Some nonparametric methods for analyzing financial data
(StatFin Conference 2025)

Soudeep Deb

Indian Institute of Management Bangalore
Bannerghatta Road, Bengaluru 560069, India.
Email: soudeep@iimb.ac.in.

December 18, 2025

Outline

1. Introduction
2. Mathematical framework
3. Nonparametric methods in multivariate time series
4. Empirical findings in a bivariate financial data
5. Nonparametric method of changepoint detection
6. Empirical findings in changepoint detection
7. Concluding remarks

Acknowledgements

Archi Roy
Asst Prof, IIM Kozhikode.

Kunal Rai
PhD student, IIM Bangalore.

Moumanti Podder
Asst Prof, IISER Pune.

Itai Dattner
Professor, Univ of Haifa.

Outline

Big picture

- A new nonparametric technique for detecting structural break in a stochastic time series regression model. (Under review)

Big picture

- A new nonparametric technique for detecting structural break in a stochastic time series regression model. (Under review)
- Multivariate extension of nonparametric estimation for mean, volatility, quantiles in a similar regression model. (Work in progress)

Big picture

- A new nonparametric technique for detecting structural break in a stochastic time series regression model. (Under review)
- Multivariate extension of nonparametric estimation for mean, volatility, quantiles in a similar regression model. (Work in progress)
- Both are applied to financial datasets, with suitable theoretical guarantees.

Changepoint detection: Motivation 1

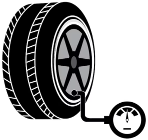
- A consumer research team at a major smartphone company launched a campaign to prevent users from overusing their phones.
- Notifications show only app icons instead of full messages during a trial run.
- Based on a trial run data, **did this campaign significantly impact phone usage behaviour**?

Changepoint detection: Motivation 1

- A consumer research team at a major smartphone company launched a campaign to prevent users from overusing their phones.
- Notifications show only app icons instead of full messages during a trial run.
- Based on a trial run data, **did this campaign significantly impact phone usage behaviour**?

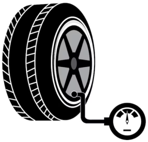
- $Y=$ screen-time, $X=$ proportion of non-essential notifications (ads).
- $Y = f(X) + \epsilon$, f is an unknown function and ϵ is random error.
- No significant change in X.
- **Is there a significant shift in the functional behavior of $f(.)$?**

Changepoint detection: Motivation 2



- A car tire manufacturing company starts receiving multiple customer complaints about inconsistent product quality.
- The company has not changed the machine settings or production protocols.
- Recent quality control reports confirm that the average quality of the sampled tires still meets the required standards.
- **Has there been a change in the consistency in the quality of the manufacturing process?**

Changepoint detection: Motivation 2



- A car tire manufacturing company starts receiving multiple customer complaints about inconsistent product quality.
- The company has not changed the machine settings or production protocols.
- Recent quality control reports confirm that the average quality of the sampled tires still meets the required standards.
- **Has there been a change in the consistency in the quality of the manufacturing process?**

- $Y=$ product quality, $X=$ operational machine setting.
- Since the average quality remains unchanged, the degradation may be due to increased variability in the quality of the product.
- $Y = g(X)\epsilon$, g is an unknown function and ϵ is random error.
- No significant change in X.
- **Is there a significant shift in the functional behavior of $g(.)$?**

Changepoint detection: Motivation 3

- An investor bases decisions on media attention toward Bitcoin, given its speculative nature.
- Recently the government imposed some regulatory restrictions to make trading more transparent.
- Based on Bitcoin returns post-restriction, **should the investor change his strategy**?

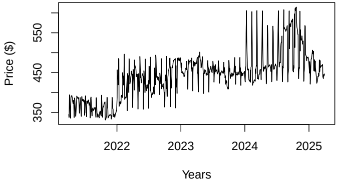
Changepoint detection: Motivation 3

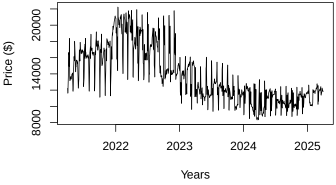
- An investor bases decisions on media attention toward Bitcoin, given its speculative nature.
- Recently the government imposed some regulatory restrictions to make trading more transparent.
- Based on Bitcoin returns post-restriction, **should the investor change his strategy**?

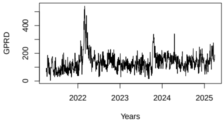
- $Y=$ Bitcoin returns, $X=$ public attention in the market
- $Y = f(X) + g(X)\epsilon$, f and g are unknown functions, ϵ is random error.
- No significant change in X.
- **Is there a significant shift in the behavior of $f(.)$ and/or $g(.)$**?

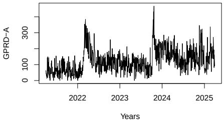
Multivariate analysis: Motivation

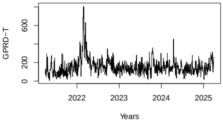
Consider the market behavior of Lockheed Martin (USA-based defense company), and Maersk (EU-based supply chain services). How do they perform under different geo-political risks (Caldara and Iacoviello, 2022)?











Outline

A general stochastic time series regression model

- Let $\{(Y_t, \boldsymbol{X}_t)\}$ be a time series data. Assume $\boldsymbol{X}_t$ has no changepoints.
- Consider the model

$$Y_t = \mu(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\,\epsilon_t$$

where $\mu(\cdot)$ are $\sigma(\cdot)$ are functions of unknown form.

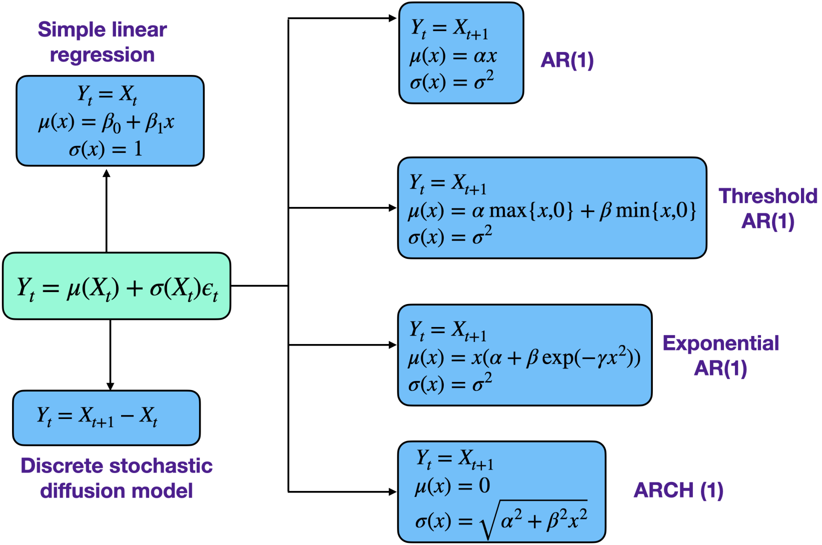
A general stochastic time series regression model

- Let $\{(Y_t, \boldsymbol{X}_t)\}$ be a time series data. Assume $\boldsymbol{X}_t$ has no changepoints.
- Consider the model

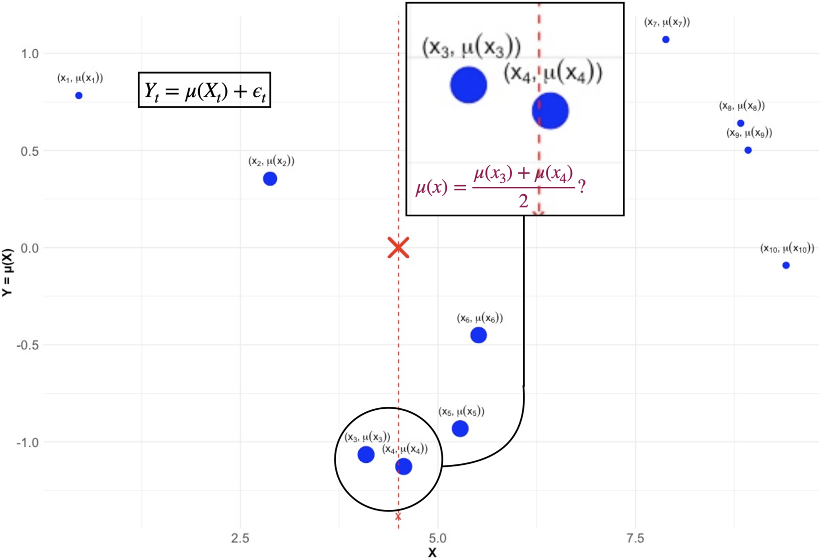
$$Y_t = \mu(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\,\epsilon_t$$

  where $\mu(\cdot)$ are $\sigma(\cdot)$ are functions of unknown form.
- Can we estimate $\mu(\cdot)$ and $\sigma(\cdot)$ with suitable theoretical properties?

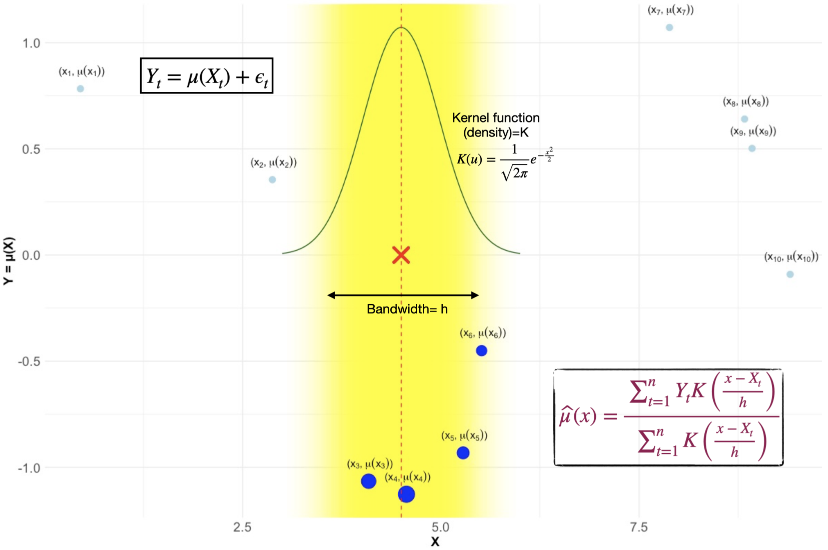
One model for all



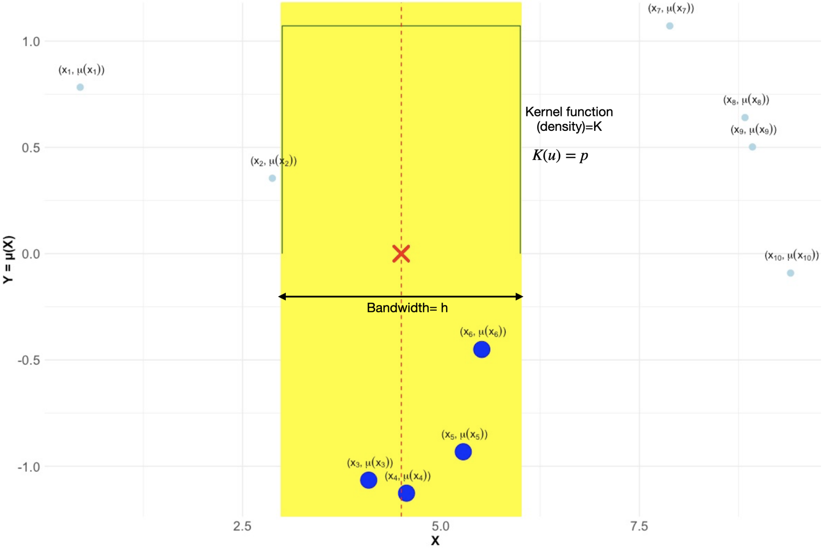
A quick introduction to nonparametric estimation



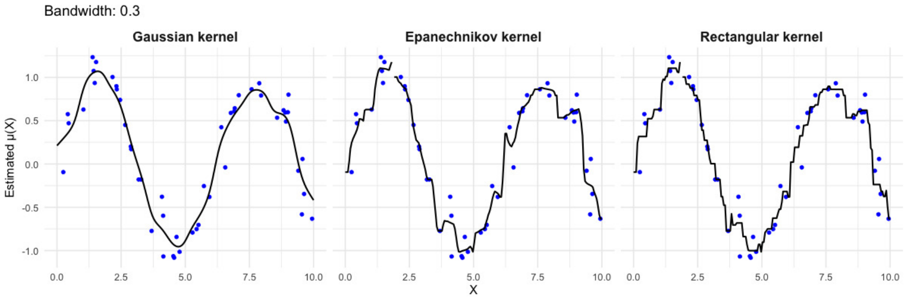
A quick introduction to nonparametric estimation



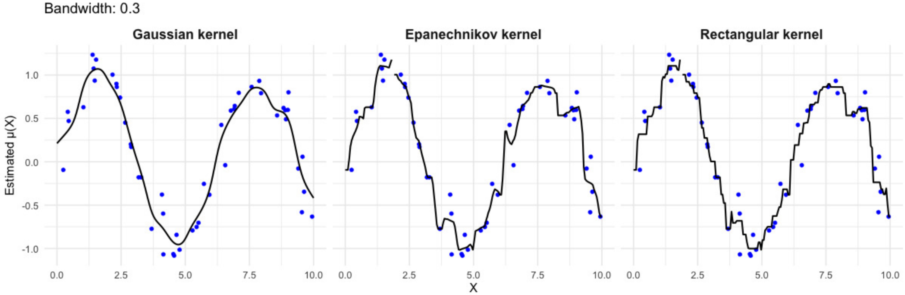
A quick introduction to nonparametric estimation



A quick introduction to nonparametric estimation



A quick introduction to nonparametric estimation



Fun fact:
The quality of the estimate is not strongly affected by the choice of the kernel, but it is highly sensitive to the choice of the bandwidth!

Outline

Multivariate framework

We consider the model:

$$\mathbf{Y}_t = \mu(\mathbf{X}_t) + \Sigma(\mathbf{X}_t)\, \boldsymbol{e}_t,\ \ t = 1, \ldots, n,$$

where $\{\mathbf{Y}_t\} \in \mathbb{R}^p$ represents a p-variate response variable, $\{\mathbf{X}_t\} \in \mathcal{X} \subseteq \mathbb{R}^k$ is a k-dimensional covariate vector, $\mathcal{X}$ is compact and $\{\boldsymbol{e}_t\} \in \mathbb{R}^p$ represents a multivariate stochastic independent and identically distributed noise. We keep $p < \infty$ and fixed, and let $n \to \infty$.

Estimation of multivariate mean and variability

The estimator for the multivariate mean and variance function are respectively given by

$$\widehat{\mu}(\boldsymbol{x}) = \sum_{t=1}^{n} \boldsymbol{Y}_t \omega_t(\boldsymbol{x}), \ \widehat{\Sigma}(\boldsymbol{x}) = \sum_{t=1}^{n} (\boldsymbol{Y}_t - \widehat{\mu}(\boldsymbol{X}_t))\, \omega_t(\boldsymbol{x})\, (\boldsymbol{Y}_t - \widehat{\mu}(\boldsymbol{X}_t))^{\intercal},$$

where

$$\omega_t(\boldsymbol{x}) = \frac{K_H(\boldsymbol{x} - \boldsymbol{X}_t)}{\sum_{t=1}^{n} K_H(\boldsymbol{x} - \boldsymbol{X}_t)}, K_H(\boldsymbol{v}) = |H|^{-\frac{1}{2}} K(H^{-1}\boldsymbol{v}),$$

with $K : \mathbb{R}^k \to \mathbb{R}$ an appropriately chosen kernel function and H as the bandwidth matrix satisfying $|H| \to 0$ and $n\,|H| \to \infty$ as $n \to \infty$.

Asymptotic theory (I)

Theorem 1: Consistency and normality of conditional mean estimate.

i) Let $\boldsymbol{x} \in \mathbb{R}^k$ be fixed, $K \in \mathcal{K}$, assume standard stationarity and smoothness conditions on $\boldsymbol{X}$, and let $\Sigma(\boldsymbol{x})$ be a positive definite matrix. Then, the mean estimate is consistent: $\widehat{\mu}(\boldsymbol{x}) \xrightarrow{\mathcal{P}} \mu(\boldsymbol{x})$.

ii) Further assume that

$$b_n + nb_n^{k+4} + \frac{1}{nb_n^k} + \Lambda_n \left(\frac{b_n^3 k}{n} + \frac{1}{n^2} \right) \to 0,$$

and let $\rho_\mu(\boldsymbol{x}) = \nabla^2 \mu(\boldsymbol{x}) + 2\nabla\mu(\boldsymbol{x}) \frac{\nabla f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})}$. Then as $n \to 0$,

$$\frac{\sqrt{nb_n^k \hat{f}_{\boldsymbol{X}}(\boldsymbol{x})}}{\sqrt{\phi_K}} \Sigma^{-1/2}(\boldsymbol{x}) \left[\widehat{\mu}_{b_n}(\boldsymbol{x}) - \mu(\boldsymbol{x}) - b_n^2 \psi_K \rho_\mu(\boldsymbol{x}) \right] \xrightarrow{\mathcal{D}} N_p(\boldsymbol{0}, \boldsymbol{I}).$$

Asymptotic theory (II)

Theorem 2: Consistency of conditional volatility estimate.

Let $\boldsymbol{x} \in \mathbb{R}^k$ be fixed, $K \in \mathcal{K}$ and under the assumptions of stationarity, smoothness ($\Sigma \in \mathcal{C}^4(\{\boldsymbol{x}^\epsilon\})$ for all $\epsilon > 0$), finite population variance, $f_{\boldsymbol{X}}(\boldsymbol{x}) > 0$ and $\Sigma(\boldsymbol{x})$ being a positive definite matrix, the sample estimate converges to the population estimate in probability,

$$\widehat{\Sigma}(\boldsymbol{x}) \xrightarrow{\mathcal{P}} \Sigma(\boldsymbol{x}).$$

Multivariate (geometric) quantiles

Multivariate (geometric) quantiles

Motivation (1D): The median solves

$$m = \arg\min_{\theta \in \mathbb{R}} \mathbb{E}\,|Y - \theta|.$$

Quantiles can be written as minimizers of an *asymmetric* absolute loss.

Multivariate (geometric) quantiles

Motivation (1D): The median solves

$$m = \arg\min_{\theta \in \mathbb{R}} \mathbb{E}\,|Y - \theta|.$$

Quantiles can be written as minimizers of an *asymmetric* absolute loss.

Geometric quantile in $\mathbb{R}^d$: Following the definition of Chowdhury and Chaudhuri (2019), for a direction vector $u \in \mathbb{R}^d$ with $\|u\| < 1$, define

$$q(u) = \arg\min_{\theta \in \mathbb{R}^d} \mathbb{E}\Big[\|Y - \theta\| + u^\top (Y - \theta)\Big].$$

Multivariate (geometric) quantiles

Special cases & interpretation

- $u = 0$ gives the spatial median (a multivariate "center"):

$$q(0) = \arg\min_{\theta} \mathbb{E}\|Y - \theta\|.$$

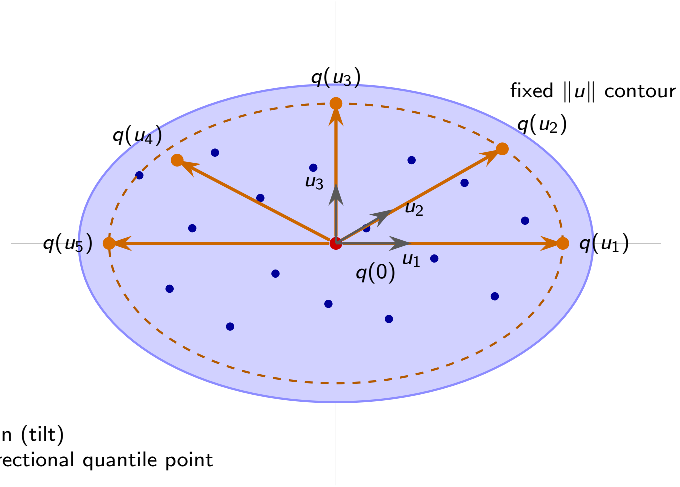
- u sets a directional tilt: $q(u)$ moves toward the "tail" in direction u.
- For fixed $\|u\|$, the set $\{q(u)\}$ traces a quantile contour.

First-order condition (intuition):

$$\mathbb{E}\left[\frac{Y - q(u)}{\|Y - q(u)\|}\right] = u,$$

i.e., the average outward unit direction from $q(u)$ matches the tilt u.

Geometric Quantiles in 2D: Direction-Indexed Points



Idea:
u sets direction (tilt)
$q(u)$ is the directional quantile point

Estimation of multivariate quantiles

- We aim to find

$$\hat{\boldsymbol{q}}_n(\boldsymbol{u}, \boldsymbol{x}) := \arg\min_{\boldsymbol{q} \in \mathbb{R}^p} \mathcal{M}^{(p)}_{\boldsymbol{u},n}(\boldsymbol{q}),$$

where

$$\mathcal{M}^{(p)}_{\boldsymbol{u},n}(\boldsymbol{q}) := \frac{1}{n} \sum_{t=1}^{n} \left\{ \| \boldsymbol{Y}_t - \boldsymbol{q} \|_{K(\cdot)} + \langle \boldsymbol{u} \mid \boldsymbol{Y}_t - \boldsymbol{q} \rangle_{K(\cdot)} \right\},$$

Estimation of multivariate quantiles

- We aim to find

$$\hat{\boldsymbol{q}}_n(\boldsymbol{u}, \boldsymbol{x}) := \arg \min_{\boldsymbol{q} \in \mathbb{R}^p} \mathcal{M}^{(p)}_{\boldsymbol{u},n}(\boldsymbol{q}),$$

  where

$$\mathcal{M}^{(p)}_{\boldsymbol{u},n}(\boldsymbol{q}) := \frac{1}{n} \sum_{t=1}^{n} \left\{ \| \boldsymbol{Y}_t - \boldsymbol{q} \|_{K(\cdot)} + \langle \boldsymbol{u} \mid \boldsymbol{Y}_t - \boldsymbol{q} \rangle_{K(\cdot)} \right\},$$

- Solution is obtained through an IRLS algorithm:

$$\hat{\boldsymbol{q}}_n^{(k+1)}(\boldsymbol{u}, \boldsymbol{x}) = \frac{\frac{1}{2} \sum_{t=1}^{n} K(\boldsymbol{X}_t - \boldsymbol{x}) \boldsymbol{u} + \sum_{t=1}^{n} w_t(\hat{\boldsymbol{q}}^{(k)}) K(\boldsymbol{X}_t - \boldsymbol{x})^2 \boldsymbol{Y}_t}{\sum_{t=1}^{n} w_t(\hat{\boldsymbol{q}}^{(k)}) K(\boldsymbol{X}_t - \boldsymbol{x})^2},$$

  where $w_t(\boldsymbol{q}^{(k)}) = \| \boldsymbol{Y}_t - \hat{\boldsymbol{q}}_n^{(k)}(\boldsymbol{u} \mid \boldsymbol{x}) \|^{-1}_{K(\boldsymbol{X}_t - \boldsymbol{x})}$ is a sequence of weights.

Asymptotic theory (III)

Theorem 3: Convergence of IRLS.

The IRLS scheme described before is guaranteed to reduce the value of the objective function at each iteration, thereby confirming that the algorithm converges to the global solution as the number of iterations increases.

Theorem 5: Consistency of conditional quantile estimate.

The estimator $\hat{q}_n(\boldsymbol{u}, \boldsymbol{x})$ of the population quantile $Q(\boldsymbol{u}, \boldsymbol{x})$ satisfies

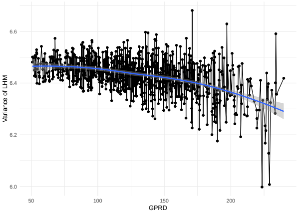
$$|| \hat{q}_n(\boldsymbol{u}, \boldsymbol{x}) - Q(\boldsymbol{u}, \boldsymbol{x}) || = \mathcal{O}_{\mathbb{P}}(n^{-1/2}), \forall\, \boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^d.$$

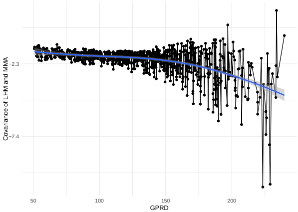
where $\mathcal{X}$ is compact and $\int_{\mathcal{X}} \|\boldsymbol{u}(s)\|^2 ds \leqslant 1$.

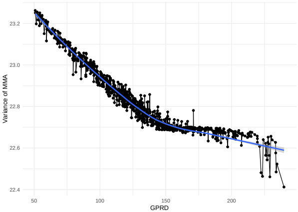
Outline

Volatility analysis

Volatility of Lockheed Martin (USA based defense company), and Maersk (EU-based logistics and supply chain services company) under different levels of geo-political risks (GPR):

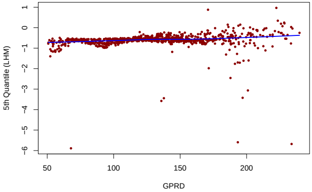


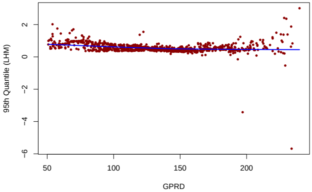




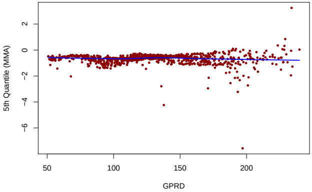
Quantile analysis

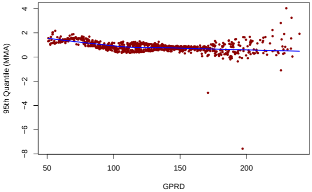
5th and 95th quantile of LHM:





5th and 95th quantile of MMA:





Outline

Formulating the problem

- Let $\{(Y_t, X_t)\}$ be a time series data. Assume X_t has no changepoints.
- Consider the model

$$Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t$$

where $\mu(.)$ are $\sigma(.)$ are functions of unknown form.

Formulating the problem

- Let $\{(Y_t, X_t)\}$ be a time series data. Assume X_t has no changepoints.
- Consider the model

$$Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t$$

where $\mu(.)$ are $\sigma(.)$ are functions of unknown form.

Detect significant changes in $\{Y_t\}$ by analyzing changes in:

- Functional behavior of $\mu(\cdot)$ (conditional mean)
- Functional behavior of $\sigma(\cdot)$ (conditional variance)

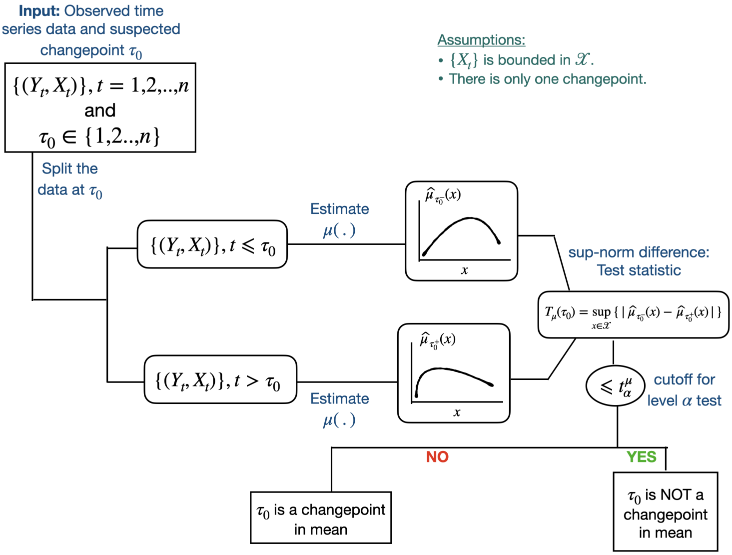
What is a "significant shift/change"?

- Given a time series $\{Z_t\}, t = 1, \ldots, n$, a changepoint is the time stamp at which an abrupt change happens in some characteristic of the process generating the data.
- Example: For some $\tau_0 \in \{1, \ldots, n\}$, if $E(Z_t) = \zeta_1, t \leqslant \tau_0$ and $E(Z_t) = \zeta_2, t > \tau_0$ then τ_0 is a changepoint in mean.
- Change in mean/variance/dependence structure etc.
- Can be induced by a "major" change in the environment (policy implementation, economic recession etc.) or some unidentified cause.

Finding the changepoint in steps

1. If the data contains just one change, can we detect it, i.e. **can we tell whether an abrupt shift has occurred at all**?
2. If we do conclude that one change exists in the data, can we pinpoint its location, i.e. **can we tell exactly when that change occurred**?
3. **If there is more than one change in the data, can we pinpoint all of them**?

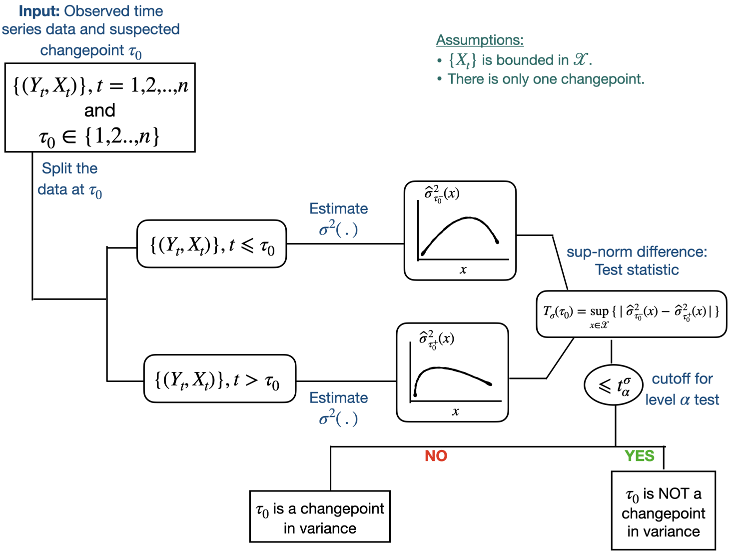
How to detect if there is a change in $\mu(\cdot)$



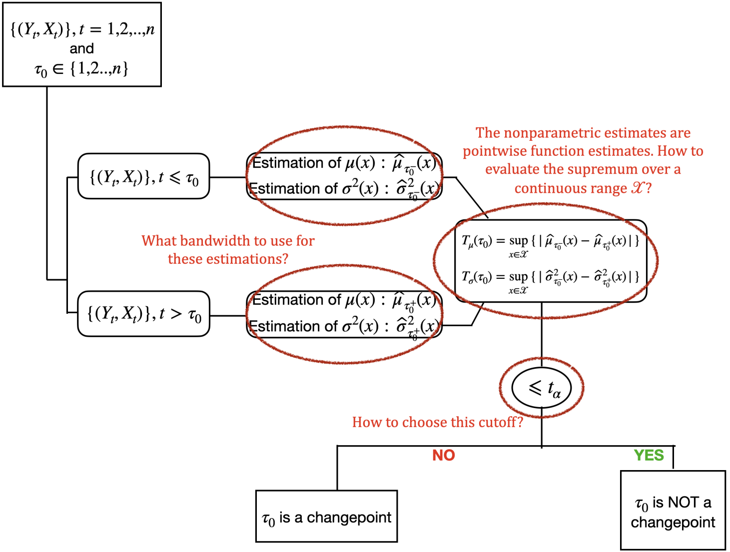
How to detect if there is a change in $\sigma(\cdot)$

Assumptions:

- $\{X_t\}$ is bounded in $\mathscr{X}$.
- There is only one changepoint.

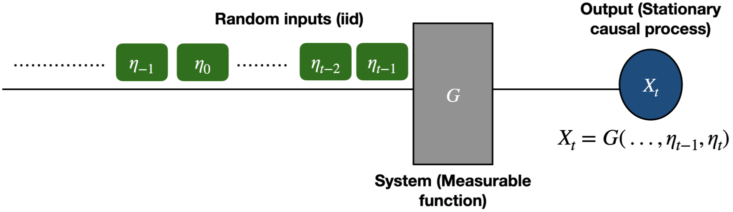


What's the catch?



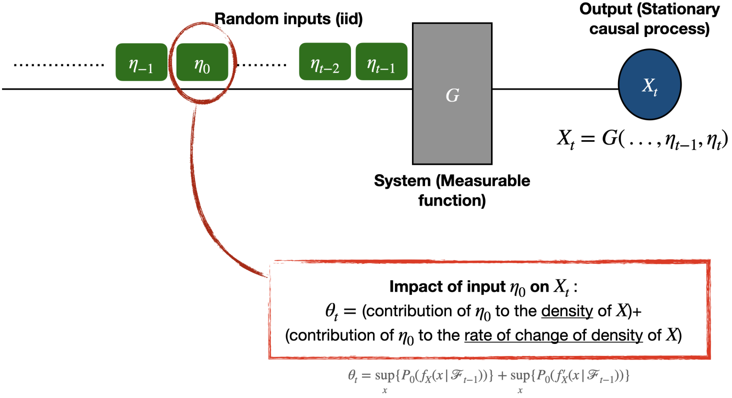
Choosing the bandwidth: Memory matters!

- Assume that $\{X_t\}$ is a **stationary causal** process.



Choosing the bandwidth: Memory matters!

- Assume that $\{X_t\}$ is a **stationary causal** process.



Choosing the bandwidth: Memory matters!

- Assume that $\{X_t\}$ is a **stationary causal** process.
- θ_t measures the impact of input η_0 on X_t.
- Define $\Theta_n = \sum_{t=1}^{n} \theta_t$. If $\Theta_n < \infty$, it is called short-range dependence (SRD), otherwise long-range dependence (LRD).
- SRD processes: $\Theta_n = O(n)$.
- LRD process with $\theta_t = t^{-\beta} l(t), \beta < 1$: $\Theta_n = O(n^{3-2\beta} l^2(n))$.

Choosing the bandwidth

How to choose the bandwidth given the sample size n and the dependence structure of $\{X_t\}$?

For a sample of size n,

$$n\mathbf{h}^9 \log n + \frac{(\log n)^3}{n\mathbf{h}^3} + \Theta_n \left\{ \frac{\mathbf{h}^3 \log n}{n} + \frac{(\log n)^2}{n^2 \mathbf{h}^{\frac{4}{3}}} \right\} \xrightarrow{n\to\infty} 0$$

- For SRD processes choose $h = O(n^{-\beta}), \beta \in (\frac{1}{9}, \frac{1}{3})$.
- For LRD processes with $\theta_t = t^{-\kappa} l(t)$,

$$h = O(n^{-\beta}), \beta \in \left(\max\left\{ \frac{1}{9}, \frac{2-2\kappa}{3} \right\}, \min\left\{ \frac{1}{3}, \frac{3(2\kappa-1)}{4} \right\} \right)$$

How to empirically evaluate the supremum?

- Assume $\{X_t\}$ is bounded in $\mathcal{X} = [\Lambda_1, \Lambda_2]$ for finite real $\Lambda_1 < \Lambda_2$.
- Define a sequence of partitions $\{\Pi_n\}$ on $\mathcal{X}$ as,

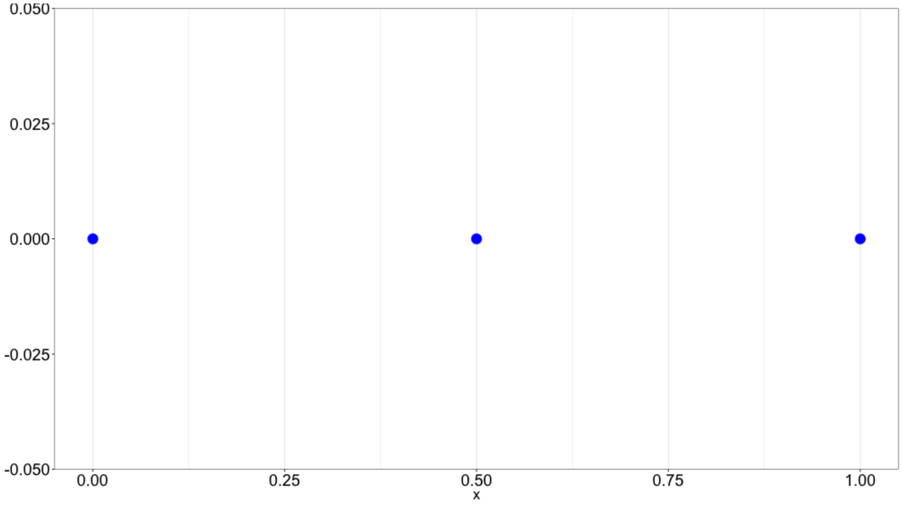
$$\Pi_n = \left\{x_{t_j} \mid x_{t_j} = \Lambda_1 + 2jh, j = 0, \ldots, m_n - 1\right\},\ m_n = \left\lceil \frac{\Lambda_2 - \Lambda_1}{2h} \right\rceil.$$

- Since $h \to 0$ as $n \to \infty$, under appropriate smoothness conditions on $\mu(\cdot)$ and $\sigma(\cdot)$:

$$\sup_{x\in\mathcal{X}}\{|\widehat{\mu}_{\tau_0^-}(x) - \widehat{\mu}_{\tau_0^+}(x)|\} = \lim_{n\to\infty} \sup_{x\in\Pi_n}\{|\widehat{\mu}_{\tau_0^-}(x) - \widehat{\mu}_{\tau_0^+}(x)|\}$$

$$\sup_{x\in\mathcal{X}}\{|\widehat{\sigma}^2_{\tau_0^-}(x) - \widehat{\sigma}^2_{\tau_0^+}(x)|\} = \lim_{n\to\infty} \sup_{x\in\Pi_n}\{|\widehat{\sigma}^2_{\tau_0^-}(x) - \widehat{\sigma}^2_{\tau_0^+}(x)|\}$$

How to empirically evaluate the supremum?



How to obtain the cutoff for a level α-test?

Theorem 3 (arXiv:2410.15173): Asymptotic distribution of test statistic for changepoint in $\mu(\cdot)$.

Assume that the functional behavior of the conditional variance $\sigma(\cdot)$ stays the same throughout the observed time domain. Take $\tau_0 = \frac{n}{2}$. Define

$$\mathcal{B}_r(p) = \sqrt{2\log p} - \frac{1}{\sqrt{2\log p}}\left[\log\log(p) + \log(2\sqrt{\pi})\right] + \frac{p}{\sqrt{2\log r}}$$

Then for any $z \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left[\frac{\sqrt{nh}}{\sqrt{\psi(K)}}\sup_{x\in\Pi_n}\left\{\frac{\sqrt{\widehat{f}_X(x)}}{\sqrt{\widehat{\sigma}^2(x)}}\left|\widehat{\mu}_{\tau_0^-}(x) - \widehat{\mu}_{\tau_0^+}(x)\right|\right\} \leqslant \mathcal{B}_{m_n}(z)\right] = e^{-2e^{-z}}$$

Double trouble? Just Bonferroni it!

Can we test if there is a changepoint in both conditional mean $\mu(\cdot)$ and variance $\sigma(\cdot)$?

Null Hypotheses:

- H_1 : There is no changepoint in $\mu(.)$.
- H_2 : There is no changepoint in $\sigma(.)$.

Steps for Bonferroni Correction at level α:

- Let p_1 and p_2 be the p-values for testing H_1 and H_2.
- Set adjusted thresholds:

$$\text{For } H_1,\ \text{compare } p_1 \text{ to } \alpha/2.$$

$$\text{For } H_2,\ \text{compare } p_2 \text{ to } \alpha/2.$$

Finding the changepoint in steps

1. If the data contains just one change, can we detect it, i.e. **can we tell whether an abrupt shift has occurred at all**?
2. If we do conclude that one change exists in the data, can we pinpoint its location, i.e. **can we tell exactly when that change occurred?**
3. **If there is more than one change in the data, can we pinpoint all of them**?

Detecting the location of detected change

- Define the changepoint estimates in $\mu(\cdot)$ and $\sigma^2(\cdot)$ respectively as

$$\widehat{\tau}_\mu = \arg\max_{\tau_0 \in \{1,2,\dots,n\}} \left\{ \sup_{x \in \Pi_n} \{ |\widehat{\mu}_{\tau_0^-}(x) - \widehat{\mu}_{\tau_0^+}(x)| \} \right\},$$

$$\widehat{\tau}_\sigma = \arg\max_{\tau_0 \in \{1,2,\dots,n\}} \left\{ \sup_{x \in \Pi_n} \{ |\widehat{\sigma}^2_{\tau_0^-}(x) - \widehat{\sigma}^2_{\tau_0^+}(x)| \} \right\}.$$

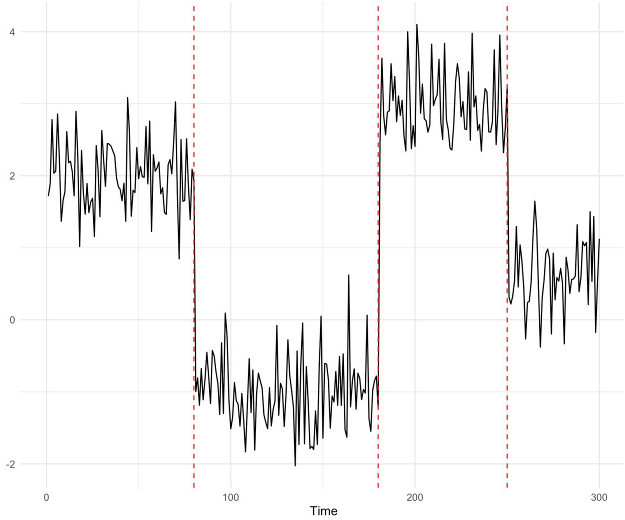
Theorem 6 (arXiv:2410.15173): Consistency of estimated changepoint.

Let τ_0^μ and τ_0^σ denote the true changepoints in $\mu(\cdot)$ and $\sigma^2(\cdot)$ respectively. Then, as $n \to \infty$, $\widehat{\tau}_\mu \xrightarrow{P} \tau_\mu$ and $\widehat{\tau}_\sigma \xrightarrow{P} \tau_\sigma$.

Finding the changepoint in steps

1. If the data contains just one change, can we detect it, i.e. **can we tell whether an abrupt shift has occurred at all**?
2. If we do conclude that one change exists in the data, can we pinpoint its location, i.e. **can we tell exactly when that change occurred**?
3. **If there is more than one change in the data, can we pinpoint all of them?**

Detection of multiple changepoints



Outline

Simulation for detection of changepoints

- $\{\mathrm{CP}_1, \mathrm{CP}_2, \ldots, \mathrm{CP}_m\}$ are the true changepoints.
- $\{\widehat{\tau}_1, \widehat{\tau}_2, \ldots, \widehat{\tau}_r\}$ are changepoints detected by our algorithm.
- Average minimum deviation measure (AMD): Mean (over all repetitions) of

$$\mathrm{MD} = \sum_{k=1}^{r} \min_{1 \leqslant i \leqslant m} \{|\widehat{\tau}_k - \mathrm{CP}_i|\},$$

- Average deviation in number (ADN): Mean (over all repetitions) of

$$\mathrm{DN} = |m - r|.$$

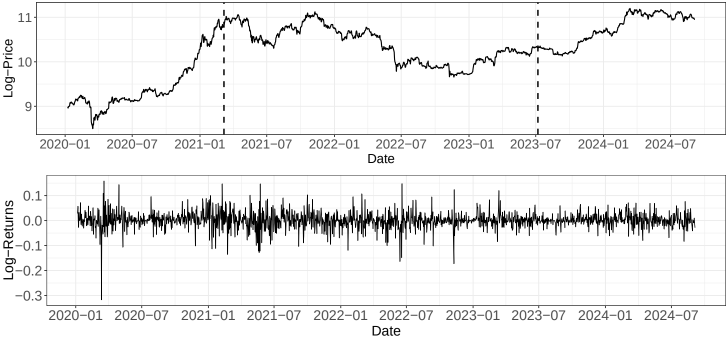
- Results are compared to PELT algorithm.

Simulation for detection of changepoints

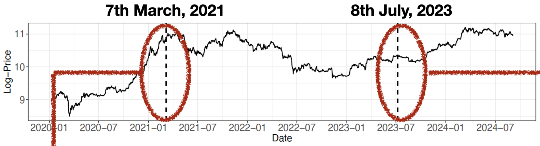
DGP	Noise	AMD		ADN	
		CPFind	**PELT**	**CPFind**	**PELT**
White Noise	$\mathcal{N}(0,1)$	57.14	67.30	1.06	1.76
White Noise	t_{10}	53.73	56.02	1.46	1.68
White Noise	Power law	24.94	91.73	1.71	1.42
ARMA-GARCH	$\mathcal{N}(0,1)$	146.88	62.57	1.10	3.58
ARMA-GARCH	t_{10}	120.00	109.77	0.69	3.54
ARMA-GARCH	Power law	115.65	113.92	0.82	3.46
TAR	$\mathcal{N}(0,1)$	72.00	68.38	1.62	5.34
TAR	t_{10}	47.31	76.23	1.88	5.26
TAR	Power law	21.24	68.02	1.70	4.96

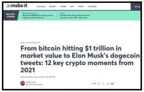
Application to Bitcoin returns data

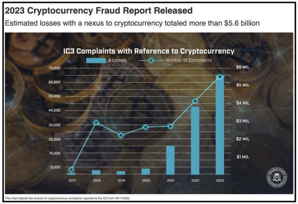
- Y_t =Bitcoin return on day t, X_t =GNIS score (frequency of Bitcoin-related searches on Google News) on day $(t-1)$.
- Changepoints found in $\mu(\cdot)$: **7th March 2021**, **8 July 2023**.
- No changepoints in the risk function $\sigma(\cdot)$.

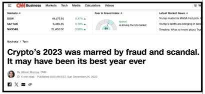


Application to Bitcoin returns data









Outline

Summary (I)

- We propose a **fully nonparametric structural-break detection** framework for regression time series.
- Designed for **minimal assumptions**: accommodates dependence in regressors and heavy-tailed noise typical in finance.
- Uses a **supremum-type, kernel-based discrepancy statistic** with supporting large-sample theory for inference.
- Provides a practical **multi-break detection algorithm** (with a confirmatory stage to reduce over-detection).
- Validated via simulations and a Bitcoin application, including comparisons against competing nonparametric approaches.

Summary (II)

- In the second work, we develop the theory for a **nonparametric multivariate location-scale** model capturing nonlinear drift and volatility/covariance dynamics.
- Theory accommodates dependence and heavy-tailed behavior relevant for high-risk financial data.
- Introducing **kernel-based multivariate conditional quantiles** with an IRLS algorithm and convergence properties. It does not suffer from quantile-crossing effects.
- We demonstrate the efficacy through an application linking geopolitical risk to asset behavior (mean/variance/extremes).

Some further questions

- Extend the changepoint detection algorithms to multivariate time series, possibly to a high-dimensional case as well.
- Spatiotemporal analysis through nonparametric techniques.
- Combine nonparametric methods with advances in deep learning (PINN) and apply in different applications.
- Develop the theory to the case of nonstationary $\{X_t\}$.
- Can we detect changepoints in higher-order conditional moment functions, or in terms of topological properties?

Selected references

- Caldara, D., Iacoviello, M. (2022). Measuring geopolitical risk. American economic review 112, 1194–1225.
- Chowdhury, J., Chaudhuri, P. (2019). Nonparametric depth and quantile regression for functional data. Bernoulli, 25(1), 395-423.
- **Deb, S.**, Jana, K. (2024) Nonparametric quantile regression for time series with replicated observations and its application to climate data. Statistical Science, 39(3), 428-448.
- **Deb, S.**, Neves, C., Roy, S. (2025+) Nonparametric quantile regression for spatio-temporal processes. `https://arxiv.org/abs/2405.13783`.
- Franke, J., Kreiss, J. P., Mammen, E. (2009). Nonparametric modeling in financial time Series. In Handbook of financial time series (pp. 927-952). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Roy, A., Podder, M., **Deb, S.** (2025+) Nonparametric method of structural break detection in stochastic time series regression model. `https://arxiv.org/abs/2410.15713`.
- Song, Y., Cai, C., Ma, D., Li, C. (2024). Modelling and forecasting high-frequency data with jumps based on a hybrid nonparametric regression and LSTM model. Expert Systems with Applications, 237, 121527.
- Zhao, Z. (2007). Parametric and nonparametric models and methods in financial econometrics. Statistics Surveys, 1, 1-42.

Thank you!

Contact: soudeep@iimb.ac.in
Webpage: https://soudeepd.github.io/