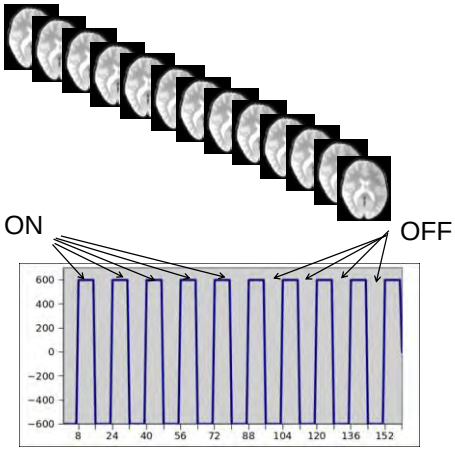
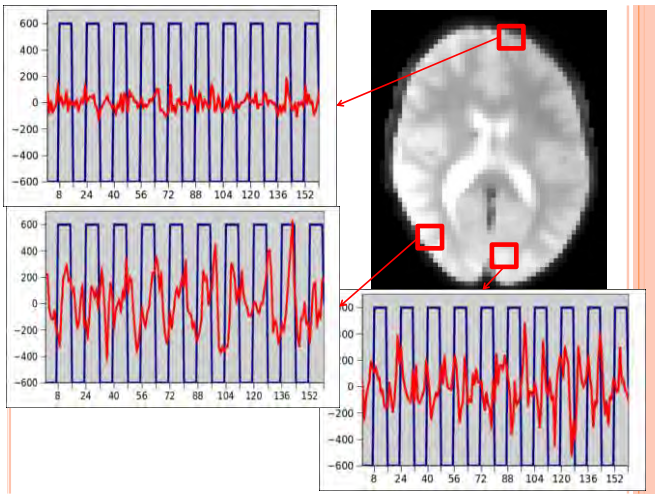
Single Subject Analysis





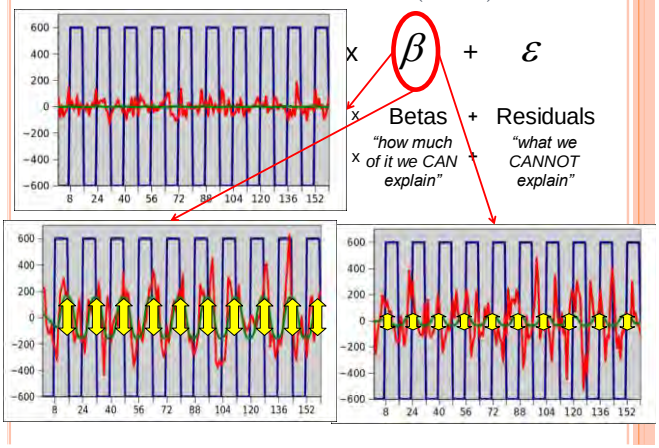
The General Linear Model (GLM)

$$y = X \times \beta + \varepsilon$$

fMRI Signal	=	Design Matrix	x	Betas	+	Residuals
"our data"	=	*"what we CAN explain"*	x	*"how much of it we CAN explain"*	+	*"what we CANNOT explain"*

→ If a voxel "cares" about the task, then you should be able to predict its time-course from the onset and offset of the task.
→ β captures "how much" of 'y' your task/design is predicting.

The General Linear Model (GLM)



$$\times \beta + \varepsilon$$

Betas	+	Residuals
x *"how much of it we CAN explain"*	+	*"what we CANNOT explain"*

The General Linear Model (GLM)

$$y = X \times \beta + \varepsilon$$

fMRI Signal	=	Design Matrix	x	Betas	+	Residuals
"our data"	=	*"what we CAN explain"*	x	*"how much of it we CAN explain"*	+	*"what we CANNOT explain"*

<u>The 2 components of statistical analysis:</u>

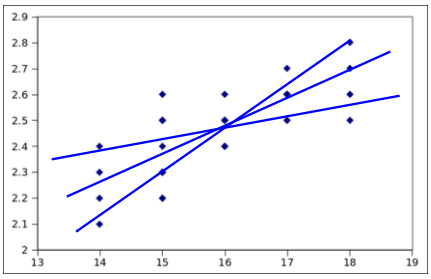
1. **Estimation**: "how big" is the effect?
2. **Inference**: is the effect we estimated statistically significant?

Estimation: The General Linear Model

- Estimation: How do we find β?
  - Many ways: Ordinary least squares (OLS), (feasible) Generalized Least Squared (GLS), Smoothing, "Sandwich", ...
  - In its simplest formulation, we use the OLS*

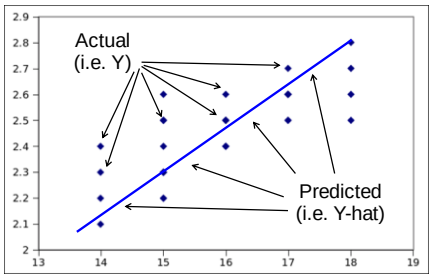
Estimation: Ordinary Least Squares (OLS)

e.g. What is the relationship tying years of MATLB experience and coding efficiency?



Estimation: Ordinary Least Squares (OLS)

e.g. What is the relationship tying years of MATLB experience and coding efficiency?



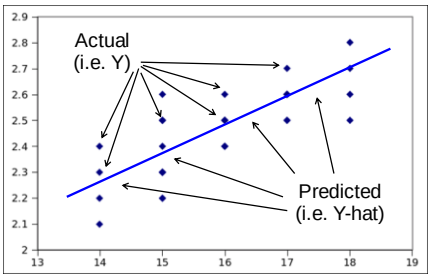
More formally, minimize:

$$\sum_{i=1}^{n} (Y - \hat{Y})^2$$

$$\sum_{i=1}^{n} (Y - X\beta)^2$$

Estimation: Ordinary Least Squares (OLS)

e.g. What is the relationship tying years of MATLB experience and coding efficiency?



More formally, minimize:

$$\sum_{i=1}^{n} (Y - \hat{Y})^2$$

$$\sum_{i=1}^{n} (Y - X\beta)^2$$

Least Squares Difference

Minimize error

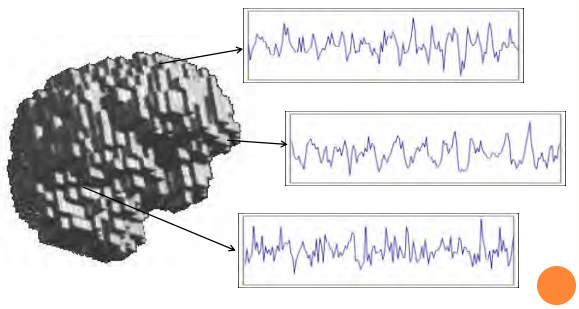
$$\sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \beta_1 X_i)^2$$

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

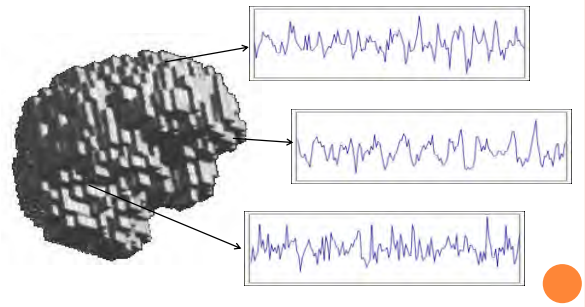
The General Linear Model (GLM)

$$y_i = \beta_0 + \beta_1 X + \varepsilon$$



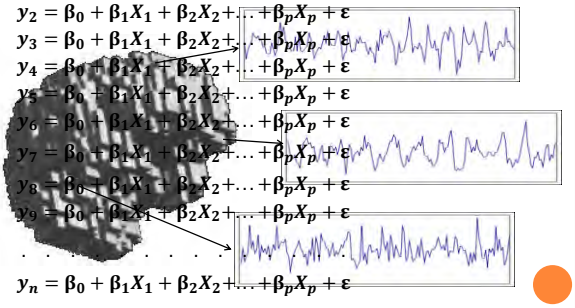
The General Linear Model (GLM)

$$y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$



The General Linear Model (GLM)

$$y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$y_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$y_3 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$y_4 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$y_5 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$y_6 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$y_7 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$y_8 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$y_9 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$
$$\ldots$$
$$y_n = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

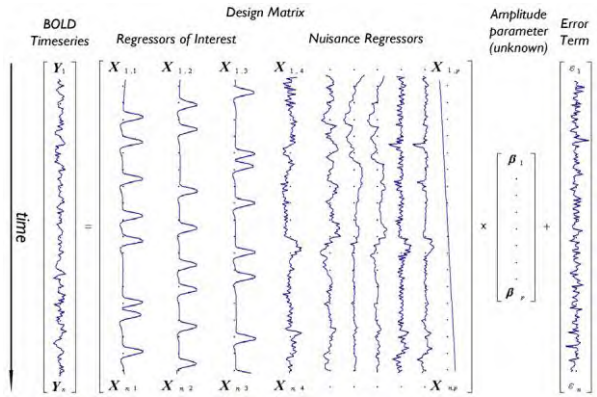


Voxel-wise GLM

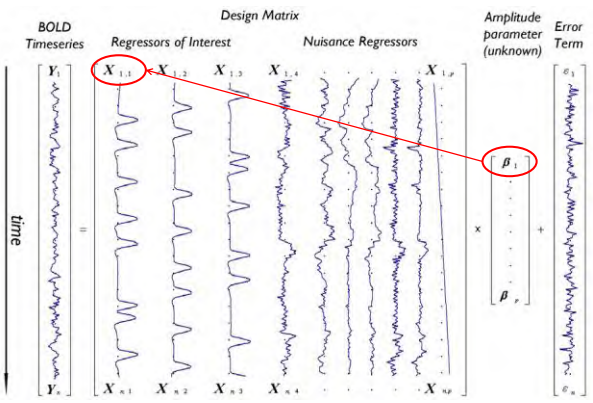
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots & x_{2,p} \\ x_{3,1} & x_{3,2} & x_{3,3} & \cdots & x_{3,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & x_{n,3} & \cdots & x_{p,n} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

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

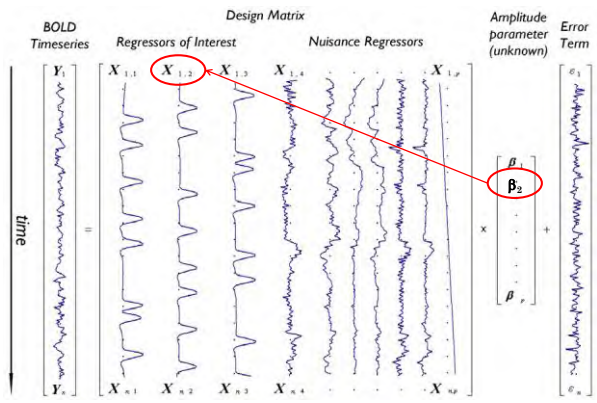
Voxel-wise GLM



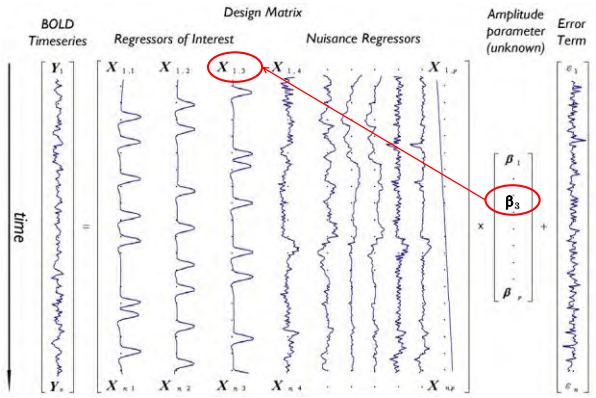
Voxel-wise GLM



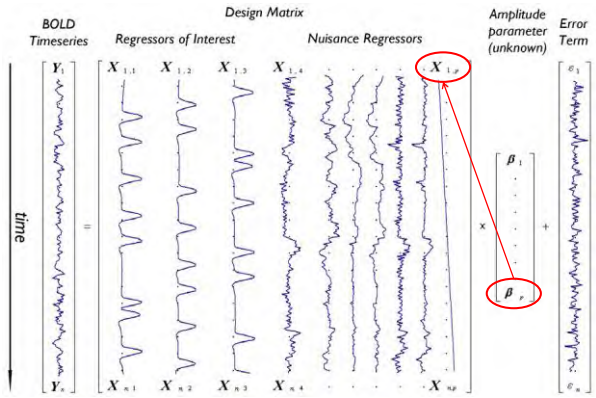
Voxel-wise GLM



Voxel-wise GLM



Voxel-wise GLM



Estimation: Ordinary Least Squares (OLS)

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

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

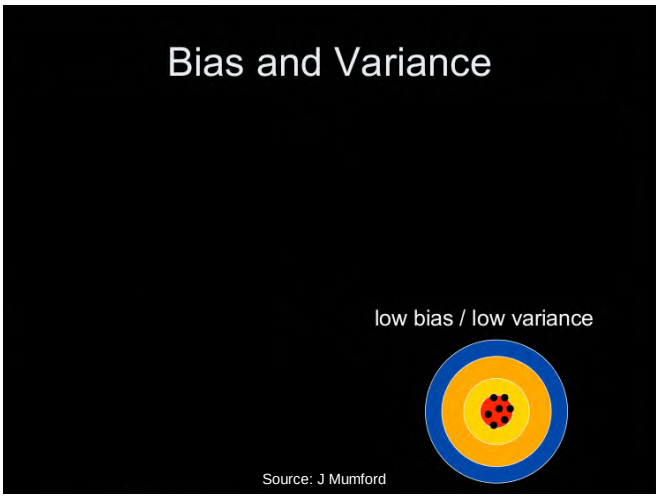
$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T Y$$
$$Var(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1}$$

Estimation: Gauss Markov Assumptions

- Under the following assumptions, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

BEST	i.e., the estimator with least variance (or the "most precise" estimator)
LINEAR	i.e., effects add up linearly
UNBIASED	i.e., the estimation, on average, yield the correct parameter

Bias and Variance



Estimation: Gauss Markov Assumptions

- Under the following assumptions, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

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

$$E(\varepsilon_i) = 0$$ — Error has mean 0

$$\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Error is Homoschedastic

$$Var(\varepsilon_i) = \sigma^2$$ — Errors have same variance

Errors are not Autocorrelated

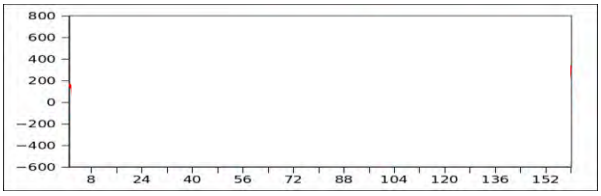
$$Cov(\varepsilon_i, \varepsilon_j) = 0$$ — Errors are independent of each other

Inference

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\hat{\beta})}} = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \sim T(N-p)$$

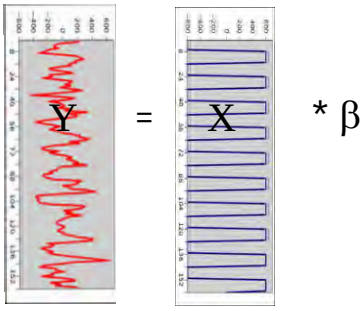
A simple experiment: the FFA

- Ss: 1 participant
- Design: 16s Faces, 16s Houses (block design)
- Take Voxel 15,19,18, in right posterior temporal cortex



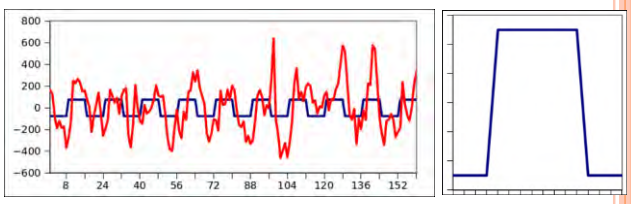
- Does this voxel "care" about faces?

Model I: Boxcar Regressor

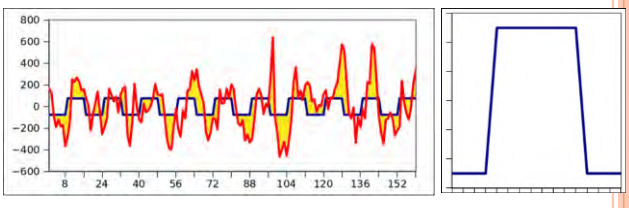


$* \beta$

Model I: Boxcar Regressor

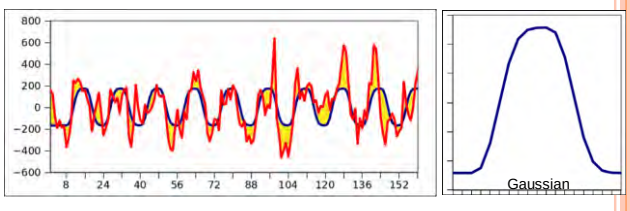


Model I: Boxcar Regressor



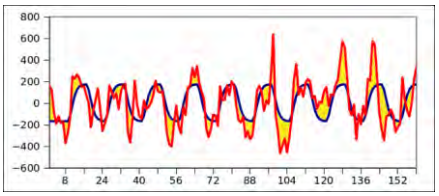
$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{150.99}{202.44 \times 0.16} = 4.66$$

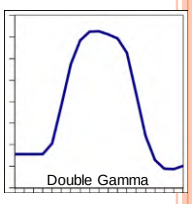
Model II: Convolved Boxcar Regressor



$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{267.02}{166.70 \times 0.16} = 10.01$$

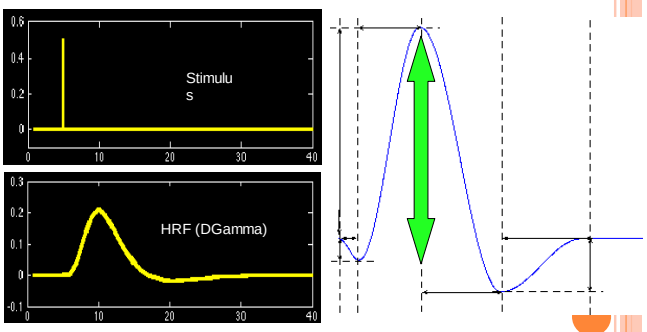
Model II: Convolved Boxcar Regressor



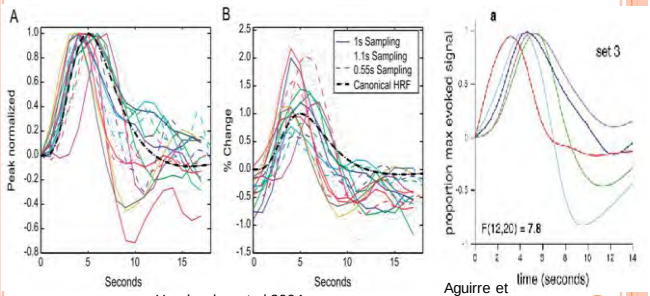


$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{344.74}{168.67 \times 0.16} = 12.77$$

Canonical HRF



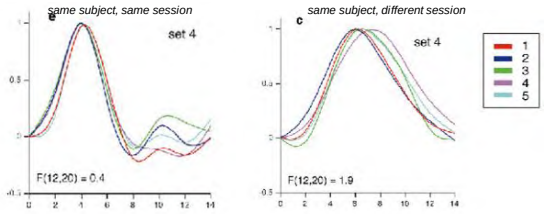
HRF Variability Across Subjects



Handwerker et al 2004

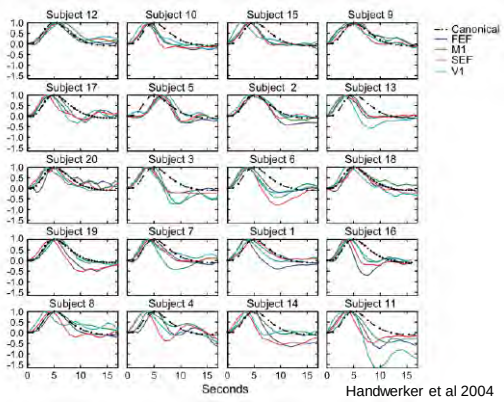
Aguirre et Al 1998

HRF Variability Within Subjects



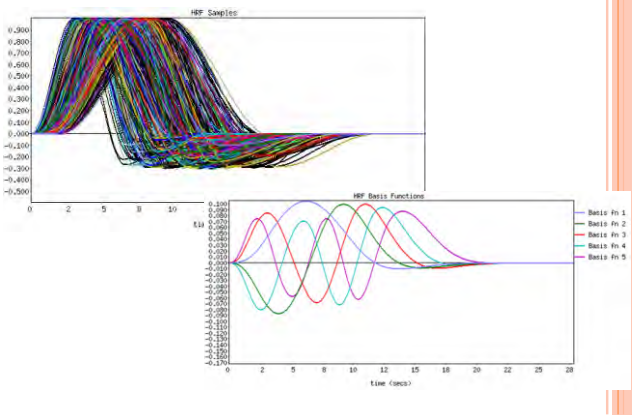
Aguirre et al 1998

HRF Variability Across Brain Area

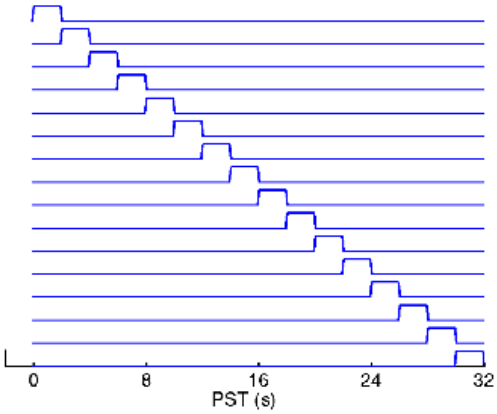


Handwerker et al 2004

Basis Functions I

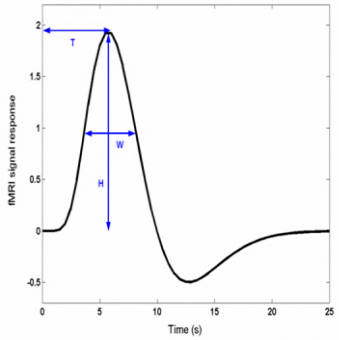


Basis Functions II: FIR

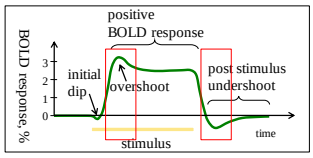


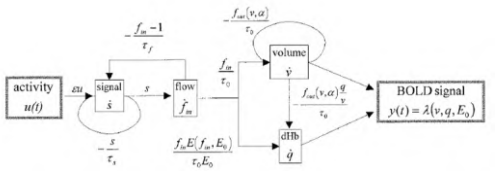
Non Linear Approaches

- Model all 3 parameters (T,W,H)



Non Linear Approaches: Balloon Model





What Convolution? - Recap

- The problem:
  - Bias/Variance trade off
  - Assume known shape v estimate
  - **X-Matrix mis-specification**: Incorrect HRF will increase σ^2

What Convolution? - Recap

Canonical HRF:
- Pros: 1 parameter only, easy to interpret; easy to integrate at the second level.
- Cons: Bias in favour of a specific filter

- **Basis Sets Functions:**
  - Pros: Flexible, can capture greater variability, unexpected and subject specific differences.
  - Cons: many parameters to estimate (degrading dfs); difficult to interpret; how to aggregate at the 2nd level?
- **Non-Linear Systems:**
  - Potentially very flexible and may better capture parameters and vacolature
  - Computationally very demanding, parameters not obvious to interpret and aggregate across subjects (e.g., balloon model)

X-Matrix Assumptions

- Deterministic & known
- No regressor is a linear transformation of one (or more) regressors

MULTICOLLINEARITY

- No column of the X matrix can be obtained by linear combination of one or more other columns.
- The issue is not whether there is/isn't multicollinearity, but rather, "how much" is there.
- Two main consequences:
  - Decreased efficiency of the design degrades power.
  - Parameter estimates can vary wildly, even in the wrong direction!

Source: J Mumford

Why is it a problem?

$$\begin{pmatrix} 12 \\ 12 \\ 12 \\ 3 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \\ \beta_3 \\ \beta_3 \\ \beta_3 \end{pmatrix}$$

Source: J Mumford

Why is it a problem?

$$\begin{pmatrix} 12 \\ 12 \\ 12 \\ 3 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \\ \beta_3 \\ \beta_3 \\ \beta_3 \end{pmatrix}$$

- There are an infinite # of solutions for β_1 and β_2

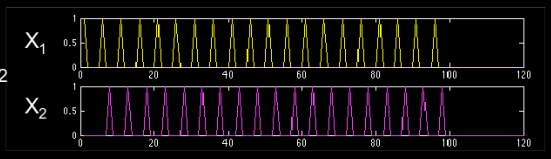
$\hat{\beta}_1 = 10, \hat{\beta}_2 = 2 \quad \hat{\beta}_1 = 100, \hat{\beta}_2 = -88$ …etc

Source: J Mumford

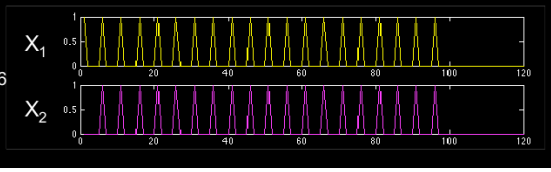
Collinearity illustration

$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon \quad \beta_0 = 1, \beta_1 = 2, \beta_2 = 4$

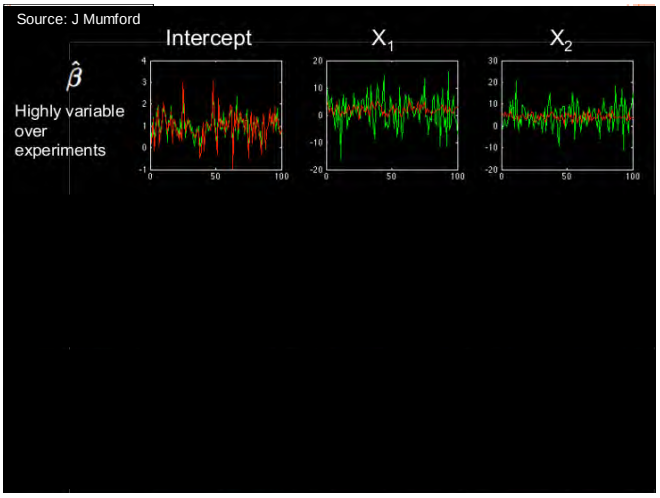
Cor= -0.2



Cor=0.96



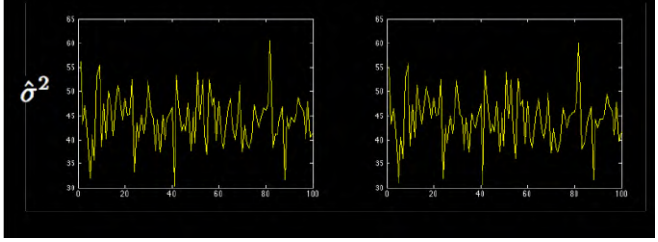
Source: J Mumford



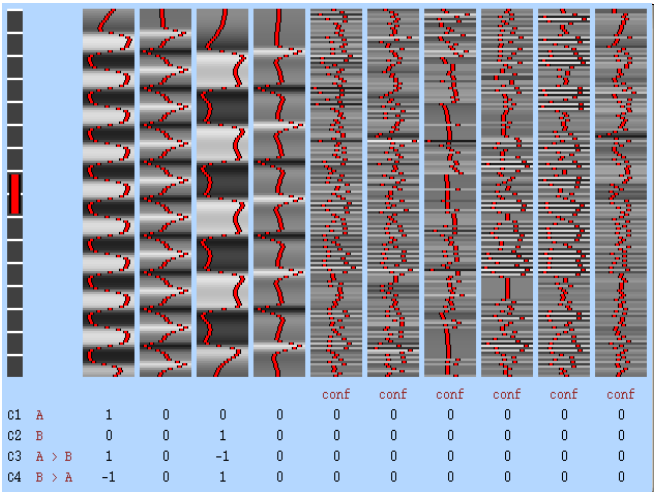
Source: J Mumford

Residuals don't change

- The designs explain the same amount of variability



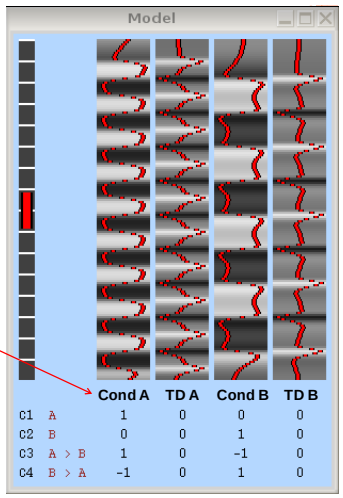
Contrasts



Contrasts

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\beta)}}$$

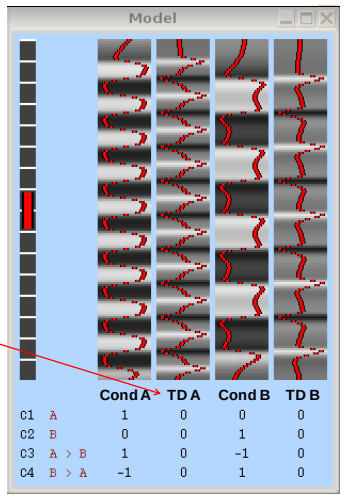
$\hat{\beta}_1$
$\hat{\beta}_2$
$\hat{\beta} = \hat{\beta}_3$
$\hat{\beta}_4$
:
$\hat{\beta}_p$



Contrasts

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\beta)}}$$

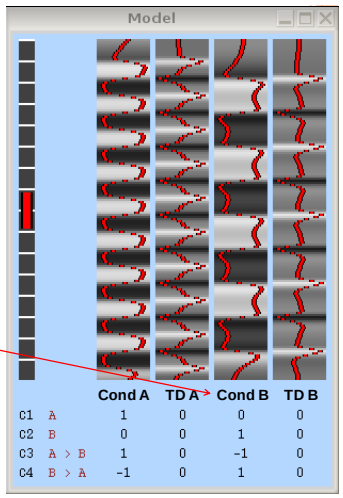
$\hat{\beta}_1$
$\hat{\beta}_2$
$\hat{\beta} = \hat{\beta}_3$
$\hat{\beta}_4$
:
$\hat{\beta}_p$



Contrasts

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\beta)}}$$

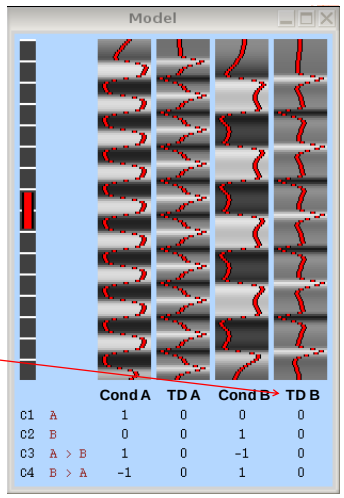
$\hat{\beta}_1$
$\hat{\beta}_2$
$\hat{\beta} = \hat{\beta}_3$
$\hat{\beta}_4$
:
$\hat{\beta}_p$



Contrasts

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\beta)}}$$

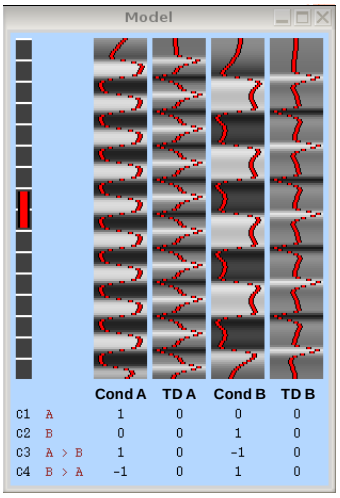
$\hat{\beta}_1$
$\hat{\beta}_2$
$\hat{\beta} = \hat{\beta}_3$
$\hat{\beta}_4$
:
$\hat{\beta}_p$



Contrasts

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

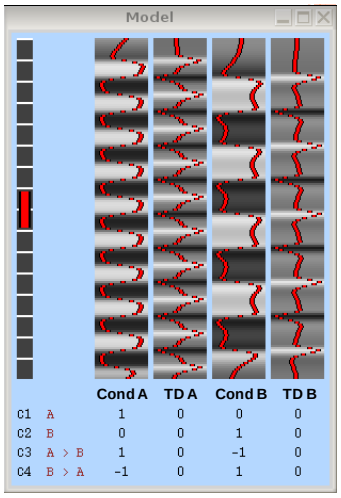
$$\hat{\beta} = \begin{matrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \hat{\beta}_4\\ \vdots\\ \hat{\beta}_p\end{matrix} \qquad c = \begin{matrix}c_1\\ c_2\\ c_3\\ c_4\\ .\\ c_p\end{matrix}$$



Contrasts

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

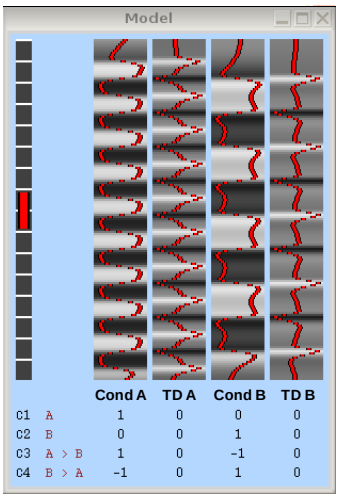
$$\hat{\beta} = \begin{matrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \hat{\beta}_4\\ \vdots\\ \hat{\beta}_p\end{matrix} \qquad c = \begin{matrix}1\\ 0\\ 0\\ 0\\ .\\ 0\end{matrix}$$



Contrasts

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

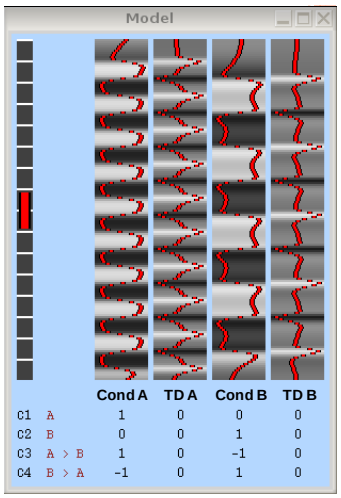
$$\hat{\beta} = \begin{matrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \hat{\beta}_4\\ \vdots\\ \hat{\beta}_p\end{matrix} \qquad c = \begin{matrix}0\\ 0\\ 1\\ 0\\ .\\ 0\end{matrix}$$



Contrasts

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

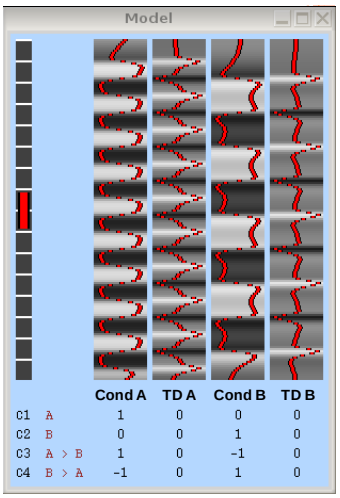
$$\hat{\beta} = \begin{matrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \hat{\beta}_4\\ \vdots\\ \hat{\beta}_p\end{matrix} \qquad c = \begin{matrix}1\\ 0\\ -1\\ 0\\ .\\ 0\end{matrix}$$



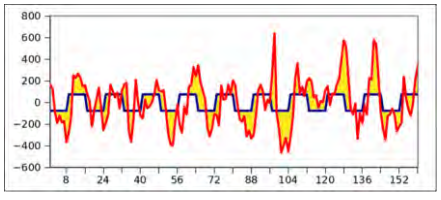
Contrasts

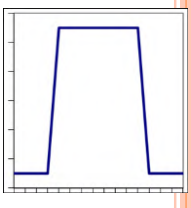
$$t = \frac{\hat{c^T\beta}}{\sqrt{\hat{Var}(c^T\beta)}}$$

$$\hat{\beta} = \begin{matrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \hat{\beta}_4\\ \vdots\\ \hat{\beta}_p\end{matrix} \qquad c = \begin{matrix}-1\\ 0\\ 1\\ 0\\ .\\ 0\end{matrix}$$



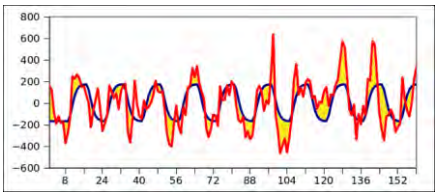
Model I: Boxcar Regressor

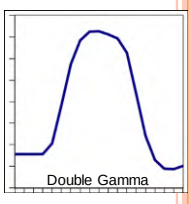




$$t(df) = \frac{(X^TX)^{-1}X^TY}{\sqrt{\hat{\sigma}^2(X^TX)^{-1}}} \qquad t(159) = \frac{150.99}{202.44 \times 0.16} = 4.66$$

Model II: Convolved Boxcar Regressor





$$t(df) = \frac{(X^TX)^{-1}X^TY}{\sqrt{\hat{\sigma}^2(X^TX)^{-1}}} \qquad t(159) = \frac{344.74}{168.67 \times 0.16} = 12.77$$

Estimation: Gauss Markov Assumptions

- Under the following assumptions, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

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

→ $E(\varepsilon_i) = 0$ — Error has mean 0

$$\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Error is Homoschedastic

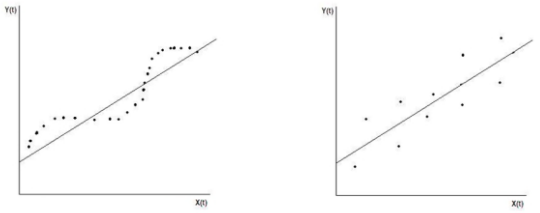
→ $Var(\varepsilon_i) = \sigma^2$ — Errors have same variance

Errors are not Autocorrelated

→ $Cov(\varepsilon_i, \varepsilon_j) = 0$ — Errors are independent of each other

Temporal Structure

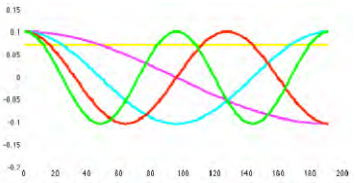
- If residuals are correlated, estimator is still unbiased but the variance estimate is not; and df are less than nominal.

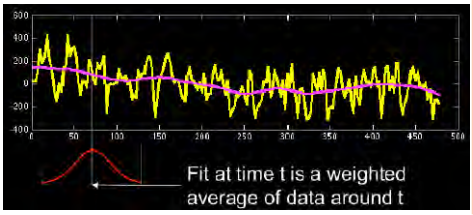


Temporal Structure

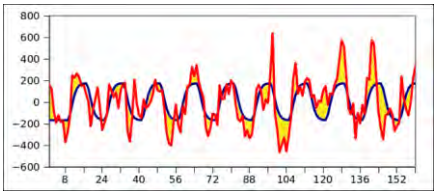
- If residuals are correlated, estimator is still unbiased but the variance estimate is not; and df are less than nominal.
- This can affect your T-test either way:
  i. Positive autocorrelation: underestimate of variance, T-tests artificially inflated (actual α up to 0.16 for a nominal of 0.05).
  ii. Negative autocorrelation: overestimate of variance, T-test artificially reduced.
- Two strategies:
  i. High Pass filtering
  ii. Autocorrelation correction ("pre-whitening"; "colouring")

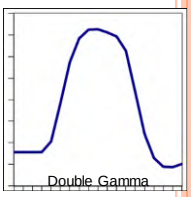
HP Filtering





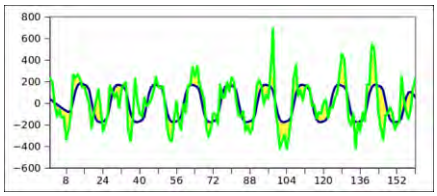
Model II: Convolved Boxcar Regressor

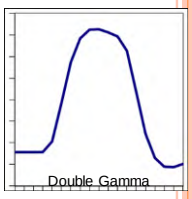




$$t(df) = \frac{(X^TX)^{-1}X^TY}{\sqrt{\hat{\sigma}^2(X^TX)^{-1}}} \qquad t(159) = \frac{344.74}{168.67 \times 0.16} = 12.77$$

Model III: High Pass Filter





$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{349}{154.08 \times 0.21} = 11.05$$

Autocorrelation

GM Assumption:

Errors are not Autocorrelated

$$\varepsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 I) \qquad Cov(\varepsilon_i, \varepsilon_j) = 0$$

Errors are independent of each other

fMRI Data:

$$\varepsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 V) \qquad Cov(\varepsilon_i, \varepsilon_j) = \sigma^2 V$$

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

Errors are **NOT** independent of each other

Autocorrelation

GM Assumption:

Errors are not Autocorrelated

$$\varepsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 I) \qquad Cov(\varepsilon_i, \varepsilon_j) = 0$$

Errors are independent of each other

$$\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} \begin{bmatrix} \sigma^2 & Cov(\varepsilon_1,\varepsilon_2) & \dots & Cov(\varepsilon_1,\varepsilon_n) \\ Cov(\varepsilon_2,\varepsilon_1) & \sigma^2 & \dots & Cov(\varepsilon_2,\varepsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(\varepsilon_n,\varepsilon_1) & Cov(\varepsilon_n,\varepsilon_2) & \dots & \sigma^2 \end{bmatrix}$$

Autocorrelation: Pre-whitening (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it:

1. Find matrix **K** such that **KVK^T=I**
2. Premultiply the GLM equation by **K:**

$$K^{-1}Y = K^{-1}X\beta + K^{-1}e \qquad (e = V\varepsilon)$$

$$\hat{\boldsymbol{\beta}} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

$$\hat{Var}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\sigma}}^2 (X^T V^{-1} X)^{-1}$$

Autocorrelation: Pre-whitening (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it**.

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 1

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]
[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

Autocorrelation: Pre-whitening (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 2

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]
[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

AUTOCORRELATION: PRE-WHITENING (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 3

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e

AUTOCORRELATION: PRE-WHITENING (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 8

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

[e01 e02 e03 e04 e05 e06

AUTOCORRELATION: PRE-WHITENING (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 9

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

[e01 e02 e03 e04 e05 e

AUTOCORRELATION: PRE-WHITENING (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel)
2. For each lag, smooth estimates spatially ('regularization')

AUTOCORRELATION: PRE-WHITENING (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

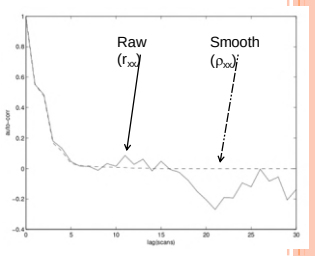
1. Estimate the autocorrelation structure (for every voxel)
2. For each lag, smooth estimates spatially ('regularization')
3. Smooth (temporally) the correlation estimate, for each voxel ('Tukey taper').

$$\hat{\rho}_{xx}(\tau) = \begin{cases} \frac{1}{2}\left(1 + cos\left(\frac{\pi\tau}{M}\right)\right) r_{xx}(\tau) & \text{if } \tau < M \\ 0 & \text{if } \tau \geq M \end{cases}$$

AUTOCORRELATION: PRE-WHITENING (FSL)

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel)
2. For each lag, smooth estimates spatially ('regularization')
3. Smooth (temporally) the correlation estimate, for each voxel ('Tukey taper').



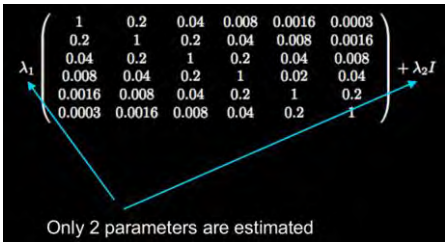
Woolrich et al, 2001 NeuroImage

Autocorrelation: Pre-whitening (FSL)

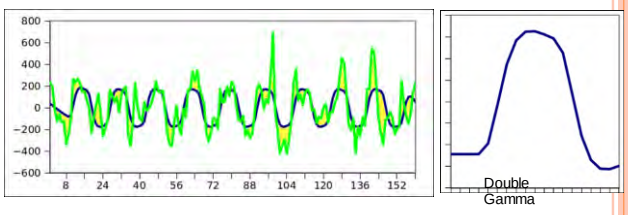
- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**
- Now that we have an estimate of the structure, we can run a second regression in which we "whiten" the residuals using the estimated **V**.
- The estimated β for the new model are BLUE (under the assumption that **V** was appropriately modelled)

Autocorrelation: pre-whitening (SPM)

- Global estimation of the correlations: the correlation of the time series is averaged over voxels.
- Imposed AR(1) model with correlation 0.2 plus white noise.

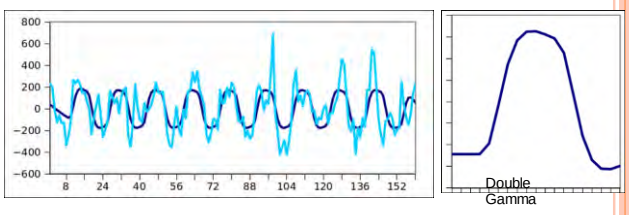


Model III: High Pass Filter



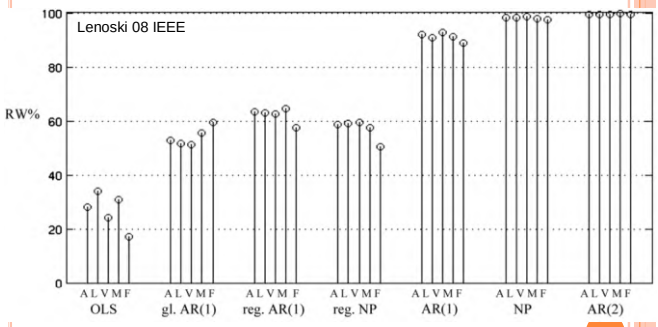
$$t(df) = \frac{(X^TX)^{-1}X^TY}{\sqrt{\hat{\sigma}^2(X^TX)^{-1}}} \quad t(159) = \frac{349}{154.08 \times 0.21} = 10.78$$

Model IV: Whitening



$$t(df) = \frac{(X^TX)^{-1}X^TY}{\sqrt{\hat{\sigma}^2(X^TX)^{-1}}} \quad t(159) = \frac{273.36}{130.20 \times 0.21} = 10.00$$

Whitening algorithms performance



Autocorrelation: Coloring

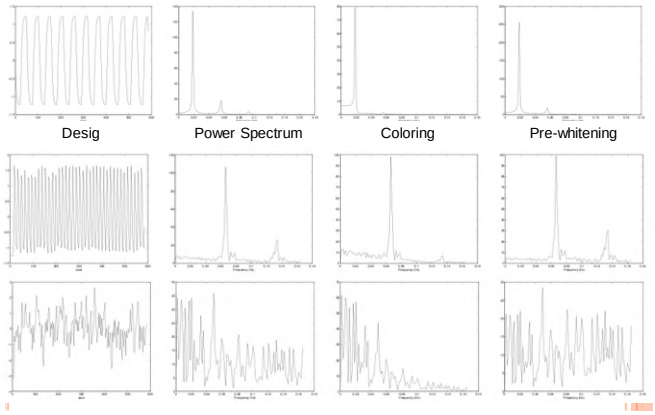
$$\varepsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 V) \qquad Cov(\varepsilon_i, \varepsilon_j) = \sigma^2 V$$

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

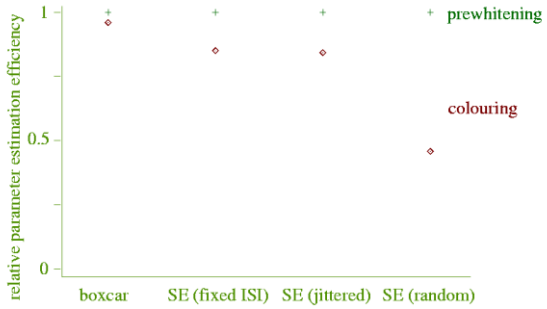
Errors are **NOT** independent of each other

- Impose a known correlation structure **S**
- If indeed $SVS^T \sim SS^T$ then the error is $N(0, \sigma^2 S)$
- So: $SY = SX\beta + S\varepsilon$
- Estimates are unbiased but no longer efficient (according to how well V is "swamped")

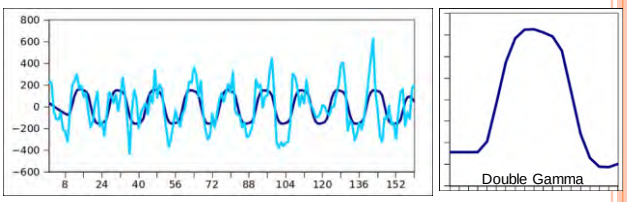
Coloring v. Whitening



Coloring v. Whitening

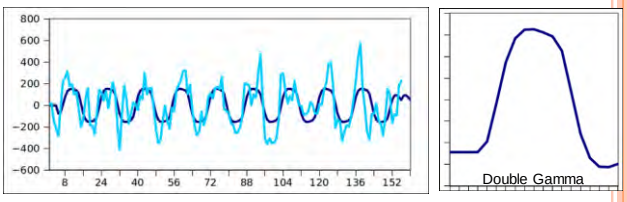


Model V: Motion Correction



$$t(df) = \frac{(X^TX)^{-1}X^TY}{\sqrt{\hat{\sigma}^2(X^TX)^{-1}}} \qquad t(159) = \frac{242.85}{102.91 \times 0.21} = 11.24$$

Model VI: Delete first 4 TRs



$$t(df) = \frac{(X^TX)^{-1}X^TY}{\sqrt{\hat{\sigma}^2(X^TX)^{-1}}} \qquad t(155) = \frac{241.79}{103.44 \times 0.21} = 11.13$$