Change-Points in High-Dimensional Settings

Claudia Kirch

Karlsruhe Institute of Technology (KIT)

(Otto-von-Guericke University Magdeburg)

joint work with

John Aston

University of Cambridge

University of Bristol, 20.02.2015

Change-Point Setting

Consider the following setup:

$$X_{i,t} = \mu_i + \delta_{i,T}\, g(t/T) + e_{i,t}, \quad 1 \leqslant i \leqslant d = d_T, 1 \leqslant t \leqslant T,$$

where $\{(e_{1,t}, \ldots, e_{d,T})^T, t = 1, \ldots, T\}$ is i.i.d., $\operatorname{E} e_{i,j} = 0$, $0 < \operatorname{var} e_{i,t} < \infty$, $g(\cdot)$ Riemann-integrable.

- Change: $\mathbf{\Delta}_d = (\delta_{1,T}, \ldots, \delta_{d,T})^T$

Change-Point Setting

Consider the following setup:

$$X_{i,t} = \mu_i + \delta_{i,T}\, g(t/T) + e_{i,t}, \quad 1 \leqslant i \leqslant d = d_T, 1 \leqslant t \leqslant T,$$

where $\{(e_{1,t}, \ldots, e_{d,T})^T, t = 1, \ldots, T\}$ is i.i.d., $\mathrm{E}\, e_{i,j} = 0$, $0 < \operatorname{var} e_{i,t} < \infty$, $g(\cdot)$ Riemann-integrable.

- Change: $\boldsymbol{\Delta}_d = (\delta_{1,T}, \ldots, \delta_{d,T})^T$
- $g(\cdot)$ describes type of mean change, e.g.

Change-Point Setting

Consider the following setup:

$$X_{i,t} = \mu_i + \delta_{i,T}\, g(t/T) + e_{i,t}, \quad 1 \leqslant i \leqslant d = d_T, 1 \leqslant t \leqslant T,$$

where $\{(e_{1,t}, \ldots, e_{d,T})^T, t = 1, \ldots, T\}$ is i.i.d., $\mathrm{E}\, e_{i,j} = 0$, $0 < \operatorname{var} e_{i,t} < \infty$, $g(\cdot)$ Riemann-integrable.

- Change: $\mathbf{\Delta}_d = (\delta_{1,T}, \ldots, \delta_{d,T})^T$
- $g(\cdot)$ describes type of mean change, e.g.
  - AMOC-Location Model: $g(u) = 1_{\{u > \vartheta\}}$.

Change-Point Setting

Consider the following setup:

$$X_{i,t} = \mu_i + \delta_{i,T}\, g(t/T) + e_{i,t}, \quad 1 \leqslant i \leqslant d = d_T, 1 \leqslant t \leqslant T,$$

where $\{(e_{1,t}, \ldots, e_{d,T})^T, t = 1, \ldots, T\}$ is i.i.d., $\operatorname{E} e_{i,j} = 0$, $0 < \operatorname{var} e_{i,t} < \infty$, $g(\cdot)$ Riemann-integrable.

- Change: $\boldsymbol{\Delta}_d = (\delta_{1,T}, \ldots, \delta_{d,T})^T$
- $g(\cdot)$ describes type of mean change, e.g.
  - AMOC-Location Model: $g(u) = 1_{\{u>\vartheta\}}$.
  - Epidemic Location Model: $g(u) = 1_{\{\vartheta_1 < u \leqslant \vartheta_2\}}$.

Change-Point Setting

Consider the following setup:

$$X_{i,t} = \mu_i + \delta_{i,T}\, g(t/T) + e_{i,t}, \quad 1 \leqslant i \leqslant d = d_T, 1 \leqslant t \leqslant T,$$

where $\{(e_{1,t}, \ldots, e_{d,T})^T, t = 1, \ldots, T\}$ is i.i.d., $\operatorname{E} e_{i,j} = 0$,
$0 < \operatorname{var} e_{i,t} < \infty$, $g(\cdot)$ Riemann-integrable.

- Change: $\mathbf{\Delta}_d = (\delta_{1,T}, \ldots, \delta_{d,T})^T$
- $g(\cdot)$ describes type of mean change, e.g.
  - AMOC-Location Model: $g(u) = 1_{\{u > \vartheta\}}$.
  - Epidemic Location Model: $g(u) = 1_{\{\vartheta_1 < u \leqslant \vartheta_2\}}$.
- Test: $H_0 : \mathbf{\Delta}_d = 0, \qquad H_1 : \mathbf{\Delta}_d \neq 0.$

Change-Point Setting

Consider the following setup:

$$X_{i,t} = \mu_i + \delta_{i,T}\, g(t/T) + e_{i,t}, \quad 1 \leqslant i \leqslant d = d_T, 1 \leqslant t \leqslant T,$$

where $\{(e_{1,t}, \ldots, e_{d,T})^T, t = 1, \ldots, T\}$ is i.i.d., $\mathrm{E}\, e_{i,j} = 0$, $0 < \operatorname{var} e_{i,t} < \infty$, $g(\cdot)$ Riemann-integrable.

- Change: $\mathbf{\Delta}_d = (\delta_{1,T}, \ldots, \delta_{d,T})^T$
- $g(\cdot)$ describes type of mean change, e.g.
  - AMOC-Location Model: $g(u) = 1_{\{u>\vartheta\}}$.
  - Epidemic Location Model: $g(u) = 1_{\{\vartheta_1 < u \leqslant \vartheta_2\}}$.
- Test: $H_0 : \mathbf{\Delta}_d = 0, \qquad H_1 : \mathbf{\Delta}_d \neq 0.$

Change-Point Setting

Consider the following setup:

$$X_{i,t} = \mu_i + \delta_{i,T}\, g(t/T) + e_{i,t}, \quad 1 \leqslant i \leqslant d = d_T, 1 \leqslant t \leqslant T,$$

where $\{(e_{1,t}, \ldots, e_{d,T})^T, t = 1, \ldots, T\}$ is i.i.d., $\mathrm{E}\, e_{i,j} = 0$, $0 < \operatorname{var} e_{i,t} < \infty$, $g(\cdot)$ Riemann-integrable.

Asymptotic Framework:

- Multivariate Setting:
  $d > 1$ fixed, i.e. small in comparison to T.
- High-Dimensional/Panel-Data Setting:
  $d = d_T \to \infty$ as $T \to \infty$.

Change-Point Setting

Consider the following setup:

$$X_{i,t} = \mu_i + \delta_{i,T}\, g(t/T) + e_{i,t}, \quad 1 \leqslant i \leqslant d = d_T, 1 \leqslant t \leqslant T,$$

where $\{(e_{1,t}, \ldots, e_{d,T})^T, t = 1, \ldots, T\}$ is i.i.d., $\operatorname{E} e_{i,j} = 0$, $0 < \operatorname{var} e_{i,t} < \infty$, $g(\cdot)$ Riemann-integrable.

Asymptotic Framework:

- Multivariate Setting:
  $d > 1$ fixed, i.e. small in comparison to T.
- High-Dimensional/Panel-Data Setting:
  $d = d_T \to \infty$ as $T \to \infty$.

This talk. . .

is not about

- how time dependency influences tests.
  - Multivariate Setting:
    Dependency structure must allow for multivariate FCLT.

This talk. . .

is not about

- how time dependency influences tests.
  - Multivariate Setting:
    Dependency structure must allow for multivariate FCLT.
  - Panel Data Setting:
    More difficult but e.g. linear processes possible.

This talk. . .

is not about

- how time dependency influences tests.
  - Multivariate Setting:
    Dependency structure must allow for multivariate FCLT.
  - Panel Data Setting:
    More difficult but e.g. linear processes possible.

This talk. . .

is not about

- how time dependency influences tests.
- how change-point tests for different types of changes are obtained.

Change-Point-Statistics usually based on functionals of the centered partial sum process $\{\mathbf{Z}(\frac{k}{T}) = \frac{1}{\sqrt{T}} \sum_{j=1}^{k} (\mathbf{X}_t - \bar{\mathbf{X}}_T)\}$, e.g.

This talk. . .

is not about

- how time dependency influences tests.
- how change-point tests for different types of changes are obtained.

Change-Point-Statistics usually based on functionals of the centered partial sum process $\{\boldsymbol{Z}(\frac{k}{T}) = \frac{1}{\sqrt{T}} \sum_{j=1}^{k} (\mathbf{X}_t - \bar{\mathbf{X}}_T)\}$, e.g.

- AMOC-Change: $\max_{1 \leqslant k \leqslant T} w(\frac{k}{T}) \boldsymbol{Z}(\frac{k}{T}) \boldsymbol{A} \boldsymbol{Z}(\frac{k}{T})$.

This talk. . .

is not about

- how time dependency influences tests.
- how change-point tests for different types of changes are obtained.

Change-Point-Statistics usually based on functionals of the centered partial sum process $\{\boldsymbol{Z}(\frac{k}{T}) = \frac{1}{\sqrt{T}}\sum_{j=1}^{k}(\mathbf{X}_t - \bar{\mathbf{X}}_T)\}$, e.g.

- AMOC-Change:
  $\max_{1\leqslant k\leqslant T} w(\frac{k}{T})\boldsymbol{Z}(\frac{k}{T})\boldsymbol{A}\boldsymbol{Z}(\frac{k}{T})$.
- Epidemic Change:
  $\max_{1\leqslant k_1<k_2\leqslant T}(\boldsymbol{Z}(\frac{k_2}{T}) - \boldsymbol{Z}(\frac{k_1}{T}))^T\boldsymbol{A}(\boldsymbol{Z}(\frac{k_2}{T}) - \boldsymbol{Z}(\frac{k_1}{T}))$.

This talk...

is not about

- how time dependency influences tests.
- how change-point tests for different types of changes are obtained.

Change-Point-Statistics usually based on functionals of the centered partial sum process $\{\boldsymbol{Z}(\frac{k}{T}) = \frac{1}{\sqrt{T}} \sum_{j=1}^{k} (\mathbf{X}_t - \bar{\mathbf{X}}_T)\}$, e.g.

- AMOC-Change:
  $\max_{1 \leqslant k \leqslant T} w(\frac{k}{T}) \boldsymbol{Z}(\frac{k}{T}) \boldsymbol{A} \boldsymbol{Z}(\frac{k}{T})$.
- Epidemic Change:
  $\max_{1 \leqslant k_1 < k_2 \leqslant T} (\boldsymbol{Z}(\frac{k_2}{T}) - \boldsymbol{Z}(\frac{k_1}{T}))^T \boldsymbol{A} (\boldsymbol{Z}(\frac{k_2}{T}) - \boldsymbol{Z}(\frac{k_1}{T}))$.

This talk. . .

is not about

- how time dependency influences tests.
- how change-point tests for different types of changes are obtained.

but about

- asymptotic properties of tests based on projections.

This talk. . .

is not about

- how time dependency influences tests.
- how change-point tests for different types of changes are obtained.

but about

- **asymptotic properties of tests based on projections.**
- **how to quantify and compare the efficiency of high dimensional tests.**

This talk. . .

is not about

- how time dependency influences tests.
- how change-point tests for different types of changes are obtained.

but about

- **asymptotic properties of tests based on projections.**
- **how to quantify and compare the efficiency of high dimensional tests.**
- **the influence of misspecification of the model covariance on size and power.**

This talk. . .

is not about

- how time dependency influences tests.
- how change-point tests for different types of changes are obtained.

but about

- **asymptotic properties of tests based on projections.**
- **how to quantify and compare the efficiency of high dimensional tests.**
- **the influence of misspecification of the model covariance on size and power.**

Motivations for Projections

Change $\boldsymbol{\Delta}_d$ is always a one-dimensional object
(no matter what d does):

If we know, where to look, we can increase signal-to-noise ratio!

Motivations for Projections

Change $\boldsymbol{\Delta}_d$ is always a one-dimensional object
(no matter what d does):

If we know, where to look, we can increase signal-to-noise ratio!

In practical situation: Certain change-scenarios expected or of particular interest,

Motivations for Projections

Change $\boldsymbol{\Delta}_d$ is always a one-dimensional object (no matter what d does):

If we know, where to look, we can increase signal-to-noise ratio!

In practical situation: Certain change-scenarios expected or of particular interest, e.g.

- Economics: Performance of several companies looking for changes caused by recession.

Motivations for Projections

Change $\boldsymbol{\Delta}_d$ is always a one-dimensional object
(no matter what d does):

If we know, where to look, we can increase signal-to-noise ratio!

In practical situation: Certain change-scenarios expected or of particular interest, e.g.

- Economics: Performance of several companies looking for changes caused by recession.
- Medical studies: Knowledge about co-regulation of genes.

Motivations for Projections

Change $\boldsymbol{\Delta}_d$ is always a one-dimensional object (no matter what d does):

If we know, where to look, we can increase signal-to-noise ratio!

In practical situation: Certain change-scenarios expected or of particular interest, e.g.

- Economics: Performance of several companies looking for changes caused by recession.
- Medical studies: Knowledge about co-regulation of genes.

Projections allow to increase power for changes that are close to those scenarios at the cost of decreasing power in different directions.

Motivations for Projections

Change $\mathbf{\Delta}_d$ is always a one-dimensional object (no matter what d does):

If we know, where to look, we can increase signal-to-noise ratio!

In practical situation: Certain change-scenarios expected or of particular interest, e.g.

- Economics: Performance of several companies looking for changes caused by recession.
- Medical studies: Knowledge about co-regulation of genes.

Projections allow to increase power for changes that are close to those scenarios at the cost of decreasing power in different directions.

Analogous assertion hold for one- or two-sample location problem.

Motivations for Projections

Change $\boldsymbol{\Delta}_d$ is always a one-dimensional object
(no matter what d does):

If we know, where to look, we can increase signal-to-noise ratio!

In practical situation: Certain change-scenarios expected or of particular interest, e.g.

- Economics: Performance of several companies looking for changes caused by recession.
- Medical studies: Knowledge about co-regulation of genes.

Projections allow to increase power for changes that are close to those scenarios at the cost of decreasing power in different directions.

Analogous assertion hold for one- or two-sample location problem.

Statistics based on projections

Let $\mathbf{p}_d$ be a (possibly random) projection vector!

Consider univariate time series:

$$\langle \mathbf{X}(t), \mathbf{p}_d \rangle = \langle \boldsymbol{\mu}, \mathbf{p}_d \rangle + \langle \boldsymbol{\Delta}_d, \mathbf{p}_d \rangle g(t/T) + \langle \boldsymbol{e}_t, \mathbf{p}_d \rangle, \quad 1 \leqslant t \leqslant T.$$

Statistics based on projections

Let $\mathbf{p}_d$ be a (possibly random) projection vector!

Consider univariate time series:

$$\langle \mathbf{X}(t), \mathbf{p}_d \rangle = \langle \boldsymbol{\mu}, \mathbf{p}_d \rangle + \langle \boldsymbol{\Delta}_d, \mathbf{p}_d \rangle g(t/T) + \langle \boldsymbol{e}_t, \mathbf{p}_d \rangle, \quad 1 \leqslant t \leqslant T.$$

Standard univariate change-point statistics based on the centered partial sum process

$$U_{d,T}(x) = \langle \boldsymbol{Z}(x), \mathbf{p}_d \rangle = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tx \rfloor} \left(\langle \mathbf{X}_d(t), \mathbf{p}_d \rangle - \frac{1}{T} \sum_{j=1}^{T} \langle \mathbf{X}_d(j), \mathbf{p}_d \rangle \right),$$

Statistics based on projections

Let $\mathbf{p}_d$ be a (possibly random) projection vector!

Consider univariate time series:

$$\langle \mathbf{X}(t), \mathbf{p}_d \rangle = \langle \boldsymbol{\mu}, \mathbf{p}_d \rangle + \langle \boldsymbol{\Delta}_d, \mathbf{p}_d \rangle g(t/T) + \langle \mathbf{e}_t, \mathbf{p}_d \rangle, \quad 1 \leqslant t \leqslant T.$$

Standard univariate change-point statistics based on the centered partial sum process

$$U_{d,T}(x) = \langle \boldsymbol{Z}(x), \mathbf{p}_d \rangle = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tx \rfloor} \left(\langle \mathbf{X}_d(t), \mathbf{p}_d \rangle - \frac{1}{T} \sum_{j=1}^{T} \langle \mathbf{X}_d(j), \mathbf{p}_d \rangle \right),$$

e.g. simplest AMOC-statistic $\sup_{0 \leqslant x \leqslant 1} |U_{d,T}(x)|$.

Statistics based on projections

Let $\mathbf{p}_d$ be a (possibly random) projection vector!

Consider univariate time series:

$$\langle \mathbf{X}(t), \mathbf{p}_d \rangle = \langle \boldsymbol{\mu}, \mathbf{p}_d \rangle + \langle \boldsymbol{\Delta}_d, \mathbf{p}_d \rangle g(t/T) + \langle \boldsymbol{e}_t, \mathbf{p}_d \rangle, \quad 1 \leqslant t \leqslant T.$$

Standard univariate change-point statistics based on the centered partial sum process

$$U_{d,T}(x) = \langle \boldsymbol{Z}(x), \mathbf{p}_d \rangle = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tx \rfloor} \left(\langle \mathbf{X}_d(t), \mathbf{p}_d \rangle - \frac{1}{T} \sum_{j=1}^{T} \langle \mathbf{X}_d(j), \mathbf{p}_d \rangle \right),$$

e.g. simplest AMOC-statistic $\sup_{0 \leqslant x \leqslant 1} |U_{d,T}(x)|$.

Null Asymptotics I

Theorem

Let (i) $\mathbf{p}_d$ independent of $\{\mathbf{e}_t\}$, (ii) $\langle \mathbf{p}_d, \mathbf{e}_t \rangle$ non-degenerate

Null Asymptotics I

Theorem

Let (i) $\mathbf{p}_d$ independent of $\{\mathbf{e}_t\}$, (ii) $\langle \mathbf{p}_d, \mathbf{e}_t \rangle$ non-degenerate

(iii) $\frac{\|\mathbf{p}_d\|_1^2}{\mathbf{p}_d^T \operatorname{cov}(\mathbf{e}_t)\mathbf{p}_d^T} = o(T^{1-2/\nu})$ $a.s.$,

Null Asymptotics I

Theorem

Let (i) $\mathbf{p}_d$ independent of $\{\mathbf{e}_t\}$, (ii) $\langle \mathbf{p}_d, \mathbf{e}_t \rangle$ non-degenerate (iii) $\frac{\|\mathbf{p}_d\|_1^2}{\mathbf{p}_d^T \operatorname{cov}(\mathbf{e}_t)\mathbf{p}_d^T} = o(T^{1-2/\nu})$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\} \overset{D[0,1]}{\longrightarrow} \{B(x) : 0 \leqslant x \leqslant 1\} \quad a.s.,$$

where $\tau^2(\mathbf{p}_d) = \operatorname{var}(\langle \mathbf{e}_t, \mathbf{p}_d \rangle)$, which can be replaced by a suitable estimator.

Null Asymptotics I

Theorem

Let (i) $\mathbf{p}_d$ independent of $\{\mathbf{e}_t\}$, (ii) $\langle \mathbf{p}_d, \mathbf{e}_t \rangle$ non-degenerate (iii) $\frac{\|\mathbf{p}_d\|_1^2}{\mathbf{p}_d^T \operatorname{cov}(\mathbf{e}_t)\mathbf{p}_d^T} = o(T^{1-2/\nu})$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\} \xrightarrow{D[0,1]} \{B(x) : 0 \leqslant x \leqslant 1\} \quad a.s.,$$

where $\tau^2(\mathbf{p}_d) = \operatorname{var}(\langle \mathbf{e}_t, \mathbf{p}_d \rangle)$, which can be replaced by a suitable estimator.

Assumption (iii) is always fulfilled in the multivariate case (with fixed $\operatorname{cov}(\mathbf{e}_t)$) or if d grows sufficiently slow.

Null Asymptotics I

Theorem

Let (i) $\mathbf{p}_d$ independent of $\{\mathbf{e}_t\}$, (ii) $\langle \mathbf{p}_d, \mathbf{e}_t \rangle$ non-degenerate (iii) $\frac{\|\mathbf{p}_d\|_1^2}{\mathbf{p}_d^T \operatorname{cov}(\mathbf{e}_t)\mathbf{p}_d^T} = o(T^{1-2/\nu})$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\} \xrightarrow{D[0,1]} \{B(x) : 0 \leqslant x \leqslant 1\} \quad a.s.,$$

where $\tau^2(\mathbf{p}_d) = \operatorname{var}(\langle \mathbf{e}_t, \mathbf{p}_d \rangle)$, which can be replaced by a suitable estimator.

Assumption (iii) is always fulfilled in the multivariate case (with fixed $\operatorname{cov}(\mathbf{e}_t)$) or if d grows sufficiently slow.

If we want to allow for $d = d_T \to \infty$ faster, we need stronger assumptions on the error sequence!

Null Asymptotics I

Theorem

Let (i) $\mathbf{p}_d$ independent of $\{\mathbf{e}_t\}$, (ii) $\langle \mathbf{p}_d, \mathbf{e}_t \rangle$ non-degenerate (iii) $\frac{\|\mathbf{p}_d\|_1^2}{\mathbf{p}_d^T \operatorname{cov}(\mathbf{e}_t)\mathbf{p}_d^T} = o(T^{1-2/\nu})$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\} \xrightarrow{D[0,1]} \{B(x) : 0 \leqslant x \leqslant 1\} \quad a.s.,$$

where $\tau^2(\mathbf{p}_d) = \operatorname{var}(\langle \mathbf{e}_t, \mathbf{p}_d \rangle)$, which can be replaced by a suitable estimator.

Assumption (iii) is always fulfilled in the multivariate case (with fixed $\operatorname{cov}(\mathbf{e}_t)$) or if d grows sufficiently slow.

If we want to allow for $d = d_T \to \infty$ faster, we need stronger assumptions on the error sequence!

Assumption on the error sequence

Let $\eta_{1,t}, \eta_{2,t}, \ldots$ independent (identically distributed across time t) with $\operatorname{E}\eta_{i,t} = 0$, $\operatorname{var}\eta_{i,t} = 1$ and $\operatorname{E}|\eta_{i,t}|^{\nu} \leqslant C < \infty$, $\nu > 2$.

The error sequence is defined as

$$\mathbf{e}_t(d) = \sum_{j \geqslant 1} \mathbf{a}_j(d)\eta_{j,t}.$$

Assumption on the error sequence

Let $\eta_{1,t}, \eta_{2,t}, \ldots$ independent (identically distributed across time t) with $\operatorname{E}\eta_{i,t} = 0$, $\operatorname{var}\eta_{i,t} = 1$ and $\operatorname{E}|\eta_{i,t}|^{\nu} \leqslant C < \infty$, $\nu > 2$.

The error sequence is defined as

$$\mathbf{e}_t(d) = \sum_{j \geqslant 1} \mathbf{a}_j(d)\eta_{j,t}.$$

Three important special cases:

$\mathcal{C}.1$ Independent case: $e_{i,t}(d) = s_i\, \eta_{i,t}$.

Assumption on the error sequence

Let $\eta_{1,t}, \eta_{2,t}, \ldots$ independent (identically distributed across time t) with $\mathrm{E}\,\eta_{i,t} = 0$, $\mathrm{var}\,\eta_{i,t} = 1$ and $\mathrm{E}\,|\eta_{i,t}|^{\nu} \leqslant C < \infty$, $\nu > 2$.

The error sequence is defined as

$$\mathbf{e}_t(d) = \sum_{j \geqslant 1} \mathbf{a}_j(d) \eta_{j,t}.$$

Three important special cases:

$\mathcal{C}.1$ Independent case: $e_{i,t}(d) = s_i\, \eta_{i,t}$.

$\mathcal{C}.2$ Fully dependent case: $e_{i,t}(d) = \Phi_i\, \eta_{1,t}$.

Assumption on the error sequence

Let $\eta_{1,t}, \eta_{2,t}, \ldots$ independent (identically distributed across time t) with $\operatorname{E}\eta_{i,t} = 0$, $\operatorname{var}\eta_{i,t} = 1$ and $\operatorname{E}|\eta_{i,t}|^{\nu} \leqslant C < \infty$, $\nu > 2$.

The error sequence is defined as

$$\mathbf{e}_t(d) = \sum_{j \geqslant 1} \mathbf{a}_j(d)\eta_{j,t}.$$

Three important special cases:

$\mathcal{C}.1$ Independent case: $e_{i,t}(d) = s_i\,\eta_{i,t}$.

$\mathcal{C}.2$ Fully dependent case: $e_{i,t}(d) = \Phi_i\,\eta_{1,t}$.

$\mathcal{C}.3$ Mixed components: $e_{i,t}(d) = s_i\,\eta_{i,t} + \Phi_i\,\eta_{d+1,t}$.

Assumption on the error sequence

Let $\eta_{1,t}, \eta_{2,t}, \ldots$ independent (identically distributed across time t) with $\operatorname{E} \eta_{i,t} = 0$, $\operatorname{var} \eta_{i,t} = 1$ and $\operatorname{E} |\eta_{i,t}|^\nu \leqslant C < \infty$, $\nu > 2$.

The error sequence is defined as

$$\mathbf{e}_t(d) = \sum_{j \geqslant 1} \mathbf{a}_j(d) \eta_{j,t}.$$

Three important special cases:

$\mathcal{C}.1$ Independent case: $e_{i,t}(d) = s_i \, \eta_{i,t}$.

$\mathcal{C}.2$ Fully dependent case: $e_{i,t}(d) = \Phi_i \, \eta_{1,t}$.

$\mathcal{C}.3$ Mixed components: $e_{i,t}(d) = s_i \, \eta_{i,t} + \Phi_i \, \eta_{d+1,t}$.

Null Asymptotics II

Theorem

Let the above error structure hold and

(i) $\mathbf{p}_d$ independent of $\{\eta_{i,t}\}$

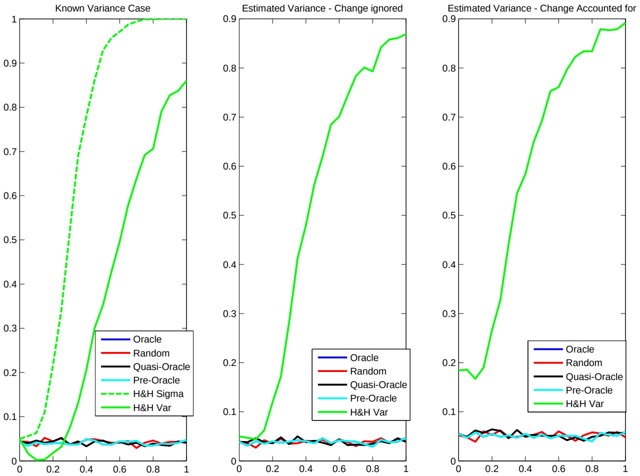
(ii) $\langle \mathbf{p}_d, \mathbf{e}_t \rangle$ non-degenerate

then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\} \xrightarrow{D[0,1]} \{B(x) : 0 \leqslant x \leqslant 1\} \quad a.s.,$$

where $\tau^2(\mathbf{p}_d) = \mathrm{var}(\langle \mathbf{e}_t, \mathbf{p}_d \rangle)$, which can be replaced by a suitable estimator.

Empirical Size, $\mathcal{C}.3$, $s_j = 1$, $T = 100$, $d = 200$



High-dimensional efficiency

Power comparison:

- Typically, large enough changes are detected by all statistics.

  Corresponding asymptotic theory: Fixed changes with $\|\boldsymbol{\Delta}_d\| = c > 0$.

High-dimensional efficiency

Power comparison:

- Typically, large enough changes are detected by all statistics.

  Corresponding asymptotic theory: Fixed changes with $\|\boldsymbol{\Delta}_d\| = c > 0$.

- To understand the small sample power using asymptotic tools contiguous alternatives need to be considered:

  $\|\boldsymbol{\Delta}_d\| \to 0$, such that asymptotic power strictly between α and 1.

High-dimensional efficiency

Power comparison:

- Typically, large enough changes are detected by all statistics.

  Corresponding asymptotic theory: Fixed changes with $\|\mathbf{\Delta}_d\| = c > 0$.

- To understand the small sample power using asymptotic tools contiguous alternatives need to be considered:

  $\|\mathbf{\Delta}_d\| \to 0$, such that asymptotic power strictly between α and 1.

  Unlike classic efficiency we obtain different rates for d increasing!

High-dimensional efficiency

Power comparison:

- Typically, large enough changes are detected by all statistics.

  Corresponding asymptotic theory: Fixed changes with $\|\boldsymbol{\Delta}_d\| = c > 0$.

- To understand the small sample power using asymptotic tools contiguous alternatives need to be considered:

  $\|\boldsymbol{\Delta}_d\| \to 0$, such that asymptotic power strictly between α and 1.

  Unlike classic efficiency we obtain different rates for d increasing!

Asymptotics for contiguous alternatives

Theorem

Let

$$\mathcal{E}_1^2(\boldsymbol{\Delta}_d, \mathbf{p}_d) := \frac{\|\boldsymbol{\Delta}_d\|^2 \|\mathbf{p}_d\|^2 \cos^2(\alpha_{\boldsymbol{\Delta}_d, \mathbf{p}_d})}{\tau^2(\mathbf{p}_d)}.$$

a) If $\sqrt{T}\,\mathcal{E}_1(\boldsymbol{\Delta}_d, \mathbf{p}_d) \to \infty$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)\sqrt{T}\,\mathcal{E}_1(\boldsymbol{\Delta}, \mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\}$$

$$\xrightarrow{D[0,1]} \left\{ \int_0^x g(t)\,dt - x \int_0^1 g(t)\,dt : 0 \leqslant x \leqslant 1 \right\} \qquad a.s.$$

Asymptotics for contiguous alternatives

Theorem

Let

$$\mathcal{E}_1^2(\boldsymbol{\Delta}_d, \mathbf{p}_d) := \frac{\|\boldsymbol{\Delta}_d\|^2 \|\mathbf{p}_d\|^2 \cos^2(\alpha_{\boldsymbol{\Delta}_d, \mathbf{p}_d})}{\tau^2(\mathbf{p}_d)}.$$

a) If $\sqrt{T}\,\mathcal{E}_1(\boldsymbol{\Delta}_d, \mathbf{p}_d) \to \infty$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)\sqrt{T}\,\mathcal{E}_1(\boldsymbol{\Delta}, \mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\}$$

$$\xrightarrow{D[0,1]} \left\{ \int_0^x g(t)\,dt - x \int_0^1 g(t)\,dt : 0 \leqslant x \leqslant 1 \right\} \qquad a.s.$$

Implications for change-point tests

Corollary (Asymptotic power one)

If $\sqrt{T}\,\mathcal{E}_1(\mathbf{\Delta}_d, \mathbf{p}_d) \to \infty$ *a.s., then for* $g(\cdot) \neq c$ *it holds*

$$P\left(\max_{0\leqslant x\leqslant 1} \frac{|U_{d,T}(x)|}{\tau(\mathbf{p}_d)} > c \,|\, \mathbf{p}_d\right) \to 1 \quad a.s.$$

Corollary

For the AMOC-situation $g(x) = 1_{\{x>\vartheta\}}$, *the estimator*

$$\widehat{\vartheta}_T = \left\lfloor \frac{\arg\ max_k |U_{d,T}(k/T)|}{T} \right\rfloor$$

is consistent, i.e. $P\left(\left|\widehat{\vartheta}_T - \vartheta\right| \geqslant \epsilon \,|\, \mathbf{p}_d\right) \to 0 \qquad a.s.$

Implications for change-point tests

Corollary (Asymptotic power one)

If $\sqrt{T}\,\mathcal{E}_1(\boldsymbol{\Delta}_d, \mathbf{p}_d) \to \infty$ *a.s., then for* $g(\cdot) \neq c$ *it holds*

$$P\left(\max_{0\leqslant x\leqslant 1} \frac{|U_{d,T}(x)|}{\tau(\mathbf{p}_d)} > c \,|\, \mathbf{p}_d\right) \to 1 \quad a.s.$$

Corollary

For the AMOC-situation $g(x) = 1_{\{x>\vartheta\}}$, *the estimator*

$$\widehat{\vartheta}_T = \left\lfloor \frac{\arg\ max_k |U_{d,T}(k/T)|}{T} \right\rfloor$$

is consistent, i.e. $P\left(\left|\widehat{\vartheta}_T - \vartheta\right| \geqslant \epsilon \,|\, \mathbf{p}_d\right) \to 0 \qquad a.s.$

Asymptotics for contiguous alternatives

Theorem

Let

$$\mathcal{E}_1^2(\mathbf{\Delta}_d, \mathbf{p}_d) := \frac{\|\mathbf{\Delta}_d\|^2 \|\mathbf{p}_d\|^2 \cos^2(\alpha_{\mathbf{\Delta}_d, \mathbf{p}_d})}{\tau^2(\mathbf{p}_d)}.$$

b) If $\sqrt{T}\,\mathcal{E}_1(\mathbf{\Delta}_d, \mathbf{p}_d) \to C_1 > 0$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)} - s_d C_1 \left(\int_0^x g(t)\,dt - x \int_0^1 g(t)\,dt \right) : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\} \xrightarrow{D[0,1]} \{B(x)\} \qquad a.s.,$$

where $s_d = \operatorname{sgn}(\mathbf{\Delta}_d^T \mathbf{p}_d)$.

Asymptotics for contiguous alternatives

Theorem

Let

$$\mathcal{E}_1^2(\boldsymbol{\Delta}_d, \mathbf{p}_d) := \frac{\|\boldsymbol{\Delta}_d\|^2 \|\mathbf{p}_d\|^2 \cos^2(\alpha_{\boldsymbol{\Delta}_d, \mathbf{p}_d})}{\tau^2(\mathbf{p}_d)}.$$

c) If $\sqrt{T}\,\mathcal{E}_1(\boldsymbol{\Delta}_d, \mathbf{p}_d) \to 0$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\} \xrightarrow{D[0,1]} \{B(x)\} \qquad a.s.$$

We call $\mathcal{E}_1(\boldsymbol{\Delta}_d, \mathbf{p}_d)$ the **absolute high dimensional efficiency** for the projection procedure!

Asymptotics for contiguous alternatives

Theorem

Let

$$\mathcal{E}_1^2(\boldsymbol{\Delta}_d, \mathbf{p}_d) := \frac{\|\boldsymbol{\Delta}_d\|^2 \|\mathbf{p}_d\|^2 \cos^2(\alpha_{\boldsymbol{\Delta}_d, \mathbf{p}_d})}{\tau^2(\mathbf{p}_d)}.$$

c) If $\sqrt{T}\, \mathcal{E}_1(\boldsymbol{\Delta}_d, \mathbf{p}_d) \to 0$ *a.s.*, then

$$\left\{ \frac{U_{d,T}(x)}{\tau(\mathbf{p}_d)} : 0 \leqslant x \leqslant 1 \,|\, \mathbf{p}_d \right\} \overset{D[0,1]}{\longrightarrow} \{B(x)\} \qquad a.s.$$

We call $\mathcal{E}_1(\boldsymbol{\Delta}_d, \mathbf{p}_d)$ the **absolute high dimensional efficiency** for the projection procedure!

Oracle Projection

Proposition

If Σ is invertible, then

$$\mathcal{E}_1(\boldsymbol{\Delta}_d, \mathbf{p}_d) = \|\Sigma^{-1/2}\boldsymbol{\Delta}_d\| \cos(\alpha_{\Sigma^{-1/2}\boldsymbol{\Delta}_d, \Sigma^{1/2}\mathbf{p}_d}).$$

High-dimensional efficiency only depends on magnitude of change $\|\Sigma^{-1/2}\boldsymbol{\Delta}_d\|$ and angle between projection and change.

Oracle Projection

Proposition

If Σ is invertible, then

$$\mathcal{E}_1(\mathbf{\Delta}_d, \mathbf{p}_d) = \|\Sigma^{-1/2}\mathbf{\Delta}_d\| \cos(\alpha_{\Sigma^{-1/2}\mathbf{\Delta}_d, \Sigma^{1/2}\mathbf{p}_d}).$$

High-dimensional efficiency only depends on magnitude of change $\|\Sigma^{-1/2}\mathbf{\Delta}_d\|$ and angle between projection and change.

The projection $\mathbf{o} = \Sigma^{-1}\mathbf{\Delta}_d$ maximizes $\mathcal{E}_1(\mathbf{\Delta}_d, \mathbf{p}_d)$ if Σ^{-1} exists. It is called **oracle**.

Oracle Projection

Proposition

If Σ is invertible, then

$$\mathcal{E}_1(\mathbf{\Delta}_d, \mathbf{p}_d) = \|\Sigma^{-1/2}\mathbf{\Delta}_d\| \cos(\alpha_{\Sigma^{-1/2}\mathbf{\Delta}_d, \Sigma^{1/2}\mathbf{p}_d}).$$

High-dimensional efficiency only depends on magnitude of change $\|\Sigma^{-1/2}\mathbf{\Delta}_d\|$ and angle between projection and change.

The projection $\mathbf{o} = \Sigma^{-1}\mathbf{\Delta}_d$ maximizes $\mathcal{E}_1(\mathbf{\Delta}_d, \mathbf{p}_d)$ if Σ^{-1} exists. It is called **oracle**.

Another way to think of it: First, standardize data via $\Sigma^{-1/2}$, then project onto the new change $\Sigma^{-1/2}\mathbf{\Delta}_d$.

Oracle Projection

Proposition

If Σ is invertible, then

$$\mathcal{E}_1(\mathbf{\Delta}_d, \mathbf{p}_d) = \|\Sigma^{-1/2}\mathbf{\Delta}_d\| \cos(\alpha_{\Sigma^{-1/2}\mathbf{\Delta}_d, \Sigma^{1/2}\mathbf{p}_d}).$$

High-dimensional efficiency only depends on magnitude of change $\|\Sigma^{-1/2}\mathbf{\Delta}_d\|$ and angle between projection and change.

The projection $\mathbf{o} = \Sigma^{-1}\mathbf{\Delta}_d$ maximizes $\mathcal{E}_1(\mathbf{\Delta}_d, \mathbf{p}_d)$ if Σ^{-1} exists. It is called **oracle**.

Another way to think of it: First, standardize data via $\Sigma^{-1/2}$, then project onto the new change $\Sigma^{-1/2}\mathbf{\Delta}_d$.

Benchmark: Correctly Scaled Random Projection

Theorem

Consider a random uniform projection $\mathbf{r}_d$ on the d-dimensional unit sphere and $\mathbf{r}_{\Sigma,d} = \Sigma^{-1/2}\mathbf{r}_d$.

Another way to think of it: Random projection on the unit sphere after standardizing the data!

Benchmark: Correctly Scaled Random Projection

Theorem

Consider a random uniform projection $\mathbf{r}_d$ on the d-dimensional unit sphere and $\mathbf{r}_{\Sigma,d} = \Sigma^{-1/2}\mathbf{r}_d$.

Then, there exist for all $\epsilon > 0$ constants $c, C > 0$, such that

$$P\left(c \leqslant \mathcal{E}_1(\mathbf{\Delta}, \mathbf{r}_{\Sigma,d})\,\frac{\sqrt{d}}{\|\Sigma^{-1/2}\mathbf{\Delta}\|} \leqslant C\right) \geqslant 1-\epsilon.$$

Another way to think of it: Random projection on the unit sphere after standardizing the data!

Benchmark: Correctly Scaled Random Projection

Theorem

Consider a random uniform projection $\mathbf{r}_d$ on the d-dimensional unit sphere and $\mathbf{r}_{\Sigma,d} = \Sigma^{-1/2}\mathbf{r}_d$.

Then, there exist for all $\epsilon > 0$ constants $c, C > 0$, such that

$$P\left(c \leqslant \mathcal{E}_1(\boldsymbol{\Delta}, \mathbf{r}_{\Sigma,d})\,\frac{\sqrt{d}}{\|\Sigma^{-1/2}\boldsymbol{\Delta}\|} \leqslant C\right) \geqslant 1-\epsilon.$$

Another way to think of it: Random projection on the unit sphere after standardizing the data!

Random high dimensional efficiency is an order $\sqrt{d}$ worse than oracle projection.

Benchmark: Correctly Scaled Random Projection

Theorem

Consider a random uniform projection $\mathbf{r}_d$ on the d-dimensional unit sphere and $\mathbf{r}_{\Sigma,d} = \Sigma^{-1/2}\mathbf{r}_d$.

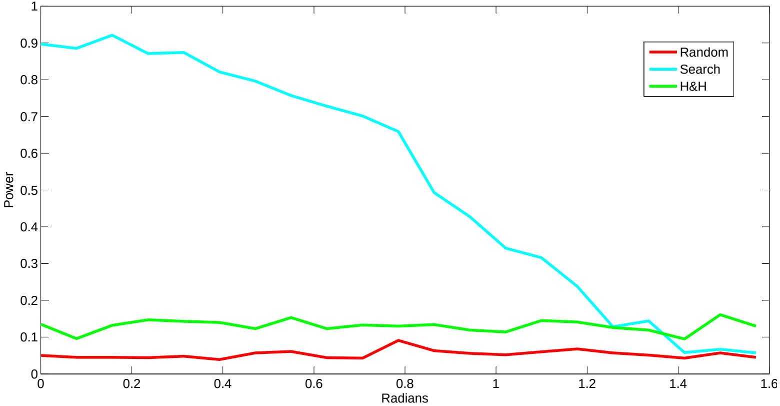
Then, there exist for all $\epsilon > 0$ constants $c, C > 0$, such that

$$P\left(c \leqslant \mathcal{E}_1(\boldsymbol{\Delta}, \mathbf{r}_{\Sigma,d}) \frac{\sqrt{d}}{\|\Sigma^{-1/2}\boldsymbol{\Delta}\|} \leqslant C\right) \geqslant 1 - \epsilon.$$

Another way to think of it: Random projection on the unit sphere after standardizing the data!

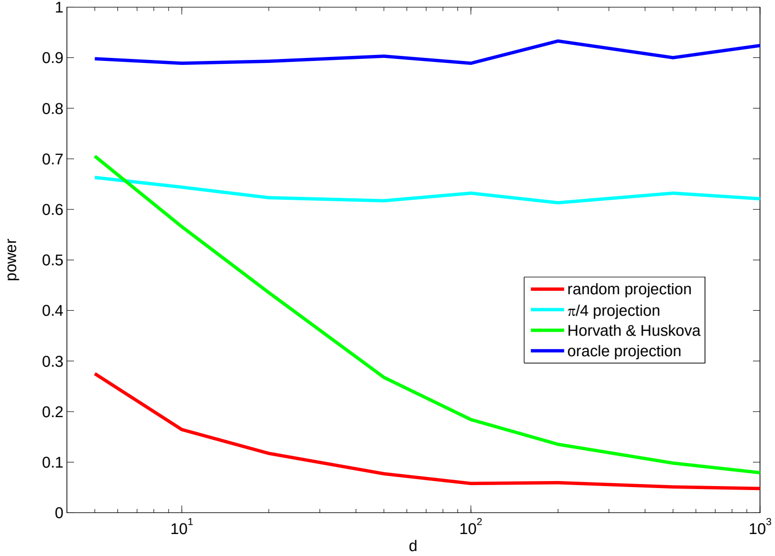
Random high dimensional efficiency is an order $\sqrt{d}$ worse than oracle projection.

Empirical Power: Different Angles, $T = 100$, $d = 200$



$\mathcal{C}.1$: $s_j = 1$, i.e. $\Sigma = I_d$

Power for increasing dimension, $T = 100$, $\|\mathbf{\Delta}\|$ constant



Misscaled Projections

In high dimensional settings: Covariance structure not known and not estimable!

Theorem

a) For a misscaled random projection $\mathbf{r}_{\mathbf{M},d} = \mathbf{M}^{-1/2}\mathbf{r}_d$:

$$P\left(c \leqslant \mathcal{E}_1^2(\boldsymbol{\Delta}_d, \mathbf{r}_{\mathbf{M},d})\,\frac{\operatorname{tr}(\mathbf{M}^{-1/2}\Sigma\mathbf{M}^{-1/2})}{\|\mathbf{M}^{-1/2}\boldsymbol{\Delta}_d\|^2} \leqslant C\right) \geqslant 1-\epsilon.$$

Misscaled Projections

In high dimensional settings: Covariance structure not known and not estimable!

Theorem

a) For a misscaled random projection $\mathbf{r}_{\mathbf{M},d} = \mathbf{M}^{-1/2}\mathbf{r}_d$:

$$P\left(c \leqslant \mathcal{E}_1^2(\boldsymbol{\Delta}_d, \mathbf{r}_{\mathbf{M},d})\,\frac{\operatorname{tr}(\mathbf{M}^{-1/2}\Sigma\mathbf{M}^{-1/2})}{\|\mathbf{M}^{-1/2}\boldsymbol{\Delta}_d\|^2} \leqslant C\right) \geqslant 1-\epsilon.$$

b) For the misscaled oracle $\mathbf{o}_{\mathbf{M}} = \mathbf{M}^{-1}\boldsymbol{\Delta}_d$, it holds

$$\mathcal{E}_1^2(\boldsymbol{\Delta}_d, \mathbf{o}_{\mathbf{M}}) \geqslant \frac{\|\mathbf{M}^{-1/2}\boldsymbol{\Delta}_d\|^2}{\operatorname{tr}(\mathbf{M}^{-1/2}\Sigma\mathbf{M}^{-1/2})},$$

with equality iff there is only one common factor and $\boldsymbol{\Delta}_d$ is a multiple of this factor.

Misscaled Projections

In high dimensional settings: Covariance structure not known and not estimable!

Theorem

a) For a misscaled random projection $\mathbf{r}_{\mathbf{M},d} = \mathbf{M}^{-1/2}\mathbf{r}_d$:

$$P\left(c \leqslant \mathcal{E}_1^2(\mathbf{\Delta}_d, \mathbf{r}_{\mathbf{M},d})\,\frac{\operatorname{tr}(\mathbf{M}^{-1/2}\Sigma\mathbf{M}^{-1/2})}{\|\mathbf{M}^{-1/2}\mathbf{\Delta}_d\|^2} \leqslant C\right) \geqslant 1-\epsilon.$$

b) For the misscaled oracle $\mathbf{o}_{\mathbf{M}} = \mathbf{M}^{-1}\mathbf{\Delta}_d$, it holds

$$\mathcal{E}_1^2(\mathbf{\Delta}_d, \mathbf{o}_{\mathbf{M}}) \geqslant \frac{\|\mathbf{M}^{-1/2}\mathbf{\Delta}_d\|^2}{\operatorname{tr}(\mathbf{M}^{-1/2}\Sigma\mathbf{M}^{-1/2})},$$

with equality iff there is only one common factor and $\mathbf{\Delta}_d$ is a multiple of this factor.

Misscaled Oracle Projection

- The projection ${}_q\mathbf{o} = (\Delta_1/\sigma_1^2, \ldots, \Delta_d/\sigma_d^2)^T$ is called **quasi-oracle**, if $\sigma_j^2 > 0$, $j = 1, \ldots, d$.
- The projection ${}_p\mathbf{o} = \boldsymbol{\Delta}_d$ is called **pre-oracle**.

Misscaled Oracle Projection

- The projection ${}_q\mathbf{o} = (\Delta_1/\sigma_1^2, \ldots, \Delta_d/\sigma_d^2)^T$ is called **quasi-oracle**, if $\sigma_j^2 > 0$, $j = 1, \ldots, d$.
- The projection ${}_p\mathbf{o} = \boldsymbol{\Delta}_d$ is called **pre-oracle**.

Proposition

If the channels are uncorrelated, the oracle and the quasi-oracle are equal.

Misscaled Oracle Projection

- The projection ${}_q\mathbf{o} = (\Delta_1/\sigma_1^2, \ldots, \Delta_d/\sigma_d^2)^T$ is called **quasi-oracle**, if $\sigma_j^2 > 0$, $j = 1, \ldots, d$.
- The projection ${}_p\mathbf{o} = \boldsymbol{\Delta}_d$ is called **pre-oracle**.

Proposition

If the channels are uncorrelated, the oracle and the quasi-oracle are equal.

If additionally $0 < c \leqslant \sigma_i^2 < C < \infty$ for $i = 1, \ldots, d$, then

$$\frac{c}{C}\mathcal{E}_1(\boldsymbol{\Delta}, {}_q\mathbf{o}) \leqslant \mathcal{E}_1(\boldsymbol{\Delta}, {}_p\mathbf{o}) \leqslant \mathcal{E}_1(\boldsymbol{\Delta}, {}_q\mathbf{o}),$$

i.e. the high dimensional efficiency for pre- and (quasi-)oracle is of the same order.

Misscaled Oracle Projection

- The projection ${}_q\mathbf{o} = (\Delta_1/\sigma_1^2, \ldots, \Delta_d/\sigma_d^2)^T$ is called **quasi-oracle**, if $\sigma_j^2 > 0$, $j = 1, \ldots, d$.
- The projection ${}_p\mathbf{o} = \boldsymbol{\Delta}_d$ is called **pre-oracle**.

Proposition

If the channels are uncorrelated, the oracle and the quasi-oracle are equal.

If additionally $0 < c \leqslant \sigma_i^2 < C < \infty$ for $i = 1, \ldots, d$, then

$$\frac{c}{C}\mathcal{E}_1(\boldsymbol{\Delta}, {}_q\mathbf{o}) \leqslant \mathcal{E}_1(\boldsymbol{\Delta}, {}_p\mathbf{o}) \leqslant \mathcal{E}_1(\boldsymbol{\Delta}, {}_q\mathbf{o}),$$

i.e. the high dimensional efficiency for pre- and (quasi-)oracle is of the same order.

Benchmark: Random Projection

Remark:

It can happen, that all the efficiency of all three oracle projections is of the same order as for the random projection!

Example:

$\mathcal{C}.3$: Mixed dependence with common factor:

If $\boldsymbol{\Delta}_d \sim \boldsymbol{\Phi}$ projection maximizes not only the signal but also noise!

Benchmark: Random Projection

Remark:

It can happen, that all the efficiency of all three oracle projections is of the same order as for the random projection!

Example:

$\mathcal{C}.3$: Mixed dependence with common factor:

If $\mathbf{\Delta}_d \sim \mathbf{\Phi}$ projection maximizes not only the signal but also noise!

Competing multivariate or panel data procedures are also of the same order!

Benchmark: Random Projection

Remark:

It can happen, that all the efficiency of all three oracle projections is of the same order as for the random projection!

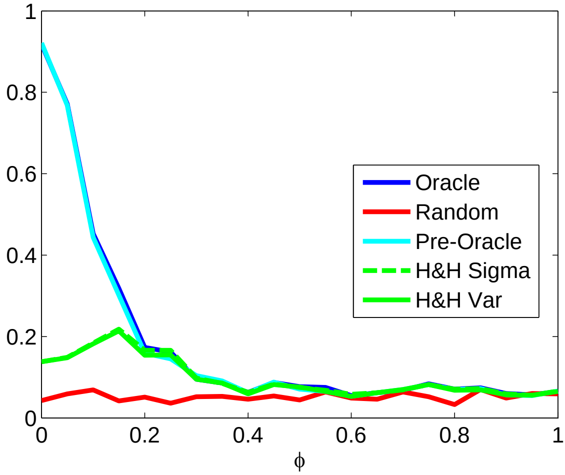
Example:

$\mathcal{C}.3$: Mixed dependence with common factor:

If $\boldsymbol{\Delta}_d \sim \boldsymbol{\Phi}$ projection maximizes not only the signal but also noise!

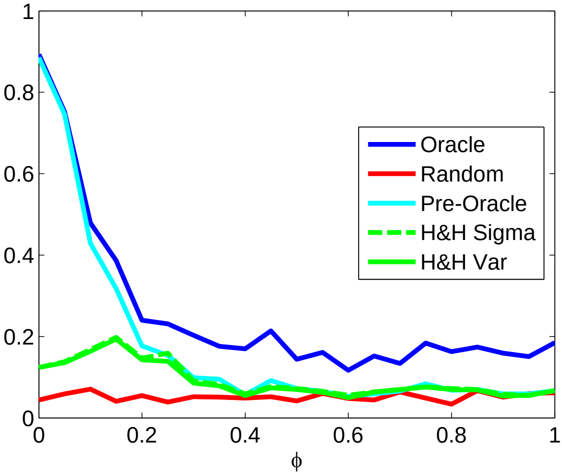
Competing multivariate or panel data procedures are also of the same order!

Empirical Power: $s_j = 1$, $T = 100$, $d = 200$



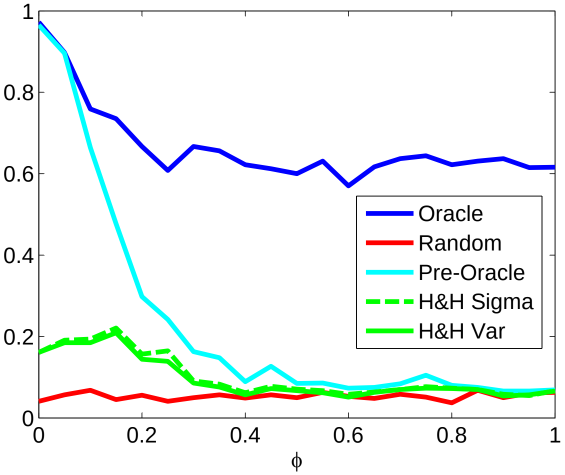
Angle between $\boldsymbol{\Delta}_d$ and $\boldsymbol{\Phi}_d$ is 0 radians.

Empirical Power: $s_j = 1,\ T = 100,\ d = 200$



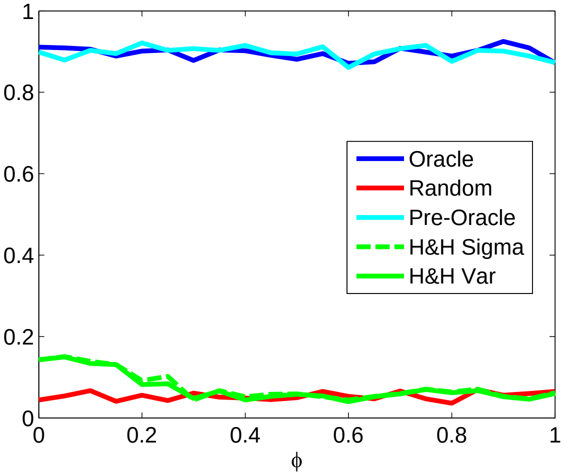
Angle between $\mathbf{\Delta}_d$ and $\mathbf{\Phi}_d$ is $\pi/8$ radians.

Empirical Power: $s_j = 1,\ T = 100,\ d = 200$



Angle between $\mathbf{\Delta}_d$ and $\mathbf{\Phi}_d$ is $\pi/4$ radians.

Empirical Power: $s_j = 1$, $T = 100$, $d = 200$



Angle between $\mathbf{\Delta}_d$ and $\mathbf{\Phi}_d$ is $\pi/2$ radians.

Comparison with Panel-Data-Statistics

Null asymptotics for independent panels

Theorem (Horváth, Hušková (2012))

If the panels are independent, $\sigma_i^2 = \operatorname{var} e_{i,t} \geqslant c > 0$ for all i and $\mathrm{E}\,|e_{i,t}|^\nu \leqslant C < \infty$ for some $\nu > 4$ and $\frac{d}{T^2} \to 0$,
then it holds under the null hypothesis of no change

$$\frac{1}{\sqrt{d}} \sum_{i=1}^{d} \left(\frac{1}{\sigma_i^2} Z_{T,i}^2(x) - \frac{\lfloor Tx \rfloor (T - \lfloor Tx \rfloor)}{T^2} \right) \xrightarrow{D[0,1]} \sqrt{2}(1-x)^2 W\left(\frac{x^2}{(1-x)^2} \right),$$

where $W(\cdot)$ is a standard Wiener process.

Comparison with Panel-Data-Statistics

Null asymptotics for independent panels

Theorem (Horváth, Hušková (2012))

If the panels are independent, $\sigma_i^2 = \operatorname{var} e_{i,t} \geqslant c > 0$ for all i and $\operatorname{E}|e_{i,t}|^{\nu} \leqslant C < \infty$ for some $\nu > 4$ and $\frac{d}{T^2} \to 0$,
then it holds under the null hypothesis of no change

$$\frac{1}{\sqrt{d}} \sum_{i=1}^{d} \left(\frac{1}{\sigma_i^2} Z_{T,i}^2(x) - \frac{\lfloor Tx \rfloor (T - \lfloor Tx \rfloor)}{T^2} \right) \xrightarrow{D[0,1]} \sqrt{2}(1-x)^2 W\left(\frac{x^2}{(1-x)^2} \right),$$

where $W(\cdot)$ is a standard Wiener process.

High-dimensional efficiency for independent panels

Theorem

In this independent setting, the high dimensional efficiency is given by

$$\mathcal{E}_2(\boldsymbol{\Delta}_d) = \frac{1}{d^{1/4}} \|\Sigma^{-1/2}\boldsymbol{\Delta}_d\|.$$

The panel high dimensional efficiency is

- an order $d^{1/4}$ better than the random projection

High-dimensional efficiency for independent panels

Theorem

In this independent setting, the high dimensional efficiency is given by

$$\mathcal{E}_2(\mathbf{\Delta}_d) = \frac{1}{d^{1/4}} \|\Sigma^{-1/2}\mathbf{\Delta}_d\|.$$

The panel high dimensional efficiency is

- an order $d^{1/4}$ better than the random projection
- an order $d^{1/4}$ worse than all three oracles.

High-dimensional efficiency for independent panels

Theorem

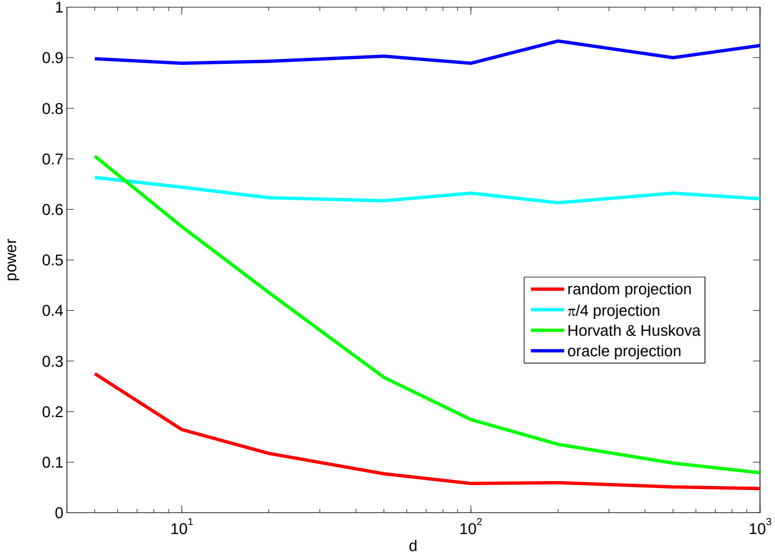
In this independent setting, the high dimensional efficiency is given by

$$\mathcal{E}_2(\mathbf{\Delta}_d) = \frac{1}{d^{1/4}} \|\Sigma^{-1/2}\mathbf{\Delta}_d\|.$$

The panel high dimensional efficiency is

- an order $d^{1/4}$ better than the random projection
- an order $d^{1/4}$ worse than all three oracles.

Power for increasing dimension, $T = 100$, $\|\mathbf{\Delta}\|$ constant



Null asymptotics for dependent panels

Theorem (Horváth, Hušková (2012))

Let Case $\mathcal{C}.3$ hold (mixed with common factor) with certain moment conditions. If

$$\frac{1}{\sqrt{d}} A_d := \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \frac{\Phi_i^2}{\sigma_i^2} \to \infty,$$

Null asymptotics for dependent panels

Theorem (Horváth, Hušková (2012))

Let Case $\mathcal{C}.3$ hold (mixed with common factor) with certain moment conditions. If

$$\frac{1}{\sqrt{d}} A_d := \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \frac{\Phi_i^2}{\sigma_i^2} \to \infty,$$

then

$$\frac{1}{A_d} \sum_{i=1}^{d} \left(\frac{1}{\sigma_i^2} Z_{T,i}^2(x) - \frac{\lfloor Tx \rfloor (T - \lfloor Tx \rfloor)}{T^2} \right) \xrightarrow{D[0,1]} B^2(x) - x(1-x),$$

where $\{B(x) : 0 \leqslant x \leqslant 1\}$ is a standard Brownian bridge.

Null asymptotics for dependent panels

Theorem (Horváth, Hušková (2012))

Let Case $\mathcal{C}.3$ hold (mixed with common factor) with certain moment conditions. If

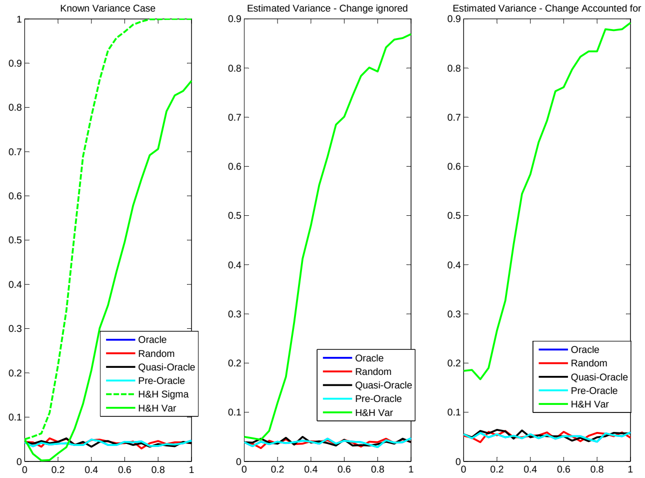
$$\frac{1}{\sqrt{d}}A_d := \frac{1}{\sqrt{d}}\sum_{i=1}^{d}\frac{\Phi_i^2}{\sigma_i^2} \to \infty,$$

then

$$\frac{1}{A_d}\sum_{i=1}^{d}\left(\frac{1}{\sigma_i^2}Z_{T,i}^2(x) - \frac{\lfloor Tx\rfloor(T-\lfloor Tx\rfloor)}{T^2}\right) \xrightarrow{D[0,1]} B^2(x) - x(1-x),$$

where $\{B(x) : 0 \leqslant x \leqslant 1\}$ is a standard Brownian bridge.

Empirical Size, $\mathcal{C}.3$, $s_j = 1$, $T = 100$, $d = 200$



High-dimensional efficiency for dependent panels

Theorem

a) In this situation the high dimensional efficiency is given by

$$\mathcal{E}_3(\boldsymbol{\Delta}_d) = \sqrt{\frac{1}{A_d}\, \boldsymbol{\Delta}_d^T \operatorname{diag}\left(\frac{1}{s_1^2 + \Phi_1^2}, \ldots, \frac{1}{s_d^2 + \Phi_d^2}\right) \boldsymbol{\Delta}_d}.$$

b) The high dimensional efficiency of the quasi-oracle is always at least as good as the one of the misspecified panel statistic.

High-dimensional efficiency for dependent panels

Theorem

a) In this situation the high dimensional efficiency is given by

$$\mathcal{E}_3(\mathbf{\Delta}_d) = \sqrt{\frac{1}{A_d}\,\mathbf{\Delta}_d^T \operatorname{diag}\left(\frac{1}{s_1^2+\Phi_1^2}, \ldots, \frac{1}{s_d^2+\Phi_d^2}\right)\mathbf{\Delta}_d}.$$

b) The high dimensional efficiency of the quasi-oracle is always at least as good as the one of the misspecified panel statistic.

c) If additionally $A_d/d \to A > 0$, then the high dimensional efficiency is of the same order as for the random projection.

High-dimensional efficiency for dependent panels

Theorem

a) In this situation the high dimensional efficiency is given by

$$\mathcal{E}_3(\mathbf{\Delta}_d) = \sqrt{\frac{1}{A_d}\,\mathbf{\Delta}_d^T \operatorname{diag}\left(\frac{1}{s_1^2+\Phi_1^2},\ldots,\frac{1}{s_d^2+\Phi_d^2}\right)\mathbf{\Delta}_d}.$$

b) The high dimensional efficiency of the quasi-oracle is always at least as good as the one of the misspecified panel statistic.

c) If additionally $A_d/d \to A > 0$, then the high dimensional efficiency is of the same order as for the random projection.

Remark:

For the oracle projections this was only the case for $\mathbf{\Delta}_d \sim \mathbf{\Phi}$.

High-dimensional efficiency for dependent panels

Theorem

a) In this situation the high dimensional efficiency is given by

$$\mathcal{E}_3(\boldsymbol{\Delta}_d) = \sqrt{\frac{1}{A_d}\, \boldsymbol{\Delta}_d^T \operatorname{diag}\left(\frac{1}{s_1^2 + \Phi_1^2}, \ldots, \frac{1}{s_d^2 + \Phi_d^2}\right) \boldsymbol{\Delta}_d}.$$

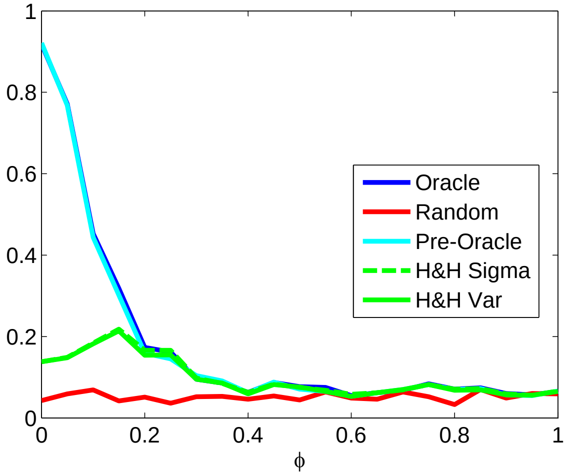
b) The high dimensional efficiency of the quasi-oracle is always at least as good as the one of the misspecified panel statistic.

c) If additionally $A_d/d \to A > 0$, then the high dimensional efficiency is of the same order as for the random projection.

Remark:

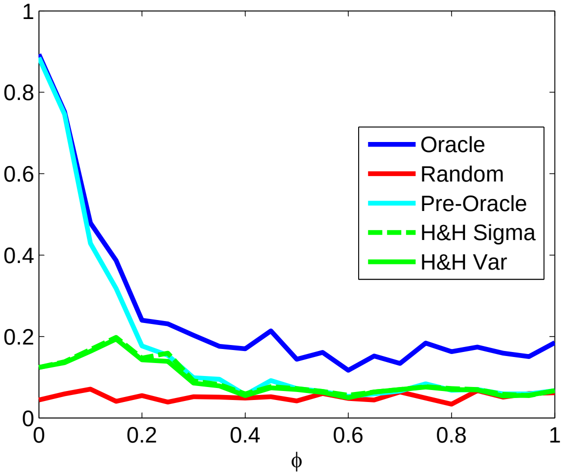
For the oracle projections this was only the case for $\boldsymbol{\Delta}_d \sim \boldsymbol{\Phi}$.

Empirical Power: $s_j = 1,\ T = 100,\ d = 200$



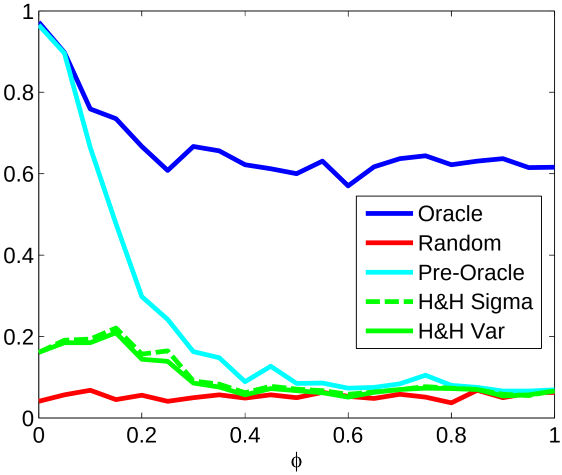
Angle between $\mathbf{\Delta}_d$ and $\mathbf{\Phi}_d$ is 0 radians.

Empirical Power: $s_j = 1,\ T = 100,\ d = 200$



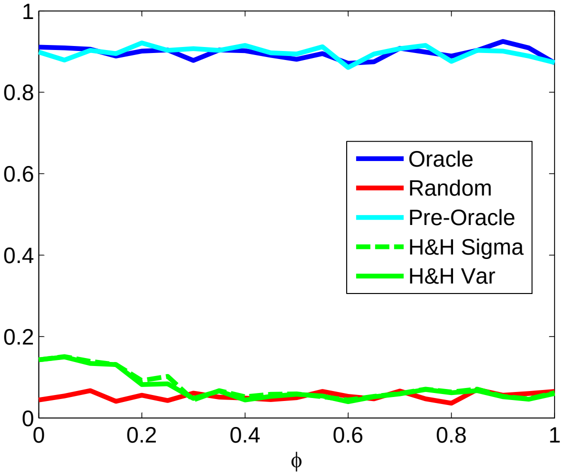
Angle between $\mathbf{\Delta}_d$ and $\mathbf{\Phi}_d$ is $\pi/8$ radians.

Empirical Power: $s_j = 1$, $T = 100$, $d = 200$



Angle between $\mathbf{\Delta}_d$ and $\mathbf{\Phi}_d$ is $\pi/4$ radians.

Empirical Power: $s_j = 1,\ T = 100,\ d = 200$



Angle between $\boldsymbol{\Delta}_d$ and $\boldsymbol{\Phi}_d$ is $\pi/2$ radians.

Outlook: Different projection methods

Multiple Projections:

- Multivariate change-point tests in lower dimensional space:
  Only information about subspace not actual directions used.

Outlook: Different projection methods

Multiple Projections:

- Multivariate change-point tests in lower dimensional space:
  Only information about subspace not actual directions used.
- Combine univariate projections in different way, e.g. as maximum of the univariate change-point statistics.

Outlook: Different projection methods

Multiple Projections:

- Multivariate change-point tests in lower dimensional space: Only information about subspace not actual directions used.
- Combine univariate projections in different way, e.g. as maximum of the univariate change-point statistics.

Multiple Changes:

- Use multiple projections, one for each possible change.

Outlook: Different projection methods

Multiple Projections:

- Multivariate change-point tests in lower dimensional space: Only information about subspace not actual directions used.
- Combine univariate projections in different way, e.g. as maximum of the univariate change-point statistics.

Multiple Changes:

- Use multiple projections, one for each possible change.

Data-Driven Projections:

Outlook: Different projection methods

Multiple Projections:

- Multivariate change-point tests in lower dimensional space:
  Only information about subspace not actual directions used.
- Combine univariate projections in different way, e.g. as maximum of the univariate change-point statistics.

Multiple Changes:

- Use multiple projections, one for each possible change.

Data-Driven Projections:

- Principal Component Analysis:
  Empirical size close to nominal size, certain power advantage for fixed alternatives if projection is necessary (Aston, K. (2012)) but not for contiguous alternatives.

Outlook: Different projection methods

Multiple Projections:

- Multivariate change-point tests in lower dimensional space: Only information about subspace not actual directions used.
- Combine univariate projections in different way, e.g. as maximum of the univariate change-point statistics.

Multiple Changes:

- Use multiple projections, one for each possible change.

Data-Driven Projections:

- Principal Component Analysis: Empirical size close to nominal size, certain power advantage for fixed alternatives if projection is necessary (Aston, K. (2012)) but not for contiguous alternatives.
- Univariate estimation of $\delta_{i,d}$: Irreparable size problems.

Outlook: Different projection methods

Multiple Projections:

- Multivariate change-point tests in lower dimensional space: Only information about subspace not actual directions used.
- Combine univariate projections in different way, e.g. as maximum of the univariate change-point statistics.

Multiple Changes:

- Use multiple projections, one for each possible change.

Data-Driven Projections:

- Principal Component Analysis: Empirical size close to nominal size, certain power advantage for fixed alternatives if projection is necessary (Aston, K. (2012)) but not for contiguous alternatives.
- Univariate estimation of $\delta_{i,d}$: Irreparable size problems.

For further reading:

Aston, Kirch
Change-points in high dimensional settings.
Preprint, 2014.

Horváth , Hušková
Change-point detection in panel data.
J. Time Ser. Anal., 33:631-648, 2012.

Thank you very much for your attention!

For further reading:

Aston, Kirch
Change-points in high dimensional settings.
Preprint, 2014.

Horváth , Husková
Change-point detection in panel data.
J. Time Ser. Anal., 33:631-648, 2012.

Thank you very much for your attention!