Evaluating stationarity via epidemic change-point alternatives for fMRI data

Claudia Kirch

Karlsruhe Institute of Technology (KIT)

joint work with John Aston (University of Warwick)

Recent Advances in Changepoint Analysis, 26-28 March, 2012

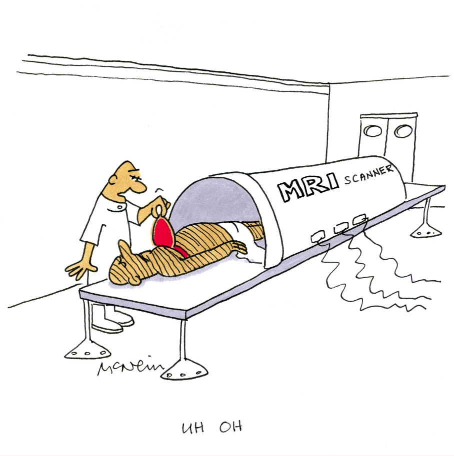
Functional Magnetic Resonance Imaging

Neuroimaging:

Measures blood flow in the brain associated with neural activity

Functional Magnetic Resonance Imaging

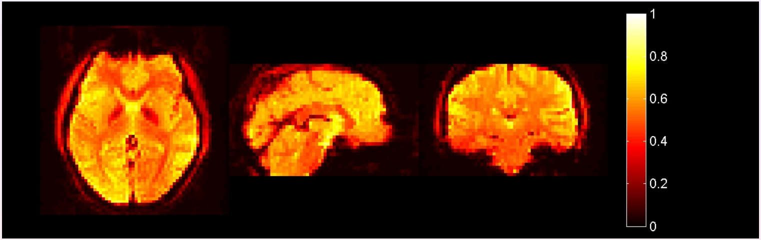
Neuroimaging:
Measures blood flow in the brain associated with neural activity



Functional Magnetic Resonance Imaging

Neuroimaging:

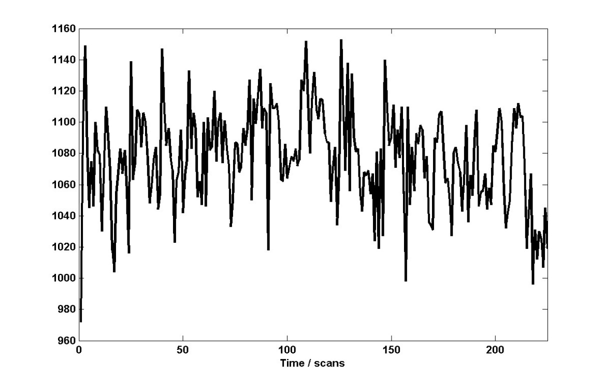
Measures blood flow in the brain associated with neural activity



Functional Magnetic Resonance Imaging

Neuroimaging:

Measures blood flow in the brain associated with neural activity



Functional Magnetic Resonance Imaging

Neuroimaging:
Measures blood flow in the brain associated with neural activity

Experimental design:

- Classical design: Visual or motion based stimulus (instantly linked to brain activity)
- Connection in the brain not due to external stimuli: Resting state data.

Functional Magnetic Resonance Imaging

Neuroimaging:

Measures blood flow in the brain associated with neural activity

Experimental design:

- Classical design: Visual or motion based stimulus (instantly linked to brain activity)
- Connection in the brain not due to external stimuli: Resting state data.
  - Is the resting-state data stationary?

Functional Magnetic Resonance Imaging

Neuroimaging:

Measures blood flow in the brain associated with neural activity

Experimental design:

- Classical design: Visual or motion based stimulus (instantly linked to brain activity)
- Connection in the brain not due to external stimuli: Resting state data.
  - Is the resting-state data stationary?
  - Are there time periods where blood flow increases to a certain level before falling back to baseline?

Functional Magnetic Resonance Imaging

Neuroimaging:

Measures blood flow in the brain associated with neural activity

Experimental design:

- Classical design: Visual or motion based stimulus (instantly linked to brain activity)
- Connection in the brain not due to external stimuli: Resting state data.
  - Is the resting-state data stationary?
  - Are there time periods where blood flow increases to a certain level before falling back to baseline?

Functional Magnetic Resonance Imaging

Statistical Difficulties:

- Functional (very high dimensional) data:
  $10^5 - 10^6$ observations for each time point,
  but only 225 time points!
- Dependent data (across time):

Functional Magnetic Resonance Imaging

Statistical Difficulties:

- Functional (very high dimensional) data:
  $10^5 - 10^6$ observations for each time point,
  but only 225 time points!
- Dependent data (across time):
  - Scanner effects.
  - Dependency in neural activity.

Functional Magnetic Resonance Imaging

Statistical Difficulties:

- Functional (very high dimensional) data:
  $10^5 - 10^6$ observations for each time point,
  but only 225 time points!
- Dependent data (across time):
  - Scanner effects.
  - Dependency in neural activity.

Lindquist et al. (2007): Pointwise testing plus correction:
fMRI-Data with emotional stimulus.

Functional Magnetic Resonance Imaging

Statistical Difficulties:

- Functional (very high dimensional) data:
  $10^5 - 10^6$ observations for each time point,
  but only 225 time points!
- Dependent data (across time):
  - Scanner effects.
  - Dependency in neural activity.

Lindquist et al. (2007): Pointwise testing plus correction:
fMRI-Data with emotional stimulus.

Berkes et al. (2009): Functional change-point analysis:
One-dimensional application such as temperature curves.

Functional Magnetic Resonance Imaging

Statistical Difficulties:

- Functional (very high dimensional) data:
  $10^5 - 10^6$ observations for each time point,
  but only 225 time points!
- Dependent data (across time):
  - Scanner effects.
  - Dependency in neural activity.

Lindquist et al. (2007): Pointwise testing plus correction:
fMRI-Data with emotional stimulus.

Berkes et al. (2009): Functional change-point analysis:
One-dimensional application such as temperature curves.

Functional data analysis

$Y_1(\cdot), \ldots, Y_n(\cdot)$ are random functions in $L^2(\mathcal{Z})$, $\mathcal{Z}$ compact.

$\{Y_i(\cdot)\}$ stationary and ergodic with $\operatorname{E} Y_i(t) = 0$ and

$$\operatorname{E} \|Y_i(\cdot)\|^2 < \infty$$

Functional data analysis

$Y_1(\cdot), \ldots, Y_n(\cdot)$ are random functions in $L^2(\mathcal{Z})$, $\mathcal{Z}$ compact.

$\{Y_i(\cdot)\}$ stationary and ergodic with $\mathrm{E}\, Y_i(t) = 0$ and

$$\mathrm{E}\, \| Y_i(\cdot) \|^2 < \infty$$

Epidemic Change:

$$X_i(t) = Y_i(t) + \mu(t) + \Delta(t) 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}},$$

i: time point, t: function variable, pixel,

$\mu(\cdot), \Delta(\cdot) \in L^2(\mathcal{Z})$ unknown, $0 < \vartheta_1 < \vartheta_2 < 1$ unknown.

Functional data analysis

$Y_1(\cdot), \ldots, Y_n(\cdot)$ are random functions in $L^2(\mathcal{Z})$, $\mathcal{Z}$ compact.

$\{Y_i(\cdot)\}$ stationary and ergodic with $\mathrm{E}\, Y_i(t) = 0$ and

$$\mathrm{E}\, \| Y_i(\cdot) \|^2 < \infty$$

Epidemic Change:

$$X_i(t) = Y_i(t) + \mu(t) + \Delta(t) 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}},$$

i: time point, t: function variable, pixel,

$\mu(\cdot), \Delta(\cdot) \in L^2(\mathcal{Z})$ unknown, $0 < \vartheta_1 < \vartheta_2 < 1$ unknown.

Test:

$H_0 : \Delta \equiv 0$ vs. $H_1 : \Delta \not\equiv 0.$

Functional data analysis

$Y_1(\cdot), \ldots, Y_n(\cdot)$ are random functions in $L^2(\mathcal{Z})$, $\mathcal{Z}$ compact.

$\{Y_i(\cdot)\}$ stationary and ergodic with $\mathrm{E}\, Y_i(t) = 0$ and

$$\mathrm{E}\, \| Y_i(\cdot) \|^2 < \infty$$

Epidemic Change:

$$X_i(t) = Y_i(t) + \mu(t) + \Delta(t) 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}},$$

i: time point, t: function variable, pixel,

$\mu(\cdot), \Delta(\cdot) \in L^2(\mathcal{Z})$ unknown, $0 < \vartheta_1 < \vartheta_2 < 1$ unknown.

Test:

$H_0 : \Delta \equiv 0$ vs. $H_1 : \Delta \not\equiv 0.$

State-of-the-art

At-most-one-change model ($\vartheta_2 = 1$):

- Berkes et al. 2009:
  Development of tests for i.i.d. (across time) data
- Aue et al. 2009:
  Estimators for change-points i.i.d. (across time) data

State-of-the-art

At-most-one-change model ($\vartheta_2 = 1$):

- Berkes et al. 2009:
  Development of tests for i.i.d. (across time) data
- Aue et al. 2009:
  Estimators for change-points i.i.d. (across time) data
- Hörmann and Kokoszka 2010:
  Test for weakly dependent (across time) data

State-of-the-art

At-most-one-change model ($\vartheta_2 = 1$):

- Berkes et al. 2009:
  Development of tests for i.i.d. (across time) data
- Aue et al. 2009:
  Estimators for change-points i.i.d. (across time) data
- Hörmann and Kokoszka 2010:
  Test for weakly dependent (across time) data

Our aim:

- Tests and estimators for epidemic change taking dependence (across time) into account.

State-of-the-art

At-most-one-change model ($\vartheta_2 = 1$):

- Berkes et al. 2009:
  Development of tests for i.i.d. (across time) data
- Aue et al. 2009:
  Estimators for change-points i.i.d. (across time) data
- Hörmann and Kokoszka 2010:
  Test for weakly dependent (across time) data

Our aim:

- Tests and estimators for epidemic change taking dependence (across time) into account.
- Estimation in hierarchical models.

State-of-the-art

At-most-one-change model ($\vartheta_2 = 1$):

- Berkes et al. 2009:
  Development of tests for i.i.d. (across time) data
- Aue et al. 2009:
  Estimators for change-points i.i.d. (across time) data
- Hörmann and Kokoszka 2010:
  Test for weakly dependent (across time) data

Our aim:

- Tests and estimators for epidemic change taking dependence (across time) into account.
- Estimation in hierarchical models.
- Application to fMRI resting state data.

State-of-the-art

At-most-one-change model ($\vartheta_2 = 1$):

- Berkes et al. 2009:
  Development of tests for i.i.d. (across time) data
- Aue et al. 2009:
  Estimators for change-points i.i.d. (across time) data
- Hörmann and Kokoszka 2010:
  Test for weakly dependent (across time) data

Our aim:

- Tests and estimators for epidemic change taking dependence (across time) into account.
- Estimation in hierarchical models.
- Application to fMRI resting state data.

Testing Problem

Main idea:
Dimension reduction via orthonormal system $\{v_k(\cdot), k = 1, \ldots, d\}$:

$$\langle X_i, v_l \rangle = \langle Y_i, v_l \rangle + \langle \mu, v_l \rangle + 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}} \langle \Delta, v_l \rangle,$$

where $\langle X_i, v_l \rangle = \int X_i(t) v_l(t)\, dt.$

Hence: Same type of level shifts present in projection if Δ not orthogonal to projection subspace.

Testing Problem

Main idea:
Dimension reduction via orthonormal system $\{v_k(\cdot), k = 1, \ldots, d\}$:

$$\langle X_i, v_l \rangle = \langle Y_i, v_l \rangle + \langle \mu, v_l \rangle + 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}} \langle \Delta, v_l \rangle,$$

$$\text{where } \langle X_i, v_l \rangle = \int X_i(t) v_l(t)\, dt.$$

Hence: Same type of level shifts present in projection if Δ not orthogonal to projection subspace.

We will use: Principal component analysis:
Use estimator $\{\widehat{v}_l : l = 1, \ldots, d\}$.

Testing Problem

Main idea:
Dimension reduction via orthonormal system $\{v_k(\cdot), k = 1, \ldots, d\}$:

$$\langle X_i, v_l \rangle = \langle Y_i, v_l \rangle + \langle \mu, v_l \rangle + 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}} \langle \Delta, v_l \rangle,$$

$$\text{where } \langle X_i, v_l \rangle = \int X_i(t) v_l(t)\, dt.$$

Hence: Same type of level shifts present in projection if Δ not orthogonal to projection subspace.

We will use: Principal component analysis:
Use estimator $\{\widehat{v}_l : l = 1, \ldots, d\}$.

Use multi-dimensional testing procedures on projected data!

Testing Problem

Main idea:
Dimension reduction via orthonormal system $\{v_k(\cdot), k = 1, \ldots, d\}$:

$$\langle X_i, v_l \rangle = \langle Y_i, v_l \rangle + \langle \mu, v_l \rangle + 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}} \langle \Delta, v_l \rangle,$$

$$\text{where } \langle X_i, v_l \rangle = \int X_i(t) v_l(t)\, dt.$$

Hence: Same type of level shifts present in projection if Δ not orthogonal to projection subspace.

We will use: Principal component analysis:
Use estimator $\{\widehat{v}_l : l = 1, \ldots, d\}$.

Use multi-dimensional testing procedures on projected data!

Resting State Data (Connectome Project)

197 persons from Beijing, China:

Each scan:

225 time points of a 3-dim. image of $64 \times 64 \times 33 \approx 10^5$ voxels.

Preliminary data processing:

- Removal of polynomial trend due to scanner drift.
- Motion correction.

Resting State Data (Connectome Project)

197 persons from Beijing, China:

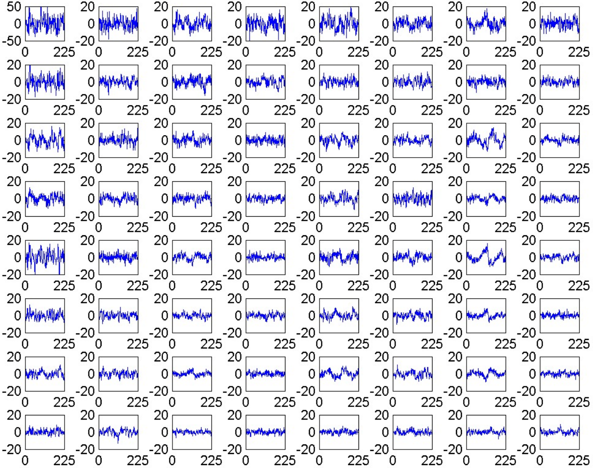
Each scan:

225 time points of a 3-dim. image of $64 \times 64 \times 33 \approx 10^5$ voxels.

Preliminary data processing:

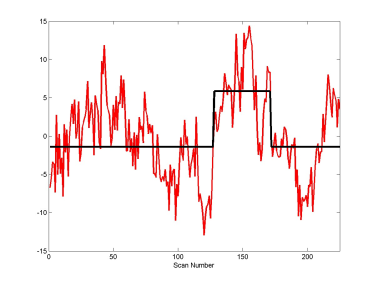
- Removal of polynomial trend due to scanner drift.
- Motion correction.

Resting State Example 1: Strong level shift

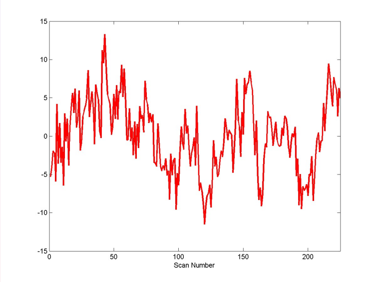


Resting State Example 1: Strong level shift

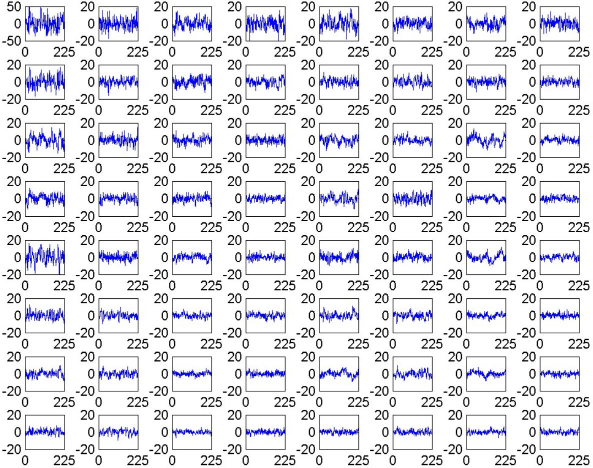
With level shifts:



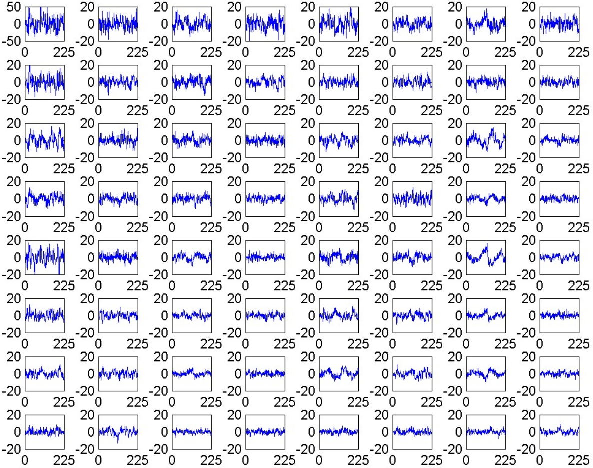
Epidemic change removed:



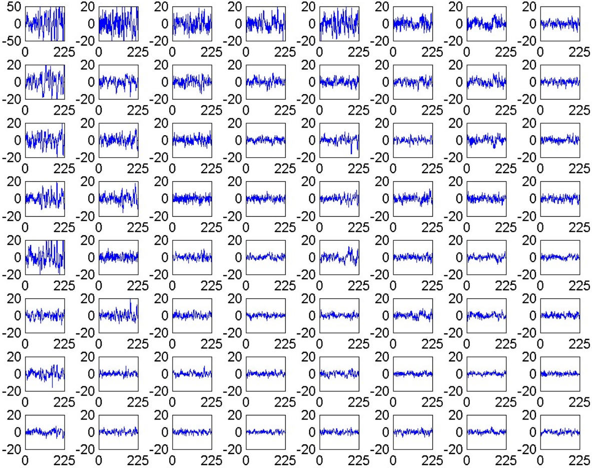
Resting State Example 1: Strong level shift



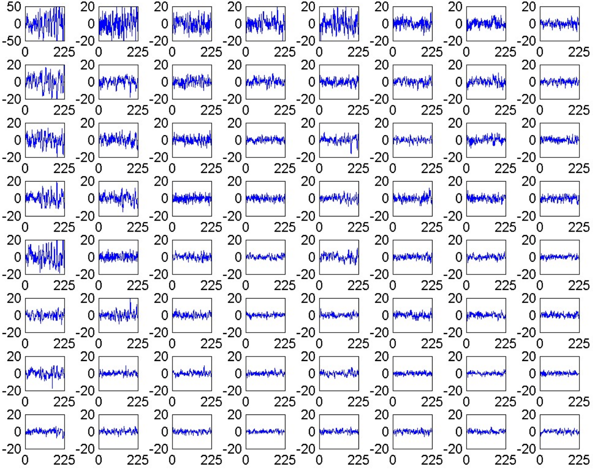
Resting State Example 1: Strong level shift



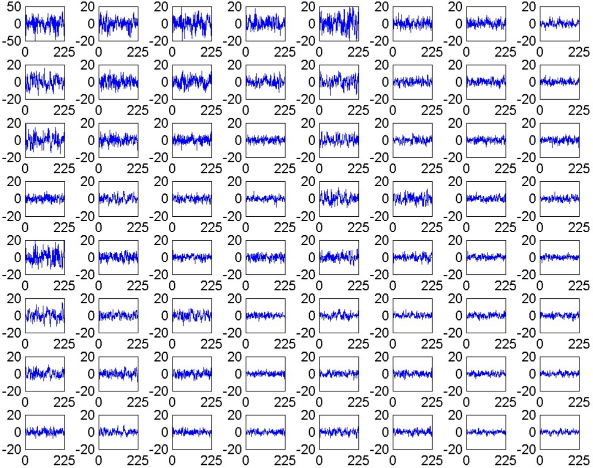
Resting State Example 2: Medium level shift



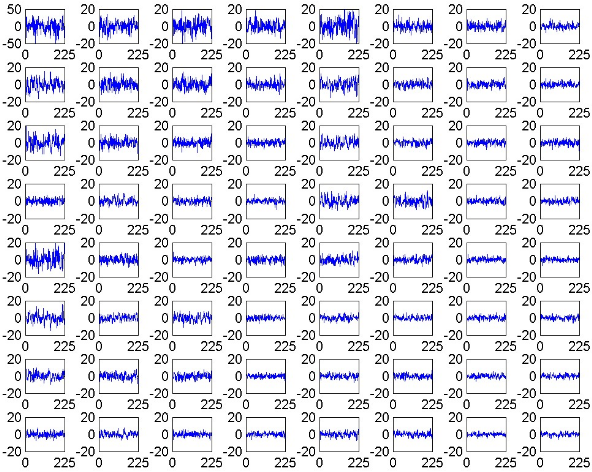
Resting State Example 2: Medium level shift



Resting State Example 3: No level shift?



Resting State Example 3: No level shift?



Assumptions under H_0

$\sqrt{n}$-Consistency of estimators: $\int(\widehat{v}_k(t) - s_k v_k(t))^2\,dt = O_P(n^{-1})$, where $s_k = \operatorname{sgn}\left(\int v_{k,j}(t)\widehat{v}_k(t)\,dt\right)$.

Principal Components:

- $d+1$ distinct eigenvalues.
- $\sqrt{n}$-Consistency of sample covariance function.

Assumptions under H_0

$\sqrt{n}$-Consistency of estimators: $\int (\widehat{v}_k(t) - s_k v_k(t))^2 \, dt = O_P(n^{-1})$, where $s_k = \operatorname{sgn}\left(\int v_{k,j}(t)\widehat{v}_k(t)\, dt\right)$.

Principal Components:

- $d+1$ distinct eigenvalues.
- $\sqrt{n}$-Consistency of sample covariance function.

Let $\eta_{i,l} = \langle Y_i, v_l \rangle$. Assume

Assumptions under H_0

$\sqrt{n}$-Consistency of estimators: $\int(\widehat{v}_k(t) - s_k v_k(t))^2\,dt = O_P(n^{-1})$,
where $s_k = \operatorname{sgn}\left(\int v_{k,j}(t)\widehat{v}_k(t)\,dt\right)$.

Principal Components:

- $d+1$ distinct eigenvalues.
- $\sqrt{n}$-Consistency of sample covariance function.

Let $\eta_{i,l} = \langle Y_i, v_l\rangle$. Assume

- $\{\boldsymbol{\eta}_i = (\eta_{i,1},\ldots,\eta_{i,d}) : i \geqslant 1\}$ (across time!) are stationary and short-range dependent

Assumptions under H_0

$\sqrt{n}$-Consistency of estimators: $\int(\widehat{v}_k(t) - s_k v_k(t))^2\,dt = O_P(n^{-1})$, where $s_k = \operatorname{sgn}\left(\int v_{k,j}(t)\widehat{v}_k(t)\,dt\right)$.

Principal Components:

- $d+1$ distinct eigenvalues.
- $\sqrt{n}$-Consistency of sample covariance function.

Let $\eta_{i,l} = \langle Y_i, v_l \rangle$. Assume

- $\{\boldsymbol{\eta}_i = (\eta_{i,1}, \ldots, \eta_{i,d}) : i \geqslant 1\}$ (across time!) are stationary and short-range dependent
- and fulfill FCLT, asymptotic long-run covariance: $\Sigma = \lim_{n\to\infty} \frac{1}{n} \operatorname{E}\left(\sum_{i=1}^n \boldsymbol{\eta}_i\right)\left(\sum_{i=1}^n \boldsymbol{\eta}_i\right)^T$.

Assumptions under H_0

$\sqrt{n}$-Consistency of estimators: $\int(\widehat{v}_k(t) - s_k v_k(t))^2\,dt = O_P(n^{-1})$, where $s_k = \operatorname{sgn}\left(\int v_{k,j}(t)\widehat{v}_k(t)\,dt\right)$.

Principal Components:

- $d+1$ distinct eigenvalues.
- $\sqrt{n}$-Consistency of sample covariance function.

Let $\eta_{i,l} = \langle Y_i, v_l \rangle$. Assume

- $\{\boldsymbol{\eta}_i = (\eta_{i,1}, \ldots, \eta_{i,d}) : i \geqslant 1\}$ (across time!) are stationary and short-range dependent
- and fulfill FCLT, asymptotic long-run covariance: $\Sigma = \lim_{n\to\infty} \frac{1}{n} \operatorname{E}\left(\sum_{i=1}^n \boldsymbol{\eta}_i\right)\left(\sum_{i=1}^n \boldsymbol{\eta}_i\right)^T$.

Examples:

- $Y_i(\cdot)$ i.i.d. (Berkes et al. 2009)
- $Y_i(\cdot)$ $L^4 - m$-approximable (Hörmann and Kokoszka 2010)
- $Y_i(\cdot)$ α-mixing (Aston, K. 2012)

Assumptions under H_0

$\sqrt{n}$-Consistency of estimators: $\int(\widehat{v}_k(t) - s_k v_k(t))^2\,dt = O_P(n^{-1})$, where $s_k = \operatorname{sgn}\left(\int v_{k,j}(t)\widehat{v}_k(t)\,dt\right)$.

Principal Components:

- $d+1$ distinct eigenvalues.
- $\sqrt{n}$-Consistency of sample covariance function.

Let $\eta_{i,l} = \langle Y_i, v_l \rangle$. Assume

- $\{\boldsymbol{\eta}_i = (\eta_{i,1}, \ldots, \eta_{i,d}) : i \geqslant 1\}$ (across time!) are stationary and short-range dependent
- and fulfill FCLT, asymptotic long-run covariance: $\Sigma = \lim_{n\to\infty} \frac{1}{n} \operatorname{E}\left(\sum_{i=1}^n \boldsymbol{\eta}_i\right)\left(\sum_{i=1}^n \boldsymbol{\eta}_i\right)^T$.

Examples:

- $Y_i(\cdot)$ i.i.d. (Berkes et al. 2009)
- $Y_i(\cdot)$ $L^4 - m$-approximable (Hörmann and Kokoszka 2010)
- $Y_i(\cdot)$ α-mixing (Aston, K. 2012)

Null Asymptotics

Change-point tests typically based on:

$\widehat{\mathbf{S}}_{k_1,k_2} = \left(\sum_{k_1<i\leqslant k_2} \left(\widehat{\eta}_{i,l} - \bar{\widehat{\eta}}_l\right)\right)_{l=1,\ldots,d'}$ $\bar{\widehat{\eta}}_l = \frac{1}{n}\sum_{i=1}^n \widehat{\eta}_{i,l}$, where $\widehat{\eta}_{i,l} = \langle X_i, \widehat{v}_l \rangle$.

Theorem (Aston, K. (2012))

Test statistic for epidemic change:

$$T_n := \frac{1}{n^3} \sum_{1\leqslant k_1<k_2\leqslant n} \widehat{\mathbf{S}}_{k_1,k_2}^T \widehat{\Sigma}^{-1} \widehat{\mathbf{S}}_{k_1,k_2}$$
$$\xrightarrow{\mathcal{D}} \sum_{1\leqslant l\leqslant d} \int\int_{0\leqslant x<y\leqslant 1} (B_l(y) - B_l(x))^2 \, dx\, dy \qquad \text{under } H_0,$$

where $\widehat{\Sigma}$ estimator for the long-run covariance matrix Σ and $B_l(\cdot)$ are independent Brownian bridges.

Null Asymptotics

Change-point tests typically based on:

$\widehat{\mathbf{S}}_{k_1,k_2} = \left(\sum_{k_1 < i \leqslant k_2} \left(\widehat{\eta}_{i,l} - \bar{\widehat{\eta}}_l\right)\right)_{l=1,\ldots,d}$, $\bar{\widehat{\eta}}_l = \frac{1}{n}\sum_{i=1}^n \widehat{\eta}_{i,l}$, where $\widehat{\eta}_{i,l} = \langle X_i, \widehat{v}_l \rangle$.

Theorem (Aston, K. (2012))

Test statistic for epidemic change:

$$T_n := \frac{1}{n^3} \sum_{1 \leqslant k_1 < k_2 \leqslant n} \widehat{\mathbf{S}}_{k_1,k_2}^T \widehat{\Sigma}^{-1} \widehat{\mathbf{S}}_{k_1,k_2}$$

$$\xrightarrow{\mathcal{D}} \sum_{1 \leqslant l \leqslant d} \int\int_{0 \leqslant x < y \leqslant 1} (B_l(y) - B_l(x))^2 \, dx \, dy \qquad \text{under } H_0,$$

where $\widehat{\Sigma}$ estimator for the long-run covariance matrix Σ and $B_l(\cdot)$ are independent Brownian bridges.

Resting state data

Problem: Estimation of the inverse of long-run covariance matrix!

$d \times d$ matrix ($d = 64, 125$) but only 225 time points:

Standard estimators are no longer statistically stable: About 1/3 negative eigenvalues of estimate for positive-definite matrix.

$\Longrightarrow$ Problematic since inverse is needed.

Resting state data

Problem: Estimation of the inverse of long-run covariance matrix!

$d \times d$ matrix ($d = 64, 125$) but only 225 time points:

Standard estimators are no longer statistically stable: About 1/3 negative eigenvalues of estimate for positive-definite matrix.

$\Longrightarrow$ Problematic since inverse is needed.

Even if inverse of small eigenvalues is set to zero (conservative):
Every resting state data set rejects the null hypothesis!

Resting state data

Problem: Estimation of the inverse of long-run covariance matrix!

$d \times d$ matrix ($d = 64, 125$) but only 225 time points:

Standard estimators are no longer statistically stable: About 1/3 negative eigenvalues of estimate for positive-definite matrix.

$\Longrightarrow$ Problematic since inverse is needed.

Even if inverse of small eigenvalues is set to zero (conservative): Every resting state data set rejects the null hypothesis!

Therefore: Use only diagonal matrix of long-run variances!

Resting state data

Problem: Estimation of the inverse of long-run covariance matrix!

$d \times d$ matrix ($d = 64, 125$) but only 225 time points:

Standard estimators are no longer statistically stable: About 1/3 negative eigenvalues of estimate for positive-definite matrix.

$\Longrightarrow$ Problematic since inverse is needed.

Even if inverse of small eigenvalues is set to zero (conservative): Every resting state data set rejects the null hypothesis!

Therefore: Use only diagonal matrix of long-run variances!

- Theory still applies, but limit no longer pivotal: Dependency between Brownian bridges.

Resting state data

Problem: Estimation of the inverse of long-run covariance matrix!

$d \times d$ matrix ($d = 64, 125$) but only 225 time points:

Standard estimators are no longer statistically stable: About 1/3 negative eigenvalues of estimate for positive-definite matrix.

$\Longrightarrow$ Problematic since inverse is needed.

Even if inverse of small eigenvalues is set to zero (conservative): Every resting state data set rejects the null hypothesis!

Therefore: Use only diagonal matrix of long-run variances!

- Theory still applies, but limit no longer pivotal: Dependency between Brownian bridges.
- Therefore: Bootstrap methods necessary.

Resting state data

Problem: Estimation of the inverse of long-run covariance matrix!

$d \times d$ matrix ($d = 64, 125$) but only 225 time points:

Standard estimators are no longer statistically stable: About 1/3 negative eigenvalues of estimate for positive-definite matrix.

$\Longrightarrow$ Problematic since inverse is needed.

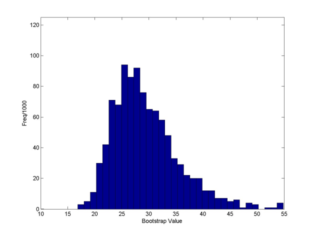
Even if inverse of small eigenvalues is set to zero (conservative): Every resting state data set rejects the null hypothesis!

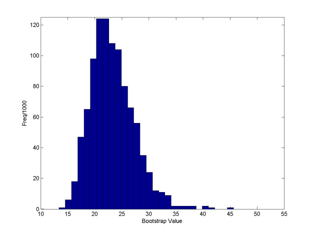
Therefore: Use only diagonal matrix of long-run variances!

- Theory still applies, but limit no longer pivotal: Dependency between Brownian bridges.
- Therefore: Bootstrap methods necessary.

Bootstrap distribution

Bootstrap distribution
with no change detected

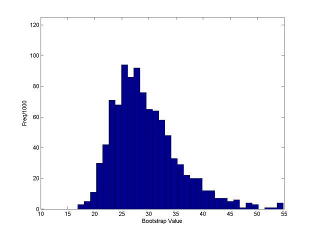


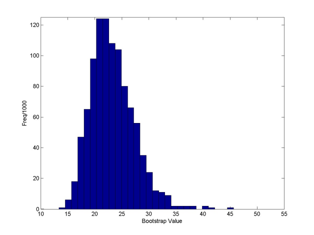


Bootstrap distributions
with change detected

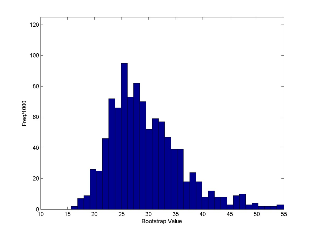
Bootstrap distribution

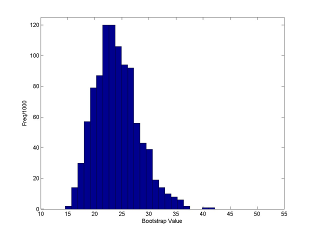
Bootstrap distribution
with no change detected



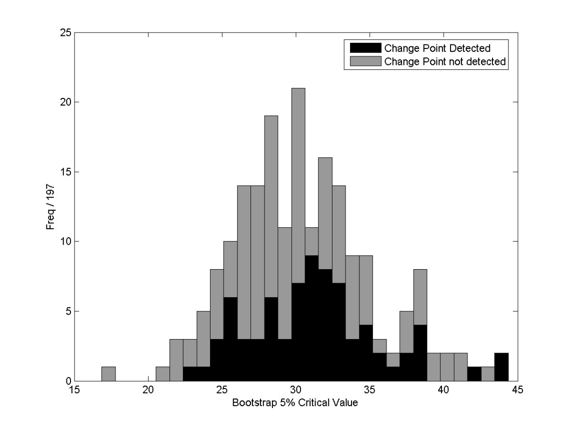


Bootstrap distributions
with change detected





Bootstrap distribution



Distribution of bootstrap 5% critical values from 197 scans

How to do PCA for such a data set?

Very high-dimensional data set (10^5 voxel),
very short time series (225 time points):

$\Longrightarrow$ inaccurate estimate of principal components (e.g. $\leqslant 225$ non-zero eigenvalues).

How to do PCA for such a data set?

Very high-dimensional data set (10^5 voxel),
very short time series (225 time points):

$\Longrightarrow$ inaccurate estimate of principal components (e.g. $\leqslant 225$ non-zero eigenvalues).

Semiparametric approach: Make use of 3-dimensional structure $\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{T}_3$.

How to do PCA for such a data set?

Very high-dimensional data set (10^5 voxel),
very short time series (225 time points):

$\Longrightarrow$ inaccurate estimate of principal components (e.g. $\leqslant 225$ non-zero eigenvalues).

Semiparametric approach: Make use of 3-dimensional structure $\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{T}_3$.

Parametric assumption: Separable covariance structure:

$c((t_1, t_2, t_3), (s_1, s_2, s_3)) = c_1(t_1, s_1)\, c_2(t_2, s_2)\, c_3(t_3, s_3).$

How to do PCA for such a data set?

Very high-dimensional data set (10^5 voxel),
very short time series (225 time points):
$\Longrightarrow$ inaccurate estimate of principal components (e.g. $\leqslant 225$ non-zero eigenvalues).

Semiparametric approach: Make use of 3-dimensional structure $\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{T}_3$.

Parametric assumption: Separable covariance structure:

$$c((t_1, t_2, t_3), (s_1, s_2, s_3)) = c_1(t_1, s_1)\, c_2(t_2, s_2)\, c_3(t_3, s_3).$$

Eigenfunctions are product of eigenfunctions of c_1, c_2 and c_3
$\Longrightarrow$ dimension reduction in solving eigenvalue problem.

How to do PCA for such a data set?

Very high-dimensional data set (10^5 voxel),
very short time series (225 time points):
$\Longrightarrow$ inaccurate estimate of principal components (e.g. $\leqslant 225$ non-zero eigenvalues).

Semiparametric approach: Make use of 3-dimensional structure $\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{T}_3$.

Parametric assumption: Separable covariance structure:

$$c((t_1, t_2, t_3), (s_1, s_2, s_3)) = c_1(t_1, s_1)\, c_2(t_2, s_2)\, c_3(t_3, s_3).$$

Eigenfunctions are product of eigenfunctions of c_1, c_2 and c_3
$\Longrightarrow$ dimension reduction in solving eigenvalue problem.

Misspecification not that problematic since still valid dimension reduction with favorable properties under alternatives.

How to do PCA for such a data set?

Very high-dimensional data set (10^5 voxel),
very short time series (225 time points):
$\Longrightarrow$ inaccurate estimate of principal components (e.g. $\leqslant 225$ non-zero eigenvalues).

Semiparametric approach: Make use of 3-dimensional structure $\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{T}_3$.

Parametric assumption: Separable covariance structure:

$$c((t_1, t_2, t_3), (s_1, s_2, s_3)) = c_1(t_1, s_1)\, c_2(t_2, s_2)\, c_3(t_3, s_3).$$

Eigenfunctions are product of eigenfunctions of c_1, c_2 and c_3
$\Longrightarrow$ dimension reduction in solving eigenvalue problem.

Misspecification not that problematic since still valid dimension reduction with favorable properties under alternatives.

Why is PCA reasonable in this context?

Resting state scan: Typically,

- relatively flat eigenvalues,
- 64 components explain less than 1% of variation.

Why is PCA reasonable in this context?

Resting state scan: Typically,

- relatively flat eigenvalues,
- 64 components explain less than 1% of variation.

Useless for standard PCA applications!

Why is PCA reasonable in this context?

Resting state scan: Typically,

- relatively flat eigenvalues,
- 64 components explain less than 1% of variation.

Useless for standard PCA applications!

Change-point analysis:

Under null hypothesis: We still obtain valid size α test!

Why is PCA reasonable in this context?

Resting state scan: Typically,

- relatively flat eigenvalues,
- 64 components explain less than 1% of variation.

Useless for standard PCA applications!

Change-point analysis:

Under null hypothesis: We still obtain valid size α test!

Under alternatives: $\widehat{v}_l$ will generally converge to contaminated subspace w_l.

Why is PCA reasonable in this context?

Resting state scan: Typically,

- relatively flat eigenvalues,
- 64 components explain less than 1% of variation.

Useless for standard PCA applications!

Change-point analysis:

Under null hypothesis: We still obtain valid size α test!

Under alternatives: $\widehat{v}_l$ will generally converge to contaminated subspace w_l.

Theorem (Aston, K. (2012))

If $\int (\widehat{v}_k(t) - s_k w_k(t))^2 \, dt = o_P(1)$, $k = 1, \ldots, d$, and $\langle \Delta, w_l \rangle \neq 0$, for some $l = 1, \ldots, d$, then $T_n \xrightarrow{\mathcal{P}} \infty$, i.e. the test has asymptotic power one.

Recall: $\langle X_i, \widehat{v}_l \rangle = \langle \mu_1, \widehat{v}_l \rangle + \langle Y_i, \widehat{v}_l \rangle + 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}} \langle \Delta, \widehat{v}_l \rangle.$

Why is PCA reasonable in this context?

Resting state scan: Typically,

- relatively flat eigenvalues,
- 64 components explain less than 1% of variation.

Useless for standard PCA applications!

Change-point analysis:

Under null hypothesis: We still obtain valid size α test!

Under alternatives: $\widehat{v}_l$ will generally converge to contaminated subspace w_l.

Theorem (Aston, K. (2012))

If $\int (\widehat{v}_k(t) - s_k w_k(t))^2 \, dt = o_P(1)$, $k = 1, \ldots, d$, and $\langle \Delta, w_l \rangle \neq 0$, for some $l = 1, \ldots, d$, then $T_n \xrightarrow{\mathcal{P}} \infty$, i.e. the test has asymptotic power one.

Recall: $\langle X_i, \widehat{v}_l \rangle = \langle \mu_1, \widehat{v}_l \rangle + \langle Y_i, \widehat{v}_l \rangle + 1_{\{\vartheta_1 n < i \leqslant \vartheta_2 n\}} \langle \Delta, \widehat{v}_l \rangle.$

Standard Principal Component Analysis

Contaminated subspace:

$\{w_k\}$ eigenfunctions of $c(t,s) + \theta(1-\theta)\Delta(t)\Delta(s)$, $\theta = \vartheta_2 - \vartheta_1$

($d+1$ distinct eigenvalues needed).

Theorem (Aston, K. (2012))

a) Any change not orthogonal to the non-contaminated eigenspace is detectable also by contaminated eigenspace.

Standard Principal Component Analysis

Contaminated subspace:

$\{w_k\}$ eigenfunctions of $c(t,s) + \theta(1-\theta)\Delta(t)\Delta(s)$, $\theta = \vartheta_2 - \vartheta_1$

($d+1$ distinct eigenvalues needed).

Theorem (Aston, K. (2012))

a) Any change not orthogonal to the non-contaminated eigenspace is detectable also by contaminated eigenspace.

Principle of PCA: Choice of subspace such that most of the sample variance is explained by few components.

Standard Principal Component Analysis

Contaminated subspace:

$\{w_k\}$ eigenfunctions of $c(t,s)+\theta(1-\theta)\Delta(t)\Delta(s)$, $\theta=\vartheta_2-\vartheta_1$

($d+1$ distinct eigenvalues needed).

Theorem (Aston, K. (2012))

a) Any change not orthogonal to the non-contaminated eigenspace is detectable also by contaminated eigenspace.

Principle of PCA: Choice of subspace such that most of the sample variance is explained by few components.

Under alternatives the variability of the sample increases in the direction of the change!

Standard Principal Component Analysis

Contaminated subspace:

$\{w_k\}$ eigenfunctions of $c(t,s) + \theta(1-\theta)\Delta(t)\Delta(s)$, $\theta = \vartheta_2 - \vartheta_1$

($d+1$ distinct eigenvalues needed).

Theorem (Aston, K. (2012))

a) Any change not orthogonal to the non-contaminated eigenspace is detectable also by contaminated eigenspace.

Principle of PCA: Choice of subspace such that most of the sample variance is explained by few components.

Under alternatives the variability of the sample increases in the direction of the change!

Change will likely be 'included' in few principal components.

Standard Principal Component Analysis

Contaminated subspace:

$\{w_k\}$ eigenfunctions of $c(t,s) + \theta(1-\theta)\Delta(t)\Delta(s)$, $\theta = \vartheta_2 - \vartheta_1$

$(d+1$ distinct eigenvalues needed$)$.

Theorem (Aston, K. (2012))

a) Any change not orthogonal to the non-contaminated eigenspace is detectable also by contaminated eigenspace.

b) Any large enough change will be detectable using only the first principal component!

Principle of PCA: Choice of subspace such that most of the sample variance is explained by few components.

Under alternatives the variability of the sample increases in the direction of the change!

Change will likely be 'included' in few principal components.

Standard Principal Component Analysis

Contaminated subspace:

$\{w_k\}$ eigenfunctions of $c(t,s) + \theta(1-\theta)\Delta(t)\Delta(s)$, $\theta = \vartheta_2 - \vartheta_1$

($d + 1$ distinct eigenvalues needed).

Theorem (Aston, K. (2012))

a) Any change not orthogonal to the non-contaminated eigenspace is detectable also by contaminated eigenspace.
b) Any large enough change will be detectable using only the first principal component!

Principle of PCA: Choice of subspace such that most of the sample variance is explained by few components.

Under alternatives the variability of the sample increases in the direction of the change!

Change will likely be 'included' in few principal components.

Standard Principal Component Analysis

Contaminated subspace:

$\{w_k\}$ eigenfunctions of $c(t,s)+\theta(1-\theta)\Delta(t)\Delta(s)$, $\theta=\vartheta_2-\vartheta_1$

($d+1$ distinct eigenvalues needed).

Theorem (Aston, K. (2012))

a) Any change not orthogonal to the non-contaminated eigenspace is detectable also by contaminated eigenspace.

b) Any large enough change will be detectable using only the first principal component!

Remark:

- The second feature works best for flat eigenvalues.
- Separable PCA: If $\Delta(t_1,t_2,t_3)=\Delta_1(t_1)\Delta_2(t_2)\Delta_3(t_3)$, then analogous assertion hold (Aston, K. (2011)).

Standard Principal Component Analysis

Contaminated subspace:

$\{w_k\}$ eigenfunctions of $c(t,s)+\theta(1-\theta)\Delta(t)\Delta(s)$, $\theta=\vartheta_2-\vartheta_1$

($d+1$ distinct eigenvalues needed).

Theorem (Aston, K. (2012))

a) Any change not orthogonal to the non-contaminated eigenspace is detectable also by contaminated eigenspace.

b) Any large enough change will be detectable using only the first principal component!

Remark:

- The second feature works best for flat eigenvalues.
- Separable PCA: If $\Delta(t_1,t_2,t_3)=\Delta_1(t_1)\Delta_2(t_2)\Delta_3(t_3)$, then analogous assertion hold (Aston, K. (2011)).

Resting State Data

Test results:

- $64\,(=4^3)$ components: 78 (of 197) reject at 5% level
  70 after FDR-Correction, FDR threshold: 0.022
- $125\,(=5^3)$ components: 82 (of 197) reject at 5% level
  76 after FDR-Correction, FDR threshold: 0.022

Resting State Data

Test results:

- $64\,(=4^3)$ components: 78 (of 197) reject at 5% level
  70 after FDR-Correction, FDR threshold: 0.022
- $125\,(=5^3)$ components: 82 (of 197) reject at 5% level
  76 after FDR-Correction, FDR threshold: 0.022

Estimators of location of change-points:

$$(\widehat{\vartheta}_1, \widehat{\vartheta}_2) = \arg\max_{x<y} \widehat{\mathbf{S}}^T_{\lfloor nx \rfloor, \lfloor ny \rfloor} \widehat{\Sigma}^{-1} \widehat{\mathbf{S}}_{\lfloor nx \rfloor, \lfloor ny \rfloor}.$$

Resting State Data

Test results:

- $64\,(=4^3)$ components: 78 (of 197) reject at 5% level
  70 after FDR-Correction, FDR threshold: 0.022
- $125\,(=5^3)$ components: 82 (of 197) reject at 5% level
  76 after FDR-Correction, FDR threshold: 0.022

Estimators of location of change-points:

$$(\widehat{\vartheta}_1, \widehat{\vartheta}_2) = \arg\max_{x<y} \widehat{\mathbf{S}}^T_{\lfloor nx \rfloor, \lfloor ny \rfloor} \widehat{\Sigma}^{-1} \widehat{\mathbf{S}}_{\lfloor nx \rfloor, \lfloor ny \rfloor}.$$

Theorem (Aston, K. (2012))

a) If the change is detectable, then $(\widehat{\vartheta}_1 - \vartheta_1, \widehat{\vartheta}_2 - \vartheta_2)^T = o_P(1)$.

Resting State Data

Test results:

- $64\,(=4^3)$ components: 78 (of 197) reject at 5% level
  70 after FDR-Correction, FDR threshold: 0.022
- $125\,(=5^3)$ components: 82 (of 197) reject at 5% level
  76 after FDR-Correction, FDR threshold: 0.022

Estimators of location of change-points:

$$(\widehat{\vartheta}_1, \widehat{\vartheta}_2) = \arg\max_{x<y} \widehat{\mathbf{S}}^T_{\lfloor nx \rfloor, \lfloor ny \rfloor} \widehat{\Sigma}^{-1} \widehat{\mathbf{S}}_{\lfloor nx \rfloor, \lfloor ny \rfloor}.$$

Theorem (Aston, K. (2012))

a) If the change is detectable, then $(\widehat{\vartheta}_1 - \vartheta_1, \widehat{\vartheta}_2 - \vartheta_2)^T = o_P(1)$.

b) If additionally a Hájek-Rényi-type condition holds for $\langle Y_i, w_l \rangle$,
then $n(\widehat{\vartheta}_1 - \vartheta_1, \widehat{\vartheta}_2 - \vartheta_2)^T = O_P(1)$.

Resting State Data

Test results:

- $64\,(=4^3)$ components: 78 (of 197) reject at 5% level
  70 after FDR-Correction, FDR threshold: 0.022
- $125\,(=5^3)$ components: 82 (of 197) reject at 5% level
  76 after FDR-Correction, FDR threshold: 0.022

Estimators of location of change-points:

$$(\widehat{\vartheta}_1, \widehat{\vartheta}_2) = \arg\max_{x<y} \widehat{\mathbf{S}}^T_{\lfloor nx \rfloor, \lfloor ny \rfloor} \widehat{\Sigma}^{-1} \widehat{\mathbf{S}}_{\lfloor nx \rfloor, \lfloor ny \rfloor}.$$

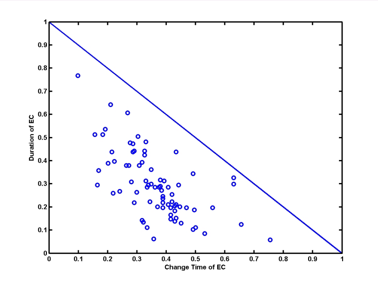
Theorem (Aston, K. (2012))

a) If the change is detectable, then $(\widehat{\vartheta}_1 - \vartheta_1, \widehat{\vartheta}_2 - \vartheta_2)^T = o_P(1)$.

b) If additionally a Hájek-Rényi-type condition holds for $\langle Y_i, w_l \rangle$,
then $$n(\widehat{\vartheta}_1 - \vartheta_1, \widehat{\vartheta}_2 - \vartheta_2)^T = O_P(1).$$

Estimated Change and Duration Distribution

Observations of several independent time series (persons) involving epidemic changes.
Of interest: Delay and duration distribution!



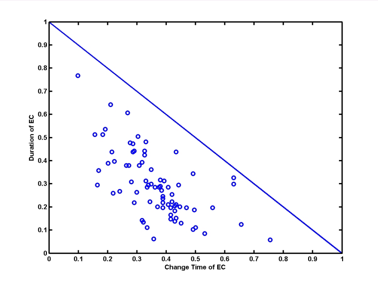
(125 separable principal components)

Estimated Change and Duration Distribution

Observations of several independent time series (persons) involving epidemic changes.

Of interest: Delay and duration distribution!

If estimators are consistent with certain rate (related to bandwidth), then Kernel density estimators based on estimated change-points are consistent (Aston, K. (2011)).

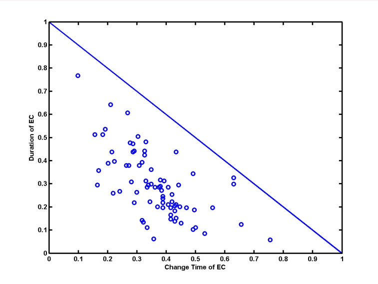


(125 separable principal components)

Estimated Change and Duration Distribution

Observations of several independent time series (persons) involving epidemic changes.
Of interest: Delay and duration distribution!

If estimators are consistent with certain rate (related to bandwidth), then Kernel density estimators based on estimated change-points are consistent (Aston, K. (2011)).



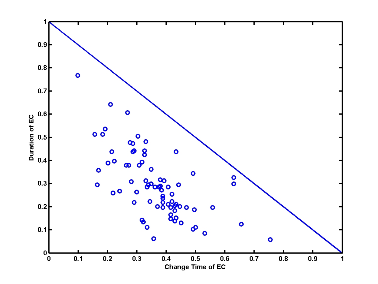
(125 separable principal components)

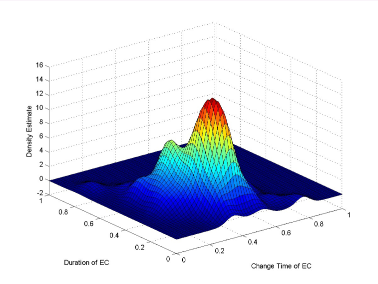
Estimated Change and Duration Distribution

Observations of several independent time series (persons) involving epidemic changes.

Of interest: Delay and duration distribution!

If estimators are consistent with certain rate (related to bandwidth), then Kernel density estimators based on estimated change-points are consistent (Aston, K. (2011)).





(125 separable principal components)

References:

J. Aston, C. Kirch.
Detecting and estimating epidemic changes in dependent functional data.
J. Multivariate Anal., 2012. To appear.

J. Aston, C. Kirch.
Estimation of the distribution of change-points with application to fMRI data.
CRiSM Research Reports, 2011.

A. Aue, R. Gabrys, L. Horváth, P. Kokoszka.
Estimation of a change-point in the mean function of functional data.
J. Multivariate Anal., 100:2254-2269, 2009.

I. Berkes, R. Gabrys, L. Horváth, P. Kokoszka.
Detecting changes in the mean of functional observations.
J. R. Stat. Soc. Ser. B, 71:927-946, 2009.

S. Hörmann, P. Kokoszka.
Weakly dependent functional data.
Ann. Statist., 38:1845-1884, 2010.

Thank you very much for your attention!

References:

J. Aston, C. Kirch.
Detecting and estimating epidemic changes in dependent functional data.
J. Multivariate Anal., 2012. To appear.

J. Aston, C. Kirch.
Estimation of the distribution of change-points with application to fMRI data.
CRiSM Research Reports, 2011.

A. Aue, R. Gabrys, L. Horváth, P. Kokoszka.
Estimation of a change-point in the mean function of functional data.
J. Multivariate Anal., 100:2254-2269, 2009.

I. Berkes, R. Gabrys, L. Horváth, P. Kokoszka.
Detecting changes in the mean of functional observations.
J. R. Stat. Soc. Ser. B, 71:927-946, 2009.

S. Hörmann, P. Kokoszka.
Weakly dependent functional data.
Ann. Statist., 38:1845-1884, 2010.

Thank you very much for your attention!