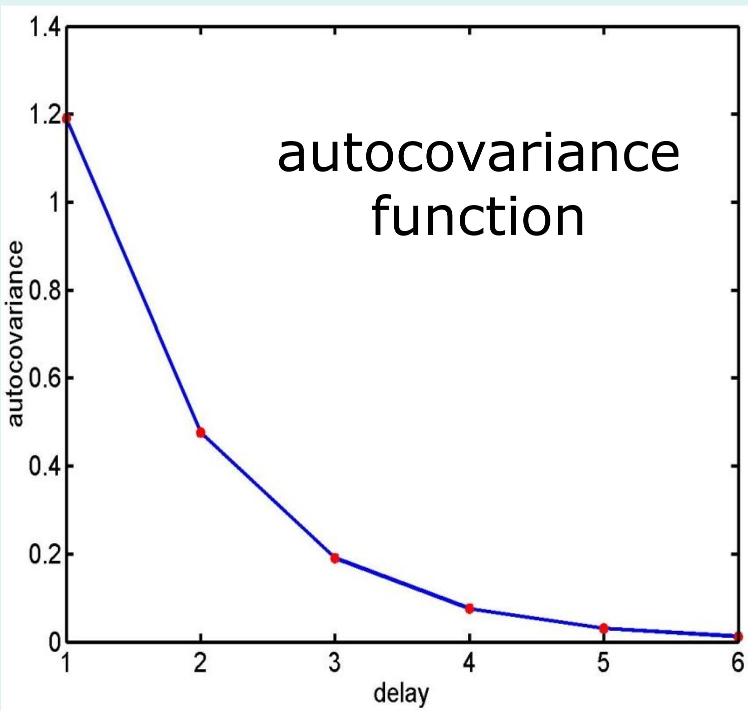
discrete cosine transform (DCT) set

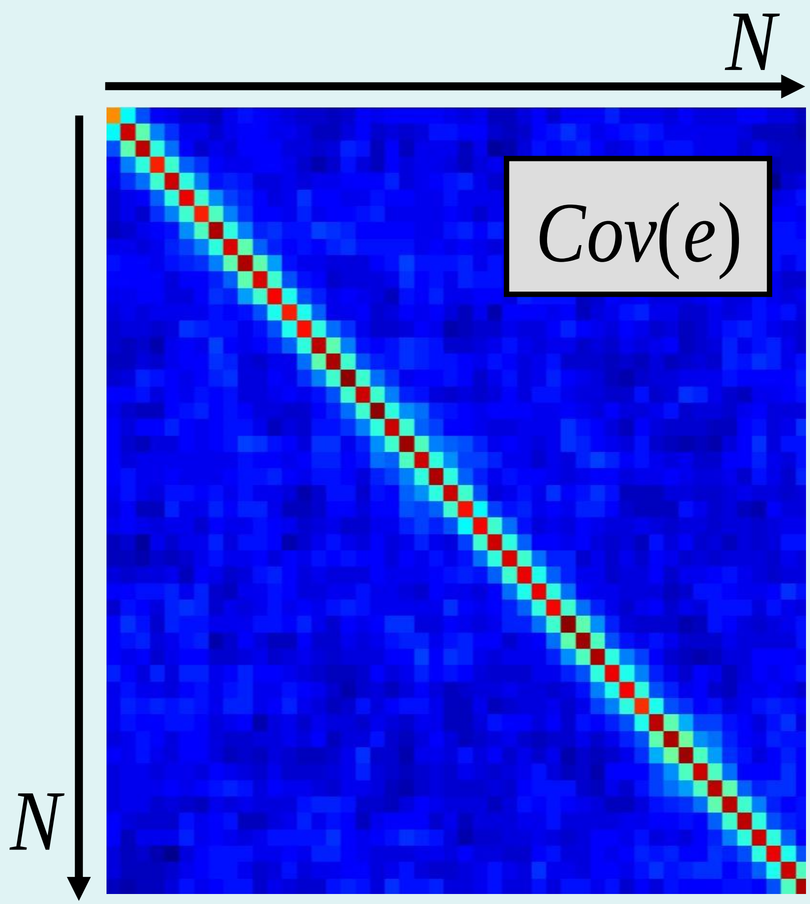


blue = data
black = mean + low-frequency drift
green = predicted response, taking into account low-frequency drift
red = predicted response, NOT taking into account low-frequency drift

3. Serial correlation

i.i.d: $\varepsilon \sim N(0, \sigma^2 I)$





Multiple covariance components

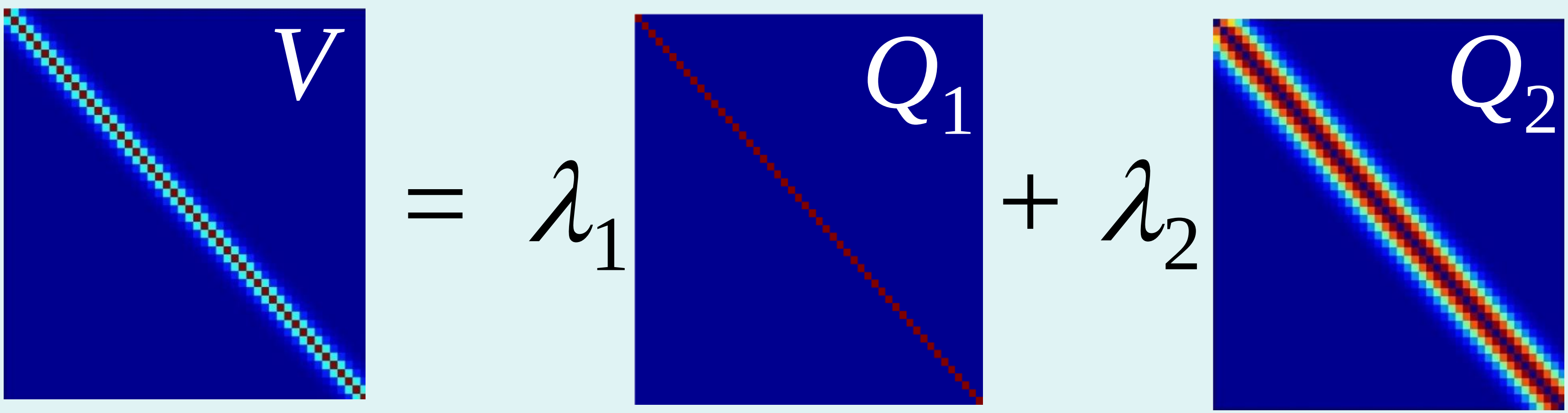
enhanced noise model at voxel *i*

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

$$C_i = \sigma_i^2 V$$

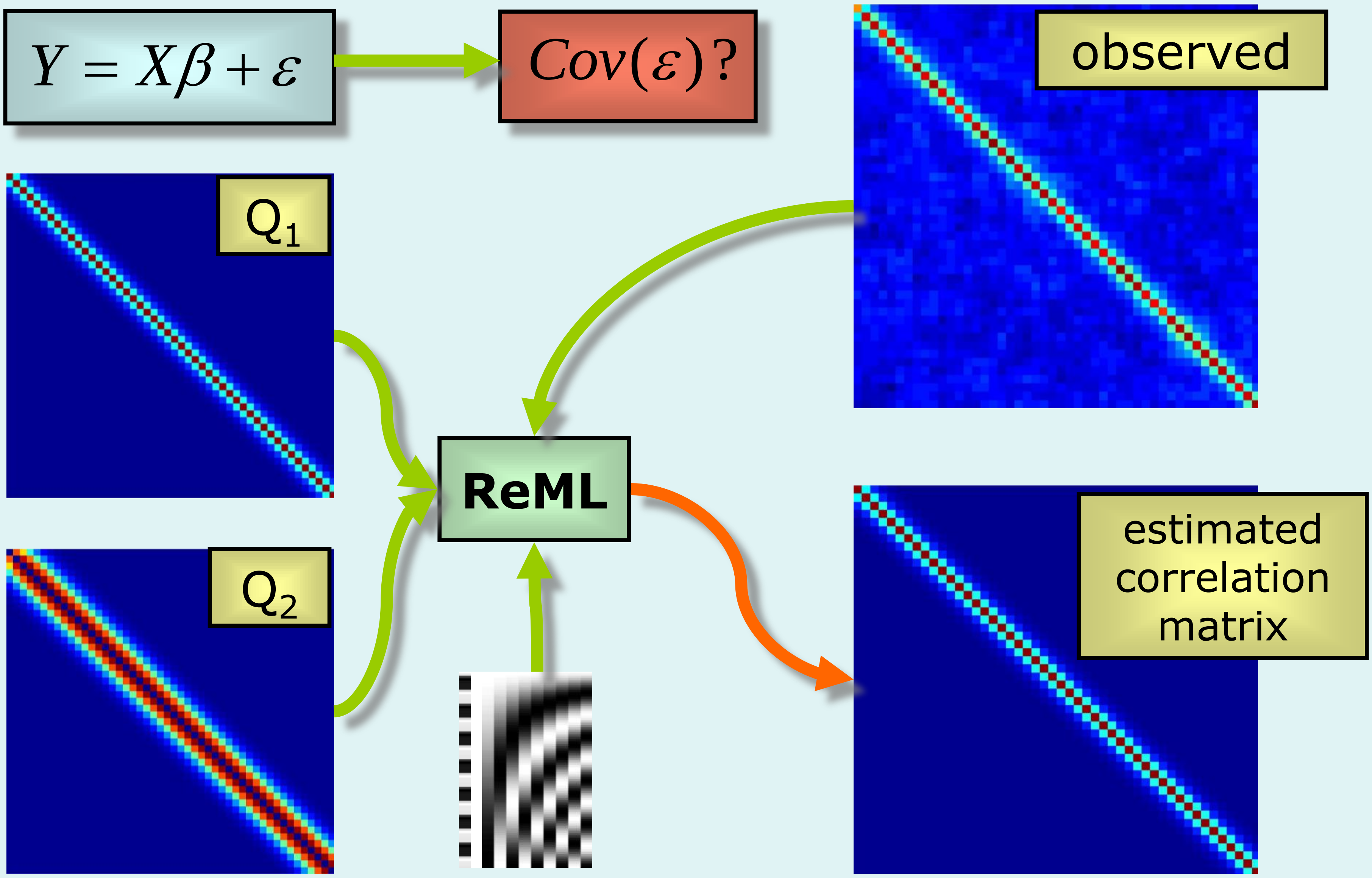
$$V = \sum \lambda_j Q_j$$

error covariance components Q and hyperparameters λ



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

Restricted Maximul Likelihood



Estimation in SPM

$$y_j = X\theta_j + \varepsilon_j$$

$$\hat{\theta}_{j,OLS} = X^+ y_j$$

Ordinary least-squares

$$\hat{C}_\varepsilon = C\hat{o}v(\varepsilon) = \text{ReML}\left(\sum_{voxel\ j} y_j y_j^T, X, Q\right)$$

ReML (pooled estimate)

$$\hat{\theta}_{j,ML} = (X^T V^{-1} X)^{-1} X^T V^{-1} y_j$$

Maximum Likelihood

- 2 passes (first pass for selection of voxels)
- more accurate estimate of V

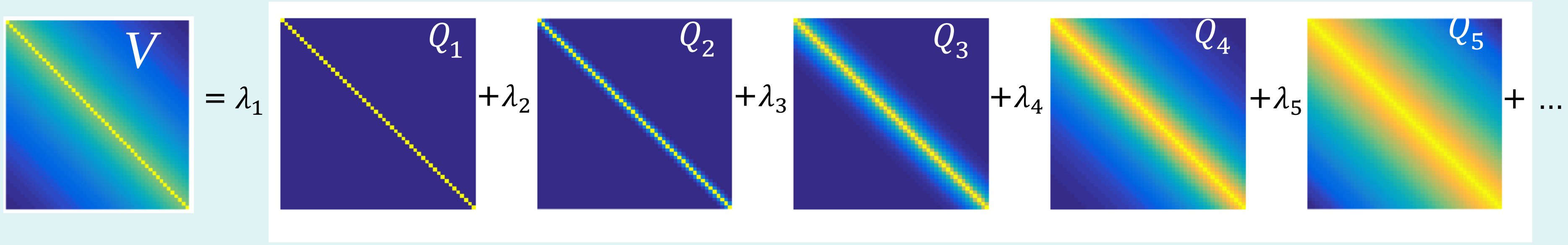
Assume, at voxel j: $C_{\varepsilon,j} = \sigma_j V$

$$t = \frac{c^T\theta}{\text{SE}(c^T\theta)}$$

$$\text{SE}(c^T\theta) = \sqrt{\hat{\sigma}^2 c^T (V^{-1/2}X)^- (V^{-1/2}X)^{-T} c}$$

Limitations

The AR(1)+white noise model may not be enough for short TR (<1.5 s)



The flexibility of the ReML enables the use of any number of components of any shape

Content

- **Introduction**
- **Contrast & Inference**
  - **Hypothesis testing**
  - **Contrast**
  - **t-Test**
  - **F-test**
- **Orthogonality issue**
- **Conclusion**

Hypothesis testing

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

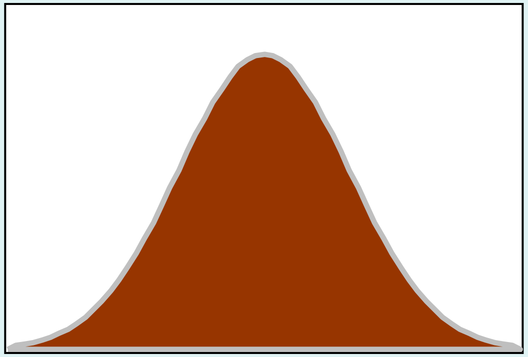
- **Null Hypothesis H_0**

  Typically what we want to disprove (no effect).
  → *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.



Null Distribution of T

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

Hypothesis testing & inference

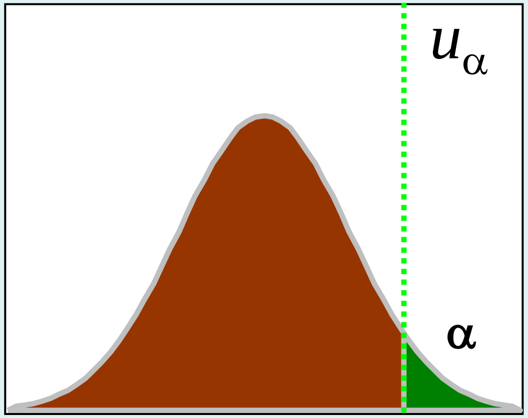
- ***Significance level α:***

  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:**

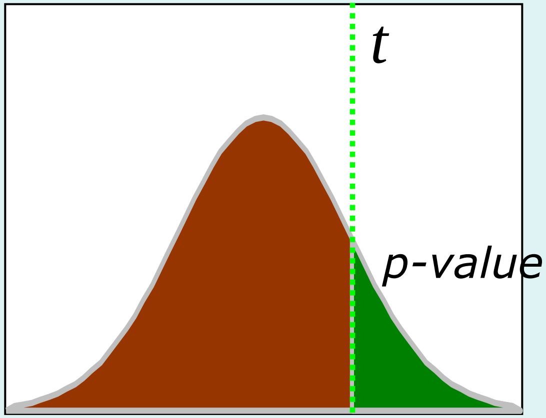
  We reject the null hypothesis in favour of the alternative hypothesis if $t > u_\alpha$

- ***p-value*:**

  A *p-value* summarises evidence against H_0.

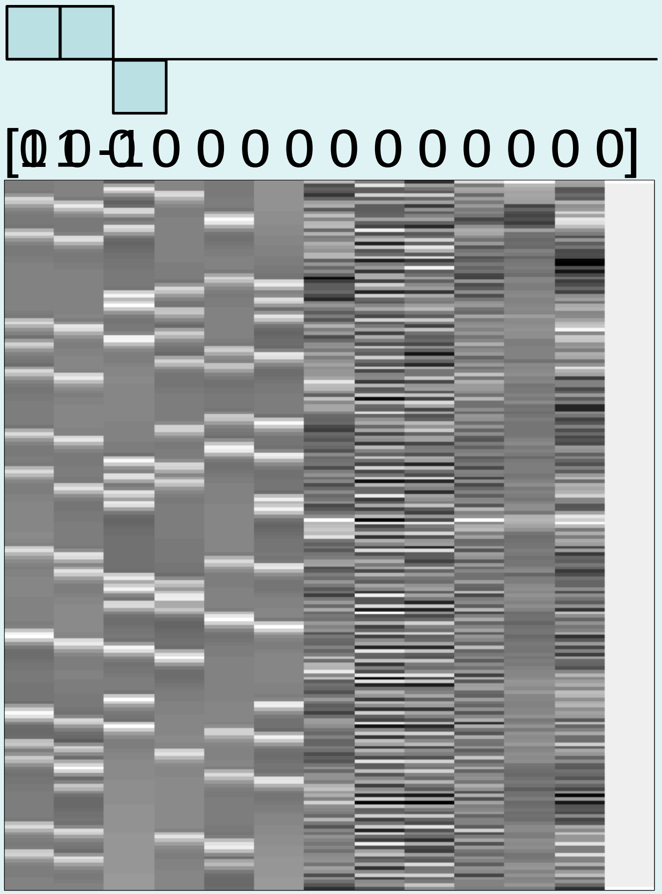
  This is the chance of observing value more extreme than t under the null hypothesis.

$$p(T > t \mid H_0)$$



Null Distribution of T

Contrast & effect of interest



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.

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

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

$$= \boldsymbol{\beta_1}$$

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

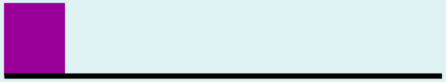
$$c^T\beta = \mathbf{0} \times \beta_1 + \mathbf{1} \times \beta_2 + \mathbf{-1} \times \beta_3 + \mathbf{0} \times \beta_4 + \cdots$$

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

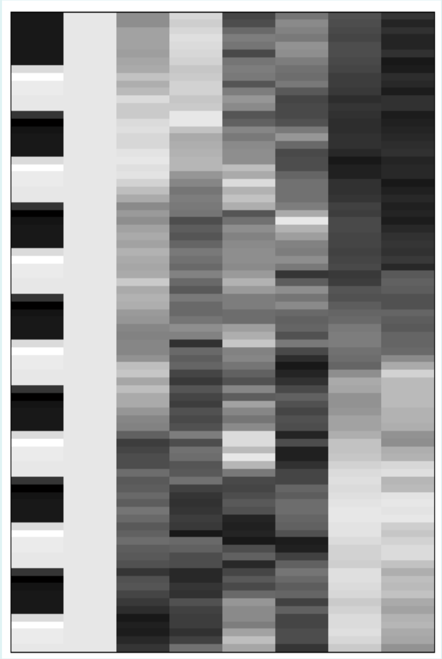
$$c^T\hat{\beta} \sim N(c^T\beta, \sigma^2 c^T (X^T X)^{-1} c)$$

t-Test, one dimensional contrast

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



$\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\ \ldots$



Question: box-car amplitude > 0 ?

=

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

Null hypothesis: $H_0\!: c^T\beta = 0$

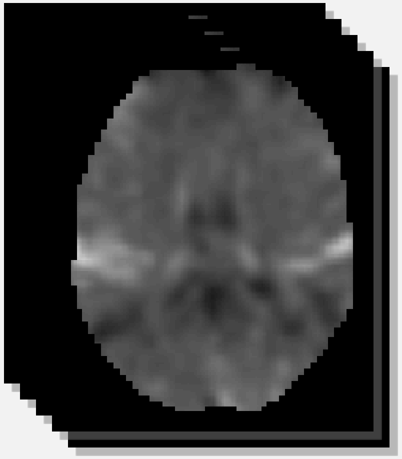
Test statistic:

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

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

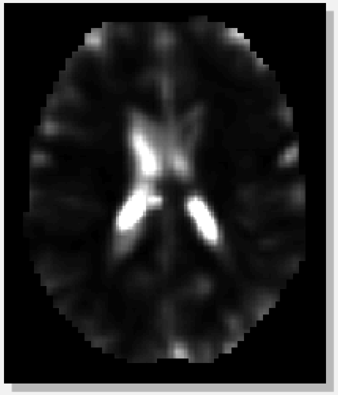
t-Test in SPM

For a given contrast *c*:



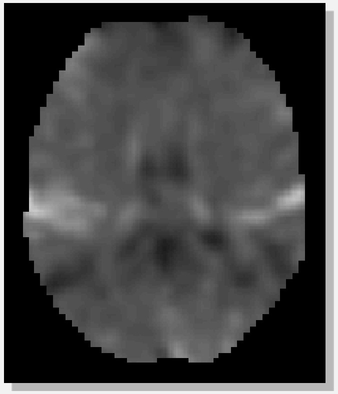
beta_???? images

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



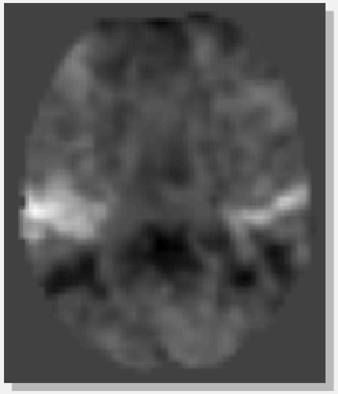
ResMS image

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N - p}$$



con_???? image

$$c^T \hat{\beta}$$



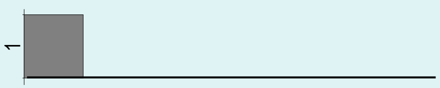
spmT_???? image

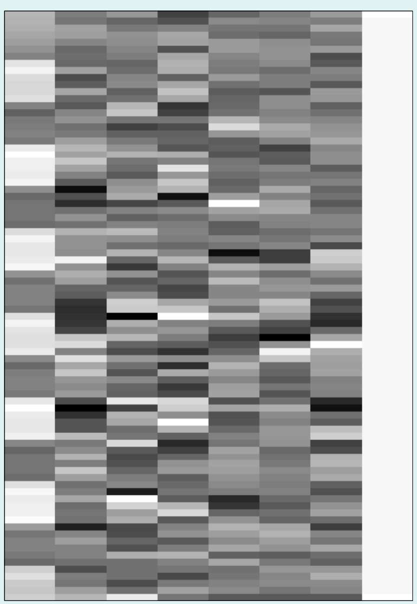
SPM{t}

t-Test, simple example

Passive word listening vs. rest

c^T = [1 0 0 0 0 0 0 0]

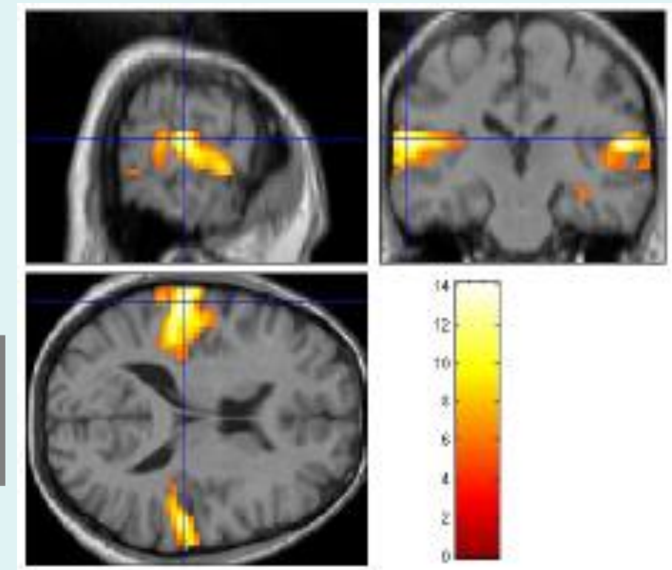




Q: activation during listening ?

Null hypothesis: $\beta_1 = 0$

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



SPMresults:

Height threshold T = 3.2057 {p<0.001}

voxel-level					
T	$(Z_{\equiv})$	$p_{\text{uncorrected}}$	mm	mm	mm
13.94	**Inf**	**0.000**	**-63**	**-27**	**15**
12.04	Inf	0.000	-48	-33	12
11.82	Inf	0.000	-66	-21	6
13.72	**Inf**	**0.000**	**57**	**-21**	**12**
12.29	Inf	0.000	63	-12	-3
9.89	7.83	0.000	57	-39	6
7.39	**6.36**	**0.000**	**36**	**-30**	**-15**
6.84	**5.99**	**0.000**	**51**	**0**	**48**
6.36	**5.65**	**0.000**	**-63**	**-54**	**-3**
6.19	**5.53**	**0.000**	**-30**	**-33**	**-18**
5.96	**5.36**	**0.000**	**36**	**-27**	**9**
5.84	**5.27**	**0.000**	**-45**	**42**	**9**
5.44	**4.97**	**0.000**	**48**	**27**	**24**
5.32	**4.87**	**0.000**	**36**	**-27**	**42**

t-Test, summary

- *T*-test is a signal-to-noise measure (ratio of estimate to standard deviation of estimate).

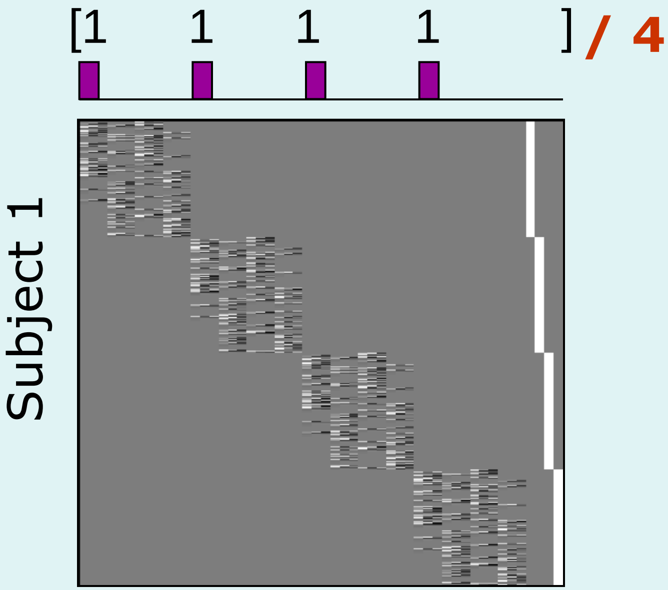
- Alternative hypothesis:

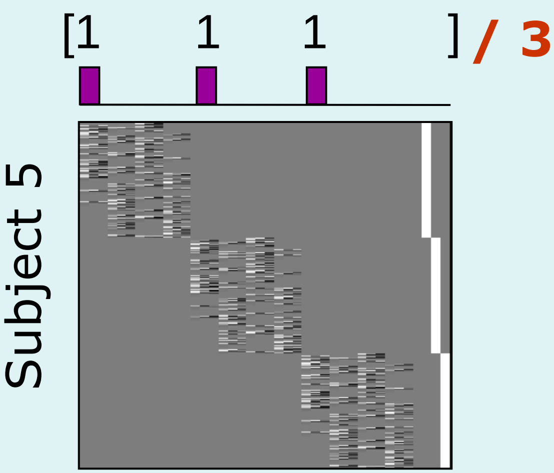
$$H_0:\ c^T\beta = 0 \quad \text{vs} \quad H_A: c^T\beta > 0$$

- *T*-contrasts are simple combinations of the betas

- *T*-statistic *does not depend* on the scaling of the regressors or the scaling of the contrast.

t-Test, scaling issue





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

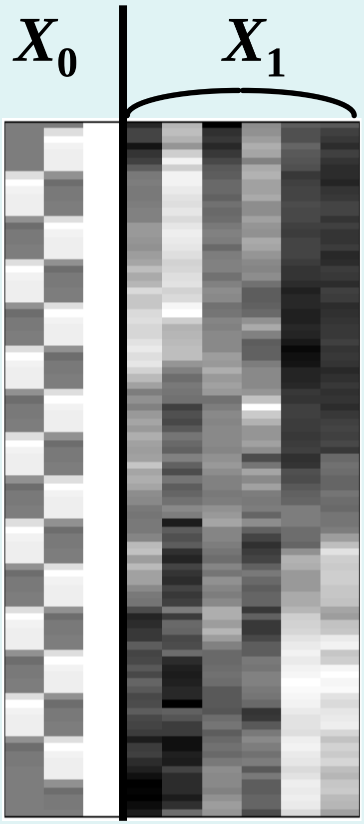
- The *T*-statistic does not depend on the scaling of the regressors neither of the contrast.
- Contrast $c^T \hat{\beta}$ does depend on scaling.
- Be careful of the interpretation of the contrasts $c^T \hat{\beta}$ themselves (e.g., for a second level analysis):

sum ≠ average

F-test, extra-sum-of-squares principles

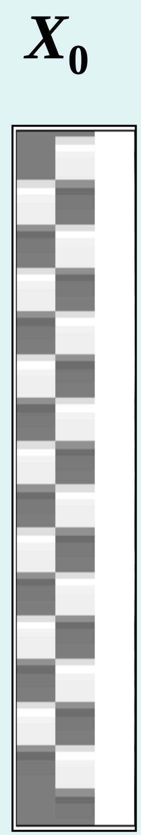
Model comparison:

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



Full model ?

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



or Reduced model?

→ $\mathbf{RSS_0}$ $\sum \hat{\varepsilon}^2_{reduced}$

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

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

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

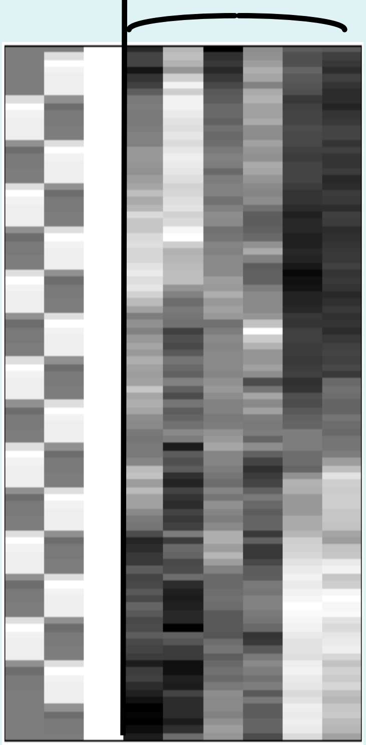
ν_1 = rank(X) – rank(X_0)
ν_2 = N – rank(X)

F-test, multidimensional contrast

Tests *multiple* linear hypotheses:

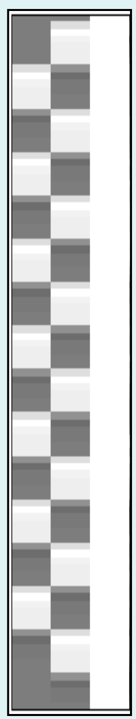
H_0: True model is X_0	**H_0: $\beta_4 = \beta_5 = \ldots = \beta_9 = 0$**	**test H_0 : $c^T\beta = 0$?**

X_0 X_1 ($\beta_{4\text{-}9}$)



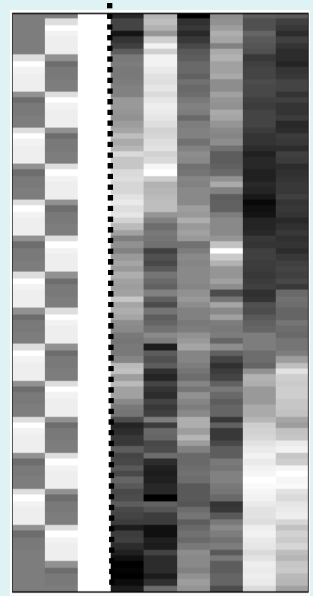
Full model?

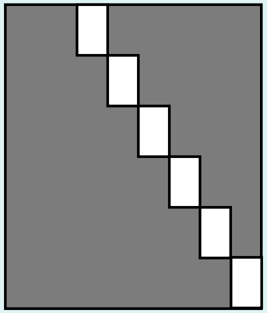
X_0

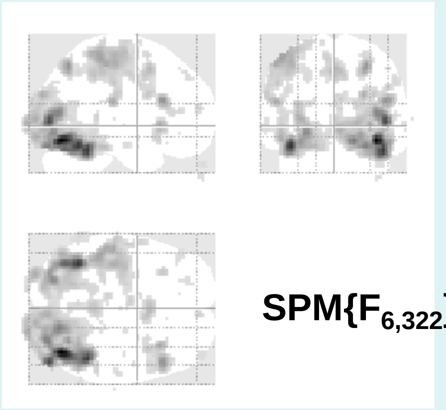


Reduced model?

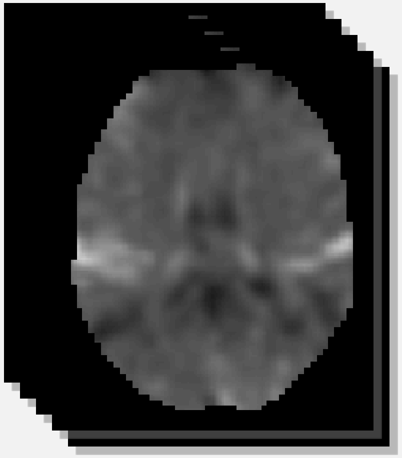
$$c^T = \left[\begin{array}{ccc|cccccc} 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&1 \end{array}\right]$$





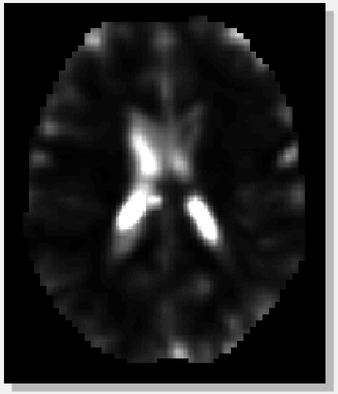


F-contrast in SPM



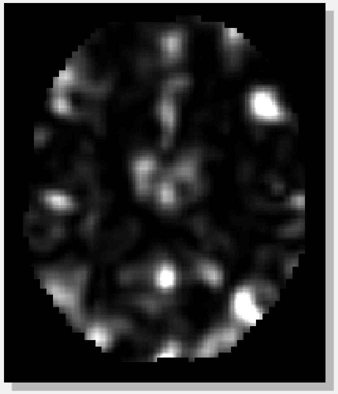
beta_???? images

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



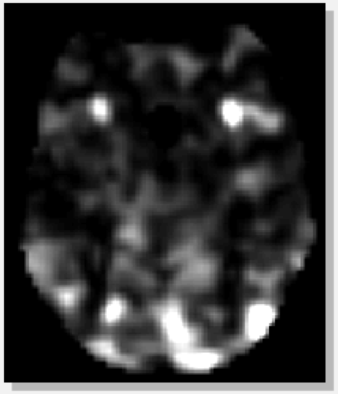
ResMS image

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N - p}$$



ess_???? images

$$(RSS_0 \text{ - } RSS)$$

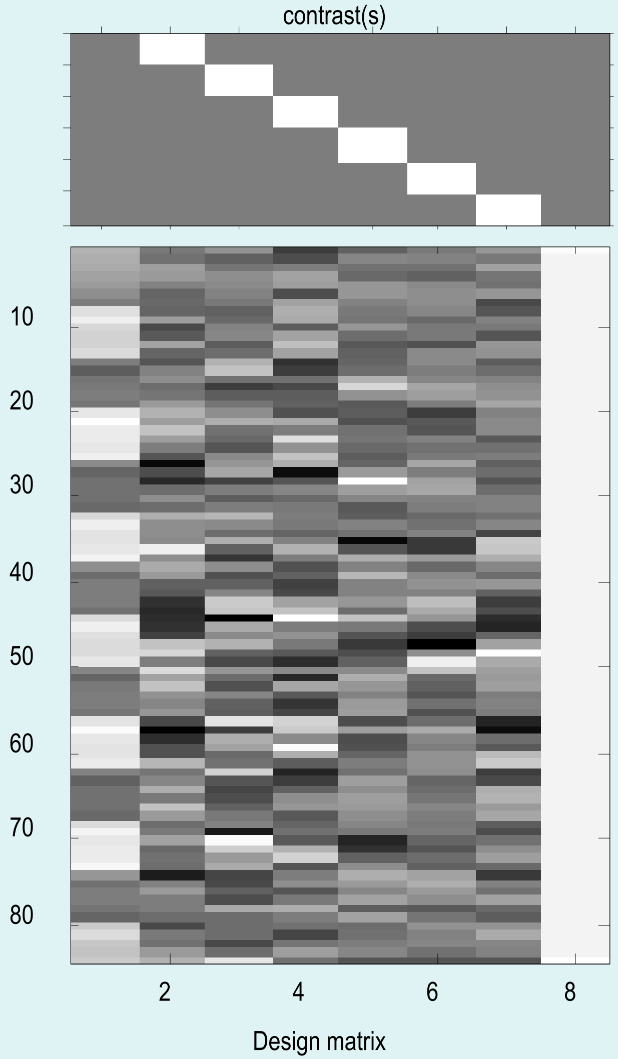


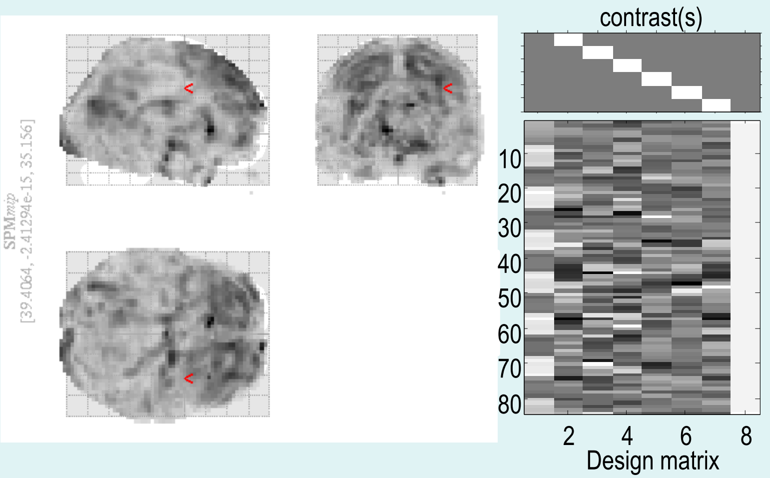
spmF_???? images

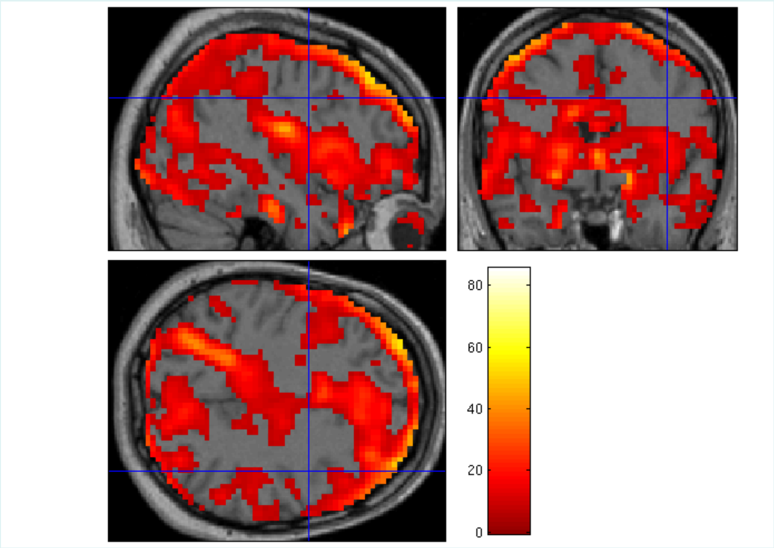
SPM{F}

F-test, example

Movement related effects







F-test summary

- F-tests can be viewed as testing for the additional variance explained by a larger model w.r.t. a simpler (*nested*) model → **model comparison**.
- F-tests a weighted ***sum of squares*** of one or several combinations of the coefficients β.
- In practice, noneed to explicitly separate X into $[X_1\ X_2]$ thanks to multidimensional contrasts.
- Hypotheses:

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

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

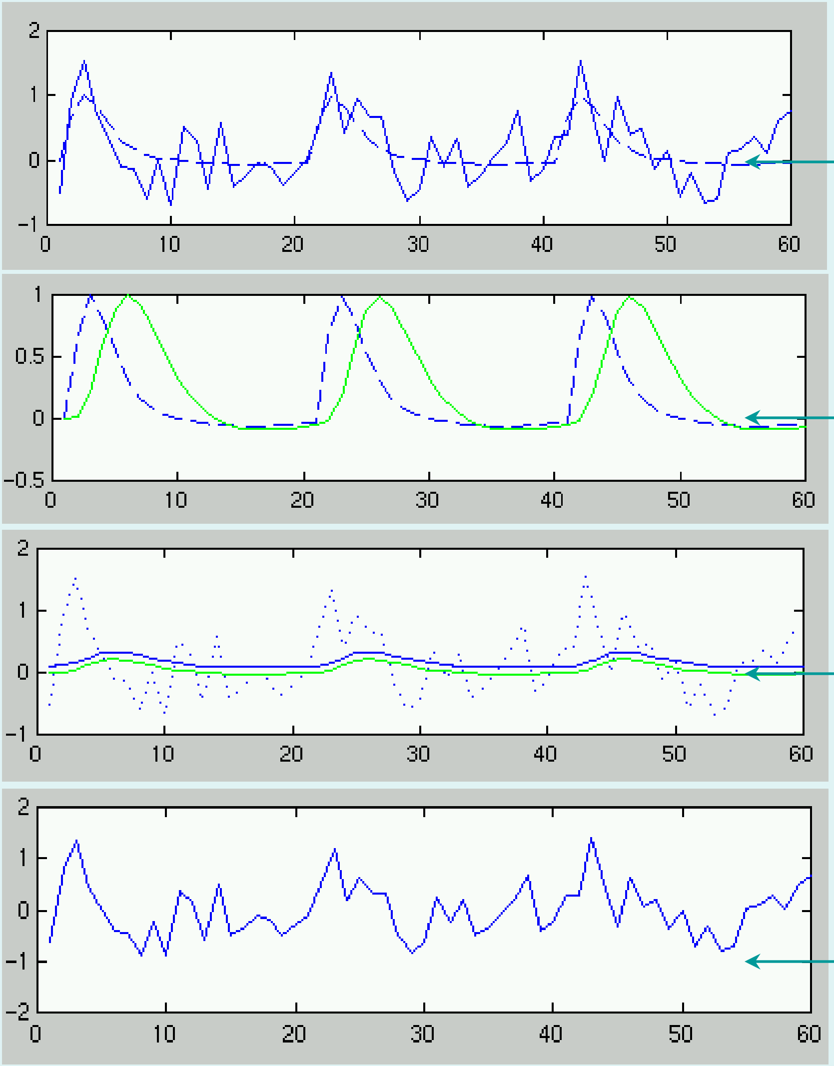
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- In 1D contrast with an F-test, testing $\beta_1 - \beta_2$ is the same as testing $\beta_2 - \beta_1$.

Content

- **Introduction**
- **Contrast & Inference**
- **Orthogonality issue**
- **Conclusion**

A bad model



True signal (---) and observed signal

Model (green, peak at 6sec) and TRUE signal (blue, peak at 3sec)

Fitting :

b1 = 0.2, mean = .11

Noise

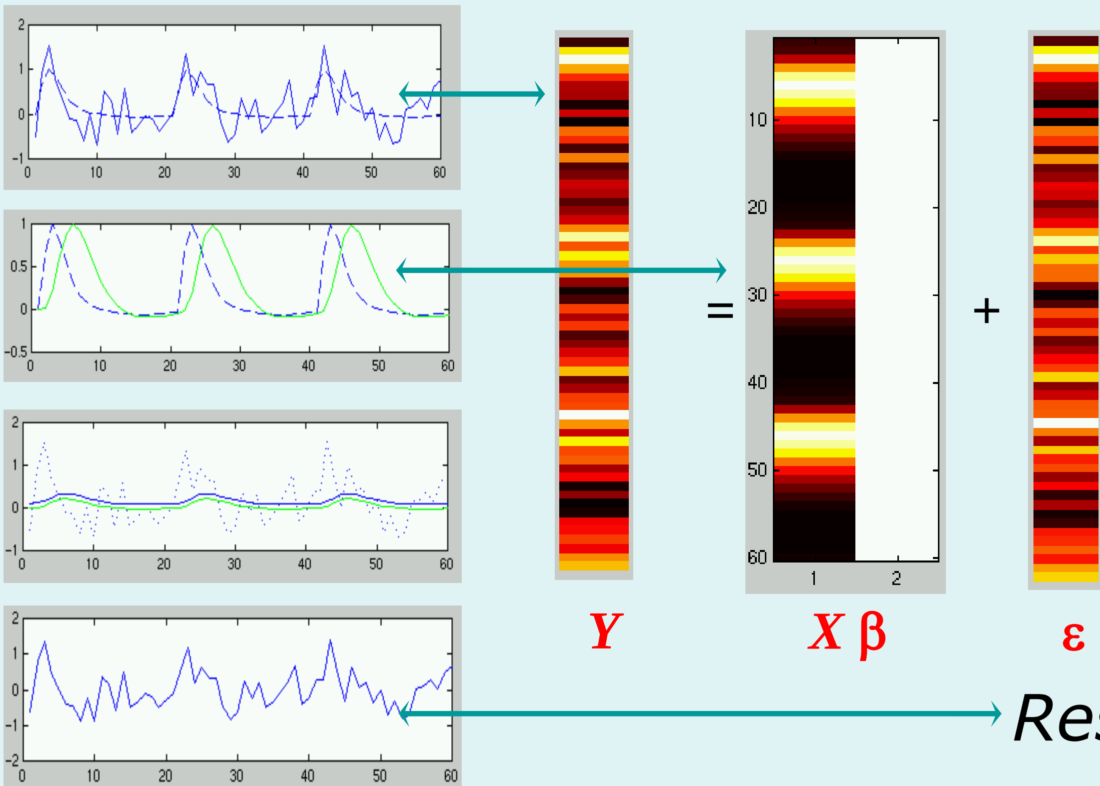
(still contains some signal)

⇒ Test for the green regressor not significant

A bad model

$b_1 = 0.22$

$b_2 = 0.11$

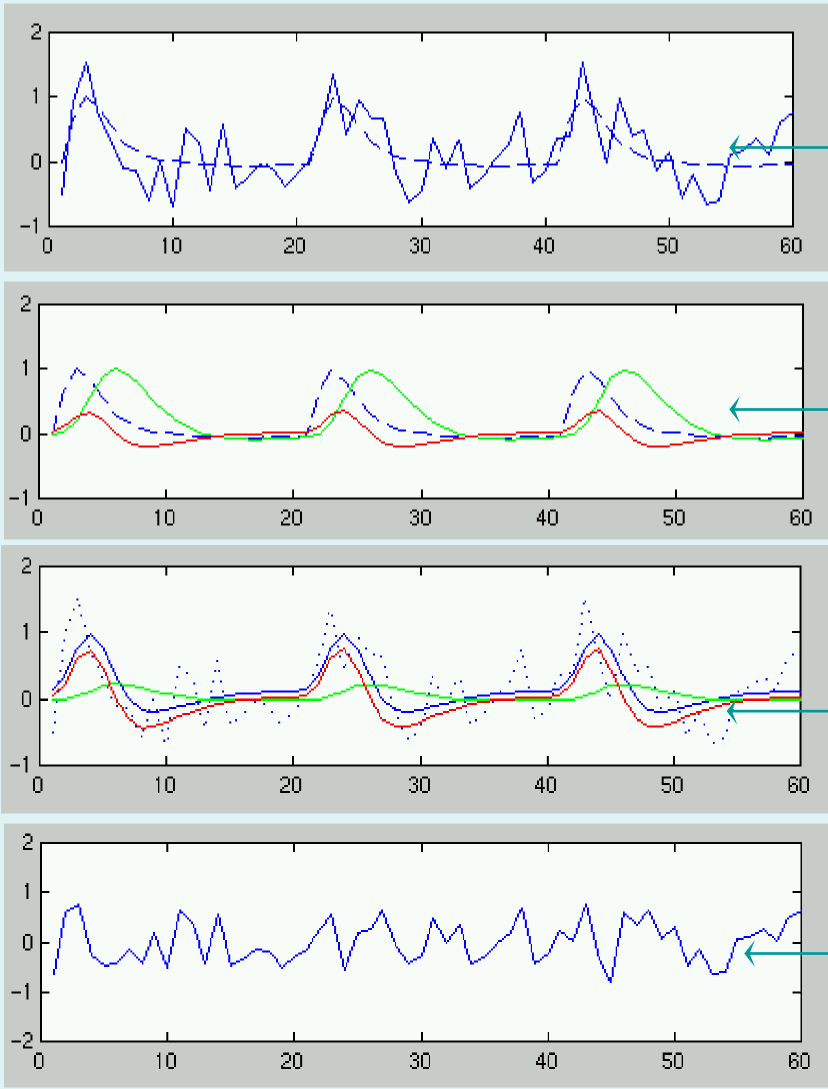


P(b1 = 0) = 0.1
(t-test b1>0)

P(b1 = 0) = 0.2
(F-test b1≠0)

Residual Variance = 0.3

A better model...



True signal + observed signal

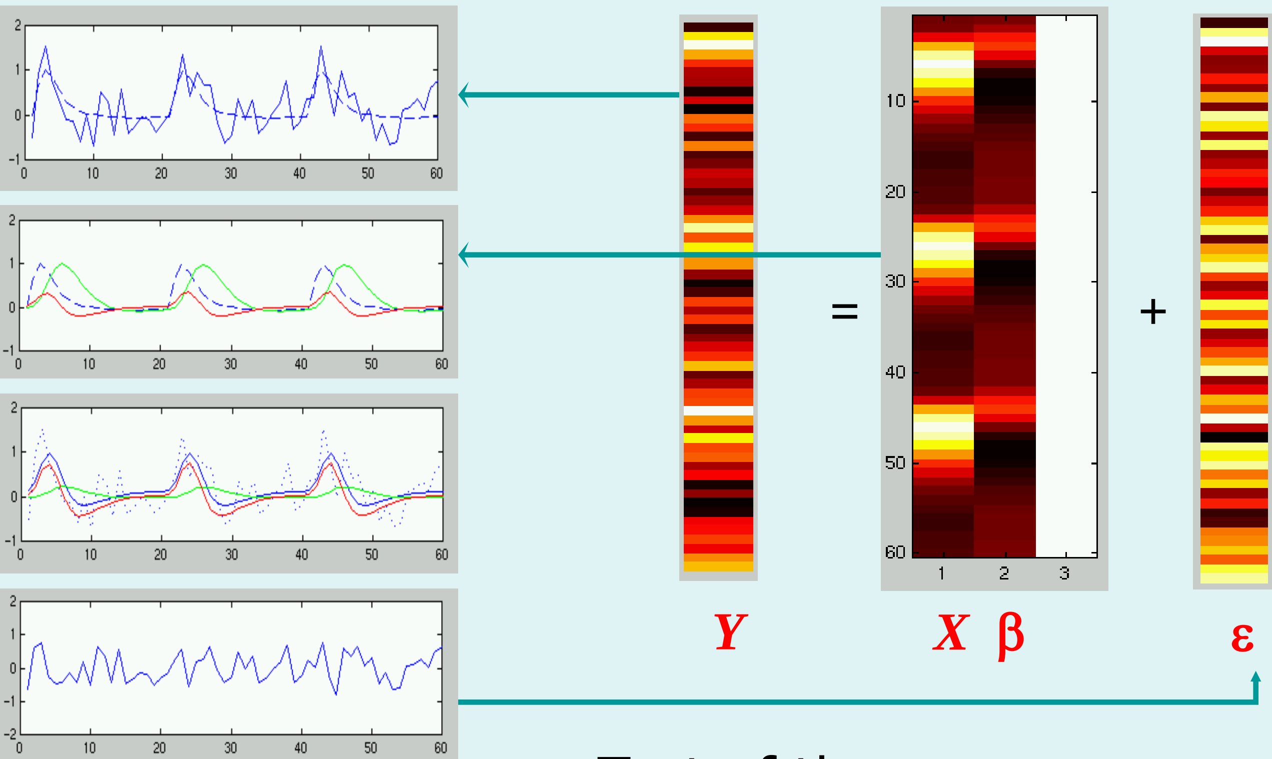
Model (green and red)
and true signal (blue ---)
Red regressor : temporal derivative of the green regressor

Global fit (blue)
and partial fit (green & red)
Adjusted and fitted signal

Noise (a smaller variance)

A better model

$b_1 = 0.22$
$b_2 = 2.15$
$b_3 = 0.11$



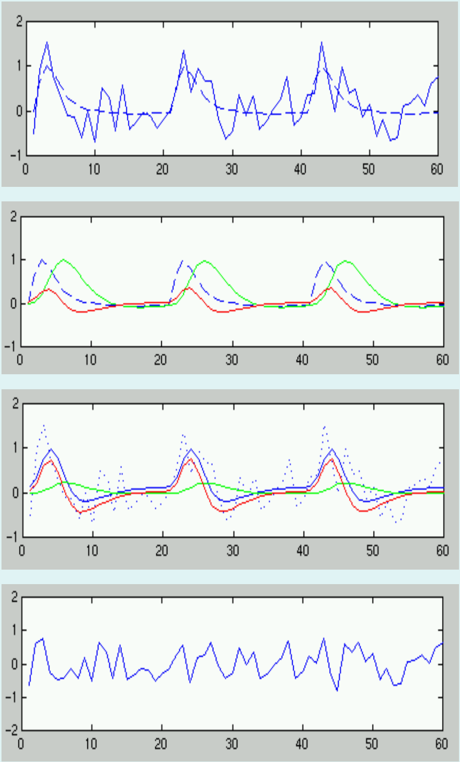
Residual Var = 0.2

P(b1 > 0) = 0.07
(t test b1>0)

P([b1 b2] = [0 0]) = 0.000001
(F test [b1 b2]≠[0 0])

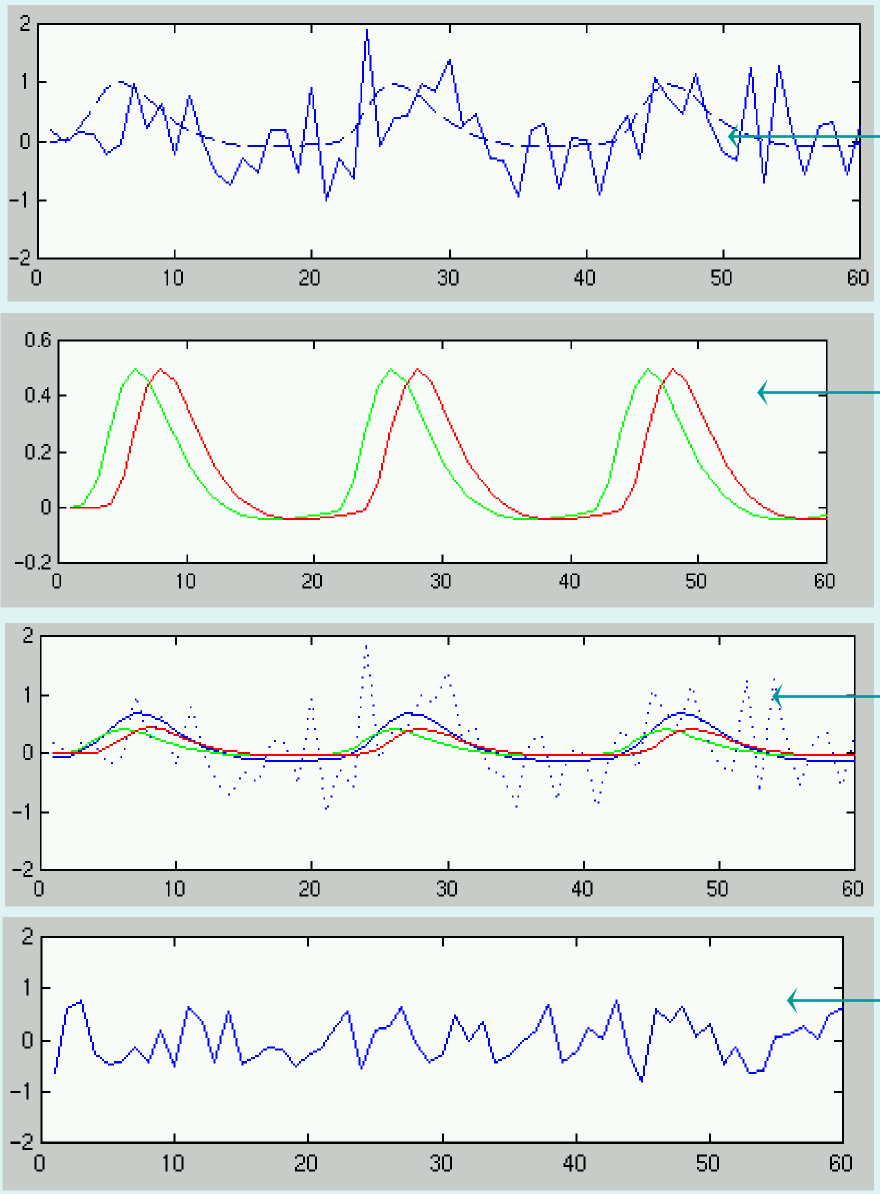
⇒ Test of the green regressor almost significant
⇒ Test F very significant
⇒ Test of the red regressor very significant

Summary



- *The residuals should be looked at ...(non random structure ?)*
- *We rather test flexible models if there is little a priori information, and precise ones with a lot a priori information*
- *In general, use the F-tests to look for an overall effect, then look at the betas or the adjusted signal to characterise the origin of the signal*
- *Interpreting the test on a single parameter (one function) can be very confusing: cf. the delay or magnitude situation*

Correlation between regressors



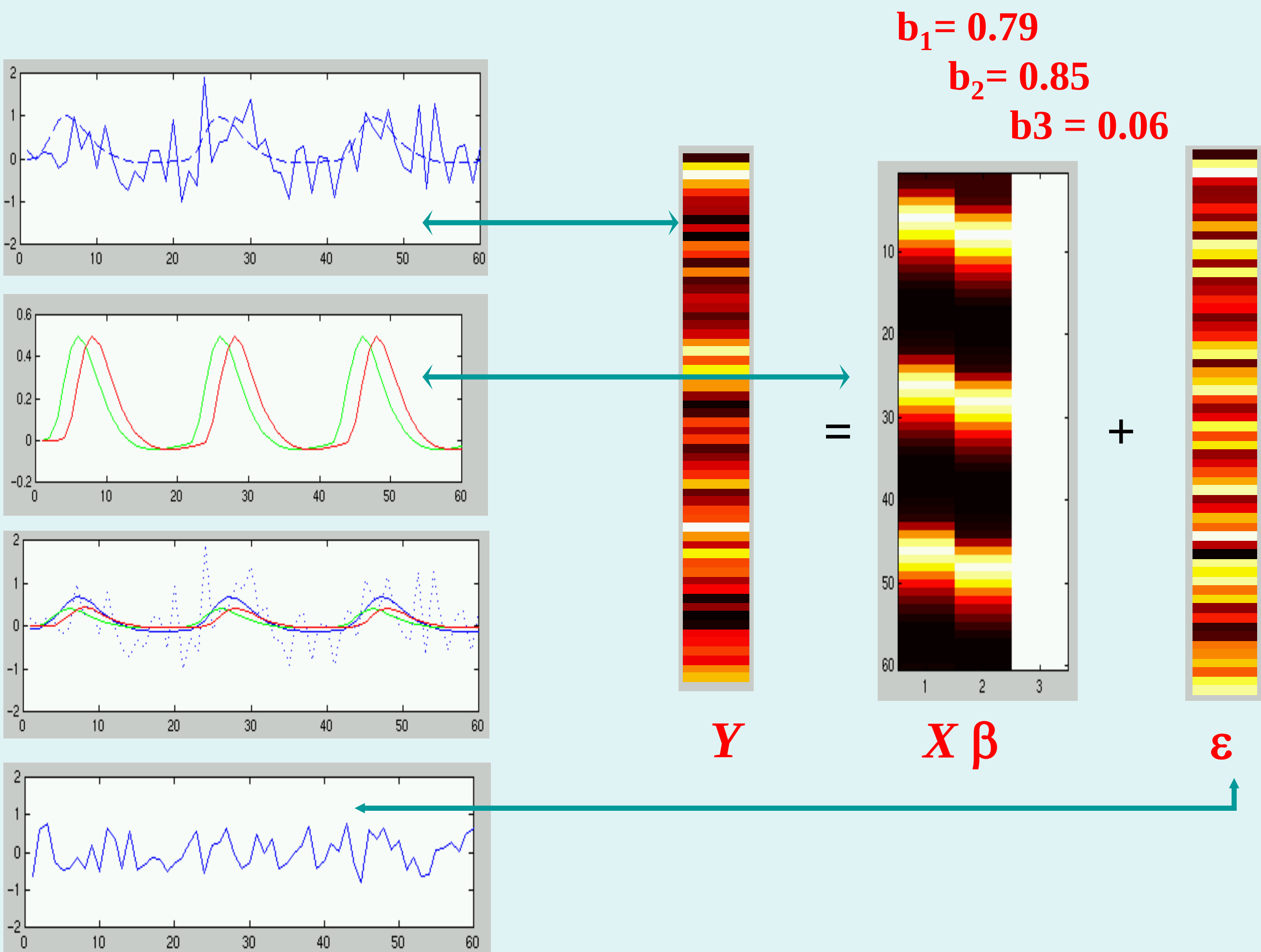
True signal

Model (green and red)

Fitting (blue : global fit)

Noise

Correlation between regressors



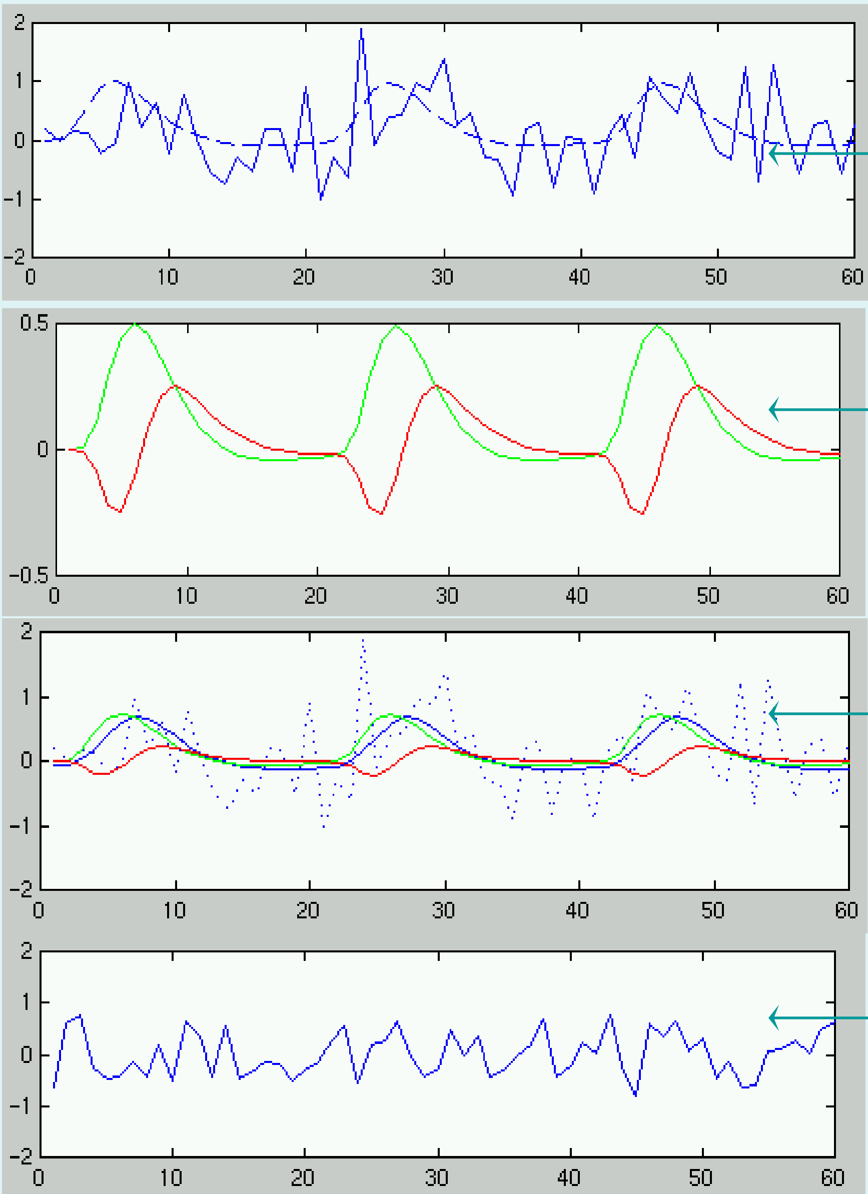
Residual var. = 0.2

P(b1 = 0) = 0.08
(t test b1>0)

P(b2 = 0) = 0.07
(t test b2>0)

P([b1 b2] = 0) = 0.002
(F test [b1 b2] ≠ 0)

Decorrelated regressors



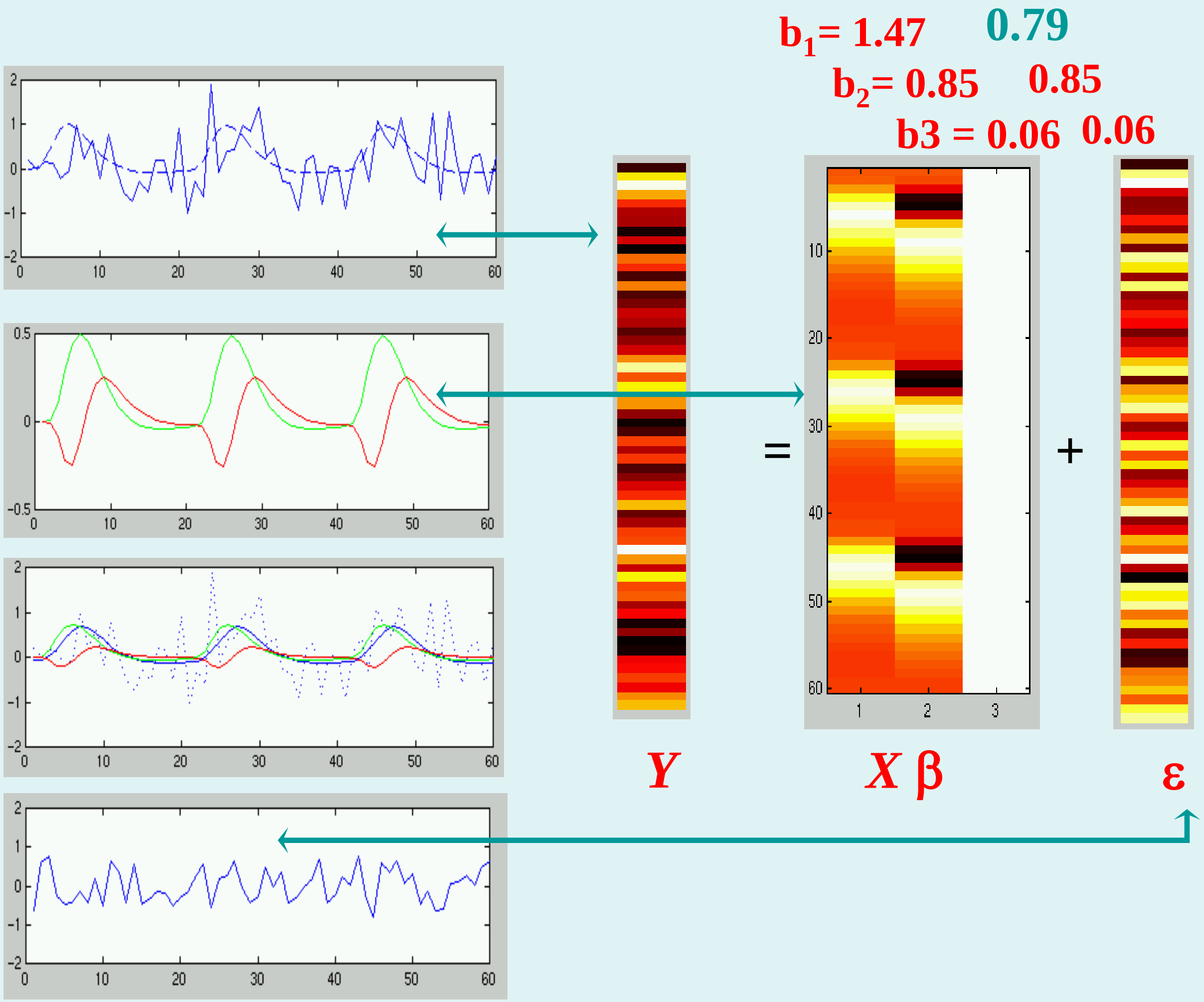
true signal

Model : red regressor orthogonalised with respect to the green one = remove every thing that can correlate with the green regressor

Fit

Noise

Decorrelated regressors



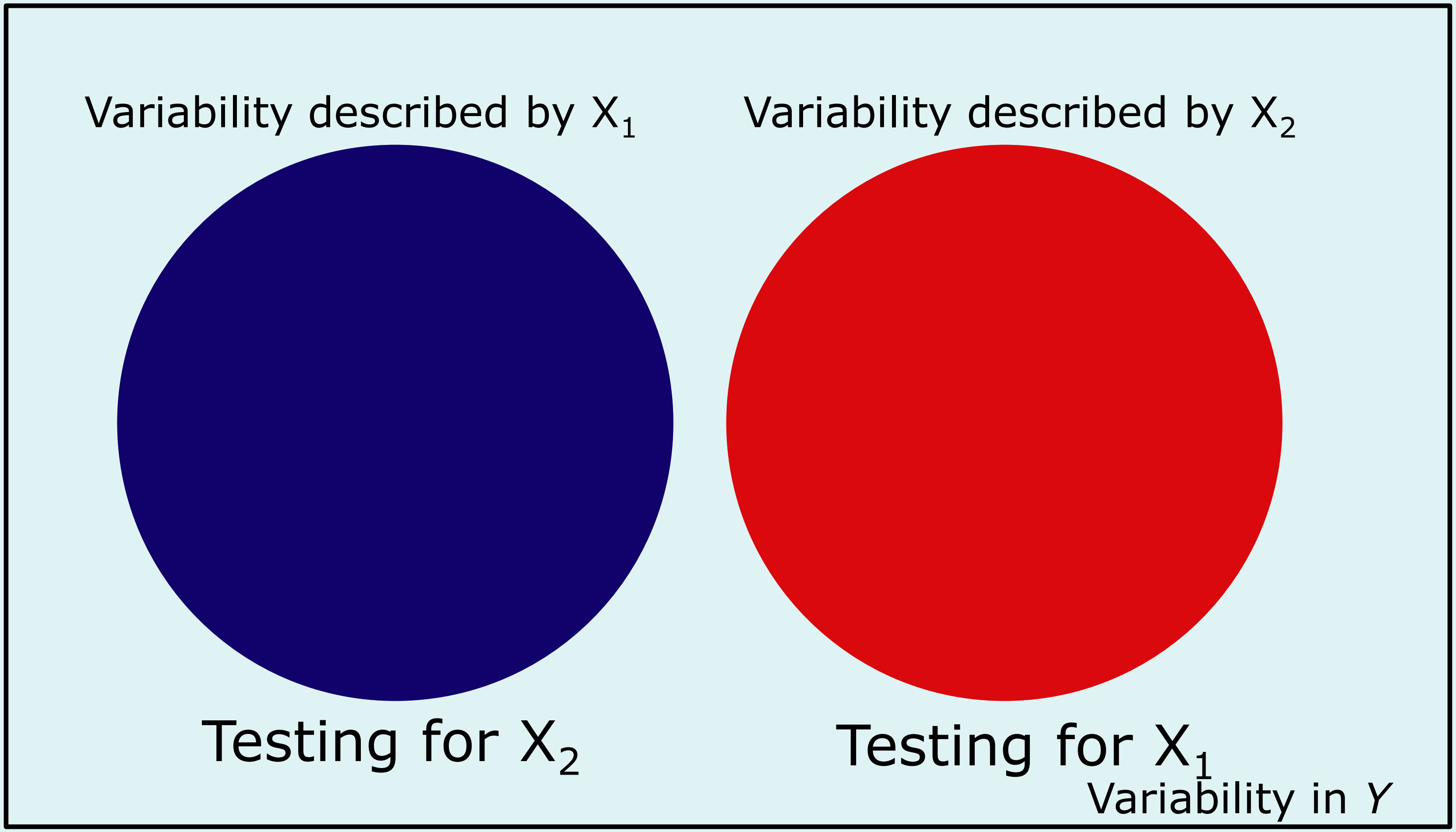
Residual var. = 0.2

P(b1 = 0) = 0.0003
(t test b1>0)

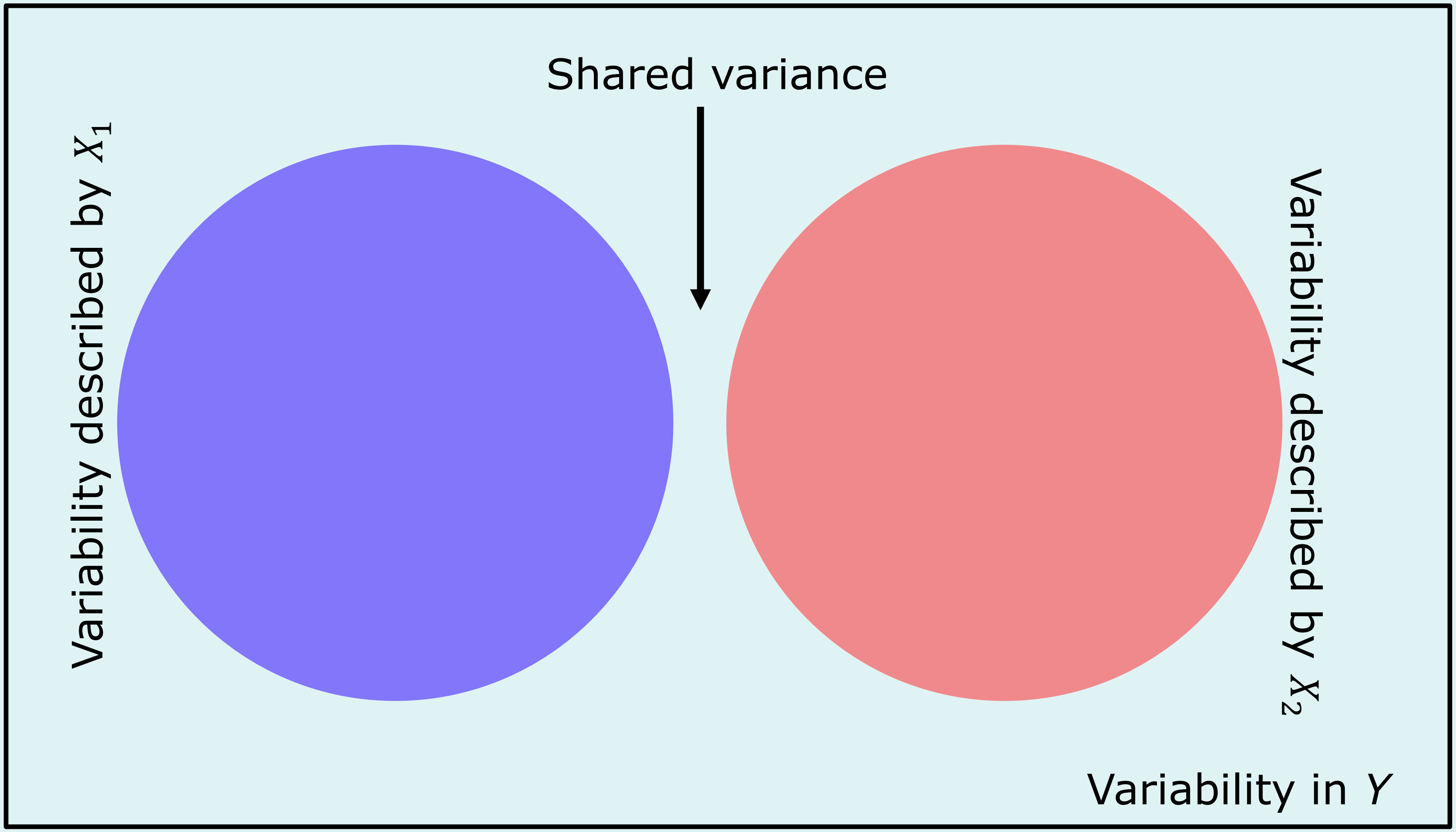
P(b2 = 0) = 0.07
(t test b2>0)

P([b1 b2] = 0) = 0.002
(F test [b1 b2] ≠ 0)

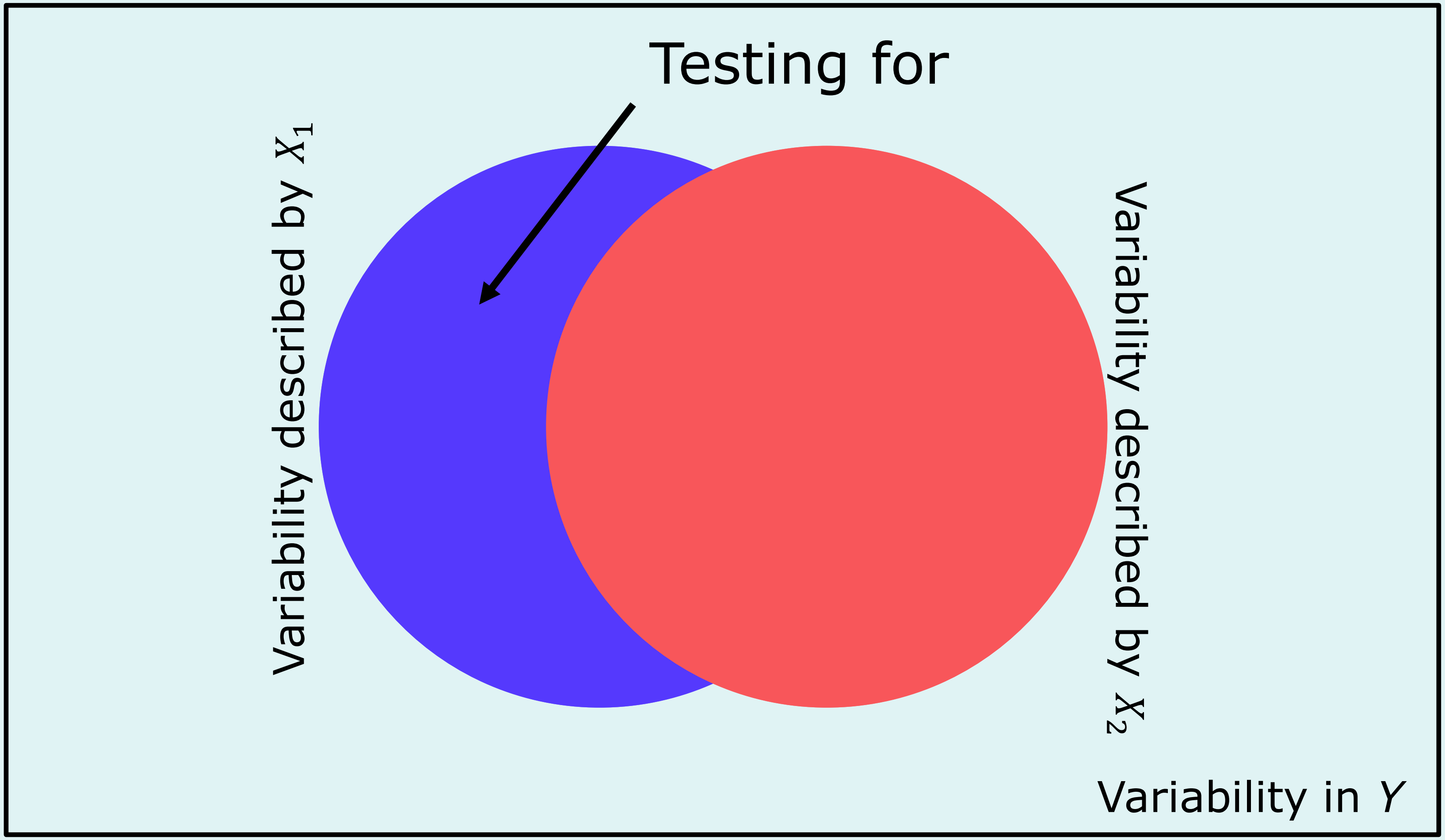
Orthogonal regressors



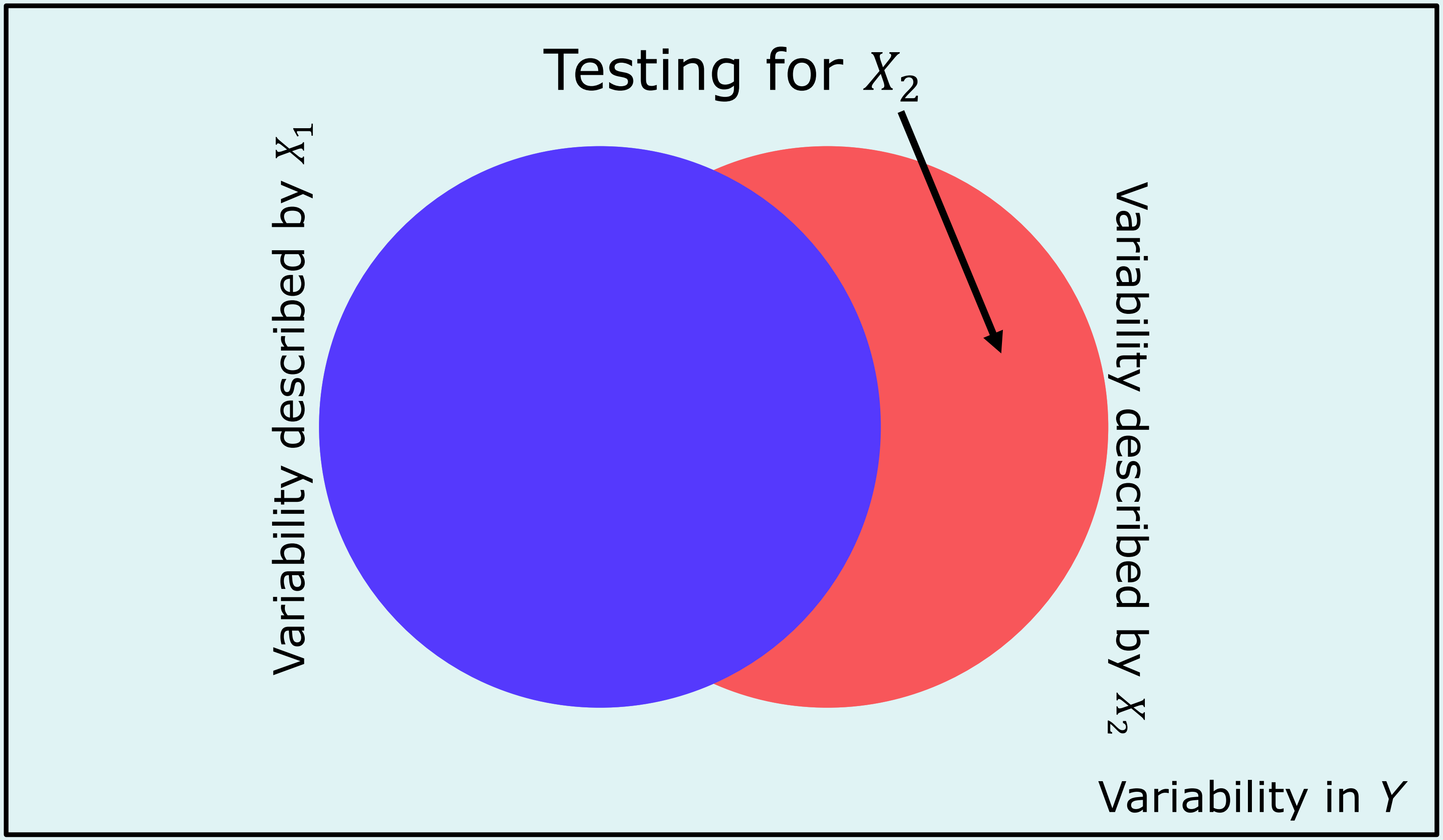
Orthogonal regressors



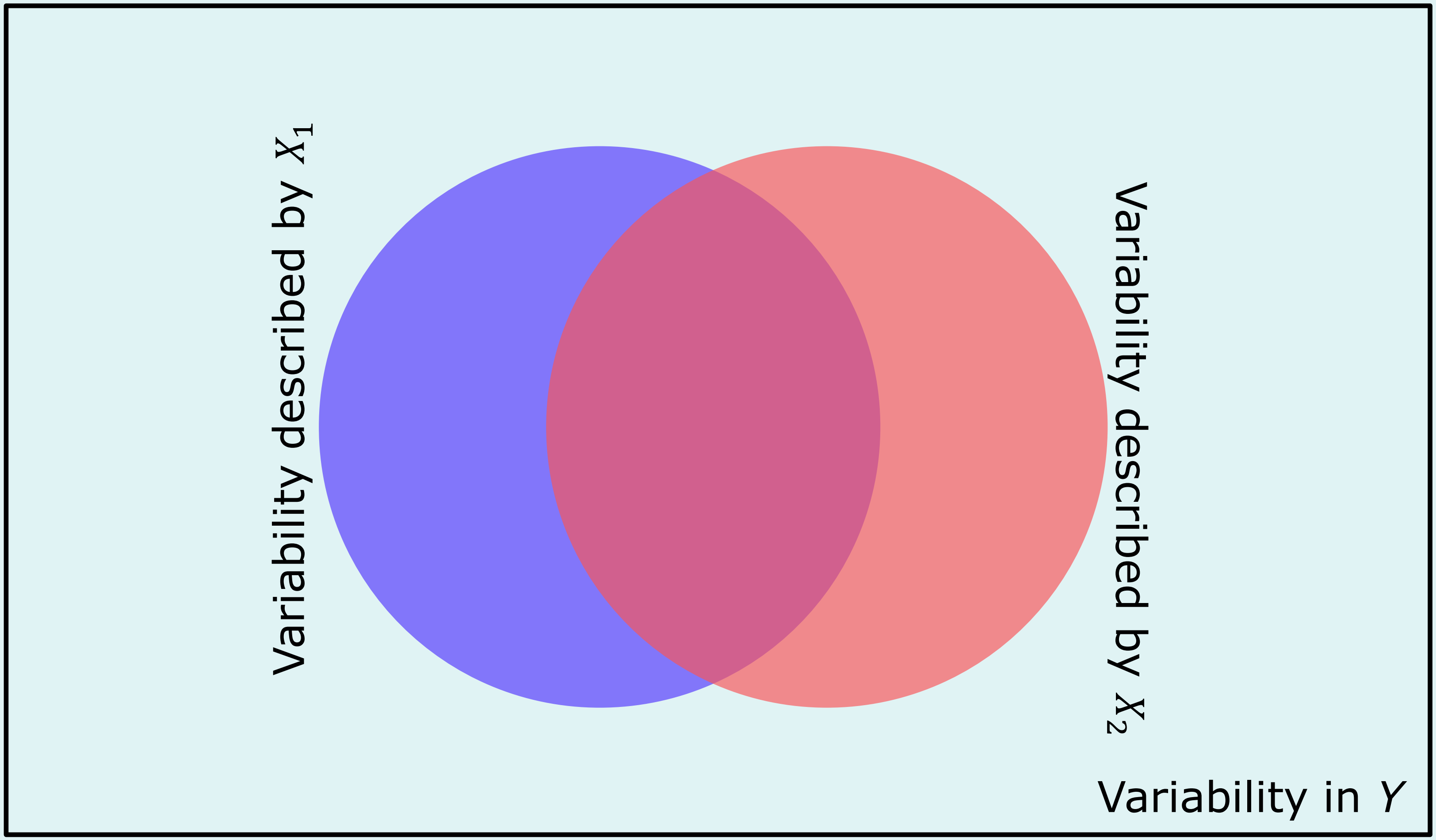
Orthogonal regressors



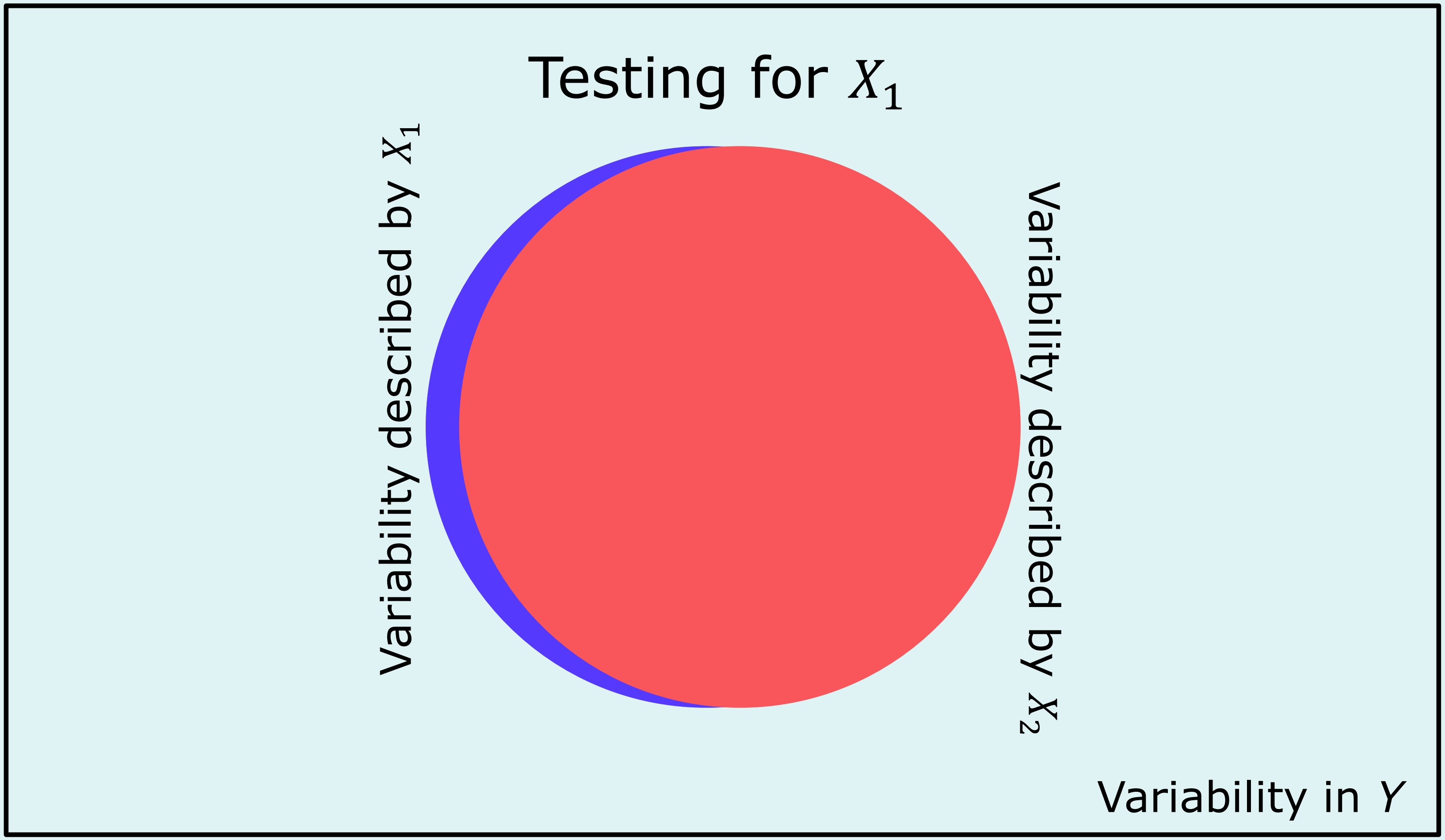
Orthogonal regressors



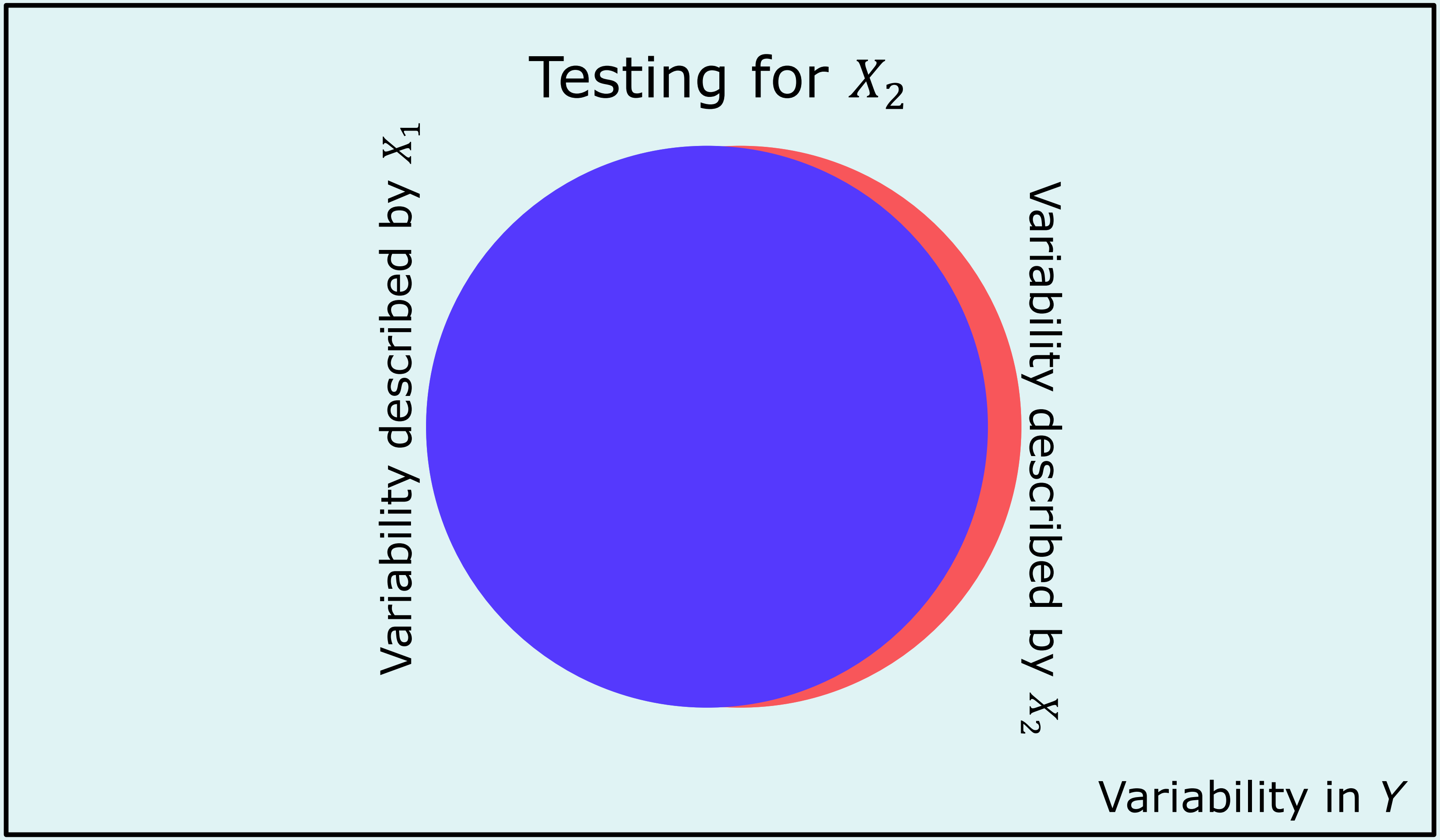
Orthogonal regressors



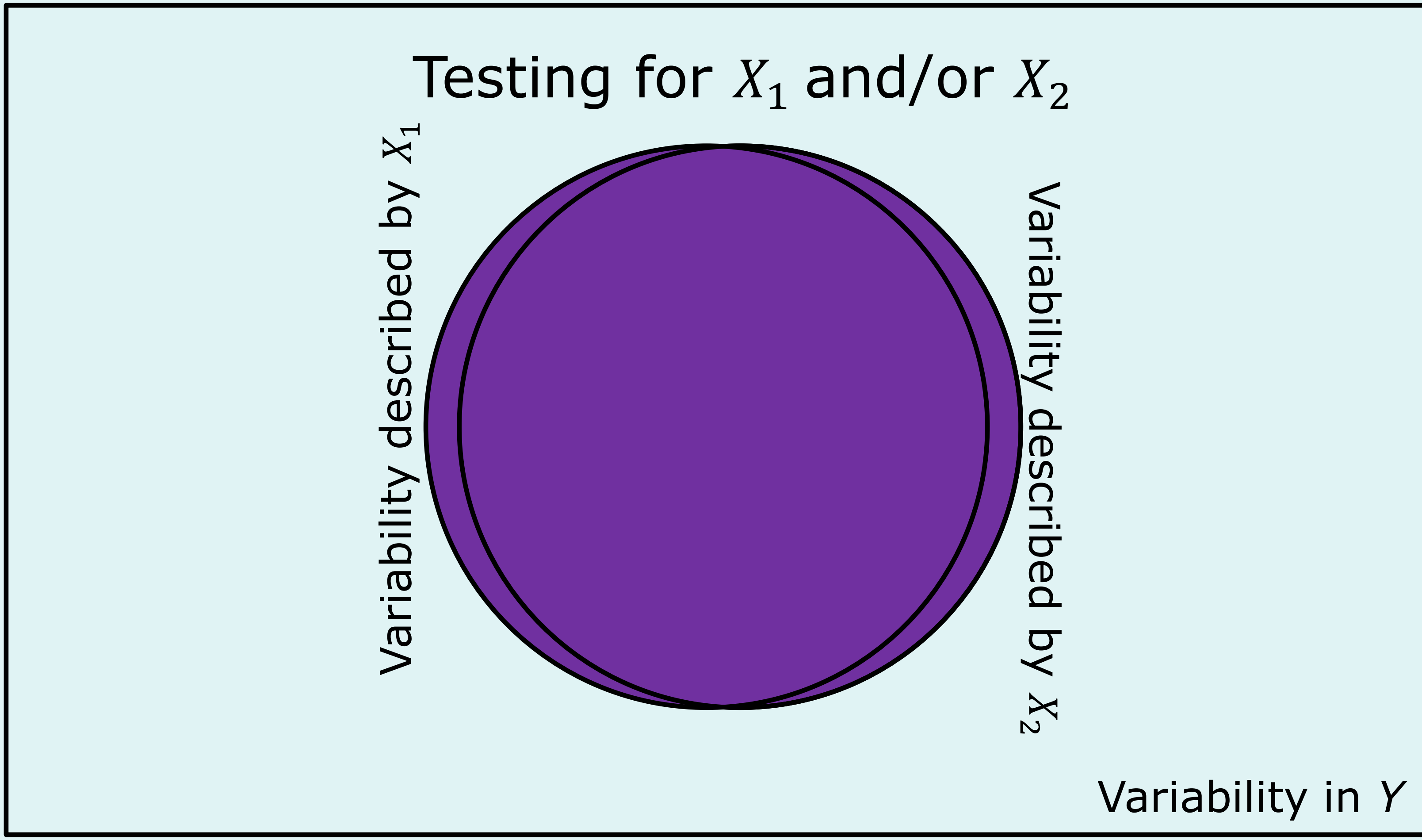
Orthogonal regressors



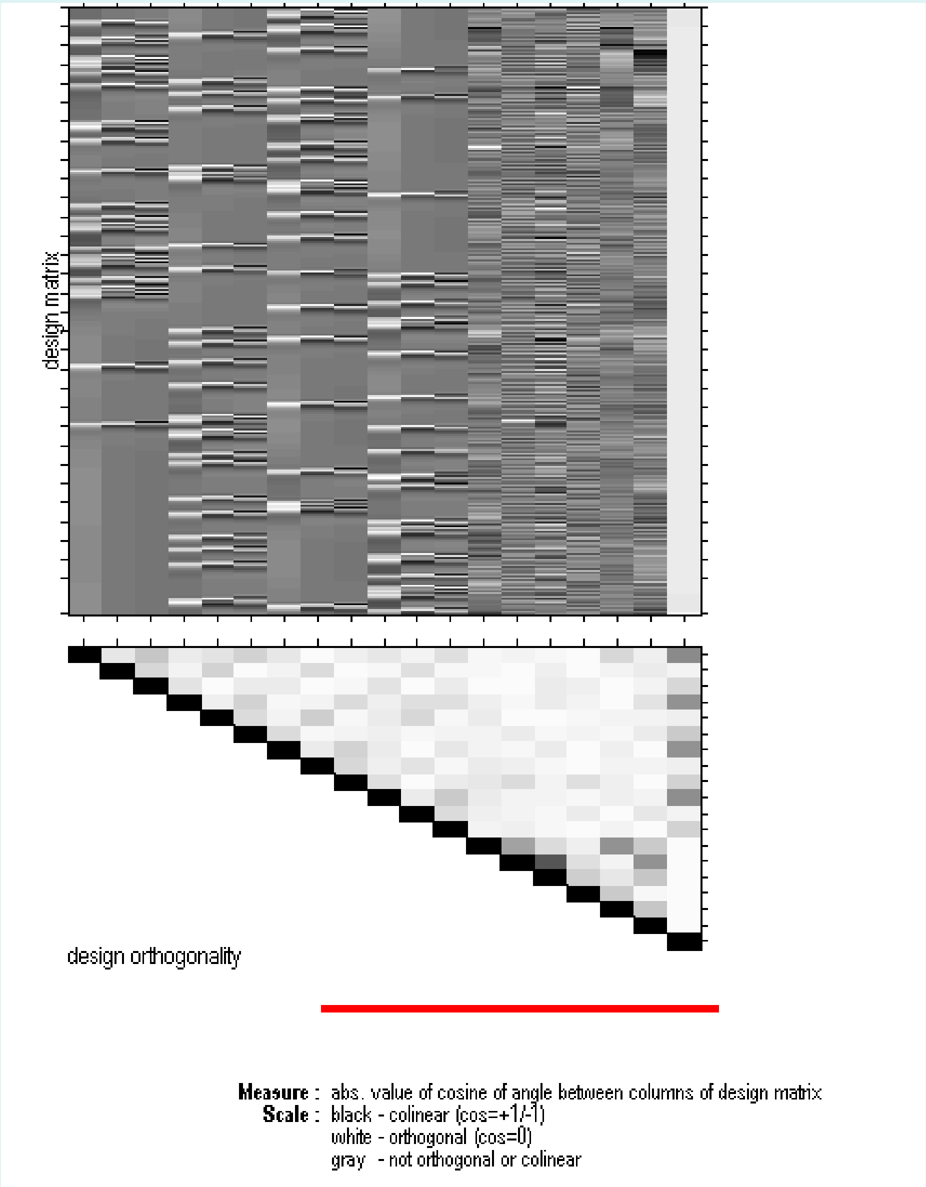
Orthogonal regressors



Orthogonal regressors



Design orthogonality

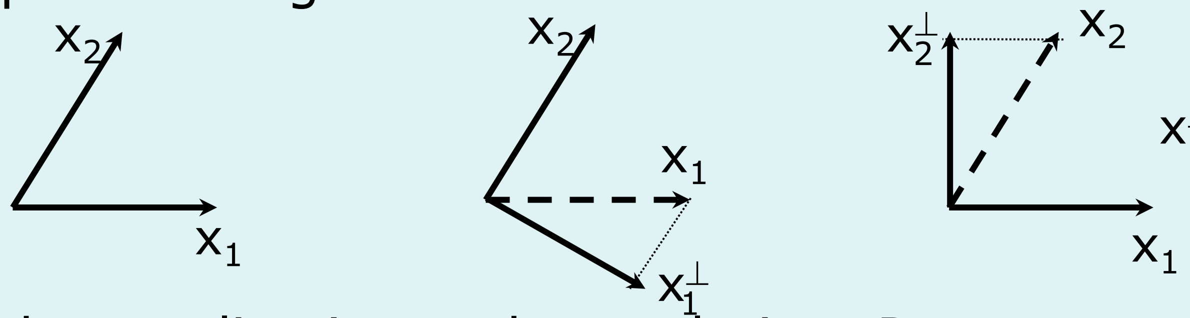


- For each pair of columns of the design matrix, the orthogonality matrix depicts the magnitude of the **cosine of the angle** between them, with the range 0 to 1 mapped from white to black.

- 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.

Correlated regressors

- We implicitly test for an additional effect only. When testing for the first regressor, we are effectively removing the part of the signal that can be accounted for by the second regressor → implicit orthogonalisation.



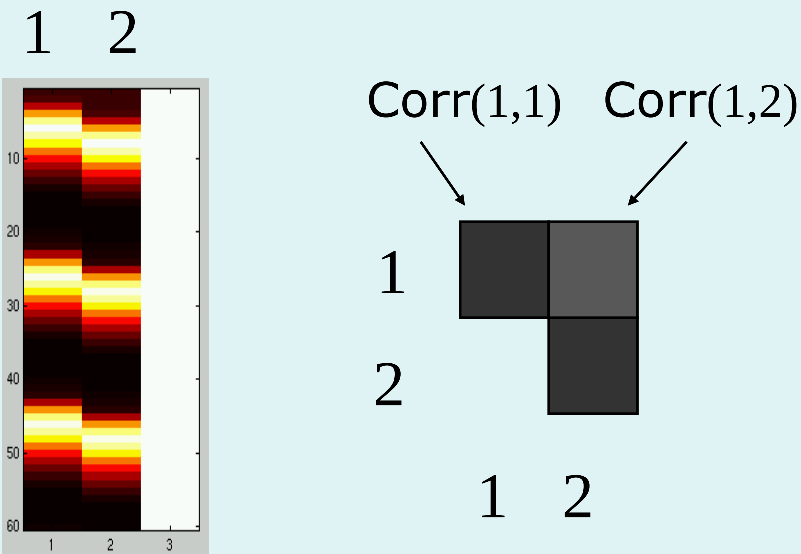
$x_2^{\perp} = x_2 - x_1.x_2\ x_1$

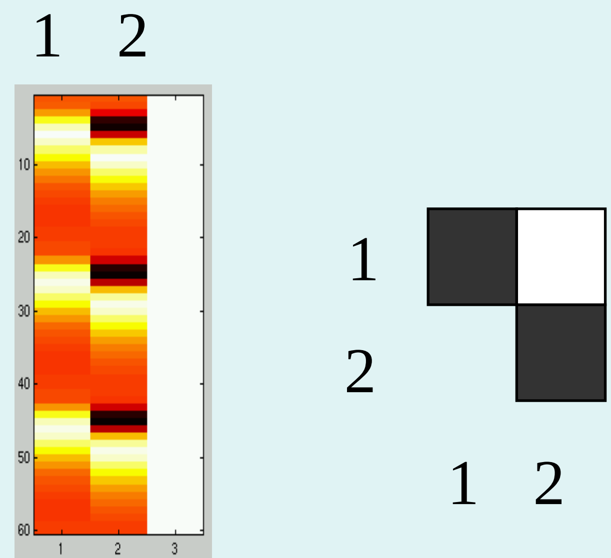
- Orthogonalisation = decorrelation. Parameters and test on the non modified regressor change.
  Rarely solves the problem as it requires assumptions about which regressor to uniquely attribute the common variance.
  → change regressors (i.e. design) instead, e.g. factorial designs.
  → use F-tests to assess overall significance.
- Original regressors may not matter: it's the contrast you are testing which should be as decorrelated as possible from the rest of the design matrix

Design orthogonality

Black = completely correlated

White = completely orthogonal



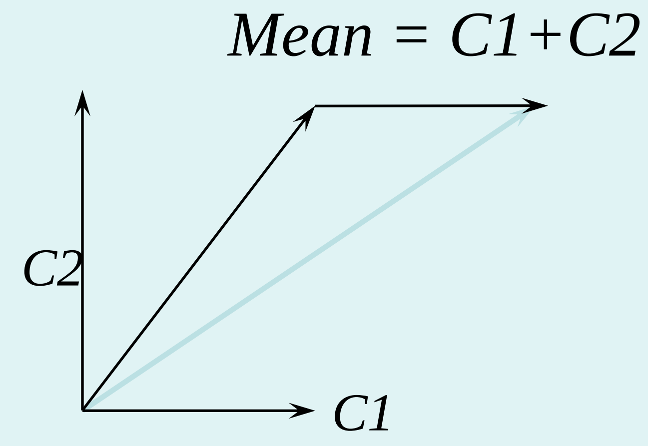


Beware : when there is more than 2 regressors (C1,C2,C3...), you may think that there is little correlation (light grey) between them, but C1 + C2 + C3 may be correlated with C4 + C5

Rank-deficient model

$$Y = Xb + e \qquad X = \begin{matrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{matrix}$$

Cond 1 *Cond 2* *Mean*



Parameters are not unique in general !
Some contrasts have no meaning: NON ESTIMABLE

Example here :

- c' = [1 0 0] is not estimable
  (= no specific information in the first regressor);
- c' = [1 -1 0] is estimable.

Summary

- We are implicitly testing additional effect only, so we may miss the signal if there is some correlation in the model using t tests
- Orthogonalisation is not generally needed - parameters and test on the changed regressor don't change
- It is always simpler (when possible !) to have orthogonal (uncorrelated) regressors
- In case of correlation, use F-tests to see the overall significance. There is generally no way to decide where the « common » part shared by two regressors should be attributed to
- In case of correlation and you need to orthogonolise a part of the design matrix, there is no need to re-fit a new model : the contrast only should change.

Content

- **Introduction**
- **Contrast & Inference**
- **Orthogonality issue**
- **Conclusion**

Way to proceed

Prepare your questions.
ALL the questions !

Find a model which
- allows contrasts that translates these questions.
- takes into account ALL the effects (interaction, sessions,etc)

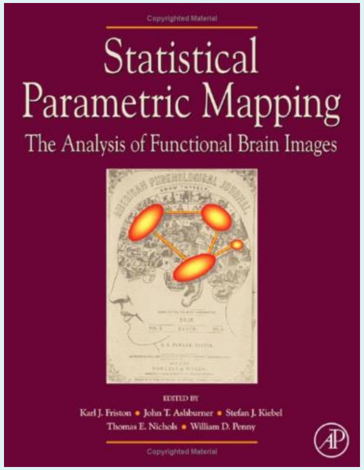
Devise task & stimulus presentation.

Acquire the data & analyse.

Not the other way round!!!

References

- Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier, 2007.



- Plane Answers to Complex Questions: The Theory of Linear Models. R. Christensen, Springer, 1996.
- Statistical parametric maps in functional imaging: a general linear approach. K.J. Friston et al, Human Brain Mapping, 1995.
- Ambiguous results in functional neuroimaging data analysis due to covariate correlation. A. Andrade et al., NeuroImage, 1999.