Equivariant Variance Estimation for Multiple Change-point Model

Han Xiao
joint with Ning Hao and Yue Selena Niu

Department of Statistics
Rutgers University

Bayesian Statistics and Statistical Learning
IMSI, 12/12/2023, Chicago

Outline

Variance estimation under the presence of change points

- Consider a sequence of random variables $\boldsymbol{X} = (X_1, \cdots, X_n)^\top$ satisfying

$$X_i = \theta_i + \varepsilon_i, \qquad 1 \le i \le n.$$

  - The mean vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)^\top$ is piecewise constant:

$$\begin{aligned} \theta_1 = \cdots = \theta_{\tau_1} &= \mu_1 \\ \theta_{\tau_1+1} = \cdots = \theta_{\tau_2} &= \mu_2 \\ &\cdots \\ \theta_{\tau_J+1} = \cdots = \theta_n &= \mu_{J+1} \end{aligned}$$

  - $\boldsymbol{\tau} = (\tau_1, ..., \tau_J)^\top$ is the location vector of change points.
  - The noises $\{\varepsilon_i\}$ are i.i.d. with $\mathbb{E}(\varepsilon_1) = 0$ and $\mathrm{Var}(\varepsilon_1) = \sigma^2$.
- Goal: estimate σ^2 WITHOUT knowing the change points.
- Most change point detection procedures requires an estimate of σ^2. E.g. SARA (Niu and Zhang 2012), SMUCE (Frick et al 2014), Wild Binary Segmentation (Fryzlewicz 2014).

Some Existing Estimators

- Median absolute deviation (MAD) estimator (Hampel 1974, Fryzlewicz 2014).

$$\hat{\sigma}_1 = 1.48 * \mathrm{med}(|\boldsymbol{X} - \mathrm{med}(\boldsymbol{X})|).$$

- Davies and Kovac (2001) and Frick et al (2014).

$$\hat{\sigma}_2 = \frac{1.48}{\sqrt{2}} * \mathrm{med}(|\boldsymbol{X}_{(-1)} - \boldsymbol{X}_{(-n)}|).$$

  - $\boldsymbol{X}_{(-1)} = (X_2, X_3, \ldots, X_n)$, $\boldsymbol{X}_{(-n)} = (X_1, X_2, \ldots, X_{n-1})$.

- The Rice estimator (Rice 1984)

$$\hat{\sigma}_3^2 = \frac{1}{2n} \|\boldsymbol{X}_{(-1)} - \boldsymbol{X}_{(-n)}\|^2 = \frac{S_1}{2n}$$

- Motivations: robustness, unbiasedness, efficiency.

Difference Based Estimator in Nonparametric Regression

- The difference based variance estimators has been studied in nonparametric regression: Rice (1984), Gasser et al (1986), Müller and Stadtmüller (1987), Hall et al (1990), etc.
- Müller and Stadtmüller (1999) innovatively built a variance estimator by regressing the lag-k Rice estimators on the lags.
- Recent developments include Tong et al (2013), Tecuapetla-Gómez and Munk (2017) and Levine and Tecuapetla-Gómez (2019).
- Our contribution: exact and non-asymptotic results regarding the unbiasedness and the minimax risk, under minimal conditions.

Outline

Rice Estimators and Beyond

- The Rice estimator is $S_1/(2n)$, where

$$S_1 = \sum_{i=1}^{n-1} (X_i - X_{i+1})^2.$$

- Let

$$V(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} (\theta_i - \theta_{i+1})^2 = \sum_{j=1}^{J} (\mu_j - \mu_{j+1})^2.$$

- Let $L(\boldsymbol{\theta})$ be the minimum segment length.
- The Rice estimator is biased since

$$\mathbb{E}[S_1/(2n)] = \sigma^2 + [V(\boldsymbol{\theta}) - 2\sigma^2]/(2n).$$

Rice Estimators and Beyond

- Consider the lag-k Rice estimators

$$S_k = \sum_{i=1}^{n-k} (X_i - X_{i+k})^2, \quad k \geq 1.$$

- Let $\kappa_4 := \mathbb{E}\varepsilon_1^4/\sigma^4$. If $\mathbb{E}\varepsilon_1^3 = 0$, then for $1 \leq k \leq h \leq L(\boldsymbol{\theta})/2$,

$$\begin{aligned} \mathbb{E}S_k &= 2n\sigma^2 + k\left[V(\boldsymbol{\theta}) - 2\sigma^2\right] \\ \mathrm{Cov}(S_k, S_h) &= (4n - 4h - 2k)(\kappa_4 - 1)\sigma^4 + 4(n-k)\sigma^4\delta_{k,h} + 8k\sigma^2 V(\boldsymbol{\theta}). \end{aligned}$$

- Müller and Stadtmüller (1999) built an estimator based on multiple S_k, by regressing S_k on k to remove the bias.
  - The formulas for the variance (of the variance estimator) is complicated.
  - The minimax theory is more cumbersome.

The Model Class

- Embed the index set $[n] = \{1, ..., n\}$ into the unit circle $\mathcal{S}^1 \subset \mathbb{R}^2$ by the exponential map $\pi_n : i \mapsto e^{\frac{2\pi i \sqrt{-1}}{n}}$.
- A segment $[k, \ell]$ with $k \geq \ell$ is also well-defined. E.g., $[n-1, 3] = \{n-1, n, 1, 2, 3\}$.
- Assume $\boldsymbol{\theta}$ consists of J segments with constant means, $[\tau_1 + 1, \tau_2]$,..., $[\tau_J + 1, \tau_1]$, which are separated by the change points $1 \leq \tau_1 < \tau_2 < \cdots < \tau_J \leq n$.
- Denote by $L(\boldsymbol{\theta})$ the minimal length of all constant segments in $\boldsymbol{\theta}$.
- Consider a family of nested model classes $\Theta_2 \supset \Theta_3 \supset \cdots$, where

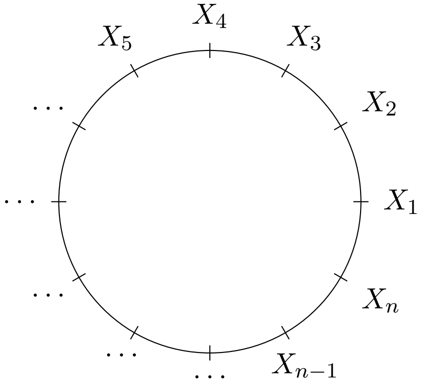
$$\Theta_L = \{\boldsymbol{\theta} \in \mathbb{R}^n : \ L(\boldsymbol{\theta}) \geq L\}.$$

- The classical model classes can be defined by

$$\Theta_L^c = \{\boldsymbol{\theta} \in \mathbb{R}^n : \ L(\boldsymbol{\theta}) \geq L, \ \tau_{J+1} = n\}.$$

- $\Theta_L \supset \Theta_L^c$.

The Model Class



Equivariant Variance Estimator

- With the convention that $\theta_i = \theta_{n+i}$ and $\mu_{J+1} = \mu_J$, define

$$W(\boldsymbol{\theta}) = \sum_{i=1}^{n} (\theta_i - \theta_{i+1})^2 = \sum_{j=1}^{J} (\mu_j - \mu_{j+1})^2.$$

- Our estimator is based on the lag-k Rice estimators (same convention that $X_i = X_{n+i}$)

$$T_k = \sum_{i=1}^{n} (X_i - X_{i+k})^2, \quad 1 \le k \le K.$$

- For $1 \le k \le L(\boldsymbol{\theta})$,

$$\mathbb{E} T_k = 2n\sigma^2 + kW(\boldsymbol{\theta}).$$

Equivariant Variance Estimator: OLS

- $T_k = 2n\sigma^2 + kW(\boldsymbol{\theta}) + (T_k - \mathbb{E}T_k)$.
- Let $Y_k = T_k/(2n)$, $(\alpha, \beta)^\top = (\sigma^2, W(\boldsymbol{\theta})/(2n))^\top$, then

$$\begin{aligned} Y_k &= \sigma^2 + k\frac{W(\boldsymbol{\theta})}{2n} + (Y_k - \mathbb{E}Y_k), \quad k = 1, ..., K; \\ Y_k &= \alpha + k\beta + e_k, \quad k = 1, ..., K. \end{aligned}$$

- The ordinary least squares (OLS) estimator of σ^2 is given by

$$\hat{\alpha}_K = (1, 0)(\boldsymbol{Z}_K^\top \boldsymbol{Z}_K)^{-1}\boldsymbol{Z}_K^\top \boldsymbol{Y}_K,$$

  - $\boldsymbol{Y}_K = (Y_1, \ldots, Y_k)^\top$, $\boldsymbol{\eta}_K = (1, 2, \ldots, K)^\top$, $\boldsymbol{Z}_K = (\mathbf{1}_K, \boldsymbol{\eta}_K)$.
- $\hat{\alpha}_K$ is unbiased when $2 \leq K \leq L(\boldsymbol{\theta})$.

Variance of the OLS Estimator

- Let κ_4 be the kurtosis of ε_i, it holds that $1 \leq k \leq h \leq L(\boldsymbol{\theta})/2$,

$$\mathrm{Cov}(T_k, T_h) = 4n(\kappa_4 - 1)\sigma^4 + 4n\sigma^4\delta_{k,h} + 8k\sigma^2 W(\boldsymbol{\theta}).$$

Theorem

If $K \leq L(\boldsymbol{\theta})/2$,

$$\mathrm{Var}(\hat{\alpha}_K) = \frac{\sigma^4}{n}\left(\kappa_4 - 1 + \frac{4K+2}{K(K-1)} + \frac{2W(\boldsymbol{\theta})}{n\sigma^2}\frac{(K+1)(K+2)(2K+1)}{15K(K-1)}\right).$$

If $K \leq L(\boldsymbol{\theta})$,

$$\mathrm{Var}(\hat{\alpha}_K) \leq \frac{\sigma^4}{n}\left(\kappa_4 - 1 + \frac{4K+2}{K(K-1)} + \frac{W(\boldsymbol{\theta})}{n\sigma^2}\frac{(K+1)(K+2)^2}{3K(K-1)}\right).$$

Another Look at the Covariance

- The covariance matrix of $(e_1, \ldots, e_K)^\top$ is

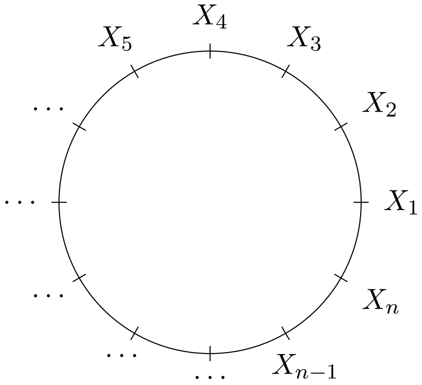
$$\frac{\sigma^4}{n}\left[\boldsymbol{I}_K + (\kappa_4 - 1)\mathbf{1}_K\mathbf{1}_K^\top + \frac{2W(\boldsymbol{\theta})}{n\sigma^2}\boldsymbol{H}_K\right],$$

  - $\boldsymbol{I}_K$ is the $K \times K$ identity matrix,
  - $\mathbf{1}_K$ is a vector of length K with all entries equal to 1,
  - $\boldsymbol{H}_K = (H_{ij})$ is a $K \times K$ matrix with $H_{ij} = \min\{i, j\}$.
- The OLS $\hat{\alpha}_K$ is the BLUE if $W(\boldsymbol{\theta}) = 0$.

Equivariance

- Embed the index set $[n] = \{1, ..., n\}$ into the unit circle $\mathcal{S}^1 \subset \mathbb{R}^2$ by the exponential map $\pi_n : i \mapsto e^{\frac{2\pi i\sqrt{-1}}{n}}$.
- Rotating the indexes by the angle $2\pi/n$ maps the n-vector $\boldsymbol{X} = (X_1, ..., X_n)^\top$ to $(X_n, X_1, X_2, ..., X_{n-1})^\top$.
- This rotation can be represented using a permutation matrix $\boldsymbol{C}$ as: $(X_n, X_1, X_2, ..., X_{n-1})^\top = \boldsymbol{C}\boldsymbol{X}$.
- A statistic $T(\boldsymbol{X})$ is said to be equivariant if $T(\boldsymbol{C}^m\boldsymbol{X}) = T(\boldsymbol{X})$ for any integer m.
- Each T_k is equivariant, and so is the OLS $\hat{\alpha}_K$.

Equivariance



Equivariance

- Define $\boldsymbol{C}_k$ as a circulant matrix with its (i, j) entry

$$C_{k,ij} = \begin{cases} 1, & j - i = k \mod n \\ 0, & \text{otherwise.} \end{cases}$$

- Treat the subscript k in $\boldsymbol{C}_k$ as a number modulo n.
  - $\boldsymbol{C}_k \boldsymbol{C}_\ell = \boldsymbol{C}_{k+\ell}$,
  - $\boldsymbol{C}_k^\top = \boldsymbol{C}_{-k} = \boldsymbol{C}_{n-k}$,
- $\mathcal{C}_n = \{\boldsymbol{C}_k,\ 1 \leq k \leq n\}$ is a group isomorphic to the order-n cyclic group.
- It gives a group action $\mathcal{C}_n \hookrightarrow \mathbb{R}^n$: $\boldsymbol{X} \mapsto \boldsymbol{C}_k \boldsymbol{X}$.
- A variance estimator $\hat{\sigma}^2$ is equivariant if $\hat{\sigma}^2(\boldsymbol{X}) = \hat{\sigma}^2(\boldsymbol{C}_k \boldsymbol{X})$.

Equivariant Unbiased Quadratic Estimators

- Consider the class of quadratic estimators $\sum_{i,j=1}^{n} a_{ij} X_i X_j$, or $\boldsymbol{X}^\top \boldsymbol{A} \boldsymbol{X}$, where $\boldsymbol{A} = (a_{ij})$ is a symmetric matrix.
- $Y_k = \frac{1}{2n} T_k = \boldsymbol{X}^\top \boldsymbol{A}_k \boldsymbol{X}$ with $\boldsymbol{A}_k = \frac{1}{n}\left(\boldsymbol{I} - \frac{1}{2}\boldsymbol{C}_k - \frac{1}{2}\boldsymbol{C}_k^\top\right)$.

Theorem

The set of all equivariant unbiased quadratic variance estimators for the model class Θ_L is

$$\mathcal{Q}_L = \left\{ \frac{1}{2n} \sum_{k=1}^{L} c_k T_k = \sum_{k=1}^{L} c_k Y_k \; : \; c_1, ..., c_L \in \mathbb{R}, \sum_{k=1}^{L} c_k = 1, \sum_{k=1}^{L} k c_k = 0 \right\}.$$

- Such an $\boldsymbol{A}$ is not positive definite.

Outline

The Model Class

- Model class, $L \geq 2$, $w \geq 0$.

$$\Theta_{L,w} = \{(\boldsymbol{\theta}, \sigma^2) : \; L(\boldsymbol{\theta}) \geq L, \, W(\boldsymbol{\theta})/(n\sigma^2) \leq w, \, \sigma^2 > 0\}.$$

  - We will also consider $\Theta_{2L,w} \subset \Theta_{L,w}$.
- For any estimator $\hat{\sigma}^2$, define the ℓ_2 risk up to a factor $\frac{\sigma^4}{n}$

$$r(\hat{\sigma}^2) = \frac{n}{\sigma^4} \mathbb{E}(\hat{\sigma}^2 - \sigma^2)^2.$$

- The minimax estimator and the exact risk are difficult to find for $L \geq 3$. We focus on lower and upper bounds of the minimax risk.

The GLS Estimator

- Recall that

$$\boldsymbol{\Sigma}_{L,w} = \frac{\sigma^4}{n}\left[\boldsymbol{I}_L + (\kappa_4 - 1)\mathbf{1}_L\mathbf{1}_L^\top + 2w\boldsymbol{H}_L\right].$$

  - $\boldsymbol{H}_L = (H_{ij})$ is a $L \times L$ matrix with $H_{ij} = \min\{i, j\}$.

- The generalized least squares (GLS) estimator $\tilde{\alpha}_{L,w}$ is given by

$$\tilde{\alpha}_{L,w} = (1, 0)(\boldsymbol{Z}_L^\top \boldsymbol{\Sigma}_{L,w}^{-1}\boldsymbol{Z}_L)^{-1}\boldsymbol{Z}_L^\top \boldsymbol{\Sigma}_{L,w}^{-1}\boldsymbol{Y}_L.$$

Theorem

On $\Theta_{2L,w}$, the GLS estimator $\tilde{\alpha}_{L,w} \in \mathcal{Q}_L$ is minimax with the risk

$$\min_{\hat{\sigma}^2 \in \mathcal{Q}_L} \max_{(\boldsymbol{\theta},\sigma^2)\in\Theta_{2L,w}} r(\hat{\sigma}^2) = \max_{(\boldsymbol{\theta},\sigma^2)\in\Theta_{2L,w}} r(\tilde{\alpha}_{L,w}) = g_L(2w).$$

- $g_L(\cdot)$ is a nonnegative increasing function on $[0, \infty)$ with

$$g_L(0) = \kappa_4 - 1 + \frac{4L+2}{L(L-1)}, \quad \text{and} \quad g_L'(0) = \frac{(L+1)(L+2)(2L+1)}{15L(L-1)}.$$

Lower Bound

- Let $\{D_k\}$ be the sequence defined recursively by $D_k = (2+\lambda)D_{k-1} - D_{k-2}$ with initial values $D_0 = 1,\ D_1 = 1+\lambda$.
- Define the matrix

$$\boldsymbol{V}_{L,\lambda} := \begin{pmatrix} \frac{1-D_{L-1}/D_L}{\lambda} & \frac{D_L-1}{\lambda D_L} \\ \frac{D_L-1}{\lambda D_L} & \frac{D_{L-1}/D_L+\lambda L-1}{\lambda^2} \end{pmatrix}.$$

- Define

$$g_L(\lambda) := \kappa_4 - 1 + \boldsymbol{V}_{L,\lambda}^{-1}[1,1].$$

  - $\boldsymbol{V}_{L,\lambda}^{-1}[1,1]$ is the top left entry of the 2×2 matrix $\boldsymbol{V}_{L,\lambda}^{-1}$.
- Lower Bound. $\min_{\hat{\sigma}^2 \in \mathcal{Q}_L} \max_{(\boldsymbol{\theta},\sigma^2)\in\Theta_{L,w}} r(\hat{\sigma}^2) \geq g_L(2w)$.
- On the class $\Theta_{L,w}$, the risk of an estimator in $\mathcal{Q}_L$ can depend on $\boldsymbol{\theta}$ NOT just through $W(\boldsymbol{\theta})$.
- Open question. What is the minimax estimator over $\Theta_{L,w}$?

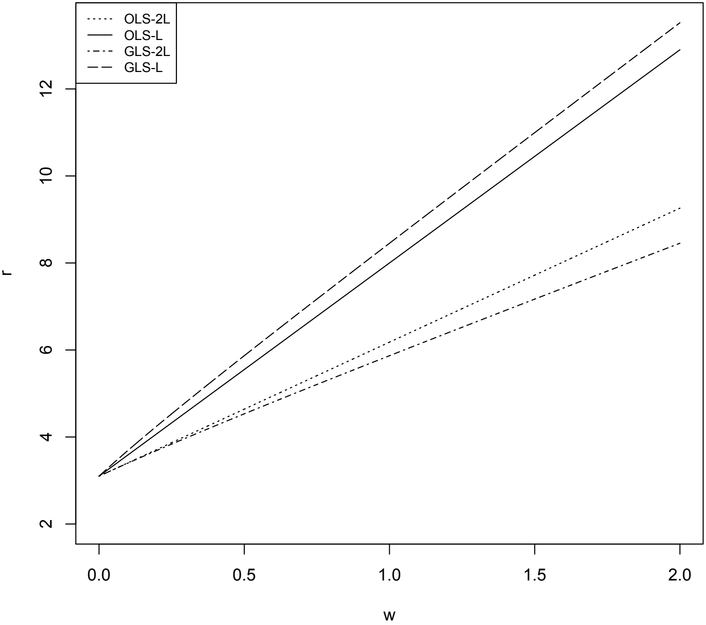
- First upper bound.

$$\min_{\hat{\sigma}^2 \in \mathcal{Q}_L} \max_{(\boldsymbol{\theta},\sigma^2)\in\Theta_{L,w}} r(\hat{\sigma}^2) \leq \max_{(\boldsymbol{\theta},\sigma^2)\in\Theta_{L,w}} r(\hat{\alpha}_L)$$
$$= \kappa_4 - 1 + \frac{4L+2}{L(L-1)} + \frac{(L+1)(L+2)^2}{3L(L-1)} w.$$

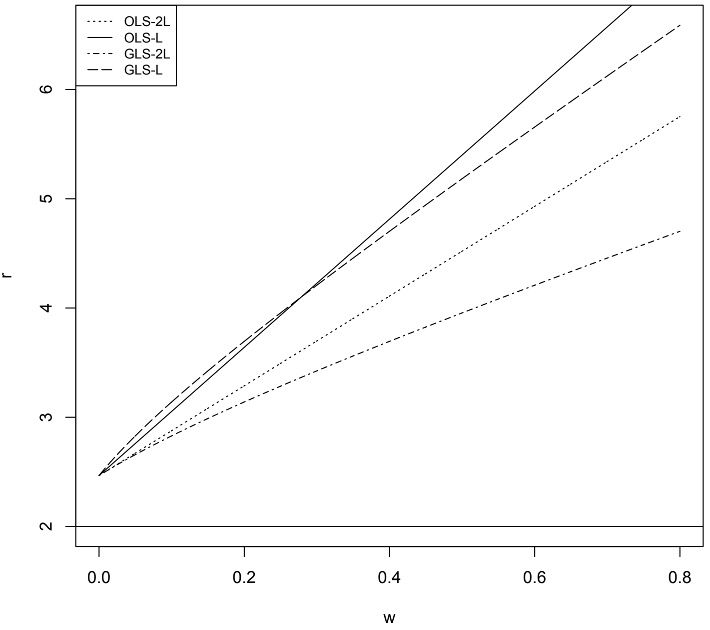
Theorem (Second Upper Bound)

$$\min_{\hat{\sigma}^2 \in \mathcal{Q}_L} \max_{(\boldsymbol{\theta},\sigma^2)\in\Theta_{L,w}} r(\hat{\sigma}^2) \leq g_L(4w)$$

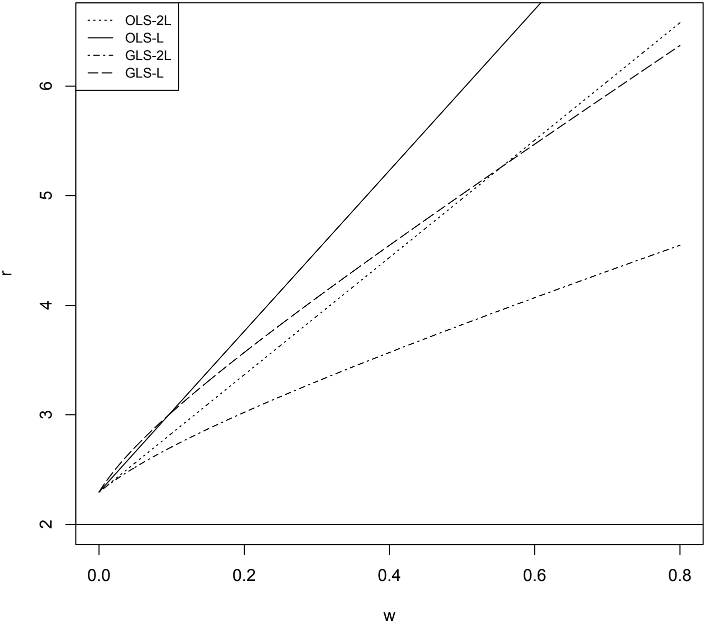
Minimax Bounds: $L = 5$



Minimax Bounds: $L = 10$



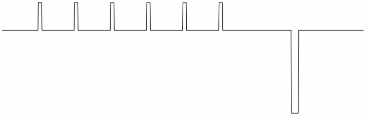
Minimax Bounds: $L = 15$



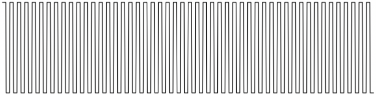
Outline

Simulation Setup

- Scenario 1: $\boldsymbol{\theta} = \mathbf{0}$.
- Scenario 2: $\theta_i = 1$ when $100m + 1 \leq i \leq 100m + 10$, $m \in \{1, 2, ..., 6\}$; $\theta_i = -3$ when $801 \leq i \leq 820$, and $\theta_i = 0$ otherwise. $L(\boldsymbol{\theta}) = 10$.



- Scenario 3: $\theta_i = 1$ when $20m + 1 \leq i \leq 20m + 10$, $m \in \{0, 1, ..., 49\}$, and $\theta_i = -1$ otherwise. $L(\boldsymbol{\theta}) = 10$.



- The noise ε_i follow: normal, standardized t_3, and centered standard exponential distributions.
- Report simulation results on the estimated standard deviations.

The Choice of K

- Given a range of K, say $K_{\min} = 5 \le K \le K_{\max} = 20$, we calculate Y_1,..., $Y_{K_{\max}+1}$ and use Y_1,..., Y_K to predict Y_{K+1} based on the linear model.
- Calculate a score defined by $SC(K) = |\hat{Y}_{K+1} - Y_{K+1}|/\hat{\sigma}_e$, where $\hat{\sigma}_e$ is estimated based on the RSS.
- The $\hat{K}$ is selected by

$$\hat{K} = \underset{\{K_{\min} \le K \le K_{\max}\}}{\text{argmax}} SC(K).$$

- This tuning process chooses $K = 10$ with high empirical probabilities (96.8%, 96.0%, and 95.2%) in S3-G, S3-T, and S3-E, respectively.

Choice of K

Table: Average values of estimators with standard errors in parenthesis over 500 replicates.

	K=5	K=10	K=15	K=20	tuned
S1-G	0.999(0.029)	1.000(0.026)	1.000(0.025)	1.000(0.024)	0.999(0.028)
S1-T	0.999(0.039)	0.999(0.037)	0.999(0.036)	0.999(0.035)	0.999(0.038)
S1-E	0.998(0.048)	0.998(0.046)	0.998(0.046)	0.998(0.046)	0.998(0.047)
S2-G	1.000(0.029)	1.000(0.026)	1.004(0.026)	1.009(0.025)	1.000(0.028)
S2-T	0.999(0.039)	0.999(0.037)	1.003(0.036)	1.008(0.035)	1.000(0.038)
S2-E	0.998(0.049)	0.998(0.046)	1.003(0.046)	1.007(0.046)	0.999(0.047)
S3-G	1.000(0.034)	1.000(0.030)	1.253(0.026)	1.468(0.031)	1.001(0.030)
S3-T	0.999(0.043)	0.999(0.040)	1.254(0.033)	1.469(0.035)	1.000(0.041)
S3-E	0.998(0.052)	0.998(0.049)	1.252(0.041)	1.467(0.041)	0.999(0.049)

Estimators for Comparison

- EVE $\hat{\alpha}_K$ with $K = 10$.
- Median absolute deviation (MAD) estimator (Hampel1974)

$$\hat{\sigma}_1 = 1.4826 * \mathrm{med}(|\boldsymbol{X} - \mathrm{med}(\boldsymbol{X})|).$$

- DK Estimator (Davies and Kovac 2001)

$$\hat{\sigma}_2 = \frac{1.48}{\sqrt{2}} * \mathrm{med}(|\boldsymbol{X}_{(-1)} - \boldsymbol{X}_{(-n)}|).$$

- Rice estimator (Rice 1984)

$$\hat{\sigma}_3^2 = \frac{1}{2n} \|\boldsymbol{X}_{(-1)} - \boldsymbol{X}_{(-n)}\|^2 = \frac{S_1}{2n}$$

Comparison: Accuracy

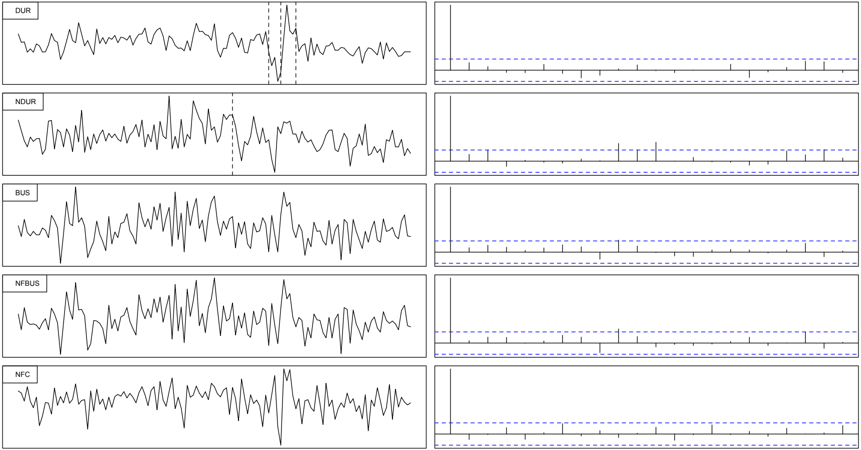
Table: Average values of estimators with standard errors in parenthesis over 500 replicates.

	EVE	MAD	DK	Rice	Orcale
S1-G	1.000(0.026)	1.001(0.040)	1.001(0.041)	0.999(0.028)	1.000(0.023)
S1-T	0.999(0.037)	0.867(0.036)	0.916(0.038)	0.999(0.039)	1.000(0.034)
S1-E	0.998(0.046)	0.714(0.033)	0.727(0.038)	0.998(0.048)	0.998(0.046)
S2-G	1.000(0.026)	1.049(0.042)	1.005(0.041)	1.007(0.028)	1.000(0.023)
S2-T	0.999(0.037)	0.921(0.036)	0.921(0.039)	1.006(0.039)	1.000(0.034)
S2-E	0.998(0.046)	0.781(0.034)	0.735(0.038)	1.006(0.048)	0.998(0.046)
S3-G	1.000(0.030)	1.557(0.052)	1.071(0.043)	1.094(0.028)	1.000(0.023)
S3-T	0.999(0.040)	1.556(0.046)	0.994(0.041)	1.095(0.038)	1.000(0.034)
S3-E	0.998(0.049)	1.575(0.066)	0.821(0.043)	1.094(0.046)	0.998(0.046)

Example: Labor Productivity

- Consider the variance estimation of the U.S. labor productivity (`https://www.bls.gov/lpc/`, quarterly growth rates in percentages, 1987 Q1 – 2019 Q4) of major sectors:
  - manufacturing/durable (DUR)
  - manufacturing/nondurable (NDUR)
  - business (BUS)
  - nonfarm business (NFBUS)
  - nonfinancial corporations (NFC)
- Compare the following estimators:
  - MAD, DK, Rice.
  - SD: sample standard deviation.
  - SD_s: sample standard deviation of the segmented series.

Example: Labor Productivity



Example: Labor Productivity

Table: Variance estimation for the US labor productivity indices.

	$\mathrm{SD_s}$	EVE	MAD	DK	Rice	SD
DUR	3.82	3.61	5.49	3.40	3.80	5.20
NDUR	3.59	3.49	3.71	3.30	3.39	3.81
BUS	2.59	2.49	2.37	2.41	2.50	2.59
NFBUS	2.60	2.54	2.37	2.62	2.55	2.60
NFC	3.62	3.60	3.11	3.40	3.76	3.62

- SD might overestimate σ for DUR and NDUR as it ignores the potential change points.
- DK often underestimates σ possibly due to non-Gaussian noise distribution.
- The MAD estimator seems to be unstable, with larger biases.
- Overall, the EVE is very close to the benchmark $\mathrm{SD_s}$, but without segmenting the series first.

Conclusion

- We consider the equivariant variance estimation for multiple change-point model, based on the equivariant lag-k Rice estimators.
- No distributional assumption.
- Minimax lower and upper bounds.
- Rule of Thumb: the number of Rice estimators K does NOT need to be very large.
- Future.
  - Covariance matrix estimation.
  - Test of independence.
  - Time series: autocovariance estimation.
  - Time series: autocorrelation tests.

Thank You!

MD **Mathias Drton**
Technical University of Munich

H **Jonathan Hauenstein**
University of Notre Dame

LL **Lek-Heng Lim**
University of Chicago

DP **Debdeep Pati**
Texas A&M University

MD **Mathias Drton**
Technical University of Munich

EG **Elizabeth Gross**
University of Hawaiʻi at Mānoa

LL **Lek-Heng Lim**
University of Chicago

SP **Sonja Petrović**
Illinois Institute of Technology

ER **Elina Robeva**
University of British Columbia

JR **Jose Rodriguez**
University of Wisconsin – Madison

BS **Bernd Sturmfels**
MPI Leipzig & UC Berkeley

PZ **Piotr Zwiernik**
University of Toronto & Pompeu Fabra University