Physique des marchés

Conditional finance

damien.challet@centralesupelec.fr

[https://s.42l.fr/PhM]

25th March 2025

Plan du cours

date	heure	sujet	salle
24.02	8 : 15	faits stylisés	MI.106
04.03	8 : 15	stratégies	sd.206
11.03	8 : 15	prévisibilité et stratégies	Amphi IV
25.03	8 : 15	prévisibilité et stratégies	Amphi III
01.04	8 : 15	modèles d'agents	sc.266
03.04	13 : 45	dynamique de la prévisibilité	mc.107
08.04	8 : 15	dynamique de la prévisibilité	sd.204
15.04	8 : 15	contrôle	Amphi I

Signal-to-noise ratio estimation

- N samples $r = \{r_i\}_{i=1,\cdots,N}$
- Signal

$$\hat{\mu} = \text{mean}(r)$$

- Noise

$$\hat{\sigma} = \text{std}(r)$$

- SNR

sacar el mu y sigma de manera separaada llevar a hacer un metodo recursivo. para obtener SNR es mejor estimarlos al mismo tiempo

$$SNR = \frac{\hat{\mu}}{\hat{\sigma}} \text{ or } \widehat{\left(\frac{\mu}{\sigma}\right)}$$

- t statistics, from Central Limit Theorem

$$\text{t-stat} = SNR\sqrt{N}$$

- Sharpe ratio, annualized

$$SR = SNR\sqrt{\text{number of periods in a year}}$$

Sharpe ratio

- Aim: to compare r with another asset returns R
- Fixed interest rate $R = \rho$

$$SR = \frac{\hat{\mu} - \rho}{\hat{\sigma}}$$

- Two stochastic processes (e.g. $R =$ index)

$$SR \equiv \text{statistical test of location differences}$$

- Robust method: studentized differences, Ledoit and Wolf (2008) [paper]

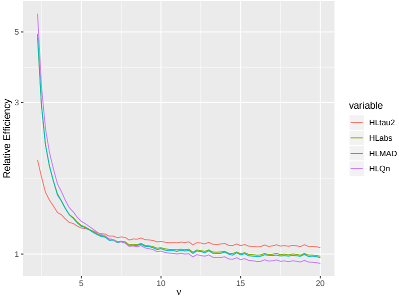
  for all data distributions

Robust estimators of Sharpe ratios

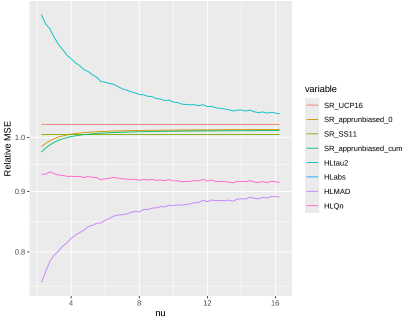
- Divide a robust location estimator by a robust scale estimator (Challet 2025)
    1. Hodges-Lehman pseudo-median $O(N^2)$
    2. Qn $[O(N^2)]$, τ^2 $[O(N)]$
- Total drawdown duration (Challet, 2017, 2025)

Robust location / robust scale

known tail



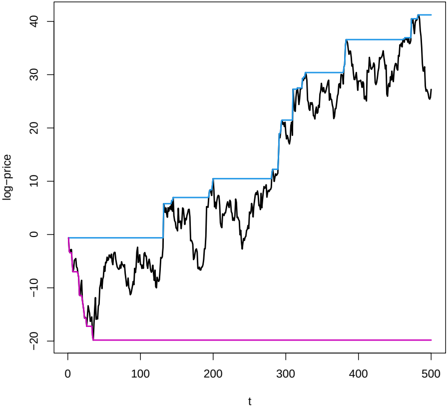
fitted tail parameter



Price records

- N i.i.d. increments r_n, $P(r)$
- Cumulative sum (price):

$$p_n = p_{n-1} + r_n$$



Records: universality

If $P(r) = P(-r)$: [Majumdar and Ziff (2008)]

- Number of records $R_\pm$

$$P(R_\pm)(N) = \frac{\binom{2N-R_\pm+1}{N}}{2^{2N-R_\pm+1}}$$
$$P(R_\pm) \simeq \mathcal{N}_+ \left[0, N\right] (R_\pm)$$
$$= \sqrt{\frac{2}{\pi N}} e^{-\frac{R_\pm^2}{2N}}$$

- Price improvement (classes)

- Age of records

Biased random walk

Gaussian case (Wergen *et al.* (2013))

$$P(r) = \mathcal{N}(\mu, \sigma^2)$$
$$P(x_n) = \mathcal{N}(n\mu, n\sigma^2)$$

$$E(R_\pm)(n, \frac{\mu}{\sigma}) \simeq 2\sqrt{\frac{n}{\pi}} + \frac{\mu}{\sigma}\frac{\sqrt{2}}{\pi}\left[n \arctan(\sqrt{n}) - \sqrt{n}\right] \quad \frac{\mu}{\sigma}\sqrt{n} \ll 1$$

Biased random walk

Student-t case (Challet 2017)

$$P(r-\mu)\propto\frac{1}{\left(r_0^2+r^2\right)^{\frac{\nu+1}{2}}}$$

$$P(x_n):\tilde{}\text{convolutions for } \nu=3$$

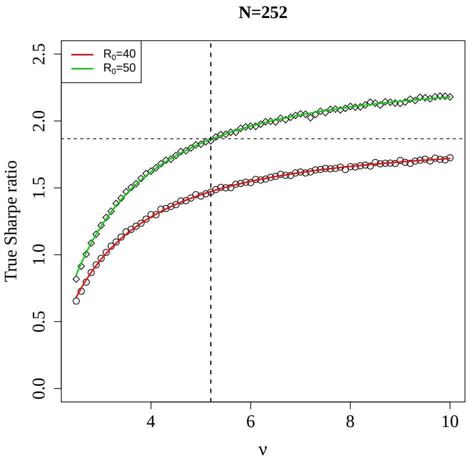
$\nu=3$, $\frac{\mu}{\sigma}\sqrt{n}\ll 1$

$$E_{\nu=3}(R_+)(n,\frac{\mu}{\sigma})\simeq\frac{2\sqrt{n}}{\sqrt{\pi}}+\frac{\mu}{\sigma}\frac{\sqrt{2}}{\pi}\left[n\arctan(\sqrt{n})-\sqrt{n}\right]$$
$$+\frac{\mu}{\sigma}\frac{\sqrt{2}}{\pi}\frac{8}{\sqrt{3}\pi^{3/2}}\sqrt{n}[\text{arctanh}\sqrt{1-\frac{1}{n}}-\sqrt{1-\frac{1}{n}}]$$

Gaussian vs Student

At *fixed* $R_0 = R_+ - R_-$

$$SR(\nu, n) = SR_{Gaussian} - b\nu^{-3/2}$$

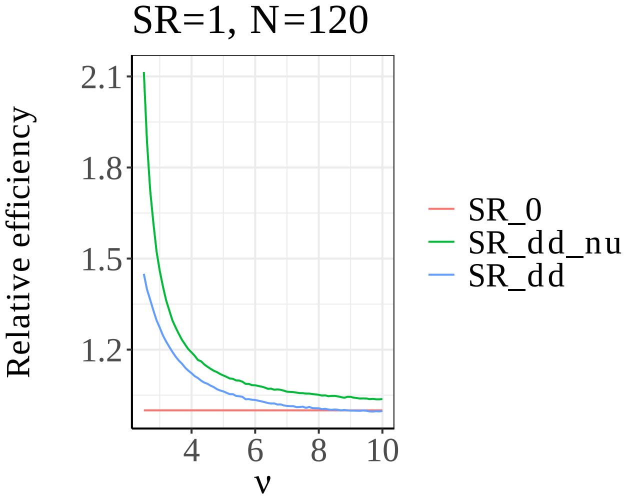


Sharpe ratio from records

- Find $P(r)$
- Measure R_+ or R_-
- Invert $E_{P_r}(R_\pm)(n, \frac{\mu}{\sigma})$

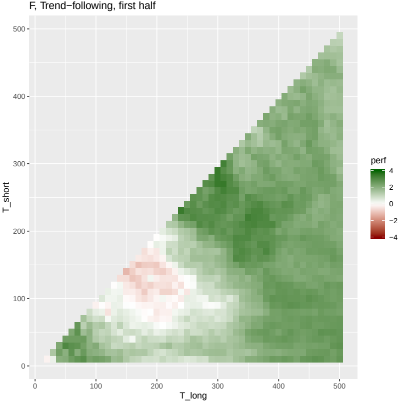
$$SR = F(R_\pm, n, P_r)$$

SR: estimation from drawdown duration

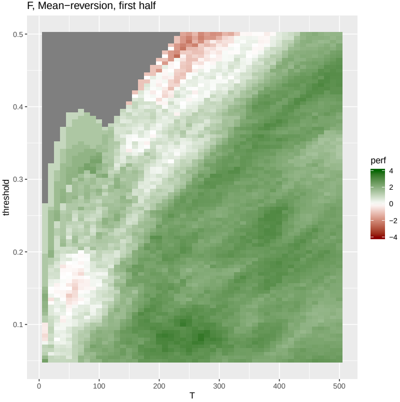


Need for simplification

Trend following



Mean-reverting



Brute-force approach

- Many strategies $\rightarrow$ many measuring tools
  - more complex, richer, market description
  - simpler description $=\rightarrow$ dimension reduction, clustering

Dimension reduction by clustering

Group objects by similarity

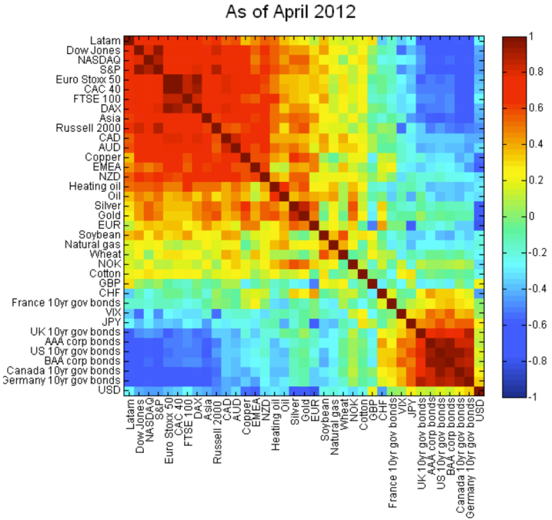
- N objects
- T features r_{it}, $i = 1, \cdots, N$, $t = 1, \cdots, T$
- Similarity $\simeq$ distance $\sim$ correlation
- Clustering $=$ classification

Clustering in finance

- N things $i = 1, \cdots, N$
- T descriptions $x_{i,t}$ $t = 1, \cdots, T$
- Normalisation $E(x_i) = 0$, $E(x_i^2) = 1$
- Group things in S clusters $s_i \in 1, \cdots, K$
- Similarity measure?

Structure of correlation matrices visualization

source



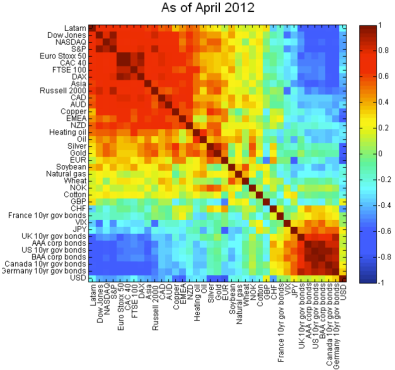
Ansatz for the correlation matrix → clustering

Giada and Marsili (2003): block-diagonal C

- Correlation matrix $C_{ij} = E(x_i x_j) \geq 0$

- Cluster = objects with $\sim$ same cross-correlation

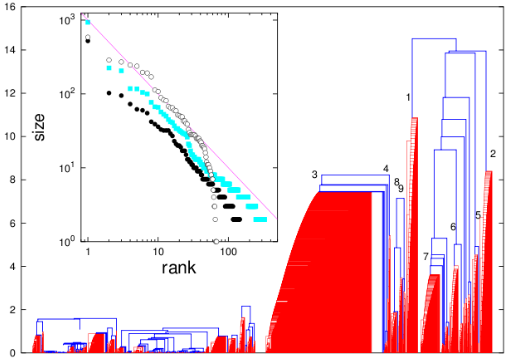
source



Clustering: assets

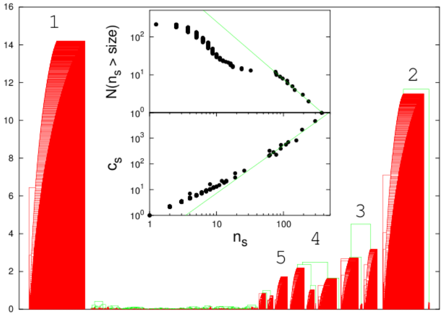
Clusters: economic sectors

1. electric and computers
2. electric and computers
3. mixed
4. gold
5. banks



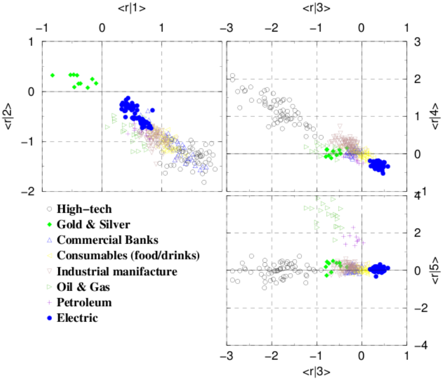
Clustering: days

- $x_{i,t}$: matrix
- $N \simeq T$: transpose and cluster
- Clusters of days $\to$ state



Clustering: days

Day states: way of sectors co-moving



Clustering: days

- Date $\rightarrow$ state
- 5 meaningful states + 1 random

state	cluster	date
1	1	1990/01/02
6	44	1990/01/03
4	5	1990/01/04
2	2	1990/01/05

Spectral Decomposition of Σ (PCA)

C being symmetric non-negative definite, one has

$$C = V^{'} \Lambda V$$

where

1. V, C's eigenvectors matrix is unitary:

$$VV^{'} = V^{'} V = I_N$$

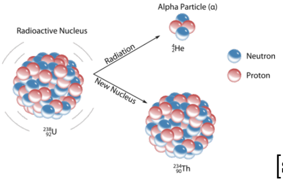
2. $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$ the matrix of eigenvalues of C is diagonal with

$$\lambda_1 \geq \cdots \geq \lambda_N \geq 0.$$

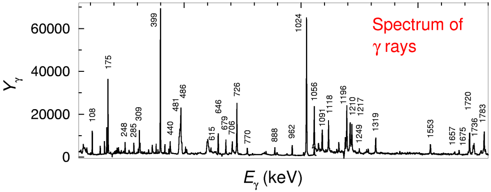
PCA: eigenvalues

- Eigenvalues λ_i
- Density of eigenvalues = spectrum $S(\lambda)$
- $X_{ti} = \frac{R_{ti} - E(R_{ti})}{\sigma_i}$: random $\rightarrow$ C random $\rightarrow$ $S(\lambda)$ random
- How much of C is random?
- How much of $S(\lambda)$ random?
- What is $E(S)(\lambda)$?

Random matrix theory (RMT)



[source]



[source]

- Energy levels: eigenvalues of the energy operator
- Wigner:
  - Energy separation δE is so complex is looks random.
  - $E[P(\delta E)]$,and $E[P(E)]$: universal for a given matrix fanily

Random matrix theory: correlation matrix spectrum

Marcenko-Pastur: Wishart matrices with Gaussian elements

- Correlation matrix: time series of length T

$$C = \frac{1}{T} X^{'} X$$

$$(X)_{i,t} = x_{i,t} \sim N(0,1)$$

- $N,\ T \to \infty$, $q = N/T$, $S \to$

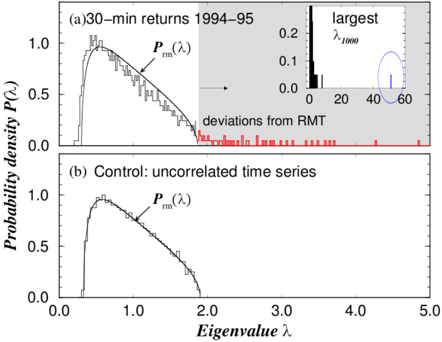
$$S(\lambda) = \frac{1/q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

$$\lambda_\pm = 1 + q \pm 2\sqrt{q} = (1 \pm \sqrt{q})^2$$

- Generalizations to weighted returns, exponentially moving averages, Student-t returns
  [review: Bouchaud and Potters (2009)]

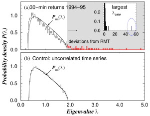
Random bulk vs structure

Laloux et al. (1999), Plerou et al. (2002)



RMT + PCA: structure = significant eigenvalues

Laloux et al. (1999), Plerou et al. (2002)



- usually $\simeq 2\%$ of significant eigenvalues
- clean correlation matrix
- note: eigenvectors are orthogonal

Borrowing from other disciplines

[source]

~~Stealing~~ Adapting tools from other disciplines

Network theory:

- A_{ij}: adjacency matrix:

$$A_{ij} = \begin{cases} 1 & \text{if link between nodes } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$$

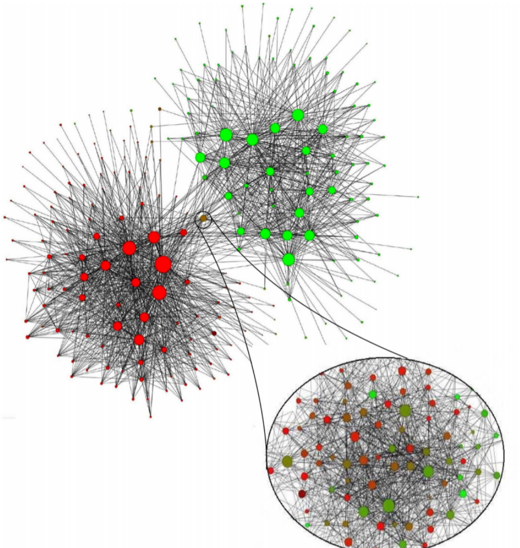
- Weighted graphs

$$A_{ij} \in \mathbb{R}_+$$

- Idea: C_{ij} is a matrix of real numbers
  - $\sqrt{1 - C} \sim$ distance
  - clustering $\equiv$ community detection, see review [link]

Community detection

[Blondel et al (2008)] (Louvain method)



Network theory: community detection

From adjacency matrix

- Null model: configuration model (shuffling links with constraints)
    1. keep empirical degree of all nodes

    $$k_i = \sum_j A_{ij}$$

    2. given degrees $k_1, \cdots, k_N$, probabilistic model

    $$P(\text{link } i \to j) \simeq \frac{k_i k_j}{A_{tot}}$$

    3. max-entropy method with constraints $\to$ ensemble of networks

Network theory: community detection

- Define null model (max-entropy)
- Maximize modularity (average excess of links with respect to null model)

$$Q(\{s_i\}) = \frac{1}{A_{tot}} \sum_{ij} [A_{i,j} - E_{\text{null model}}(A_{ij})] \delta_{s_i, s_j}$$

$$A_{tot} = \sum_{ij} A_{ij}$$

$$\delta_{ab} = 1 \text{ if } a = b, \text{ 0 otherwise}$$

- Louvain clustering method is recommended for finance

Louvain clustering

1. Start with each node in its own cluster.
2. Draw a node a random
   2.1 remove it from its community
   2.2 try to add it to the community of each its neighbour
   2.3 add it to the community with max modularity improvement, if positive
3. Repeat 2 until nothing can be done

N.B.: stochastic algorithm, some randomness in clustering

Consensus clustering

Solution to too-stochastic clustering

E.g. Tandon *et al. (2019)* [link]

Fast iterative method:

1. Perform clustering n_p times $\to$ matrix D_{ij}: if ij are neighbours in graph, compute fraction of times i and j are in the same cluster
2. Thresholding $D_{ij} \to D_{ij} \times (D_{ij} > \theta)$ but keep D connected
3. Triadic closure: for a subset of m nodes, if ij are neighbours and $D_{ij} = 0$, set D_{ij} to fraction of clusterings in which i and j are in the same cluster.
4. Apply clustering method to D_{ij}

Leiden algorithm

Traag, Waltman, van Eck (2019) [paper]

- Extends Louvain algo
- Guarantees connectivity within communities

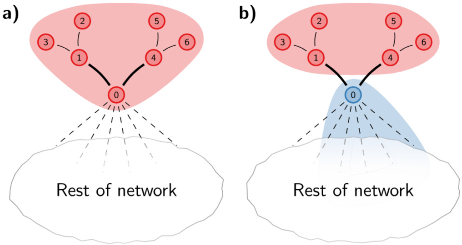


Figure 2. Disconnected community. Consider the partition shown in (**a**). When node 0 is moved to a different community, the red community becomes internally disconnected, as shown in (**b**). However, nodes 1–6 are still locally optimally assigned, and therefore these nodes will stay in the red community.

Leiden algorithm

Traag, Waltman, van Eck (2019) [paper]

- Extends Louvain algo
- Refines communities before merge
  $\rightarrow$ guarantees connectivity within communities

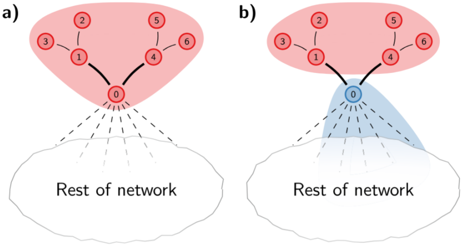
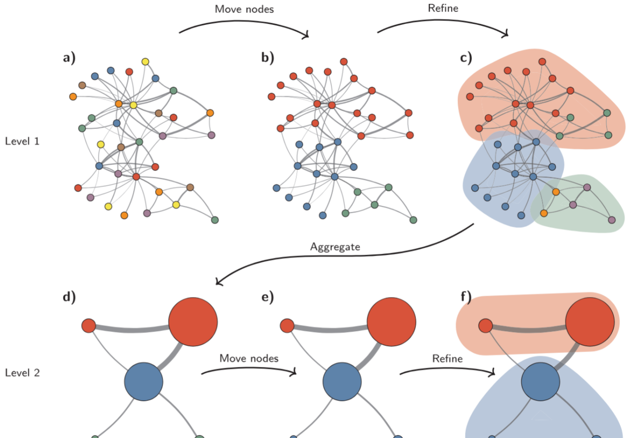


Figure 2. Disconnected community. Consider the partition shown in (**a**). When node 0 is moved to a different community, the red community becomes internally disconnected, as shown in (**b**). However, nodes 1–6 are still locally optimally assigned, and therefore these nodes will stay in the red community.

Leiden algorithm

Traag, Waltman, van Eck (2019) [paper]

- Extends Louvain algo
- Refines communities before merge
  $\rightarrow$ guarantees connectivity within communities



Applying community detection to correlation matrix

Reference: MacMahon & Garlaschelli (2013)[link]

$$Q(\{s_i\}) \underset{=}{?} \frac{1}{C_{tot}} \sum_{ij} [C_{i,j} - E_{\text{null model}}(C_{ij})] \delta_{s_i,s_j}$$

What is the constraint? What is the null model?

$E_{\text{null model}}(C_{ij}) = C^0 = 1$? $C^{(random)}$? $C^{(random)} + C^{(market\ mode)}$?

Proper definition

$$Q(\{s_i\}) _= \frac{1}{C_{tot}} \sum_{ij} [C_{i,j} - C^0_{i,j}] \delta_{s_i,s_j}$$

Clustering correlation matrices the Louvain/Leiden way

- Using $C^0 = C^{(r)} + C^{(m)}$ where

$$C^{(r)} = \sum_{i \text{ s.th } \lambda_i \leq \lambda_+} \lambda_i v_i^{(t)} v_i$$

$$C^{(m)} = \lambda_M v_M^{(t)} v_M \quad \text{where } \lambda_M = \max_i \lambda_i$$

- Optimization:

$$C - C^0 = \sum_{i \text{ s.th } \lambda_+ < \lambda < \lambda_M} \lambda_i v_i^{(t)} v_i$$

- Note: L/L algorithms require positive coefficients

$$C - C^0 \to \mathsf{abs}(C - C^0)$$

Clustering @ 60 minutes

[Hendricks, Gebbie and Wilcox (2015)]

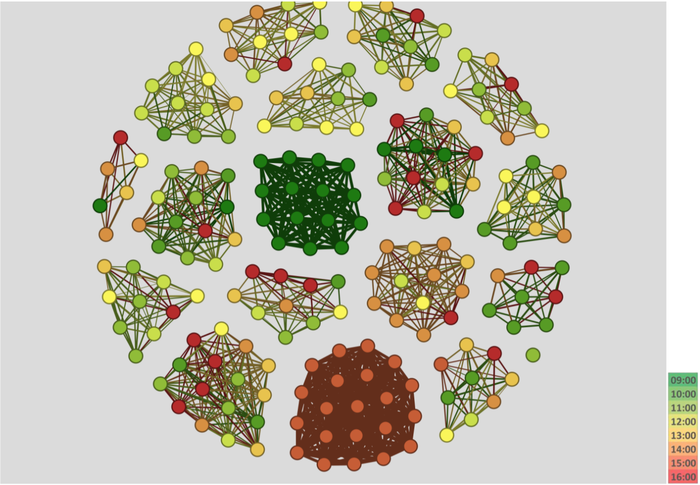


Figure 3.: JSE TOP40 60-minute temporal clusters for the period 01-Nov-2012 to 30-Nov-2012, representing 184 distinct periods. Each node represents a 60-minute period during a trading day, with the colour shading indicating the time-of-day (Morning = green, Lunch = yellow, Afternoon = red) and node connectedness indicating cluster membership.

Clustering @ 60 minutes

[Hendricks, Gebbie and Wilcox (2015)]
Dynamics of market states is not totally random

		state_{t+1}					
		1	**2**	**3**	**4**	**5**	**6**
	1	0.13	0.49	0.32	0.00	0.06	0.00
	2	0.41	0.41	0.09	0.00	0.09	0.00
state_t	**3**	0.00	0.00	0.00	0.52	0.05	0.43
	4	0.25	0.07	0.00	0.25	0.43	0.00
	5	0.32	0.59	0.05	0.05	0.00	0.00
	6	0.00	0.00	0.00	1.00	0.00	0.00

Clustering @ 5 minutes

[Hendricks, Gebbie and Wilcox (2015)]

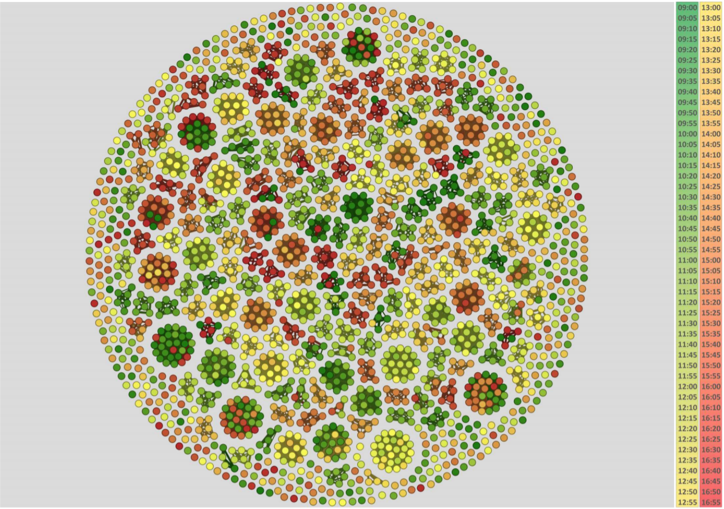


Figure 9.: JSE TOP40 5-minute temporal clusters for the period 01-Nov-2012 to 30-Nov-2012, representing 2208 distinct periods. Each node represents a 5-minute period during a trading day, with the colour shading indicating the time-of-day (Morning = green, Lunch = yellow, Afternoon = red) and node connectedness indicating cluster membership.

Clustering: 5 minutes

[Hendricks, Gebbie and Wilcox (2015)]

													state$_{t+1}$												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
1	0.04	0.08	0.17	0.06	0.12	0.14	0.07	0.00	0.19	0.00	0.10	0.01	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.08	0.01	0.16	0.14	0.08	0.23	0.03	0.01	0.13	0.01	0.04	0.02	0.01	0.03	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.01
3	0.04	0.06	0.12	0.12	0.11	0.33	0.05	0.02	0.07	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.04	0.07	0.24	0.03	0.17	0.21	0.02	0.02	0.08	0.01	0.05	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00
5	0.04	0.03	0.16	0.13	0.06	0.16	0.08	0.03	0.23	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.07	0.07	0.32	0.12	0.08	0.09	0.02	0.01	0.13	0.02	0.03	0.01	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.14	0.07	0.24	0.04	0.23	0.10	0.04	0.01	0.05	0.00	0.02	0.00	0.00	0.01	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.02	0.05	0.33	0.10	0.10	0.18	0.03	0.03	0.10	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.11	0.04	0.15	0.04	0.20	0.30	0.04	0.02	0.03	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.06	0.12	0.24	0.00	0.00	0.18	0.06	0.06	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.19	0.01	0.26	0.21	0.07	0.13	0.00	0.00	0.04	0.01	0.03	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
state$_t$ 12	0.11	0.11	0.00	0.00	0.22	0.22	0.00	0.22	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	0.03	0.03	0.17	0.00	0.10	0.37	0.00	0.03	0.10	0.00	0.07	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
15	0.00	0.00	0.33	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.20	0.40	0.20	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	0.20	0.00	0.20	0.20	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	0.00	0.12	0.24	0.24	0.04	0.12	0.04	0.08	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.04	0.00	0.00	0.00
19	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.14	0.14	0.14	0.14	0.00	0.14	0.00	0.00	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22	0.00	0.00	0.25	0.00	0.00	0.50	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
23	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
24	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 8.: Empirical 1-step transition probability matrix for 5-minute states, based on identified temporal cluster configuration. State transitions with a probability > 0 are highlighted in green.

Custering of days: predictability?

- At the close of time t, state μ_t
- $E(r_t|\mu_t)$ very significantly non-zero
- Is $E(r_{t+1,1}|\mu_t)$ significantly non-zero?

$$E(r_{t+1,1}|\mu_t) = \sum_{\nu} W(\mu_t \to \nu) E(r_{t+1}|\nu)$$

where $W(\mu_t \to \nu)$ changes slowly a function of time

- Raw Sharpe ratio

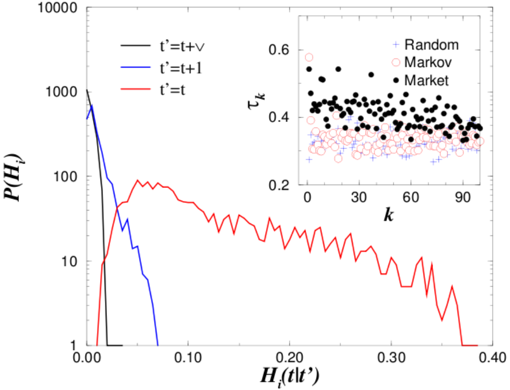
$$S_{\mu,\mathrm{raw}} = \frac{E(r_{t+1,1}|\mu)}{\sqrt{\mathrm{var}(r_{t+1,1}|\mu)}}$$

- Benchmark: $E(r_{t+1,1})$, $\delta r_{t+1,1} = r_{t+1,1} - E(r_{t+1,1})$

$$S_{\mu} = \frac{E(\delta r_{t+1,1}|\mu)}{\sqrt{\mathrm{var}(\delta r_{t+1,1}|\mu)}}$$

Custering of days: predictability?

For stock i : $H_i = E(S_{i,\mu})$ (note: in-sample only)



Conditional finance: single asset

- Strategy: signal x_t, real return R_t

$$x_t \rightarrow g_{t+1} = x_t R_{t+1}$$

- Conditional strategies

$$x_t = \operatorname{sgn} R_t$$
$$x_t = \operatorname{sgn}(R_t R_{t-1})$$

$$\mu_t = (\operatorname{sgn} R_t, \operatorname{sgn} R_{t-1})$$
$$\mu_t = (\operatorname{sgn} R_t, \cdots, \operatorname{sgn} R_{t-M})$$

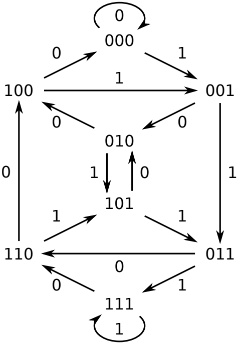
μ_t: state of the asset $\rightarrow$ conditional signal

Asset state evolution

Bit-shift register:

$$x_t = (\mathrm{sgn} R_t, \cdots, \mathrm{sgn} R_{t-M})$$

$x_t \to x_{t+1}$: De Bruijn graph, constrained evolution



Conditional investements

Conditional signal

1. compute states μ_t in-sample $t \in [t_0, t_1]$
2. compute performance of assets/strategies/ML conditional on μ_{t_1}
3. Rank / select (FDR) assets / strategies
4. Invest out-of-sample
5. Shift in-sample period, repeat.

Financially-motivated clustering

- Giada and Marsili (2003): block-diagonal C
- Correlation matrix $C_{ij} = E(x_i x_j) \geq 0$
- Cluster = objects with $\sim$ same cross-correlation
  Ansatz: C diagonal by blocks

$$C_{i,j} = \begin{cases} 1 & i = j \\ c_s & s_i = s_j,\ i \neq j \\ 0 & s_i \neq s_j \end{cases}$$

$$n_s = \sum_i \delta_{s,s_i}$$

$$c_s = \sum_{i,j} \delta_{s,s_i} \delta_{s,s_j} C_{i,j}$$

Clustering

- Stochastic model for $x_{i,t}$

$$\tilde{x}_{i,t} = \frac{\sqrt{g_{s_i}}\eta_{s_i,t} + \epsilon_{i,t}}{\sqrt{1 + g_{s_i}}} \qquad \eta \text{ and } \epsilon \text{ iid and } \sim \mathcal{N}(0,1)$$

- Total cross-correlation inside cluster s

$$C_s = \frac{g_s \delta_{s_i,s_j} + \delta_{i,j}}{1 + g_{s_i}}$$

- Model of time series given by

$$G = \{g_1, \cdots, g_K\} \text{ how many clusters, correlation}$$
$$S = \{s_1, \cdots, s_N\} \quad \text{cluster attribution}$$

Clustering

- Model of time series given by

$$G = \{g_1, \cdots, g_K\} \text{ how many clusters, correlation}$$
$$S = \{s_1, \cdots, s_N\} \text{ cluster attribution}$$

- Likelihood

$$P(x|S,G) = \prod_{t=1}^{T} E_{\eta,\epsilon}\left[\prod_{i=1}^{N} \delta\left(x_{i,t} - \tilde{x}_{i,t}\right)\right] \propto P(S,G|x)$$

- Exponentiation of Dirac functions with

$$\delta(x) = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{ikx}$$

Clustering: maximum likelihood

- Gaussian integration $\rightarrow$

$$P(S,G|x) \propto e^{T\mathcal{L}\{S,G\}}$$

$$\mathcal{L}\{S,G\} = -\frac{1}{2}\sum_s [(1+g_s)(n_s - \frac{g_s c_s}{1+g_s n_s}) + n_s \ln(1+g_s) - \ln(1+g_s n_s)]$$

- Log-likelihood $\mathcal{L}$; maximisation: $\frac{\partial \mathcal{L}}{\partial g_s} = 0$

$$\hat{g}_s = \begin{cases} \frac{c_s - n_s}{n_s^2 - c_s} & n_s > 0 \\ 0 & n_s = 0 \end{cases}$$

$$\mathcal{L}_c(S) = \frac{1}{2}\sum_{s, n_s > 0}\left[\ln\frac{n_s}{c_s} + (n_s - 1)\ln\frac{n_s^2 - n_s}{n_s^2 - c_s}\right]$$

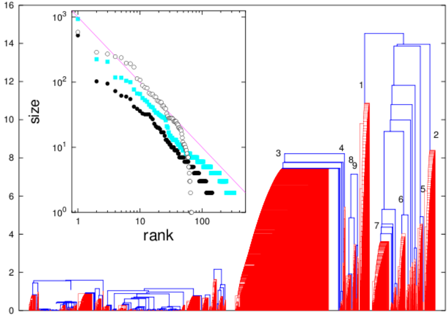
Clustering: maximum log-likelihood

- Problem: s_i is discrete. Maximize $\mathcal{L}_c$ w.r.t S?
- Enumerate: $O(K^N)$: NP-Complete
- Random search
  1. Start with arbitrary $\mathcal{S}$
  2. Propose $s_i \to s$ for all i
  3. Compute differences in $\mathcal{L}_c$ for each i
  4. Keep single move that improves $\mathcal{L}_c$ the most
  5. Stop when no move improves $\mathcal{L}_c$
- Merging algorithm
  1. Start with N clusters, $s_i = i$
  2. Merge two clusters s', s'' so that $\mathcal{L}_c$ is the most improved
  3. Repeat $N - 1$ times

Clustering

Merging algorithm

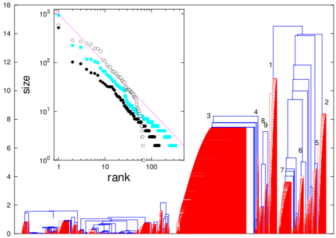
1. Start with N clusters, $s_i = i$
2. Merge two clusters i, j so that $\mathcal{L}_c$ is the most improved
3. Repeat $N - 1$ times



Clustering

- $\mathcal{L}_c = \sum_s l_s$: superposition of terms
- merge r and s into q:
  1. $n_q = n_r + n_s$
  2. c_q: recompute from x_i
  3. merge l_s and $l_{s'}$into l_q

$$\begin{cases} l_q & > l_r + l_s \\ l_q & < l_r + l_s, \quad l_q > \max(l_r, l_s) \\ l_q & < l_r + l_s, \quad l_q < \max(l_r, l_s): \quad \text{no links in dendrogram} \end{cases}$$



Ansätze: the good and the bad

The good

- Structure $\rightarrow$ good filtering

The bad

- Structure: arbitrariness