Group analyses & Hierarchical Models

Based on slides from Will Penny & Tom Nichols

NORTHWESTERN UNIVERSITY

Darren Gitelman, MD

Associate Professor of Neurology and Radiology

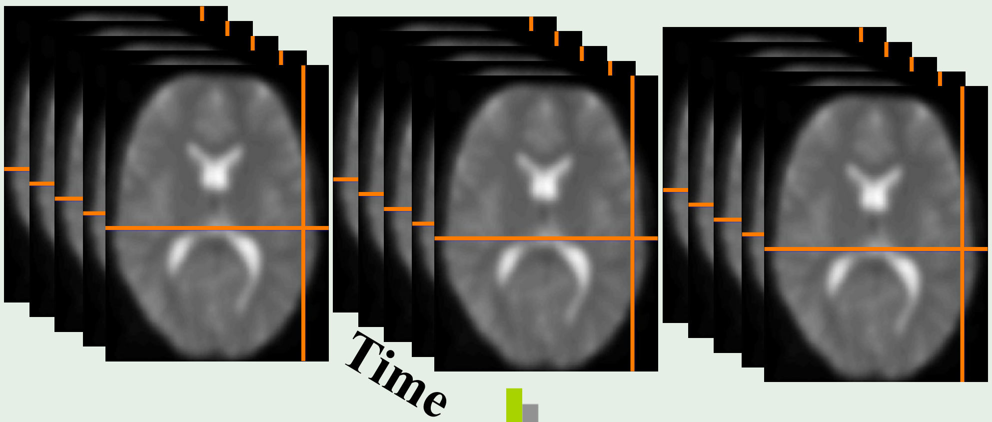
Northwestern University, Feinberg School of Medicine

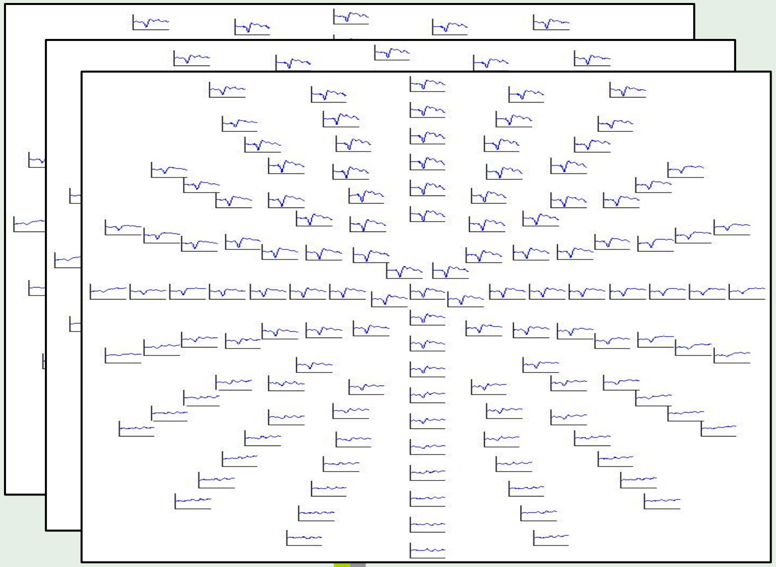
d-gitelman@northwestern.edu

Data

fMRI, single subject

EEG/MEG, single subject



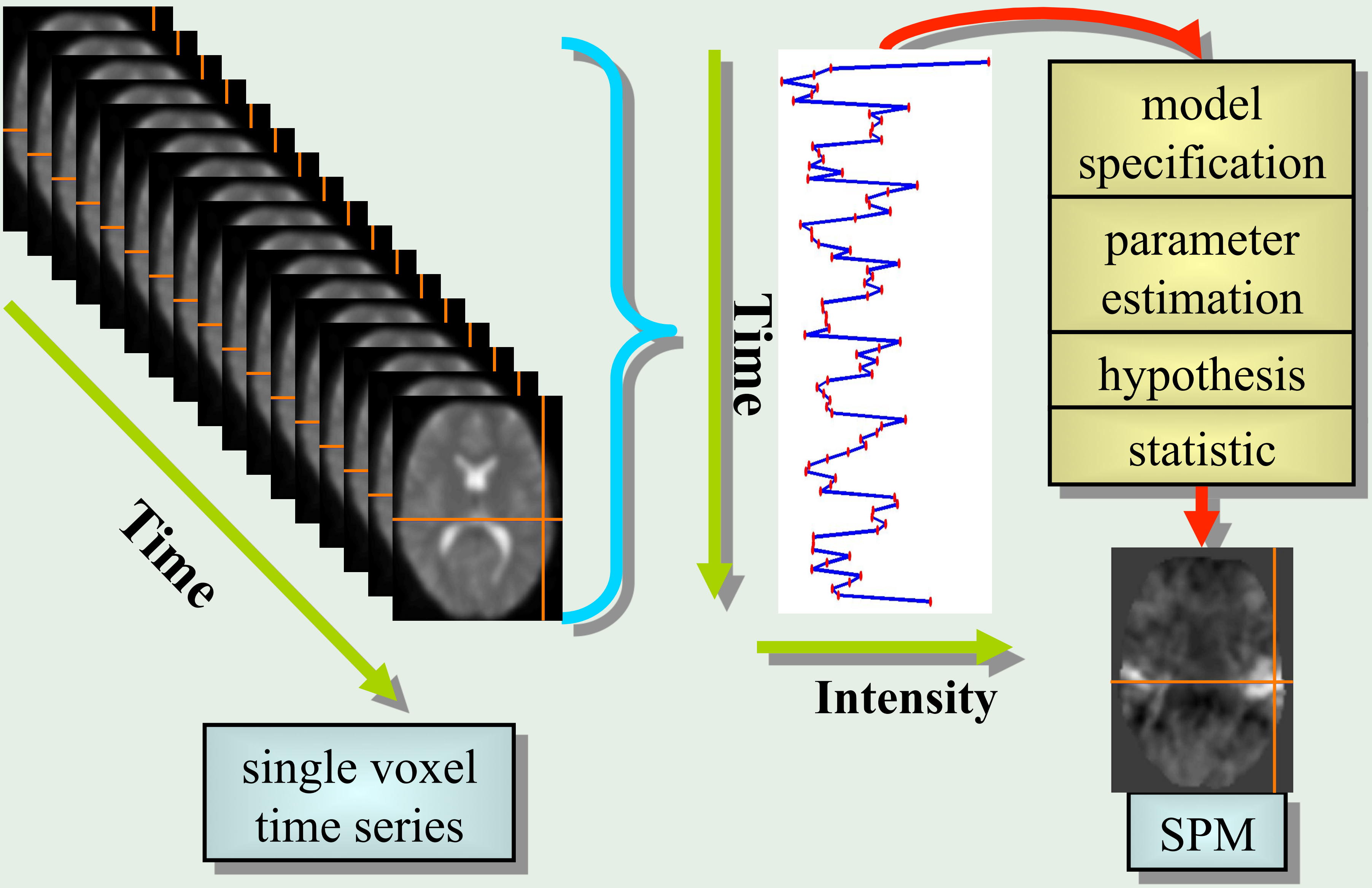


fMRI, multi-subject

ERP/ERF, multi-subject

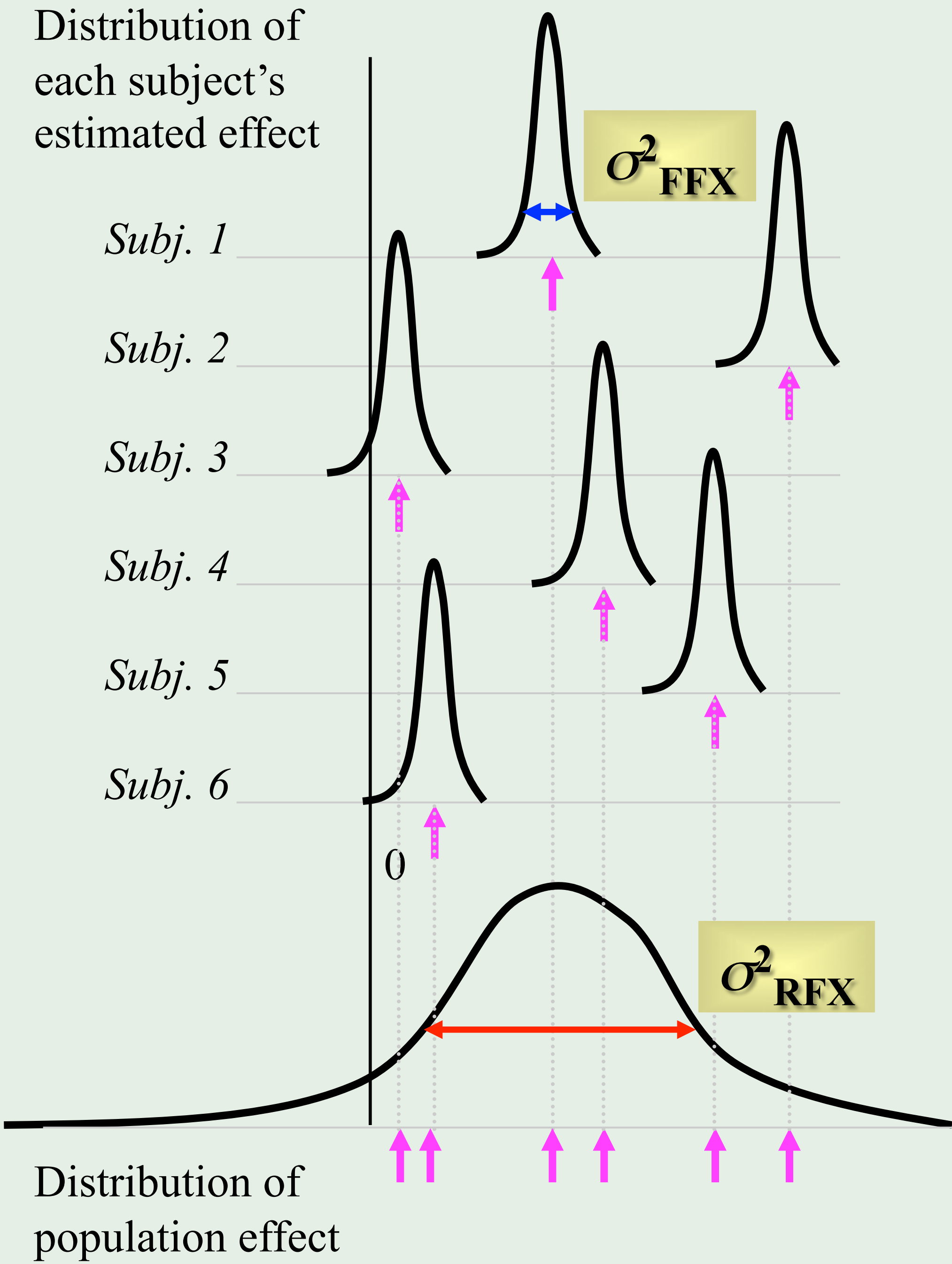
Hierarchical modeling for all imaging data

Reminder: voxel by voxel

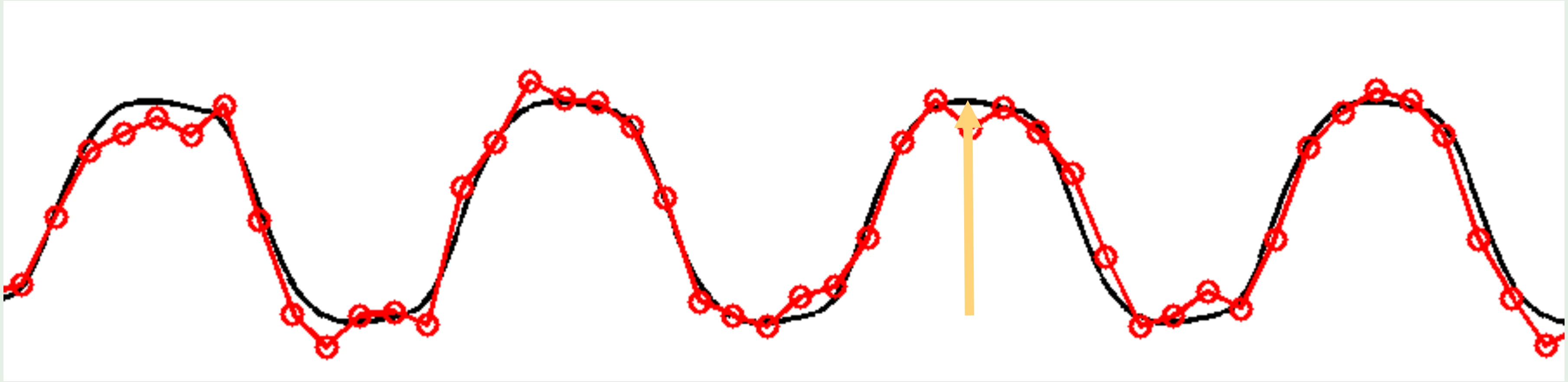


Fixed vs. Random Effects in fMRI

- Fixed Effects
  - Intra-subject variation suggests *all these subjects* different from zero
- Random Effects
  - Intersubject variation suggests *population* not very different from zero

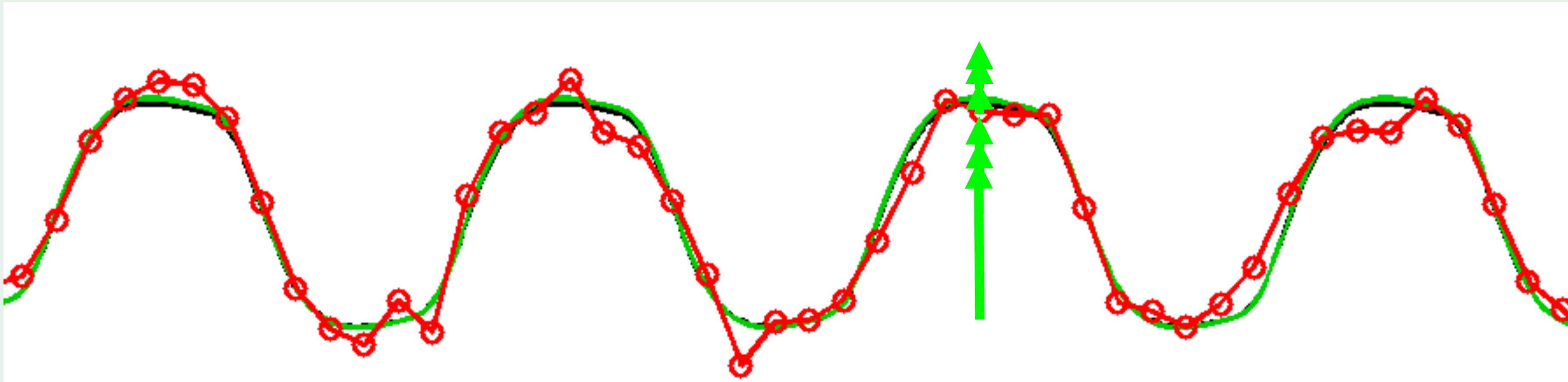


Fixed Effects



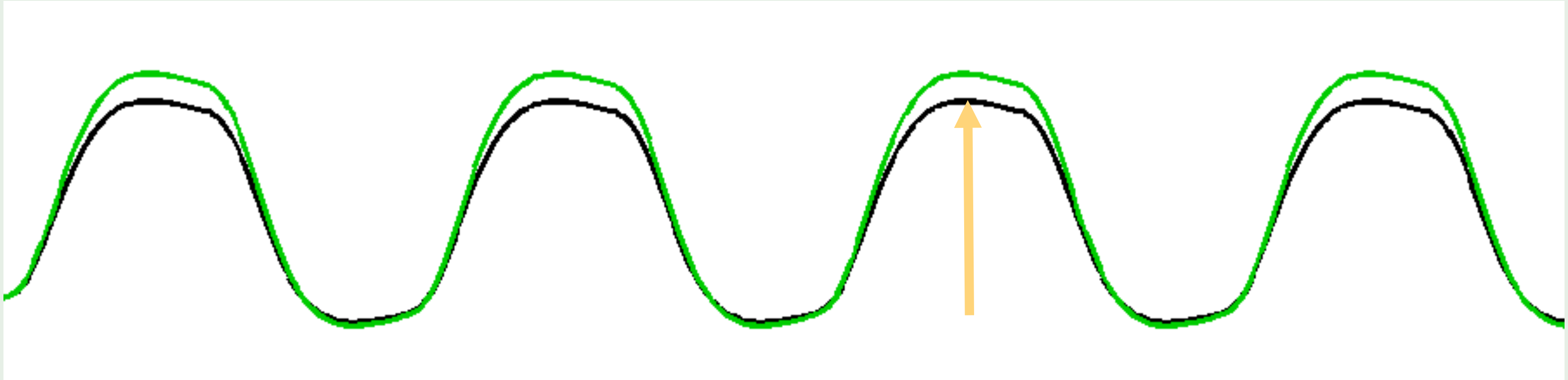
- Only variation (over sessions) is measurement error
- True Response magnitude is ***fixed***

Random/Mixed Effects



- Two sources of variation
  - Measurement error
  - Response magnitude
- Response magnitude is ***random***
  - Each subject/session has random magnitude

Random/Mixed Effects



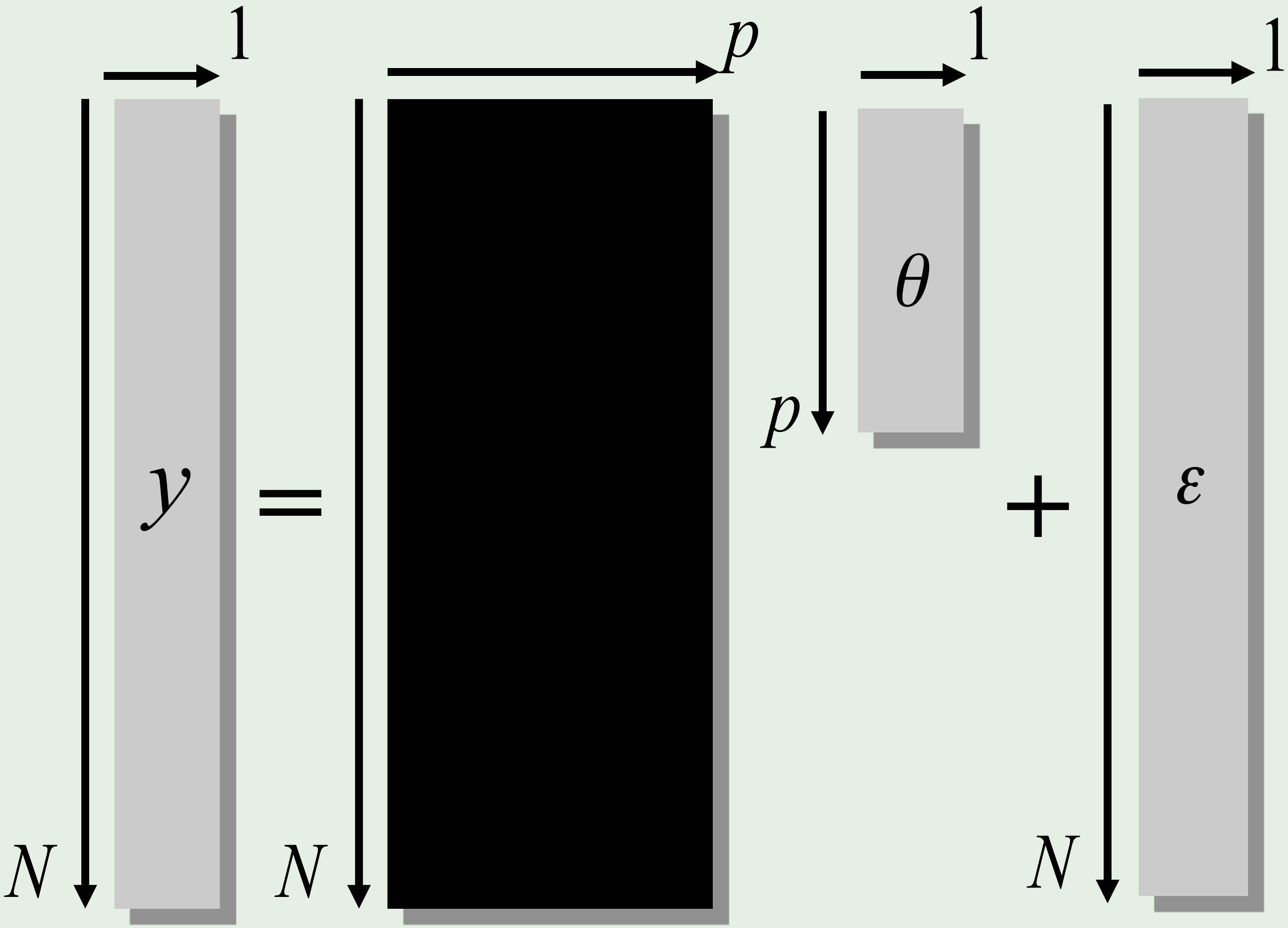
- Two sources of variation
  - Measurement error
  - Response magnitude
- Response magnitude is *random*
  - Each subject/session has random magnitude
  - But note, population mean magnitude is ***fixed***

Fixed vs. Random

- A group fixed effects analysis isn't "wrong," just usually isn't of interest across a population
- Fixed Effects Inference
  - "I can see this effect in this cohort"
  - Fixed effects might be used in a case study.
- Random Effects Inference
  - "If I were to sample a new cohort from the population I would get the same result"

General Linear Model

$$y = X\theta + \varepsilon$$



N: number of scans
p: number of regressors

Model is specified by

1. Design matrix X
2. Assumptions about ε

Linear hierarchical model

Hierarchical model

$$
\begin{aligned}
y &= X^{(1)}\theta^{(1)} + \varepsilon^{(1)} \\
\theta^{(1)} &= X^{(2)}\theta^{(2)} + \varepsilon^{(2)} \\
&\vdots \\
\theta^{(n-1)} &= X^{(n)}\theta^{(n)} + \varepsilon^{(n)}
\end{aligned}
$$

Multiple variance components at each level

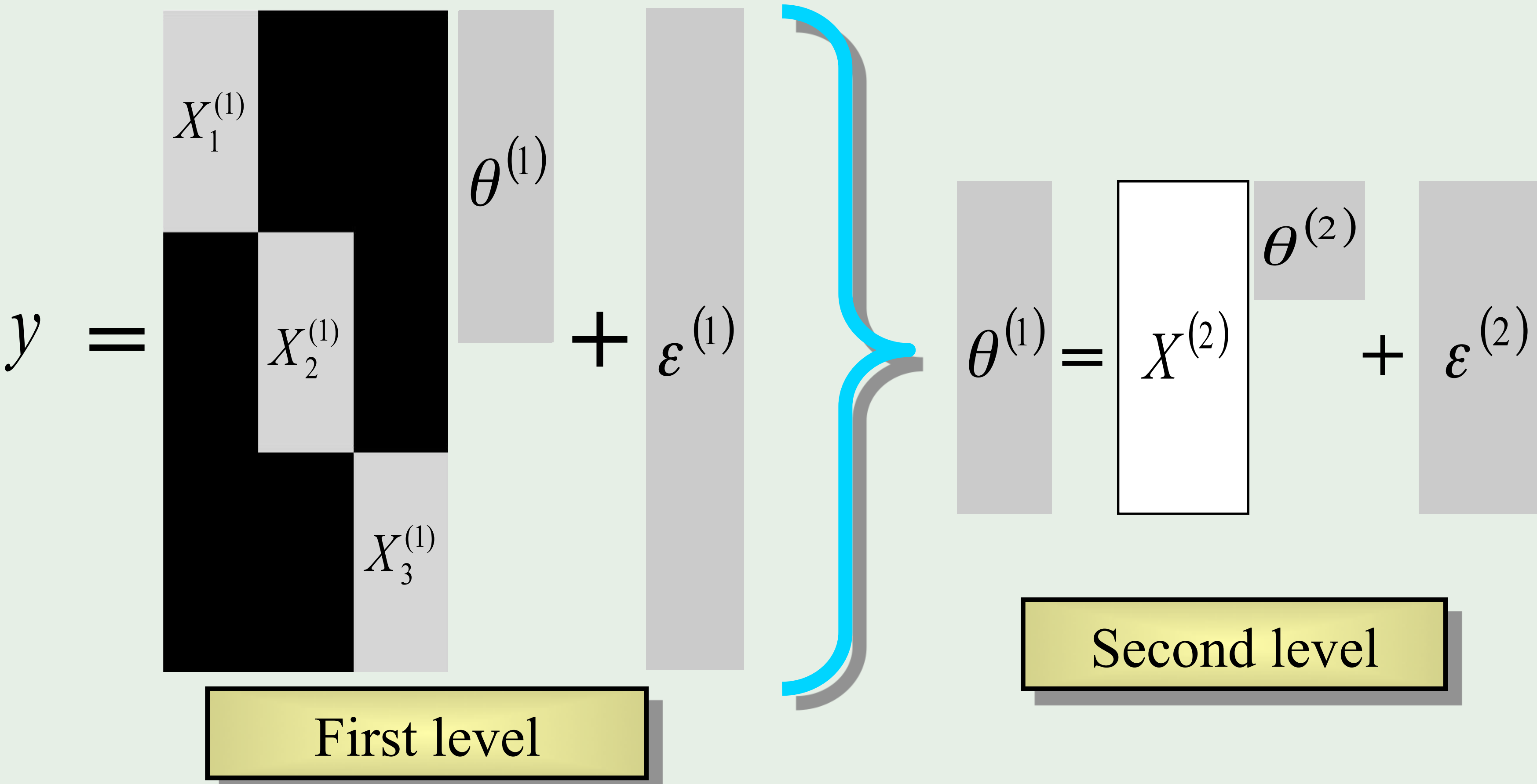
$$C_{\varepsilon}^{(i)} = \sum_k \lambda_k^{(i)} Q_k^{(i)}$$

- At each level, distribution of parameters is given by level above.
- What we don't know: distribution of parameters and variance parameters.

Example: Two level model

$$y = X^{(1)}\theta^{(1)} + \varepsilon^{(1)}$$

$$\theta^{(1)} = X^{(2)}\theta^{(2)} + \varepsilon^{(2)}$$



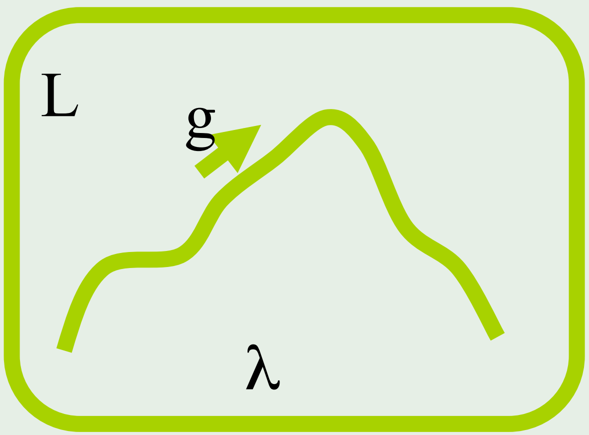
Estimation

$$\underset{N\times 1}{y} = \underset{N\times p}{X}\;\underset{p\times 1}{\theta} + \underset{N\times 1}{\varepsilon}$$

$$C_\varepsilon = \sum_k \lambda_k Q_k$$

ReML-algorithm

Maximise $L = \ln p(y \mid \lambda) = \ln \int p(y \mid \theta, \lambda)\, d\theta$



$$g = \frac{dL}{d\lambda}$$

$$J = \frac{d^2L}{d\lambda^2}$$

$$\lambda = \lambda + J^{-1} g$$

OrdinaryLeast Squares

$$\theta = \left(X^T X\right)^{-1} X^T y$$

Correct for non-sphericity

Weighted Least Squares

$$\theta = (X^T C_e^{-1} X^T) X^T C_e^{-1} y$$

WLS equivalent to OLS on whitened data and design

Friston et al., Neuroimage, 2002

Algorithmic Equivalence

Hierarchical model

$$y = X^{(1)}\theta^{(1)} + \varepsilon^{(1)}$$
$$\theta^{(1)} = X^{(2)}\theta^{(2)} + \varepsilon^{(2)}$$
$$\vdots$$
$$\theta^{(n-1)} = X^{(n)}\theta^{(n)} + \varepsilon^{(n)}$$

Single-level model

$$y = \varepsilon^{(1)} + X^{(1)}\varepsilon^{(2)} + \ldots + X^{(1)}\ldots X^{(n-1)}\varepsilon^{(n)} + X^{(1)}\ldots X^{(n)}\theta^{(n)}$$

Restricted Maximum Likelihood (ReML)

Group analysis in practice

Many 2-level models are just too big to compute.

And even if estimable, it takes a long time!

And if subjects are added it must be completely re-estimated.

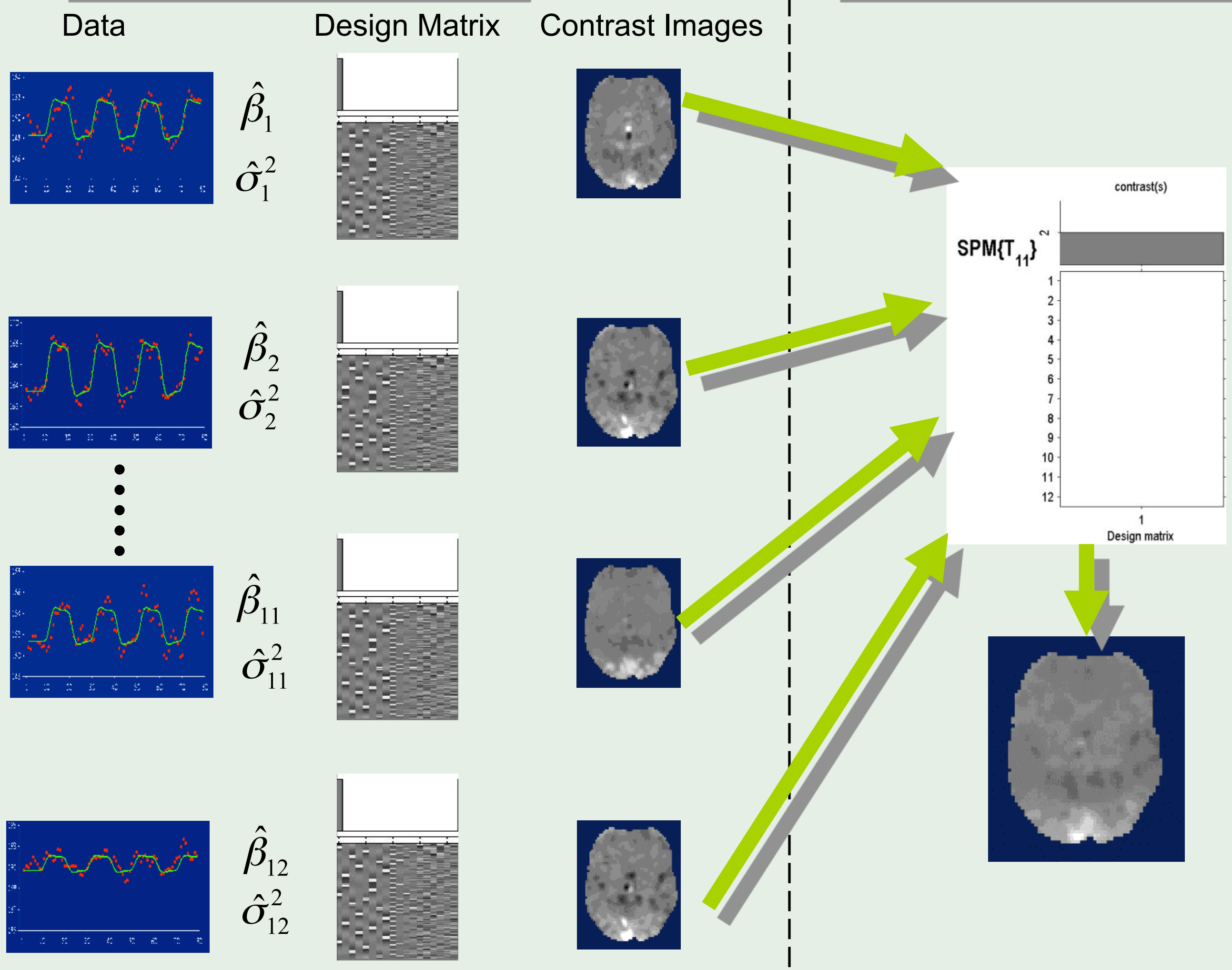
Is there a fast & valid approximation?

Summary Statistics approach

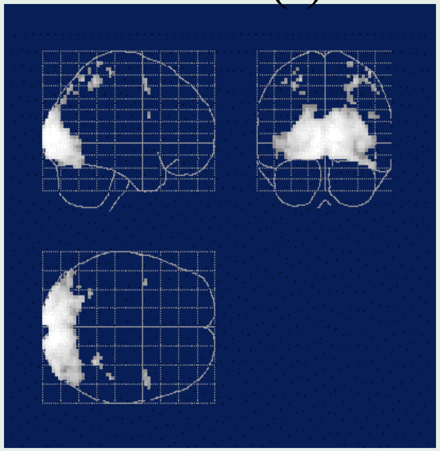
First level
(within subject)

Second level
(between subject)

$$t = \frac{c^T \hat{\beta}}{\sqrt{V\hat{a}r(c^T \hat{\beta})}}$$



SPM(t)



One-sample t-test @ 2^{nd} level

Validity of approach

The summary stats approach is exact if for each session/subject:

Within-session covariance the same

First-level design the same

Errors are normally distributed

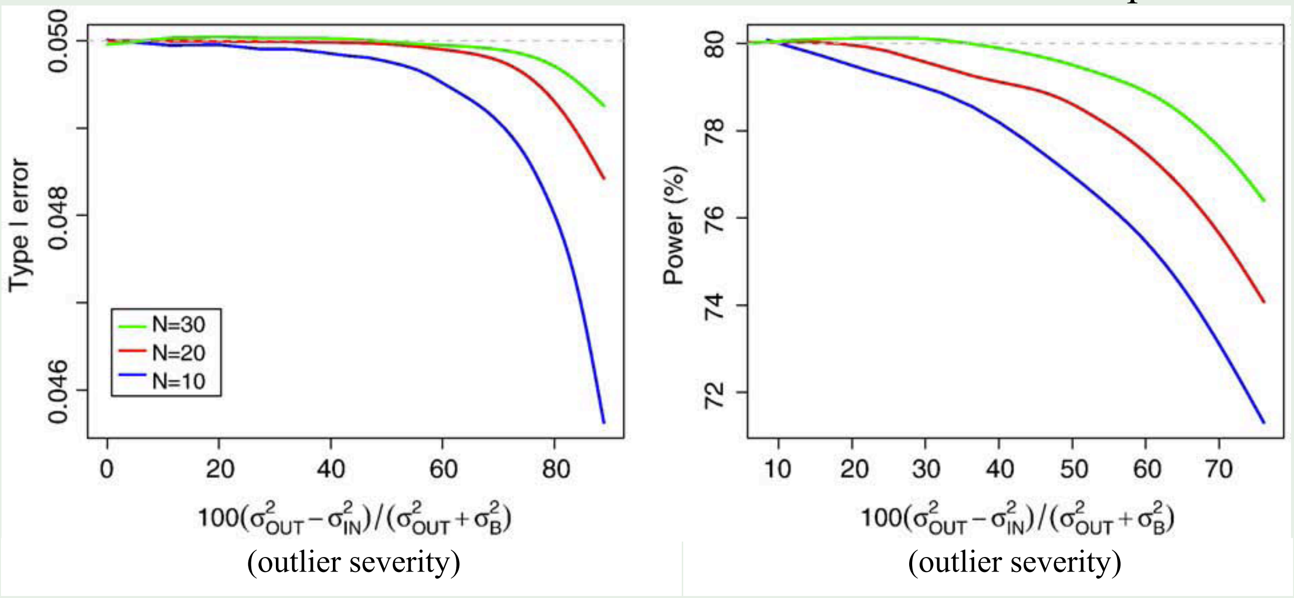
Original specification of summary statistics approach (Holmes & Friston, 1996) was limited to 1 contrast image per subject.

If >1 contrast image per subject need to estimate the effects of correlated errors: non-sphericity

Holmes & Friston Robustness

- In practice, Validity & Efficiency are excellent
  - For one sample case, HF almost impossible to break

False Positive Rate   Power Relative to Optimal

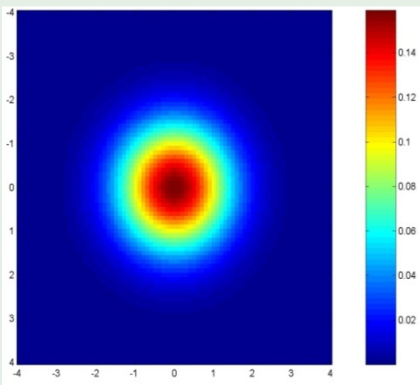


Mumford & Nichols. Simple group fMRI modeling and inference. *Neuroimage*, 47(4):1469--1475, 2009.

- 2-sample & correlation might give trouble
  - Dramatic imbalance or heteroscedasticity

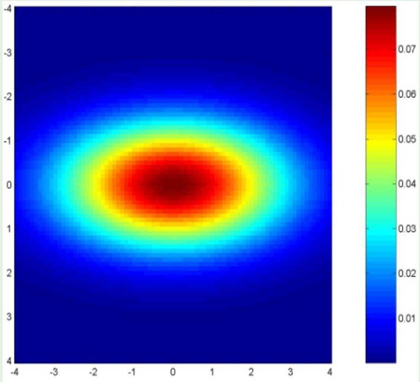
GLM assumes Gaussian “spherical” (i.i.d.) errors

sphericity = iid:
error covariance is
scalar multiple of
identity matrix:
Cov(e) = $\sigma^2 I$



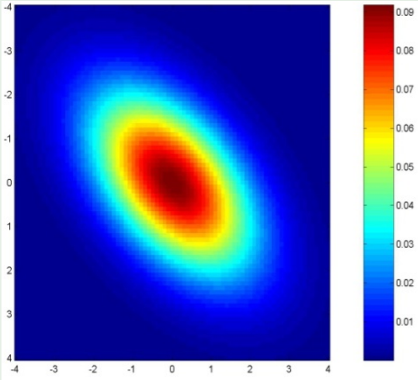
$$Cov(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Examples of non-sphericity:



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

non-identity



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

non-independence

Multiple Variance Components

$$\underset{N\times 1}{y} = \underset{N\times p}{X}\ \underset{p\times 1}{\theta} + \underset{N\times 1}{\varepsilon}$$

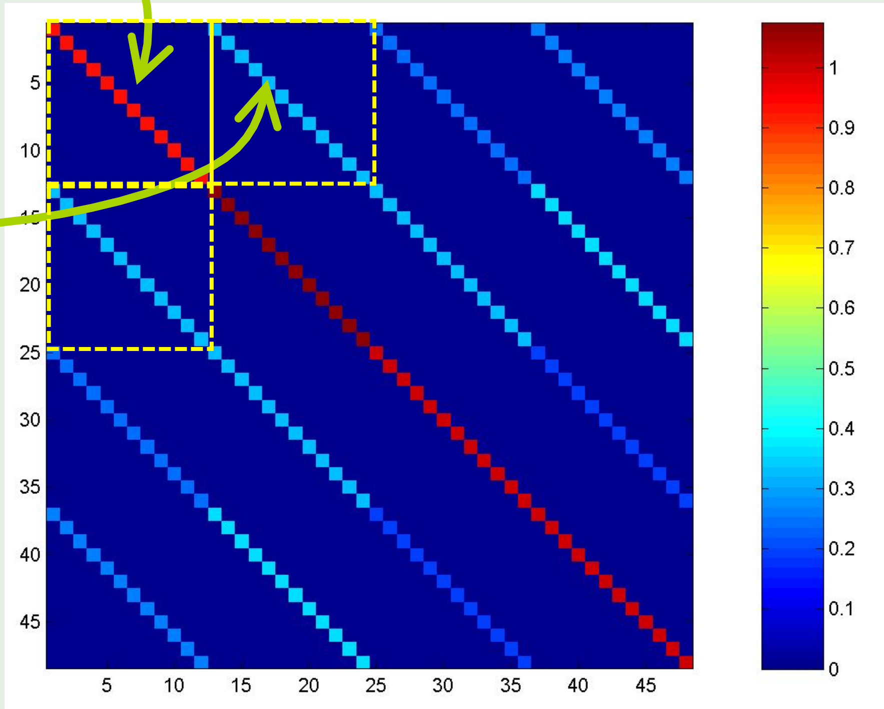
$$\mathrm{Cov}(\varepsilon) = \Sigma_k \lambda_k Q_k$$

Error covariance

- 12 subjects, 4 conditions
- Measurements btw subjects uncorrelated
- Measurements w/in subjects correlated
- Errors can now have
- different variances and
- there can be correlations
- Allows for ‘non-sphericity’

N

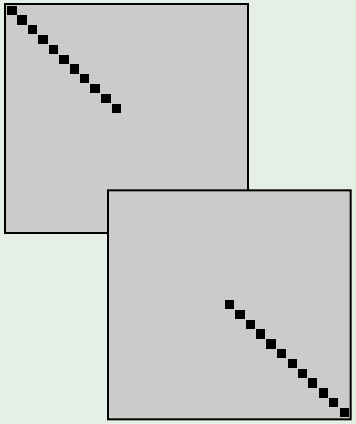
N



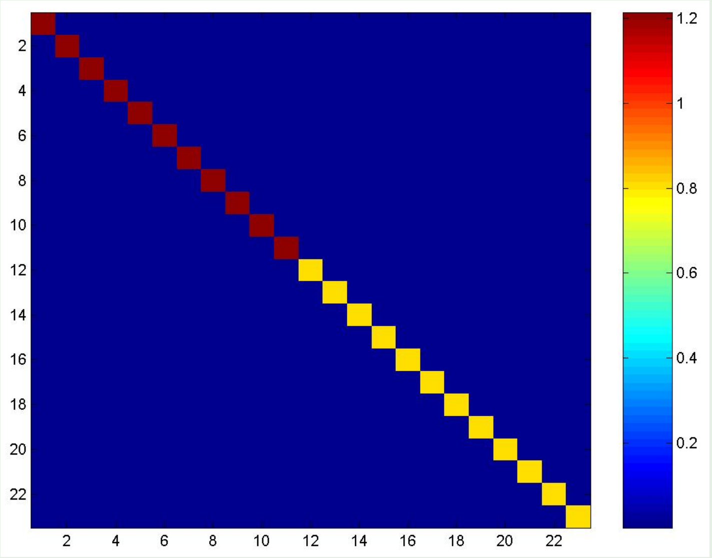
Non-Sphericity Modeling

- Errors are independent but not identical
  - Eg. Two Sample T-test
    Two basis elements

Q_k's:



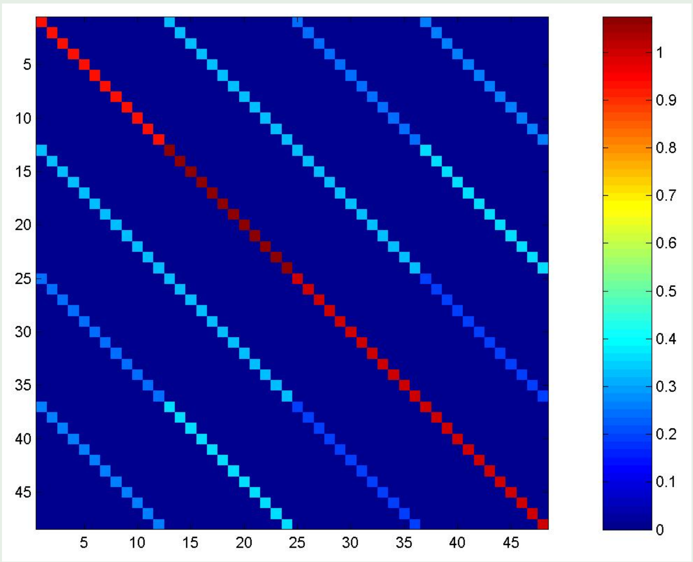
Error Covariance



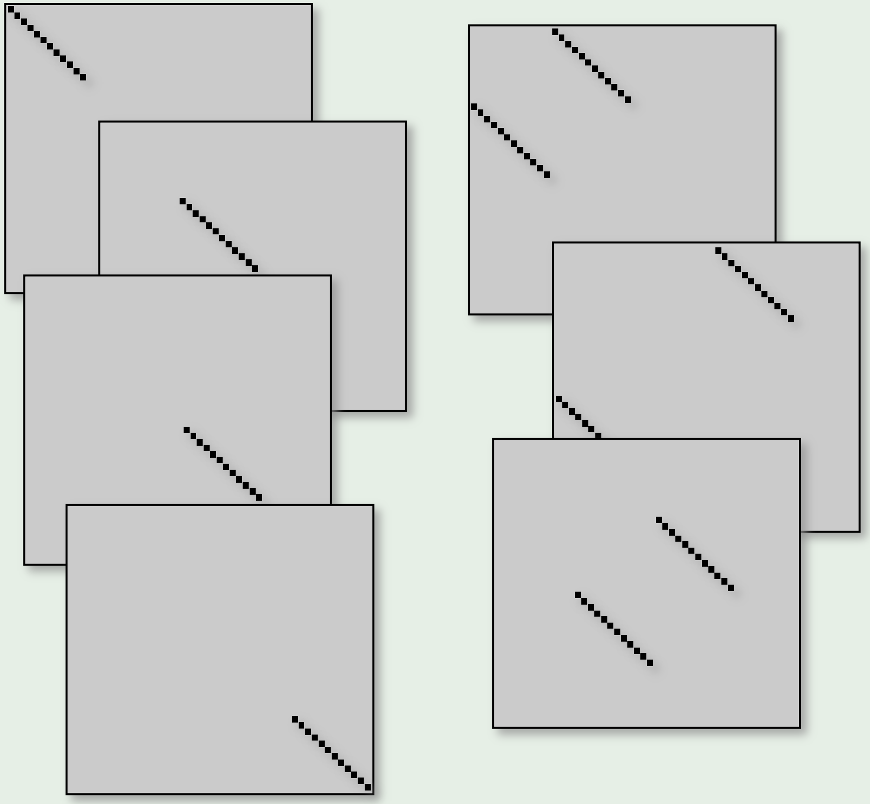
Non-Sphericity Modeling

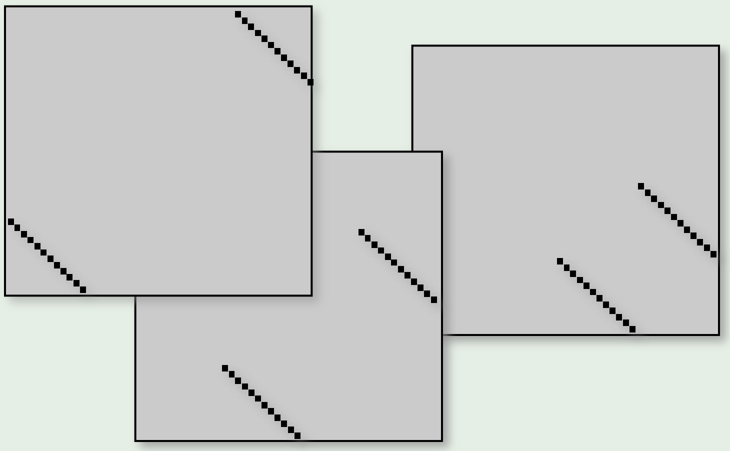
- Errors are not independent and not identical

Error Covariance



Q_k's:



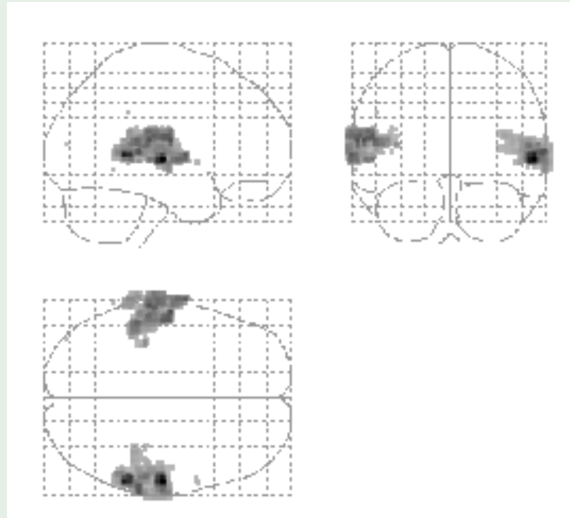


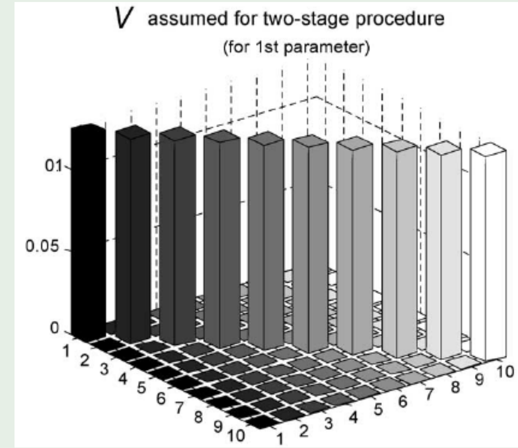
SPM8 Nonsphericity Modelling

- Assumptions & Limitations
  - $\mathrm{Cov}(\varepsilon) = \Sigma_k \lambda_k Q_k$ assumed to be globally homogeneous
  - λ_k's only estimated from voxels with large F (>0.001 unc)
  - Most realistically, $\mathrm{Cov}(\varepsilon)$ spatially heterogeneous
  - Intrasubject variance assumed homogeneous

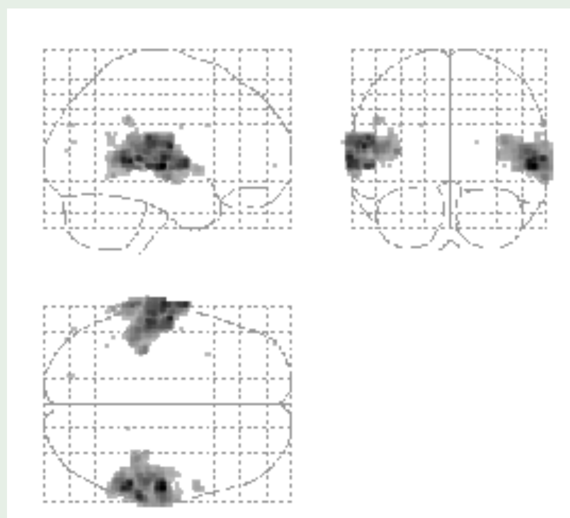
Auditory fMRI Data

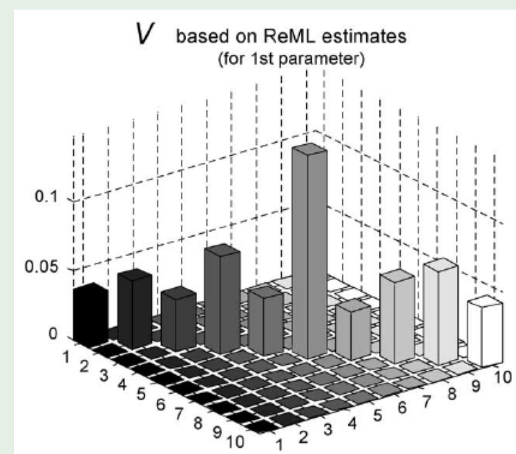
Summary statistics





Hierarchical Model





Friston et al., Neuroimage, 2005

Example 1: non-identical groups

Stimuli:

Auditory Presentation (SOA = 4 secs) of
(i) words and (ii) words spoken backwards

e.g.
"Book"
and
"Koob"

Subjects:

(i) 12 control subjects
(ii) 11 blind subjects

Scanning:

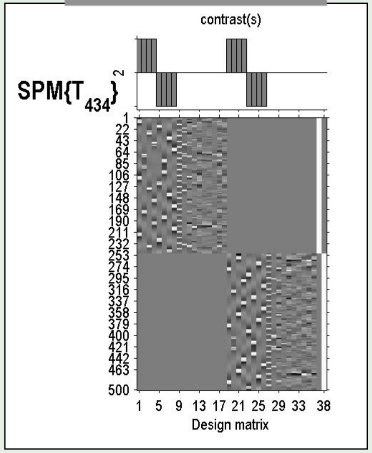
fMRI, 250 scans per
subject, block design

Noppeney et al., Brain, 2003

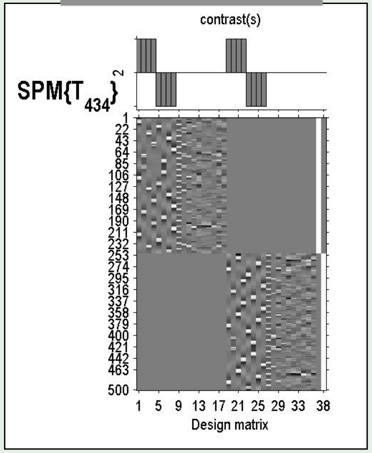
Population differences

1st level:

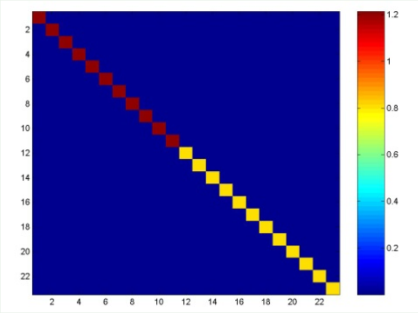
Controls



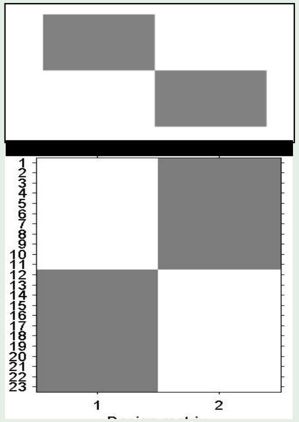
Blinds



2nd level:

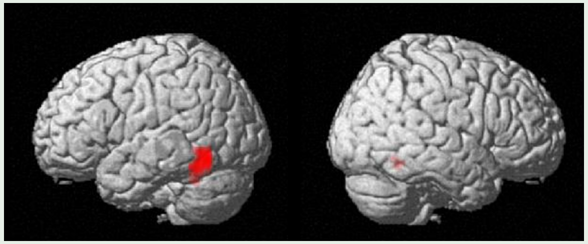


V



$c^T = [1 \quad -1]$

X



Example 2: Multiple contrasts per subject

Stimuli:

Auditory Presentation (SOA = 4 secs) of words

Motion	Sound	Visual	Action
“jump”	“click”	“pink”	“turn”

Subjects:

(i) 12 control subjects

Scanning:

fMRI, 250 scans per subject, block design

Question:

What regions are affected by the semantic content of the words?

Noppeney et al. Brain & Lang, 2003

ANOVA

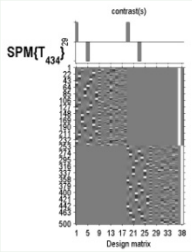
1st level:

1.Motion

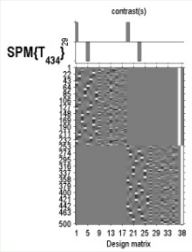
2.Sound

3.Visual

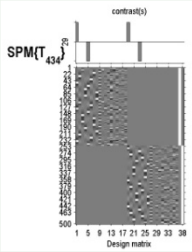
4.Action



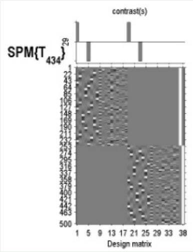
$\stackrel{?}{=}$



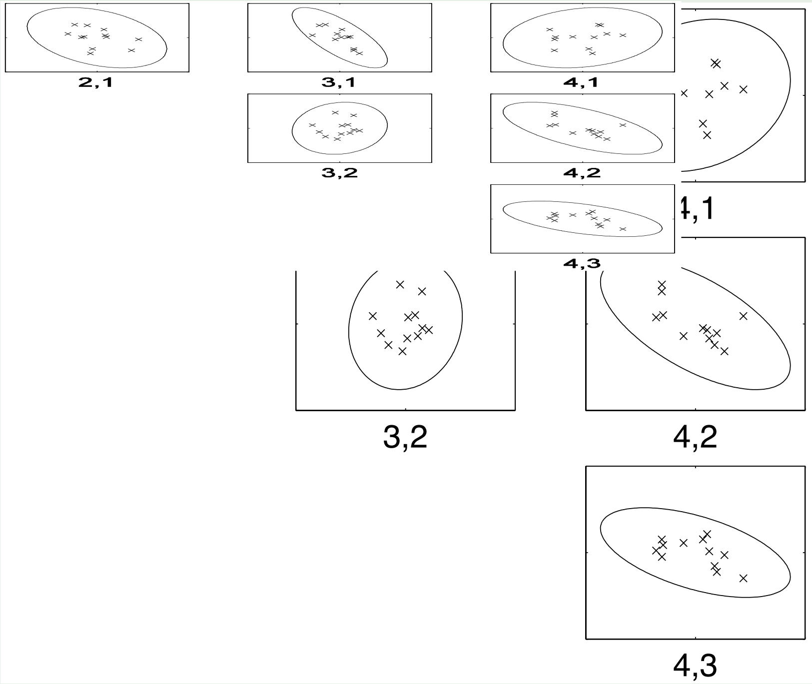
$\stackrel{?}{=}$



$\stackrel{?}{=}$



2nd level:

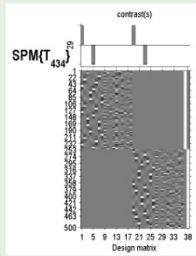


Noppeney et al. Brain & Lang, 2003

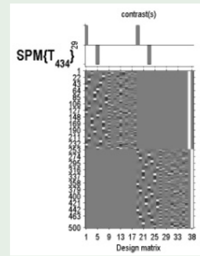
ANOVA

1st level:

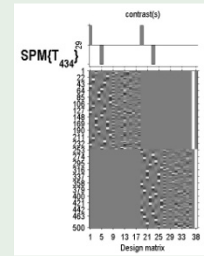
Motion | Sound | Visual | Action



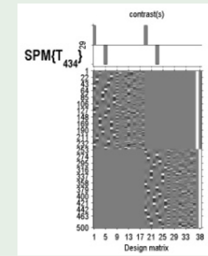
$\stackrel{?}{=}$



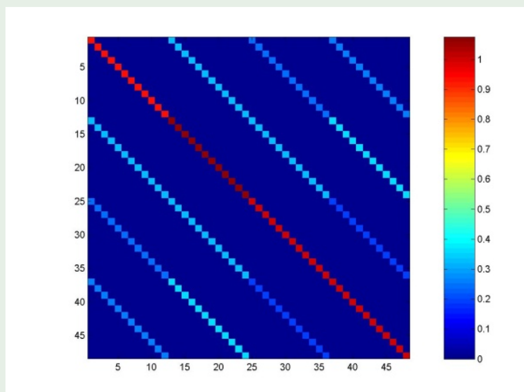
$\stackrel{?}{=}$



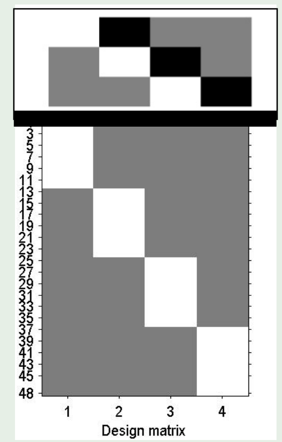
$\stackrel{?}{=}$



2nd level:

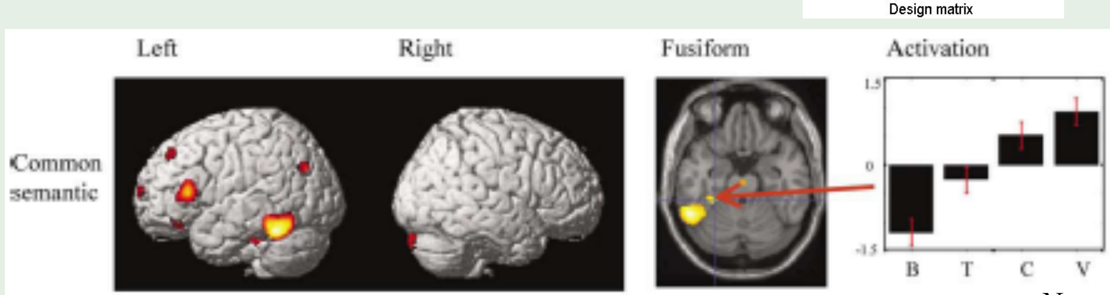


V



X

$$c^T = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$



Noppeney et al. Brain & Lang, 2003

Summary

Linear hierarchical models are general enough for typical multi-subject imaging data (PET, fMRI, EEG/MEG).

Summary statistics are a robust approximation for group analysis.

Modeling non-sphericity at the second level accommodates multiple contrasts per subject.

Use mixed-effects model only if seriously in doubt about validity of summary statistics approach.

The End