General Linear Model for fMRI: bases of statistical analyses

Cyril Pernet, PhD
University of Edinburgh

Objectives

- Intuitive understanding of the GLM
- Get an idea how t-tests, ANOVA, regressions, etc .. are instantiation of the GLM
- Learn key concepts: linearity, model, design matrix, collinearity and orthogonalization

Overview

- What is linearity?
- Why do we speak of models?

- A simple fMRI model
- A more complex model
- Issues with regressors

What is linearity?

Linearity

- Means created by lines
- In maths it refers to equations or functions that satisfy 2 properties: additivity (also called superposition) and homogeneity of degree 1 (also called scaling)
- Additivity → $y = x1 + x2$ (output is sum of inputs)
- Scaling → $y = \beta x1$ (output is proportional to input)

http://en.wikipedia.org/wiki/Linear

Examples of linearity / non linearity

X = sort(randn(10,1))

- <u>Linear correlation</u>

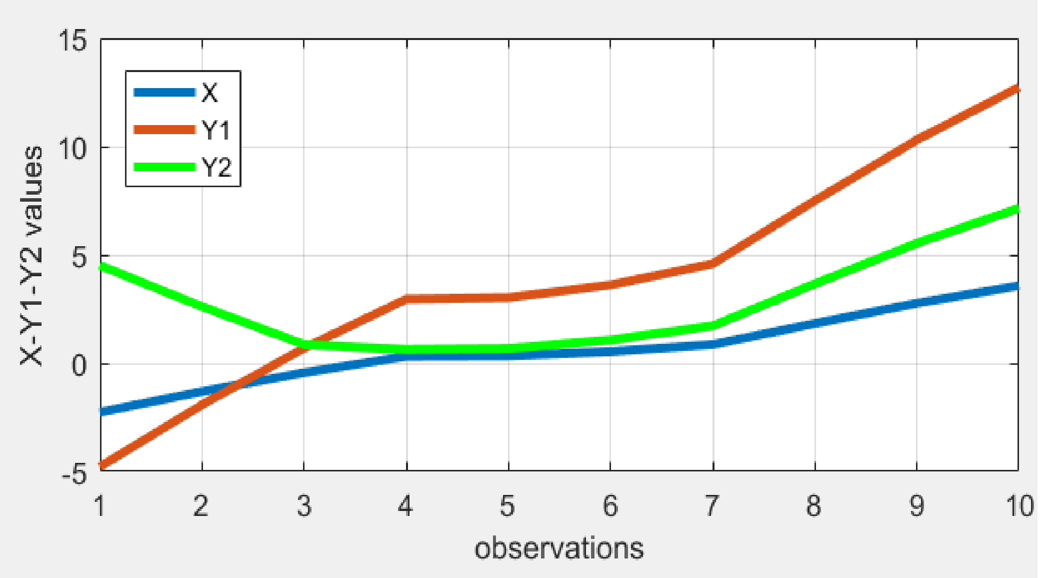
Y1 = 3*X + 2

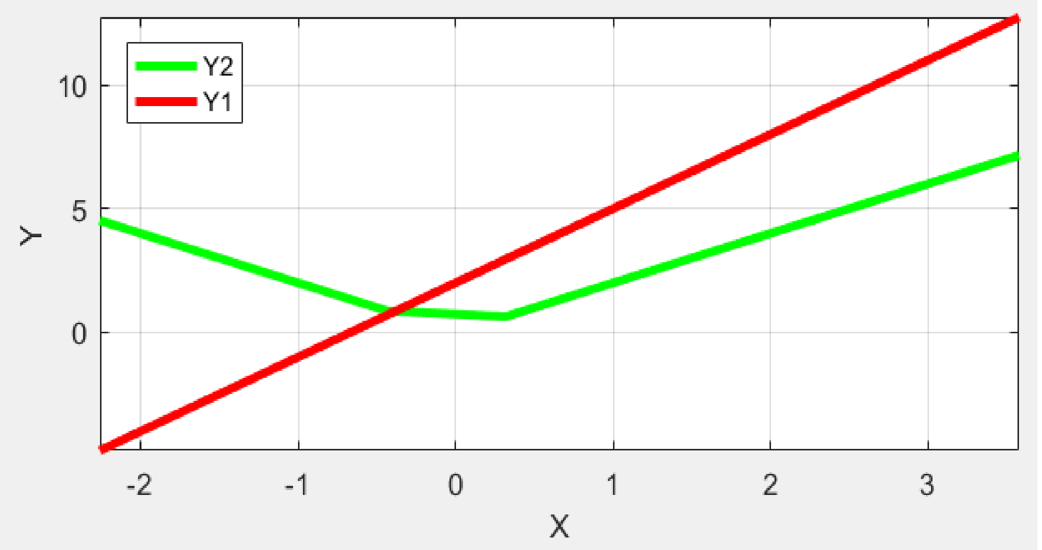
Pearson r = 1

- <u>Non linear correlation</u>

Y2 = abs(2*X)

Pearson r = 0.5





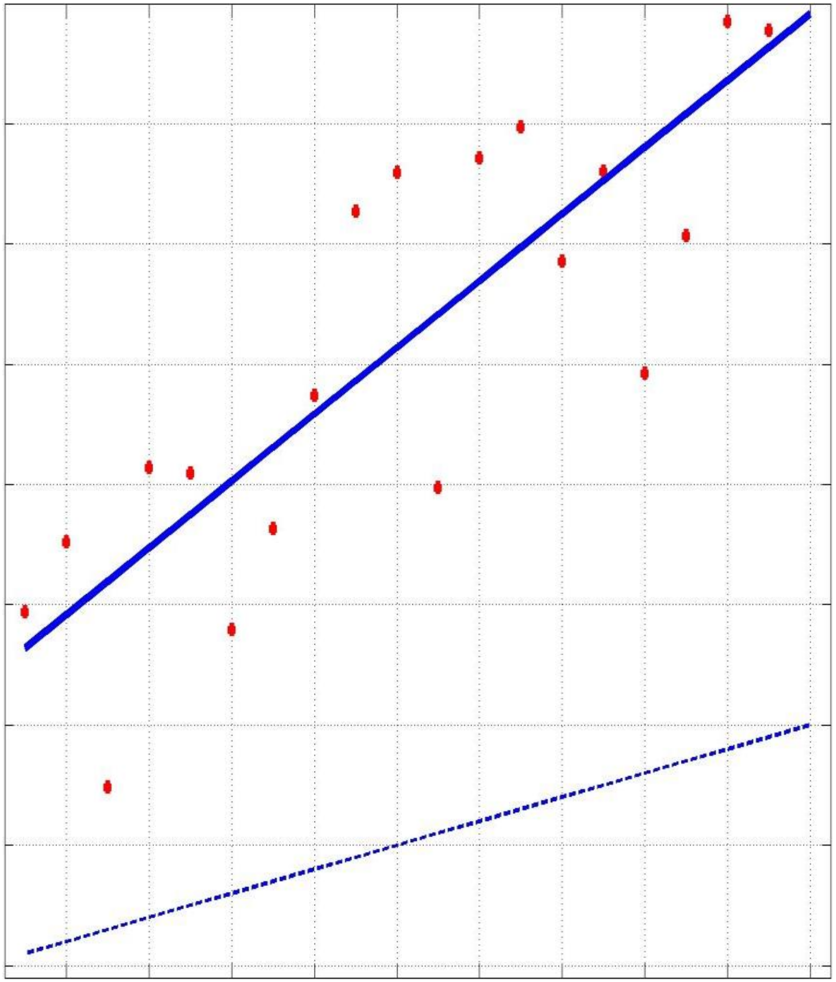
What is a linear model?

What is a linear model?

- An equation or a set of equations that models data and which corresponds geometrically to straight lines, plans, hyperplans and satisfy the properties of additivity and scaling.
- Simple regression: $y = \beta 1x1 + \beta 2 + \varepsilon$
- Multiple regression: $y = \beta 1x1 + \beta 2x2 + \beta 3 + \varepsilon$
- One way ANOVA: $y = \mu + \alpha i + \varepsilon$
- Repeated measure ANOVA: $y = u + Si + \alpha i + \varepsilon$
- …

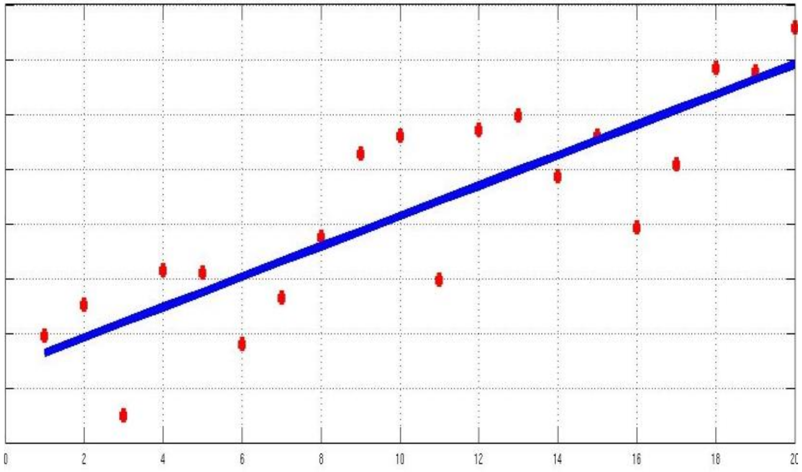
A regression is a linear model

- We have an experimental measure x (e.g. stimulus intensity from 0 to 20)
- We then do the expe and collect data y (e.g. RTs)
- Model: $y = \beta 1x+\beta 2$
- Do some maths / run a software to find $\beta 1$ and $\beta 2$
  $y\hat{} = 2.7x+23.6$

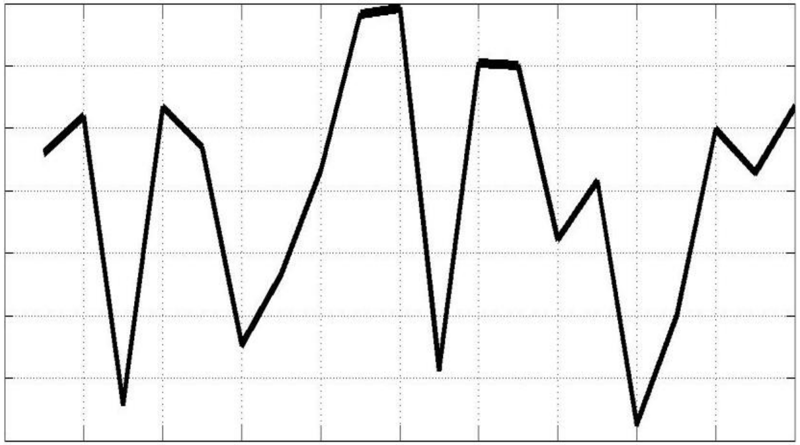


A regression is a linear model

- The error is the distance between the data and the model
- F = (SSeffect / df) / (SSerror / df_error)
- <u>SSeffect</u> = norm(model - mean(model)).^2;
- <u>SSerror</u> = norm(residuals).^2;

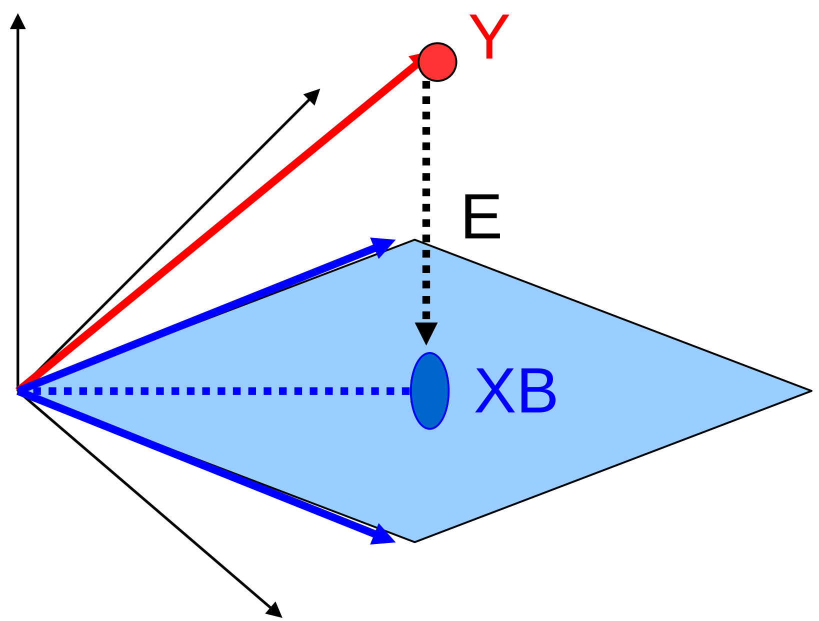


Error = Y – XB



Summary

- Linear model: $y = x1\beta1 + x2\beta2$ (output = additivity and scaling of input)

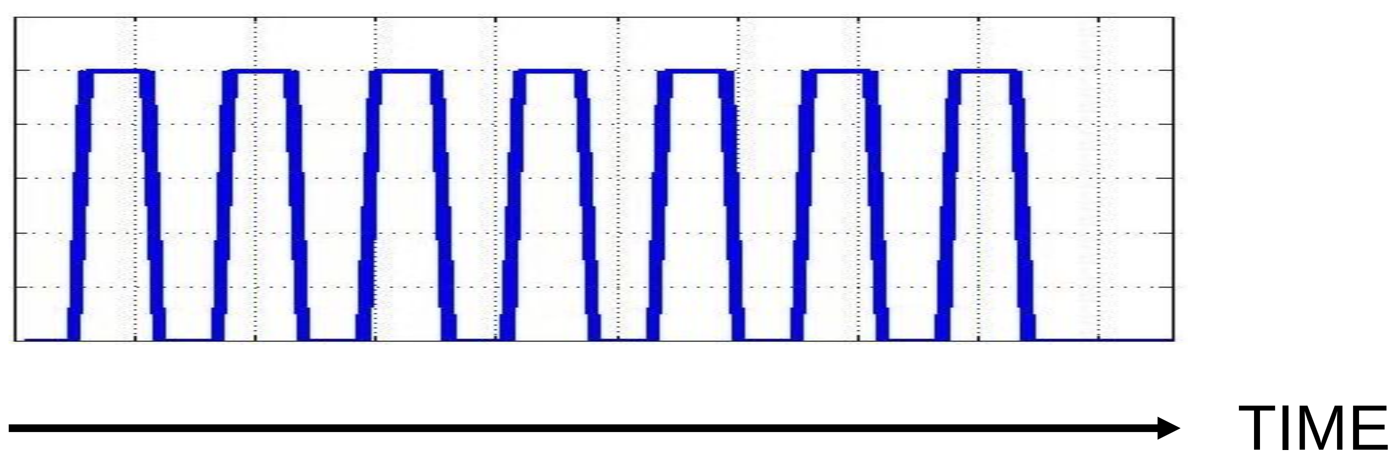


A simple fMRI model

http://www.fil.ion.ucl.ac.uk/spm/data/auditory/

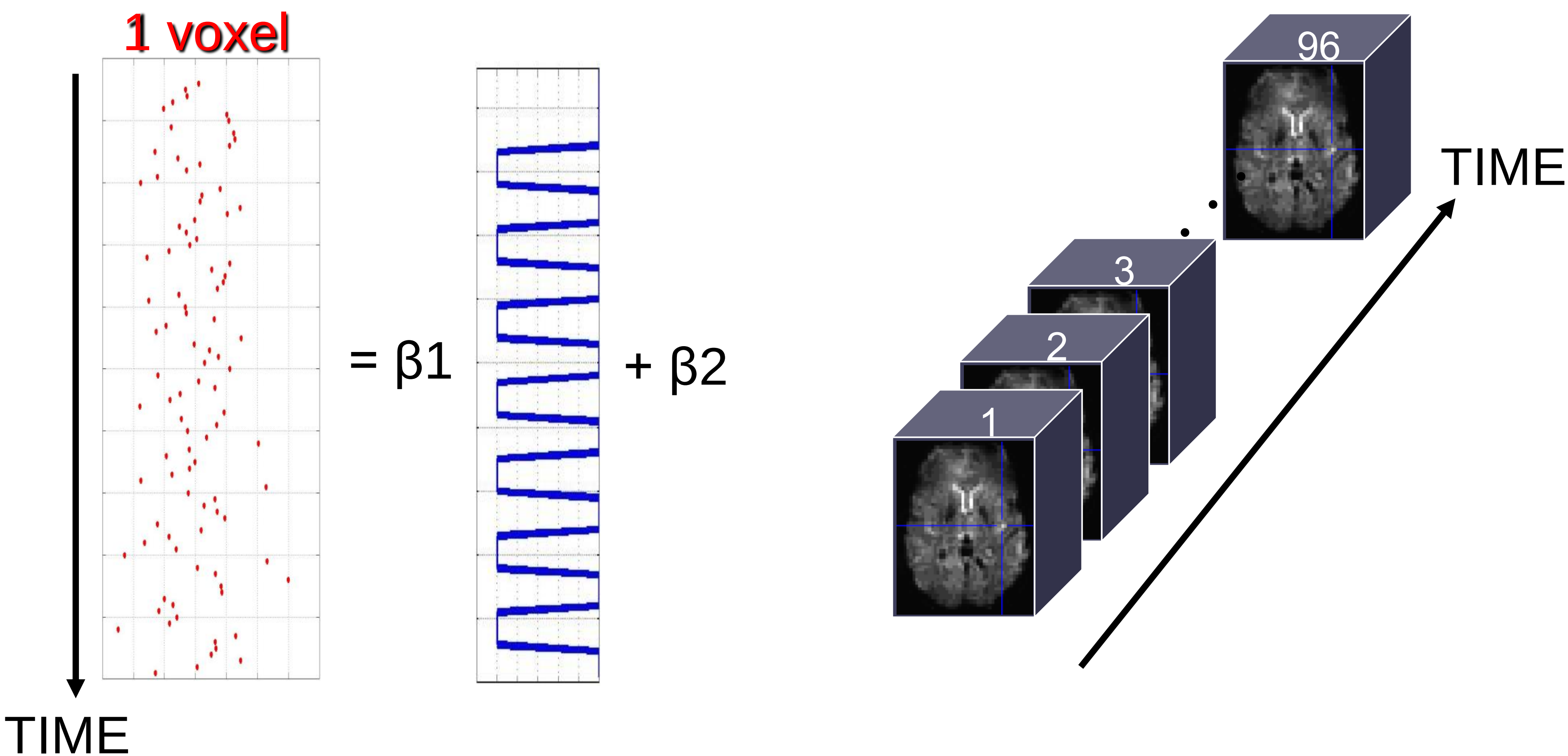
FMRI experiment

- *SPM auditory data set*: which areas are activated by the presentation of bi-syllabic words presented binaurally (60 per minute)
- Experimental measure x: 7 blocks of 42 sec of stimulation

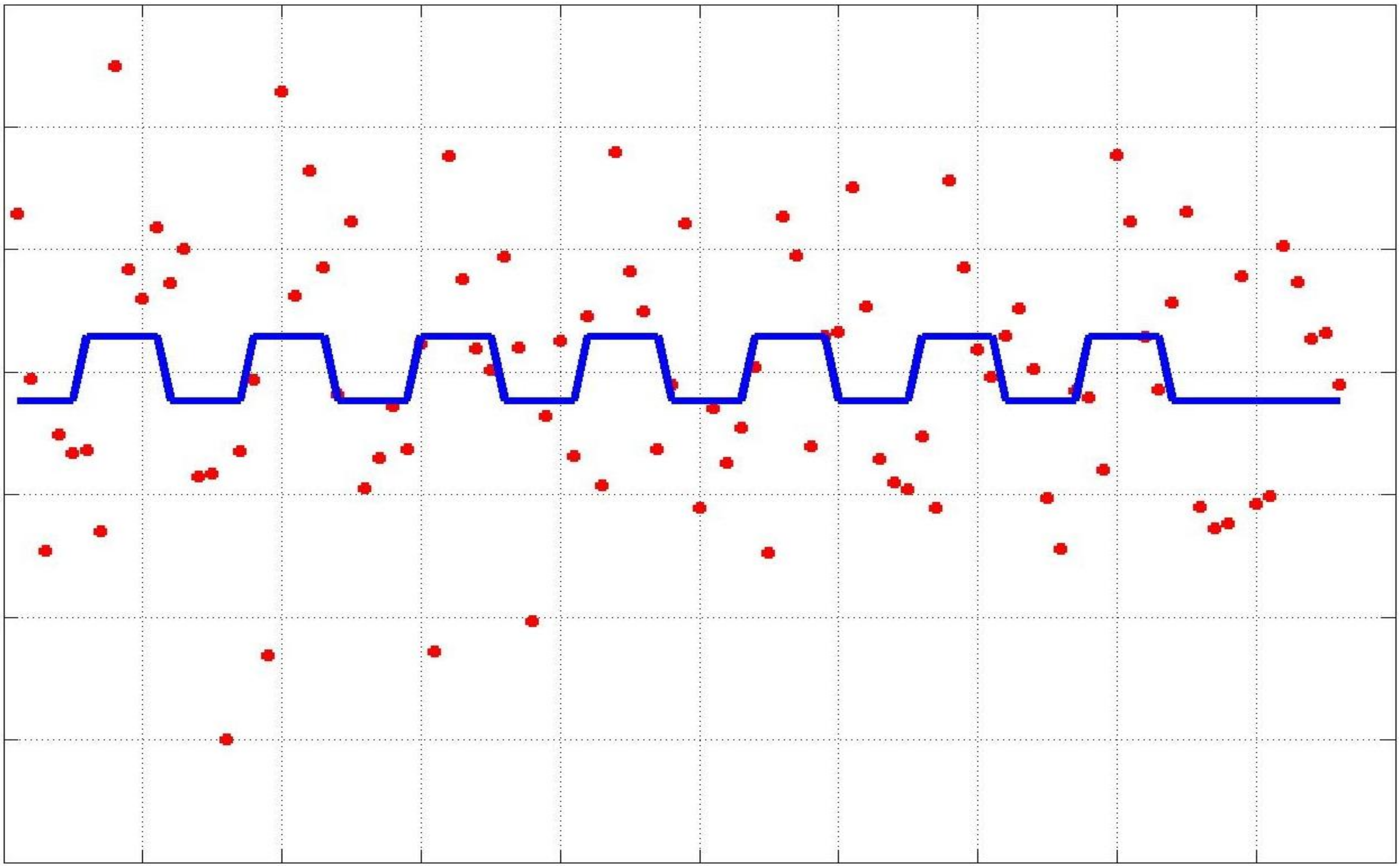


FMRI experiment

- Collect the data : 96 fMRI volumes (TR=7s)
- Model: y = β1x+ β2

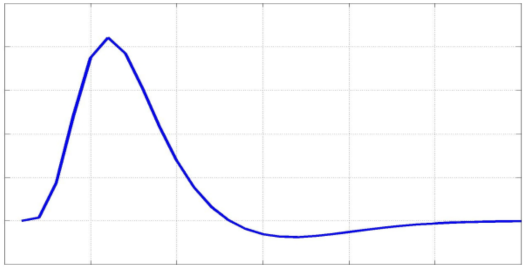


FMRI experiment

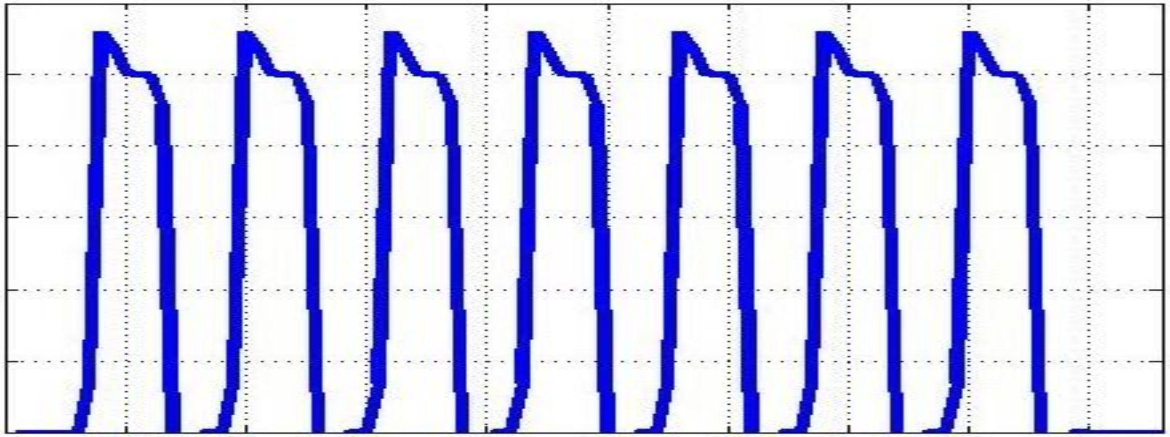


FMRI experiment

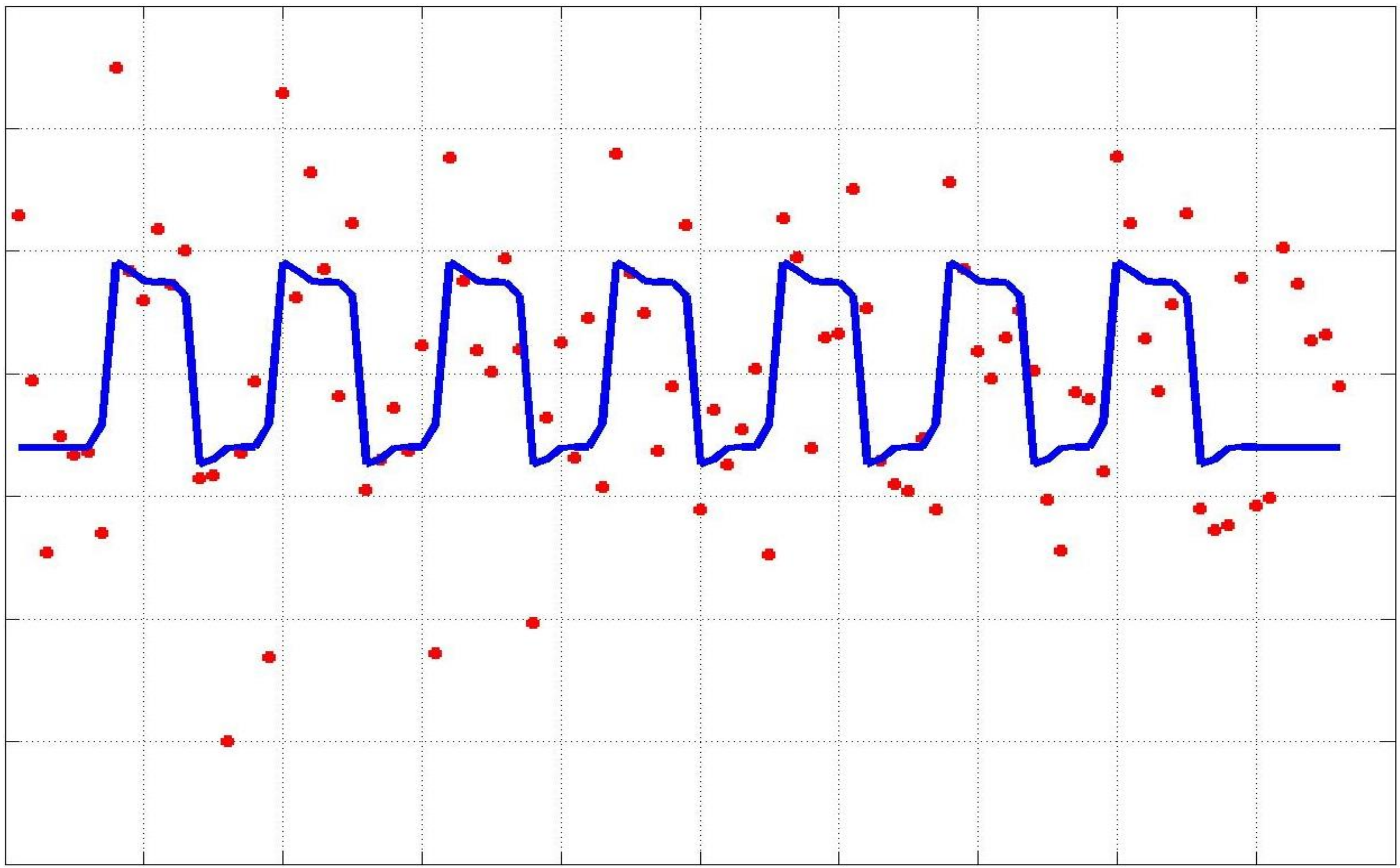
- A better model: we know the shape of the BOLD response



- Convolution by the hrf: x ⊗ hrf



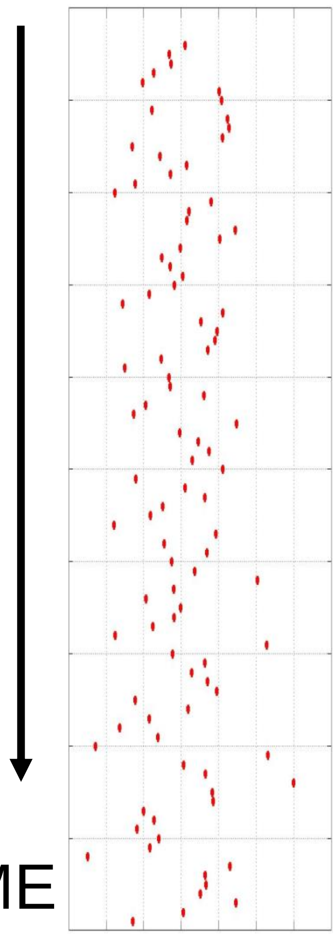
FMRI experiment



FMRI experiment

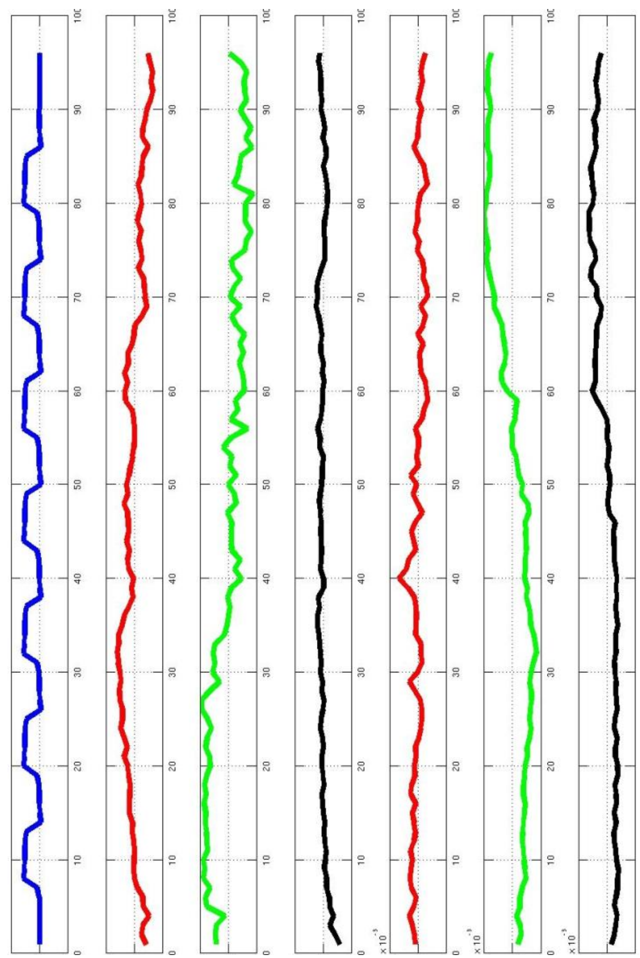
- An even better model: add motion parameters

1 voxel



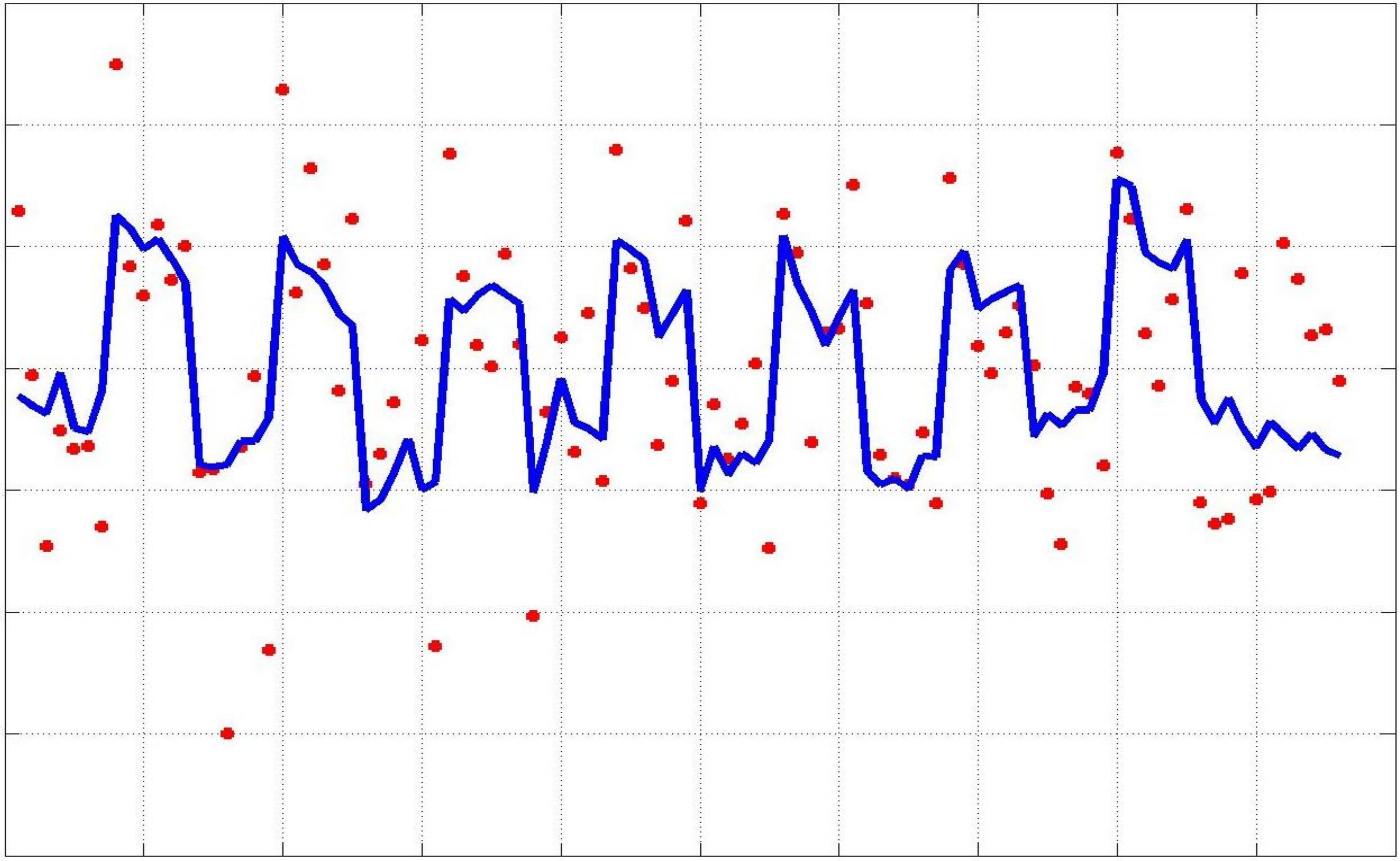
TIME

=



* [β1 β2 β3 β4 β5 β6 β7 β8]+ β8

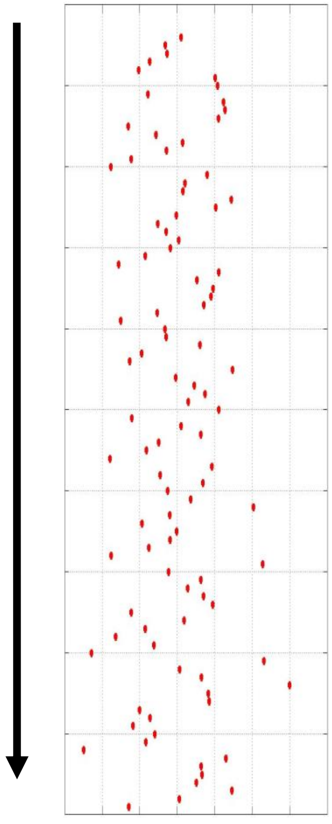
FMRI experiment



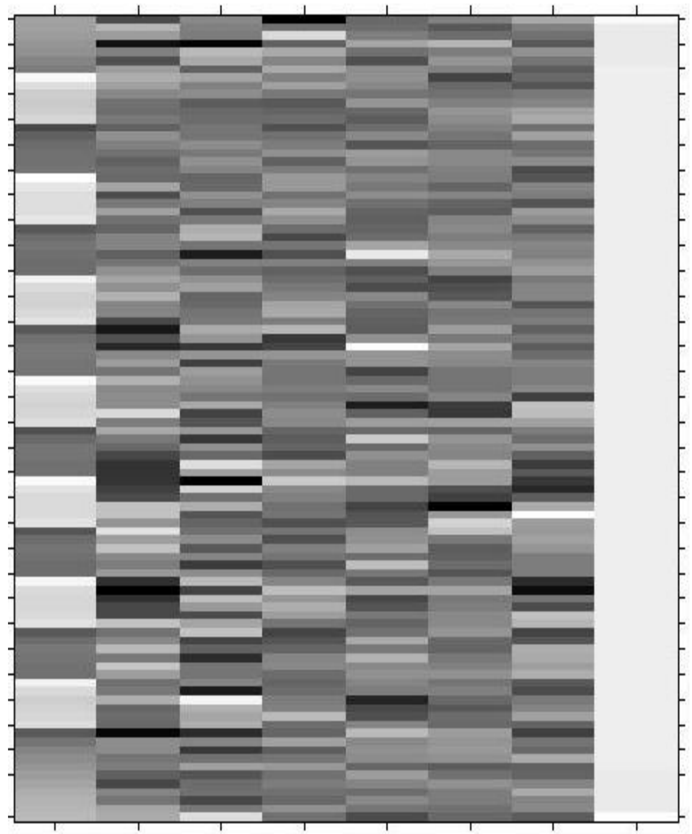
FMRI experiment

- Matrix formulation and SPM colour coding

1 voxel



=



* [β1 β2 β3 β4 β5 β6 β7 β8]

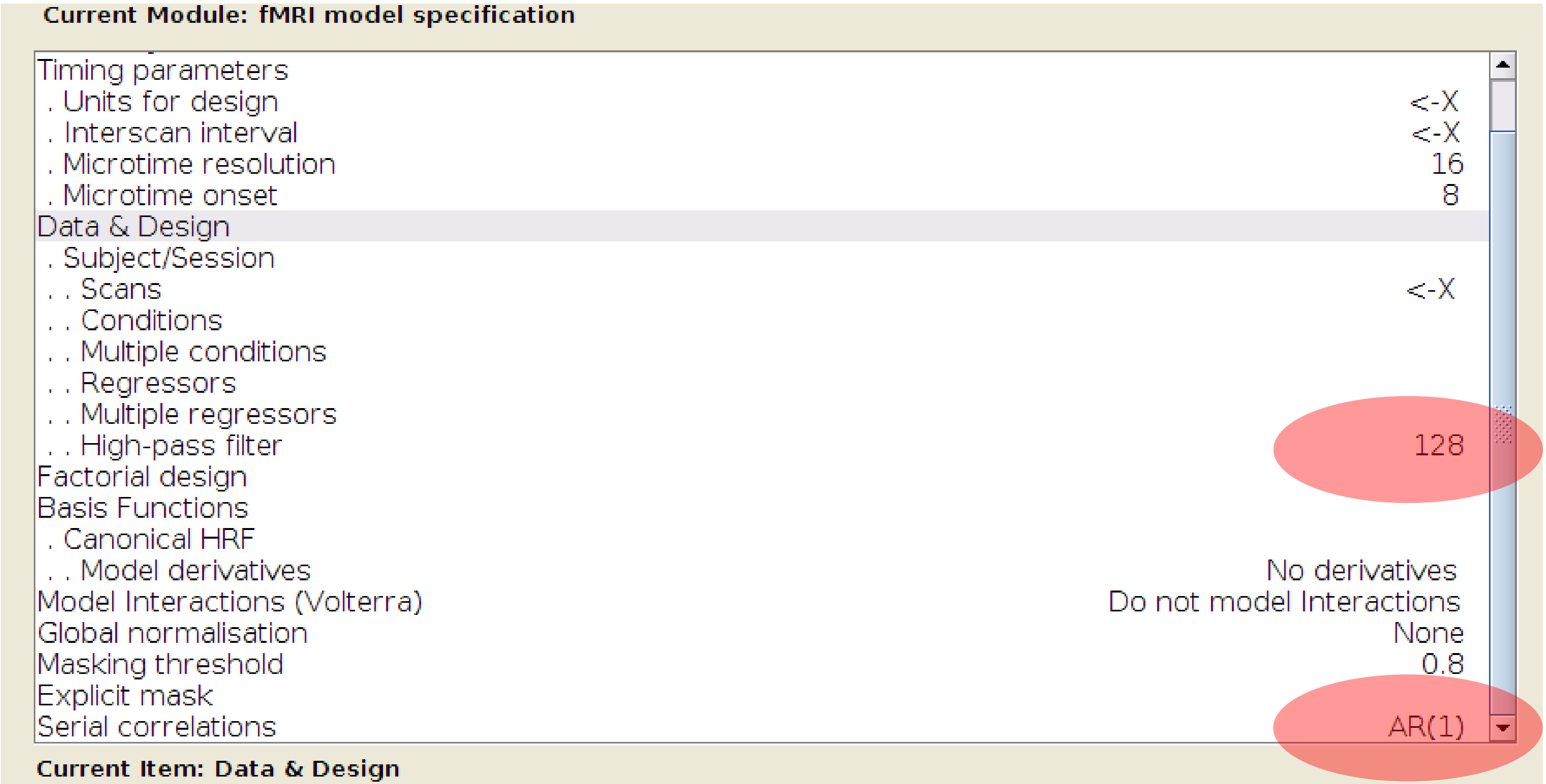
BetaXXXX.hdr
BetaXXXX.img

FMRI data (Y) = Design matrix (X = SPM.mat) * B + E (ResMS)

Noise modelling

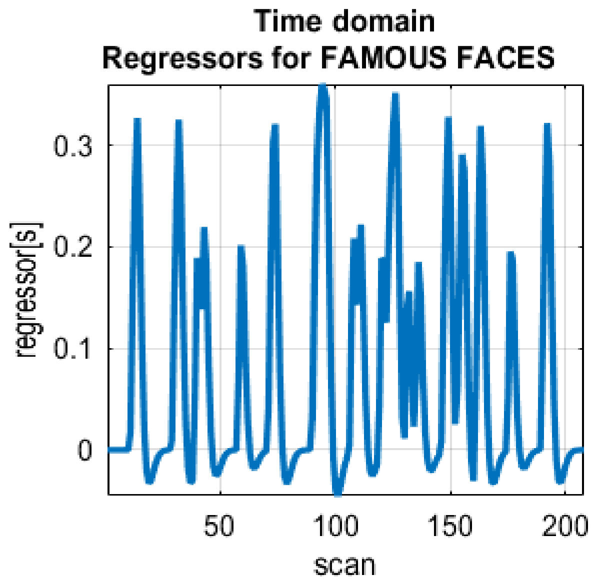
- GLM: $Y = XB+E$ with $E \sim N(0,V)$
- Because of scanner issue, data often have a drift, that we can incorporate into the model X (high pass filter)
- V is the covariance matrix whose depends on your model – in SPM we use an AR(1) model + white noise to remove dependencies related to physiological artefacts (e.g. cardiac aliasing) and model residual noise

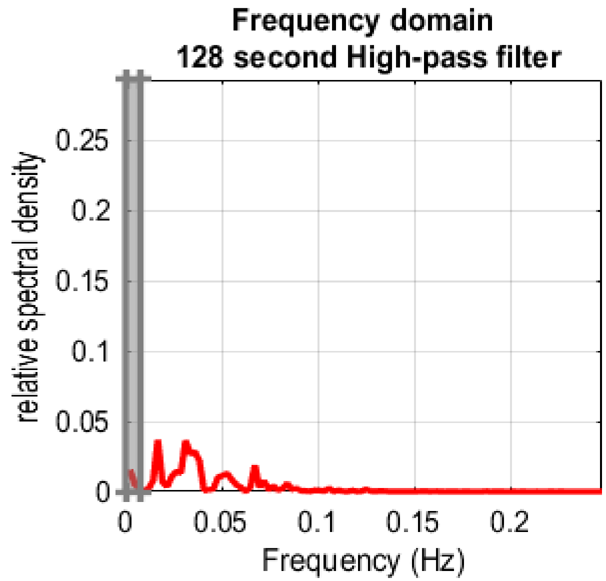
FMRI experiment

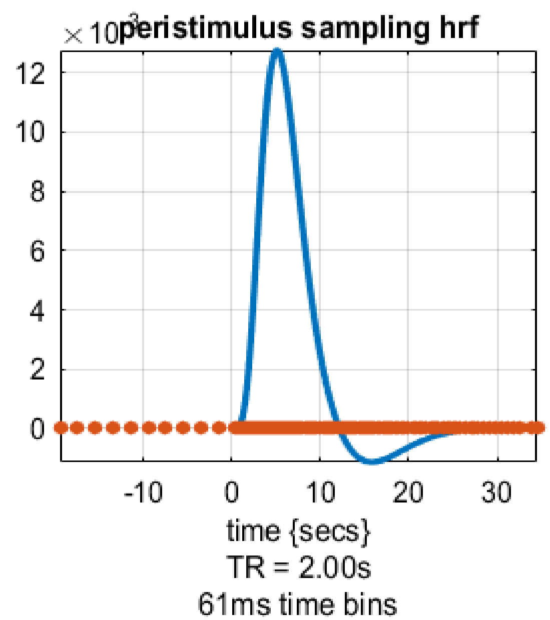


Note: High pass filter in seconds 1/128 sec = 0.007 Hz
The hrf itself peaks at ~0.04Hz

When you specify a design you can check it !







What should be your filter?

- the old rule of thumb (a la SPM99) is two take the highest freq (longest periods between repeats) and multiply by 2 (to have a filter away from your peak) – ideally toward high frequencies
- Because noise is low frequency – e.g. resting is 0.01 Hz - 0.1Hz → 100sec / 10sec → can't really get much lower than the 128 sec (0.007Hz)

Summary

- Linear model: $y = \beta 1 x1 + \beta 2 x2$ (output = additivity and scaling of input)
- GLM: Y = XB+E (matrix formulation, works for any statistics, express the data Y as a function of the design matrix X)

A more complex model

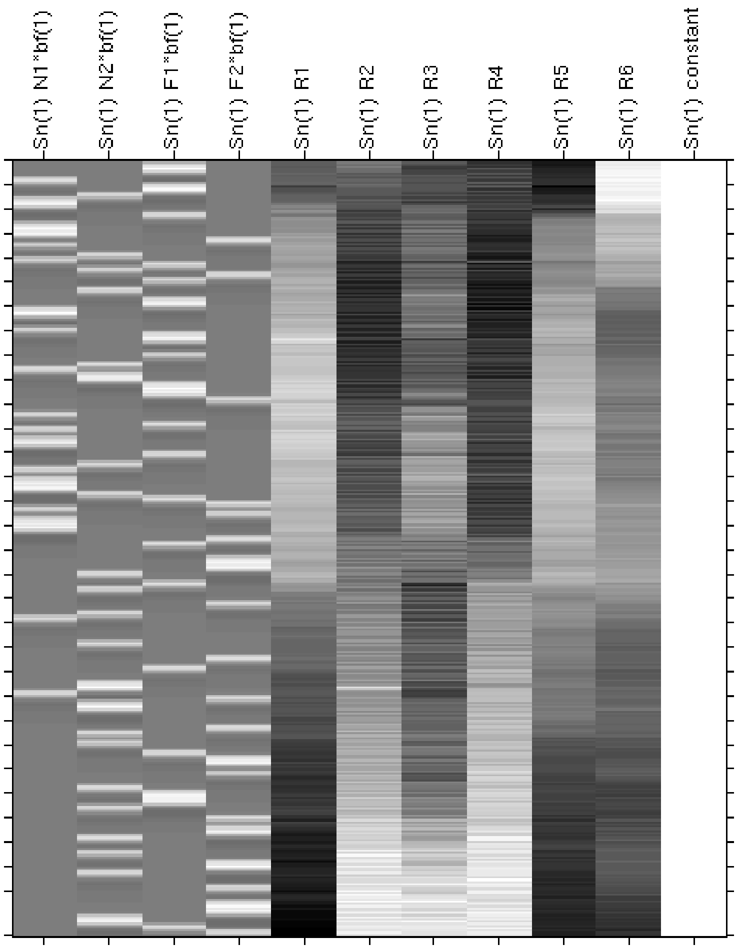
2 x 2 factorial design

- New experiment: (Famous vs. Nonfamous) x (1st vs 2nd presentation) of faces against baseline of chequerboard
- 2 presentations of 26 Famous and 26 Nonfamous Greyscale photographs, for 0.5s, randomly intermixed, for fame judgment task (one of two right finger key presses).

http://www.fil.ion.ucl.ac.uk/spm/data/face_rep/face_rep_SPM5.htm

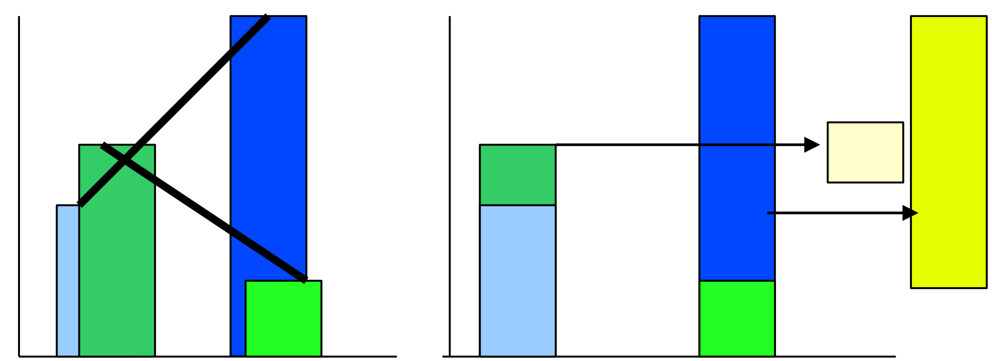
Contrasts (more tomorrow)

- SPM design matrix



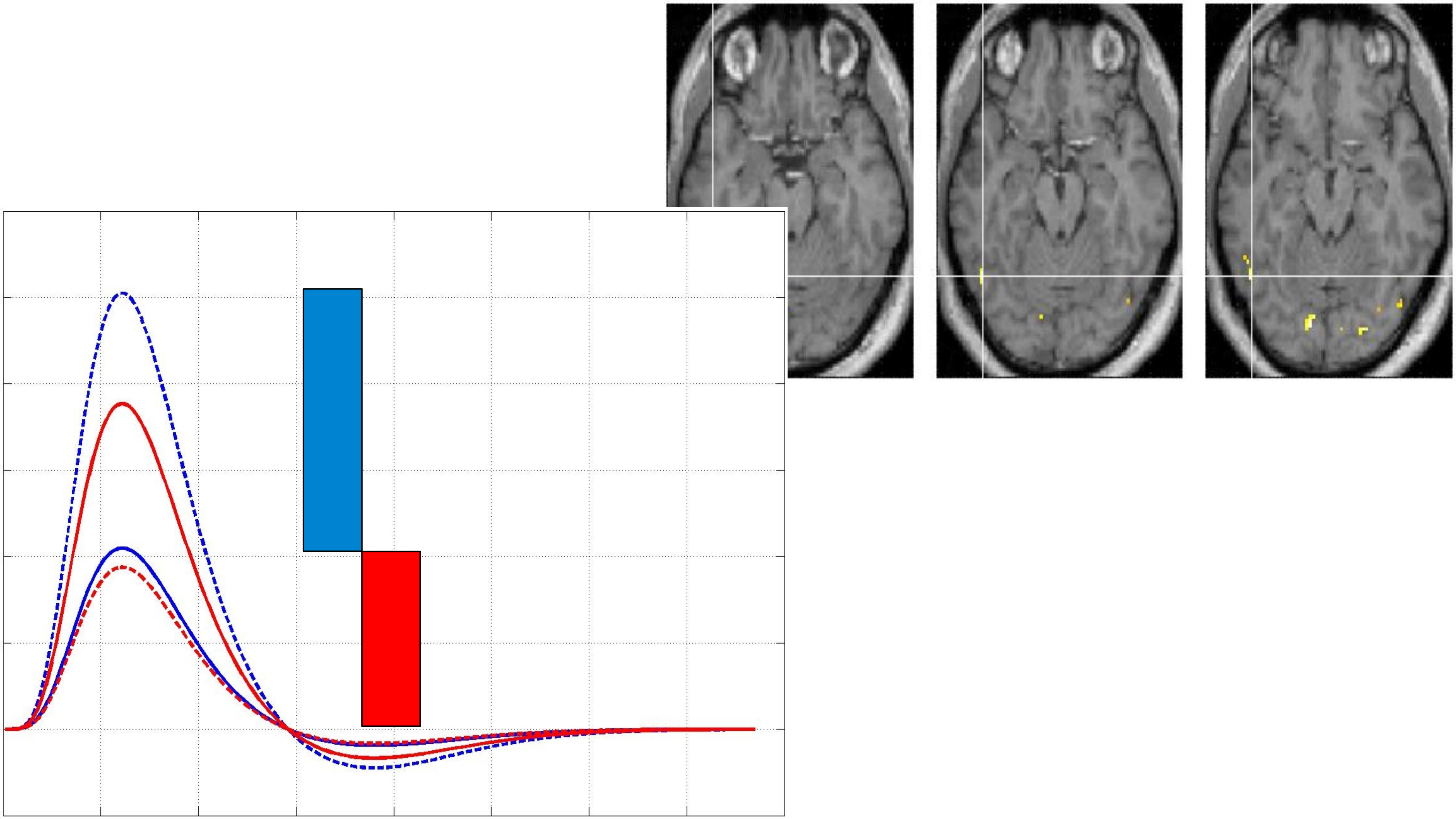
<u>Use contrasts</u>

- Main effects
- Fame: [1 1 -1 -1 0 0 ...]
- Rep: [1 -1 1 -1 0 0 ...]
- Interaction [1 -1 -1 1 0 0 ..]



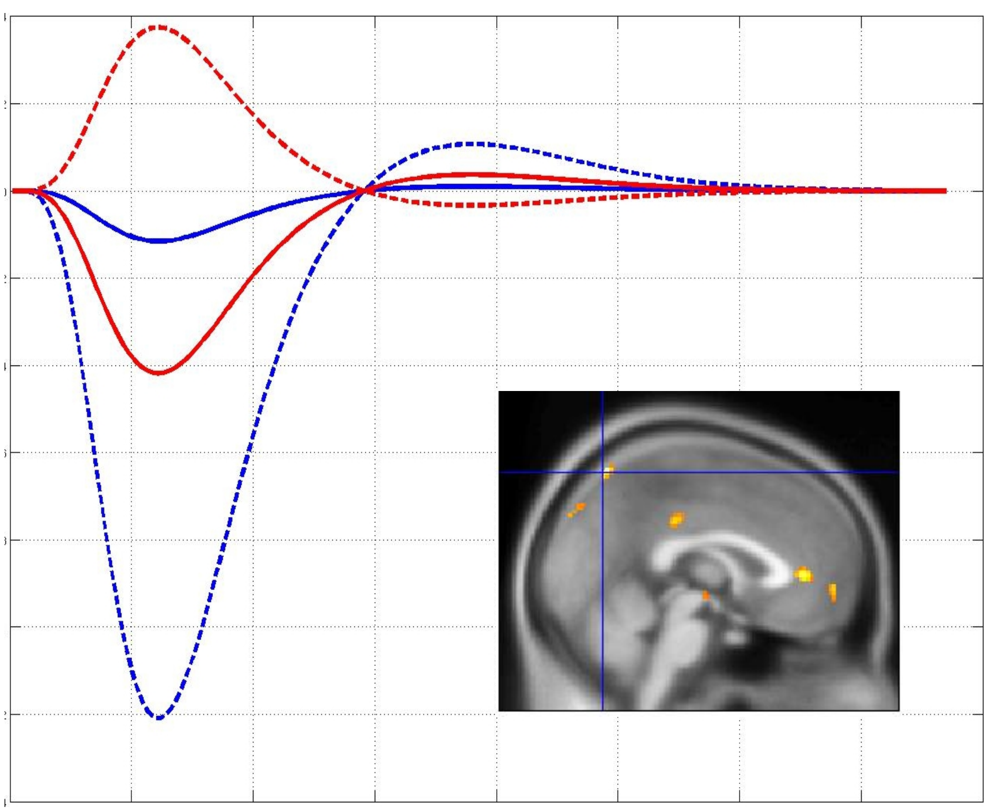
(1 -1) – (1 -1)

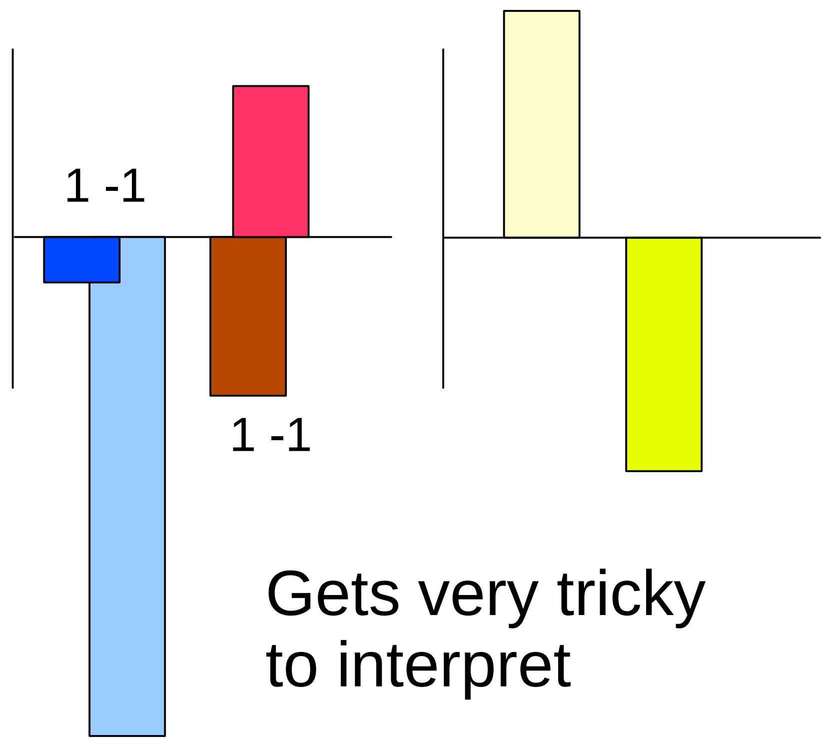
Always check interactions



Always check interactions

- Search in areas where all regressors are positive or all negative (i.e. use masking) otherwise ...



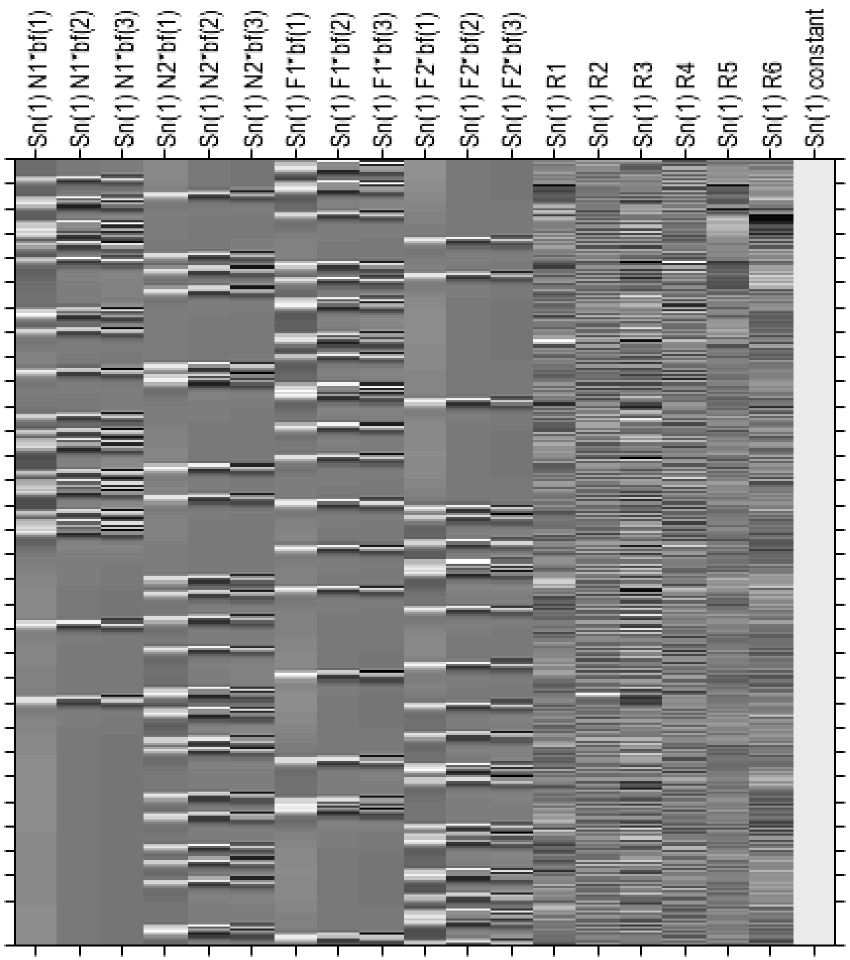


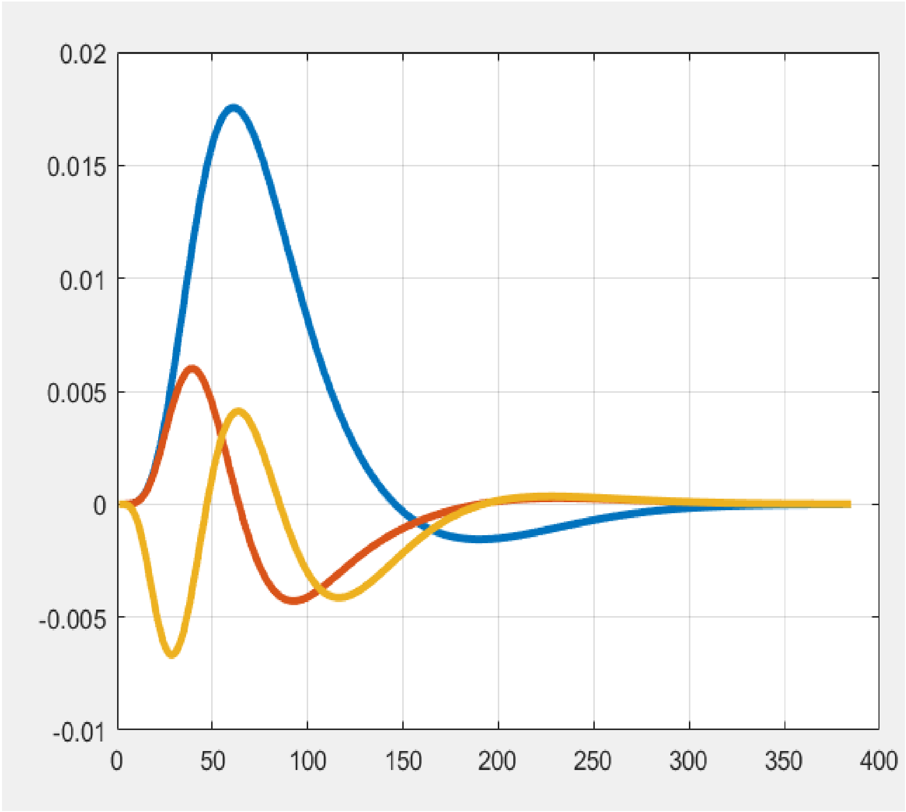
Gets very tricky to interpret

Plot the results !

A more complete model

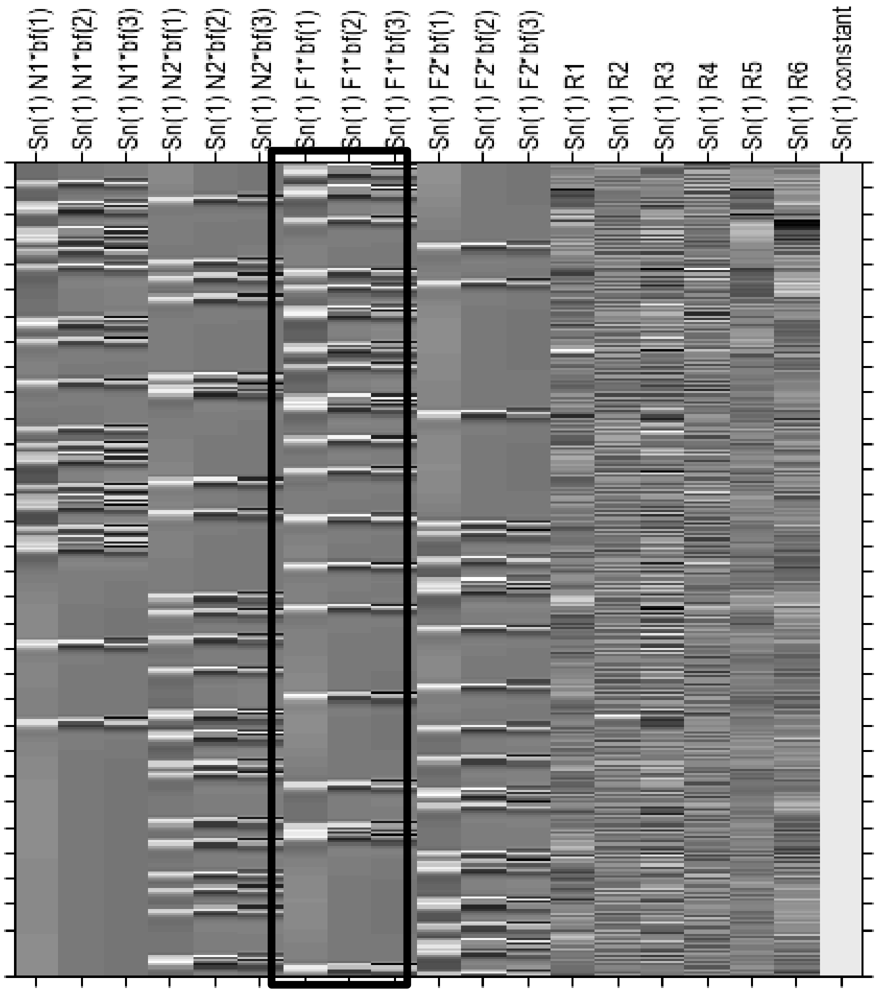
- Same design as before but added hrf derivatives.

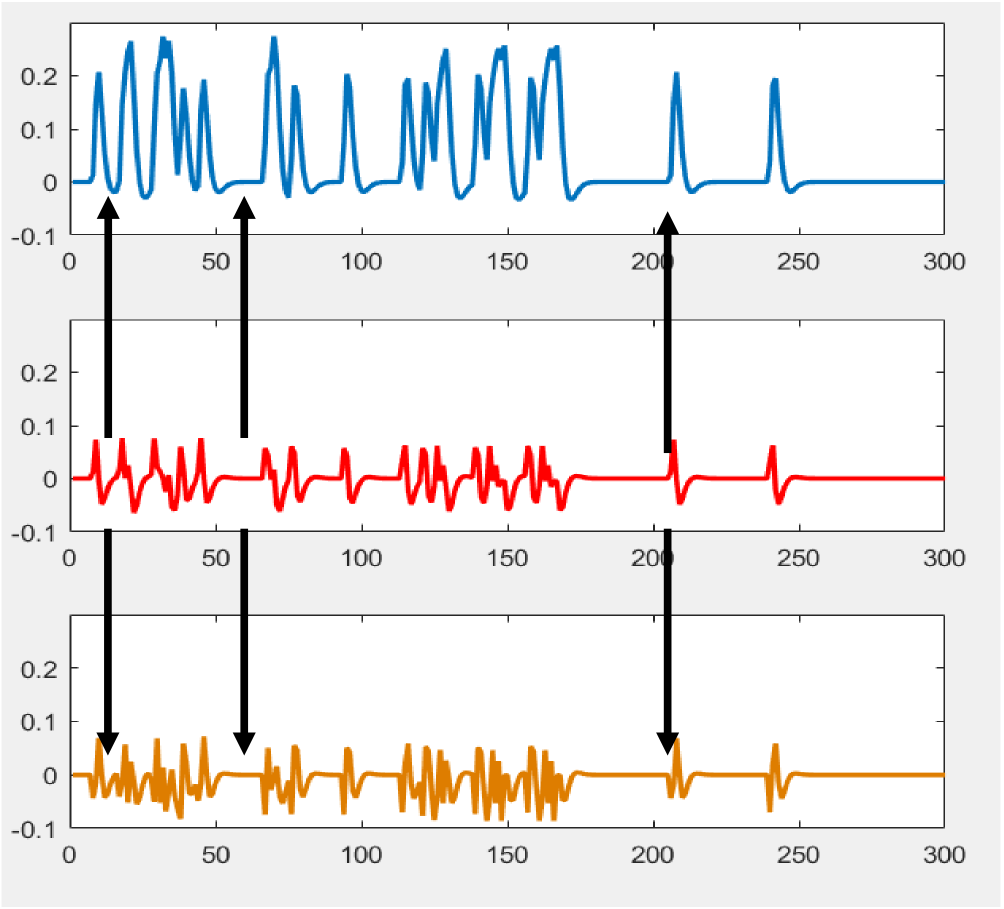




A more complete model

- Same design as before but added hrf derivatives.

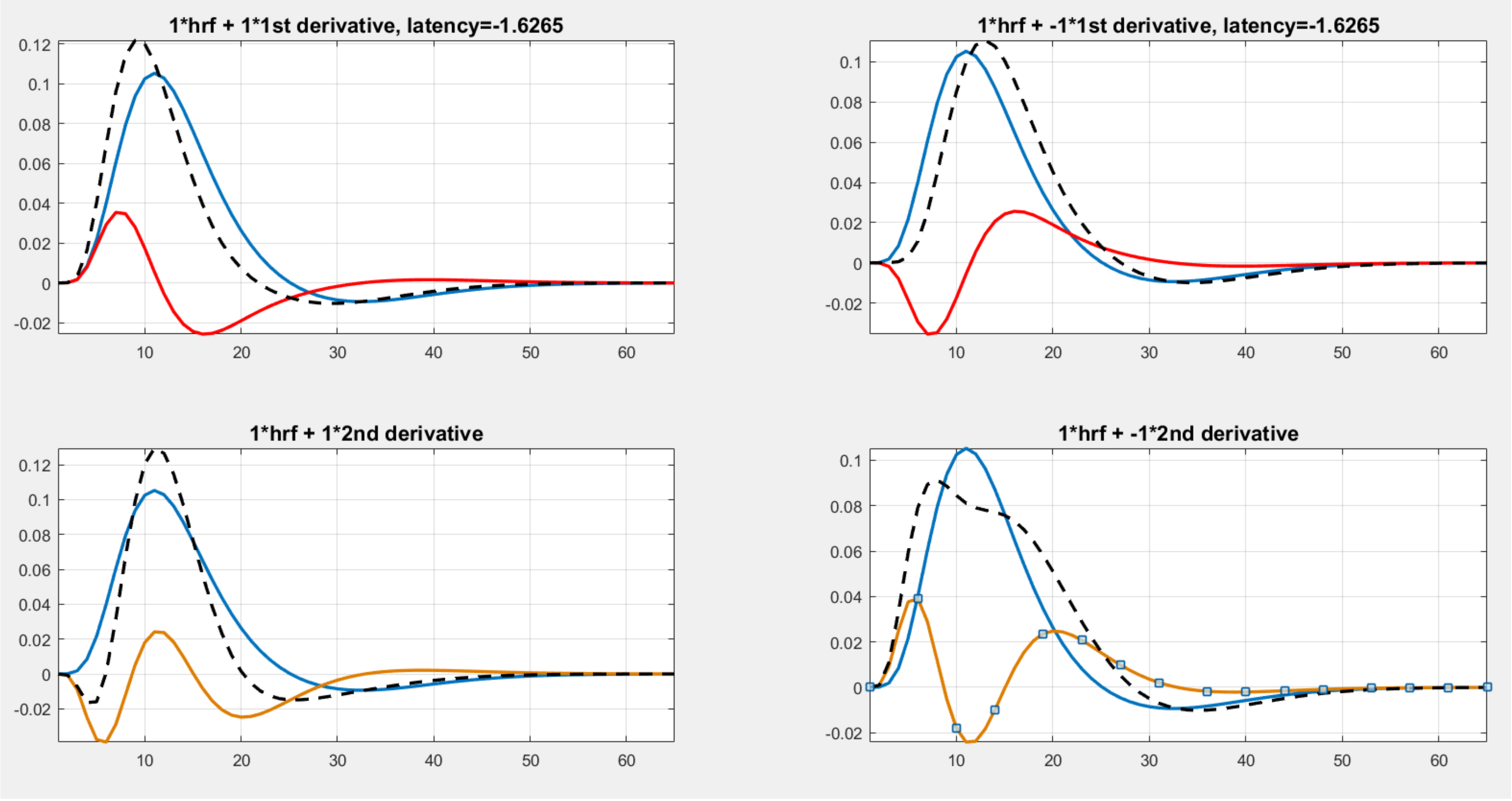




3 regressors all at the same time (diff. values)

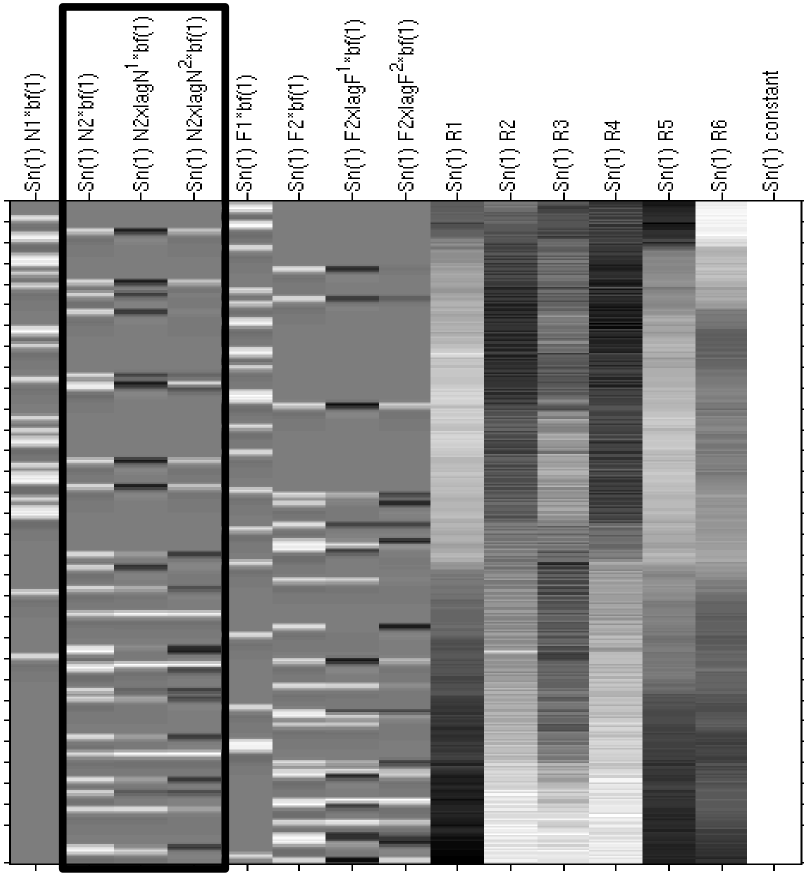
A more complete model

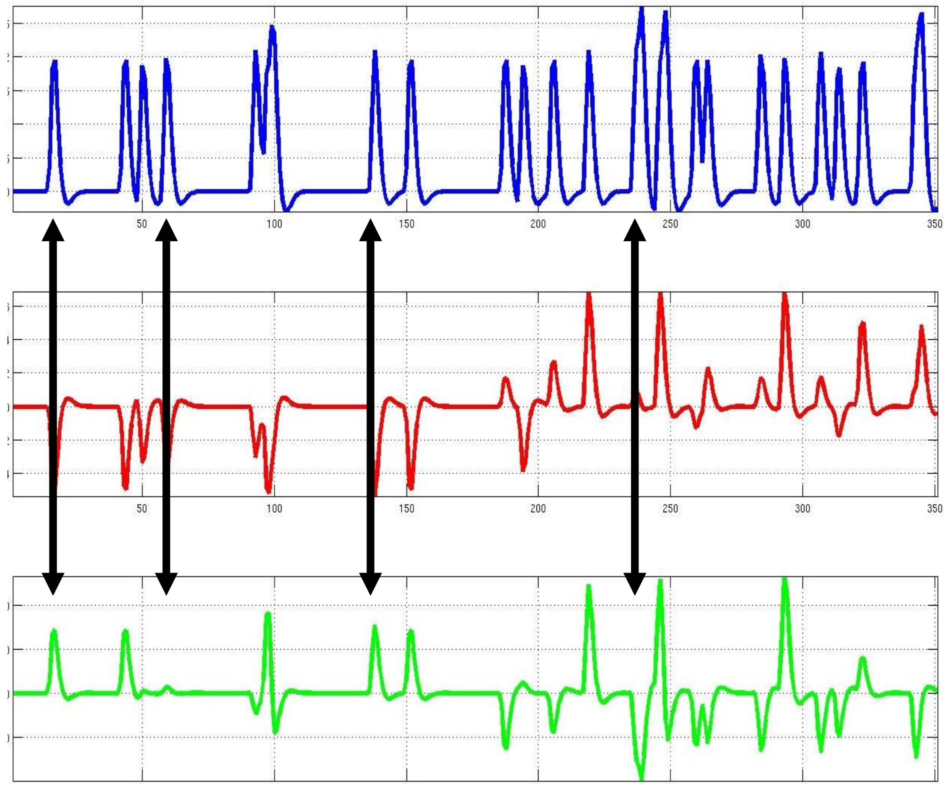
- What's a linear model already?
- We simply have to add the columns * coefficients betas



A more complex model

- Same design as before but added a 'parametric' regressor – here the lag and lag^2 between presentations

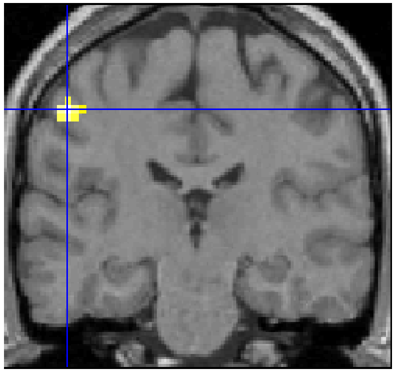


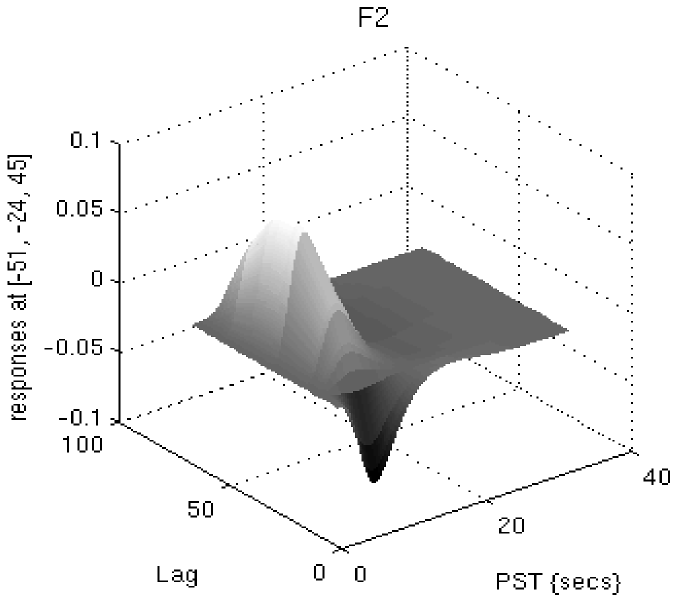


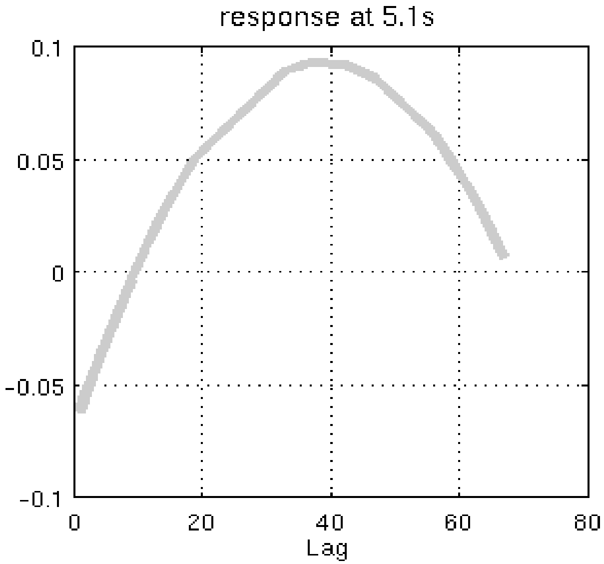
3 regressors all at the same time (diff. values)

A more complex model

- The parametric regressors express the amplitude of signal as a function of the lag, i.e. the signal amplitude changes from trial to trial







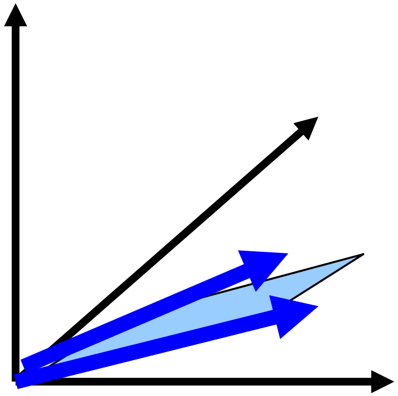
Issues with regressors

More Regressors: collinearity

- By default, SPM *orthogonalizes* parametric regressors making the regressors *non collinear (but you can disable that)*

- Three or more points are said to be collinear if they lie on a single straight line.

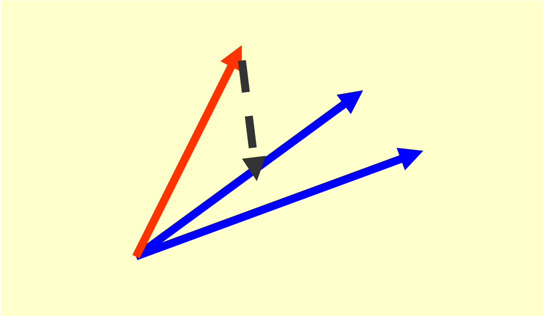
- Regressors are collinear if they are perfectly correlated (note corr of 2 vectors = cosθ)

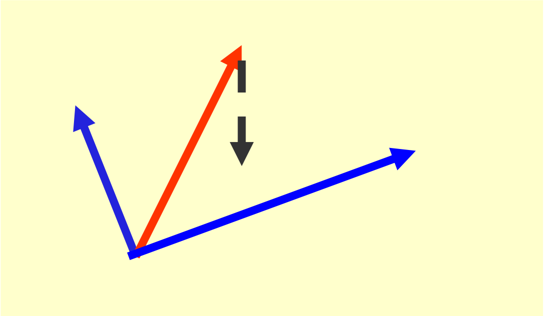


- Can make solution impossible
- Often make the model ok but individual regression values unstable
- Classical height and weight regression pblm

http://en.wikipedia.org/wiki/Multicollinearity
http://mathworld.wolfram.com/Collinear.html

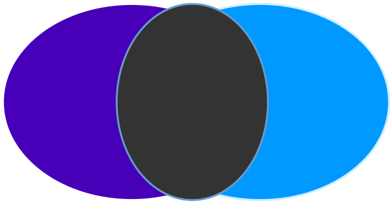
More Regressors: orthogonalization

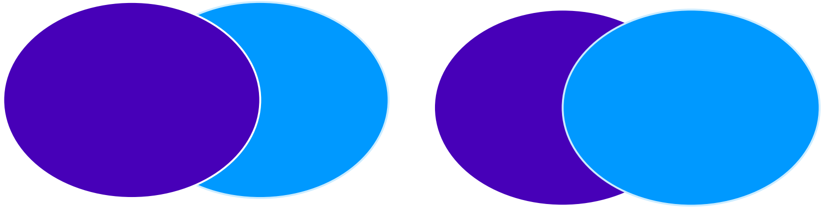




Lot of variance shared – because we look for the unique part of variance, the shared part goes into the error

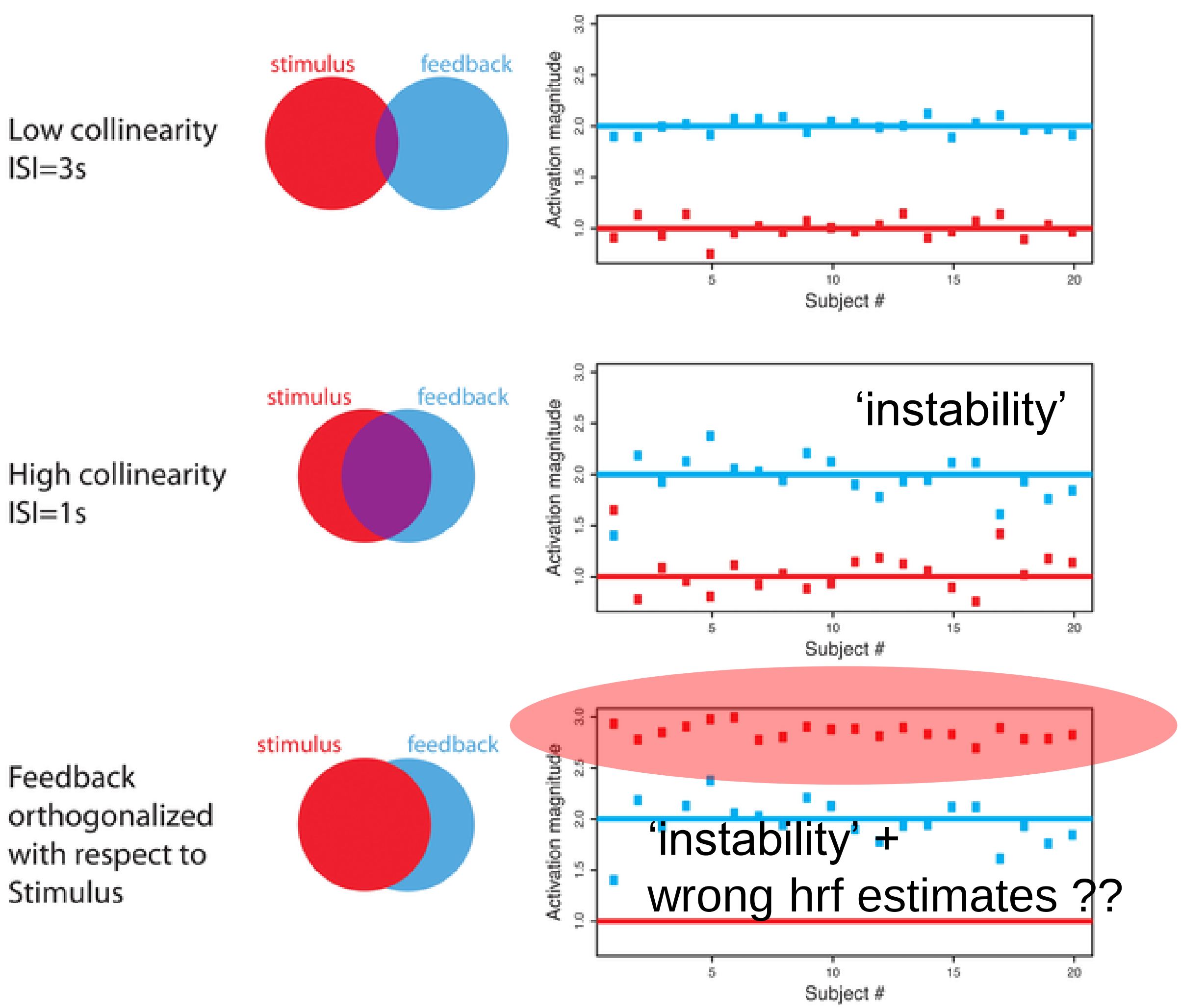
Orthogonalization (θ = 90°) removes shared variance BUT order matters !





http://en.wikipedia.org/wiki/Gram-Schmidt_process

More Regressors: orthogonalization



It's an interpretation issue, what do we want to look at ?

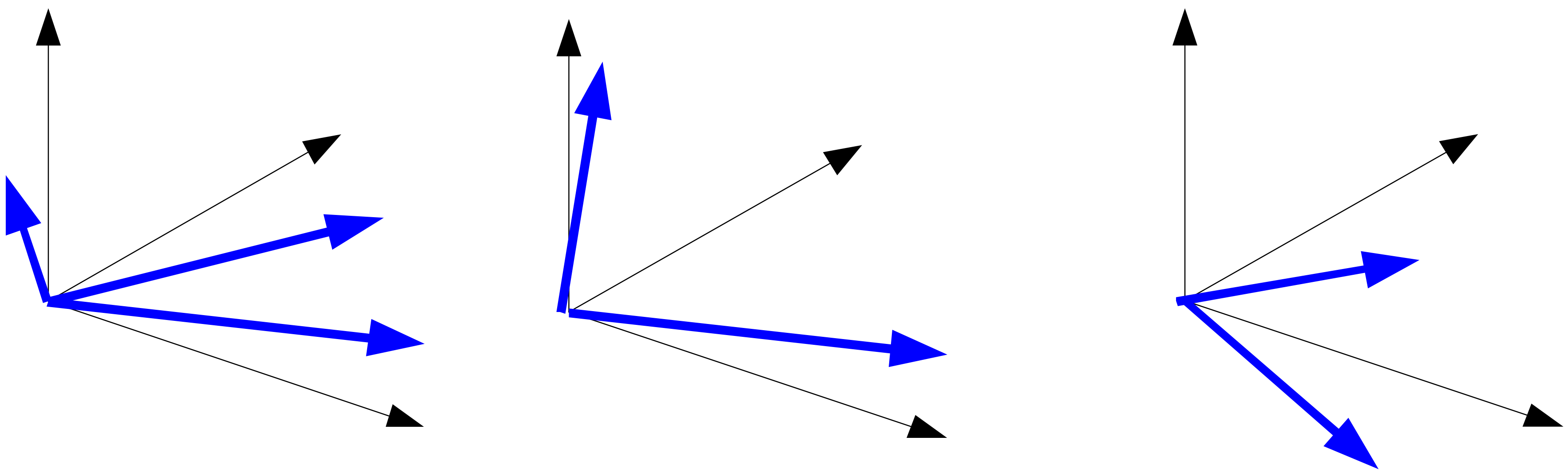
Y = x1b1 + x2b2
→ b1 is the 'baseline' activation
→ b2 is the modulation above baseline

Y = x1b1 + ⊥ x2b2
→ b1 is the mean activation
→ b2 is the modulation around the mean

Mumford JA, Poline JB, Poldrack RA (2015)

More regressors

- Linearly independent (X2 ≠ aX1), orthogonal (X1'Y2 = 0) and uncorrelated (X1-mean(X1))'(X2-mean(X2))=0) variables



More regressors

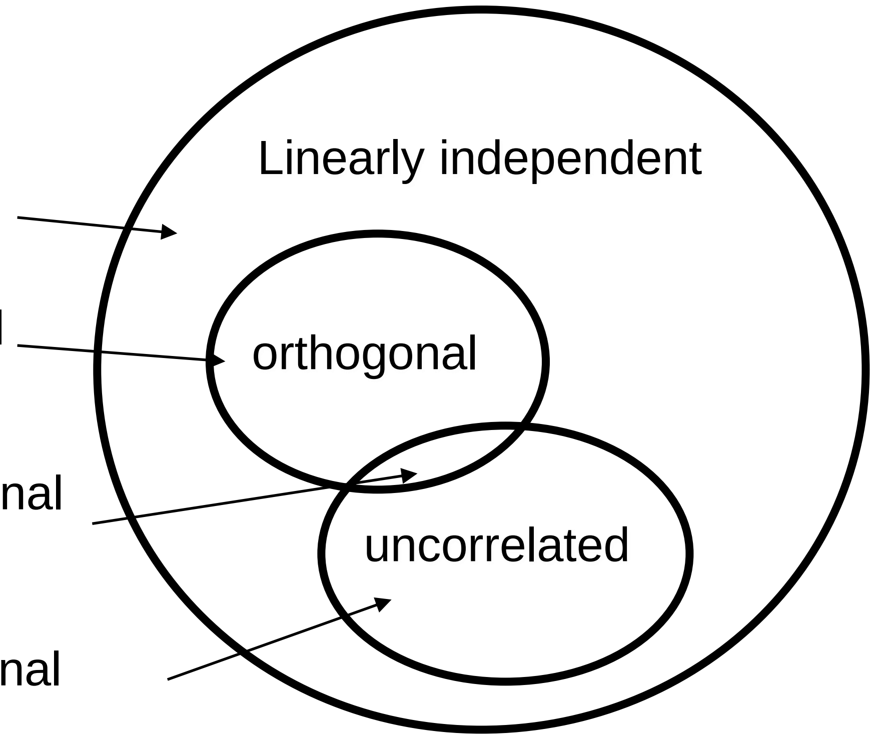
- Linearly independent (X2 ≠ aX1), orthogonal (X1'Y2 = 0) and uncorrelated (X1-mean(X1))'(X2-mean(X2))=0) variables

[1 1 2 3] and [2 3 4 5]
Independent, correlated, not orthogonal

[1 -5 3 -1] and [5 1 1 3]
Independent, correlated and orthogonal

[-1 -1 1 1] and [1 -1 1 -1]
Independent, uncorrelated and orthogonal

[0 0 1 1] and [1 0 1 0]
Independent, uncorrelated, not orthogonal



http://www.jstor.org/pss/2683250

Summary

- Linear model: $y = \beta 1 x1 + \beta 2 x2$ (output = additivity and scaling of input)
- GLM: Y = XB+E (matrix formulation, works for any statistics, express the data Y as a function of the design matrix X)
- More regressors is better as it captures more of the signal but it may bring instability if regressors are collinear (and cost df). Orthogonalization makes sense for parametric regressors but not always.

Selected References

- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp*. 2, 189–210. doi: 10.1002/hbm.460020402
- Friston, K. J., Holmes, A., Poline, J.-B., Grasby, P., Williams, S., Frackowiak, R. S. J., et al. (1995). Analysis of time-series revised. *Neuroimage* 2, 45–53. doi: 10.1006/nimg.1995.1007
- Worsley, K. J., and Friston, K. (1995). Analysis of fMRI time-series revised - again. *Neuroimage* 2, 173–181. doi: 10.1006/nimg.1995.1023
- Poline, J.-B., and Brett, M. (2012). The general linear model and fMRI: does love last forever? *Neuroimage* 62, 871–880. doi: 10.1016/j.neuroimage.2012.01.133
- Pernet CR (2014) Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers. *Front. Neurosci*. **8**:1. doi: 10.3389/fnins.2014.00001
- Mumford JA, Poline JB, Poldrack RA (2015) Orthogonalization of Regressors in fMRI Models. PLOS ONE 10(4): e0126255. doi: 10.1371/journal.pone.0126255
- Rodgers, J. L., Nicewander, W. A., and Toothaker, L. (1984). Linearly independent, orthogonal, and uncorrelated variables. *Am. Stat*. 38, 133–134.