

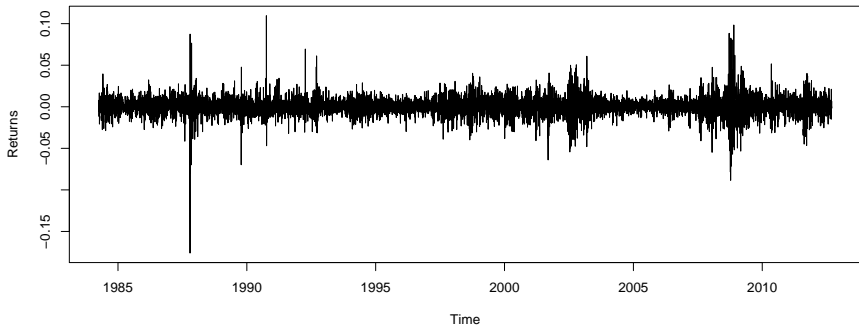
Financial Time Series: Changepoints, structural breaks, segmentations and other stories.

“City Lecture” hosted by NAG in partnership with CQF Institute and Fitch Learning

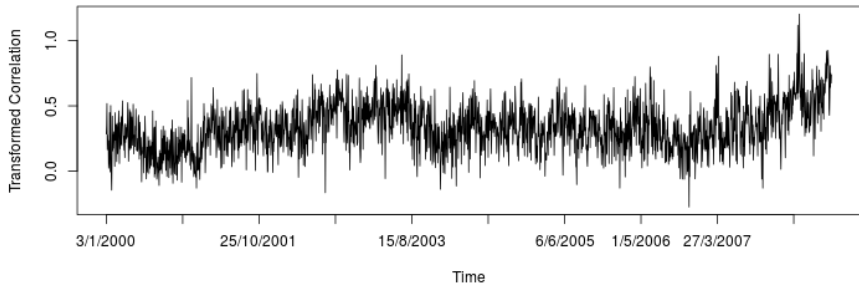
Rebecca Killick
r.killick@lancs.ac.uk
Feb 08, 2017

- Motivation
- Review existing search methods
- Introduce the PELT (Pruned Exact Linear Time) method
- Simulation Study
- Examples
- Extensions

Application Motivation



Cross correlations



What are changepoints?

Changepoints are also known as:

- breakpoints
- segmentation
- structural breaks
- regime switching
- detecting disorder

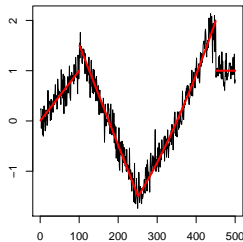
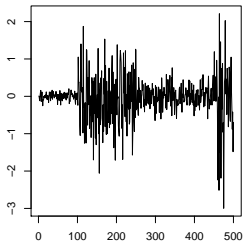
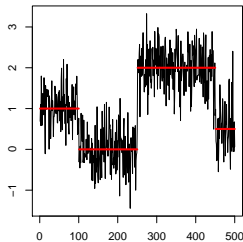
and can be found in a wide range of literature including

- quality control
- finance
- medicine
- environment
- linguistics
- ...

What are changepoints?

For data y_1, \dots, y_n , if a changepoint exists at τ , then y_1, \dots, y_τ differ from $y_{\tau+1}, \dots, y_n$ in some way.

There are many different types of change.

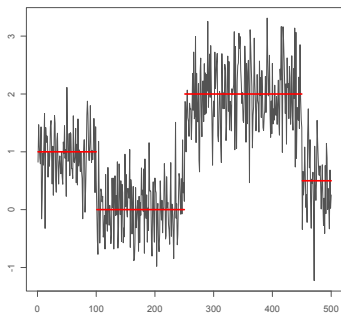


Example: Change in mean

Assume we have time-series data
where

$$Y_t | \theta_t \sim N(\theta_t, 1),$$

but where the means, θ_t , are
piecewise constant through time.



We want to infer the number and position of the points at which the mean changes. One approach:

Likelihood Ratio Test

To detect a single changepoint we can use the likelihood ratio test statistic:

$$LR = \max_{\tau} \{ \ell(y_{1:\tau}) + \ell(y_{\tau+1:n}) - \ell(y_{1:n}) \}.$$

We infer a changepoint if $LR > \beta$ for some (suitably chosen) β . If we infer a changepoint its position is estimated as

$$\tau = \arg \max \{ \ell(y_{1:\tau}) + \ell(y_{\tau+1:n}) - \ell(y_{1:n}) \}.$$

This can test can be repeatedly applied to new segments to find multiple changepoints.

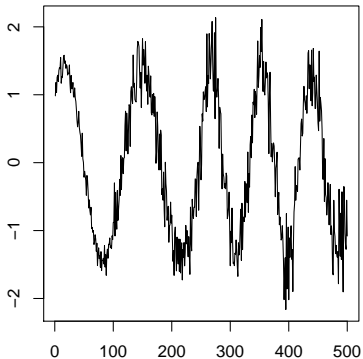
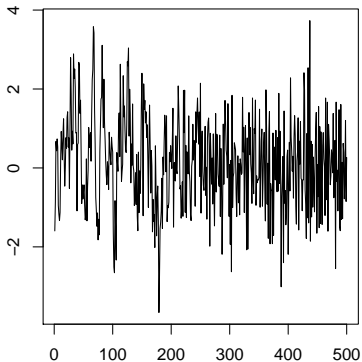
Define m to be the number of changepoints, with positions $\tau = (\tau_0, \tau_1, \dots, \tau_{m+1})$ where $\tau_0 = 0$ and $\tau_{m+1} = n$.

Then one application of the Likelihood ratio test can be viewed as

$$\min_{m \in \{0,1\}, \tau} \left\{ \sum_{i=1}^{m+1} [-\ell(y_{\tau_{i-1}:\tau_i})] + \beta m \right\}$$

Repeated application is thus aiming to minimise

$$\min_{m, \tau} \left\{ \sum_{i=1}^{m+1} [-\ell(y_{\tau_{i-1}:\tau_i})] + \beta m \right\}$$



The Challenge

- What are the values of τ_1, \dots, τ_m ?
- What is m ?

- What are the values of τ_1, \dots, τ_m ?
- What is m ?

- For n data points there are 2^{n-1} possible solutions

- If m is known there are still $\binom{n-1}{m-1}$ solutions
- If $n = 1000$ and $m = 10$, 2.634096×10^{21} solutions

- What are the values of τ_1, \dots, τ_m ?
- What is m ?

- For n data points there are 2^{n-1} possible solutions

- If m is known there are still $\binom{n-1}{m-1}$ solutions
- If $n = 1000$ and $m = 10$, 2.634096×10^{21} solutions

- How do we search the solution space efficiently?

Old methods are either *fast* but *approximate*.

Such as Binary Segmentation (Scott and Knott (1974)) which is $\mathcal{O}(n \log n)$ in CPU time.

Or they are *slower* but *exact*.

These methods used dynamic programming. For example, Segment Neighbourhood (Auger and Lawrence (1989)) is $\mathcal{O}(n^3)$.

For linear penalties $f(m) = m$, Optimal Partitioning (Jackson et al. (2005)) is $\mathcal{O}(n^2)$.

Consider $y_{1:2}$, either

1. There is no changepoint, or
2. There is a changepoint at y_1

Both scenarios are calculated and the optimal kept

Now consider $y_{1:3}$,

1. No changepoint
2. A changepoint at y_1
3. A changepoint at y_2
4. A changepoint at y_1 and y_2

But the decision between 3 and 4 has already been decided (at the previous iteration)!

The decision for $y_{1:3}$ becomes

1. No changepoint
2. Most recent changepoint at y_1 ,
i.e. a single change at y_1
3. Most recent changepoint at y_2 , and the optimal partition of $y_{1:2}$.

In a similar fashion, the decision for $y_{1:4}$ becomes

1. No changepoint
2. Most recent changepoint at y_1 ,
i.e. a single change at y_1
3. Most recent changepoint at y_2 , and the optimal partition of $y_{1:2}$.
4. Most recent changepoint at y_3 , and the optimal partition of $y_{1:3}$.

If we define

$$\mathcal{P}_t = \{\boldsymbol{\tau} : 0 < \tau_1 < \dots < \tau_m < t\}$$

$$F(t) = \min_{\boldsymbol{\tau} \in \mathcal{P}_t} \left\{ \sum_{i=1}^{m+1} \mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) + \beta \right\}$$

i.e. $f(m) = m$ in original minimisation.

So,

$$F(t) = \min_{\tau^*} \left\{ \min_{\boldsymbol{\tau} \in \mathcal{P}_{\tau^*}} \left[\sum_{i=1}^m \mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) + \beta \right] + \mathcal{C}(y_{(\tau^*+1):t}) + \beta \right\}$$

Thus we minimise,

$$F(t) = \min_{\tau^*} \left\{ F(\tau^*) + \mathcal{C}(y_{(\tau^*+1):t}) + \beta \right\}$$

Recursively solving the minimisation for $t = 1, \dots, n$ gives an algorithm that is $\mathcal{O}(n^2)$.

The PELT Method

(Pruned Exact Linear Time)

By eye there is often an obvious changepoint at (or by) a time-point s .

This means that for any $T > s$ the most recent changepoint cannot be at time $t < s$.

Thus we could prune the search step: and avoid searching over $t < s$.

ASSUMPTION: adding a changepoint reduces the overall cost

This means that for $t < s < T$:

$$C(y_{t+1:T}) \geq C(y_{t+1:s}) + C(y_{s+1:T})$$

This holds for costs based on the negative log-likelihood; and often can be made to hold for costs based on the negative log-marginal-likelihood.

Let $0 < t < s < T$

Theorem

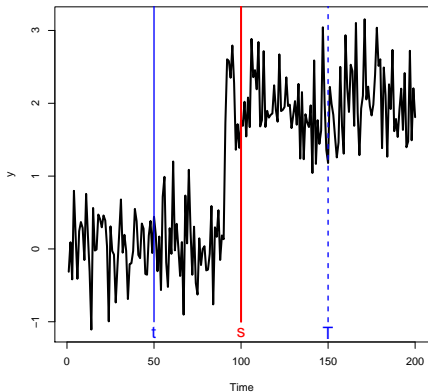
If

$$F(t) + C(y_{(t+1):s}) < F(s)$$

then at any future time $T > s$, t can never be the optimal last changepoint prior to T .

The condition in the theorem just means that for any $T > s$ the best partition which involves a changepoint at s will be better than one which have $[t, T]$ as a single segment.

Thus t can never be the (optimal) most recent changepoint prior to T for all $T > s$.





If many t are pruned, excluded from the minimisation then computational time will be drastically reduced.

We can prove that, under certain regularity conditions, the expected computational complexity will be $\mathcal{O}(n)$.

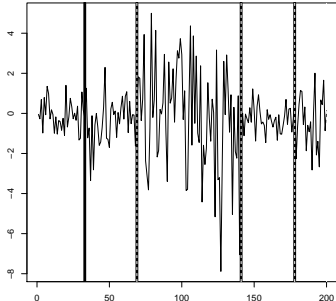
The most important condition is that *the number of changepoints increases linearly with n .*

This is natural in many applications: e.g. as you collect time-series data over larger time-periods; or genomic data over larger regions of the genome.

Simulation Study

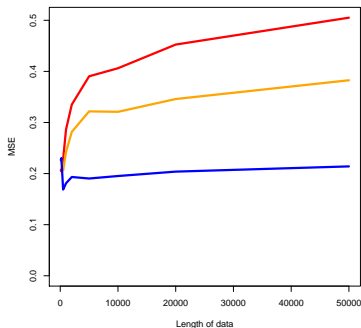
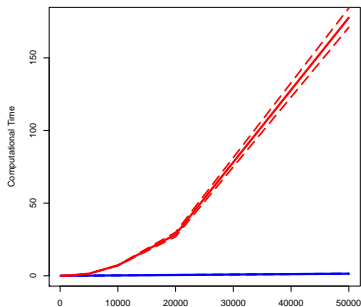
- 9 scenarios with lengths 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000
- Uniform distributed changepoints, subject to > 30 observations per segment
- Each scenario has 1,000 repetitions, $\frac{n}{50}$ changepoints

- Cost function: Negative log-likelihood
- Mean set to 0
- Variances simulated from a Log-Normal distribution



θ is a parameter that changes, i.e. the variance

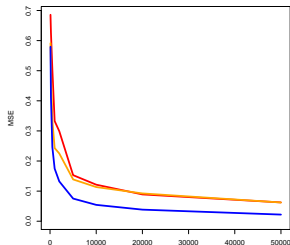
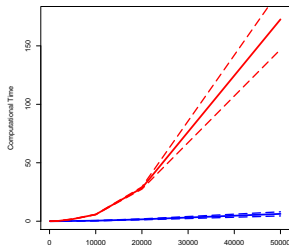
$$\text{MSE} = \frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2}{n}$$



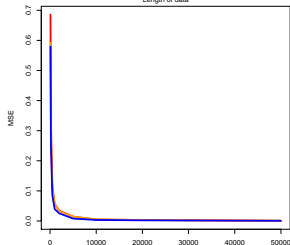
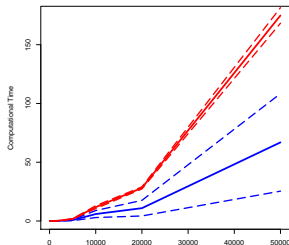
Breaking Assumptions

A key assumption for linear time is a linearly increasing number of changepoints.

Square Root



Fixed



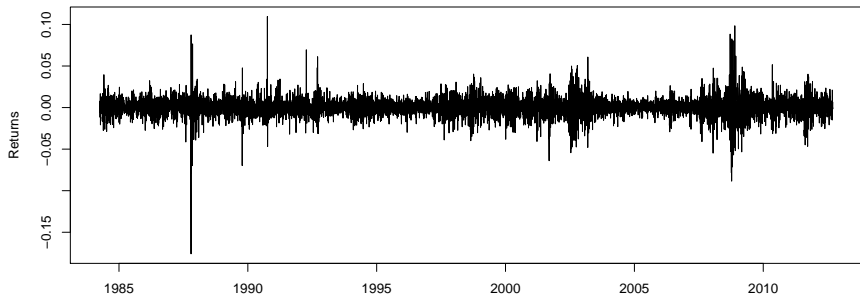
Applications

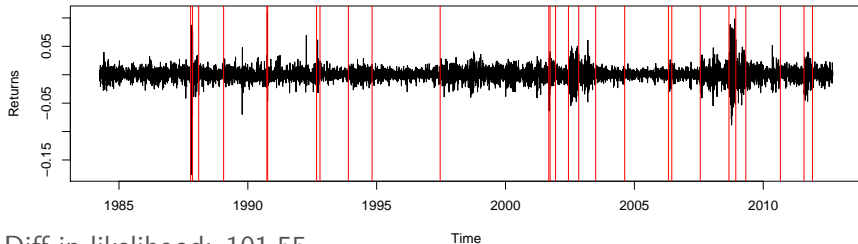
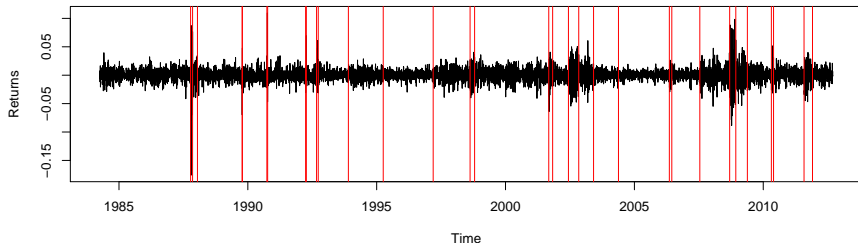
Data from Yahoo! Finance.

Daily returns for the history of the FTSE100 index

Return defined as

$$r_t = \frac{X_{t+1} - X_t}{X_t}$$

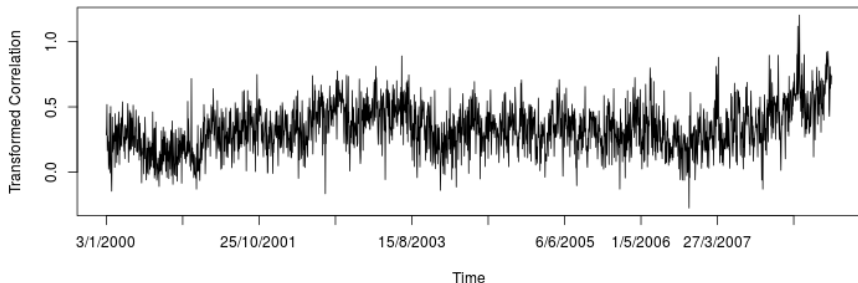


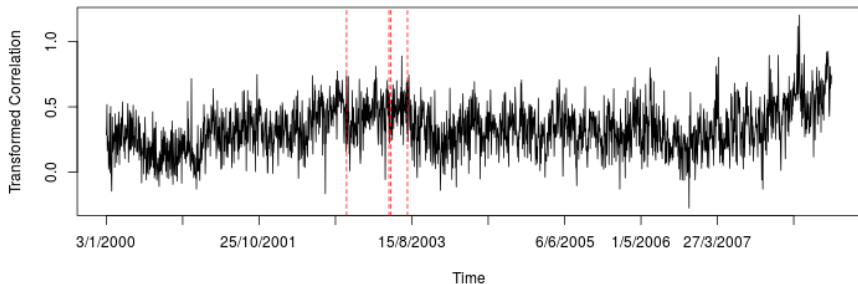


Diff in likelihood: 101.55

Data from Chiriac, Voev (2011)

Cross correlation of the Fourier transformed open to close returns for American Express and Home Depot.

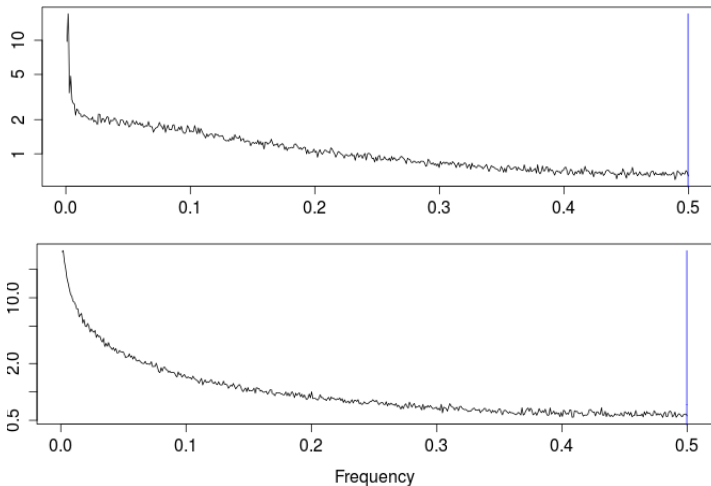




Changes at 15/12/2002, 20/04/2003, 26/04/2003 and 14/06/2003.

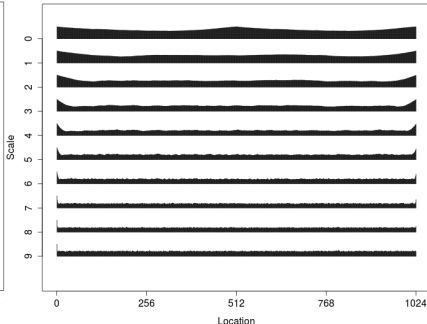
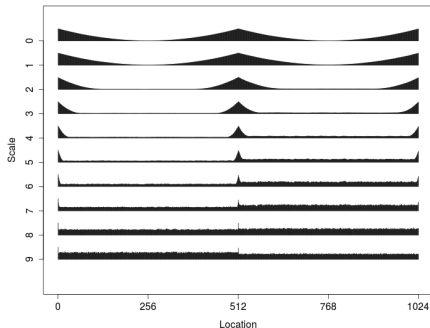
Changepoint or Long Memory?

Average periodograms of 1,000 changepoint and long memory processes



Changepoint or Long Memory?

Average wavelet periodograms of 1,000 changepoint and long memory processes



Changepoint or Long Memory?

Use a classification algorithm based on wavelet spectra to give decision.

Long Memory Model	Score	Changepoint Model	Score
ARFIMA(1, 0, 1) $\phi = 0.30$ $\theta = 0.58$ $d = 0.47, \quad \mu = 0.35$	12128	ARMA(4, 4) $\phi = (-0.05, -0.002, 0.07, 0.94)$ $\theta = (0.17, 0.14, 0.04, -0.83)$ $\tau = 715, \quad \mu = 0.31$	Segment 1
		AR(1) $\phi = 0.29, \quad \tau = 841, \quad \mu = 0.42$	Segment 2
		ARMA(2, 1) $\phi = (0.09, 0.09)$ $\theta = -0.58, \quad \tau = 847, \quad \mu = 0.30$	Segment 3
		MA(1) $\theta = -0.72, \quad \tau = 896, \quad \mu = 0.50$	Segment 4
		ARMA(3, 1) $\phi = (1.12, -0.06, -0.07)$ $\theta = -0.90, \quad \tau = 2156, \quad \mu = 0.35$	Segment 5
		5348047	

- Created the changepoint R package on CRAN
- Donated the code to NAG for wider reach

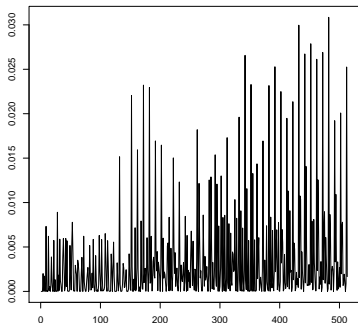
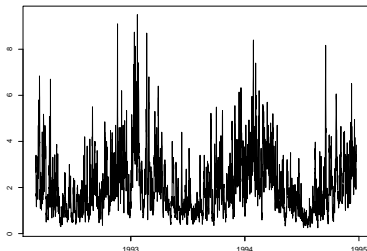
- Easy to donate
- Feedback of code improvements for changepoint
- Great for impact

- Being able to find changepoints quickly is important
- Existing methods were either inefficient or approximate
- PELT is $\mathcal{O}(n)$ under certain conditions and is exact
- Code is available in the NAG software
- Newer work is forthcoming to NAG, email me for pre-release.

- Time series are ubiquitous; used in many settings for forecasting, process control, environment.
- As data is collected over longer periods the statistical properties are changing
- There are models for describing the changing mean structure e.g. regression, functional data analysis.
- There are models for describing the changing variance / autocovariance structure e.g. GARCH, locally stationary (LS).

NAG have a variety of statistical routines but currently do not have coverage of nonstationary time series.

- Bring together the two worlds of changing mean and autocovariance structure.
- Create a model that can estimate both structures in one step.
- Implement the model in software to be added to the NAG libraries.



We compared the accuracy of PELT with a genetic algorithm approach (Davis et al.) for optimising under an MDL criteria.

The underlying model within each segment was $AR(p)$, with p unknown. Average Improvement in MDL for varying lengths of data.

n	1,000	2,000	5,000	10,000	20,000
Improvement	9	14	60	250	900

MDL fit for $AR(p)$ models

Realisation of a piecewise stationary autoregressive process.
Smoothed number of segments identified by PELT (thick line) and
Auto-PARM (thin line) algorithms for (b) $n = 5,000$ and (c) $n = 10,000$.

