Detecting changes in the covariance structure of functional time series with application to fMRI data [9]

David Veitch

University of Toronto

Nov 18 2022

UNIVERSITY OF TORONTO

Agenda

Visual Depiction

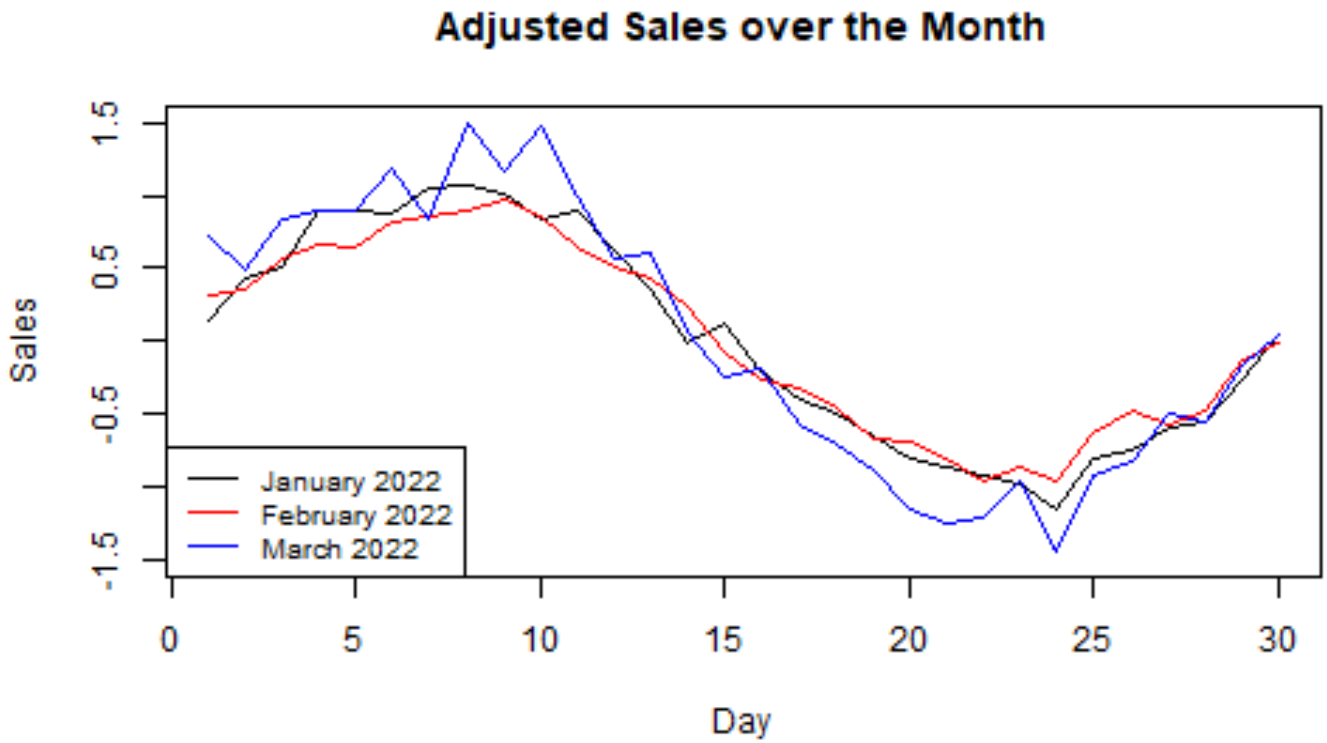

Figure: Example of 3 samples of functional time series.

In a **functional time series** your observations $X_t,\ t = 1, \ldots, n$ are functions.

Visual Depiction

It is harder to visualize for the brain, but we can treat the data from MRI images as a 3D function collected over time.

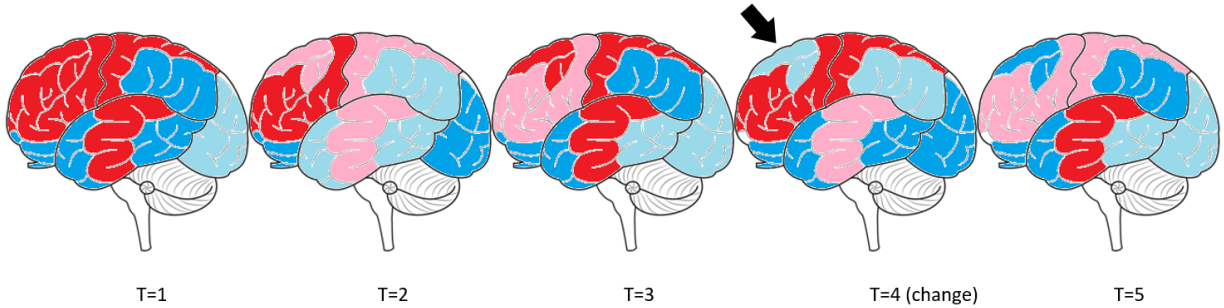

Figure: Visual representation of change in covariance between brain regions [1].

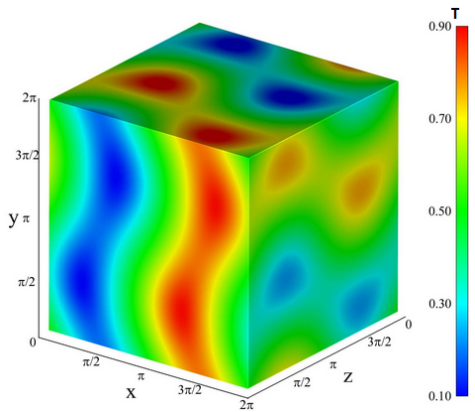

Figure: Visualization of a smooth function of 3 variables[3].

Paper Overview

Stoehr, Christina, John AD Aston, and Claudia Kirch. "Detecting changes in the covariance structure of functional time series with application to fMRI data." *Econometrics and Statistics* 18 (2021): 44-62.

- Many fMRI studies examine correlation betweeen brain regions
- These studies rely on strong assumption that fMRI data is both first (mean) and second (covariance) order stationary (i.e. not changing over time)
- Previous work by authors [5] on testing for changes in the mean of functional time series.
- **This paper:** model brain as a functional time series $\rightarrow$ look for changes covariance structure of these functions

Why Functional Time Series Approach?

- Data being discretized when it is not originally discrete
- Voxelwise approach leads to multiple testing problem, and low power for small, spatially-distributed signals
- Functional approach $\rightarrow$ leverage functional properties (e.g. spatial smoothness) and then utilize a variety of statistical tools

Definition 1 (Linear Operator)

A mapping $A : E \to F$ where E, F vector spaces over field K called a linear operator from E to F if

$$A(x + y) = Ax + Ay \tag{1}$$

$$A(\lambda x) = \lambda Ax. \tag{2}$$

Definition 2 ($\mathcal{L}$)

Space of bounded linear operators on separable Hilbert space H with norm

$$||\Phi||_{\mathcal{L}} = \sup\{||\Phi|| \mid ||x|| \leq 1\}.$$

Definition 3 (Compact)

A operator $\Phi \in \mathcal{L}$ is compact if $\exists$ two orthonormal bases $\{v_j\}, \{f_j\}$ and a real sequence $\{\lambda_j\} \to 0$ such that

$$\Phi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle f_j.$$

Example

$$\Phi = \Sigma = Q \Lambda Q^T \tag{3}$$

$$\Sigma x = \lambda_1 \langle v_1, x \rangle v_1 + \lambda_2 \langle v_2, x \rangle v_2 \tag{4}$$

Definition 4 (Symmetric)

Operator $\Phi \in \mathcal{L}$ is symmetric if

$$\langle \Phi(x), y \rangle = \langle x, \Phi(y) \rangle.$$

Definition 5 (Positive-Definite)

Operator $\Phi \in \mathcal{L}$ is positive-definite if

$$\langle \Phi(x), x \rangle \geq 0.$$

Fact 6

A symmetric positive-definite Hilbert-Schmidt operator Φ admits the decomposition

$$\Phi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j$$

with orthonormal v_j which are eigenfunctions of Φ (i.e. $\Phi(v_j) = \lambda_j v_j$).

Paper's Theoretical Setup

Consider the model with a constant mean funciton

$$X_t(s) = \mu(s) + Y_t(s) \quad 1 \leq t \leq n \tag{5}$$

where t a time point and s a spatial coordinate on a set $\mathcal{Z}$ (e.g $[0,1]^3$). Constant (in time) mean function is $\mu(s)$, $Y_t(s)$ are mean 0 random fluctuations with a possibly time-dependent covariance structure.

Definition 7 ($\mathcal{L}^2(\mathcal{Z})$)

$\mathcal{L}^2(\mathcal{Z})$ is the space of random functions X on $\mathcal{Z}$ with $\int_{\mathcal{Z}} E[X^2(s)]ds$.

Definition 8 (Covariance Operator)

Let $\{X_t(\cdot) \mid 1 \leq t \leq n\} \in \mathcal{L}^2(\mathcal{Z})$ be a functional time series, where $\mathcal{Z}$ a compact set. The covariance operator $C_t : \mathcal{L}^2(\mathcal{Z}) \to \mathcal{L}^2(\mathcal{Z})$ is defined by

$$C_t(z) = \int c_t(\cdot, s) z(s) ds$$

where

$$c_t(u, s) = \text{Cov}(X_t(u), X_t(s))$$

is the covariance kernel of $X_t(\cdot)$.

Changepoint Model

Paper looks at two changepoint models, the AMOC (at most one change), and the epidemic ($C_1 \to C_2 \to C_1$). I will focus on AMOC.

$$Y_t(s) = Y_t^{(1)}(s)1_{\{1 \le t \le \theta n\}} + Y_t^{(2)}(s)1_{\{\theta n < t \le n\}} \quad 1 \le t \le n \tag{6}$$

$$c(u,s) = \text{Cov}\left(Y_t^{(1)}(u), Y_t^{(1)}(s)\right) \tag{7}$$

$$c(u,s) + \delta(u,s) = \text{Cov}\left(Y_t^{(2)}(u), Y_t^{(2)}(s)\right) \tag{8}$$

for $1 \le t \le n$ and $0 < \theta < 1$ and $c(u,s), \delta(u,s) \in \mathcal{L}^2(\mathcal{Z} \times \mathcal{Z})$.

Under AMOC model we have

$$H_0 : \ \theta = 1 \qquad H_A : \ 0 < \theta < 1. \tag{9}$$

Paper Methdology - Dimension Reduction

General Idea: Project data to lower dimension $\rightarrow$ run changepoint testing on covariances of lower dimensional projection.

Project data into d dimensional space spanned by orthonormal basis

$$\{v_k(\cdot) \mid k = 1, \ldots, d\}.$$

And then obtain projection scores via

$$\langle X_t, v_l \rangle = \int X_t(s) v_l(s) ds \quad t = 1, \ldots, n, \; l = 1, \ldots, d. \tag{10}$$

Under the data model

$$\text{Cov}(\langle X_t, v_{l_1} \rangle, \langle X_t, v_{l_2} \rangle) = \int \int c(u, s) v_{l_1}(u) v_{l_2}(s) du ds \tag{11}$$

$$+ 1_{\{\theta n < t \leq n\}} \int \int \delta(u, s) v_{l_1}(u) v_{l_2}(s) du ds. \tag{12}$$

Therefore need Equation 12 to be large to detect change.

Paper Methdology - Functional PCA

Where $\{\lambda_l \mid l \geq 1\}$ a non-negative decreasing sequence of eigenvalues of the covariance operator, and $\{v_l(\cdot) \mid l \geq 1)\}$ the eigenfunctions defined by

$$\int c(u,s)v_l(s)ds = \lambda_l v_l(u) \quad l = 1, 2, \ldots \quad u \in \mathcal{Z}. \tag{13}$$

Mercer's Lemma and Karhunen-Loeve expansion (Lemma 1.3, Theorem 1.5) [7]

$$c(u,s) = \sum_{l=1}^{\infty} \lambda_k v_l(u) v_l(s) \tag{14}$$

$$Y_t(s) = X_t(s) - \mu(s) \tag{15}$$

$$= \sum_{l=1}^{\infty} \eta_{t,l} v_l(s) \tag{16}$$

$$\eta_{t,l} = \int (X_t(s) - \mu(s)) v_l(s) ds \tag{17}$$

with the scores $\eta_{t,l}$ are uncorrelated and centred with variance λ_j.

Paper Methdology - Functional PCA in Practice

Since do not have covariance kernel, PCA based on empirical covariance function

$$\hat{c}_n(u,s) = \frac{1}{n}\sum_{t=1}^{n}(X_t(u) - \bar{X}_n(u))(X_t(s) - \bar{X}_n(s)) \tag{18}$$

$$\bar{X}_n(s) = \frac{1}{n}\sum_{t=1}^{n} X_t(s). \tag{19}$$

Once obtain eigenfunctions $\{\hat{v}_l(\cdot) \mid l = 1, \ldots, d\}$ and eigenvalues determine projection scores by

$$\hat{\eta}_{t,l} = \int (X_t(s) - \bar{X}_n(s))\hat{v}_l(s)ds \tag{20}$$

$$= \langle X_t, \hat{v}_l \rangle - \overline{\langle X, \hat{v}_l \rangle}_n \tag{21}$$

$$\overline{\langle X, \hat{v}_l \rangle}_n = \frac{1}{n}\sum_{t=1}^{n}\langle X_t, \hat{v}_l \rangle. \tag{22}$$

Paper Methdology - Obtaining Eigenfunctions

- In very high dimensions PCA on covariance matrix infeasible due to computation.
- To get around this authors assume a **separable covariance structure**

$$c((u_1, u_2, u_3), (s_1, s_2, s_3)) = c(u_1, s_1)c(u_2, s_2)c(u_3, s_3) \tag{23}$$

  which is same as if independent X, Y, Z have covariance kernels $c_X(u_1, s_1), c_Y(u_2, s_2), c_Z(u_3, s_3)$ then $A = X(u_1)Y(u_2)Z(u_3)$ has covariance kernel $c_X(u_1, s_1)c_Y(u_2, s_2)c_Z(u_3, s_3)$.
- Significantly reduces dimension. If data 100x100x100, full covariance matrix is 10^7 x 10^7 matrix. But using separable structure means estimate 3 100x100 covariance matrices.

Paper Methodology - Separable Covariance

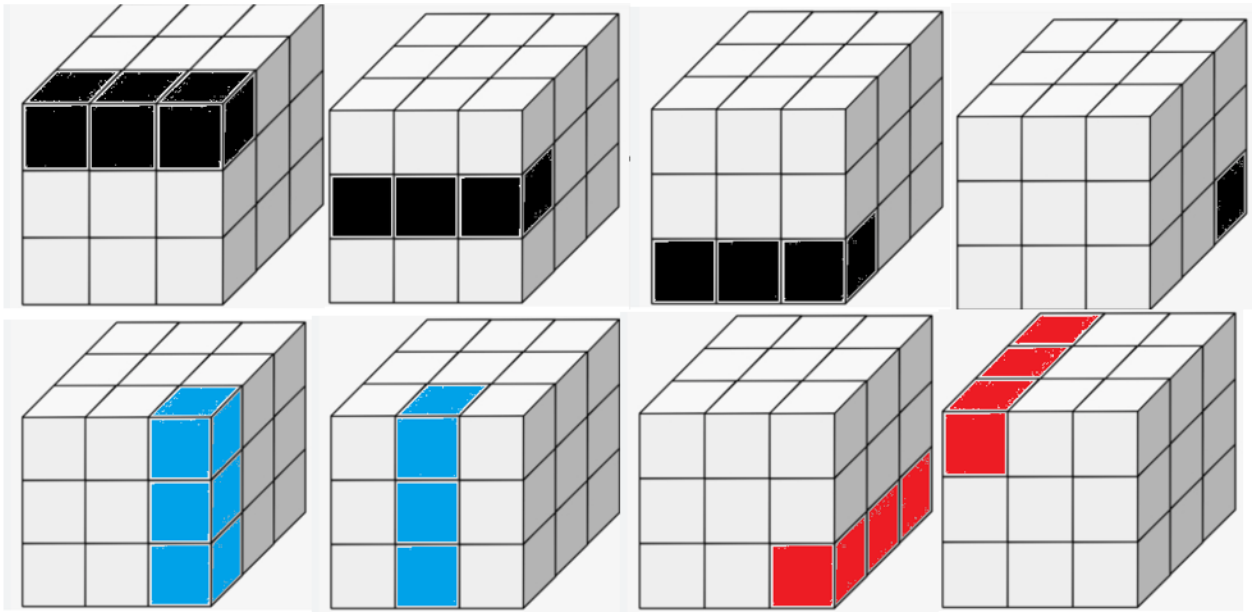

Figure: Visualization of separable covariance structure estimation. For covariance kernel in black direction, calculate empirical covariance of all black rows and then take an average. Repeat process to obtain covariance kernel in blue and red direction.

Paper Methdology - Obtaining Eigenfunctions

- Denote $\mathcal{U}_x, \mathcal{U}_y, \mathcal{U}_z$ each being a set of points in the x, y, z dimensions.
- For c_1

$$c_1(u_1, s_1) = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{|\mathcal{U}_y \times \mathcal{U}_z|} \sum_{y,z \in \mathcal{U}_y \times \mathcal{U}_z} \left(X_t(u_1, y, z) - \bar{X}_n(u_1, y, z)\right) \left(X_t(s_1, y, z) - \bar{X}_n(s_1, y, z)\right). \tag{24}$$

- Calculate c_1, c_2, c_3 as above and for each covariance matrix obtain eigenfunctions and eigenvalues $\{\hat{v}_{1,j}, \hat{\lambda}_{1,j}\}_{j=1}^{p_1}$, $\{\hat{v}_{2,j}, \hat{\lambda}_{2,j}\}_{j=1}^{p_2}$, $\{\hat{v}_{3,j}, \hat{\lambda}_{3,j}\}_{j=1}^{p_3}$.
- Order the $\hat{\lambda}_{i,j}$ and for each i select top d_i (maybe $\sqrt[3]{p_i}$) eigenfunctions for each dimension.
- Take tensor product of eigenfunctions to obtain eigenbasis[1]

$$\{\hat{v}_{1,j_1} \otimes \cdots \otimes \hat{v}_{k,j_k},\ j_1 = 1, \ldots, d_l,\ l = 1, \ldots, k\}. \tag{25}$$

[1] Fairly certain this is the correct equation, believe there is small typo in paper.

Paper Methdology - Obtaining Eigenfunctions

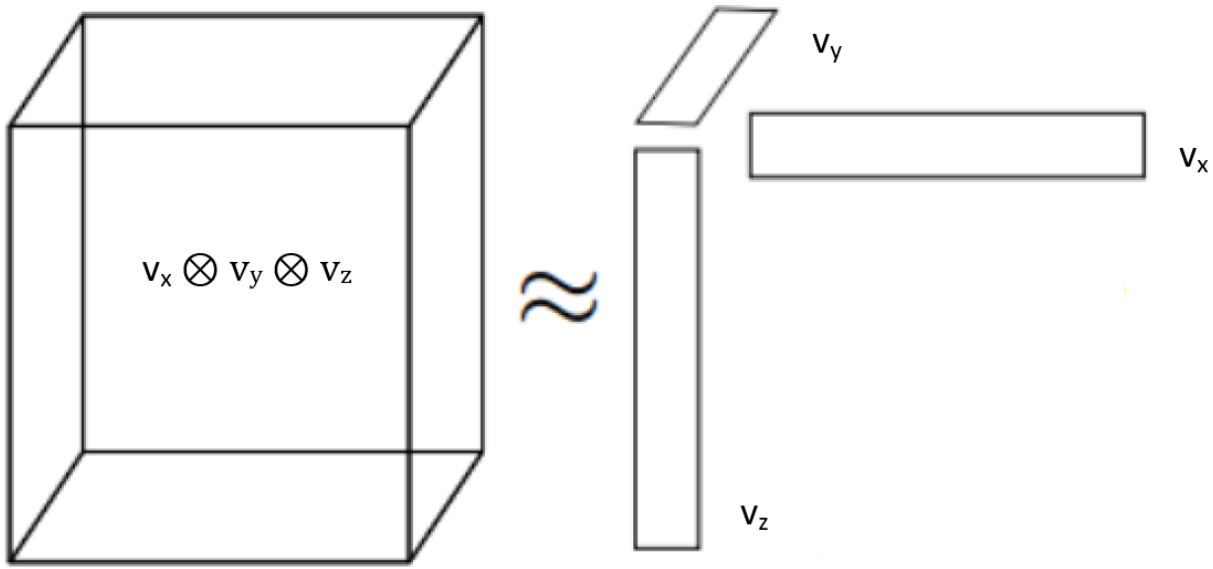

Figure: Visualization of tensor product where a_1, b_1, c_1 are the eigenvectors from dimension x, y, and z respectively and $\mathcal{X}$ is a 3D eigenfunction [4].

- In paper only need to estimate 64x64, 64x64, 33x33 covariance matrices
- Suggest looking at up to top 3 eigenfunctions in each direction

Paper Methodology - Test Statistic

Turn the functional testing problem into one testing for change in covariance structure of d-dimensional estimated score vectors. Use CUSUM-statistic as proposed in [6]

$$S_k = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{k} \text{vech}[\hat{\eta}_t \hat{\eta}_t^T] - \frac{k}{n} \sum_{t=1}^{n} \text{vech}[\hat{\eta}_t \hat{\eta}_t^T] \right) \tag{26}$$

$$\Omega_n = \frac{1}{n} \sum_{k=1}^{n} S_k^T \hat{\Sigma}_n^{-1} S_k \tag{27}$$

where $\hat{\Sigma}_n$ an estimator of long run covariance (long run covariance is $\approx$ variance of the sample mean when time dependence exists). Finally the changepoint location estimated as

$$\hat{\theta} = \frac{\hat{k}^*}{n} \tag{28}$$

$$\hat{k}^* = \underset{1 \leq k \leq n}{\arg\max}\, S_k^T \hat{\Sigma}_n^{-1} S_k. \tag{29}$$

Paper Methodology - Test Statistic

Theorem 1

Under some assumptions on data generating process, and if $\hat{\Sigma}$ *a consistent estimator of long-run covariance of* Σ *then*

$$\Omega_n \xrightarrow{\mathcal{D}} \sum_{l=1}^{d(d+1)/2} \int_0^1 B_l^2(x)dx \tag{30}$$

where $B_l(x)$ *are independent standard Brownian bridges*

$$B_l(x) = W_l(x) - xW_l(1).$$

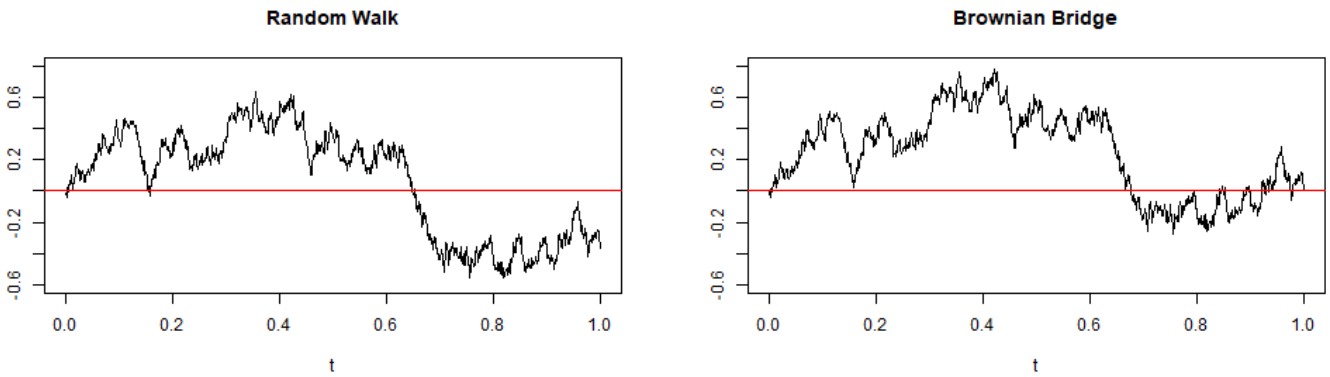

Figure: Realization of a random walk, and its corresponding Brownian Bridge.

Paper Methodology - Test Statistic

Can then use Monte Carlo simulations to estimate $1 - \alpha$ quantile of test statistic by simulating many Brownian Bridges.

Problem: Very difficult to estimate long-run covariance $\hat{\Sigma}$

$$\Sigma = \sum_{t \in \mathbb{Z}} \text{Cov}(\text{vech}[\eta_0 \eta_0^T], \text{vech}[\eta_t \eta_t^T]). \tag{31}$$

Why is this?

- Dimension of projection subspace large, and time series short leads to need to estimate large number of parameters relative to sample size
- Most long-run covariance estimation methods assume no changepoints

Paper Methodology - Long-Run Covariance

Assuming the η_0 scores are Gaussian and have no time dependence

$$\Sigma = \text{Cov}(\text{vech}[\eta_0 \eta_0^T]) = \text{diag}(2\lambda_1^2, \lambda_1\lambda_2, \ldots, 2\lambda_2^2, \lambda_2\lambda_3, \ldots, 2\lambda_d^2). \quad (32)$$

Authors believe changepoint procedure more robust/conservative if only consider diagonals of long-run covariance (i.e. long-run variance), therefore test statistic changed to

$$\tilde{\Omega} = \frac{1}{n} \sum_{k=1}^{n} S_k^T \hat{D}_n^{-1} S_k \quad (33)$$

where D_n constructed using estimated eigenvalues of covariance kernel.

Paper Methodology - Resampling

Using the new test statistic $\tilde{\Omega}$, the limiting distribution no longer pivotal (i.e. cannot be easily written as Brownian Bridges), therefore a resampling procedure is used.

1. Estimate changepoint in each coordinate of $\text{vech}[\hat{\eta}_0 \hat{\eta}_0^T]$ using max statistic.
2. Use this changepoint location to turn $\text{vech}[\hat{\eta}_0 \hat{\eta}_0^T]$ into mean 0 vector.
3. Split time series into overlapping blocks of length K, and use this blocks to estimate variances of score statistics (by taking an average of the blocks' variances).
4. Use these blocks to construct an aproximation of the test statistic

$$\tilde{\Omega}_n^* = \frac{1}{n} \sum_{k=1}^{n} S_k^{*T} D_n^{*-1} S_k^*. \tag{34}$$

5. Repeat this process B times to get $\tilde{\Omega}_n^{*(1)}, \ldots, \tilde{\Omega}_n^{*(B)}$, and take $1 - \alpha$ quantile to get critical value.

Application to fMRI Data

Setup

- 198 scans from 1000 Connectome Resting State Data [2] at same location in Beijing. In previous paper, tested for mean change, here they use 118 scans where no epidemic mean change was detected.
- Each scan 64x64x33 recorded every 2 seconds for 225 time points.
- Polynomial trend removed from each voxel (remove scanner drift).
- Use a test statistic that puts more weight on the change in scores related to eigenfunctions with largest eigenvalues.

Application to fMRI Data - Low Dimensional Projection

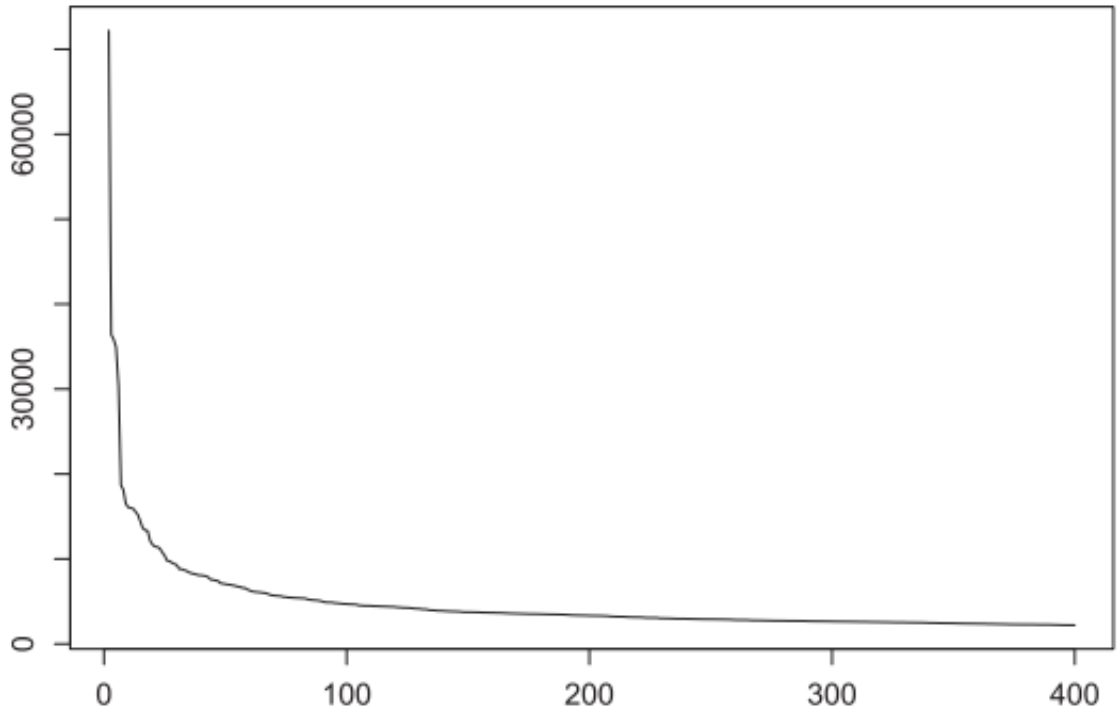

Fig. 3. sub06880: 2nd to 400 largest variance of score products in decreasing order.

Figure: 2nd to 400th largest variance of score products for one subject. Score product of first component 10x larger than second one (removed for better visibility).

Demonstrates most of variability in data driven by relatively few eigenfunctions.

Application to fMRI Data - Covariance Separability

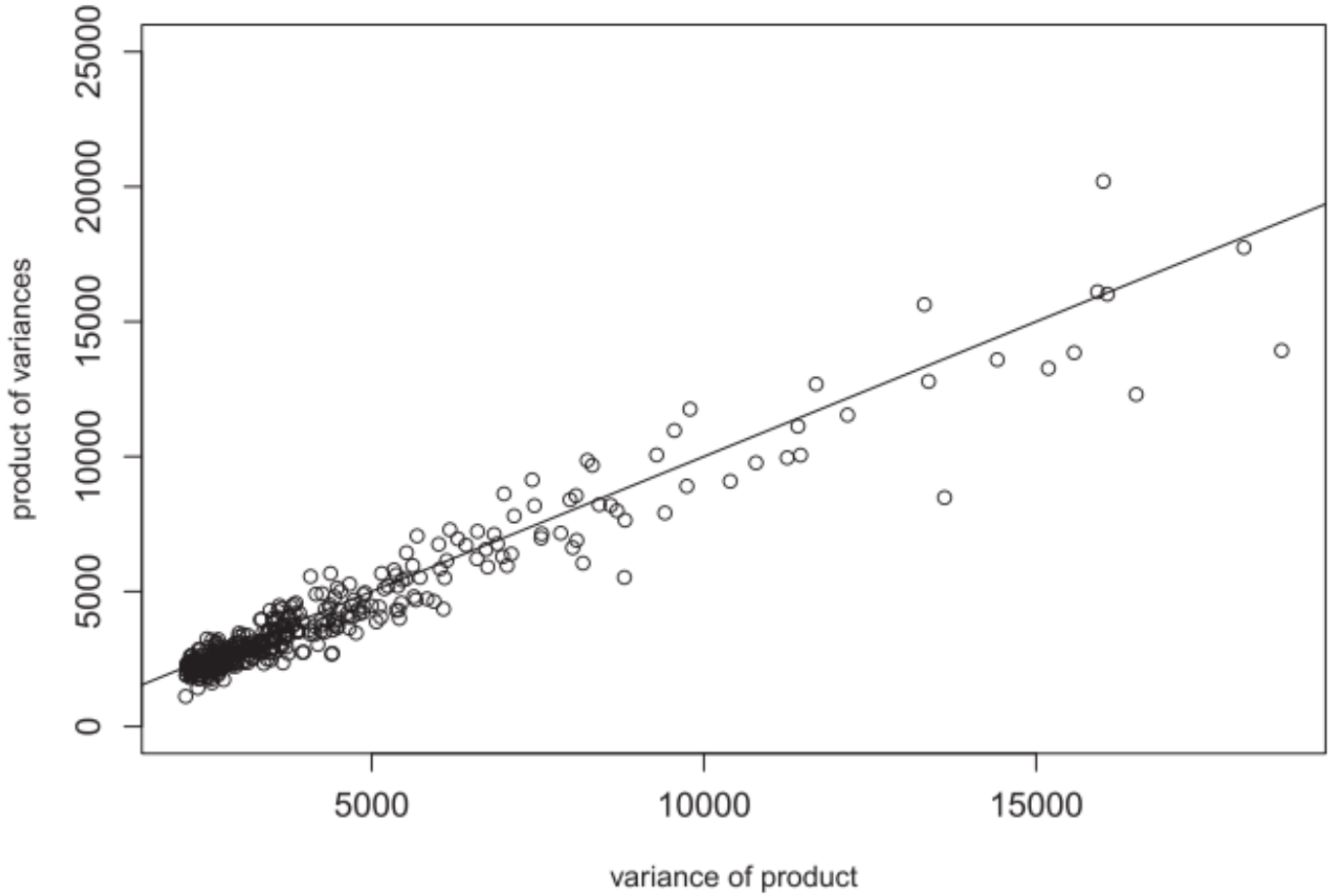

Figure: Product of variances relative to variance of products. These valuse are roughly equal, suggesting that the covariance could infact be treated as separable.

Application to fMRI Data - Changepoints Found

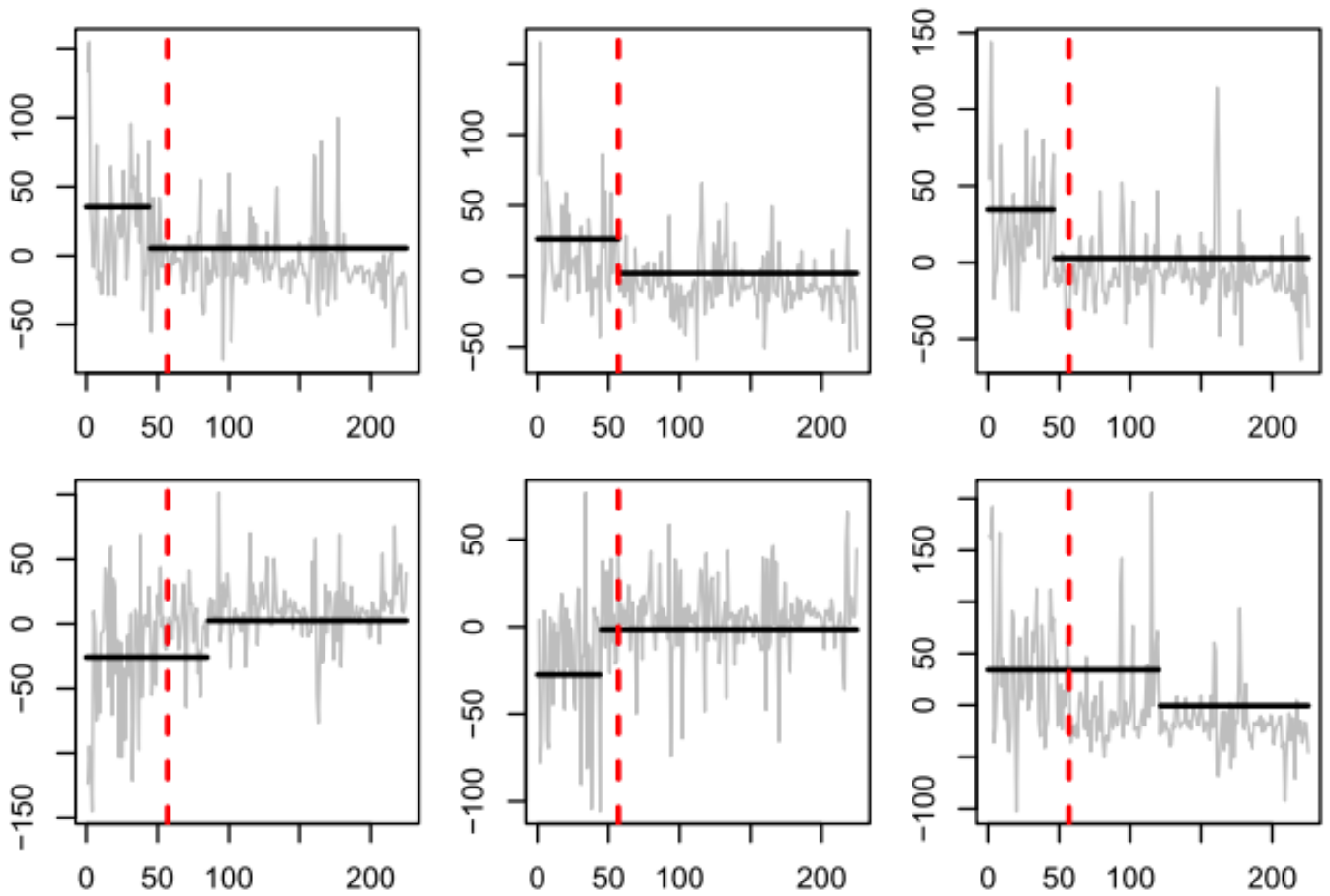

Figure: Score proucts with high evidence of changepoint. Red dotted line is global estimated changepoint. Black line of is mean before and after single dimension changepoint detected.

Thoughts on the Paper

Strengths

- Functional data analysis an intuitive way to approach fMRI data
- Some valid theory for distribution of test statistic
- Dimension reduction seems like a promising approach for fMRI data

Weaknesses

- Much of methods developed in paper not actually used in application section
- Does not seem to constrain lower dimensional basis functions to be smooth, might be subject to noise
- Some of theory falls apart in finite sample high dimensional setting

References I

[1] *Blank Brain Template*. Super Coloring. URL https://www.supercoloring.com/coloring-pages/brain-regions.

[2] International neuroimaging data-sharing initiative. URL https://fcon_1000.projects.nitrc.org/.

[3] *4D Plot*. . URL https://stackoverflow.com/questions/38218431/4d-plot-3dcolor-from-4-row-vectors.

[4] *Tensor multiplication of rank 3*. . URL https://stackoverflow.com/questions/46723652/tensor-multiplication-of-rank-3.

[5] John A. D. Aston and Claudia Kirch. Evaluating stationarity via change-point alternatives with applications to fMRI data. *The Annals of Applied Statistics*, 6(4):1906 – 1948, 2012. doi: 10.1214/12-AOAS565. URL https://doi.org/10.1214/12-AOAS565.

[6] Alexander Aue, Siegfried Hörmann, Lajos Horváth, and Matthew Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087, 2009.

References II

[7] Denis Bosq. *Linear processes in function spaces: theory and applications*, volume 149. Springer Science & Business Media, 2000.

[8] Lajos Horváth and Piotr Kokoszka. *Inference for functional data with applications*, volume 200. Springer Science & Business Media, 2012.

[9] Christina Stoehr, John AD Aston, and Claudia Kirch. Detecting changes in the covariance structure of functional time series with application to fmri data. *Econometrics and Statistics*, 18:44–62, 2021.