# Region-Based and Voxel-Wise Analysis of Medical Images Using Latent Variables

Brice Ozenne[1,2], Esben Budtz-Jørgensen[2], Martin Nørgaard[1]

[1] Neurobiology Reasearch Unit, Rigshospitalet, Copenhagen, Denmark.
[2] University of Copenhagen, Section of Biostatistics, Copenhagen, Denmark.

NBBC19 - June 5th

# Context

Neurobiology Research Unit, Cophenagen University Hospital

# Aim of the talk

Present two strategies used to relate brain imaging data to clinical outcomes:

- regional analysis using Latent variable models (LVM)
- voxel-wise analysis using Partial Least Squares (PLS)

Discuss our concerns and methodological developments.

# Typical neuroimaging study (Ebert et al., 2019)

Association study:

- After a mild traumatic brain injury,
  is there a neuroinflammatory response in the brain?

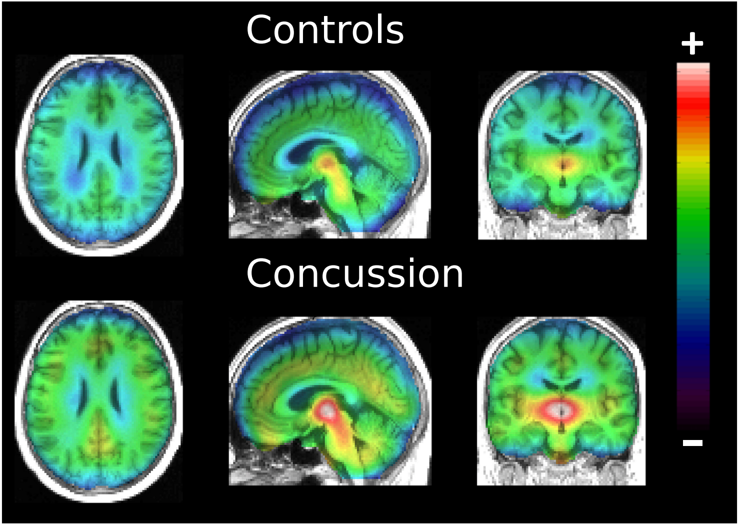
Many measurements, few subjects:

- SPECT measurement over the whole brain
- 22 healthy controls and 14 patients

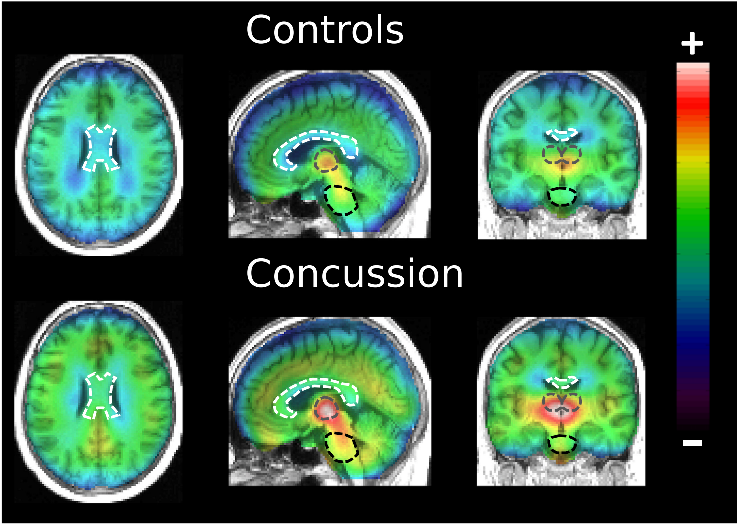
Often, some complications:

- genetic factors can influence the SPECT measurements

# Processed data "averaged" over individuals



# Example of ROIs




Corpus Callosum (white), thalamus (grey), and pons (black).

# LVM

# Latent variable models (LVM)

Consider:

- $m$ endogenous random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$
- $l$ exogenous random variables $\boldsymbol{X} = (X_1, \ldots, X_l)$
- a set of latent variables $\boldsymbol{\eta}$

A LVM is defined by

- a measurement model:

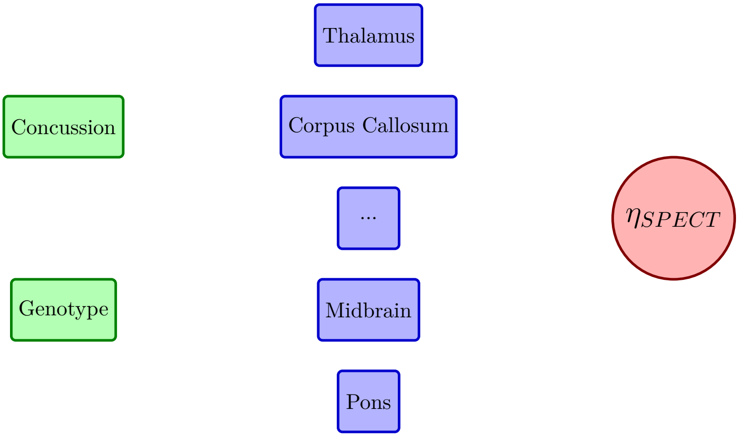
$$\boldsymbol{Y}_i = \boldsymbol{\nu} + \boldsymbol{\eta}_i \Lambda + \boldsymbol{X}_i K + \boldsymbol{\varepsilon}_i, \text{ where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, \Sigma_\varepsilon\right)$$

- a structural model:

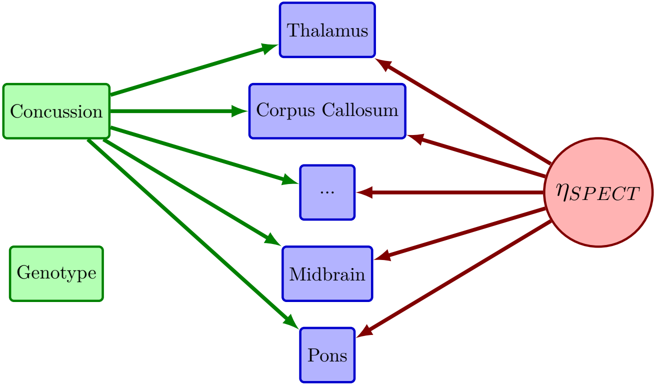
$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\eta}_i B + \boldsymbol{X}_i \Gamma + \boldsymbol{\zeta}_i, \text{ where } \boldsymbol{\zeta}_i \sim \mathcal{N}\left(0, \Sigma_\zeta\right)$$

- identifiability constrains, e.g. $\nu_1 = 0$, $\lambda_1 = 1$, $diag(B) = \mathbf{0}$

# Back to the example



# Path diagram




$$Y_{ir} = \nu_r + \eta_i \lambda_r + C_i k_r^C + G_i k_r^G + \varepsilon_{ir} \text{ where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, diag(\sigma_1^2, \ldots, \sigma_9^2)\right)$$

$$\eta_i = \alpha + \zeta_i, \text{ where } \zeta_i \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$$

# LVM as a Gaussian model

$$\boldsymbol{Y}_i|\boldsymbol{X}_i \sim \mathcal{N}\left(\mu(\Theta, \boldsymbol{X}_i), \Omega(\Theta)\right)$$

with a specific structure for the conditional mean:

$$\mu(\Theta, \boldsymbol{X}_i) = \boldsymbol{\nu} + \boldsymbol{\alpha}(1 - B)^{-1}\Lambda + \boldsymbol{X}_i\Gamma(1 - B)^{-1}\Lambda + \boldsymbol{X}_i K$$

and the conditional variance:

$$\Omega(\Theta) = \Lambda^{\intercal}(1 - B)^{-\intercal}\Sigma_{\zeta}(1 - B)^{-1}\Lambda + \Sigma_{\varepsilon}$$

# LVM as a Gaussian model

$$\boldsymbol{Y}_i|\boldsymbol{X}_i \sim \mathcal{N}\left(\mu(\Theta, \boldsymbol{X}_i), \Omega(\Theta)\right)$$

with a specific structure for the conditional mean:

$$\mu(\Theta, \boldsymbol{X}_i) = \boldsymbol{\nu} + \boldsymbol{\alpha}(1-B)^{-1}\Lambda + \boldsymbol{X}_i\Gamma(1-B)^{-1}\Lambda + \boldsymbol{X}_i K$$

and the conditional variance:

$$\Omega(\Theta) = \Lambda^{\intercal}(1-B)^{-\intercal}\Sigma_{\zeta}(1-B)^{-1}\Lambda + \Sigma_{\varepsilon}$$

In our example, the conditional variance is:

$$\mathbb{C}ov\left[Y_{ir}, Y_{ir*}\right] = \begin{cases} \lambda_r^2\tau + \sigma_r^2 \text{ if } r = r* \\ \lambda_r\lambda_{r*}\tau \text{ if } r \neq r* \end{cases}$$

# LVM as a Gaussian model

$$\boldsymbol{Y}_i|\boldsymbol{X}_i \sim \mathcal{N}\left(\mu(\Theta, \boldsymbol{X}_i), \Omega(\Theta)\right)$$

with a specific structure for the conditional mean:

$$\mu(\Theta, \boldsymbol{X}_i) = \boldsymbol{\nu} + \boldsymbol{\alpha}(1-B)^{-1}\Lambda + \boldsymbol{X}_i\Gamma(1-B)^{-1}\Lambda + \boldsymbol{X}_i K$$

and the conditional variance:

$$\Omega(\Theta) = \Lambda^{\intercal}(1-B)^{-\intercal}\Sigma_{\zeta}(1-B)^{-1}\Lambda + \Sigma_{\varepsilon}$$

In our example, the conditional variance is:

$$\mathbb{C}ov\left[Y_{ir}, Y_{ir*}\right] = \begin{cases} \lambda_r^2\tau + \sigma_r^2 \text{ if } r = r* \\ \lambda_r\lambda_{r*}\tau \text{ if } r \neq r* \end{cases}$$

(between random intercept model and unstructured covariance)

# Clinical hypotheses

Is there any effect at all of concussion?

$$(\mathcal{H}_0^a) : \beta_1 = \beta_2 = \ldots = \beta_9 = 0$$

Is the effect of concussion the same in all regions?

$$(\mathcal{H}_0^b) : \beta_1 = \beta_2 = \ldots = \beta_9$$

In which region is there an effect of concussion?

$$(\mathcal{H}_0^{c1}) : \beta_1 = 0$$

$$\ldots$$

$$(\mathcal{H}_0^{c9}) : \beta_9 = 0$$

# Test the clinical hypotheses

Easy. Estimate the LVM by maximum likelihood (ML):

- $\hat{\Theta}$: ML estimate of the model parameters
- $\hat{\Sigma}_{\hat{\Theta}}$: Estimate of the variance-covariance of the model parameters
- $C$: Contrast matrix

Wald test:

$$\left(C\hat{\Theta}\right)^{\intercal}\left(C\hat{\Sigma}_{\hat{\Theta}}C^{\intercal}\right)^{-1}\left(C\hat{\Theta}\right) \xrightarrow[n\rightarrow\infty]{} \chi^2_r$$

```
data: chisq = 24.193, df = 9, p-value = 0.004006
```

# Test the clinical hypotheses

Easy. Estimate the LVM by maximum likelihood (ML):

- $\hat{\Theta}$: ML estimate of the model parameters
- $\hat{\Sigma}_{\hat{\Theta}}$: Estimate of the variance-covariance of the model parameters
- $C$: Contrast matrix

Wald test:

$$\left(C\hat{\Theta}\right)^{\intercal}\left(C\hat{\Sigma}_{\hat{\Theta}}C^{\intercal}\right)^{-1}\left(C\hat{\Theta}\right) \xrightarrow[n\to\infty]{} \chi^2_r$$

```
data: chisq = 24.193, df = 9, p-value = 0.004006
```

Permutation test: p=0.0166 (1h, 10 000 samples, false cv 0.13%)

# Test the clinical hypotheses

Easy. Estimate the LVM by maximum likelihood (ML):

- $\hat{\Theta}$: ML estimate of the model parameters
- $\hat{\Sigma}_{\hat{\Theta}}$: Estimate of the variance-covariance of the model parameters
- $C$: Contrast matrix

Wald test:

$$\left(C\hat{\Theta}\right)^{\intercal}\left(C\hat{\Sigma}_{\hat{\Theta}}C^{\intercal}\right)^{-1}\left(C\hat{\Theta}\right) \xrightarrow[n\rightarrow\infty]{} \chi^2_r$$

```
data: chisq = 24.193, df = 9, p-value = 0.004006
```

Permutation test: p=0.0166 (1h, 10 000 samples, false cv 0.13%)
Proposed correction: p=0.011 (1s)

# Back to univariate linear regression

For the Wald test, asymptotically:

$$\frac{\hat{\theta}}{\hat{\sigma}_{\hat{\theta}}} = \frac{\hat{\theta}}{\sqrt{(X^\intercal X)^{-1}\hat{\sigma}^2}} \underset{\mathcal{H}_0}{\sim} \mathcal{N}(0,1)$$

where ˆ stands for the maximum likelihood estimate (MLE).

# Back to univariate linear regression

For the Wald test, asymptotically:

$$\frac{\hat{\theta}}{\hat{\sigma}_{\hat{\theta}}} = \frac{\hat{\theta}}{\sqrt{(X^{\intercal}X)^{-1}\hat{\sigma}^2}} \underset{\mathcal{H}_0}{\sim} \mathcal{N}(0,1)$$

where ˆ stands for the maximum likelihood estimate (MLE).

Standard corrections:

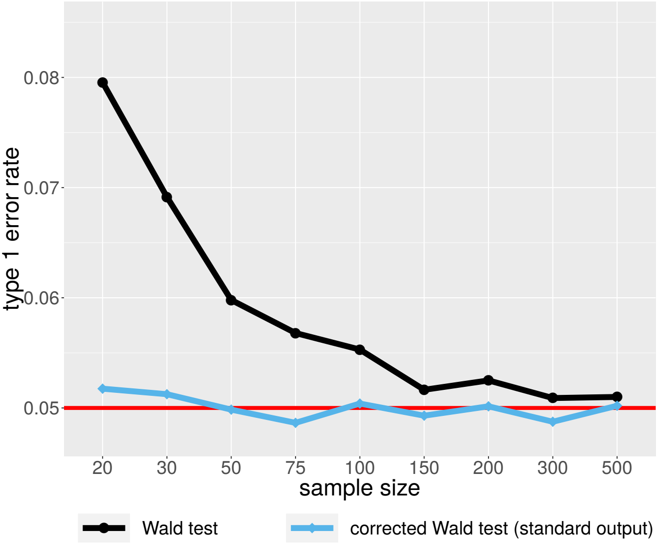
(A) correct finite sample bias of MLE:

$$\hat{\sigma}^{2,c} = \frac{n}{n-p}\hat{\sigma}^2$$

(B) use a t-distribution.

(C) correct the degrees of freedom: $n-p$ instead of $n$.

# Back to univariate linear regression



# (A) bias correction in LVM (1/2)

Denoting:

$$\xi_i(\hat{\Theta}) = \boldsymbol{Y}_i - \mu(\hat{\Theta}, \boldsymbol{X}_i)$$

the observed residuals, we can show that their variance is smaller than the (true) conditional variance of $Y$:

$$\mathbb{E}\left[\xi_i(\hat{\Theta})^\intercal \xi_i(\hat{\Theta})\right] = \Omega(\Theta) - \Psi_i + o_p(n^{-1})$$

where $\Psi_i$ is the first order bias:

$$\Psi_i = \frac{\partial \mu(\Theta, \boldsymbol{X}_i)}{\partial \Theta}^\intercal \Sigma_{\hat{\Theta}} \frac{\partial \mu(\Theta, \boldsymbol{X}_i)}{\partial \Theta}$$

# (A) bias correction in LVM (2/2)

Assuming that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i(\hat{\Theta})^{\mathsf{T}}\xi_i(\hat{\Theta})\right]$ and $\Omega(\Theta)$ are subject to the same type of bias

- we can use $\Psi = \frac{1}{n}\sum_{i=1}^{n}\Psi_i$ to correct $\Omega(\Theta)$
- and the new $\Omega(\Theta)$ to better estimate $\frac{1}{n}\sum_{i=1}^{n}\Psi_i$

# (A) bias correction in LVM (2/2)

Assuming that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i(\hat{\Theta})^{\mathsf{T}}\xi_i(\hat{\Theta})\right]$ and $\Omega(\Theta)$ are subject to the same type of bias

- we can use $\Psi = \frac{1}{n}\sum_{i=1}^{n}\Psi_i$ to correct $\Omega(\Theta)$
- and the new $\Omega(\Theta)$ to better estimate $\frac{1}{n}\sum_{i=1}^{n}\Psi_i$

**Example**: $Y_i = \beta X_i + \varepsilon_i$, with $\varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$

$$\hat{\Psi}_i = \hat{\sigma}^2 \frac{p}{n} \text{ and } \tilde{\sigma}^2 = \hat{\sigma}^2 + \hat{\Psi} = \left(1 + \frac{p}{n}\right)\hat{\sigma}^2$$

Iterating the procedure gives:

$$\tilde{\sigma}^{\infty} = \sum_{k=0}^{\infty}\left(\frac{p}{n}\right)^k \hat{\sigma}^2 = \frac{n}{n-p}\hat{\sigma}^2$$

# (B) Satterthwaite approximation

We model the distribution of the variance of our estimator:

$$k\hat{\sigma}^2_{\hat{\theta}} \sim \chi^2(df)$$

We identify $k$ and $df$ using the method of moments:

$$\mathbb{E}\left[k\hat{\sigma}^2_{\hat{\theta}}\right] = \mathbb{E}\left[\chi^2(df)\right] = df$$
$$\mathbb{V}ar\left[k\hat{\sigma}^2_{\hat{\theta}}\right] = \mathbb{V}ar\left[\chi^2(df)\right] = 2df$$

i.e.

$$df = 2\frac{\mathbb{E}\left[\hat{\sigma}^2_{\hat{\theta}}\right]^2}{\mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\theta}}\right]}$$

## (B) Estimating $\mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\theta}}\right]$

We can relate the estimated variance of our estimator to the model parameters:

$$\hat{\sigma}_{\hat{\theta}} = f(\hat{\Theta})$$

# (B) Estimating $\mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\theta}}\right]$

We can relate the estimated variance of our estimator to the model parameters:

$$\hat{\sigma}_{\hat{\theta}} = f(\hat{\Theta})$$

Estimates from maximum likelihood estimator satisfies:

$$n^{1/2}(\hat{\Theta} - \Theta) \sim \mathcal{N}\left(0, \mathcal{I}_1(\Theta)^{-1}\right)$$

# (B) Estimating $\mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\theta}}\right]$

We can relate the estimated variance of our estimator to the model parameters:

$$\hat{\sigma}_{\hat{\theta}} = f(\hat{\Theta})$$

Estimates from maximum likelihood estimator satisfies:

$$n^{1/2}(\hat{\Theta} - \Theta) \sim \mathcal{N}\left(0, \mathcal{I}_1(\Theta)^{-1}\right)$$

Then, the delta method gives:

$$n^{1/2}(\hat{\sigma}_{\hat{\Theta}} - \sigma_{\hat{\Theta}}) \sim \mathcal{N}\left(0, \nabla_{\Theta} f(\Theta)\mathcal{I}_1(\Theta)^{-1}\nabla_{\Theta} f(\Theta)\right)$$

# (C) Effective sample size

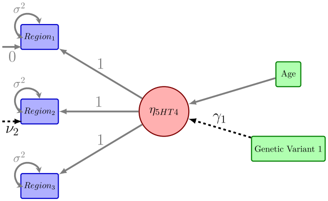
$$\mathbf{n}^c = \sum_{i=1}^{n} \frac{\partial \xi_i(\hat{\Theta})}{\partial \boldsymbol{Y}_i} = n - \sum_{i=1}^{n} \frac{\partial \mu(\hat{\Theta}, \boldsymbol{X}_i)}{\partial \boldsymbol{Y}_i}$$

where $\frac{\partial \mu(\hat{\Theta}, \boldsymbol{X}_i)}{\partial \boldsymbol{Y}_i}$ are the generalized leverage defined by (Wei et al., 1998).

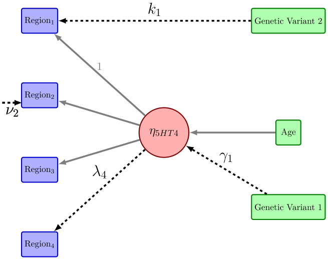
**Example**: univariate linear regression

$$\hat{n}^c = n - \sum_{i=1}^{n} \boldsymbol{X}_i \frac{\partial \beta}{\partial Y_i} = n - \sum_{i=1}^{n} \boldsymbol{X}_i \left(\boldsymbol{X}^\intercal \boldsymbol{X}\right)^{-1} \boldsymbol{X}_i{}^\intercal = n - p$$

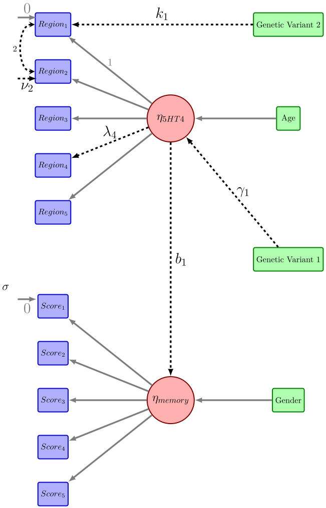
# Simulation study




Scenario A : mixed model

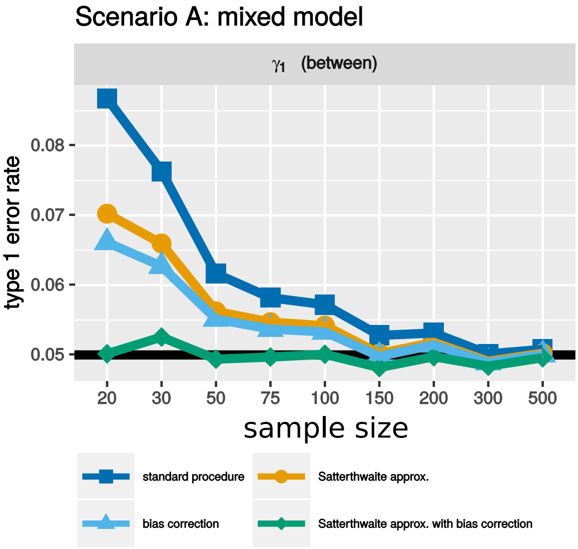



Scenario B : single factor model

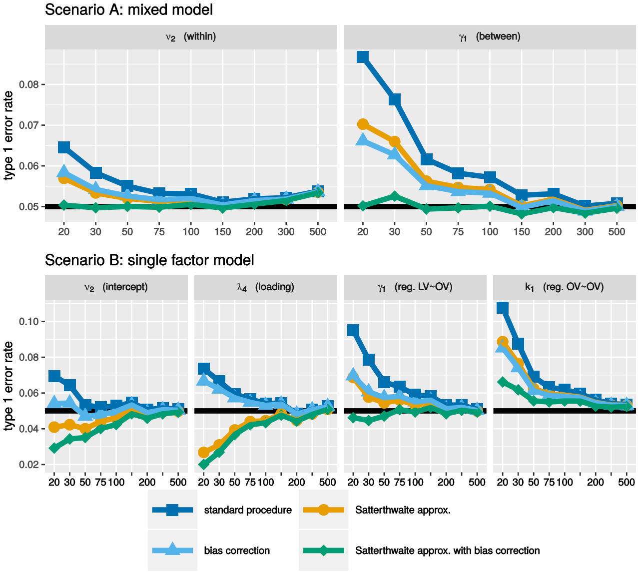



Scenario C : two latent variable model

# Type 1 error rate



# Type 1 error rate



# Diagnostic - small sample correction

| | | Empirical Student | | Modeled Student | | KS-test |
|---|---|---|---|---|---|---|
| Scenario | parameter | dispersion | df | dispersion | df | p-value |
| A | $\nu_2$ | 1.006 | 40.8 | 1 | 36.7 | 0.659 |
| | $\gamma_1$ | 1.004 | 17.9 | 1 | 18.3 | 0.961 |
| B | $\nu_2$ | 0.998 | 39.2 | 1 | 9.1 | 0.513 |
| | $\lambda_4$ | 1.067 | 117.2 | 1 | 5.9 | 0.885 |
| | $\gamma_1$ | 1.088 | 311.8 | 1 | 10.5 | 0.455 |
| | $k_1$ | 1.063 | 14.8 | 1 | 17.1 | 0.792 |
| C | $\nu_2$ | 1.026 | 45.5 | 1 | 10.1 | 0.536 |
| | $k_1$ | 1.106 | 12.6 | 1 | 17.1 | 0.936 |
| | $\lambda_4$ | 1.080 | 145.9 | 1 | 6.0 | 0.237 |
| | $\gamma_1$ | 1.152 | 49.6 | 1 | 10.1 | 0.135 |
| | $b_1$ | 1.127 | 783296.7 | 1 | 3.6 | $<0.001$ |
| | $\sigma_{12}$ | 1.074 | 934003.5 | 1 | 7.8 | $<0.001$ |

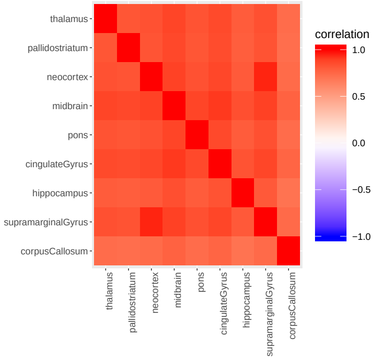
# Handling multiple comparisons

In which region is there an effect of concussion?

$$(\mathcal{H}_0^{c1}) : \beta_1 = 0$$
$$\ldots$$
$$(\mathcal{H}_0^{c9}) : \beta_9 = 0$$

# Handling multiple comparisons

Correlation between the test statistics:




In which region is there an effect of concussion?

$$(\mathcal{H}_0^{c1}) : \beta_1 = 0$$

$$\dots$$

$$(\mathcal{H}_0^{c9}) : \beta_9 = 0$$

# Notations

From maximum likelihood theory, we know that:

$$\sqrt{n}\left(\hat{\Theta} - \Theta\right) \overset{d}{\sim} \mathcal{N}\left(0, \mathcal{I}_1(\Theta)^{-1}\right)$$

We consider the contrast matrix $C$ such that we want to test:

$$C\Theta = \mathbf{0}$$

We denote the vector of Wald statistics by:

$$\mathbf{T} = diag(C\Sigma_{\hat{\Theta}}C^{\intercal})^{-\frac{1}{2}}C\Theta$$

# Max test procedure (Hothorn et al., 2008)

The vector of Wald statistics is asymptotically normally distributed:

$$\sqrt{n}\,\mathbf{T} \underset{\mathcal{H}_0}{\overset{d}{\sim}} \mathcal{N}(0, \Sigma_{\mathbf{T}})$$

with $\Sigma_{\mathbf{T}} = f(C, \mathcal{I}_1(\Theta))$

We define

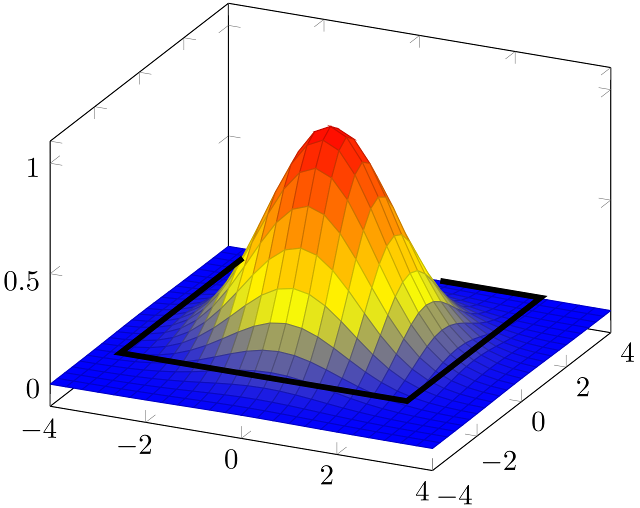
- the max statistic: $|T|_{max} = max(\mathbf{T})$
- the observed max statistic: $|t|_{max}$ (e.g. $t_1$)

We obtain an adjusted p-value for the largest observed $T$-statistic by computing (under the null):

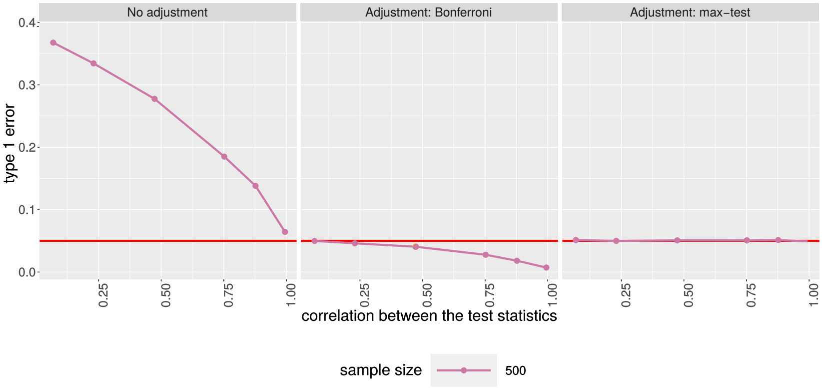
$$1 - \mathbb{P}\left[|T|_{max} < |t|_{max}\right]$$

# Integration step using `mvtnorm`

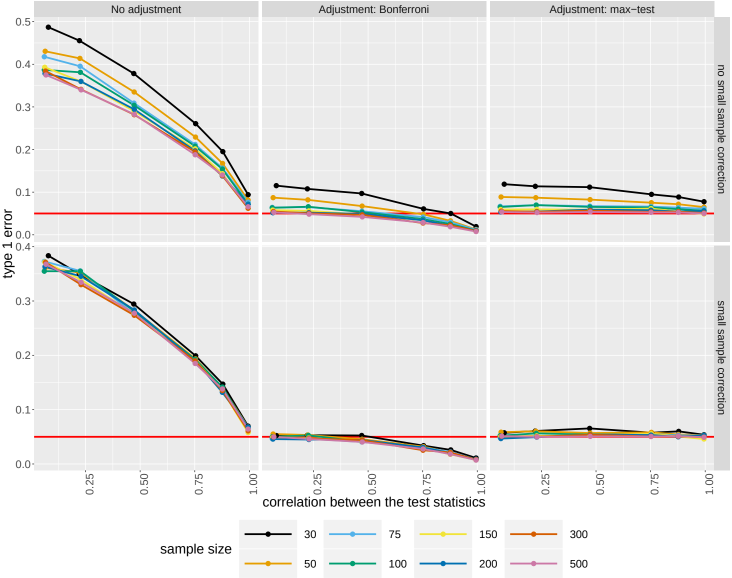
Black thick line: $|t|_{max}$



# Simulation study



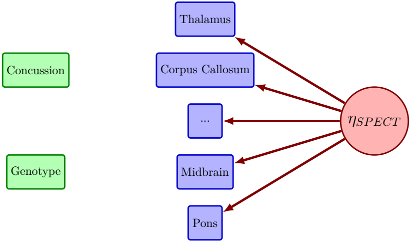
# Simulation study



# Back to the example

| | | p-value | | |
|---|---|---|---|---|
| region | effect (%) | unadjusted | Bonferroni | max-test |
| thalamus | 12.75 | 0.23 | 1.00 | 0.53 |
| pallidostriatum | 12.03 | 0.18 | 1.00 | 0.43 |
| neocortex | 4.38 | 0.53 | 1.00 | 0.97 |
| midbrain | 10.40 | 1.25 | 1.00 | 0.51 |
| pons | 1.56 | 0.18 | 1.00 | 1.00 |
| cingulate gyrus | 17.28 | 0.03 | 0.30 | 0.11 |
| hippocampus | 12.64 | 0.14 | 1.00 | 0.36 |
| supramarginal gyrus | 5.22 | 0.55 | 1.00 | 0.94 |
| corpus callosum | 19.02 | 0.03 | 0.23 | 0.09 |

# Extension to score tests



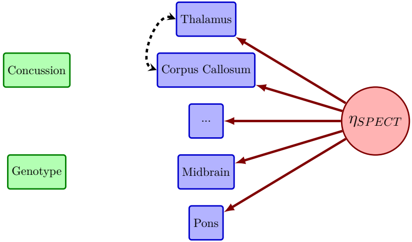

Initially:

$$\boldsymbol{Y}_i = \boldsymbol{\nu} + \boldsymbol{\eta}_i \Lambda + \boldsymbol{X}_i K + \boldsymbol{\varepsilon}_i, \text{ where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, \Sigma_\varepsilon = diag(\sigma_1^2, \ldots, \sigma_9^2)\right)$$

Investigate model mispecification in $\Sigma_\varepsilon$ using score tests

- adjustment for multiple comparisons

# Extension to score tests



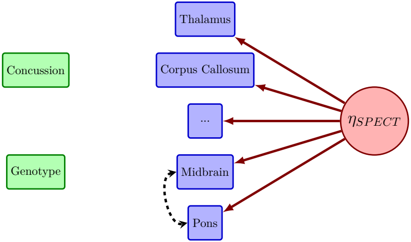

Initially:

$$\boldsymbol{Y}_i = \boldsymbol{\nu} + \boldsymbol{\eta}_i \Lambda + \boldsymbol{X}_i K + \boldsymbol{\varepsilon}_i, \text{ where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, \Sigma_\varepsilon = diag(\sigma_1^2, \dots, \sigma_9^2)\right)$$

Investigate model mispecification in $\Sigma_\varepsilon$ using score tests

- adjustment for multiple comparisons

# Extension to score tests




Initially:

$$\boldsymbol{Y}_i = \boldsymbol{\nu} + \boldsymbol{\eta}_i \Lambda + \boldsymbol{X}_i K + \boldsymbol{\varepsilon}_i, \text{ where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, \Sigma_\varepsilon = diag(\sigma_1^2, \dots, \sigma_9^2)\right)$$

Investigate model mispecification in $\Sigma_\varepsilon$ using score tests

- adjustment for multiple comparisons

# Conclusion

LVM are a flexible framework to analyze regional imaging data:

- R package `lava`

We propose inference tools in the R package `lavaSearch2`:

- inference in small samples (not perfect but better than `lava`)
- adjustment for multiple comparisons (via `multcomp` / `mvtnorm`)

Perspectives

- influence of model selection on the statistical inference
- LVM for large number of $X$ and $Y$

# Partial least square (PLS) for voxel-wise analysis

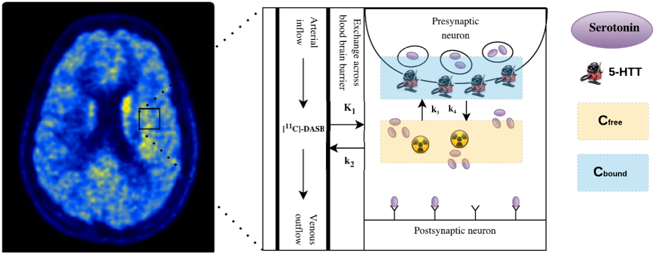
Nørgaard et al. (2017) based on Churchill et al. (2013); Krishnan et al. (2011)

# Context

Finding areas of the brain associated with seasonal affective disorder.

We observe realizations of:

- a 3-dimensional variable $X$ (here serotonin transporter binding)
- a bivariate variable $Y$ (here season and disease status)



# PLS - General idea

Find the linear combination of $X$ that correlates best with $Y$:

- $l_x = Xv_x$

i.e. maximizing $\mathbb{C}ov(l_x, Y)$

# PLS - General idea

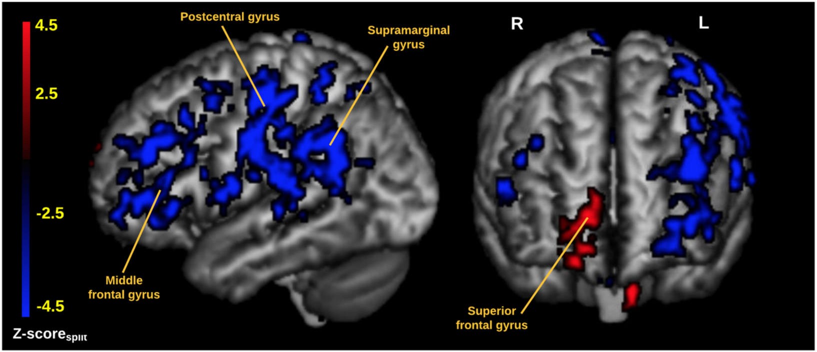
Find the linear combination of $X$ that correlates best with $Y$:

- $l_x = Xv_x$

i.e. maximizing $\mathbb{C}ov(l_x, Y)$

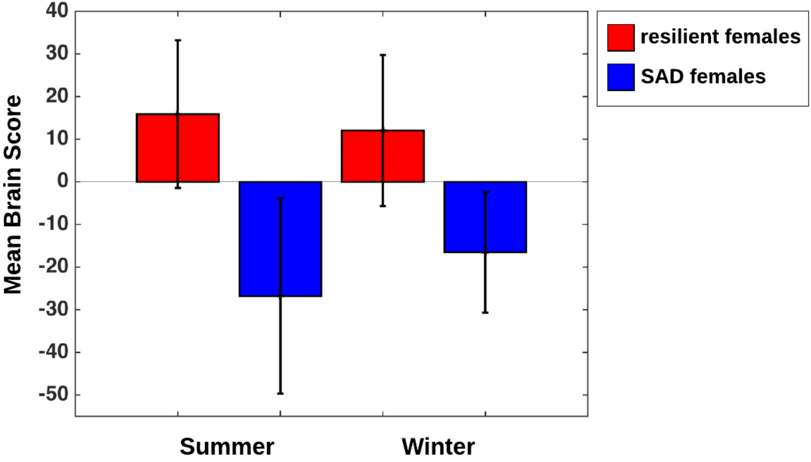
- compute the SVD of $Y^\intercal X = U\Delta V^\intercal$
- $l_x = V[,1]$
- covariance: $\Delta[1,1]$

# Brain network according to PLS

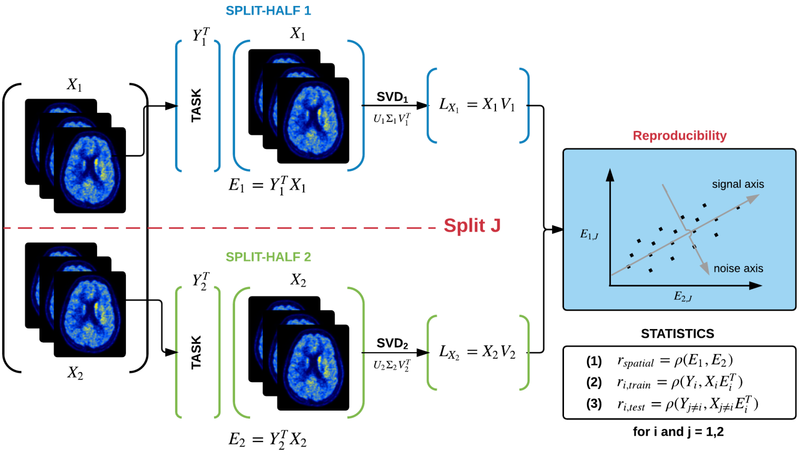



$$Z_{score} = \frac{\text{loading of the voxel}}{sd(loading)}$$

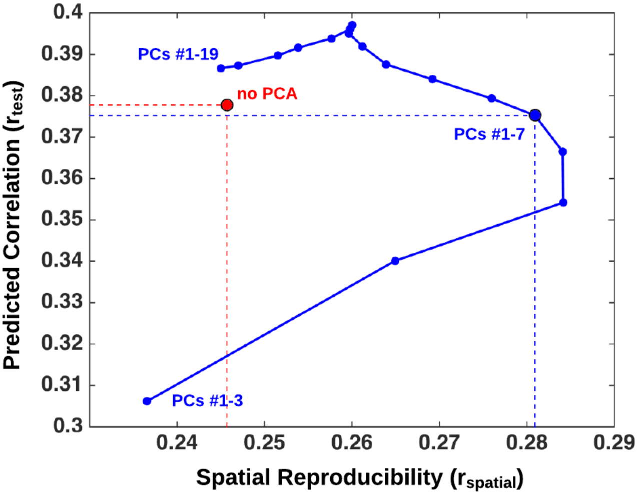
# Individual latent scores (mean, 95% CI)



# Split half resampling



# Reproducibility vs. correlation




$p_{\text{test}} = 0.011, \; p_{\text{spatial}} = 0.016$

# Reference I


Churchill, N., Spring, R., Abdi, H., Kovacevic, N., McIntosh, A. R., and Strother, S. (2013). The stability of behavioral pls results in ill-posed neuroimaging problems. In *New perspectives in partial least squares and related methods*, pages 171–183. Springer.

Ebert, S. E., Jensen, P., Ozenne, B., Armand, S., Svarer, C., Stenbaek, D. S., Moeller, K., Dyssegaard, A., Thomsen, G., Steinmetz, J., Forchhammer, B. H., Knudsen, G. M., and Pinborg, L. H. (2019). Molecular imaging of neuroinflammation in patients after mild traumatic brain injury; a longitudinal 123i-clinde spect study. *European Journal of Neurology*.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical journal*, 50(3):346–363.

# Reference II

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial least squares (pls) methods for neuroimaging: a tutorial and review. *Neuroimage*, 56(2):455–475.

Nørgaard, M., Ganz, M., Svarer, C., Fisher, P. M., Churchill, N. W., Beliveau, V., Grady, C., Strother, S. C., and Knudsen, G. M. (2017). Brain networks implicated in seasonal affective disorder: a neuroimaging pet study of the serotonin transporter. *Frontiers in neuroscience*, 11:614.

Wei, B.-C., Hu, Y.-Q., and Fung, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of statistics*, 25(1):25–37.

# Information matrix

$$\mathcal{I}(\theta,\theta') = \frac{n}{2} tr \left( \Omega(\Theta)^{-1} \frac{\partial \Omega(\Theta)}{\partial \theta} \Omega(\Theta)^{-1} \frac{\partial \Omega(\Theta)}{\partial \theta'} \right) + \sum_{i=1}^{n} \frac{\partial \mu(\Theta, \boldsymbol{X}_i)}{\partial \theta} \Omega(\Theta)^{-1} \frac{\partial \mu(\Theta, \boldsymbol{X}_i)}{\partial \theta'}^{\mathsf{T}}$$

# Example Bias correction

$$\hat{\Psi}_i = \begin{bmatrix} X_i & 0 \end{bmatrix} \begin{bmatrix} \frac{\hat{\sigma}^2}{X^\intercal X} & 0 \\ 0 & \frac{2\hat{\sigma}^2}{n} \end{bmatrix} \begin{bmatrix} X_i \\ 0 \end{bmatrix} = X_i \left(X^\intercal X\right)^{-1} X_i{}^\intercal \hat{\sigma}^2$$

$$\hat{\Psi} = \hat{\sigma}^2 \frac{p}{n}$$

$$\text{so } \tilde{\sigma}^2 = \hat{\sigma}^2 + \hat{\Psi} = \left(1 + \frac{p}{n}\right) \hat{\sigma}^2$$

So we have removed the first order bias in $\hat{\sigma}^c$:

$$\mathbb{E}\left[\tilde{\sigma}^2\right] = \frac{n+p}{n} \frac{n-p}{n} \sigma^2 = \left(1 - \frac{p^2}{n^2}\right) \sigma^2$$

# (C) Example Satterthwaite correction

The variance covariance matrix of $\Theta = (\beta, \sigma)$ is:

$$\Sigma_\Theta = \begin{bmatrix} \frac{\sigma^2}{X^\intercal X} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

$$\text{so } \nabla_\Theta \Sigma_{\hat{\Theta}} = (\nabla_\beta \Sigma_{\hat{\Theta}}, \nabla_{\sigma^2} \Sigma_{\hat{\Theta}}) = \left( \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{X^\intercal X} & 0 \\ 0 & \frac{4\sigma^2}{n} \end{bmatrix} \right)$$

$$\text{and } \mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\beta}}\right] = \nabla_\Theta \sigma^2_{\hat{\theta}}(\hat{\Theta}) \quad \hat{\Sigma}_{\hat{\Theta}} \quad \nabla_\Theta \sigma^2_{\hat{\beta}}(\hat{\Theta})^\intercal$$

$$= \begin{bmatrix} 0 & \frac{1}{X^\intercal X} \end{bmatrix} \begin{bmatrix} \frac{\hat{\sigma}^2}{X^\intercal X} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{X^\intercal X} \end{bmatrix} = 2\frac{\hat{\sigma}^4}{(X^\intercal X)^2 n} = 2\frac{\left(\hat{\sigma}^2_{\hat{\beta}}\right)^2}{n}$$

$$\text{so } \widehat{df} = 2\left(\hat{\sigma}^2_{\hat{\beta}}\right)^2 \Bigg/ \left(2\frac{\left(\hat{\sigma}^2_{\hat{\beta}}\right)^2}{n}\right) = n$$