Dimension-agnostic Change Point Testing

Xiaofeng Shao

Department of Statistics
University of Illinois at Urbana-Champaign

November 2024 (University of Kentucky)
https://arxiv.org/abs/2303.10808
with Hanjia Gao (UC Irvine) and Runmin Wang (Texas A&M)

Setting

- **Data:** Given the observations $\{X_t\}_{t=1}^n \in \mathbb{R}^p$ with both temporal and cross-sectional dependence. We denote $\mu_t = \mathbb{E}(X_t)$ for $t = 1, \cdots, n$.
- **Single change point testing:**
  To test $H_0: \ \mu_1 = \cdots = \mu_n$ versus

$$H_1: \ \mu_1 = \cdots = \mu_{k_0} \neq \mu_{k_0+1} = \cdots = \mu_n$$

  where $k_0 = n\varepsilon_0$ with $\varepsilon_0 \in (0, 1)$ is an unknown location.
- p can be low or high relative to n.

Applications/Motivation:

- Genomics: Wang, T. and Samworth (2018)
- Finance: Wang, G. and Feng (2023)
- Climatology: Lund and Shi (2023)
- Neuroinformatics: Toutounji and Durstewitz (2018)

- **Existing work:**

Low/fixed-dimensional	High-dimensional
Page (1954, 1955), $\cdots$	Horvath and Huskova (2012)
Shao and Zhang (2010)	Jirak (2015), Cho (2016)
Zhang and Lavitas (2018)	Wang and Samworth (2018)
Aue and Horvath (2013, review)	Enikeeva and Harchaoui (2019)
Casini and Perron (2019, review) etc	Wang, Zhu, Vogulshev and Shao (2022)
	Liu et al. (2022, review) etc

- **Limitations:**
  - $\ast$ Low/fixed-dimensional setup
    - p is fixed and small
    - theory is justified specifically for small/fixed p
    - not applicable when $p > n$ and may exhibit serious size distortion when p is moderate
  - $\ast$ High-dimensional setup
    - p is high and is comparable to or exceeds n
    - theory is justified when $p \to \infty$
    - the approximation accuracy may rely on the central limit effect from the high dimension, often has serious size distortion for data of low or moderate dimension

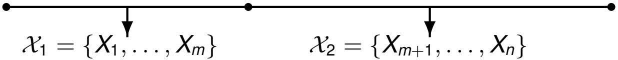
- **Motivation:**
  - When $p = 20, n = 100$, shall we
    - use a low-dimensional method (i.e., calibrating a statistic by fixing the dimension p while letting $n \to \infty$)?
    - use a high-dimensional method (i.e., calibrating a statistic by letting $p \to \infty$ as $n \to \infty$)?
  - Calibration of a test depends on the assumption of how p scales with n, which is usually pre-decided but unverifiable.
    1. $p = 20$ fixed?
    2. $p/n = 0.2$ fixed?
    3. $p/\sqrt{n} = 2$ fixed?

Dimension-agnostic inference (Kim and Ramdas (2024)):

$\cdots$ *the goal of dimension-agnostic inference: developing methods whose validity does not depend on any assumption on p versus n.*

Review of Kim and Ramdas (2024)

- **Data:** iid samples $\{X_t\}_{t=1}^{n} \in \mathbb{R}^p$ with mean $\mathbb{E}[X_t] = \mu$.
- **One sample mean testing:** $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$.
- **Sample splitting:**



- **Projection:** project sample mean from $\mathcal{X}_2$ onto $\mathcal{X}_1$: let

$$f_m := \frac{1}{n-m} \sum_{j=m+1}^{n} X_j, \quad Z_j = f_m^\top X_j, \; j = 1, \cdots, m.$$

- **Studentization**:

$$T_m = \frac{\sqrt{m}\bar{Z}_m}{\sqrt{m^{-1} \sum_{i=1}^{m} (Z_i - \bar{Z}_m)^2}}, \text{ where } \bar{Z}_m = m^{-1} \sum_{j=1}^{m} Z_j$$

Review of Kim and Ramdas (2024) cont'd

Let $X \in \mathbb{R}^p$ be a continous random vector with distribution P and mean μ. For a given $L > 1$, define

$$\mathcal{Q}_{0,p}(L) = \left\{ P : \sup_{u \in \mathcal{S}^{p-1}} \frac{\{E_P[|u^T(X-\mu)|^3]\}^{1/3}}{\{E_P[|u^T(X-\mu)|^2]\}^{1/2}} \leq L^{1/3} \right\},$$

where $\mathcal{S}^{p-1}$ denotes the $(p-1)$-dimensional unit sphere. Then there exists an absolute constant $C > 0$ such that under H_0,

$$\sup_{p \geq 1} \sup_{P \in \mathcal{Q}_{0,p}(L)} \sup_{-\infty < z < \infty} |P(T_m \leq z) - \Phi(z)| \leq \frac{CL}{\sqrt{m}}.$$

— $\mathcal{Q}_{0,p}(L)$ can be satisfied by multivariate sub-Gaussianity or sub-Exponentiality

Review of Kim and Ramdas (2024) cont'd

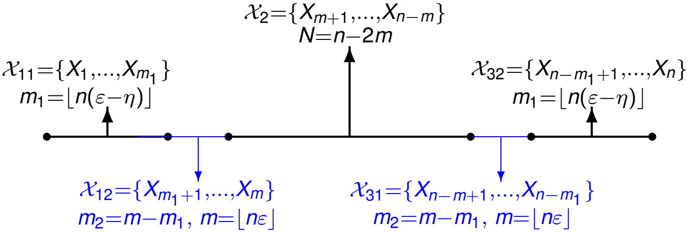
— Power results

- The power of T_m is lower by a $\sqrt{2}$ factor than the usual U-statistic-based test [Chen and Qin (2010)] in high dimension (a price to pay for dimension-agnostic property).
- The test is minimax rate optimal in terms of Euclidean norm deviations.
- Same phenomenon for covariance testing.

— Aggregating test statistics from multiple sample splitting may lose the dimension-agnostic property!

Proposed Method

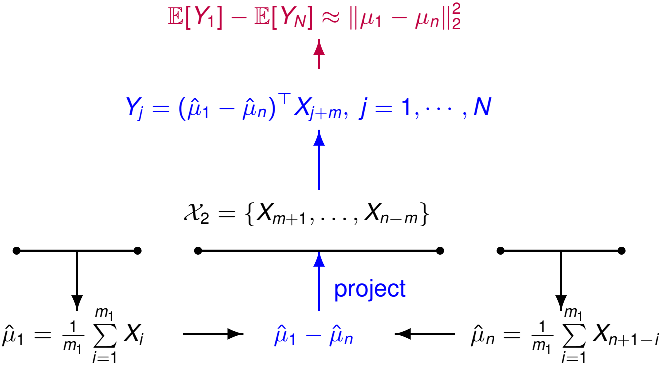
- **Methodology (Step 1: Splitting and trimming)**:



  * $\varepsilon \in (0, 1)$: the splitting ratio, satisfying that $\varepsilon_0 \in (\varepsilon, 1 - \varepsilon)$.
  * $\eta \in (0, \varepsilon)$: the trimming ratio.

Proposed Method

- **Methodology (Step 2: Projection)**:



SS-SN test statistic

- **Test statistic for the single change point** [Shao and Zhang (2010)]:
  - For $k = 1, 2, \ldots, N-1$, define the CUSUM statistic as

$$T_n(k/N) = N^{-1/2} \sum_{t=1}^{k} \left(Y_t - \bar{Y}_N\right), \text{ where } \bar{Y}_N = \frac{1}{N} \sum_{j=1}^{N} Y_j \qquad (1)$$

  - For $k = 1, 2, \ldots, N-1$, define the self-normalizer as

$$V_n(k/N) = N^{-2} \left(\sum_{t=1}^{k} \left(S_{1,t} - \frac{t}{k} S_{1,k} \right)^2 + \sum_{t=k+1}^{N} \left(S_{t,N} - \frac{N-t+1}{N-k} S_{+1,N} \right)^2 \right), \qquad (2)$$

  where $S_{a,b} = \sum_{j=a}^{b} Y_j$ denotes the cumulative sum.
  - The proposed test statistic is defined as

$$G_n = \sup_{k=1,\ldots,N-1} T_n(k/N) V_n^{-1/2}(k/N). \qquad (3)$$

Back

- **Comparison with Kim and Ramdas (2024)**

	Kim and Ramdas (2024)	Proposed
Data	**iid sequence** non-unique sample splits	**time series** unique sample split
Problem	**mean testing** one-sample problem	**change-point testing** two-sample problem
Techniques	t studentizer (Conditional) Berry-Esseen bound	self-normalizer new conditioning arguments
Agnostic Property	dimension	dimension cross-sectional dependence
Asymptotic size	Uniform	Pointwise

Theoretical Framework

- Three data generating processes:

Dimensionality	Data Generating Process	Temporal Dependence	Cross-sectional Dependence
fixed p	$X_t - \mu_t$ **stationary sequence (DGP1)** $\mathbb{E}[X_t] = \mu_t$, LRV Ω positive definite	weak	arbitrary
diverging p	**linear process (DGP2)** $X_t = \mu_t + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, $\{\varepsilon_t\}_{t\in\mathbb{Z}} \overset{iid}{\sim} (0, \Gamma)$	weak	weak
	static factor model (DGP3) $X_t = \mu_t + \Lambda F_t + Z_t$, $\Lambda \in \mathbb{R}^{p\times s}$, $s \ll p$ $\{F_t\}_{t=1}^n \in \mathbb{R}^s \sim (0, \Omega)$ $\{Z_t\}_{t=1}^n \in \mathbb{R}^p \sim (0, \Sigma)$ linear process $\{F_t\}_{t=1}^n \perp\!\!\!\perp \{Z_t\}_{t=1}^n$	weak	strong

Theory

- Limiting null distribution

p	$\{X_t\}_{t=1}^n$	Key assumptions	Limiting Null Distribution
Fixed p	Stationary (DGP1)	Functional CLT	$G_n \to^d \bar{G}$
$p \to \infty$	Linear process (DGP2)	$\|a_j\| \lesssim \rho^j$ for $\rho \in (0,1)$ and $j \geq 0$	
		$\rho^{m_2/4}\|\Gamma\|_F = o\left(\frac{n}{\log(n)}\right)$	
	Factor model (DGP3)	Functional CLT for $\{F_t\}$	
		$\rho^{m_2/4}\|\Gamma\|_F = o\left(\frac{n}{\log(n)}\right)$	

—- $\bar{G} := \sup_{r\in[0,1]} (B(r) - rB(1))\, V^{-1/2}(r)$ with $B(r)$ being the standard Brownian motion.

Quantile Level	90%	95%	97.5%	99%	99.5%	99.9%
Critical Value	4.32	5.39	6.38	7.58	8.49	10.40

—- $\log p = O(n)$ if $\Gamma = I_p$.

Theoretical Power Analysis

$$\mu_1 = \cdots = \mu_{k_0} = \mu,\ \mu_{k_0+1} = \cdots = \mu_n = \mu + \delta$$

- Theoretical asymptotic power against the local alternative
  - $\delta = \Delta/\sqrt{n}$, where $\Delta \in \mathbb{R}^p$.

	Stationary (DGP1)	Linear process (DGP2) (with $A^{(0)} = \sum_{\ell=0}^{\infty} a_\ell$)	Factor model (DGP3)
$\mathbb{P}(G_n > G_{1-\alpha}) \to \alpha$	$\|\Delta\|_2 \to 0$	$\frac{\|\Delta\|_2}{\|A^{(0)}\Gamma(A^{(0)})^\top\|_F^{1/2}} \to 0$	$\|\Delta\|_2 = o\left(\max\{\|\Lambda\|, \|\Gamma\|_F^{1/2}\}\right)$
$\mathbb{P}(G_n > G_{1-\alpha}) \to c_0$ $c_0 \in (\alpha, 1)$	$\|\Delta\|_2 \to c_1$ $c_1 \in (0, \infty)$	$\frac{\|\Delta\|_2}{\|A^{(0)}\Gamma(A^{(0)})^\top\|_F^{1/2}} \to c_2$ $c_2 \in (0, \infty)$	$\|\Delta\|_2 \sim \max\{\|\Lambda\|, \|\Gamma\|_F^{1/2}\}$
$\mathbb{P}(G_n > G_{1-\alpha}) \to 1$	$\|\Delta\|_2 \to \infty$	$\frac{\|\Delta\|_2}{\|A^{(0)}\Gamma(A^{(0)})^\top\|_F^{1/2}} \to \infty$	$\max\{\|\Lambda\|, \|\Gamma\|_F^{1/2}\} = o(\|\Delta\|_2)$

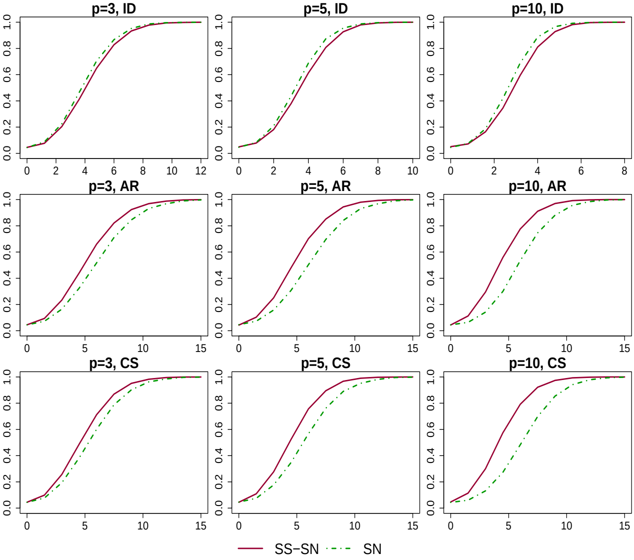
Power comparison

- Local asymptotic power exhibits a complex form, but numerical comparison possible with simulated local asymptotic power.
  - We generate

$$\begin{cases} X_1, \cdots, X_{k_0} \overset{iid}{\sim} \mathcal{N}_p(\mathbf{0}, \Omega), \\ X_{k_0+1}, \cdots, X_n \overset{iid}{\sim} \mathcal{N}_p(\delta, \Omega), \end{cases}$$

    where $n = 1000$, $k_0 = 500$, $p \in \{3, 5, 10\}$ and $\delta = c(1, \cdots, 1)^\top / \sqrt{n}$.
  - $\Omega = (\Omega_{ij}) \in \mathbb{R}^{p \times p}$ takes the following forms:
    - ID: $\Omega_{ij} = \mathbf{1}\{i = j\}$;
    - AR: $\Omega_{ij} = 0.8^{|i-j|}$;
    - CS: $\Omega_{ij} = 0.5 + 0.5\mathbf{1}\{i = j\}$.
  - **Method Comparison:**
    - SN test in Shao and Zhang (2010)
  - Significance level $\alpha = 0.05$
  - 5000 Monte-Carlo replicates

Power comparison



Power comparison

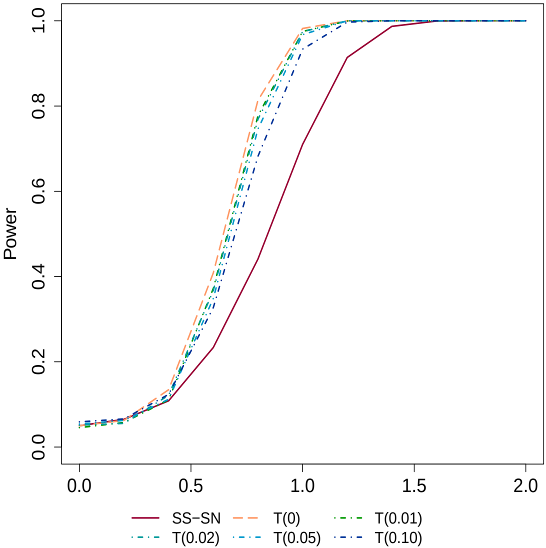
- **Local asymptotic power comparison under DGP2:**
  - We generate

$$\begin{cases} X_1, \cdots, X_{k_0} \overset{iid}{\sim} \mathcal{N}_p(\mathbf{0}, I_p), \\ X_{k_0+1}, \cdots, X_n \overset{iid}{\sim} \mathcal{N}_p(\delta, I_p), \end{cases}$$

    where $n = 1000$, $k_0 = 500$, $p = 1000$ and $\delta = c(1, \cdots, 1)^\top / \sqrt{n}$.
  - **Method Comparison:**
    - Trimmed U-statistic-based method $T(\eta_0)$ by Wang et al. (2022) with the trimming parameter $\eta_0 \in \{0, 0.01, 0.02, 0.05, 0.10\}$.
  - Significance level $\alpha = 0.05$
  - 2000 Monte-Carlo replicates

Power comparison

- **Local asymptotic power comparison under DGP2:**



Simulation Studies

- **DGP:** p-dimensional AR(1) process

$$X_t - \mu_t = \kappa(X_{t-1} - \mu_{t-1}) + \epsilon_t \in \mathbb{R}^p, \qquad 1 \le t \le n,$$

  where $\{\epsilon_t\} \overset{iid}{\sim} \mathcal{N}_p(0, \Sigma)$, $\kappa = 0.7$ and Σ takes the following forms:
  - AR ($\Sigma_{i,j} = 0.8^{|i-j|}$)
  - CS ($\Sigma_{i,j} = 0.5 + 0.5\mathbf{1}\{i = j\}$)
  - ID ($\Sigma_{i,j} = \mathbf{1}\{i = j\}$)
- **Method Comparison:**
  - test statistic targeting at dense alternatives: $G_{n,2}$ (splitting ratio $\varepsilon = 0.1$, trimming ratio $\eta = 0.04$).
  - Wang et al. (2022), denoted by $T_{SN}(\eta_0)$ where η_0 is a trimming parameter selected from $\{0, 0.01, 0.02, 0.05, 0.1\}$
- Significance level $\alpha = 0.05$, 5000 Monte-Carlo replicates
- **Dense change point:**
  - $k = \lfloor n/2 \rfloor + 1$ and $\mu_t = \begin{cases} (0, \ldots, 0)^\top, & 1 \le t < k \\ c(1, \ldots, 1)^\top/\sqrt{p}, & k \le t \le n \end{cases}$

Size

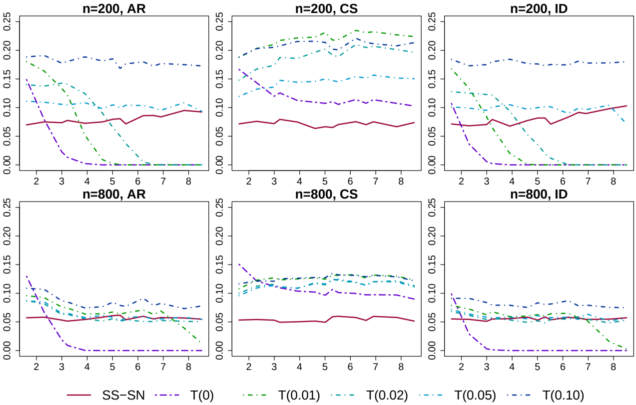


Figure: Empirical size curves versus the logarithm of p

Power

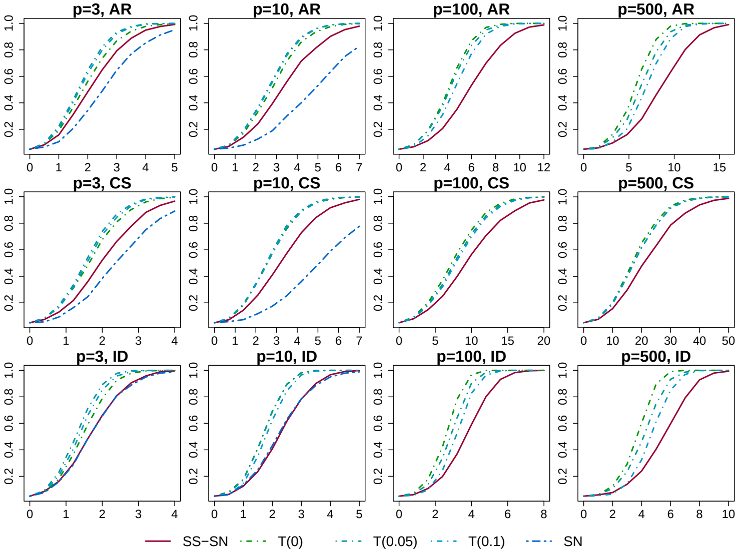


Figure: Size-adjusted Power Curves: single dense change

Generalization and Open Problems

- Single **dense** alternative $\rightarrow$ Single **sparse** alternative
  - * Use a sparse direction for projection
  - Theory is wide open
- **Single** change point $\rightarrow$ **Multiple** change points
  - * Use scanning-based tests by Zhang and Lavitas (2018)
  - Done (method, theory and simulations)
- Sensitivity with respect to η
- Comparison with approaches based on LRV consistent estimation

Summary

- Dimension-agnostic testing for (weakly dependent) time series is feasible and natural since sample splitting is unique for given splitting proportions.
- Size accuracy with respect to a range of dimensions; agnostic to the magnitude of cross-sectional dependence.
- Computational complexity $O(n(n+p))$
- Tradeoff between robustness and efficiency:
  - Robustness: dimension, cross-sectional and temporal dependence
  - Efficiency: power loss (true in high dimension, not necessary in low dimension)

Thank you very much!