TECHNISCHE UNIVERSITÄT DRESDEN

Chair of Econometrics and Statistics esp. in Transport, Dresden University of Technology

A NONPARAMETRIC CONTROL CHART FOR FINANCIAL SURVEILLANCE

Ostap Okhrin, Ya-Fei Xu

Dresden, April 8, 2017

DRESDEN concept

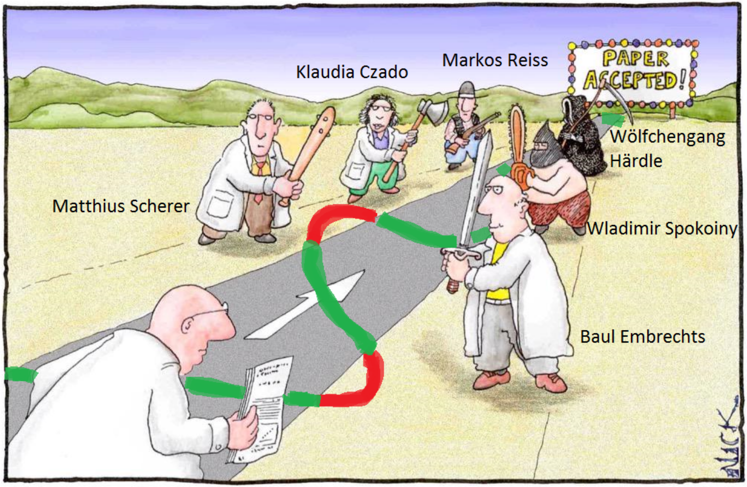
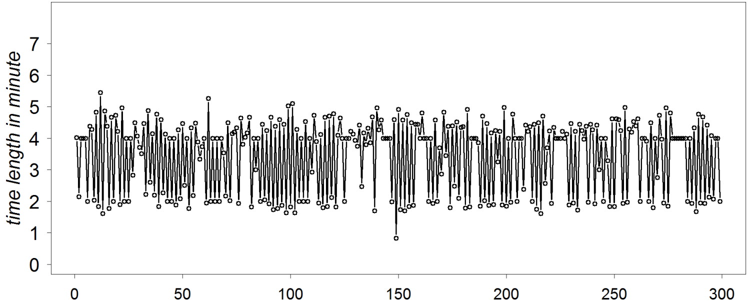


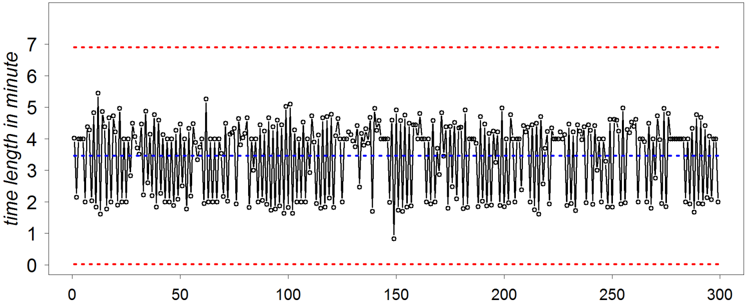
Figure 1: IF (quality out of control), THEN (kill).


© Mike Baldwin/Cornered
DOLL LINE
QUALITY CONTROL
TOO CREEPY
BALDWIN

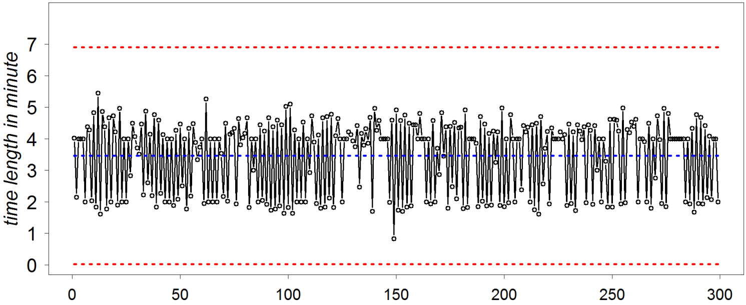
A Real Data Example



A Real Data Example



A Real Data Example



Old Faithful Geyser Ejection Duration at Yellowstone National Park, USA

Data Source: `geyser{MASS}`, Azzalini and Bowman (1990, *Appl Stat*) 299 readings in period of 1985.08.01-1985.08.15.

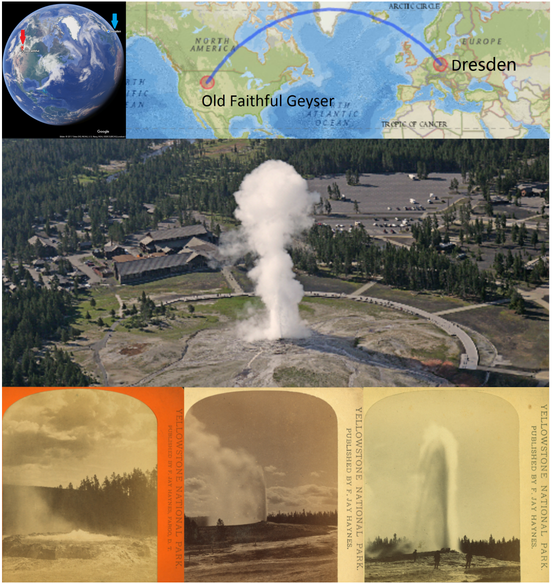


Figure 2: Old Faithful Geyser@Yellowstone National Park, USA

A Simulation Example: Jump

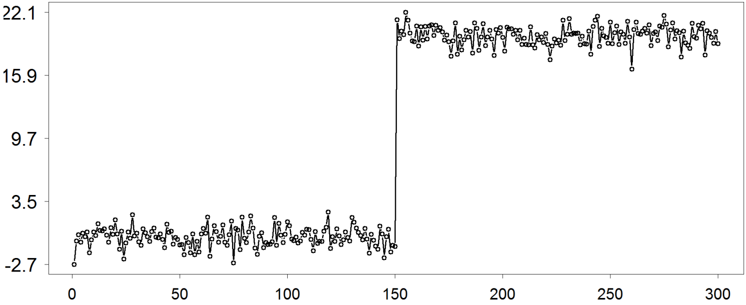
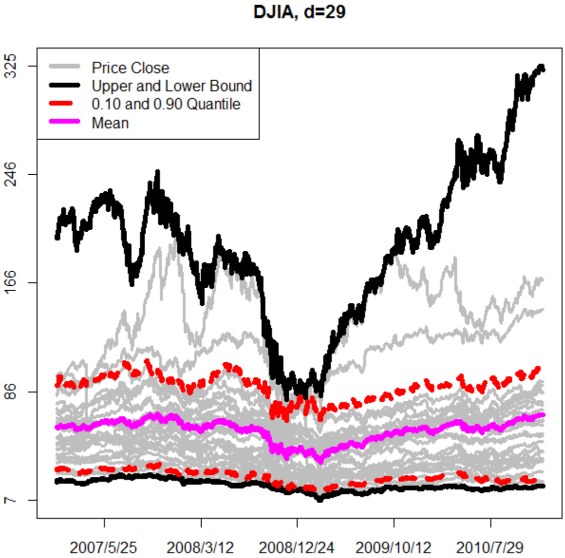
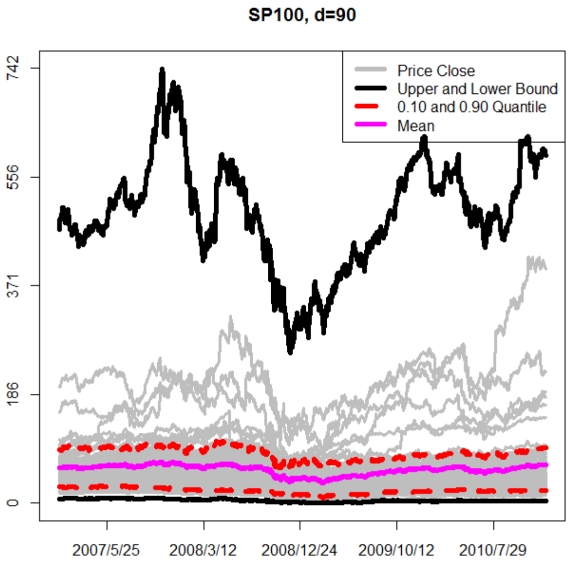


Figure 3: Mean jump from $N(0,1)$ to $N(20,1)$

Real Data in 29 Dim.



Real Data in 90 Dim.



Three Questions

1. How to monitor many processes simultaneously?

Three Questions

1. How to monitor many processes simultaneously?
2. How to monitor many processes simultaneously and onlinely?

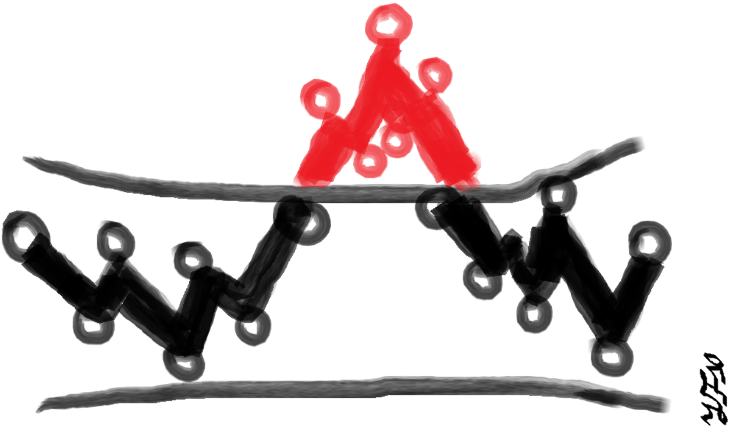
Three Questions

1. How to monitor many processes simultaneously?
2. How to monitor many processes simultaneously and onlinely?
3. How to monitor many processes simultaneously, onlinely and nonparametrically?

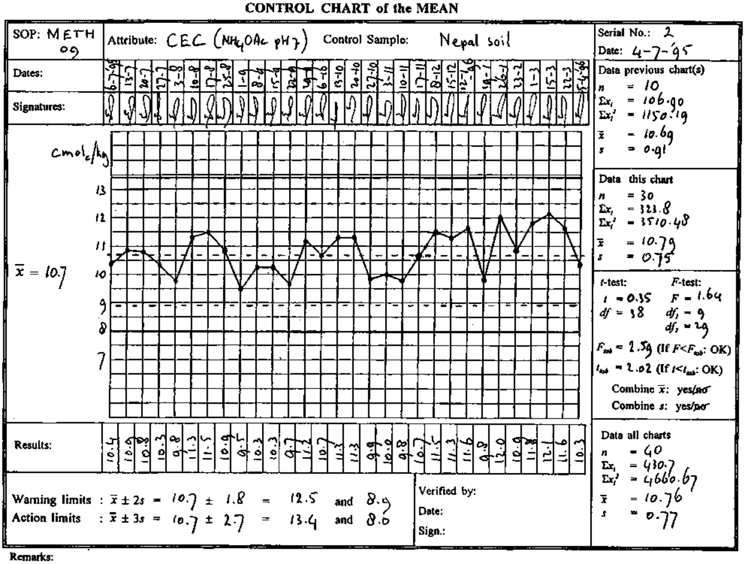
Outline

1. Motivation ✓
2. Methodology
3. Introduction of Copula
4. Simulation Study
5. Empirical Study
6. `R` Package `"EnergyOnlineCPM"`
7. Conclusion

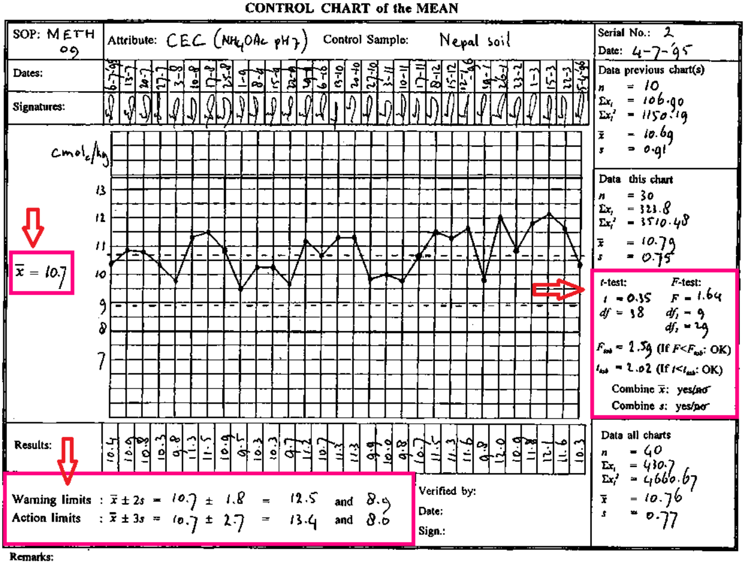
02 Methodology



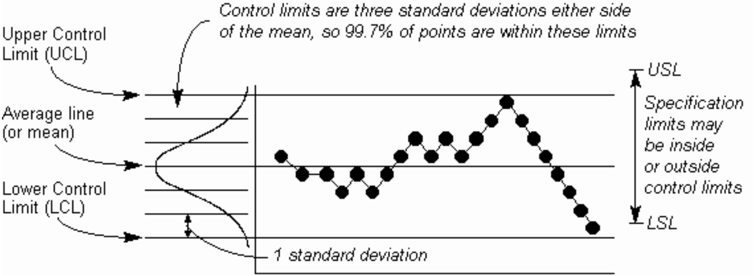
An Intuition



An Intuition



An Intuition



Classical Control Chart

Classical SPC tools:

- Shewhart Chart, developed in 1920s@Bell Telephone Laboratories and published in Shewhart (1931, *Economic Control of Quality of Manufactured Product*).

Classical Control Chart

Classical SPC tools:

- Shewhart Chart, developed in 1920s@Bell Telephone Laboratories and published in Shewhart (1931, *Economic Control of Quality of Manufactured Product*).
- CUSUM (cumulative sum), Page (1954, *Biometrika*).

Classical Control Chart

Classical SPC tools:

- Shewhart Chart, developed in 1920s@Bell Telephone Laboratories and published in Shewhart (1931, *Economic Control of Quality of Manufactured Product*).
- CUSUM (cumulative sum), Page (1954, *Biometrika*).
- EWMA (exponentially weighted moving average), Roberts (1959, *Technometrics*).

Classical Control Chart

Classical SPC tools:

- Shewhart Chart, developed in 1920s@Bell Telephone Laboratories and published in Shewhart (1931, *Economic Control of Quality of Manufactured Product*).
- CUSUM (cumulative sum), Page (1954, *Biometrika*).
- EWMA (exponentially weighted moving average), Roberts (1959, *Technometrics*).

Restrictions of the classical models:

Classical Control Chart

Classical SPC tools:

- Shewhart Chart, developed in 1920s@Bell Telephone Laboratories and published in Shewhart (1931, *Economic Control of Quality of Manufactured Product*).
- CUSUM (cumulative sum), Page (1954, *Biometrika*).
- EWMA (exponentially weighted moving average), Roberts (1959, *Technometrics*).

Restrictions of the classical models:

- Most center on mean and variance monitoring (first and second moment).

Classical Control Chart

Classical SPC tools:

- Shewhart Chart, developed in 1920s@Bell Telephone Laboratories and published in Shewhart (1931, *Economic Control of Quality of Manufactured Product*).
- CUSUM (cumulative sum), Page (1954, *Biometrika*).
- EWMA (exponentially weighted moving average), Roberts (1959, *Technometrics*).

Restrictions of the classical models:

- Most center on mean and variance monitoring (first and second moment).
- Distributional assumptions are "skewed" (Gaussian world is ideal).

Classical Control Chart

Classical SPC tools:

- Shewhart Chart, developed in 1920s@Bell Telephone Laboratories and published in Shewhart (1931, *Economic Control of Quality of Manufactured Product*).
- CUSUM (cumulative sum), Page (1954, *Biometrika*).
- EWMA (exponentially weighted moving average), Roberts (1959, *Technometrics*).

Restrictions of the classical models:

- Most center on mean and variance monitoring (first and second moment).
- Distributional assumptions are "skewed" (Gaussian world is ideal).
- Large historical sample is required (difficult in some cases).

Change Point Model

- Hawkins, Qiu, and Kang (2003, JQT) proposed parametric change point formulation for Phase-II SPC, requiring no knowledge on univariate in- or out-of-control process parameters for shifts in location.

Change Point Model

- Hawkins, Qiu, and Kang (2003, JQT) proposed parametric change point formulation for Phase-II SPC, requiring no knowledge on univariate in- or out-of-control process parameters for shifts in location.
- Zamba and Hawkins (2006, Technometrics) employed likelihood ratio test for shift in multivariate normal data for shifts in location and variance.

Change Point Model

- Hawkins, Qiu, and Kang (2003, JQT) proposed parametric change point formulation for Phase-II SPC, requiring no knowledge on univariate in- or out-of-control process parameters for shifts in location.
- Zamba and Hawkins (2006, Technometrics) employed likelihood ratio test for shift in multivariate normal data for shifts in location and variance.
- Hawkins and Deng (2010, JQT) used univariate Wilcoxon-Mann-Whitney nonparametric test for changes in location.

Change Point Model

- Hawkins, Qiu, and Kang (2003, JQT) proposed parametric change point formulation for Phase-II SPC, requiring no knowledge on univariate in- or out-of-control process parameters for shifts in location.
- Zamba and Hawkins (2006, Technometrics) employed likelihood ratio test for shift in multivariate normal data for shifts in location and variance.
- Hawkins and Deng (2010, JQT) used univariate Wilcoxon-Mann-Whitney nonparametric test for changes in location.
- Holland and Hawkins (2014, JQT) used multivariate Choi-Marden (1997) nonparametric test for shifts in location.

Change Point Model

- Hawkins, Qiu, and Kang (2003, JQT) proposed parametric change point formulation for Phase-II SPC, requiring no knowledge on univariate in- or out-of-control process parameters for shifts in location.
- Zamba and Hawkins (2006, Technometrics) employed likelihood ratio test for shift in multivariate normal data for shifts in location and variance.
- Hawkins and Deng (2010, JQT) used univariate Wilcoxon-Mann-Whitney nonparametric test for changes in location.
- Holland and Hawkins (2014, JQT) used multivariate Choi-Marden (1997) nonparametric test for shifts in location.
- Zhou, Zi, Geng and Li (2014, Comm. in Stat.) used multivariate nonparametric number of runs test for shifts in location.

Change Point Model

- Hawkins, Qiu, and Kang (2003, JQT) proposed parametric change point formulation for Phase-II SPC, requiring no knowledge on univariate in- or out-of-control process parameters for shifts in location.
- Zamba and Hawkins (2006, Technometrics) employed likelihood ratio test for shift in multivariate normal data for shifts in location and variance.
- Hawkins and Deng (2010, JQT) used univariate Wilcoxon-Mann-Whitney nonparametric test for changes in location.
- Holland and Hawkins (2014, JQT) used multivariate Choi-Marden (1997) nonparametric test for shifts in location.
- Zhou, Zi, Geng and Li (2014, Comm. in Stat.) used multivariate nonparametric number of runs test for shifts in location.
- Chen, Zi and Zou (2016, Technometrics) used multivariate nonparametric Wilcoxon rank-sum test for shifts in location.

Change Point Model

- Hawkins, Qiu, and Kang (2003, JQT) proposed parametric change point formulation for Phase-II SPC, requiring no knowledge on univariate in- or out-of-control process parameters for shifts in location.
- Zamba and Hawkins (2006, Technometrics) employed likelihood ratio test for shift in multivariate normal data for shifts in location and variance.
- Hawkins and Deng (2010, JQT) used univariate Wilcoxon-Mann-Whitney nonparametric test for changes in location.
- Holland and Hawkins (2014, JQT) used multivariate Choi-Marden (1997) nonparametric test for shifts in location.
- Zhou, Zi, Geng and Li (2014, Comm. in Stat.) used multivariate nonparametric number of runs test for shifts in location.
- Chen, Zi and Zou (2016, Technometrics) used multivariate nonparametric Wilcoxon rank-sum test for shifts in location.

toward multivariate+nonparametric+CDF?

Energy Test

Theorem (Gábor J. Székely and Maria L. Rizzo, 2005)

Let $X \overset{\mathcal{L}}{\sim} F$ *and* $Y \overset{\mathcal{L}}{\sim} G$ *be two d-dimensional random vectors.* X', Y' *are copies of* X *and* Y*. The corresponding characteristic functions of the two samples are* ϕ_X *and* ϕ_Y*. If* $0 < \alpha < 2$ *with* $E||X||_2{}^{\alpha} < \infty$ *and* $E||Y||_2{}^{\alpha} < \infty$ *then*

$$\int_{\mathbb{R}^d} \frac{|\phi_X(\mathbf{t}) - \phi_Y(\mathbf{t})|^2}{||\mathbf{t}||_2^{d+\alpha}} d\mathbf{t} = W(d,\alpha)\mathcal{E}^{\alpha}(X,Y), \tag{1}$$

where

$$\begin{aligned}
\phi_X(\mathbf{t}) &= E\{\exp[i(X,\mathbf{t})]\} = E[\cos(X,\mathbf{t}) + i\,\sin(\mathrm{X},\mathbf{t})], \\
W(d,\alpha) &= \frac{2\Pi^{\frac{d}{2}}\Gamma(1-\frac{\alpha}{2})}{\alpha 2^{\alpha}\Gamma(\frac{\alpha+d}{2})}, \\
\Gamma(r) &= \int_0^{\infty} t^{r-1}e^{-t}dt,\ r \neq 0,-1,-2,\dots, \\
\mathcal{E}^{\alpha}(X,Y) &= 2E||X-Y||_2^{\alpha} - E||X-X'||_2^{\alpha} - E||Y-Y'||_2^{\alpha}. \qquad (2)
\end{aligned}$$

Energy Test

Theorem (Gábor J. Székely and Maria L. Rizzo, 2005)

Following set-up in Theorem 1, $\mathcal{E}^\alpha(X, Y) = 0$ iff X and Y are identically distributed.

Therefore the metric $\mathcal{E}^\alpha(X, Y)$ can be used to measure the divergence between two distributions. The empirical counterpart of Equation (2) is derived as

$$\begin{aligned}\hat{\mathcal{E}}^\alpha(X, Y) \quad &= \quad \frac{\tau(T-\tau)}{T}\left(\frac{2}{\tau(T-\tau)}\sum_{i=1}^{\tau}\sum_{j=1}^{T-\tau}||x_i - y_j||_2^\alpha\right. \\ &\quad \left. -\frac{1}{\tau^2}\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}||x_i - x_j||_2^\alpha - \frac{1}{(T-\tau)^2}\sum_{i=1}^{T-\tau}\sum_{j=1}^{T-\tau}||y_i - y_j||_2^\alpha\right),\end{aligned}$$

where $\{x_i,\ i = 1, 2, \ldots, \tau\}, \{y_j,\ j = 1, 2, \ldots, T-\tau\}$ are observations.

Phase I SPC

- Let $x = \{x_1, ..., x_T\}$ denote a sample of observations with length of T. In Phase I detection, the sample and its size T are fixed, i.e. no new in-comer.

Phase I SPC

- Let $x = \{x_1, ..., x_T\}$ denote a sample of observations with length of T. In Phase I detection, the sample and its size T are fixed, i.e. no new in-comer.
- Assume there is a change occurs at k, then the problem can be represented in the following test hypotheses,

$$
\begin{aligned}
H_0 : &\quad X_t \overset{\mathcal{L}}{\sim} F_0(x|\theta_0),\ 1 \leq t \leq T, \\
H_1 : &\quad X_t \overset{\mathcal{L}}{\sim} \begin{cases} F_0(x|\theta_0), & 1 \leq t \leq k, \\ F_1(x|\theta_1), & k < t \leq T. \end{cases}
\end{aligned}
$$

Phase I SPC

- Since the change point location is unknown, hence the two-sample test will be performed at every point $1 < k < T$. The test statistic is changed to

$$Z_T = \max_{1<k<T} Z_{k,T}.$$

Phase I SPC

- Since the change point location is unknown, hence the two-sample test will be performed at every point $1 < k < T$. The test statistic is changed to

$$Z_T = \max_{1<k<T} Z_{k,T}.$$

- The change location can be estimated by

$$\hat{\tau} = \arg \max_{1<k<T} Z_{k,T}.$$

Phase II SPC

- Phase II detection considers the sample x with not fixed T size but an increasing size, i.e. T is increased with time proceeding.

Phase II SPC

- Phase II detection considers the sample x with not fixed T size but an increasing size, i.e. T is increased with time proceeding.
- Phase II can be extended from the Phase I with increasing sample size. That is whenever a new observation arrives then a new sample $x = \{x_1, \ldots, x_T, x_{T+1}, \ldots, x_t\}$ is constructed with size t.

Phase II SPC

- Phase II detection considers the sample x with not fixed T size but an increasing size, i.e. T is increased with time proceeding.
- Phase II can be extended from the Phase I with increasing sample size. That is whenever a new observation arrives then a new sample $x = \{x_1, \ldots, x_T, x_{T+1}, \ldots, x_t\}$ is constructed with size t.
- For every new observation the Phase I process will be performed. Hence the null hypothesis is rejected if $Z_t > h_t$. The Type I error α can be represented with

$$\begin{aligned} Pr(Z_1 > h_1) &= \alpha, \ t = 1, \\ Pr(Z_t > h_t | Z_{t-1} \leq h_{t-1}, \ldots, Z_1 \leq h_1) &= \alpha, \ t > 1. \end{aligned}$$

03 Introduction of Copula

Literal Copula

Copula is a word in Latin as a noun which means a "link" or a "band" utilized to connect two things together.

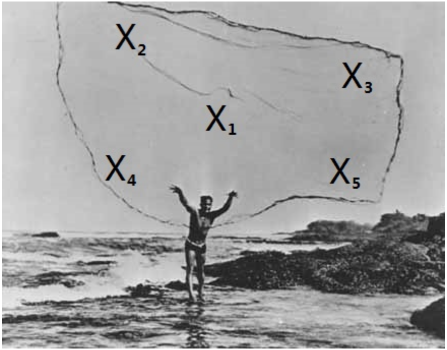


Figure 4: The ancient Hawaiian tradition of net fishing.

Sklar's Theorem

Theorem (Sklar 1959)

Given a d-dimensional joint CDF F such that $F(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$ of a random vector $(X_1, X_2, \ldots, X_d)^\top$ with margins $F_k(x) = \mathbb{P}(X_k \leq x)$, there exists a d-dimensional copula C such that

$$F(x_1, \ldots, x_d) = C\left\{F_1(x_1), \ldots, F_d(x_d)\right\}.$$

The copula C is unique if every F_k, $k = 1, 2, \ldots, d$, is continuous, otherwise C is uniquely defined on $\prod_{k=1}^{d} Range(F_k)$.

Gaussian Copula

Definition (Gaussian Copula)

For a K-dimensional uniform vector $u = (u_1, \ldots, u_K) \in [0,1]^K$ the Gaussian copula can be represented as follows,

$$C_{gs}(u;\rho) = \Phi_K\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_K); \rho\right),$$

where ρ is a $(K \times K)$ correlation matrix, Φ_K is a K-dimensional standard normal distribution function and Φ is a 1-dimensional standard normal distribution function.

Example: CDO Pricing Model (see Li 1999, 2000):

$$C(u_1, \ldots, u_K; \theta) \quad = \quad \Phi_K\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_K); \rho\right).$$

Student-t Copula

Definition (Student-t Copula)

Let $\nu \in (1, +\infty)$ the degree of freedom and $\rho := (1 - \frac{2}{\nu})var(X)$ the $(K \times K)$ matrix, $X := (X_1, \dots, X_K)^\top \in \mathbb{R}^K$, the dispersion matrix in an K-dimensional Student-t distribution $t_K(\nu, \rho)$ such that

$$t_K(x) = \frac{\Gamma(\frac{\nu+K}{2})}{\frac{\nu}{2}\sqrt{(\Pi\nu)^K \mid \rho \mid}} \left(1 + \frac{(x-\Delta)^T \rho^{-1}(x-\Delta)}{\nu}\right)^{-\frac{\nu+K}{2}},$$

where $\Gamma(y) = \int_0^{+\infty} x^y e^{-x} \frac{\mathrm{d}x}{x}$, $\Delta = \mathbb{E}(X)$ and $x = (x_1, \dots, x_K)$ then the Student-t Copula can be represented as follows,

$$C_t(u; \nu, \rho) = t_K\left(t^{-1}(u_1; \nu), \dots, t^{-1}(u_K; \nu); \nu, \rho\right).$$

Papers: Lindskog and McNeil (2001), Embrechts et al. (2001), Frey and McNeil (2003), Andersen et al. (2003), Demarta and McNeil (2004), Greenberg et al. (2004), Mashal et al. (2004).

Archimedean Copula

Definition (Archimedean Copula)

Let $\varphi\colon [0,1] \to [0,+\infty)$ is the generator function, which satisfies the following three properties,
1. $\varphi(1) = 0$
2. $\varphi(+\infty) = 1$
3. $\varphi\colon [0,1] \to [0,+\infty)$ is a decreasing function.

And let $\varphi^{[-1]}$ the pseudo-inverse of the generator function φ such that

$$\varphi^{[-1]}(t;\theta) = \begin{cases} \varphi^{-1}(t;\theta) & \text{if } 0 \le t \le \varphi(0;\theta) \\ 0 & \text{if } \varphi(0;\theta) \le t \le \infty \end{cases}$$

where φ^{-1} is n-monotone on $[0,+\infty)$ that it can be $K-2$ times differentiated and the derivatives fulfill $(-1)^j \varphi^{-1,j}(y;\theta) \ge 0$. $\forall\ y \ge 0$ and $j = 0, 1, \ldots, K-2$, and $(-1)^{K-2}\varphi^{-1,K-2}(y;\theta)$ is non-increasing and convex. Then the Archimedean copula can be represented as follows,

$$C(u_1, \ldots, u_n; \theta) = \varphi^{[-1]}\left(\varphi(u_1;\theta) + \cdots + \varphi(u_K;\theta);\theta\right).$$

Archimedean Copula: Examples

Archimedean Copula	Representation $C(u_1, \dots, u_d; \theta)$	Generator Function $\varphi(t; \theta)$	Parameter θ
Clayton	$\left(\sum_{k=1}^{d} u_k^{-\theta} - d + 1\right)^{-\theta^{-1}}$	$\frac{1}{\theta}(t^{-\theta} - 1)$	$[-1/(d-1), \infty) \backslash \{0\}$
Gumbel	$\exp\left\{-\sum_{k=1}^{d}(-\log u_k)^{\theta}\right\}^{\theta^{-1}}$	$\{-\log(t)\}^{\theta}$	$[1, +\infty)$
Joe	$1 - \left\{\sum_{k=1}^{d}(1-uk)^{\theta} - \prod_{k=1}^{d}(1-u_k)^{\theta}\right\}^{\frac{1}{\theta}}$	$-\log\left\{1-(1-t)^{\theta}\right\}$	$[1, +\infty)$

Table 1: Structures of common Archimedean copulas.

Exchangeable Gaussian Copula

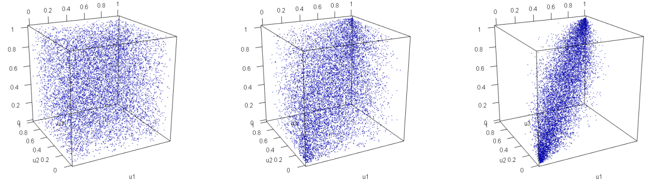


Figure 5: 10000 Monte Carlo simulations for 3-dimensional Gaussian copula $C_{gs}(u_1, u_2, u_3; \rho_{\mathcal{P}})$ with an exchangeable correlation matrix, where from left to right $\rho_{\mathcal{P}} = 0.1, 0.5, 0.9$.

Exchangeable Gaussian Copula

Figure 6: 10000 Monte Carlo simulations for 3-dimensional Gaussian copula $C_{gs}(u_1, u_2, u_3; \rho_{\mathcal{P}})$ with an exchangeable correlation matrix, where from left to right $\rho_{\mathcal{P}} = 0.1, 0.5, 0.9$.

Archimedean Copula

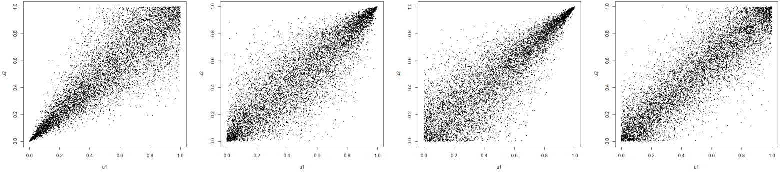


Figure 7: Plots from left to right are 2-dimensional scatter plots drawing from Clayton, Gumbel and Joe copulas. The Archimedean copulas are employed here with $\rho_{\mathcal{K}} = 0.7$.

Archimedean Copula

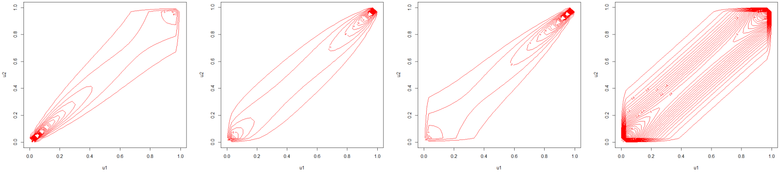


Figure 8: Plots from left to right are 2-dimensional contour plots for Clayton, Gumbel, Joe and Frank copulas. The Archimedean copulas are employed here with $\rho_{\mathcal{K}} = 0.7$.

Archimedean Copula

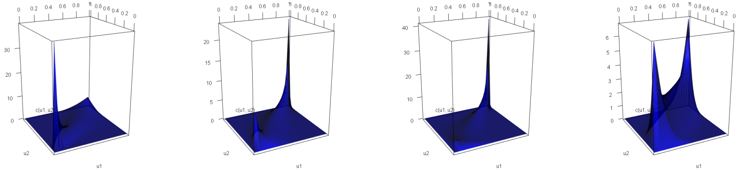
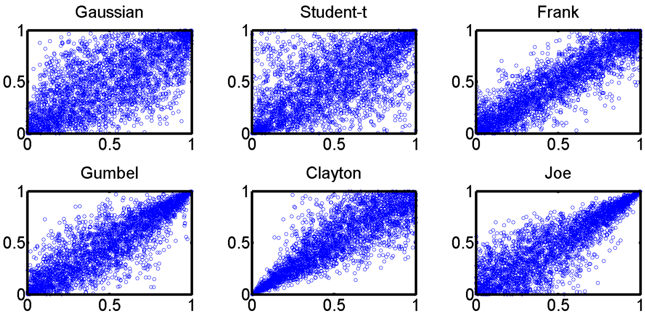


Figure 9: Plots from left to right are 2-dimensional copula density plots for Clayton, Gumbel, Joe and Frank copulas. The Archimedean copulas are employed here with $\rho_{\mathcal{K}} = 0.7$.

Archimedean Copula

Figure 10: Plots from left to right are 2-dimensional copula density plots for Clayton, Gumbel, Joe and Frank copulas. The Archimedean copulas are employed here with $\rho_{\mathcal{K}} = 0.7$.

A Comparison of Simulation



Vine Copula

- Vine copula or pair-copula constructions are originally proposed in Joe (1996) and developed in-depth by Bedford & Cooke (2001), Bedford & Cooke (2002), Kurowicka & Cooke (2006) and Aas, Czado, Frigessi & Bakken (2009).
- The catchy name is due to similarities of the graphical representation of vine copulae and botanical vines.
- The fundamental idea of the vine copula is to construct a d-dimensional copula by decomposing the dependence structure into $d(d-1)/2$ bivariate copulas.

Vine Copula-Derivation 1

Some notations:

- S: the index subset of $D = \{1, \ldots, d\}$ referring to the index set of conditioning variables.
- T: the index set of conditioned variables with $T \cup S = D$.
- $\sharp M$: the cardinality of set M.
- F_S: the CDF of variables with index in S, i.e. $F(x) = F_D(x)$.
- $F_{T|S}$: the conditional CDF of variables with index in T conditional on S.

To derive a vine copula for a given $x = (x_1, \ldots, x_d)^\top$ from a d-dimensional distribution function, i.e.

$$F(x) = \int_{(-\infty, x_S]} F_{T|S}(x_T | y_S) dF_S(y_S),$$

Vine Copula-Derivation 2

A replacement:

- The d-dimensional distribution function,

$$F(x) = \int_{(-\infty, x_{\boldsymbol{S}}]} F_{T|S}(x_T|y_S) dF_S(y_S). \tag{3}$$

- Replace the conditional distribution $F_{T|S}(x_T|y_S)$ by the corresponding $\sharp T$-dimensional copula $F_{T|S}(x_T|x_S) = C_{T;S}\{F_{j|S}(x_j|y_S) : j \in T\}$ such that

$$F(x) = \int_{(-\infty, x_{\boldsymbol{S}}]} C_{T;S}\{F_{j|S}(x_j|y_S) : j \in T\} dF_S(y_S). \tag{4}$$

Vine Copula-Derivation 3

A converting:

- Converting all univariate margins to uniform distributed random variables allows rewriting $F(x)$ as a d-dimensional copula

$$C(u) = \int_{[0,\boldsymbol{u_S}]} C_{T;S}\{G_{j|S}(u_j|v_S) : j \in T\} dC_S(v_S), \tag{5}$$

  where $G_{j|S}(u_j|v_S)$ is a conditional distribution from copula $C_{S\cup\{j\}}$.
- If $T = \{i_1, i_2\}$, then

$$C_{S\cup\{i_1,i_2\}}(u_{S\cup\{i_1,i_2\}}) = \int_{[0,\boldsymbol{u_S}]} C_{i_1,i_2;S}\{G_{i_1|S}(u_{i_1}|v_S), G_{i_2|S}(u_{i_2}|v_S)\} dC_S(v_S).$$

A converting:

- In case of continuous random variables, the d-dimensional distribution function admits a density function $f(x_1, \ldots, x_d)$, which can be decomposed and represented by bivariate copula densities in an analogue manner.
- An example of density decompositions for the 5-dimensional case related to so called R-vine (regular vine) copula is given as follows

$$
\begin{aligned}
& c\{F_1(x_1), \ldots, F_5(x_5)\} \\
& = c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{13}\{F_1(x_1), F_3(x_3)\} \cdot c_{14}\{F_1(x_1), F_4(x_4)\} \\
& \cdot c_{45}\{F_4(x_4), F_5(x_5)\} \\
& \cdot c_{15;4}\{F(x_1|x_4), F(x_5|x_4)\} \cdot c_{24;1}\{F(x_2|x_1), F(x_4|x_1)\} \cdot c_{34;1}\{F(x_3|x_1), F(x_4|x_1)\} \\
& \cdot c_{23;14}\{F(x_2|x_{14}), F(x_3|x_{14})\} \cdot c_{35;14}\{F(x_3|x_{14}), F(x_5|x_{14})\} \\
& \cdot c_{25;314}\{F(x_2|x_{314}), F(x_5|x_{314})\}.
\end{aligned}
$$

Drawable Vine Copula Graph

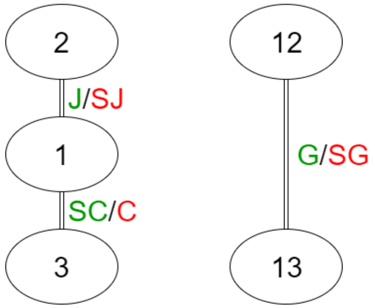
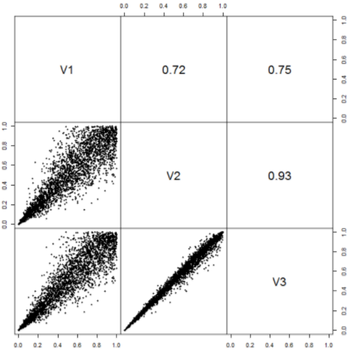


Figure 11: The structure of a five dimensional regular vine copula. The employed bivariate copula family for upper tail dependence structure is in green, lower tail dependence in red. Abbreviation: J: Joe, G: Gumbel, C: Clayton, S: Survival.

Simulation of Vine in 3d



Regular Vine Copula Graph

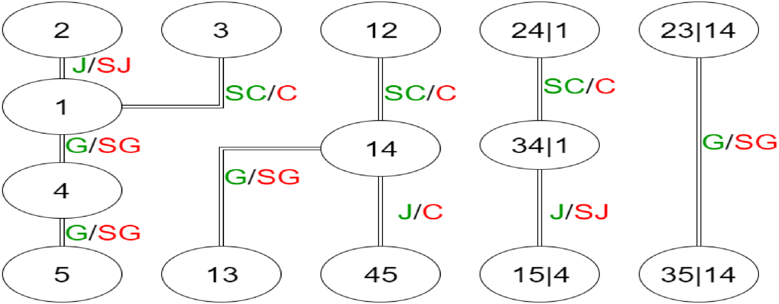
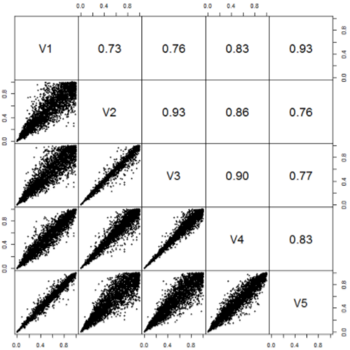
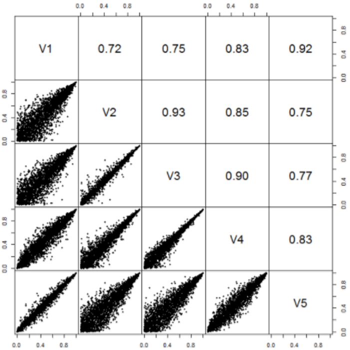


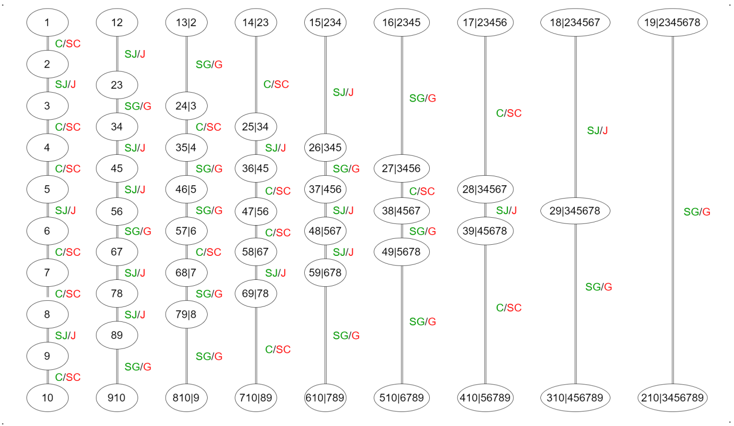
Figure 12: The structure of a five dimensional regular vine copula. The employed bivariate copula family for upper tail dependence structure is in green, lower tail dependence in red. Abbreviation: J: Joe, G: Gumbel, C: Clayton, S: Survival.

Simulation of Vine in 5d

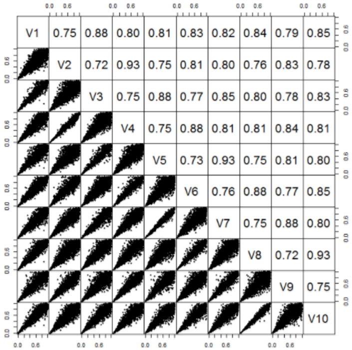


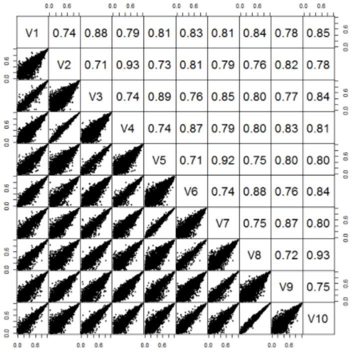


Drawable Vine Copula Graph



Simulation of Vine in 10d





04 Simulation Study

05 Empirical Study

06 R Package

Change Point Packages in R

- strucchange, Zeileis, Leisch, Hornik, and Kleiber 2002, univariate, parametric, Phase I, mean
- bcp, Erdman and Emerson 2007, univariate, Bayesian single change point analysis, Phase I, mean
- changepoint, Killick and Eckley 2011, univariate, (non)paramtric, Phase I, mean and variance
- cpm, Ross 2015, univaraite, (non)paramtric, Phase II, CDF
- NPMVCP, Holland 2014, multivariate, nonparametric, Phase II, mean
- ecp, James and Matteson 2014, (uni)multivariate, nonparametric, Phase I, CDF

Comparison with the State-of-Art

- `NPMVCP`, Holland 2014 JQT, multivariate, nonparametric, Phase II, mean
- `ecp`, James and Matteson 2014 JSS, (uni)multivariate, nonparametric, Phase I, CDF
- `EnergyOnlineCPM`, Okhrin and Xu 2017, multivariate, nonparametric, Phase II, CDF

An R Package "EnergyOnlineCPM"

- The package requires the R version $>=$ 3.3.2
- The package is now hosted in Github: https://github.com/YafeiXu/EnergyOnlineCPM/tree/master
- The web page for the package: https://sites.google.com/site/EnergyOnlineCPM/

Installation

```
install.packages("devtools")
library(devtools)
install_github("YafeiXu/EnergyOnlineCPM")
library(EnergyOnlineCPM)
```

A Toy Example in "EnergyOnlineCPM"

Simulation of two DGPs, i.e. a 5-dimensional standard Gaussian and a non-standard Gaussian with 555 in mean shift.

Example using "EnergyOnlineCPM"

```
library(MASS)
simNr=300 # simulate 300 length time series
# simulate 300 length 5 dimensonal standard Gaussian series
Sigma2 <- matrix(c(1,0,0,0,0, 0,1,0,0,0, 0,0,1,0,0, 0,0,0,1,0,
0,0,0,0,1),5,5)
Mean2=rep(1,5)
sim2=(mvrnorm(n = simNr, Mean2, Sigma2))
# simulate 300 length 5 dimensonal standard Gaussian series
Sigma3 <- matrix(c(1,0,0,0,0, 0,1,0,0,0, 0,0,1,0,0, 0,0,0,1,0,
0,0,0,0,1),5,5)
Mean3=rep(0,5)
sim3=(mvrnorm(n = simNr, Mean3, Sigma3))
```

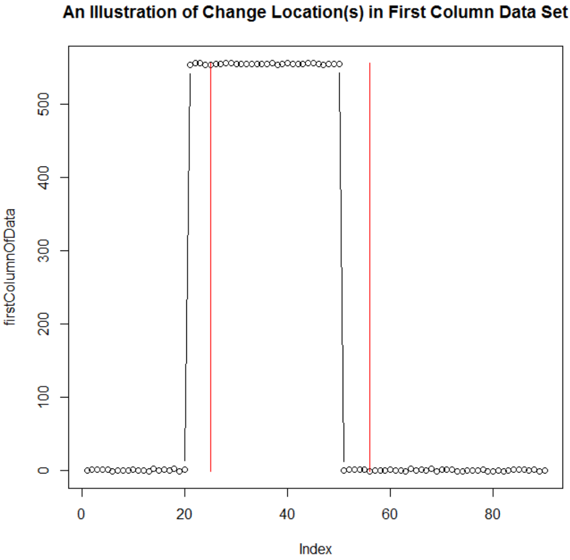
A Toy Example in "EnergyOnlineCPM"

Apply the function `maxEnergyCPMv` for multiple change points detection. The DGP has three segments, from a five dimensional standard Gaussian to a shift Gaussian then again back to the standard Gaussian.

Example using "EnergyOnlineCPM"

```
# construct a data set of length equal to 90.
# first 20 points are from standard Gaussian.
# second 30 points from a Gaussian with a mean shift with 555.
# last 40 points are from standard Gaussian.
data1=sim6=rbind(sim2[1:20,],(sim3+555)[1:30,],sim2[1:40,])
# set warm-up number as 20, permutation 200 times, significant
level 0.005
wNr=20
permNr=200
alpha=1/200
maxEnergyCPMv(data1,wNr,permNr,alpha)
```

A Toy Example in "EnergyOnlineCPM"



Conclusion: Model Features

- It is a nonparametric change point model which requires no pre-knowledge on the process compared to the classical parametric control chart.
- It is oriented to Phase II change point detection which is central for real time surveillance of stream data and can be applied extensively, e.g. in industrial quality control, finance, medical science, geology et al.
- The model is designed for multivariate time series, which is more practical and informative for catching the essence of data as a whole than univariate time series.

Conclusion: Model Performance

- In simulation study the mean, variance and tail change have been investigated and shown superior performance compared to a benchmark model.
- In empirical application, the model is implemented to surveillance 5-dimensional ETF data, 29-dimensional DJIA data and 90-dimensional SP100 data separetely. The model show the ability to detect the change of the market from quiescent period to volatile period which provides reference to financial institutions to prevent from market risk.
- An R package "`EnergyOnlineCPM`" for Phase II nonparametric multivariate change point detection is contributed.

For Further Reading

Douglas M. Hawkins, Peihua Qiu and Chang Wook Kang
The changepoint model for statistical process control
Journal of Quality Technology, 2003

Mark D. Holland and Douglas M. Hawkins
A control chart based on a nonparametric multivariate change-point model
Journal of Quality Technology, 2014

Gábor J. Székely and Maria L. Rizzo
Energy statistics: A class of statistics based on distances
Journal of Statistical Planning and Inference, 2013

David S. Matteson and Nicholas A. James
A nonparametric approach for multiple change point analysis of multivariate data
Journal of the American Statistical Association, 2014

Harry Joe
Dependence Modeling with Copulas
Chapman & Hall/CRC, 2014

A Nonparametric Control Chart for High Dimensional Financial Surveillance

Ostap Okhrin
Ya-Fei Xu

Chair of Econometrics and Statistics
Institute for Transport and Economics
Faculty of Transport and Traffic Sciences
Technische Universität Dresden

Einführung

figures/olf10.png

Figure 13:

$$
\begin{aligned}
4x + 8 &= (3 - 2)^2 && (6) \\
4x &= -7 \\
x &= -\frac{7}{4} && (7)
\end{aligned}
$$

Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- Itemize environments

(i) First Roman point.

Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- Itemize environments
- can be uncovered or hidden

(i) First Roman point.

(ii) Second Roman point, uncovered on second slide.

Shown on second and third slide.

- Still shown on 2nd and 3rd slide.

Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- can be uncovered or hidden
- piecewise.

(i) First Roman point.
(ii) Second Roman point, uncovered on second slide.
(iii) Last Roman point.

Shown on second and third slide.

- Still shown on 2nd and 3rd slide.

- Shown on slides 3 and 5.

Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- can be uncovered or hidden
- piecewise.

(i) First Roman point.
(ii) Second Roman point, uncovered on second slide.
(iii) Last Roman point.

- Itemize environments

- Shown from slide 4 on.

Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- can be uncovered or hidden
- piecewise.

(i) First Roman point.
(ii) Second Roman point, uncovered on second slide.
(iii) Last Roman point.

- Itemize environments
- can be uncovered or hidden

- Shown from slide 4 on.
- Shown on slides 3 and 5.

Piecewise Uncovering I

The following example uses $< 1 - 2 >$ commands to piecewise hide and uncover text. $< 1 - 2 >$ makes the first item appear only on slides 1 and 2, $< 2- >$ has the second item visible from slide 2 onwards.

- can be uncovered or hidden
- piecewise.

(i) First Roman point.
(ii) Second Roman point, uncovered on second slide.
(iii) Last Roman point.

- Itemize environments
- can be uncovered or hidden
- piecewise.

- Shown from slide 4 on.