Analysis of longitudinal imaging data with OLS & Sandwich Estimator standard errors

Bryan Guillaume

Supervisors: Thomas Nichols (Warwick University) and Christophe Phillips (Liège University)

Outline

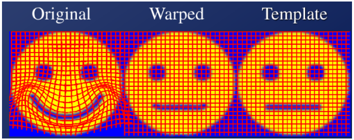
An example of longitudinal studies in neuroimaging

The ADNI study

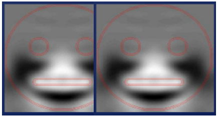
- Tensor-Based Morphometry (TBM) images from the Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Available scans:

	AD	MCI	N	Total
0 Mo	188	400	229	817
6 Mo	159	346	208	713
12 Mo	138	326	196	660
18 Mo	n/a	286	n/a	286
24 Mo	105	244	172	521
36 Mo	n/a	170	147	317

TBM images?



$det(J)$ of the spatial deformation



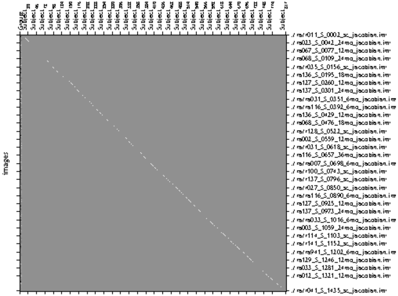
$det(J) > 1$: expansion
$det(J) < 1$: contraction

The Linear Mixed Effect (LME) model

- The gold standard in the biostatistics literature
- Iterative method
  - Generally slow
  - May fail to converge
    - E.g., 12 subjects, 8 visits, Toeplitz, LME with unstructured intra-visit correlation fails to converge 95 % of the time.
    - E.g., 12 subjects, 8 visits, Compound Symmetry (CS), LME with random int. and random slope fails to converge 2 % of the time.
- LME model with a random intercept per subject
  - May be slow (iterative method)
  - Only valid under Compound Symmetric (CS) intra-visit covariance structure

The Naive Ordinary Least Squares (N-OLS) model

- Design matrix in the ADNI design



- Much faster than a LME model
- No inference possible on between subject effects (e.g., group intercept, gender, age at first visit)
- Only valid under Compound Symmetric (CS) intra-visit covariance structure

The Summary Statistics OLS (SS-OLS) model

- Procedure
  1. Extraction of summary statistics for each subjects
     - E.g., intercept, slope
  2. Use of an OLS model for each summary statistics
- Transformation of correlated data into uncorrelated data
- Important loss of information
  - Will affect negatively the power
  - Are we able to get accurate inferences?

Outline

The Sandwich Estimator (SwE) method

- Use of a simple OLS model (without subject indicator variables)
- The fixed effects parameters β are estimated by

$$\hat{\beta}_{OLS} = \left(\sum_{i=1}^{M} X_i' X_i \right)^{-1} \sum_{i=1}^{M} X_i' y_i$$

- The fixed effects parameters covariance $\text{var}(\hat{\beta}_{OLS})$ are estimated by

$$S = \underbrace{\left(\sum_{i=1}^{M} X_i' X_i \right)^{-1}}_{\text{Bread}} \underbrace{\left(\sum_{i=1}^{M} X_i' \hat{V}_i X_i \right)}_{\text{Meat}} \underbrace{\left(\sum_{i=1}^{M} X_i' X_i \right)^{-1}}_{\text{Bread}}$$

Property of the Sandwich Estimator (SwE)

$$S = \left(\sum_{i=1}^{M} X_i' X_i\right)^{-1} \left(\sum_{i=1}^{M} X_i' \hat{V}_i X_i\right) \left(\sum_{i=1}^{M} X_i' X_i\right)^{-1}$$

If $m^{-1} \sum_{i=1}^{m} X_i' \hat{V}_i X_i$ consistently estimates $m^{-1} \sum_{i=1}^{m} X_i' V_i X_i$, the SwE tends **asymptotically** (Large samples assumption) towards the true variance var$(\hat{\beta}_{OLS})$. (Eicker, 1963; Eicker, 1967; Huber, 1967; White, 1980)

The classical Heterogeneous SwE $S_{Het,0}$

In practice, V_i is generally estimated from the residuals $r_i = y_i - X_i\hat{\beta}$ by

$$\hat{V}_i = r_i r_i^{'}$$

and the SwE becomes

$$S_{Het,0} = \left(\sum_{i=1}^{M} X_i' X_i\right)^{-1} \left(\sum_{i=1}^{M} X_i' r_i r_i^{'} X_i\right) \left(\sum_{i=1}^{M} X_i' X_i\right)^{-1}$$

The adjusted Heterogeneous SwE $S_{Het,1}$

Hinkley (1977) proposed to estimate V_i in the SwE by

$$\hat{V}_i = \frac{N}{N-p} r_i r_i^{'}$$

where N is the number of observations and p the number of parameters

$$\begin{aligned} S_{Het,1} &= \frac{N}{N-p} \left(\sum_{i=1}^{M} X_i^{\prime} X_i \right)^{-1} \left(\sum_{i=1}^{M} X_i^{\prime} r_i r_i^{'} X_i \right) \left(\sum_{i=1}^{M} X_i^{\prime} X_i \right)^{-1} \\ &= \frac{N}{N-p} S_{Het,0} \end{aligned}$$

The adjusted Heterogeneous SwE $S_{Het,2}$

- In an OLS model, we have

$$(I - H)\text{var}(y)(I - H) = \text{var}(r)$$

where $H = X(X'X)^{-1}X'$

- Under independent homoscedastic errors,

$$(I - H)\sigma^2 = \text{var}(r)$$
$$(1 - h_{ik})\sigma^2 = \text{var}(r_{ik})$$
$$\sigma^2 = \text{var}\left(\frac{r_{ik}}{\sqrt{1 - h_{ik}}}\right)$$

- This suggests to estimate V_i by

$$\hat{V}_i = r_i^* r_i^{*'} \text{ where } r_{ik}^* = \frac{r_{ik}}{\sqrt{1 - h_{ik}}}$$

The adjusted Heterogeneous SwE $S_{Het,2}$

Using

$$\hat{V}_i = r_i^* r_i^{*'} \text{ where } r_{ik}^* = \frac{r_{ik}}{\sqrt{1-h_{ik}}},$$

We obtain

$$S_{Het,2} = \left(\sum_{i=1}^{M} X_i' X_i\right)^{-1} \left(\sum_{i=1}^{M} X_i' r_i^* r_i^{*'} X_i\right) \left(\sum_{i=1}^{M} X_i' X_i\right)^{-1}$$

(Mckinnon and White, 1985)

The adjusted Heterogeneous SwE $S_{Het,3}$

Mckinnon and White (1985) also proposed to use standardised residuals

$$r_{ik}^{**} = \frac{r_{ik}}{1 - h_{ik}}$$

instead of the raw residuals. V_i is then estimated by

$$\hat{V}_i = r_i^{**} r_i^{**'}$$

where r_i^{**} is the vector of standardized residuals for subject i

$$S_{Het,3} = \left(\sum_{i=1}^{M} X_i' X_i \right)^{-1} \left(\sum_{i=1}^{M} X_i' r_i^{**} r_i^{**'} X_i \right) \left(\sum_{i=1}^{M} X_i' X_i \right)^{-1}$$

The Homogeneous SwE $S_{Hom,*}$

- In the standard SwE $S_{Het,*}$, each V_i is usually estimated from only the residuals of subject i.
- Assumption of a common covariance matrix V_{0g} for subjects belonging to the same group

$$(\hat{V}_{0g})_{kk} = \frac{1}{M_{gkk}} \sum_{i \in A(g,k,k)}^{M} r_{ik}^2.$$

$$(\hat{V}_{0g})_{kk'} = \frac{\sum_{i \in A(g,k,k')}^{M} r_{ik} r_{ik'}}{\sqrt{\left(\sum_{i \in A(g,k,k')}^{M} r_{ik}^2\right)\left(\sum_{i \in A(g,k,k')}^{M} r_{ik'}^2\right)}} \sqrt{(\hat{V}_{0g})_{kk} (\hat{V}_{0g})_{k'k'}}$$

M_{gkk}: number of subjects in group g with data at visit k
$A(g,k,k')$: subset of subjects in group g who have data at both visit k and k'

Test statistics with the SwE

- $H_0 : C\hat{\beta} = 0, H_1 : C\hat{\beta} \neq 0$
  C: contrast matrix of rank q
- Using multivariate statistics theory:

$$\frac{\nu - q + 1}{\nu q}(C\hat{\beta})'(CSC')^{-1}(C\hat{\beta}) \sim F(q, \nu - q + 1)$$

- In practice, ν can be estimated by

$$\hat{\nu} = \frac{tr(CSC')^2 + (tr(CSC'))^2}{\sum_{g=1}^{G} \frac{tr(CSC')_g^2 + (tr(CSC')_g)^2}{\nu_g}}$$

$$(CSC')_g = C\left(\sum_{i=1}^{M} X_i' X_i\right)^{-1} \left(\sum_{i \in A(g)}^{M} X_i' \hat{V}_i X_i\right) \left(\sum_{i=1}^{M} X_i' X_i\right)^{-1} C'$$

Outline

Simulations: setup

- Designs considered:
  - Balanced design (2 gr.; 12, 25, 50, 100 $\&$ 200 subj.; 3, 5 $\&$ 8 vis.)
  - ADNI design (3 gr.; 51, 102, 204, 408 $\&$ 817 subj.; max. 6 vis.)
- Monte Carlo Gaussian null simulation (10,000 realizations)
- For each realization,
  1. Generation of longitudinal Gaussian null data (no effect) with intra-visit covariance structures:

Compound Symmetric

$$\begin{pmatrix} 1 & 0.8 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 0.8 & 1 \end{pmatrix}$$

Toeplitz

$$\begin{pmatrix} 1 & 0.8 & 0.6 & 0.4 \\ 0.8 & 1 & 0.8 & 0.6 \\ 0.6 & 0.8 & 1 & 0.8 \\ 0.4 & 0.6 & 0.8 & 1 \end{pmatrix}$$

Heterogeneous visit variances

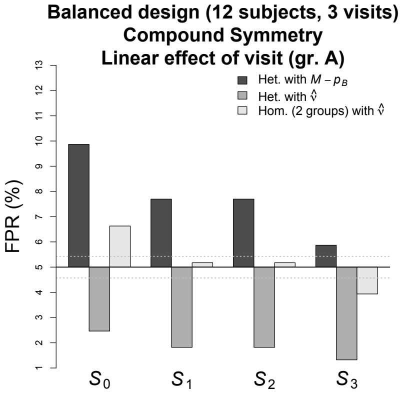
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

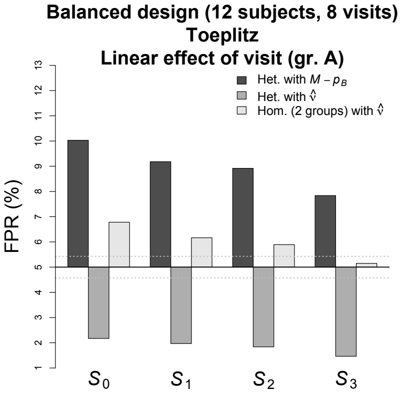
Heterogeneous group variances

$$\alpha_g \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

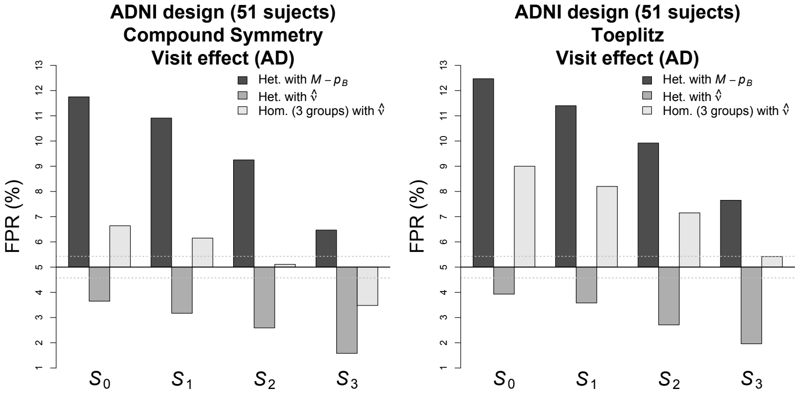
  2. Statistical test (F-test at 5%) on the parameters of interest and estimation of the FPR

SwE versions comparison: balanced design

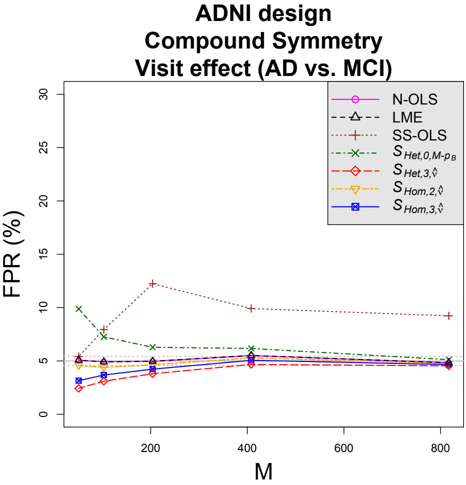


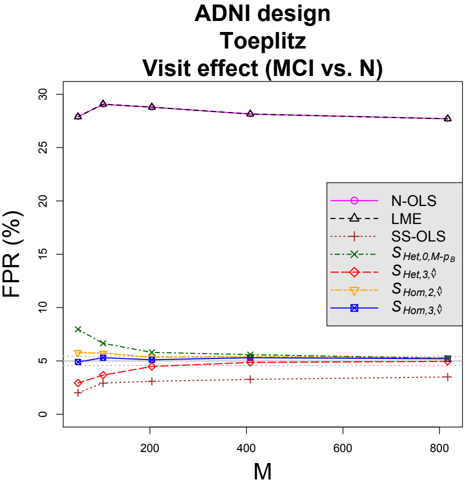


SwE versions comparison: ADNI design

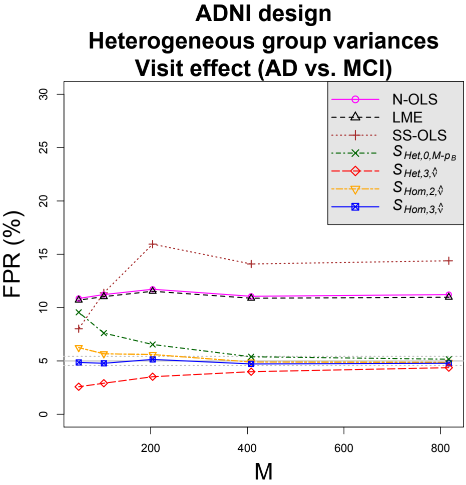


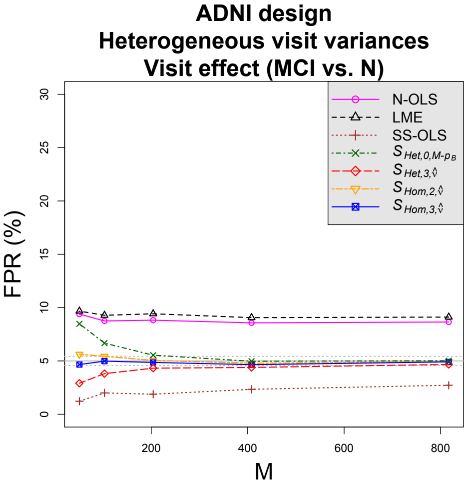
Comparison with the other methods





Comparison with the other methods



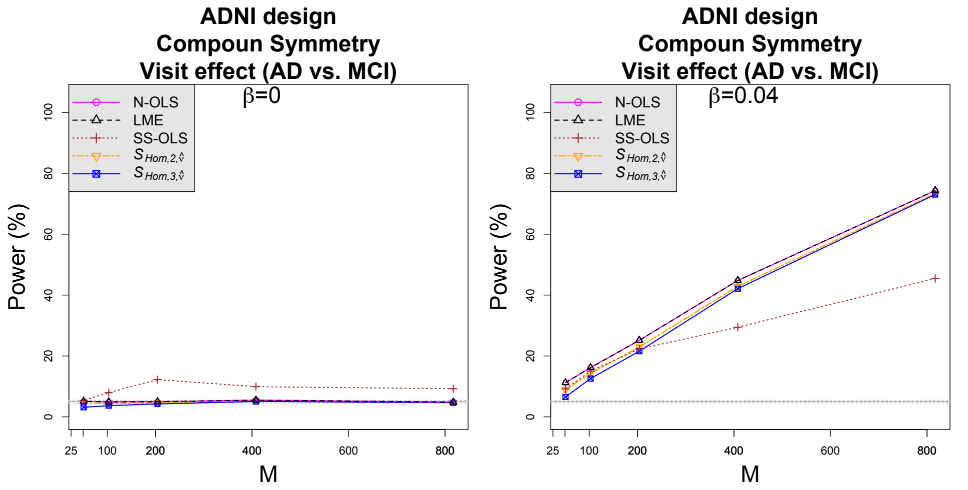


FPR control with enough subjects: results summary

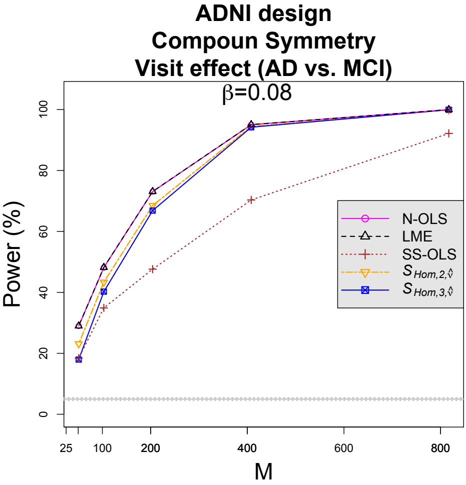
Design	Cov. type	Effect type	NOLS	LME	SSOLS	SwE
Balanced	CS	Between	n/a	·	·	·
		Within	·	·	·	·
	Toeplitz	Between	n/a	·	·	·
		Within	$++/-$	$++/-$	·	·
	Het. visits	Between	n/a	$-$	·	·
		Within	·	·	·	·
	Het. groups	Between	n/a	$+/-$	$+/-$	·
		Within	$+/-$	$+/-$	$+/-$	·
ADNI	CS	Between	n/a	·	·	·
		Within	·	·	$+/-$	·
	Toeplitz	Between	n/a	·	·	·
		Within	$+$	$+$	$+/-$	·
	Het. visits	Between	n/a	$-$	·	·
		Within	$+$	$+$	$+/-$	·
	Het. groups	Between	n/a	$+/-$	$+/-$	·
		Within	$+/-$	$+/-$	$+/-$	·

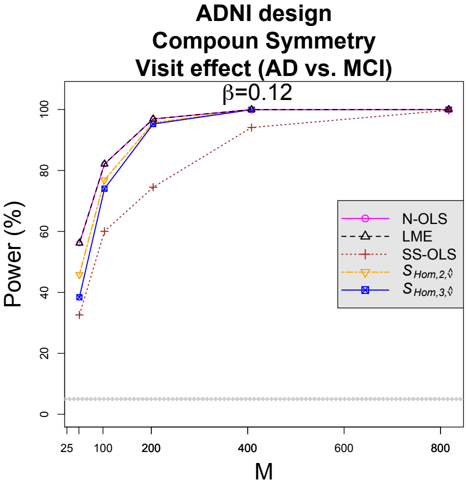
"·": accurate "$+$": overconfident "$-$": overconservative

Power analysis

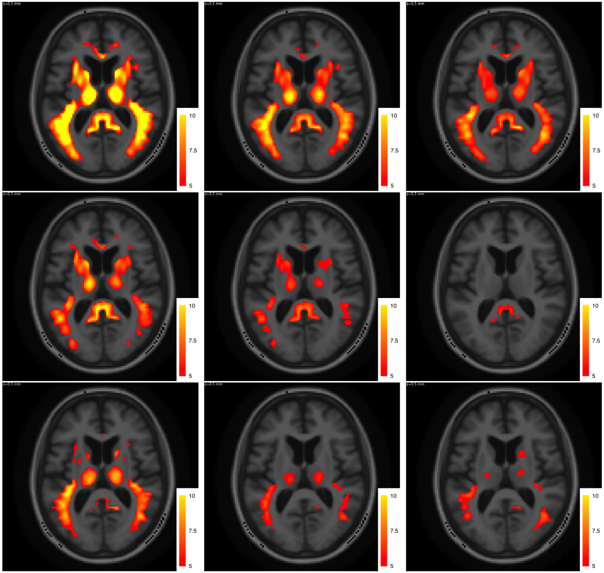


Power analysis

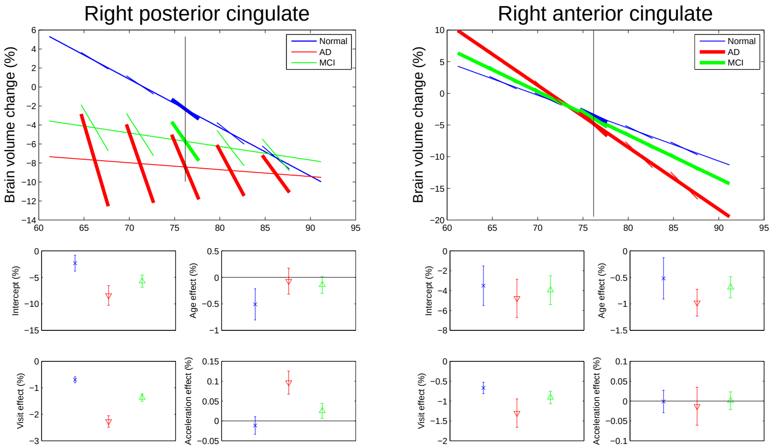




Real ADNI data: Visit effect on the brain atrophy



Fitted models in 2 particular voxels



Outline

Summary

- Longitudinal standard methods not really appropriate to neuroimaging data:
  - Convergence issues with LME
  - N-OLS & LME with random intercepts: issues when CS does not hold
  - Under unbalanced design, SS-OLS may be inaccurate and its power quite poor
- The SwE method
  - Accurate in a large range of settings
  - Easy to specify
  - No iteration needed
    - Quite fast
    - No convergence issues
  - Can accommodate pure between covariates
  - But, careful in small samples:
    - Adjustment essential
    - less powerful than N-OLS or LME (Solution: spatial regularisation)

Thanks for your attention!