On a bayesian change point detection model for multivariate data

Third chapter: Real data applications

Chiara Bianchi, Lorenzo Dominoni, Piergiuseppe Pezzoli

Tutor: Riccardo Corradin

17th February 2021

POLITECNICO
MILANO 1863

Outline

1. Brief Recap
2. Real applications
   - Covid-19
   - Financial data
3. Conclusions

The Problem

- What?
  Identify changes in time series underlying structure
- Why?
  Modeling different behaviours over time, anomaly detection
- Where?
  Several applications in many fields: medical, financial, etc.

POLITECNICO MILANO 1863

The Model - Prior distribution

Prior distribution induced by the two-parameter Poisson-Dirichlet process restricted on the set of partitions that preserve time ordering

$$\mathbb{P}(\rho_n = (n_1, ..., n_k)) = \frac{n!}{k!} \frac{\prod_{i=1}^{k-1}(\theta + i\sigma)}{(\theta + 1)_{n-1\uparrow}} \prod_{j=1}^{k} \frac{(1-\sigma)_{n_j-1\uparrow}}{n_j!}$$

Prior distributions of hyperparameters

- $\sigma \sim Beta(a, b)$
- $\theta | \sigma \sim ShiftedGamma(c, d, -\sigma)$

POLITECNICO
MILANO 1863

The Model - Likelihood

Integrated Regime Likelihood given by an Ornstein-Uhlenbeck Process

$$\mathbb{P}(\underline{y}|\rho_n) = \prod_{j=1}^{k} \frac{k_{n_j}^{-\frac{d}{2}}}{k_0^{-\frac{d}{2}}} \cdot \frac{\pi^{-\frac{n_j d}{2}}}{(1-\gamma^2)^{\frac{n_j-1}{2}}} \cdot \frac{|\psi_0|^{\frac{\nu_0}{2}}}{\Gamma_d(\frac{\nu_0}{2})} \cdot \frac{\Gamma_d(\frac{\nu_{n_j}}{2})}{|\psi_{n_j}|^{\frac{\nu_{n_j}}{2}}}$$

$$\Gamma_d(a) = \pi^{\frac{d(d-1)}{4}} \prod_{j=1}^{d} \Gamma\left(a + \frac{1-j}{2}\right)$$

$$\nu_{n_j} = \nu_0 + n_j$$

$$k_{n_j} = \left[k_0 + \frac{(1-\gamma)^2}{(1-\gamma^2)}(n_j - 1) + 1\right]$$

$$m_{n_j} = \frac{1}{k_{n_j}}\left[m_0 k_0 + y_1 + \frac{(1-\gamma)}{(1-\gamma^2)} \sum_{i=2}^{n_j} (y_i - \gamma y_{i-1})\right]$$

$$\psi_{n_j} = \left(\psi_0 + y_1 y_1^\top + \sum_{i=2}^{n_j} \frac{(y_i - \gamma y_{i-1})(y_i - \gamma y_{i-1})^\top}{(1-\gamma^2)} + k_0(m_0 m_0^\top) + k_{n_j}(m_{n_j} m_{n_j}^\top)\right)$$

POLITECNICO MILANO 1863

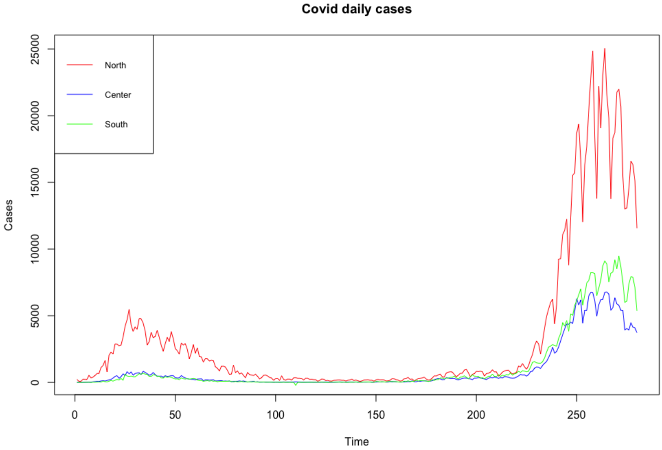
The Algorithm

Split and Merge MCMC algorithm to simulate from the posterior distribution

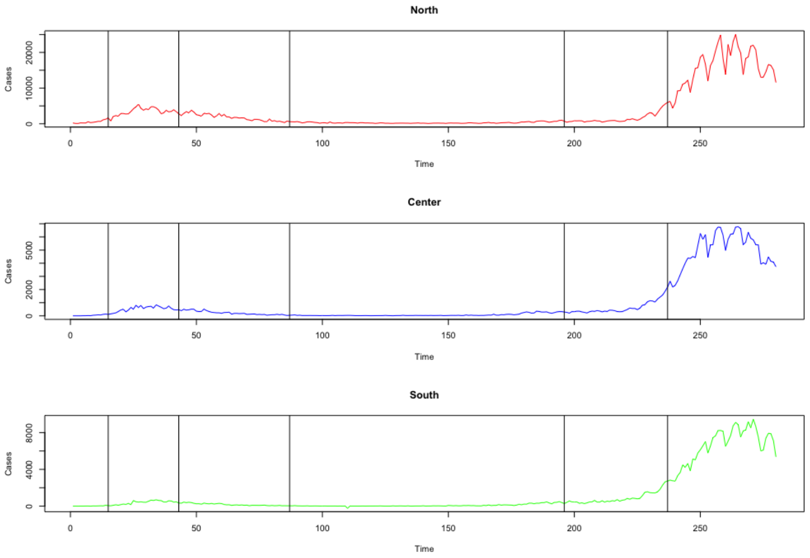
- At each step two possible choices are available:
  - Split: divide a group into two consecutive ones
  - Merge: combines two consecutive groups into a single one
- Shuffle: a random pair of adjacent groups is updated proposing a different change point
- Simulation of additional parameters with the ARMS method

Covid-19

- Covid-19 in North, Centre and South Italy
- New daily cases from 24/02/20 to 29/11/20

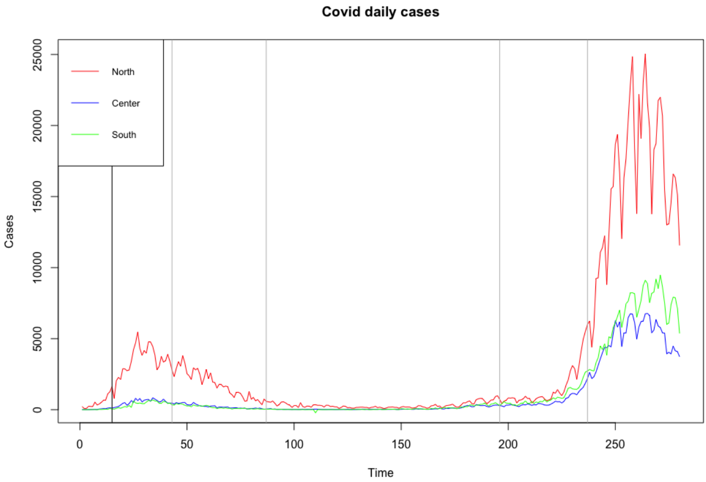


Change Points - Mode



Important events

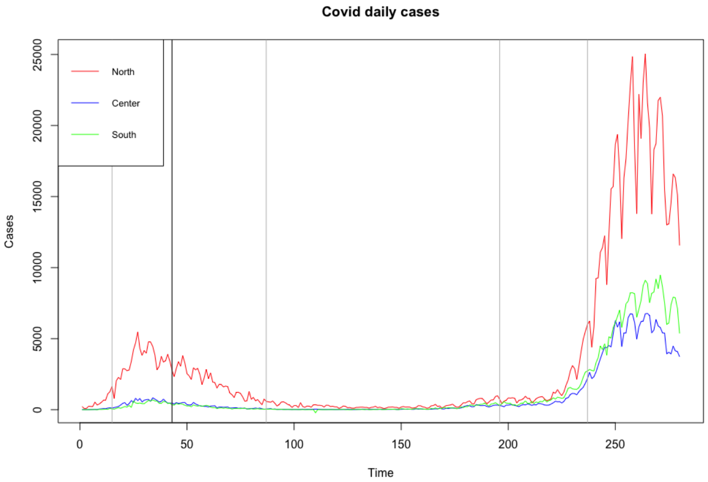
09/03/20 $\rightarrow$ Outbreak of Coronavirus



POLITECNICO MILANO 1863

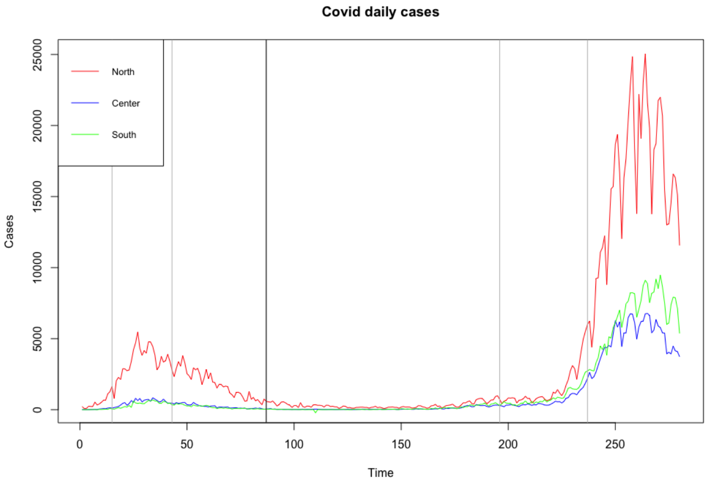
Important events

06/04/20 $\rightarrow$ Two weeks after lockdown DPCM



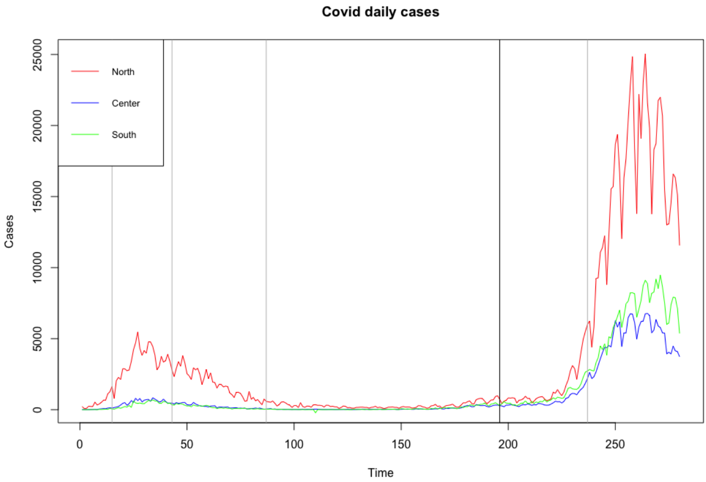
Important events

20/05/20 → Two weeks after phase-2 start



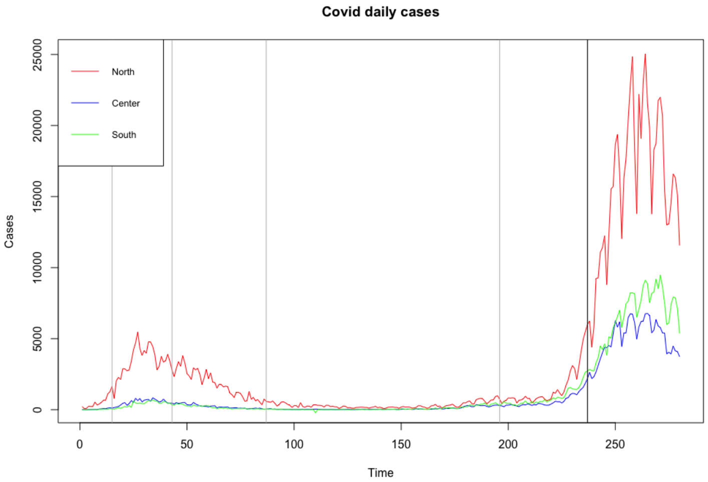
Important events

06/09/20 → Second outbreak after summer holidays

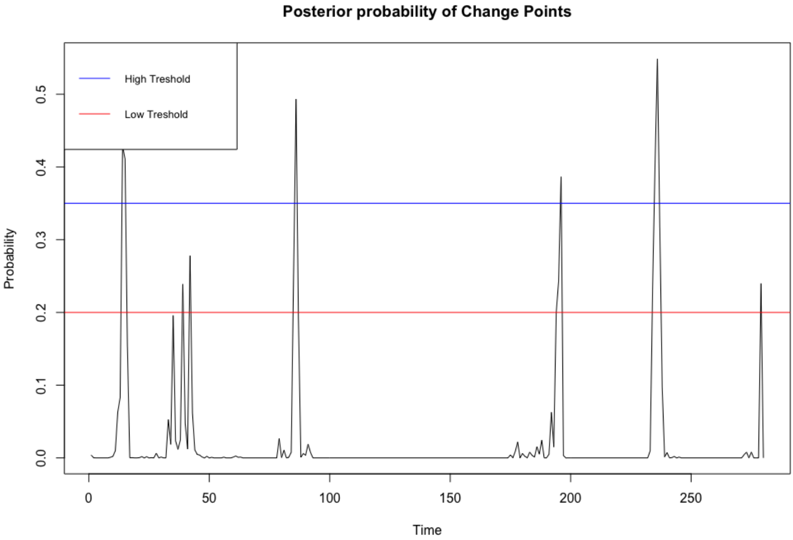


Important events

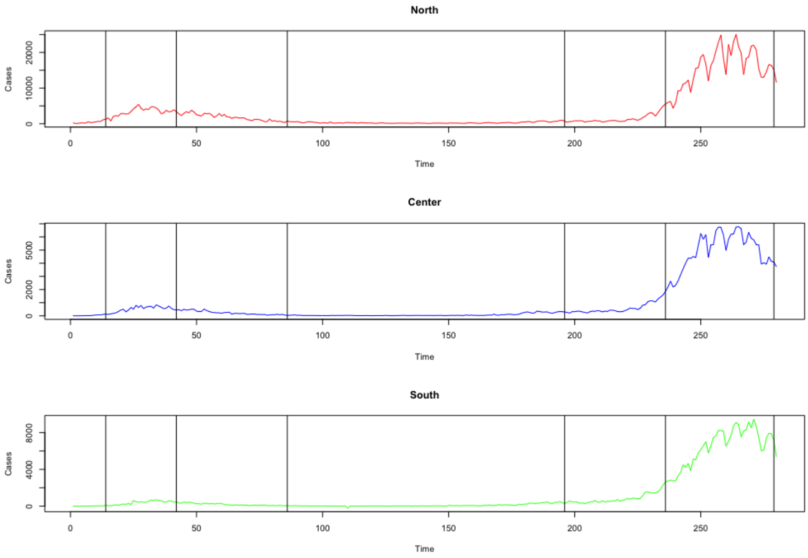
17/10/20 $\rightarrow$ Close of business



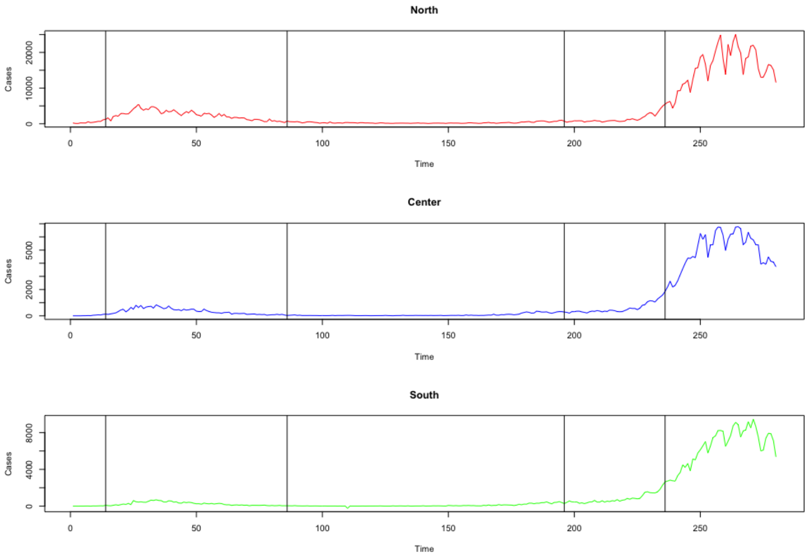
Posterior inference



Change Points - Low threshold

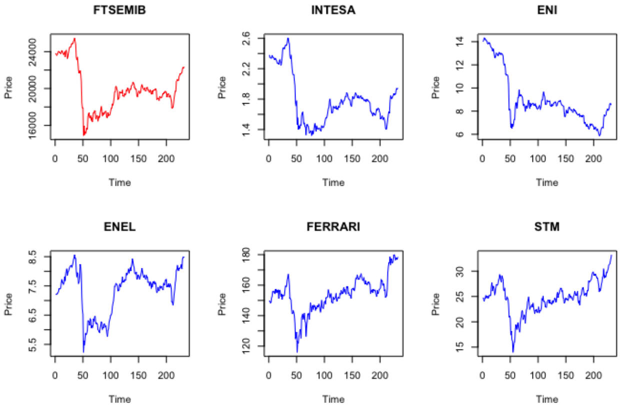


Change Points - High threshold

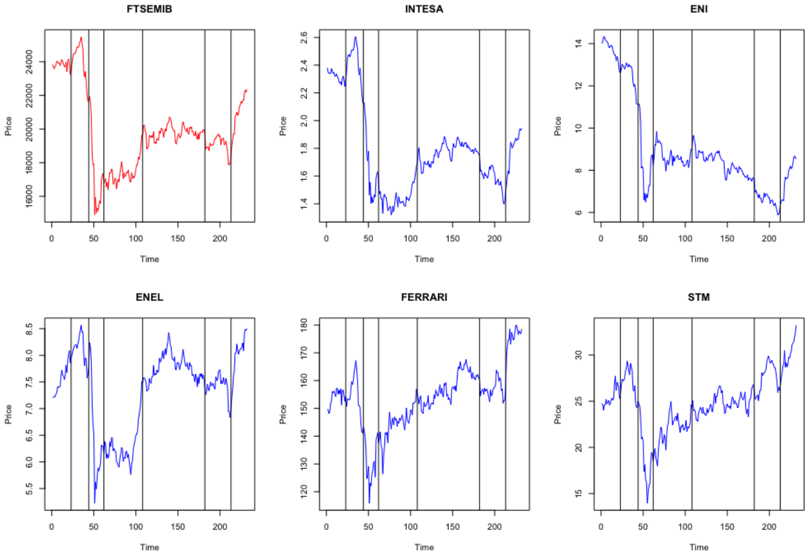


Financial data

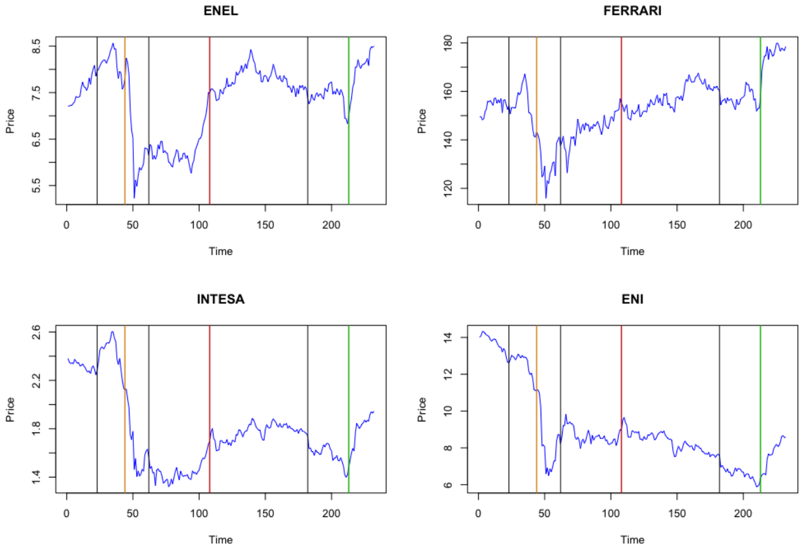
- Ftsemib, Enel, Intesa, Ferrari, Eni, STM
- Daily stock closing prices from 01/01/20 to 29/11/20



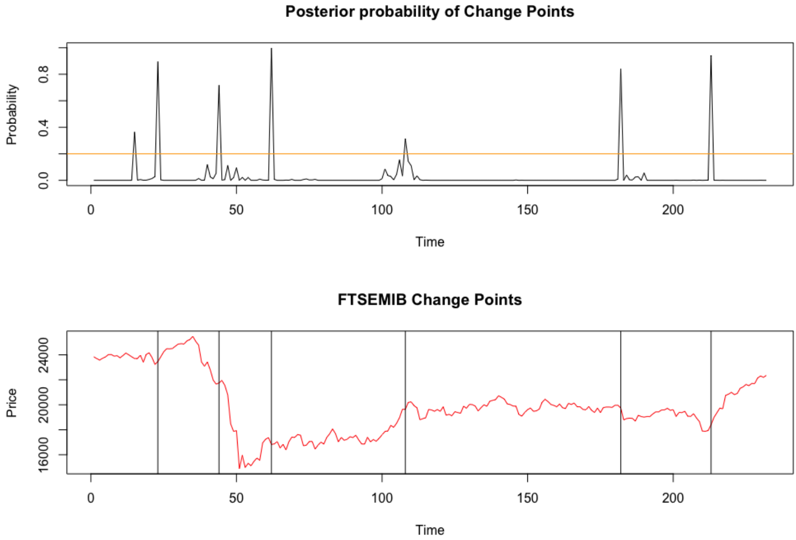
Change Points - Comparison between companies



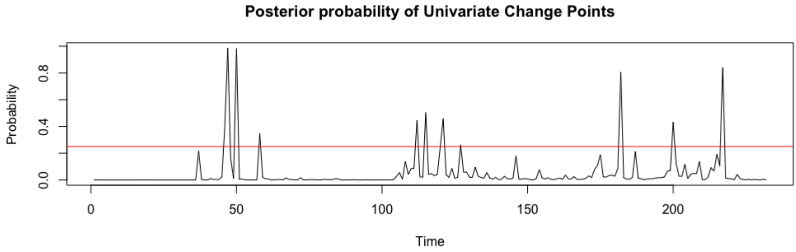
Focus on specific change points

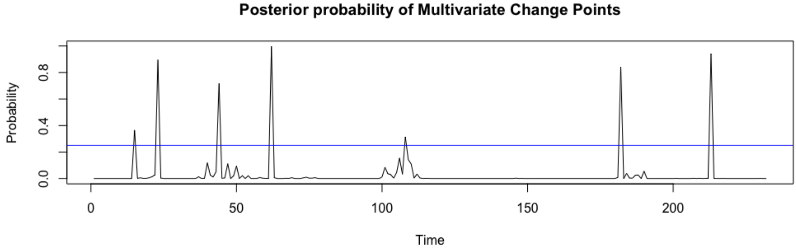


Change Points - Ftsemib

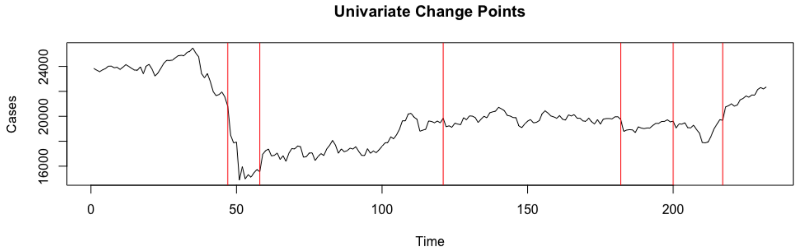


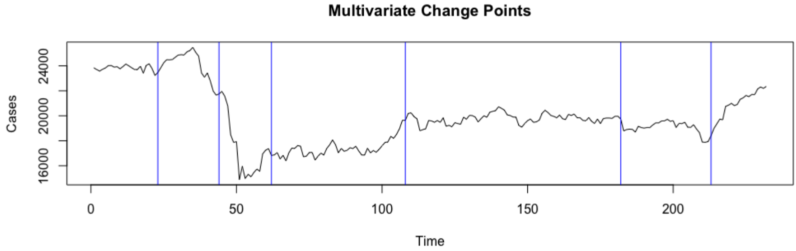
Comparison with univariate case





Comparison with univariate case





Conclusions

- Robustness with respect to parameters choice

Conclusions

- Robustness with respect to parameters choice
- Estimate flexibility and interpretation

Conclusions

- Robustness with respect to parameters choice
- Estimate flexibility and interpretation
- Correspondence with important events

Conclusions

- Robustness with respect to parameters choice
- Estimate flexibility and interpretation
- Correspondence with important events
- More informative with respect to univariate case

Conclusions

- Robustness with respect to parameters choice
- Estimate flexibility and interpretation
- Correspondence with important events
- More informative with respect to univariate case
- Computationally expensive

Conclusions

- Robustness with respect to parameters choice
- Estimate flexibility and interpretation
- Correspondence with important events
- More informative with respect to univariate case
- Computationally expensive
- Not suited for too high dimensional data

THANK YOU FOR YOUR ATTENTION

References

- Martinez, F., Mena, R. (2014, 04). On a nonparametric change point detection model in markovian regimes. Bayesian Analysis, TBA. $DOI : 10.1214/14 - BA878$
- Fuentes–Garcıa, R., Mena, R. H., Walker, S. G. (2010). A probability for classification based on the dirichlet process mixture model. Journal of Classification, 27(3), 389-403. $DOI : 10.1007/s00357 - 010 - 9061 - 9$
- Vatiwutipong, P., Phewchean, N. (2019, 07). Alternative way to derive the distribution of the multi-variate ornstein–uhlenbeck process. Advances in Difference Equations, 2019. $DOI : 10.1186/s13662 - 019 - 2214 - 1$

POLITECNICO
MILANO 1863