On Robust Estimation of Multiple Change Points in Multivariate Processes

Yana Melnykov

Joint work with M. Perry and V. Melnykov

October 5, 2017

Talk outline

- Introduction and notation
- Motivation
- Methodology
- Experiments on simulated data
- Applications
- Discussion and future work

Introduction and notation

- Objective of change point estimation: identifying changes in processes observed over given time interval
- Simplest change point setting:
  - data observed over T time points
  - $K+1$ processes: $(0), (1), \ldots, (K)$
  - K change points at times $t_1, t_2, \ldots, t_K$
  - dataset

$$\boldsymbol{Y}_1^{(0)}, \boldsymbol{Y}_2^{(0)}, \ldots, \boldsymbol{Y}_{t_1-1}^{(0)}, \boldsymbol{Y}_{t_1}^{(1)}, \boldsymbol{Y}_{t_1+1}^{(1)}, \ldots, \boldsymbol{Y}_{t_K-1}^{(K-1)}, \boldsymbol{Y}_{t_K}^{(K)}, \boldsymbol{Y}_{t_K+1}^{(K)}, \ldots, \boldsymbol{Y}_T^{(K)}$$

Motivation

- The majority of methods
  - focus of univariate settings
  - assume processes following normal distribution
    - or p-variate normal distribution in multivariate setting
  - assume independence of observations observed over time
  - focus on single subject
    - or assume independence of subjects in multisubject studies
- We aim at developing procedures that
  - relax distributional assumptions (robust to deviations from normality)
  - are capable to handle multivariate processes
  - take into account dependence among observations
  - take into account dependence among variables
  - take into account dependence among subjects in multisubject studies
  - model joint covariance structure for variables, subjects, and observations

Matrix normal distribution

- Consider a generalization of p-variate normal to $p \times T$ matrix normal distribution (see, *e.g.*, Gupta and Nagar (2000)) with pdf

$$\phi_{p\times T}(\boldsymbol{Y};\boldsymbol{M},\boldsymbol{\Sigma},\boldsymbol{\Psi}) = (2\pi)^{-\frac{pT}{2}}|\boldsymbol{\Sigma}|^{-\frac{T}{2}}|\boldsymbol{\Psi}|^{-\frac{p}{2}} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{M})\boldsymbol{\Psi}^{-1}(\boldsymbol{Y}-\boldsymbol{M})^{\top}\right\}\right\}$$

  - $\boldsymbol{Y}$ is $p \times T$ matrix of data
  - $\boldsymbol{M}$ is $p \times T$ mean matrix
  - $\boldsymbol{\Sigma}_{p\times p}$ and $\boldsymbol{\Psi}_{T\times T}$ are covariance matrices associated with rows and columns, respectively
  - $\mathrm{tr}\{\cdot\}$ represents the trace operator
- Matrix normal distribution is effective in modeling variability associated with rows and columns

Matrix normal distribution

- It can be shown that $\text{vec}(\boldsymbol{Y}) \sim \phi_{pT}(\text{vec}(\boldsymbol{M}), \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma})$
  - *i.e.*, rows and columns are not assumed independent
- Taking into account matrix structure of data allows reducing the number of parameters associated with covariances from $pT(pT+1)/2$ to $T(T+1)/2 + p(p+1)/2$
- Note that $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma} = a\boldsymbol{\Psi} \otimes a^{-1}\boldsymbol{\Sigma}$, for any $a \in (0, \infty)$
  - to avoid non-identifiability, a constraint should be employed

Dealing with deviations from normality

- Data are often skewed – normality assumption is violated
- Possible remedies include
  - tranforming data to near-normality
    (*e.g.*, Box-Cox power or Manly exponential transformations)
  - employing more appropriate models
    (*e.g.*, skew-normal, log-normal, gamma, *etc.*)
- Manly transformation (1986) has several advantages over Box-Cox
  - not restricted to positive numbers
  - flexible for modeling left and right skewness
  - $\mathcal{M}(y;\lambda) = \lambda^{-1}(\exp(\lambda y) - 1)I(\lambda \neq 0) + yI(\lambda = 0)$
    - y: original observation
    - $\mathcal{M}(\cdot;\lambda)$: transformation operator with parameter λ
    - $I(\mathcal{A})$: indicator function that returns 1 if $\mathcal{A}$ is true and yields 0 otherwise

Multivariate Manly transformation

- For p-variate vector $\boldsymbol{y}$, it is commonly assumed that coordinatewise transformation leads to joint near-normality (*e.g.*, Andrews et al (1971), Velilla (1993))
  - $\mathcal{M}(\boldsymbol{y};\boldsymbol{\lambda}) = (\mathcal{M}(y_1;\lambda_1), \mathcal{M}(y_2;\lambda_2), \ldots, \mathcal{M}(y_p;\lambda_p))^\top \sim Normal(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
    - $\boldsymbol{y} = (y_1, y_2, \ldots, y_p)^\top$
    - $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_p)^\top$
- This transformation idea can be generalized to matrices
  - $\mathcal{M}(\boldsymbol{Y}, \boldsymbol{\Lambda}) \sim MatrixNormal(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$
  - specific form of $\boldsymbol{\Lambda}$ depends on the problem
    - if rows represent p variables and columns correspond to T times, each variable can have its own skewness parameter, *i.e.*, $\boldsymbol{\Lambda} = \boldsymbol{\lambda}_p \mathbf{1}_T^\top$
    - if rows represent N subjects and columns correspond to T times, common parameter λ need to be applied to all rows, *i.e.*, $\boldsymbol{\Lambda} = \lambda \mathbf{1}_N \mathbf{1}_T^\top$

Multivariate processes

- p variables, T times
- Log-likelihood function:

$$\begin{aligned}\log \mathcal{L}(\boldsymbol{Y}; \boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \boldsymbol{\lambda}) = &-\frac{pT}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}| - \frac{p}{2}\log|\boldsymbol{\Psi}| \\ &- \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}(\mathcal{M}(\boldsymbol{Y};\boldsymbol{\lambda}) - \boldsymbol{M})\boldsymbol{\Psi}^{-1}(\mathcal{M}(\boldsymbol{Y};\boldsymbol{\lambda}) - \boldsymbol{M})^\top\right\} \\ &+ \boldsymbol{\lambda}^\top \boldsymbol{Y}\mathbf{1}_T\end{aligned}$$

  - $\boldsymbol{\lambda}^\top \boldsymbol{Y}\mathbf{1}_T$ is the log of Jacobian corresponding to transformation

Modeling mean matrix

- Note that mean matrix $\boldsymbol{M}$ has pT parameters
- In case of K shift change points

$$\boldsymbol{M} = \Big(\underbrace{\boldsymbol{\mu}_0, \boldsymbol{\mu}_0, \ldots, \boldsymbol{\mu}_0, \boldsymbol{\mu}_1}_{t_1}, \underbrace{\boldsymbol{\mu}_1 \ldots, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2}_{t_2 - t_1}, \ldots,$$

$$\underbrace{\boldsymbol{\mu}_{K-1}, \ldots, \boldsymbol{\mu}_{K-1}, \boldsymbol{\mu}_K}_{t_K - t_{K-1}}, \underbrace{\boldsymbol{\mu}_K, \ldots, \boldsymbol{\mu}_K}_{T - t_K} \Big) = \sum_{k=0}^{K} \boldsymbol{\mu}_k \boldsymbol{m}_k^\top$$

  - $\boldsymbol{m}_k$ is vector of 0's and 1's, with 1's in positions corresponding to the k^{th} process

Modeling covariance matrices

- Choice of $\mathbf{\Psi}$ depends on specific problem
- Matrix $\mathbf{\Psi}$ corresponding to AR_1 process is given by

$$\mathbf{\Psi} = \frac{\delta^2}{1-\phi^2}\boldsymbol{R}_\phi \quad \text{with} \quad \boldsymbol{R}_\phi = \begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & 1 & \phi & \dots & \phi^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \dots & 1 \end{pmatrix}$$

  - δ^2 and ϕ are corresponding variance and AR_1 parameters
  - $\boldsymbol{R}_\phi$ represents the correlation matrix associated with AR_1
- Recall non-identifiability issue due to $a\mathbf{\Psi} \otimes a^{-1}\mathbf{\Sigma} = \mathbf{\Psi} \otimes \mathbf{\Sigma}$
- One convenient constraint in the considered setting is $\delta^2 = 1 - \phi^2$
  - $\mathbf{\Psi}$ reduces to correlation matrix $\boldsymbol{R}_\phi$

Modeling covariance matrices

- It can be shown that

$$|\boldsymbol{\Psi}| = |\boldsymbol{R}_\phi| = (1-\phi^2)^{T-1}$$

$$\boldsymbol{\Psi}^{-1} = \boldsymbol{R}_\phi^{-1} = \frac{1}{1-\phi^2}(\boldsymbol{I}_T - \phi \boldsymbol{J}_1 + \phi^2 \boldsymbol{J}_2)$$

$$\boldsymbol{J}_1 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \quad \boldsymbol{J}_2 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

- This helps completely avoid inverting potentially high-dimensional $T \times T$ matrix $\boldsymbol{\Psi}$

Parameter estimation

- Taking partial derivatives of log-likelihood with respect to $\boldsymbol{\Sigma}$ yields

$$\boldsymbol{\Sigma} = \frac{(\boldsymbol{Y} - \boldsymbol{M})\boldsymbol{R}_\phi^{-1}(\boldsymbol{Y} - \boldsymbol{M})^\top}{T}.$$

- Taking partial derivatives of log-likelihood with respect to $\boldsymbol{\mu}_k$ yields

$$\boldsymbol{\mu}_k = \left(\mathcal{M}(\boldsymbol{Y};\boldsymbol{\lambda}) - \sum_{\substack{k'=0\\k'\neq k}}^{K} \boldsymbol{\mu}_{k'}\boldsymbol{m}_{k'}^\top\right)\boldsymbol{R}_\phi^{-1}\boldsymbol{m}_k\left(\boldsymbol{m}_k^\top\boldsymbol{R}_\phi^{-1}\boldsymbol{m}_k\right)^{-1}$$

  - this system of $K+1$ equations can be solved for $\boldsymbol{\mu}_k$, $k = 0, 1, \ldots, K$
- Analytical expressions for other parameters (ϕ, $\boldsymbol{\lambda}$) unavailable
  - plug the expressions for $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}$ into log-likelihood
  - maximize log-likelihood numerically to find estimates for ϕ and $\boldsymbol{\lambda}$

Multisubject studies

- one variable, N subjects (rows) , T times (columns)
- Log-likelihood function:

$$\log \mathcal{L}(\boldsymbol{Y}; \boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \lambda) = -\frac{NT}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}\log|\boldsymbol{\Psi}| \\ -\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}(\mathcal{M}(\boldsymbol{Y};\lambda) - \boldsymbol{M})\boldsymbol{\Psi}^{-1}(\mathcal{M}(\boldsymbol{Y};\lambda) - \boldsymbol{M})^\top\right\} \\ +\lambda \mathbf{1}_N^\top \boldsymbol{Y} \mathbf{1}_T$$

- Note that mean matrix $\boldsymbol{M}$ has NT parameters
  - N can be very large
- Often explanatory variables are available and $\boldsymbol{M}$ can be modeled as $\boldsymbol{M} = \boldsymbol{X}\boldsymbol{B}$
  - $\boldsymbol{X}$ is $N \times q$ design matrix
  - $\boldsymbol{B}$ is $q \times T$ matrix of linear model coefficients

Change point modeling

- For K shift change points at times $t_1, t_2, \ldots, t_K$, $\boldsymbol{B}$ can be written as

$$\boldsymbol{B} = \Big(\underbrace{\boldsymbol{\beta}_0, \boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1}_{t_1}, \underbrace{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2}_{t_2 - t_1}, \ldots,$$
$$\underbrace{\boldsymbol{\beta}_{K-1}, \ldots, \boldsymbol{\beta}_{K-1}, \boldsymbol{\beta}_K}_{t_K - t_{K-1}}, \underbrace{\boldsymbol{\beta}_K, \ldots, \boldsymbol{\beta}_K}_{T - t_K} \Big)$$
$$= \sum_{k=0}^{K} \boldsymbol{\beta}_k \boldsymbol{b}_k^\top$$

  - $\boldsymbol{\beta}_k$ is a vector of linear model coefficients of length q that corresponds to k^{th} process
  - $\boldsymbol{b}_k$ is a vector of length T that consists of 0's and 1's, with 1's at positions corresponding to the k^{th} process

Modeling covariance matrices

- $N \times N$ matrix $\boldsymbol{\Sigma}$ can be a problem if N is large
- Specific choice of $\boldsymbol{\Sigma}$ depends on a particular application
- We illustrate further model development in random effect setting
- Then, covariance matrix $\boldsymbol{\Sigma}$ is given by $\boldsymbol{\Sigma} = \sigma^2 \text{diag}\{\boldsymbol{V}_1, \dots, \boldsymbol{V}_M\}$ with

$$\boldsymbol{V}_m = \begin{bmatrix} \eta + 1 & \eta & \dots & \eta \\ \eta & \eta + 1 & \dots & \eta \\ \dots & \dots & \dots & \dots \\ \eta & \eta & \dots & \eta + 1 \end{bmatrix}_{n_m \times n_m}$$

  - m - block number, $m = 1, 2, \dots, M$
  - n_m - size of m^{th} block
  - $\eta = \frac{\sigma_b^2}{\sigma^2}$ - ratio of between and within block variances

Parameter estimation

- To avoid inverting potentially large $N \times N$ $\boldsymbol{\Sigma}$, the following expressions can be obtained

$$|\boldsymbol{\Sigma}| = \sigma^{2N} |\mathcal{D}_M(\boldsymbol{V})| = \sigma^{2N} (n\eta + 1)^M$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2(\eta n + 1)} \left(\eta \left(n\boldsymbol{I}_N - \mathcal{D}_N(\mathbf{1}_n \mathbf{1}_n^\top) \right) + \boldsymbol{I}_N \right)$$

  - $\mathcal{D}_M(\boldsymbol{V})$ is block-diagonal matrix, consisting of M blocks $\boldsymbol{V}$

Parameter estimation

- Taking partial derivatives of log-likelihood with respect to $\boldsymbol{\beta}_k$ yields

$$\boldsymbol{\beta}_k = \left(\boldsymbol{X}^\top \mathcal{D}_M^{-1}(\boldsymbol{V})\boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \mathcal{D}_M^{-1}(\boldsymbol{V}) \times \left(\mathcal{M}(\boldsymbol{Y};\lambda) - \boldsymbol{X}\sum_{\substack{k'=0\\k'\neq k}}^{K} \boldsymbol{\beta}_{k'}\boldsymbol{b}_{k'}^\top\right) \boldsymbol{R}_\phi^{-1}\boldsymbol{b}_k \left(\boldsymbol{b}_k^\top \boldsymbol{R}_\phi^{-1}\boldsymbol{b}_k\right)^{-1}$$

  - this system of $K+1$ equations can be solved for $\boldsymbol{\beta}_k$, $k = 0, 1, \ldots, K$
- Taking partial derivatives of log-likelihood with respect to σ^2 yields

$$\sigma^2 = \frac{\mathrm{tr}\left\{\mathcal{D}_M^{-1}(\boldsymbol{V})(\mathcal{M}(\boldsymbol{Y};\lambda) - \boldsymbol{XB})\boldsymbol{R}_\phi^{-1}(\mathcal{M}(\boldsymbol{Y};\lambda) - \boldsymbol{XB})^\top\right\}}{TN}$$

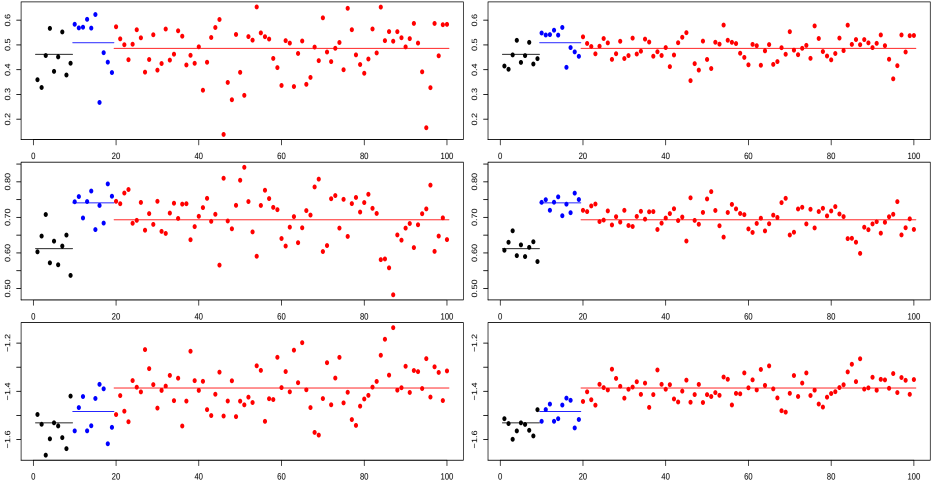
- Analytical expressions for other parameters (η, ϕ, λ) unavailable
- Plug $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ and σ^2 into log-likelihood and numerically optimize it for η, ϕ, λ

Simulation study in multivariate setting

- Assume p-variate observations observed over T time points
- Simulated datasets with $p = 3$, and $T = 100$
  - $\boldsymbol{\mu}_0 = (1, 1.2, -2.3)^\top$
  - $\boldsymbol{\mu}_1 = (1.2, 1.7, -2.2)^\top$
  - $\boldsymbol{\mu}_2 = (1.1, 1.5, -2.0)^\top$
  - $\lambda = (3, 2, -0.5)^\top$
  - $\phi = 0.1, 0.5, 0.9$
  - $\boldsymbol{\Sigma} = \frac{1}{3}\begin{bmatrix} 0.4 & -0.1 & 0.0 \\ -0.1 & 0.2 & -0.1 \\ 0.0 & -0.1 & 0.1 \end{bmatrix}$
  - $\boldsymbol{\Sigma}, \boldsymbol{\Sigma}/2, \boldsymbol{\Sigma}/4$
- Competitors (available through ECP R package):
  - probabilistic pruning with Energy statistic as goodness-of-fit measure
  - probabilistic pruning with Kolmogorov-Smirnov statistic as goodness-of-fit measure
- Model selection based on BIC

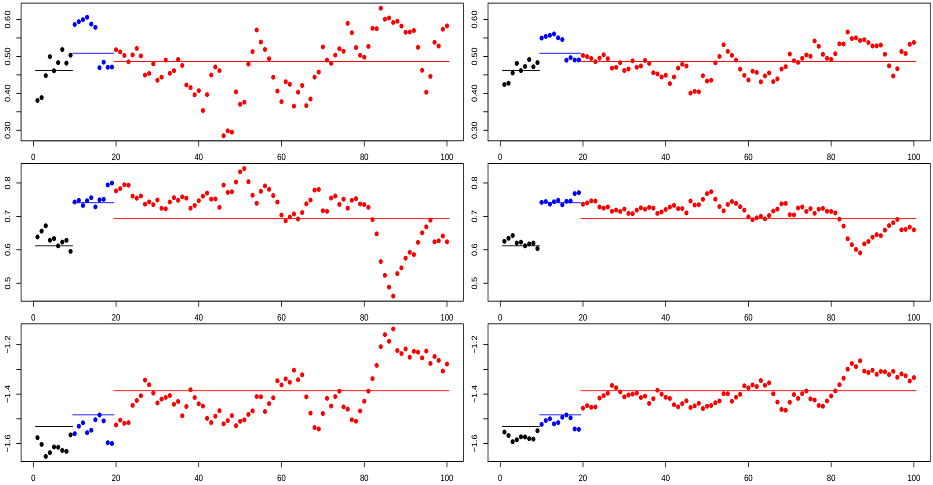
Simulation study

- Simulated datasets
  - challenging settings
  - first change point is easier to detect



$\phi = 0.1, \mathbf{\Sigma}$ $\phi = 0.1, \mathbf{\Sigma}/4$

Simulation study



$\phi = 0.9, \mathbf{\Sigma}$ $\qquad$ $\phi = 0.9, \mathbf{\Sigma}/4$

Simulation study

- Two change points at $t_1 = 10$, and $t_2 = 20$
  - First two processes are relatively short compared to the third one

$K = 2$		$\boldsymbol{\Sigma}$			$\boldsymbol{\Sigma}/2$			$\boldsymbol{\Sigma}/4$		
$t_1 = 10$, $t_2 = 20$		Method	Energy	KS	Method	Energy	KS	Method	Energy	KS
$\phi = 0.1$	$\{t_1, t_2\}$	0.061	0	0.024	0.353	0	0.020	0.660	0.036	0.044
	$\{t_1, t_2, x\}$	0	0	0	0	0	0	0	0.016	0
	$\{t_1, \tilde{t}_2\}/\{\tilde{t}_1, t_2\}$	0.236	0	0.116	0.326	0.004	0.148	0.277	0.028	0.224
	$\{t_1\}/\{t_2\}$	0.577	0	0.04	0.213	0	0.056	0.016	0.012	0.064
	$\{t_1, !t_2\}/\{!t_1, t_2\}$	0.126	0.024	0.024	0.108	0.080	0.016	0.047	0.192	0.028
$\phi = 0.5$	$\{t_1, t_2\}$	0.032	0	0	0.194	0	0.016	0.600	0.020	0.028
	$\{t_1, t_2, x\}$	0	0	0	0	0	0	0	0.008	0.004
	$\{t_1, \tilde{t}_2\}/\{\tilde{t}_1, t_2\}$	0.101	0	0.076	0.166	0	0.092	0.185	0.012	0.132
	$\{t_1\}/\{t_2\}$	0.734	0	0.032	0.479	0	0.092	0.128	0.004	0.132
	$\{t_1, !t_2\}/\{!t_1, t_2\}$	0.133	0.020	0.016	0.161	0.060	0.024	0.087	0.176	0.020
$\phi = 0.9$	$\{t_1, t_2\}$	0.148	0.004	0.004	0.492	0.028	0.012	0.932	0.356	0.024
	$\{t_1, t_2, x\}$	0	0.004	0	0	0.008	0	0	0.044	0.004
	$\{t_1, \tilde{t}_2\}/\{\tilde{t}_1, t_2\}$	0.012	0	0.044	0.016	0.004	0.052	0	0.016	0.076
	$\{t_1\}/\{t_2\}$	0.692	0	0.020	0.424	0.004	0.040	0.040	0.068	0.060
	$\{t_1, !t_2\}/\{!t_1, t_2\}$	0.136	0.120	0.032	0.068	0.188	0.048	0.028	0.148	0.060

Simulation study

- Two change points at $t_1 = 10$, and $t_2 = 50$
  - First process is relatively short compared to the other two

$K = 2$		$\boldsymbol{\Sigma}$			$\boldsymbol{\Sigma}/2$			$\boldsymbol{\Sigma}/4$		
$t_1 = 10, t_2 = 50$		Method	Energy	KS	Method	Energy	KS	Method	Energy	KS
$\phi = 0.1$	$\{t_1, t_2\}$	0.230	0	0	0.392	0	0	0.661	0.008	0.008
	$\{t_1, t_2, x\}$	0	0	0	0	0.004	0	0	0.008	0
	$\{t_1, \tilde{t}_2\}/\{\tilde{t}_1, t_2\}$	0.383	0	0	0.446	0	0.004	0.307	0	0.012
	$\{t_1\}/\{t_2\}$	0.068	0.068	0.036	0	0.128	0.056	0	0.296	0.060
	$\{t_1, !t_2\}/\{!t_1, t_2\}$	0.319	0.012	0.104	0.162	0.052	0.112	0.031	0.084	0.176
$\phi = 0.5$	$\{t_1, t_2\}$	0.117	0	0	0.332	0	0.004	0.629	0.004	0.004
	$\{t_1, t_2, x\}$	0	0	0	0	0.004	0	0	0	0
	$\{t_1, \tilde{t}_2\}/\{\tilde{t}_1, t_2\}$	0.170	0	0	0.230	0	0.008	0.222	0	0.004
	$\{t_1\}/\{t_2\}$	0.377	0.036	0.028	0.097	0.116	0.024	0	0.284	0.068
	$\{t_1, !t_2\}/\{!t_1, t_2\}$	0.336	0.024	0.056	0.341	0.052	0.096	0.149	0.088	0.104
$\phi = 0.9$	$\{t_1, t_2\}$	0.216	0	0.004	0.576	0.008	0.004	0.948	0.128	0
	$\{t_1, t_2, x\}$	0	0	0	0	0	0	0	0.064	0
	$\{t_1, \tilde{t}_2\}/\{\tilde{t}_1, t_2\}$	0.008	0.004	0.004	0.008	0.012	0.008	0	0.012	0.016
	$\{t_1\}/\{t_2\}$	0.600	0.156	0.016	0.316	0.412	0.044	0.044	0.580	0.076
	$\{t_1, !t_2\}/\{!t_1, t_2\}$	0.168	0.076	0.064	0.100	0.088	0.096	0.008	0.152	0.136

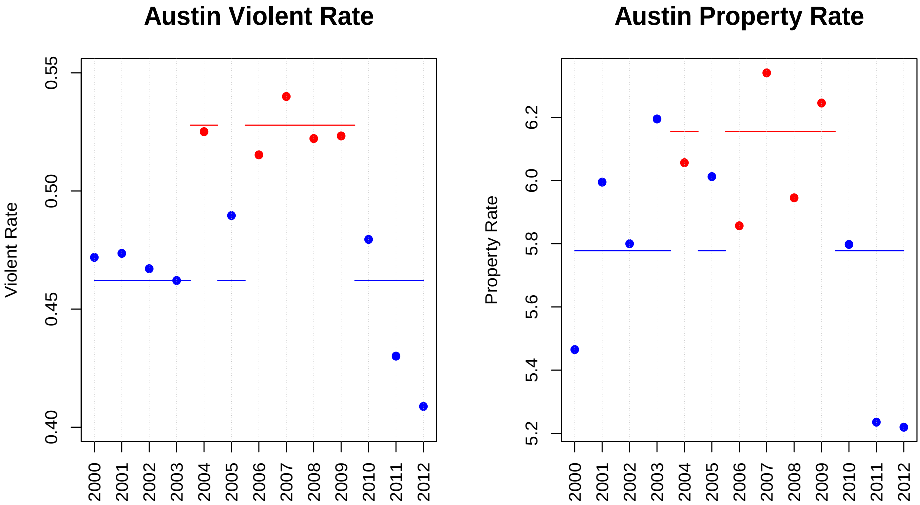
Simulation study - results

- Proposed method outperforms competitors substantially
- Best performance for $\phi = 0.9$ and $\mathbf{\Sigma}/4$:
  - 93.2% for $t_1 = 10$ and $t_2 = 20$
  - 94.8% for $t_1 = 10$ and $t_2 = 50$
- Nearly always identifies at least one change point correctly
- No tendency to overestimate the number of change points
  - BIC assumes large penalty for extra parameters
- Results are better in all cases when $t_2 = 50$

Illustration: crime rates in two US cities

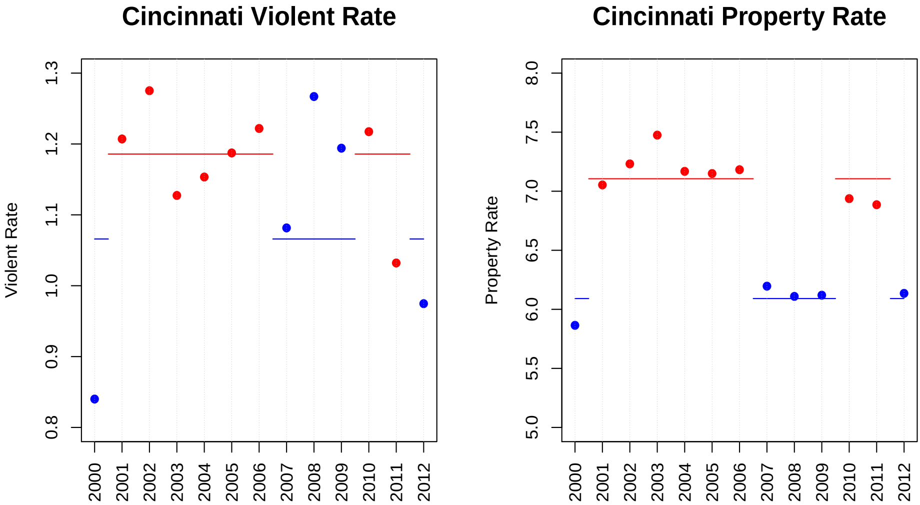
- Crime rates obtained at the US Department of Justice, FBI web-site (http://www.ucrdatatool.gov/Search/Crime/Crime.cfm)
- Two categories of crime:
  - Violent crime
    - Murder, Rape, Robbery, and Aggravated Assault
  - Property crime
    - Burglary, Larceny, and Moto-Vehicle Theft
- 13-year time period (2000-2012)
- Consider Austin and Cincinnati
- Assume all permutations of processes

Illustration: crime rates in Austin



- Solution is driven by Violent crimes primarily

Illustration: crime rates in Cincinnati

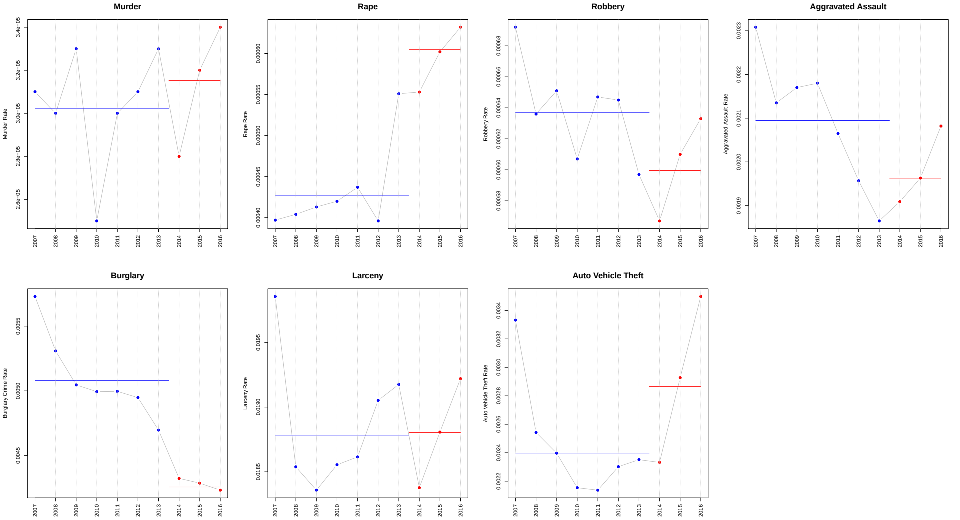


- Solution is driven by Property crimes primarily

Application: effect of Amendment 64 in Colorado

- The same seven crime variables are considered
- For Colorado, we found data for 10-year period: 2007 to 2016
- Amendment 64: legalization of marijuana
  - added to Colorado constitution in 2012
  - first official sellers appeared in January, 2014
- Goal: compare crimes in 2007-2013 versus those in 2014-2016
  - study crime variables

Application: effect of Amendment 64 in Colorado

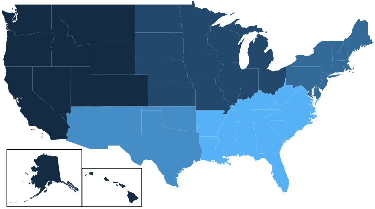


Application: effect of Amendment 64 in Colorado

- BIC of model without change point in 2014 is -996.2
- BIC of model with change point in 2014 is -1,006.1
- Likelihood ratio test yields p-value 1.47×10^{-6}
- Considered all combinations of variables to detect one leading to the most significant change point
  - Rape, Burglary, and Murder

Application: Burglary rates for 125 cities in the US

- Burglary rates are studied
- 13-year time period (2000-2012)
- US regions: *West*, *MidWest*, *NorthEast*, *SouthWest*, *SouthEast*



- Selected 25 most populated cities in each region
  - $N = 125$, $M = 5$, $n = 25$, and $T = 13$
- Multisubject setting

Application: Burglary rates for 125 cities in the US

- Change point is found for Burglary in 2012

$\hat{\boldsymbol{\beta}}_0$	$\hat{\boldsymbol{\beta}}_1$	$\hat{\eta}$	$\hat{\lambda}$	$\hat{\sigma}^2$	$\hat{\phi}$
$(5.70, 7.19, 5.77, 6.93, 8.28)^\top$	$(5.97, 6.99, 5.48, 6.52, 7.81)^\top$	0.010	-0.778	0.034	0.967

- In 2012, decrease in Burglary rates observed in all regions except *West*
- $\hat{\phi}$ high, $\hat{\sigma}^2$ low - according to simulation studies, change point estimation is rather accurate in such cases

Discussion and future work

- Novel approach to robust change point analysis proposed
- Capable of incorporating various dependence structures for
  - observations
  - subjects
  - variables
- Generalization: multisubject multivariate processes over time *i.e.*, N, p, T
  - requires tensors instead of matrices
- R package under development