

# HIERARCHIKUS KLASZTEREZÉS ÉS A PORTFÓLIÓ-KIVÁLASZTÁS PROBLÉMA<sup>1</sup>

GERA IMRE – LONDON ANDRÁS

*Szegedi Tudományegyetem és Poznań University of Economics and Business*

Ezen tanulmányban hierarchikus klaszterezésen alapuló portfólió-kiválasztási eljárásokat hasonlítunk össze. A hierarchikus klaszterezést egyrészt a korrelációs mátrix tisztítására használjuk, és a szakirodalomban bevett módszerek mellett egy ún. konfigurációs modellel alapuló módszert is bemutatunk. A portfólió kiválasztó eljárások közül a tisztított korrelációs mátrixot használó Markowitz-modellt vizsgáljuk. Másrészt a hierarchikus klaszterezést a portfólió kialakítására közvetlenül használó HRP és HERC algoritmusokat mutatjuk be. A valós adatokon végzett kísérleteinkkel összeállított portfóliókat több kritérium szerint is kiértékeljük. Vizsgálataink azt mutatják, hogy az új portfólió kiválasztó módszerek ugyan több metrika szerint is jobban teljesítenek az eredeti Markowitz-modellnél, nincs egyértelműen legjobb választás a befektető számára.

## 1 Bevezetés

A hagyományos portfólió-kiválasztási eljárások a részvényhozamok varianciáján keresztül mérik a portfólió kockázatát, melyet minimalizálni szeretnének. Ezen optimalizálási eljárások a gyakorlatban akkor működhetnek jól, ha a részvényhozamokból képzett kovariancia (vagy korrelációs) mátrix megfelelő módon tisztított. Ehhez egyrészt ki kell szűrni a kovarianciák becsléséből adódó statisztikai bizonytalanságot (röviden zajt), melynek fő okozója, hogy a vizsgált  $n$  részvény idősorainak  $T$  hossza nem elegendően nagy  $n$ -hez viszonyítva. Általában ez a helyzet, mivel a hozamok közti összefüggések nem-stacionárius jellege miatt célszerű  $T$ -t minél kisebbre választani (Bongiorno és Challet 2020). A kovariancia becsléséből származó zaj csökkentésére már korábban is léteztek megoldások, például Laloux és tsai. 1999; Tola és tsai. 2008, illetve összefoglaló tanulmány jelent meg a témában (Bun, Bouchaud és Potters 2017). Ezen módszerek az eredeti Markowitz-féle portfólió modell javítására is használhatók (London, Gera és Bánhelyi 2018). Egy jó tisztítási eljárásnak olyannak kell lennie, ami megtartja a kovariancia mátrix stabil, tényleges információt tartalmazó szerkezetét. Ezért az, hogy éppen melyik tisztítási eljárás működik legjobban, függhet az adott értékpapírpiac, illetve a vizsgált értékpapírok sajátosságaitól.

---

<sup>1</sup>E-mail: [gerai@inf.u-szeged.hu](mailto:gerai@inf.u-szeged.hu), [london@inf.u-szeged.hu](mailto:london@inf.u-szeged.hu). Beérkezett 2020. december 11.

Számos megfigyelés szerint a legtöbb komplex rendszer, így a tőkepiacok is, hierarchikus felépítést követnek (Simon 1991). Egy kézenfekvő megközelítés szerint ezt a szerkezetet valamilyen hierarchikus klaszterező eljárást alkalmazva határozhatjuk meg. Ekkor feltételezzük, hogy a hozamkorrelációk közötti függőségi (hierarchikus) viszonyok “egymásba ágyazottak”, vagyis egy dendrogram segítségével jól leírhatók (Tumminello, Lillo és Mantegna 2007). Ebben a tanulmányban a hierarchikus klaszterezés lehetséges alkalmazásait mutatjuk be a portfólió-kiválasztási feladat esetén: egyrészt mint kovarianciamátrix szűrési eljárást alkalmazzuk a Markowitz-modellben, másrészt közvetlenül használjuk a portfólió összeállítására, megkerülve ezzel az optimalizálási feladat megoldását. Kísérleteinket valós adatokon hajtjuk végre, a módszerek teljesítményét több mutatón keresztül hasonlítjuk össze.

## 1.1 A portfólió probléma

Portfólió-kiválasztásról akkor beszélünk, amikor egy befektető a tőkét szerte szétosztani a piacon elérhető részvények között. Ennek elsődleges célja egy elvart hozamszint biztosítása mellett a befektetés kockázatának minimalizálása, illetve adott kockázati szinten a lehető legnagyobb hozam elérése. A feladat megoldásához jellemzően egy adott részvényhalmaz (példának okán legyenek ezek a Budapesti Értéktőzsde részvényei) historikus adataiból indulunk ki, amely tartalmazza az egyes részvények árfolyamait tetszőleges időrendi lebontásban. Itt napi záróárakkal dolgozunk, de kisebb (pl. heti) és nagyobb (pl. óránkénti) felbontású adatokat is szokás vizsgálni. A napi záróárakból gyakran logaritmikus hozamokat képzünk. Az  $i$ -edik részvény logaritmikus hozama a  $t$  időpontban

$$x_{i,t} = \log \frac{P_i(t)}{P_i(t-1)} = \log P_i(t) - \log P_i(t-1), \quad (1)$$

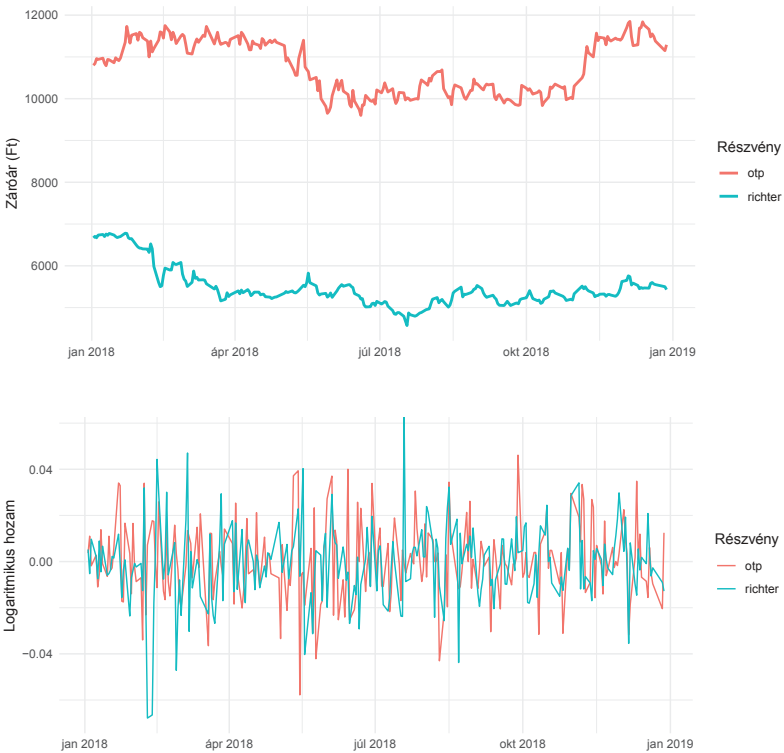
ahol  $P_i(t)$  az  $i$ -edik részvény  $t$ -edik napi záróárát jelöli. A logaritmikus hozamok használatának fő oka, hogy ezek a hozamok idő-additívak (vagyis  $x_{i,t} + x_{i,t+1} + \dots + x_{i,t+j} = \log P_i(t+j) - \log P_i(t)$ ), és ha az 1-periódusos logaritmikus hozamok normális eloszlásúak (ez egy gyakori feltételezés, mely a mi adataink esetén is teljesül), a  $j$ -periódusosak is azok lesznek. Egy példa napi záróárakra és logaritmikus hozamokra az 1. ábrán látható. A vizsgálataink során a záróárakat tehát ezekkel fogjuk helyettesíteni. Ebből az  $i$ -edik részvény átlagos hozama, ha az idősor  $T$  hosszú,

$$\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{i,t}. \quad (2)$$

Az  $i$  és  $j$  részvények közötti kovarianciát a

$$\sigma_{ij}^2 = \frac{1}{T-1} \sum_{t=1}^T (x_{i,t} - \bar{x}_i)(x_{j,t} - \bar{x}_j) \quad (3)$$

összefüggéssel határozhatjuk meg. Ezeket mátrixba rendezve kapjuk a  $\Sigma = (\sigma_{ij}^2)_{i,j}$  kovarianciamátrixot, amely tartalmazza az összes részvény hozamidősrának páronkénti kovarianciáját, valamint a főátlóban az egyes részvények hozamainak varianciáját is. A  $\mathbf{C}$  korrelációs mátrixot egyszerűen kapjuk a kovarianciamátrixból a  $C_{ij} := \sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}}$  normalizálással. Fontos megemlíteni, hogy a korreláció értéke nem invariáns a hozamok logaritmikus transzformációjára, ezért portfóliókiválasztási modellek esetén általában normál hozamokkal számolnak (Embrechts, McNeil és Straumann 2002). Ugyanakkor a valóságban megjelenő napi hozamok esetén a normál és logaritmikus hozamok közötti eltérés elhanyagolható, a logaritmikus hozamok használatának fentebb említett előnye miatt mi ezt preferáltuk.



1. ábra. Az OTP és Richter részvények 2018-as napi záróárai (felül) és logaritmikus hozamai (alul)

A továbbiakban a következő szituációt feltételezzük. A befektető az adott napon mindig az előző  $T$  nap árfolyamainak figyelembevételével határozza meg a portfólióját. Ekkor történik a kockázatelemzés és a diverzifikáció, azaz a portfólió részvényeibe történő befektetés is. Ezt követően vár újabb  $T$  napot, amikor is kiértékeli azok tényleges („realizált”) hozamát és kockázatát.

## 2 Hierarchikus klaszterezés a Markowitz-modellben

A kockázatsökkentés egyik fő eszköze a diverzifikáció, vagy kockázatmegosztás, miszerint egyszerre nem pusztán egy, hanem több értékpapírba fektetünk be. Ezt a módszert intuitíven az motiválja, hogy ha nem korrelálnak egymással a birtokunkban lévő értékpapírok hozamai, akkor egy-egy értékpapír árfolyamának csökkenésekor nem törvényszerű, hogy a többi árfolyama is csökken, tehát kisebb annak a veszélye, hogy összességében veszteségesek leszünk. Az első, diverzifikációt is magába foglaló portfólió-kiválasztási modellt Harry Markowitz fogalmazta meg (Markowitz 1952). Lévéen, hogy az elvárt hozam és kockázat jelentése, ezáltal mérése sem egyértelmű, ezt a két mérőszámot vagy a modell módszertana (ilyen pl. a *Markowitz-modell*) vagy a befektető határozza meg (mint például a *Hierarchical Equal Risk Contribution* (HERC) modell esetében (Raffinot 2018)).

Az eredeti Markowitz-modell a portfólió várható hozamát a részvények átlagos hozamainak súlyozott átlagaként, a kockázatot pedig a portfólió hozamának varianciájaként definiálja. Markowitz megmutatta, hogy amennyiben felső korlátot adunk a kockázatra vagy alsó korlátot a hozamra, az optimális portfóliót egy konvex kvadratikusan programozási feladat megoldásával kaphatjuk meg. Itt a hozamra fogunk alsó korlátot megadni, azaz a befektető számára elfogadható legkisebb hozamot lehet majd paraméterként kezelni. A megoldandó kvadratikusan optimalizálási feladat a következő:

$$\begin{aligned} \min_{\mathbf{p}} \quad & \mathbf{p} \Sigma \mathbf{p}^{\top} \\ \text{feltéve} \quad & \sum_{i=1}^n p_i = 1 \\ & \sum_{i=1}^n p_i \bar{r}_i \geq R, \end{aligned} \tag{4}$$

ahol  $\bar{r}_i$  az  $i$ -edik részvény ( $i = 1, \dots, n$ ) becsült hozama valamilyen statisztikai becslés szerint, mint például az átlagos hozam, azaz  $\bar{r}_i = \bar{x}_i$  (ld. (2)-es képlet),  $R$  pedig a minimálisan elvárt hozam. A  $\sum_i p_i = 1$  technikai feltétel biztosítja, hogy a tőkénk pontosan 100%-át fektessük be. A feladat megoldásaként egy  $\mathbf{p}$  súlyvektort kapunk, amelyben az egyes részvényekbe fektetendő tőkearányok szerepelnek. Megjegyezzük, hogy megengedünk negatív súlyokat is, azaz az ún. *rövidre eladás* is lehetséges. A feladat megoldásakor kikötjük továbbá, hogy  $-1 \leq p_i \leq 1$ . Ez segít elkerülni olyan, kevésbé realisztikus helyzeteket, ahol az optimális megoldásban egy-egy részvénybe a tőkénk többszörösét fektetnénk be (pl. hitelfelvétel segítségével).

## 2.1 Kovarianciamátrixok szűrése

Markowitz modellje több korlátba is ütközik. Megmutatható például, hogy a kapott megoldás (az optimális portfólió) rendkívül érzékeny a feladat paramétereinek változásaira. A kvadratikusan optimalizálási feladat a részvények hozamaiból képzett  $\Sigma$  kovarianciamátrix invertálását is igényli, ami egy közel szinguláris mátrix esetében instabil megoldásokhoz vezethet. Továbbá, mivel véges ( $T$ -hosszú) idősorok alapján becsüljük a kovarianciákat, a hozamok közötti összefüggések nem-stacionárius jellege miatt a kovarianciamátrix gyakran hordoz magában statisztikai bizonytalanságot (egyszerűen mondva zajt), és ez rossz kockázatbecsléshez vezethet (De Prado 2016). Utóbbit igyekeznek orvosolni a különböző szűrési eljárások, mint például a véletlen mátrixok elmélete (*Random Matrix Theory*, RMT) (Bun, Bouchaud és Potters 2017), vagy a hierarchikus klaszterezés (Tola és tsai. 2008). A módszerek eredményeképpen egy tisztított (vagy „szűrt”)  $\Sigma_f$  kovarianciamátrixot kapunk, amelyet a Markowitz-modell célfüggvényében használhatunk az eredeti  $\Sigma$  helyett.

## 2.2 Hierarchikus klaszterezés mint szűrési eljárás

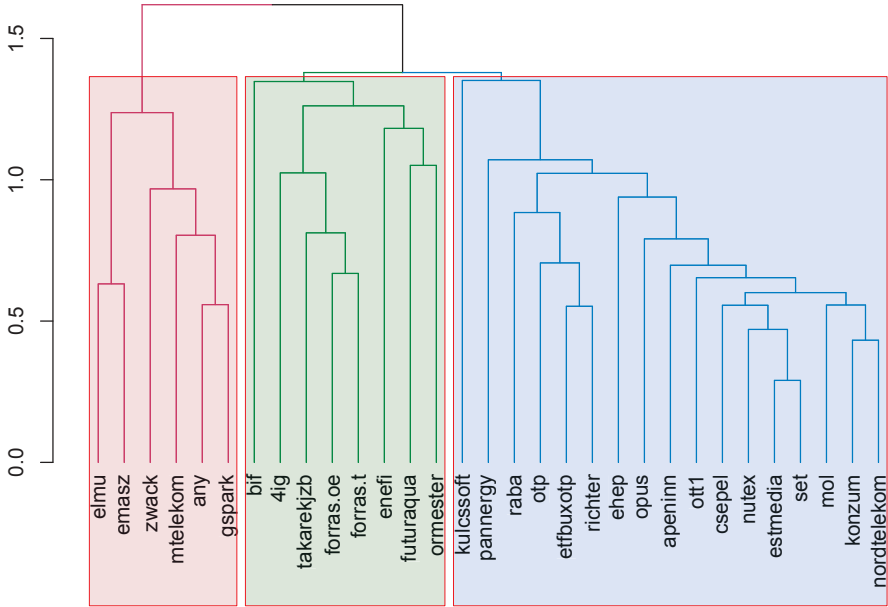
Egy  $n$  részvényből képzett  $\Sigma$  kovarianciamátrix egy  $n$  csúcsú teljes gráf szomszédsági mátrixaként is értelmezhető. Ugyanez igaz a  $\mathbf{C} = (C_{ij})_{i,j}$  korrelációs mátrixra, mely a normalizált kovarianciákat tartalmazza. A szűrési eljárás leírásához ezt fogjuk használni. Első lépésben a korrelációs mátrixból egy távolságmátrixot készítünk. Ezt Tumminello, Lillo és Mantegna (2007) és Tola és tsai. (2008) mentén a

$$D_{ij} = \sqrt{2(1 - C_{ij})} \quad (5)$$

ultrametrikus távolságfogalom segítségével tesszük meg (mely esetén a  $D_{ij} \leq \max\{D_{ik}, D_{kj}\}$  ( $\forall i, j, k$ ) feltétel teljesül), a nagy korrelációkat kis távolsággá konvertálva így. A hierarchikus klaszterezést a  $\mathbf{D} = (D_{ij})_{i,j}$  mátrixon hajtjuk végre.

A klaszterezéshez egy agglomeratív módszert alkalmazunk, ami kezdetben minden csúcsot külön klaszterbe sorol, majd lépésenként a két legkisebb távolságú klasztert összeolvasztja egy új klaszterré, mindezt addig ismételve, amíg a végén egyetlen, az összes részvényt tartalmazó klasztert kapunk (Nielsen 2016). Lényeges kérdés, hogy hogyan definiáljuk két klaszter távolságát, amely alapján az összevonást végezzük. A népszerű távolságfogalmak között szerepel a *“single-linkage”*, amely szorosan kötődik minimális feszítőfák kereséséhez (Mantegna 1999), a *“complete-linkage”* és az *“average-linkage”*, melyek rendre az alábbi módon definiálják egy  $\mathcal{C}_1$  és  $\mathcal{C}_2$  klaszter távolságát:

$$\begin{aligned} d_{\text{SL}}(\mathcal{C}_1, \mathcal{C}_2) &= \min_{i \in \mathcal{C}_1, j \in \mathcal{C}_2} D_{ij}, \\ d_{\text{CL}}(\mathcal{C}_1, \mathcal{C}_2) &= \max_{i \in \mathcal{C}_1, j \in \mathcal{C}_2} D_{ij}, \\ d_{\text{AL}}(\mathcal{C}_1, \mathcal{C}_2) &= \frac{1}{|\mathcal{C}_1| \cdot |\mathcal{C}_2|} \sum_{i \in \mathcal{C}_1} \sum_{j \in \mathcal{C}_2} D_{ij}. \end{aligned} \quad (6)$$



2. ábra. Egy hierarchikus klaszterezés összeolvasztási fája dendrogramként ábrázolva,  $k = 3$  színnel színeztve

Az algoritmus egy bináris összeolvasztási fát ad eredményül, amely tárolja, hogy mely távolságnál mely csúcsokat vagy klasztereket olvasztotta össze. A fa levelei az egyes csúcsok (részvények) lesznek, a gyökere pedig az összes csúcsot tartalmazó klaszter. Ezt ábrázolhatjuk dendrogramként (ld. 2. ábra), illetve elvághatjuk a megfelelő hierarchia szinten úgy, hogy pontosan  $k$  darab (az ábrán látható példán  $k = 3$ ) klasztert kapjunk.

A klaszterezés eredményének egy fontos tulajdonságát használhatjuk ki: felépítettünk egy fát a csúcsainkból  $n - 1$  összevonással. Ekkor helyettesíthetjük a korábbi  $\mathbf{D}$  távolságmátrixunkat az összevonási távolságokat tartalmazó, összesen  $n - 1$  különböző értékkel rendelkező  $\mathbf{D}_{\text{hclust}}$  mátrixszal, majd az (5) képletet felhasználva visszaírhatjuk azt egy  $\mathbf{C}_{\text{hclust}}$  korrelációs mátrixszá, amit egy  $\mathbf{\Sigma}_{\text{hclust}}$  kovarianciamátrixszá visszaalakítva használhatunk a (4) optimalizálási feladatban.

### 2.3 Egy null modell és mohó algoritmus

A hierarchikus klaszterezés eredményét alternatív módon is használhatjuk a Markowitz-modell esetén. Az általunk alkalmazott módszer lényege, hogy létrehozunk egy  $n \times n$ -es  $\mathbf{C}^0$  korrelációs mátrixot, amelyben  $C_{ij}^0$  az  $i$  és  $j$  részvények közötti átlagos korrelációt jelöli valamilyen null modell esetén. Többféle null modellt is alkalmazhatunk. Egy lehetőség, hogy feltesszük, hogy minden részvény korrelálatlan – ekkor  $\mathbf{C}^0$  az  $n \times n$ -es egységmátrix lenne. Itt egy, a  $\mathbf{C}^0$ -t generáló konfigurációs mátrixot használunk, amelyet a Masuda, Kojaku és Sano (2018) cikkben megadott iteratív eljárással lehet előállítani.

A feltétel az, hogy  $\mathbf{C}^0$ -ban minden  $i$  részvény esetén a  $C_i^0 = \sum_j C_{ij}$  ún. korrelációs erősség legyen állandó, viszont az egyes  $C_{ij}$  ( $j = 1, \dots, n$ ) korrelációkat véletlenül generáljuk.

A két mátrixot ( $\mathbf{C}^0$ -t és  $\mathbf{C}$ -t) egy különbségmátrix képzésére használjuk fel és újraszkalázzuk:

$$\begin{aligned} \mathbf{C}' &= |\mathbf{C} - \mathbf{C}^0|, \\ \mathbf{D}_C &= -\mathbf{C}' + |\min \mathbf{C}'| + |\max \mathbf{C}'|. \end{aligned} \quad (7)$$

Az így kapott  $\mathbf{D}_C$  mátrix egy, a korrelációs mátrixhoz köthető súlyozott gráfként értelmezhető. Az újraszkalázást azért végezzük el, mert a hierarchikus klaszterezés minden lépésben a legkisebb távolságok alapján végez el összevonást, ami éppen a korrelációs mátrix és a null modellje közötti legnagyobb eltéréseket jelenti.

A  $D_c$  mátrixon végrehajtott hierarchikus klaszterezési eljárás tulajdonképpen ekvivalens az alábbi, *modularitás jellegű függvény* (Newman 2006)

$$M = \sum_{i,j} |C_{ij} - C_{ij}^0| \cdot \delta_{ij} \quad (8)$$

mohó algoritmus szerinti maximalizálásával. A  $\delta_{ij}$  egy bináris változó, amely megmondja, hogy  $i$  és  $j$  egy klaszterbe tartoznak-e, azaz  $\delta_{ij} = 1$ , ha  $i$  és  $j$  azonos klaszterben vannak, különben  $\delta_{ij} = 0$ .

Ezzel az interpretációval nemcsak egy klaszterezést tudunk megadni, hanem egyetlen számmal tudjuk jellemezni ennek a klaszterezésnek a „jóságát”. Azt a klaszterezést tekintjük, amelyre  $M$  értéke maximális, majd úgy állítjuk össze a kovarianciamátrixot a Markowitz-modellhez, hogy egy klaszterből csak egy részvényt választunk véletlenszerűen (itt ezt a megoldást választottuk), vagy valamilyen más megfontolás mentén.

### 3 Hierarchikus klaszterezés közvetlen portfólió-kiválasztásra

Amennyiben szeretnénk átlépni a Markowitz-modell bizonyos korlátozásait, más diverzifikációs technikákra van szükségünk. Több, a közelmúltban megjelent módszer is a hierarchikus klaszterezésre épít, ld. De Prado (2016), Raffinot (2017) és Raffinot (2018), akárcsak a korábban ismertetett szűrési eljárások. Ezek fő céljai, hogy elkerüljék a kovarianciamátrix invertálását, és közvetlenül használják ki a vizsgált rendszer hierarchikus szerkezetét a portfólió építésekor. Ennek ugyanakkor ára is van. A módszerek hátránya ugyanis, hogy a Markowitz-modellhez képest kevesebb a paraméterezési lehetőség: ezen módszerek esetén nem tudunk elvárt minimális hozamot ( $R$ ) specifikálni.

Fontos kiemelni azt is, hogy gráf klaszterező módszerek közvetlen alkalmazása, eljárástól függetlenül, kovariancia vagy korrelációs mátrixok esetében torzított eredményekhez vezethet, mivel az erősen korreláló csúcspárok nem

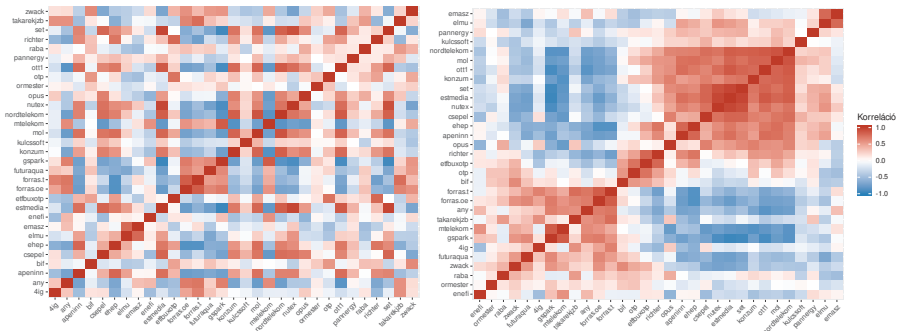
feltétlen kerülnek egy klaszterbe. Ezt a problémát többféleképpen is orvosolhatjuk (MacMahon és Garlaschelli 2013). Először a Gera és London (2020) cikk nyomán mutatunk be egy lehetséges megközelítést.

### 3.1 A HRP módszer

A 2016-ban bemutatott *Hierarchical Risk Parity* (HRP) algoritmus (De Prado 2016) három lépésben számolja ki az optimálisnak vélt portfólió súlyozást: először végrehajt egy hierarchikus klaszterezést, majd egy kvázi-diagonalizálást a kovarianciamátrixon, végül egy rekurzív vágást a klaszterezés eredményén, amely során minden részvényhez azok és a klaszterek teljes kockázathoz való hozzájárulása alapján súlyt rendel.

Az első lépésben a 2.2 szakaszban bemutatott eljárást alkalmazzuk, de nem hajtjuk végre a kovarianciamátrix szűrését, azaz csak az összevonási fát állítjuk elő a korrelációs mátrixból kapott távolságmátrixból.

A második lépésben átrendezzük a  $\Sigma$  kovarianciamátrix sorait és oszlopait úgy, hogy a legnagyobb értékek a főátló mentén helyezkedjenek el. Mindezt úgy érhetjük el, hogy az első lépésben végrehajtott klaszterezés eredménye alapján a dendrogram leveleinek sorrendjét használjuk a kovarianciamátrix átrendezésére is. A pontos algoritmust a De Prado (2016) cikk ismerteti, a 3. ábra illusztrálja a művelet eredményét egy korrelációs mátrixon.



3. ábra. Korrelációs mátrix átrendezés előtt és után

A harmadik lépésben az összevonási fát a gyökértől indulva lépésenként kettévágjuk. Minden kettévágáskor a kapott két új klaszter között aszerint osztjuk szét a szülő súlyát, hogy mekkora egy-egy klaszter hozzájárulása a szülő varianciájához. Mindezt addig ismételjük, amíg minden részvény külön klaszterbe nem esik. Az eljárás végén megkapjuk az egyes részvények súlyát az összeállítandó portfólióban.

### 3.2 A HERC módszer

A *Hierarchical Equal Risk Contribution Portfolio* (HERC) (Raffinot 2018) az előbb bemutatott HRP algoritmus egy módosítása.



Ez az algoritmus is hierarchikus klaszterezést hajt végre első lépésként, viszont ezt követően egy optimális klaszterszámot határoz meg, ami megmondja, hogy hányféle kell elvágni a bináris összeolvasztási fát. Mindezt egy statisztikai mutató alapján teszi meg – a Raffinot (2018) munkában az ún. *Gap indexet* (Tibshirani, Walther és Hastie 2001) számítják ki minden lehetséges  $k$  klaszterszámra és az algoritmus azt a  $k$ -t választja, ahol a mutató a legnagyobb értéket veszi fel.

Miután meghatároztuk a klaszterek optimális számát, először szétszjtjuk a tőkét a klaszterek között (a HRP-ből megismert variancia szétszjtással), majd a klasztereken belül. Erre a HRP-ben is alkalmazott rekurzív vágást alkalmazhatjuk, méghozzá úgy, hogy csak addig vágunk és osztjuk ketté a súlyokat valamilyen kockázatmetrika (pl. variancia) szerint, amíg elértük a  $k$  darab klasztert. Miután ez megtörtént, a klaszterekben lévő részvények között egyenlően osztjuk szét a klaszterre eső súlyt.

## 4 Az eljárások összehasonlítása

### 4.1 Adatok és módszertan

Kísérleteinket valós adatokon, a Budapesti Értéktőzsde (BSE) 31 részvényének napi záróárainak felhasználásával végeztük. Az adathalmaz részvényenként 1962 rekorddal rendelkezik, a 2011. november 29-től 2019. október 18-ig megfigyelt napi záróárakat tartalmazza.

A befektetési folyamat modellezésére egy szimulációs környezetet állítottunk fel, amely egy adott  $t_0$  kezdőidőpontban elvégzi a különböző portfóliók összeállítását és kiértékelését is. Ezt követően  $t_0$  értékét 10-zel növeltük és újra elvégeztük a kiértékelést. Gördülőablak megközelítés szerint a növelést addig ismételtük, amíg az adathalmaz végére értünk.

### 4.2 Értékelési kritériumok

Egy eredményül kapott  $\mathbf{p}$  portfólióra három fő kategóriában számoltunk mutatókat: meghatároztuk a portfólió hozamait, kockázatait és a méretét. Hozamok és kockázatok esetében beszélhetünk *becsült* és *realizált* értékekről. A becsült értékeket a múltbeli megfigyelések alapján a portfólió összeállításakor, míg a realizált értékeket a portfólió eladásakor (a befektetés után  $T$  nappal) számítjuk ki. Ezek viszonya segíthet feltárni, hogy az egyes módszerek mennyire becsülik jól a kockázatot vagy a hozamot.

#### 4.2.1 Hozamok

Egy  $\mathbf{p}$  portfólió esetében kétféle hozamot számítunk. A *becsült* hozam megmutatja, hogy a befektetés pillanatában a kapott portfóliósúlyokkal milyen hozamra lehetne szert tenni. Ezt meghatározhatjuk a múltbeli hozamok

súlyozott átlagaként:

$$r_{\text{pre}} = \sum_{i=1}^n p_i \cdot \bar{x}_i, \quad (9)$$

ahol  $\bar{x}_i$  a befektetési napot megelőző  $T$  napra számolt hozam.

Eladáskor (a befektetés után  $T$  nappal) meghatározható a portfólió utólagos (vagy *realizált*) hozama, amely megmondja, hogy a részvény ára eladáskor hogyan viszonyult a vételárhoz:

$$r_{\text{post}} = \sum_{i=1}^n p_i \cdot \frac{P_i(t_0 + T) - P_i(t_0)}{P_i(t_0)}. \quad (10)$$

Ez az érték akkor lesz 0, ha pontosan a befektetési árat kaptuk vissza, pozitív esetben nyereségesnek, negatív esetben veszteségesnek mondható a befektetés. Természetesen a befektetés időbeli alakulását is nyomon követhetjük a kumulatív hozamok segítségével (ld. 4. ábra).

#### 4.2.2 Kockázatok

A portfóliók egy másik fontos tulajdonsága a kockázatuk, valamint az őket előállító módszerek kockázatbecslésének pontossága. Egy portfólió becsült és realizált kockázata meghatározható a

$$\sigma_{\text{pre}} = \mathbf{p} \Sigma_{\text{pre}} \mathbf{p}^{\top}, \quad (11)$$

$$\sigma_{\text{post}} = \mathbf{p} \Sigma_{\text{post}} \mathbf{p}^{\top} \quad (12)$$

összefüggésekkel, ahol  $\Sigma_{\text{pre}}$  a  $[t_0 - T, t_0]$  időszakon számolt és a portfólió-kiválasztás során is felhasznált,  $\Sigma_{\text{post}}$  pedig a  $[t_0, t_0 + T]$  időszakon számolt kovarianciamátrix.

Ebből a két metrikából meghatározhatjuk, hogy egy portfólió esetében hogyan viszonyult egymáshoz a realizált és becsült hozam. A *kockázati hányados*

$$\sigma_r = \left| \frac{\sigma_{\text{post}}}{\sigma_{\text{pre}}} \right| \quad (13)$$

a két mutató arányát méri, ami annál jobb, minél közelebb van az 1-hez. Ha a kockázati hányados 1-nél nagyobb, akkor alulbecslésről, ellenkező esetben túlbecslésről beszélünk.

Hasonló módon definiálhatjuk a *kockázat változását* is:

$$\sigma_g = \left| \frac{\sigma_{\text{post}} - \sigma_{\text{pre}}}{\sigma_{\text{pre}}} \right|, \quad (14)$$

amely annál jobb, minél közelebb van a 0-hoz.

### 4.2.3 Portfólióméret

Egy befektetés során nem csak a részvények árát kellhet megfizetnünk, hanem például tranzakciós költségek is felmerülhetnek. Ebben az esetben segíthet, ha a portfóliónk kevesebb részvényből áll, azonban ez a kockázatot is megnövelheti a diverzifikáció mértékének csökkentésével. Egy  $\mathbf{p}$  portfólió mérete meghatározható a portfóliósúlyok négyzetösszegének reciprokaként, de ez torzított eredményekhez vezethet olyan helyzetekben, ahol negatív portfóliósúlyok is megengedettek. Ennek kompenzálására a súlyokat először leosztjuk a portfólióvektor 1-es normájával (az elemeinek abszolút értékének összegével):

$$\mathcal{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^n \left(\frac{p_i}{\|\mathbf{p}\|_1}\right)^2} \quad (15)$$

A portfólióméret akkor lesz 1, ha az összeállított portfólióban egyetlen részvénybe van koncentrálna a tőkénk, és akkor lesz  $n$ , ha minden részvénybe egyenlő arányban fektettünk be. Ezt lehet még normalizálni, hogy különböző méretű adathalmazokon is összehasonlítható legyen az értéke, és ne függjön  $n$ -től:

$$\hat{\mathcal{N}}_{\text{eff}} = \frac{\mathcal{N}_{\text{eff}} - 1}{n - 1}. \quad (16)$$

Ez az érték így 0 és 1 között mozog majd, ahol 0 az 1-részvényes portfóliót, 1 pedig az  $n$ -részvényes egyenletes súlyozást jelenti.

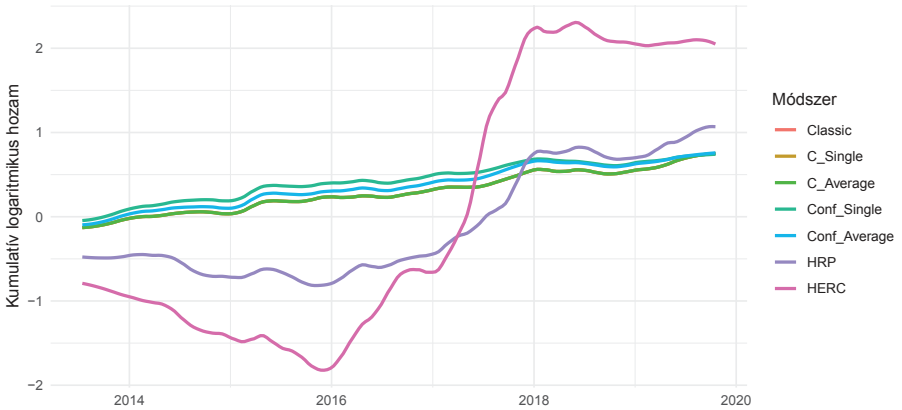
## 4.3 Eredmények

Az alábbiakban mutatónkénti lebontásban hasonlítjuk össze a vizsgált módszerek teljesítményét. A számításokban a mutatók (összes összeállított portfólióra vett) átlagait és szórásait vizsgáljuk, a Markowitz-modell elvárt hozam paraméterét ( $R$ ) az átlagos hozamra rögzítettük.

### 4.3.1 Hozamok

Becsült hozamok esetében a hierarchikus klaszterezésen alapuló portfólió kiválasztó eljárások átlagosan gyengébben teljesítettek (sokszor negatív várható hozammal), a realizált hozamaik viszont sokkal jobbak voltak. Megfigyelhető volt, hogy két részvény (az OPUS és KONZUM) árfolyama 2016 és 2018 között több nagyságrenddel megnőtt, például az OPUS ára a 2016. január közepe 4,4 Ft-ról két évvel később ugyanezen a napon 708 Ft volt. A HRP és HERC algoritmusok ezt a növekedést jól előrejelezték, és nagy arányban fektettek be ezekbe a részvényekbe (akár 30-40%-os súlyokkal), aminek köszönhetően a realizált hozam mutató is óriásira nőtt (akár 2000%-ra is a HERC esetében, 365%-ra a HRP esetében).

Ehhez képest a Markowitz-modell és kiterjesztései rosszabbul teljesítettek: az eredeti modell átlagosan 40,9%-os, a hierarchikus klaszterező szűrő módszerek ennél kicsit jobb, 41%-os, a konfigurációs modell alapú módszer pedig average-linkage esetben 1,1%-os, single-linkage esetben -35%-os realizált hozamot produkált.



4. ábra. Kumulatív logaritmusos hozamok egy 2013. június 12-i befektetésre, simítva a trendek láthatóbbá tétele végett. Az eredeti Markowitz-modell („Classic”) és a hagyományos hierarchikus klaszterező szűrések („C\_Single” és „C\_Average”) rendkívül hasonló eredményeket produkáltak, ezért azok fedik egymást.

### 4.3.2 Kockázatok

Az eredeti Markowitz modell és kiegészítései alacsony becslült kockázatokkal számoltak, míg a HRP egy nagyságrenddel, a HERC három nagyságrenddel nagyobb kockázatot jósolt. A realizált kockázatok hasonló képet mutattak a becslésekhez: többségében azok a módszerek, amelyek magasra becsülték a kockázatot (HERC, HRP), magasabb kockázatot is realizáltak a végén. Kivételt képzett az eredeti Markowitz-modell, amely a legalacsonyabb becslést adta, viszont ennek ellenére az összes szűrési eljárásnál magasabb realizált kockázattal rendelkezett.

Ennél tisztább képet kaphatunk a módszerek kockázatbecsléséről, ha megnézzük a kockázati hányadost ( $\sigma_r$ ) is. Megfigyelhető, hogy ez az eredeti Markowitz-modell esetén volt a legmagasabb (3,56), amin keveset javított a hierarchikus klaszterezés (3,53) és lényegesen többet a konfigurációs modell (average-linkage esetén 2,05, single-linkage esetén 1,76). A HRP és HERC algoritmusok nagyon eltérően viselkedtek, a HRP ugyanis (a Markowitz-moddal együtt) alulbecsülte, bár minden módszer közül a legkevésbé – itt 1,55 lett a kockázati hányados –, míg a HERC jelentősen túlbecsülte a kockázatot ( $\sigma_r = 0,12$ ). A kockázat változása tekintetében azt láthatjuk, hogy az alulbecslés ellenére a HERC esetén volt a legkisebb a kockázat változása ( $\sigma_g = 0,88$ ), ezt követte a single-linkage konfigurációs modell ( $\sigma_g = 0,95$ ) és average-linkage konfigurációs modell ( $\sigma_g = 1,27$ ). A HRP 1,35-ös értéket produkált, míg a legnagyobb változást az eredeti Markowitz-modell mutatta (2,63).

### 4.3.3 Portfólióméret

A különböző módszerek hasonló méretű portfóliókat állítottak össze, több módszer normalizált mutatója is 0,35 körül helyezkedett el. A kivételek kö-

zött szerepelt a HRP, ami kicsit nagyobb portfóliókat állított össze ( $\hat{N}_{\text{eff}} = 0,4$ ), minden közül ennek a módszernek volt a legnagyobb a portfóliómérete. A másik hierarchikus klaszterezésen alapuló portfólió kiválasztó eljárás is kivételt képezett, a HERC ugyanis a többi eljáráshoz képest átlagosan koncentráltabb portfóliókat állított össze ( $\hat{N}_{\text{eff}} = 0,1$ ), ami összevethető a felismeréssel, hogy ez a módszer úgy tudott a HRP-nél is nagyobb hozamokat elérni, hogy még jobban koncentráta a tőkét a nagy áremelkedésen áteső részvényekbe.

	C_Average	HERC	HRP	Markowitz	Conf_Average	Conf_Single
$r_{\text{pre}}$	0,0004 (0,00)	-0,0017 (0,00)	-0,0003 (0,00)	0,0004 (0,00)	0,0005 (0,00)	0,0006 (0,00)
$r_{\text{post}}$	0,4104 (0,55)	20,5447 (31,04)	3,6523 (5,17)	0,4089 (0,56)	0,0109 (0,44)	-0,3498 (0,78)
$\sigma_{\text{post}}$	0,0001 (0,00)	0,0011 (0,00)	0,0004 (0,00)	0,0001 (0,00)	0,0001 (0,00)	0,0001 (0,00)
$\hat{N}_{\text{eff}}$	0,3350 (0,07)	0,1248 (0,10)	0,3967 (0,07)	0,3353 (0,07)	0,3546 (0,05)	0,2816 (0,04)
$\sigma_r$	3,5336 (3,93)	0,1174 (0,31)	1,5467 (3,33)	3,5631 (3,94)	2,0546 (1,95)	1,7578 (1,34)
$\sigma_g$	2,5984 (3,86)	0,8828 (0,28)	1,3548 (2,93)	2,6296 (3,88)	1,2690 (1,79)	0,9533 (1,18)

1. táblázat. A Budapesti Értéktőzsde adatain végzett elemzés eredményeinek összesítése: a cellákban az átlagok, zárójelben a szórások szerepelnek. A Markowitz-modell esetében az átlagos várt hozamokra összeállított portfóliók adatai szerepelnek.

## 4.4 Összegzés

Ezen munkában hierarchikus klaszterezésen alapuló, portfólió-kiválasztásra alkalmazható eljárásokat hasonlítottunk össze, melyek között szerepeltek más portfólió kiválasztó modellek kiegészítésére szolgáló módszerek (szűrések) és tisztán hierarchikus klaszterezésre épülő, új portfólió kiválasztó módszerek is. Valós adatokon végzett vizsgálataink azt mutatják, hogy az új portfólió kiválasztó eljárások egy kisebb (31 részvényes) adathalmazon képesek voltak sokkal magasabb hozamokat produkálni, mint az eredeti Markowitz-modell és annak kiterjesztései, azonban ehhez magasabb kockázat és néhol kevésbé diverzifikált portfólió is társult. A módszerek nagyobb adathalmazokon (pl. FTSE-100, S&P500), illetve más eljárásokkal (véletlen mátrixok elméletén alapuló kovariancia mátrix szűrések, részvény-volatilitás alapú modellek) való átfogó összehasonlítását egy lehetséges továbblépésnek gondoljuk.

## Köszönetnyilvánítás

A szerzők köszönetet mondanak a cikk bírálójának az értékes tanácsokért. London András a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal az NKFI-SSN-135643 számú projekt keretében támogatta.

## Irodalom

1. Bongiorno, Christian és Challet, Damien (2020) „Nonparametric sign prediction of high-dimensional correlation matrix coefficients”. *EPL (Europhysics Letters)*, 133(4), 48001. doi: 10.1209/0295-5075/133/48001
2. Bun, Jo l, Bouchaud, Jean-Philippe és Potters, Marc (2017) „Cleaning large correlation matrices: tools from random matrix theory”. *Physics Reports*, 666, 1–109. doi: 10.1016/j.physrep.2016.10.005.

3. De Prado, Marcos Lopez (2016) „Building diversified portfolios that outperform out of sample”. *The Journal of Portfolio Management*, 42(4), 59–69. doi: 10.3905/jpm.2016.42.4.059.
4. Embrechts, Paul, McNeil, Alexander és Straumann, Daniel (2002) „Correlation and dependence in risk management: properties and pitfalls”. In Dempster, M. A. (ed.) *Risk management: value at risk and beyond*, 176–223. doi: 10.1017/CBO9780511615337.008.
5. Gera, Imre és London, András (2020) „Gráf alapú dimenzióredukciós heurisztikák részvénytársi korrelációs mátrixokra”. *Alkalmazott Matematikai Lapok* 37(2), 1–14. doi: 10.37070/AML.2020.37.2.01.
6. Laloux, Laurent és tsai. (1999) „Noise dressing of financial correlation matrices”. *Physical Review Letters*, 83(7), 1467. doi: 10.1103/PhysRevLett.83.1467.
7. London, András, Gera, Imre és Bánhelyi, Balázs (2018) „Markowitz Portfolio Selection Using Various Estimators of Expected Returns and Filtering Techniques for Correlation Matrices”. *Acta Polytechnica Hungarica*, 15(1), 217–229.
8. MacMahon, Mel és Garlaschelli, Diego (2013) „Community detection for correlation matrices”. *Physical Review*, E 5, 21006. doi: 10.1103/PhysRevX.5.021006.
9. Mantegna, Rosario N (1999) „Hierarchical structure in financial markets”. *The European Physical Journal B*, 11(1), 193–197. doi: 10.1007/s100510050929.
10. Markowitz, Harry (1952) „Portfolio selection”. *The Journal of Finance*, 7(1), 77–91.
11. Masuda, Naoki, Kojaku, Sadamori és Sano, Yukie (2018) „Configuration model for correlation matrices preserving the node strength”. *Physical Review E*, 98(1), 12312. doi: 10.1103/PhysRevE.98.012312.
12. Newman, Mark EJ (2006) „Modularity and community structure in networks”. *Proceedings of the National Academy of Sciences* 103.23, 8577–8582. doi: 10.1073/pnas.0601602103.
13. Nielsen, Frank (2016) „Hierarchical Clustering”. 195–211. isbn: 978-3-319-21902-8. doi: 10.1007/978-3-319-21903-5\_8.
14. Raffinot, Thomas (2017) „Hierarchical clustering-based asset allocation”. *The Journal of Portfolio Management*, 44(2), 89–99. doi: 10.3905/jpm.2018.44.2.089.
15. Raffinot, Thomas (2018) „The hierarchical equal risk contribution portfolio”. Available at SSRN 3237540. doi: 10.2139/ssrn.3237540.
16. Simon, Herbert A (1991) *The architecture of complexity*. Facets of Systems Science. Springer, 457–476. doi: 10.1007/978-1-4899-0718-9\_31.
17. Tibshirani, Robert, Walther, Guenther és Hastie, Trevor (2001) „Estimating the number of clusters in a data set via the gap statistic”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2), 411–423. doi: 10.1111/1467-9868.00293.
18. Tola, Vincenzo és tsai. (2008) „Cluster analysis for portfolio optimization”. *Journal of Economic Dynamics and Control*, 32(1), 235–258. doi: 10.1016/j.jedc.2007.01.034.
19. Tumminello, Michele, Lillo, Fabrizio és Mantegna, Rosario N (2007) „Hierarchically nested factor model from multivariate data”. *Europhysics Letters*, 78(3), 30006. doi: 10.1209/0295-5075/78/30006.

HIERARCHICAL CLUSTERING AND THE PORTFOLIO SELECTION  
PROBLEM

Classical portfolio selection techniques measure risk through the variance of asset returns, which investors want to minimise. These methods only work well in practice if the covariance (or correlation) matrix of the returns is properly de-noised. A good de-noising (or filtering) method conserves only the useful and informative parts of the matrix' structure. Since stock markets may also be viewed as hierarchical structures (Simon 1991), we can use hierarchical clustering to identify this hierarchy and use the detected structure to filter the covariance matrix. In this work we show possible applications of hierarchical clustering both for using it as a matrix filtering technique in the Markowitz portfolio model and directly as a way to assemble portfolios. We perform our experiments on real world data and evaluate using different performance indicators. We work with a data set containing daily closing prices and transform these into daily logarithmic returns from which we create the covariance matrix. We model the following situation: the investor decides the length  $T$  of the investment. On day  $t_0$  they create a portfolio based on the previous  $T$  days of returns and then they wait for  $T$  days, selling all of their assets afterwards.

We present two ways of applying hierarchical clustering for the portfolio problem. Firstly it is considered as a de-noising technique of the estimated covariance matrix, secondly it is used directly to create portfolios.

Diversification is one of the key tools of risk reduction, where we invest into multiple assets instead of just one. One of the first models to incorporate diversification was the Markowitz-model, which is a quadratic optimisation problem that seeks to minimise the risk for a given expected return. As a result we obtain a portfolio vector containing the fraction of wealth to be invested in each individual asset. To explicitly solve the problem, the use of an inverted covariance matrix, calculated from the logarithmic returns, is required. Since covariance is estimated from finite length time series and because of the non-stationary nature of the returns and their relationship, the covariance matrix frequently contains statistical uncertainty (it is often called noise), which can lead to wrong risk estimation (De Prado 2016). Various filtering techniques have been created to solve this issue, including the use of random matrix theory (Bun, Bouchaud, and Potters 2017) or hierarchical clustering (Tola et al. 2008). We perform filtering by first converting covariance to correlation, and then correlation to distances, where higher correlation means smaller distance. We then perform hierarchical clustering on the resulting distance matrix using an agglomerative method. The distance matrix contains  $n - 1$  different distances (instead of the original  $n \cdot (n - 1) / 2$ ) that we then convert back to covariances. We can use multiple distance definitions when performing clustering. In this study we use single-linkage, complete-linkage and average-linkage distances. Another approach is to use a null model to generate a so-called configuration matrix. We calculate the difference between the correlation and the configuration matrix and use that as a distance matrix to perform hierarchical clustering.

In order to avoid limiting ourselves to the constraints of the Markowitz-model and solve a quadratic optimisation problem, we may use the Hierarchical Risk Parity (HRP) (De Prado 2016) or Hierarchical Equal Risk Contribution Portfolio (HERC) (Raffinot 2018) algorithms to directly assemble an optimal portfolio. Both of these algorithms use hierarchical clustering to assign weights to the assets in the portfolio. While the HRP algorithm assigns different weight to each asset, the HERC algorithm finds an optimal number of clusters and distributes weights equally among the assets of each cluster.

We performed our experiments on historical data from Budapest Stock Exchange (BSE). The data set contained 1962 daily closing prices for each of the 31 stocks, from 29th November 2011 to 18th October 2019. We used a rolling window approach

to evaluate optimal portfolios assembled at different starting points. We calculated metrics in three main categories: portfolio return, risk and size. We measured both the estimated and realised risks and returns, where the estimated measure was calculated based on past observations (before day  $t_0$ ) and the realised measure was calculated at the time of selling the portfolio. The ratio of these two helps us define how well a method estimates the return and risk. We also define a modified portfolio size measure. Instead of simply taking the reciprocal sum of the squared portfolio weights (which can lead to distorted results when negative portfolio weights, e.g. in the case of short selling, are allowed), we first divide the weights by the  $L_1$ -norm of the portfolio vector. We also normalise the result by the number of assets ( $n$ ) to get a value that is always in the  $[0, 1]$  range.

In our experiments the hierarchical clustering-based portfolio selection methods performed worse (often reaching negative values) in terms of estimated returns, but extremely well in terms of realised returns. This is thanks to the fact that these methods were able to predict the huge increase in the price of two assets and assigned large weights to them. Regarding risk the Markowitz-model and its extensions predicted lower risks than the HRP and HERC algorithms. Realised risks showed a trend consistent with the estimations: methods that estimated lower risks usually realised lower risks as well. An exception was the Markowitz-model that predicted the lowest risk but realised the highest one compared to all the filtering methods. The risk ratio was thus the highest in the case of the Markowitz-model, followed by the hierarchical clustering filtering methods and the configuration model. The HRP algorithm also underestimated the risk, although by the lowest amount, while the HERC method highly overestimated it. Portfolio sizes were similar (near 0.35) across almost all methods, where the only differences were made by the HRP algorithm that assembled larger portfolios (with a size of 0.4) and the HERC that created more concentrated ones (with a size of 0.1). This is consistent with the fact that the HERC method could achieve higher realised returns by concentrating capital into stocks whose prices rose greatly.