Taxonomy of a Stock Portfolio

Chapter 13

Mantegna, Stanley: An Introduction to Econophysics

Emlékeztető: Pearson-(korrelációs)-koefficiens

$$S_i \equiv \ln Y_i(t) - \ln Y_i(t-1), \tag{12.1}$$

$$\rho_{ij} = \frac{\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle}{\sqrt{\langle S_i^2 - \langle S_i \rangle^2 \rangle \langle S_j^2 - \langle S_j \rangle^2 \rangle}}. \tag{12.2}$$

Here Y_i is the daily closure price of stock i at time t, and S_i is the daily change of the logarithm of the price of stock i. The angular brackets indicate a time average over all the trading days within the investigated time period. With this definition, the correlation coefficient ρ_{ij} can assume values ranging from -1 to 1, with three special values

$$\rho_{ij} = \begin{cases} 1 & \text{completely correlated changes in stock price,} \\ 0 & \text{uncorrelated changes in stock price, and} \\ -1 & \text{completely anticorrelated changes in stock price.} \end{cases} \tag{12.3}$$

A mai téma lényege

Miért hasznos a Pearson-koefficiens?

- Lehetővé teszi egy olyan metrika meghatározását, amely megadja egy adott portfólió részvényei közötti relatív távolságot.
- A részvényárfolyamok idősorában tárolt gazdasági információk kinyerésére szolgáló módszert kínál

Részvények közötti „távolságok”

A method of determining a distance between stocks i and j evolving in time in a synchronous fashion is the following. Let us consider

$$\tilde{S}_i \equiv \frac{S_i - \langle S_i \rangle}{\sqrt{\langle S_i^2 \rangle - \langle S_i \rangle^2}}, \qquad (13.1)$$

where S_i, the logarithmic price difference of stock i, is given by Eq. (12.1). Hence $\tilde{S}_i$ is the same variable subtracted from its mean, and divided by its standard deviation computed over a given time interval. Let us consider the n records of $\tilde{S}_i$ present in the same time interval as the components $\tilde{S}_{ik}$ of an n-dimensional vector $\tilde{\mathbf{S}}_i$.

Részvények közötti „távolságok”

The Euclidean distance d_{ij} between vectors $\tilde{\mathbf{S}}_i$ and $\tilde{\mathbf{S}}_j$ is obtainable from the Pythagorean relation

$$d_{ij}^2 = ||\tilde{\mathbf{S}}_i - \tilde{\mathbf{S}}_j||^2 = \sum_{k=1}^{n} (\tilde{S}_{ik} - \tilde{S}_{jk})^2. \tag{13.2}$$

The vector $\tilde{\mathbf{S}}_i$ has unit length because, from definition (13.1),

$$\sum_{k=1}^{n} \tilde{S}_{ik}^2 = 1. \tag{13.3}$$

Részvények közötti „távolságok”

Hence Eq. (13.2) can be rewritten as

$$d_{ij}^2 = \sum_{k=1}^{n} (\tilde{S}_{ik}^2 + \tilde{S}_{jk}^2 - 2\tilde{S}_{ik}\tilde{S}_{jk}) = 2 - 2\sum_{k=1}^{n} \tilde{S}_{ik}\tilde{S}_{jk}. \qquad (13.4)$$

The sum on the right side of Eq. (13.4), $\sum_{k=1}^{n} \tilde{S}_{ik}\tilde{S}_{jk}$, coincides with ρ_{ij} (see Eq. (12.2)). Hence Eq. (13.4) leads to (D. Sornette, private communication)

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}. \qquad (13.5)$$

Részvények közötti „távolságok”

The sum on the right side of Eq. (13.4), $\sum_{k=1}^{n} \tilde{S}_{ik} \tilde{S}_{jk}$, coincides with ρ_{ij} (see Eq. (12.2)). Hence Eq. (13.4) leads to (D. Sornette, private communication)

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}. \tag{13.5}$$

Because Eq. (13.2) defines a Euclidean distance, the following three properties must hold:

$$\begin{aligned}
\text{Property (i)}: \quad & d_{ij} = 0 \iff i = j \\
\text{Property (ii)}: \quad & d_{ij} = d_{ji} \\
\text{Property (iii)}: \quad & d_{ij} \leq d_{ik} + d_{kj}
\end{aligned} \tag{13.6}$$

Egy példa a d_{ij} metrikára

- **Hat vállalatból álló példát vizsgálunk: Chevron (CHV), General Electric (GE), Coca Cola (KO), Procter & Gamble (PG), Texaco (TX) és Exxon (XON)**
- **Az d_{ij} mátrix az 1990-es évben:**

	CHV	GE	KO	PG	TX	XON
CHV	0	1.15	1.18	1.15	0.84	0.89
GE		0	0.86	0.89	1.26	1.16
KO			0	0.74	1.27	1.11
PG				0	1.26	1.10
TX					0	0.94
XON						0

- **Minél nagyobb a szám, annál kevésbé korreláltan változnak a részvények.**

Ultrametrikus terek (ultrametric spaces)

- **Ultrametrikus terekben ultrametrikus távolságok vannak.**
- **Ultrametrikus távolságokra a háromszög-egyenlőtlenségnél egy szigorúbb feltételnek kell teljesülnie.**
- **Mire jó? Komplex rendszerekben hierarchikus struktúrákat vezethetünk be.**

An ultrametric space is a space in which the distance between objects is an ultrametric distance. An ultrametric distance $\hat{d}_{ij}$ must satisfy the first two properties of a metric distance, (i) $\hat{d}_{ij} = 0 \iff i = j$ and (ii) $\hat{d}_{ij} = \hat{d}_{ji}$, while the usual triangular inequality of Eq. (13.6) is replaced by a stronger inequality, called an ultrametric inequality,

$$\hat{d}_{ij} \leq \max\{\hat{d}_{ik}, \hat{d}_{kj}\}. \tag{13.7}$$

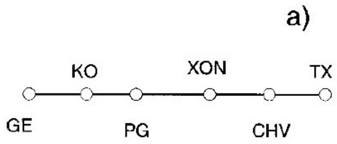
MST (minimal-spanning tree)

- **Egy metrikus térben számos ultrametrikus tér található.**
- **Ezekből egy különlegesnek tekinthető: szubdomináns ultrametrika.**
- **A szubdomináns ultrametrikát azt MST gráf választja ki.**

Az MST egy speciális gráf. Hogyan található meg?

- **Egy N elemű rendszerben N-1 éle lesz.**
- **Az N-1 db. él hosszának az összegét minimalizálni kell.**

Példa: MST és IHT (indexelt hierarchikus fa)



The MST associated with the Euclidean matrix can be obtained as follows. First find the pair of stocks separated by the smallest distance: KO and PG ($d = 0.74$). Then find the pair of stocks with the next-smallest distance: CHV and TX ($d = 0.84$). We now have two separate regions in the MST. If we continue, we find next the KO and GE pair ($d = 0.86$). At this point, the regions of the MST are GE–KO–PG and CHV–TX. The next pairs of closest stocks are GE–PG and CHV–XON ($d = 0.89$). The connection GE–PG is not considered because both stocks have already been sorted, while XON is linked to CHV in the MST. Now the two regions are XON–CHV–TX and GE–KO–PG. The smallest distance connecting the two regions is observed for PG–XON ($d = 1.10$). This PG–XON link completes the MST.

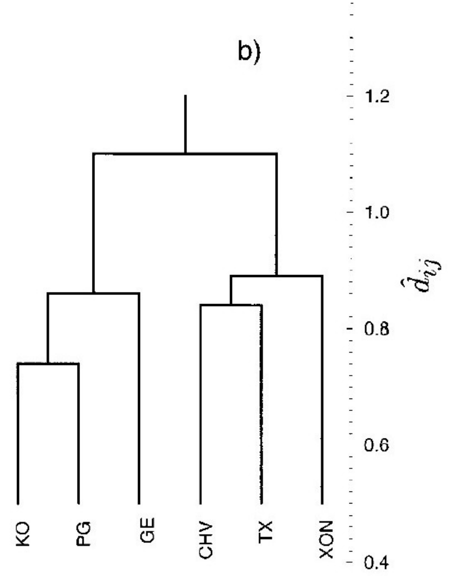
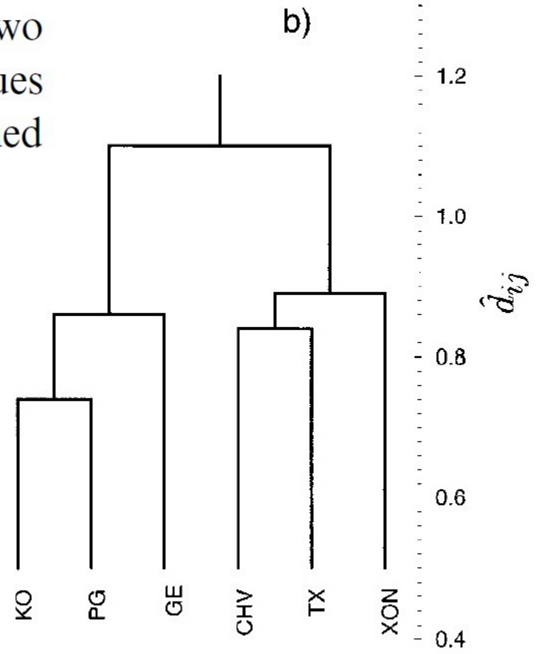


Fig. 13.1. (a) MST and (b) indexed hierarchical tree obtained for the example of six firms, identified by their tick symbols CHV, GE, KO, PG, TX and XON.

Példa: Ultrametrikus metrika meghatározása

Each element in the $\hat{d}_{ij}$ matrix is equal to the maximal distance between two successive objects encountered when moving from the starting object to the ending object over the shortest path of the MST connecting the two objects. In contrast to the d_{ij} matrix, the number of different element values in the ultrametric distance matrix $\hat{d}_{ij}$ cannot exceed $n-1$, as is confirmed by the present example.

	CHV	GE	KO	PG	TX	XON
CHV	0	1.10	1.10	1.10	0.84	0.89
GE		0	0.86	0.86	1.10	1.10
KO			0	0.74	1.10	1.10
PG				0	1.10	1.10
TX					0	0.89
XON						0



Példa: az IHT időfüggése

- **Ami nem változott: két fő klaszter rajzolódik ki, ugyanazokkal a vállalatokkal.**
- **Ami változott: a két fő klaszter belső strukúrája.**

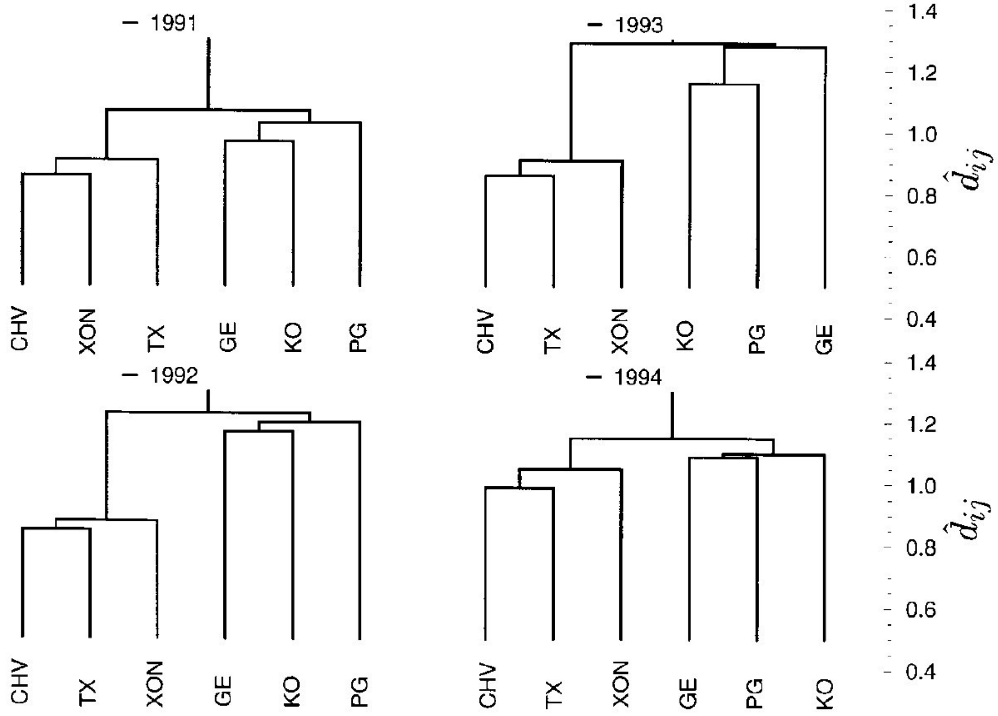
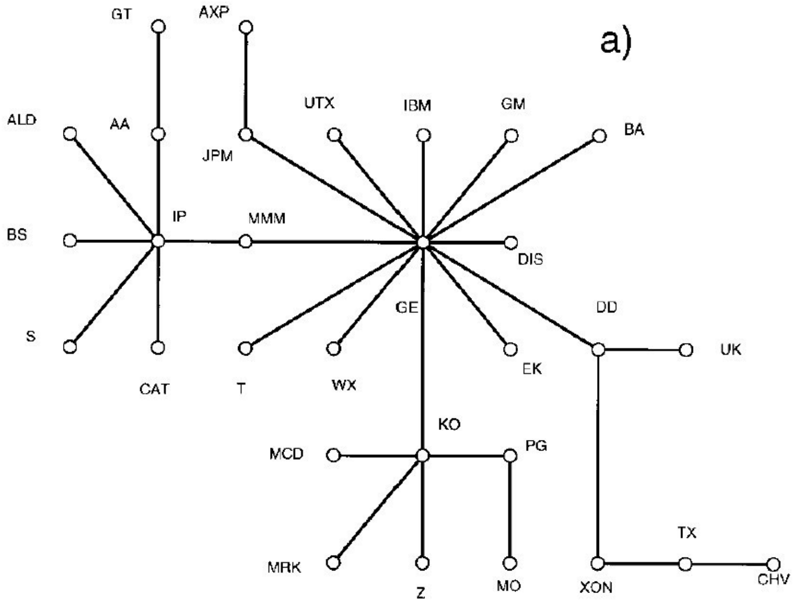


Fig. 13.2. Indexed hierarchical trees obtained during the calendar years from 1991 to 1994 for the portfolio of six firms (CHV, GE, KO, PG, TX, and XON).

Komplex példa 1

A Dow-Jones Industrial Average (DJIA) index kiszámításához használt 30 részvény.



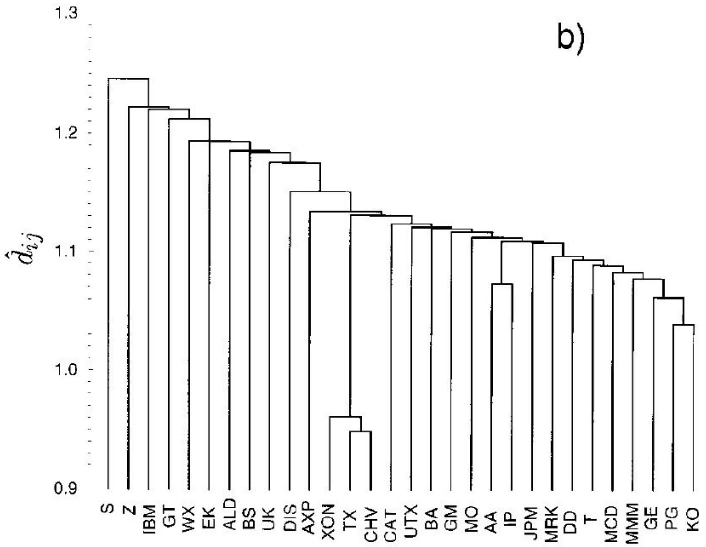


Fig. 13.3. (a) MST and (b) indexed hierarchical tree obtained for the DJIA portfolio during the time period 7/89 to 10/95. Adapted from [108].

Komplex példa 2

- A második portfólió az S&P 500.
- Sokkal nagyobb halmaz.
- Különböző gazdasági tevékenységek jelennek meg.
- Balra csak a fő struktúra látszik, a diagram „csonkított”.

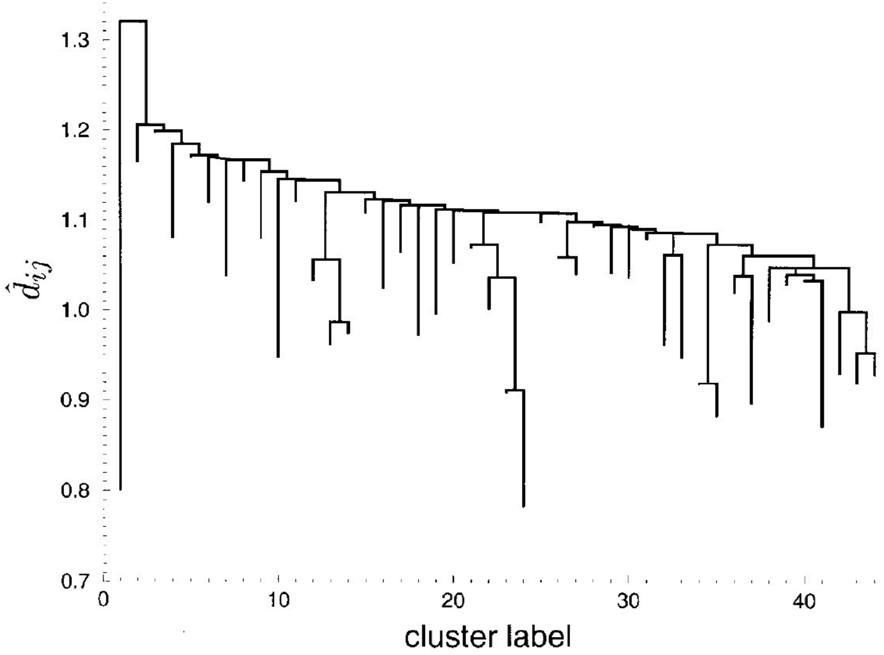


Fig. 13.4. Main structure of the MST of the S&P 500 portfolio for the time period 7/89 to 10/95. Adapted from [108].

Összefoglalás

- **Lehetséges olyan stratégiákat kidolgozni, amelyek lehetővé teszik, hogy ésszerű taxonómiákat (hierarchikus rendszereket) kapjunk, ha egynél több részvényárfolyam-idősor szinkron elemzéséből indulunk ki.**
- **Az egyes részvényárfolyamok idősoraiban tárolt gazdasági információk egy részét kinyerhetjük, ha kiszámítjuk a portfólió egyes részvénypárjai közötti távolságot, és feltételezzük, hogy a szubdomináns ultrametrikus tér egy megfelelő topológia.**