

Budapest Corvinus Egyetem

Gazdálkodástudományi Kar

Befektetések és Vállalati Pénzügy Tanszék

A piaci hatékonyság nyomában

A webes hírszoment automatizált elemzésének lehetőségei

Készítette: Gajdos András

Pénzügy MSc

Befektetéselemző szakirány

2014

Tartalom

Tartalom.....	2
1. Bevezetés	4
1.1. A dolgozat felépítése.....	5
2. Korlátlan és korlátozott racionalitás.....	6
2.1. Kritikák és anomáliák – az elméletek hiányossága.....	8
2.1.1. Stilizált tények, anomáliák.....	8
3. A webes tartalmak: hír és szentiment.....	13
3.1. A szentiment.....	13
3.2. A hírek világa, a szentiment forrása.....	14
3.2.1. A szentiment másodlagos forrásai.....	14
3.2.2. A hírek: a szentiment elsődleges forrásai.....	15
3.3. A piaci hírek és a piaci szentiment, mint alfa forrás.....	17
3.4. A szentiment kinyerésének technikai oldala.....	24
3.4.1. Első lépés: a források meghatározása és a tartalmak indexelése.....	24
3.4.2. Második lépés: szentimentelemzés.....	25
3.4.3. Néhány alpmódszer és további szempontok.....	26
4. Piaci tapasztalatok és néhány konkrét szentimentstratégia.....	29
4.1.1. Első stratégia: Iparág-rotáció.....	29
4.1.2. Második stratégia: Extrém szentiment nap (ESZN).....	32
4.1.3. Harmadik stratégia: Szentiment-visszafordulás.....	37
5. Egy saját stratégia: A globális szentiment és a devizaárfolyamok.....	42
5.1. Esettanulmányok.....	43
5.1.1. Következtetések az esettanulmányok alapján.....	45
5.2. Portfóliószimulációk.....	46
5.2.1. Monte-Carlo Portfóliószimulációk.....	48
5.2.2. Következtetések a portfóliószimulációk alapján.....	49
6. Összefoglalás és végkövetkeztetések.....	51

7. Irodalomjegyzék.....	53
8. Melléletek.....	57
8.1. Első melléklet: Portfóliószimuláció eredmények	57
8.2. Második melléklet: A saját stratégia kódjai.....	57

1. Bevezetés

Dolgozatomban különböző pénzügyi piacok hatékonyságát vizsgálom. A nyilvánosan elérhető információkra koncentrálva és elfogadva, hogy az új információk mozgatják a piacot, a vizsgálatot tipikusan 3 irányból lehet megközelíteni:

1. tudok-e extrahozamot elérni **több** információ feldolgozásával?
és/vagy
2. tudok-e extrahozamot elérni az információ **gyorsabb** feldolgozásával
és/vagy
3. tudok-e extrahozamot elérni az információ **pontosabb** feldolgozásával, azaz az (igaz) hírek jobb, mélyebb értelmezésével? Ami a kereskedés szempontjából ekvivalens, hogy jobban tudom-e előre jelezni, hogy hogyan fog reagálni a piac a hírekre?

A vizsgálathoz elvégzéséhez egy újszerű eszköz, a webes források (híportálok, blogok, közösségi oldalak...) tartalmainak automatizált elemzésének lehetőségeit fogom bemutatni és alkalmazni. A téma azonban nem a *hírkereskedés*¹. A dolgozatban a hangsúly ugyanis a hírek egy inkább kvalitatív tulajdonságán, a hangulatán van. A hangulatnak az algoritmizált meghatározásával a *szentimentelemzés* foglalkozik, de már itt fontos leszögezni, hogy ezt önmagában nem alkalmazhatjuk, párhuzamosan mindig szükség van a hírek konkrét tartalmának az elemzésére is (*tartalomelemzés*)². Ennek a két területnek a kombinálásával lehetőség nyílik arra, hogy egy adott termék (például vállalat) esetében a termékkel foglalkozó hírek elemzésével kialakítsunk a termék körüli hangulatot mutató időben dinamikus kvantitatív faktort.

Az intuícióm az volt, hogy a kialakult hangulat (érzet) egy stabil viszonyítási pont egy olyan világban, ahol csak a változás állandó. Az, hogy éppen hogyan nevezik azt az eszközt amire felfújódik egy piaci buborék, az mindig más, de a hangulat alakulása az esemény körül szabályszerűségeket mutat. Azt is mondják, hogy az emberek azt felejtik el legnehezebben, hogy hogyan érezték magukat, és ez még annál is nagyobb hatással van rájuk, mint hogy konkrétan mi is történt³. Ezt igaznak elfogadva az a kérdés adódik, hogy nem viselkednek-e hasonlóan a szereplők azokban az időszakokban, amikor hasonló impulzusok érik őket és

¹ Többnyire előre ismert időpontban bejelentésre kerülő statisztikák empirikus árfolyamhatásainak lekereskedése a bejelentés körül.

² Ugyanis nem mindegy, hogy egy hír az eladások esése fölött kesereg, vagy a vállalat menzájának romló kínálata fölött. Arról nem is beszélve, hogy első körben azt is el kell tudni dönteni, hogy melyik vállalatról is van szó.

³ "I've learned that people will forget what you said, people will forget what you did, but people will never forget how you made them feel." -Maya Angelou, saját fordítás.

emiatt hasonló lelkiállapotba kerülnek. Elfogadva továbbá, hogy ezek a szereplők képesek az árfolyamok mozgatására, arra vagyok kíváncsi, hogy a szentimentnek a beépítése egy otthoni eszközökkel is megvalósítható kereskedési stratégiába hatékonyabbá teszi-e azt.

1.1.A dolgozat felépítése

A dolgozat következő, **Hiba! A hivatkozási forrás nem található.** fejezetében hivatkozás jelleggel áttekintem a hagyományos, racionális, matematizált modellkereteket és azokat a közgazdasági elgondolásokat, amelyek a piacok és a gazdaság viselkedését magyarázni próbálják. Megértve ezek gyenge pontjait, megalapozom a korlátoltan racionális modellek létjogosultságát, melyek már elfogadják az emberi faktor, konkrétan az irracionális és a szentiment fontosságát. További megalapozásként áttekintem a legfontosabb anomáliákat, melyek megkérdőjelezhetik a piacok hatékonyságát, legalábbis az eddigi keretekben.

A 3. fejezetben részletesen bemutatom, hogy mit jelent és honnan származik az általam vizsgált szentiment. Végigveszem az automatizált hír és szentimentelemzés eddigi felhasználásait és eredményeit, itt kiemelve, hogy *az irodalom teljes áttekintésére törekedtem*. Szándékaim szerint itt tehát az összes releváns és elérhető kutatás eredményét bemutatom, hozzátéve, hogy a szakirodalom viszonylag fiatal és körülhatárolható, magyar szakirodalom pedig gyakorlatilag nem létezik. A fejezet egy szükségszerű technikai résszel zárul, ahol röviden bemutatom a módszertan gyakorlati és informatikai vonatkozásait.

A 4. fejezetben három eredményes és általam érdekesnek tartott, szentimentre építő stratégiát és portfólió szimuláció kapott helyet. Ezt követően (és tanulva hibáikból), az 5. fejezetben bemutatom és értékelem a saját stratégiámat, illetve annak eredményeit. Az értékelés során maximális objektivitásra törekedtem, és semmiképpen sem az általam használt faktor ténylegesnél jobb színben való feltüntetésére. Kiemelt igyekezettel szándékoztam elkerülni a kvantitatív stratégiák gyakori csapdáját, az adott időszakra és termékre való túltanítást, mely általában a mintán kívüli eredmények drasztikus romlását eredményezi.

Az utolsó, 6. fejezetben, összefoglaltam eredményeimet és levontam következtetéseket.

2. Korlátlan és korlátozott racionalitás

Az első pénzügyi-piaci modellek nagyon egyszerűen kezelték a piaci hatékonyságot: nem foglalkoztak vele. Erre nem is volt szükség, mivel a feltevéseik implicit módon garantálták azt.

Ezen modellek a legfontosabb közös vonása a korlátlan racionalitás, amivel szorosan összefügg a modellek matematizálása. Ez folyamat valahol talán Louis Bachelier francia matematikussal kezdődött, aki 1900-ban már alkalmazta a sztochasztikus folyamat fogalmát a pénzügyi termékekre⁴, igaz, erre akkor még nem figyelt fel a szakma. Ezt a molekulák közé került apró porszemekre Robert Brown 1827-ben már megtette. Norbert Wiener a 20-as években jelentős eredményeket ért el sztochasztikus folyamatok területén, és a róla elnevezett Wiener folyamat lett a Bolyongáselemélet alappillére, mely összeköti a normális eloszlást a Centrális Határeloszlás tételén keresztül a részvényárfolyamokkal. Erre épülve megszületett a martingálelmélet, a „fair play” modell, melynek kicsúcsosodása híres a BSM⁵ modell szigorú feltételeivel és analitikus árazóképleteivel, a Nobel díjas Black-Scholes Formulával az élen. Ennek – a benne rejlő lehetőségek miatt – gyorsan Midasz formula lett a neve, gazdaggá téve kapcsolódó LTCM⁶-et és a névadó matematikusokat. Szomorú véget értő története, ha nem is a modell hibásságára, de hiányosságaira mindenképpen felhívja a figyelmet. A piaci hatékonyságot a feltevések és az árfolyammozgások markovi jellege kétszeresen is garantálja, itt ugyanis szükséges a hatékonyság, különben a képletek nem lennének alkalmazhatóak. Ezen modell elsődleges célja ugyanis nem a spekuláció és az árfolyammozgások előrejelzése, hanem – az úgynevezett delta fedezésen keresztül – termékek replikálása és új, szintetikus termékek (derivatívák) létrehozása, amiket azután, mint pénzügyi szolgáltatást, fix díjért értékesítettek. Hogy ezek értéke hogyan alakul nem releváns, a termék kiírójának szempontjából csak annak van jelentősége, hogy a fedezeti stratégia számára garantálja a kockázatmentességet.

A homo oeconomicusra épít Markowitz mikroökonómia személetű Modern Portfólióelmélete (Markowitz, 1952), mely az árfolyamok mozgásától egyel hátrébb lépve, de szintén a pénzügyi piac egyensúlyára koncentrál. A modell alapja az a mikroökonómiai alapfeltevés, hogy a befektetőknek, akik a hasznosságukat maximalizálják, a hozam jó, a kockázat pedig rossz. Markowitz eredménye, hogy egyensúlyban minden befektető a kockázatmentes termék és a piaci portfólió kombinációját tartja (Mutual Fund Theorem), az arányokat pedig a

⁴ The Theory of Speculation, 1900

⁵ Black-Scholes-Merton modell.

⁶ Long Term Capital Management. A történetéről kiváló olvasmány Nicolas Dunbar Talált pénz című műve.

befektető kockázatkerülése határozza meg. A hatékonyság itt többek között a modell információs szerkezetéből és a szereplők racionalitására vonatkozó feltételeiből következik.

Fama a Hatékony piacok elméletében már részben feloldja a hatékonyságot és három szintben formalizálja azt: gyenge, közepes és erős (Fama, 1964). Fama továbbá egységes keretbe is foglalta a piacok hatékony működésének feltételeit, erős esetben a kritériumok szinte azonosak a Modern Portfólióelméletével, de még gyenge esetben is hatékony a piac abban az értelemben, hogy szisztematikus többletthozamra a múltbeli nyilvános információk ismeretével nem lehet szert tenni. Fama a lazítások ellenére azonban azt továbbra is feltételezi, hogy a befektetők korlátlanul racionálisak, azaz azt, hogy az információkra mindig megfelelően reagál a piac, azonnal és hibátlanul árazzák be ezeket. Az esetlegesen hibás, lassú reakciók nem merülnek fel.

A faktormodellek a túllépnek a teljesen véletlenszerű bolyongáselméleten, mert felteszik, hogy az árfolyamokat valamilyen külső tényezők (faktorok) befolyásolják. Ez bár még nem mond ellent a hatékonyságnak (például a Fama-féle gyenge esetben), mert nem állítják, hogy a faktorok a jövőbeli árfolyamokra hatnak, de sejthetően itt már cél lehet az előrejelzés is, a piac megverése. Az APT (Arbitrage Pricing Theory (Ross, 1976)) és a CAPM (Capital Asset Pricing Model) a talán legszélesebb körben ismert faktormodellek. Számtalan konkrét formájuk elképzelhető, a faktorokra ugyanis különösebb megkötés nincs, azonban ezek a faktorok tipikusan nem szubjektív, emberi faktorok, hanem jól számszerűsíthető statisztikák, vagy mutatók. Az irracionalitás mértéke (mint faktor) felmerülhetne, de ez eddig nem volt tipikus.

Összességében a hagyományos modellek a statikusságban vagy egyensúlyban való gondolkodás, a többnyire lineáris kapcsolatrendszer, az információk azonnali beépülése és azok azonnali és tökéletes feldolgozása, végül pedig a szereplők korlátlan racionalitásának feltételezése miatt gyakran nem képesek magyarázni a piaci mozgásokat, úgynevezett anomáliákat hagyva maguk mögött. Emiatt szükséges a korlátlan racionalitás feloldása és az emberi (szubjektív) tényező bevezetése, amit a hírekből kinyerhető hangulattal szándékozók megragadni.

A pénzügyi viselkedéstan tudomány (behavioral finance) a közgazdaságtan és a pszichológia kombinációja. Modelljei túllépnek a racionális befektető premisszáján: korlátozzák azt, de megengedik a tanulást vagy az evolúció jelenségét. Az közgazdasági gondolkodás első ez irányú kiterjesztése Herbert A. Simonhoz köthető, aki bevezette *korlátolt racionalitás* (bounded rationality) fogalmát. Ezen irányzat legjelentősebb most csak említés szintjén

Kahneman és Tversky Kilitáselmélete, Soros György Reflexivitáselmélete, a túlzott önbizalom vagy a megbánáselmélet.

2.1. Kritikák és anomáliák⁷ – az elméletek hiányossága

A rigorózan matematikai, korlátlanul racionális szemlélet nagyon szép, intuitív, és általánosan alkalmazható eredményeket szállít, de egy konkrét helyzetre általában javítható. A közgazdasági iskolák és a ciklusok elméletei ehhez adják hozzá a saját magyarázataikat, míg a pszichológiai modellek és a pénzügyi magatartás figyelembevételével tovább finomítja az elemzést az emberi tényezőkön keresztül. Még ezek után is felmerül néhány szabálytalanság és érdekesség, melyekre az adott magyarázat hiányos vagy nincs is, így a piaci hangulat és a hírek jellege azok a tényezők, melyektől további segítséget várok. Ebben a fejezetben tehát először bemutatom a problémás stilizált tényeket vagy anomáliákat, de megfogalmazok néhány kritikát is a kockázati mértékre, a likviditásra és a befektetési teljesítmény mérésére vonatkozóan.

2.1.1. Stilizált tények, anomáliák⁸

Fontos megemlítenem, hogy a stilizált tények, anomáliák esetében nem lehet egyértelműen eldönteni, hogy ezek inkább a racionális és következetes – de összetett – piaci mechanizmusok eredményei, vagy az irracionális emberi magatartásai, esetleg mindkettő együtt. Néhányra közülük viszonylag jó minőségű magyarázat adható az eddigi elméletekkel, de ha ezt nem mindig érezzük, az nem a véletlen műve.

Időszakokhoz köthető jelenségek

Januárhatás: Főleg a kisvállalatok körében volt megfigyelhető, hogy papírjaik kimagaslóan magas hozamokat produkáltak ebben hónapban. Régebben jelentősebb volt, mára a piac alkalmazkodott hozzá, "learbitrált". Legtöbbször adózási megfontolásokkal magyarázzák: az év végi eladásokkal realizálódó veszteséggel ugyanis csökkenthetik a tőkejövedelem adóalapját. Ezt a felszabadult pénzt aztán januárban fektetik be újra, felhajtva az árfolyamot. Ami azonban arra utal, hogy nem csupán ez lehet a háttérben, az az, hogy ott is kimutatták, ahol az adózási és a naptári év nem esik egybe (Barberis, 2001). Hasonló, naptári időszakhoz

⁷ Az anomália szó jelentése rendellenesség, szabálytalanság, angolul a *noise* kifejezés is használatos helyette a pénzügyi zsargonban. A stilizált tény pedig egy empirikus megfigyelés leegyszerűsített, általános megfogalmazása. Nem minden itt felsorolt jelenség szigorúan anomália vagy stilizált tény: de azt mondhatjuk, hogy „olyan jelenségek, amik arra utalnak, hogy a piacok nem tökéletesen hatékonyak, vagy nem abban az értelemben, ahogy a hatékonyságok értjük”.

⁸ A fejezet megírásakor (Molnár, 2006)-ra is támaszkodtam. Ennek 51. oldalán az anomáliáknak egy hasonló bemutatása található további példákkal, többek között a P/E hatás, a Kisvállalat hatás és a Value Line Investment Survey talánya.

köthető megfigyelések a péntek-hatás, a május-hatás, a Mikulás-rali vagy a Mark-Twain hatás (október), de elfogadásuk erős fenntartásokkal ajánlott.

Eseményekhez kapcsolódó jelenségek

Bejelentési hatás (announcement vagy signal effect): Olyan bejelentések előtt, melyek időpontja előre ismert (statisztikák, központi banki, gazdaságpolitikai döntések, osztalékpolitikát érintő vagy más, egy vállalat szempontjából fontos kérdések) a piac viselkedése gyakran nehezen magyarázható. A dolgozatban később az 3. fejezetben foglalkozom még a kérdéssel. Külön vizsgálatot igényelnek a bejelentés előtti és a bejelentés utáni időszakok történései.

További hatások: Lásd például *merger-hatás* (más néven fúziós-hatás, fúziós-arbitrázs), *visszavásárlási-hatás* etc. Nagyon sok minden elképzelhető ebben a kategóriában: ide sorolhatnánk minden olyan eseményt, amiről az információk nem tökéletesek, a várakozások nem homogének. Itt nem adott a helyzet egy jól átgondolt racionális döntés és árazás meghozatalára, ezért statisztikailag szisztematikus félrearázásokat figyelnek meg, amikre számtalan stratégia építhető (volt).

A győzelem átka (the winner's curse): (akár az előző kategórián belül is lehetne) Egy aukciós licit győztesére utal. Az ilyen eseményeken ugyanis az összes résztvevő az eszköz valós értékét próbálja meghatározni hiányos információk alapján, és feltéve, hogy olyan eszközről van szó, amely mindegyik szereplőnek ugyanannyit ér, a győztesre az eszközzel együtt az az átok is rászáll, hogy ő volt az, aki a legjobban túlbecsülte annak az értékét. Ez tehát valójában rossz üzletet, veszteséget sugall számára. IPO⁹-k kapcsán gyakran felmerülő jelenség, hogy az indokolhatónál magasabb ár születik, mely az ezt követő másodlagos kereskedésben erősen visszaesik. Abban, hogy a szereplők a valós ár fölé mennek, közrejátszhat, hogy az erős licitálással próbálják a többi szereplőt elrettenteni, illetve ezzel párhuzamosan a félelem egy jó üzletről való lemaradástól. (Thaler, 1988) tanulmánya szerint az irracionális ármegállapítás magyarázó tényezői ezeken az eseményeken a résztvevők száma és agresszivitása – az információk hiányosságán túlmenően.

Statisztikai – matematikai jelenségek

Túlzott volatilitás: Ez alatt azt értjük, hogy a fundamentumok, pénzügyi mutatószámok változását alapul vevő matematikai modellek eredményeihez képest túlzó a volatilitás

⁹ Initial Public Offering azaz első nyilvános kibocsátás, például egy újonnan létrehozott részvénytársaság papírjai esetében.

mértéke. Egy klasszikus magyarázat a túlzott piaci reakció, a *túlreagálás*¹⁰. Ez maga is egy következmény, és részben a **Hiba! A hivatkozási forrás nem található.** fejezetben bemutatott irracionalitások következménye, csakúgy, mint az emellett párhuzamosan megfigyelhető *alulreagálás*.

A „volatilitás tömörülése” (volatility clustering)¹¹: A nagy árfolyamváltozással járó napok nem egyenletesen oszlanak el, hanem időszakosan, tömbökben. Ez azt jelenti, hogy egy nagy hozamú nap után (akár pozitív vagy negatív) ceteris paribus valószínűbb, hogy holnap is ilyen nap lesz, azonban az elmozdulás előjelére vonatkozóan nem rendelkezünk extra információval. Precízebben fogalmazva autokorreláció mérhető a hozamokra, de csak ezek abszolút értékeire (vagy négyzeteire)¹². Ezért az ilyen tulajdonságú idősorokhoz – mint a hozamok is – az egyik leginkább elterjedt modellcsalád a GARCH¹³.

Vastag szélek (heavy tails): Bár a hozamok alakulása jól közelíthető normális eloszlású változókkal, a valós hozamok eloszlása a széleken jóval vastagabb, azaz a kiemelkedően nagy elmozdulások (akár negatív akár pozitív) előfordulásának valószínűsége magasabb annál, mint ami a normális eloszlásból következne. A jelenség helytől és időtől független, nagyon általános a pénzügyi eszközök esetében. Ennek a kis eltérésnek súlyos következményei vannak a kockázatkezelésben, érdemes figyelembe venni korrekcióval vagy vastagabb szélű eloszlások például GPD (Generalized Pareto Distribution) alkalmazásával. Két további különbség a haranggörbével szemben, hogy a historikus eloszlás csúcsosabb és nem is teljesen szimmetrikus (lásd a nyereség-veszteség aszimmetriát a következő szakaszban).

Volatilitás aszimmetria vagy nyereség-veszteség aszimmetria: Recesszió esetén megnő a volatilitás, vagy ami ezzel erősen összecseng, hogy a negatív és a pozitív hozamok nem szimmetrikusak, a veszteségoldalra erősebb ingadozások és vastagabb szélek figyelhetők meg. Ez a devizapiacok esetén (zéró-összegű játék) nem áll fenn, inkább igaz a részvény- és más eszközpiacoknál. Az egyik hagyományos magyarázat a *tőkeáttételi hatás*, miszerint az árfolyamesés következtében romlik a saját tőke / adósság arány, és a saját tőke kockázatosabbá válása további nyomást gyakorol az árfolyamra, azaz a tőketulajdonosokra. (Dzielinski, et al., 2011) ezzel szemben inkább a kis (egyéni) befektetők túlzott reakciójához

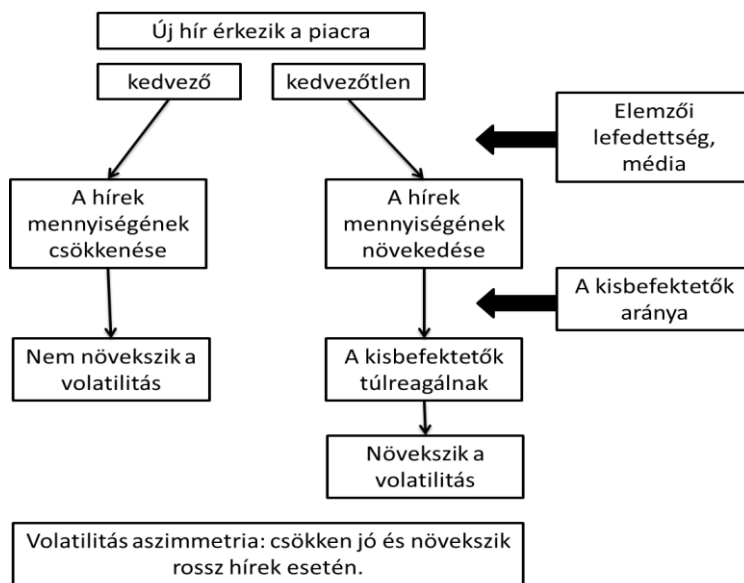
¹⁰ Az egyik legismertebb, mára már klasszikussá vált dolgozat a témában az 1985-ös "Does the Market Overreact?" Werner De Bondt és Richard Thaler tollából, ami a Journal of Finance-ban jelent meg.

¹¹ Egy részletes bemutatás például itt (Cont, 2005): *Volatility Clustering in Financial Markets: Empirical Facts and Agent Based Models*.

¹² A jelenség miatt a különböző megfigyelési gyakorisággal számított (például) havi hozamvolatilitások eltérnek. Lásd (Shah, 2008). Ugyanitt ennek kezelésére egy elsőrendű autoregresszív folyamatot javasol, módszerével azonos eredményeket kapunk megfigyelési frekvenciától függetlenül.

¹³ Generalized autoregressive conditional heteroskedasticity (Bollerslev, 1986)

társítja a jelenséget, melyben a hírek és a média is szerepet játszik, lásd 1. ábra. Tanulmányukban ugyanis megmutatták, hogy a jelenség azokon a piacokon erősebb, ahol magasabb a kisbefektetők aránya. További faktor a piaci elemzők száma, akik a hírek alapjául szolgáló jelentéseket készítik, amik alapján aztán a befektetők tipikusan kereskednek. Hozzá kell azonban tenni azt is, hogy már a hírek szintjén is torzítás van: egyrészt több a negatív hír, másrészt minél több¹⁴ a hír, annál nagyobb a torzulás mértéke a negatív irányba. Összességében tehát a hatás a fejlett pénzügyi piacokon a legerősebb (USA, UK és Japán).



1. ábra A hírek és a volatilitás aszimmetria kapcsolata, forrás: (Dzielinski, et al., 2011) 269.olda, saját fordítás

További irracionális jelenségek

A részvényprémium rejtélye (Equity Premium Puzzle): Talán a legismertebb jelenség mind közül, hogy az elmúlt 70 évben a részvényhozamok átlagosan évente 6%-kal magasabb hozamot produkáltak az államkötvényeknél (10% szemben a 3-4%-kal, USA) és egy ekkora prémium a racionális modellekkel nehezen indokolható. A pénzügyi viselkedésen a rövidlátó veszteségkerüléssel, rövidlátással magyarázza, hogy miért alakult ki ekkora különbség: a részvények rövid távú volatilitása túl nagy figyelmet kap ugyanis.

A zártvégű alapok rejtélye (Close-End Fund Puzzle): Lásd (Lee, et al., 1991). Ezen alapok befektetési jegyeinek értéke esetenként eltér a nettó eszközértéküktől, azaz attól, amennyiből a tőzsdén össze lehetne vásárolni az alap portfólióját. Az induláskori tipikusan prémium 120 nap alatt diszkonttá alakul át, ezt követően az eltérés iránya változó. Az adókkal, ügynöki költségekkel és illikvid eszközökkel dolgozó magyarázatok hibásak és hiányosak, az irracionális, rövid távon kereskedő, úgynevezett zajkereskedőké viszont kielégítőbb. E szerint

¹⁴ Egy adott vállalattal adott időszak alatt foglalkozó hírek száma.

ennek a rétegnek a nagyobb jelenléte a zártvégű befektetési alapok jegyeinek piacán, mint a mögöttes papírokéin egy olyan szisztematikus kockázatot jelent, ami miatt diszkont alakul ki, a rövid táv miatt ugyanis kereskedésüket a hangulat jobban befolyásolja. Ez jól magyarázza, hogy a diszkont a hangulatváltozásokkal, az alapok teljesítményével korrelál. Nagy optimizmus és "bikapiac" esetén kicsi a diszkont, és az új alapok indulása is az ilyen időszakokra esik. A diszkont korrelációját és az eltűnését a nyíltvégűvé alakuláskor is magyarázza, hiszen ekkor eltűnik az a többletkockázat, amit a zajkereskedők és a zártvégűség együttesen eredményeznek¹⁵.

Mániák és pánikok: Három klasszikus példa a Tulipánhagyma őrület, a Déltengeri buborék és a Mississippi buborék szoktak lenni. Ezek sem szigorúan vett anomáliák, de ha martingálokkal akarnánk modellezni ezeknek az időszakoknak az árfolyam alakulását, akkor igen sok futtatást kellene végeznünk, amíg egy hasonló trajektória adódna. A tulipánhagyma (Hollandia, XVII század) és a Dél-tengeri társaság (Nagy Britannia, XVII század) értéke igen rövid idő alatt sokszorosára - a tulipánhagymáé majdnem 200 szorosára - növekedett, majd még gyorsabban a töredékére esett vissza, rengeteg piaci szereplőt tönkretéve. Egy kiváló gyakorlati kereskedési megközelítésű könyv Zsiday Viktor: *Mániák és Válságok a tőzsde hőskorában* (2010), ahol ezeken túl kevésbé közismert mániákat is megismerhetünk.

¹⁵ Hiszen ekkor az alap bármikor eszközértéken visszavásárolja a jegyet.

3. A webes tartalmak: hír és szentiment

A hír¹⁶, a szentiment és az árfolyam nem állítható egyértelmű logikai sorrendbe, a későbbiek során látni fogjuk, hogy kölcsönösen hatnak egymásra. Például, ha esik az árfolyam, attól is elromlik a hangulat. Ennek a kapcsolatrendszernek most azt a részét szeretném bemutatni, ahogyan a hírek – és leginkább az új hírek – hatnak a szentimentre és ezen keresztül az árfolyamokra. Ebben a részben bemutatom a szentimentet, különbséget teszek a szentiment általam elsődleges forrásának nevezett hírek és a másodlagos források között. Ezután bemutatom a szakirodalom – reményeim szerint – közel teljes körű eredményeit az automatizált hír- és különösen a szentimentelemzés pénzügyi alkalmazásaiban. Végül kitérek a szentimentelemzés technikai oldalára.

3.1.A szentiment

A *szentiment* vagy *szentiment pontszám* a dolgozatban egyetlen hír hangulatának a mérőszáma vagy kontextustól függően az általános piaci hangulat. A dolgozat szempontjából azonban igazán a szentimentekből célirányosan képzett mutatószám, a szentimentindex a lényeges, ez jelen esetben tehát bizonyos hírek valamilyen gépi módszerrel kinyert hangulata, melyből azután bizonyos szempontok szerint egy mutatót, *szentimentindexet* képzünk. Az értékét adott időpontra simítás céljából mozgóátlag jelleggel számítjuk. Igény szerint képezhetünk *cégspecifius szentimentindexeket*, *iparági szentimentindexeket*, vagy teljes *piaci szentimentindexet*, a különbség abban van, hogy milyen híreket veszünk figyelembe, cégspecifikus esetben csak adott vállalattal foglalkozóakat. A hírek két kulcsmutatója a *relevancia* és az *újdonságérték*. Relevancia alatt azt értem, hogy az adott hír milyen mértékben releváns a szentimentindex szempontjából. Piaci szentimentindex esetén tipikusan valamilyen kosarat nézünk, S&P 500 vagy BUX és ekkor minket azok a hírek érdekelnek csak, amik megfelelő mértékben egy, a kosárban szereplő vállalattal foglalkoznak, a többi nem. Az újdonságérték a hír újszerűségét mutatja, ugyanis lehet elavult: a hír megjelenése és a hírben feldolgozott esemény időpontja között eltérés van, ráadásul még elavultabbá teszi a hírt, hogyha más hírek már korábban beszámoltak az adott eseményről, esetleg a hír csak egy másik hírt idéz, citál. A kettő többnyire együtt jár, később még látni fogjuk, hogy ezek igen fontos tényezők, és az ebbéli szűrés nagyban javítja a szentimentre épülő stratégiák hatásfokát.

¹⁶ A legtágabb fogalom a webes tartalmak, elfogadva azt az alátámasztatlan állítást, hogy a pénzügyi világban minden megjelenik elektronikus formában, ami megjelenik nyomtatva. Dolgozatomban a hírek, webes tartalmak, piaci információk fogalmainak összemosása nem okoz problémát. Ahol kifejezetten csak egy szűkebb információforrást használtak, ott utalok erre, de az a jellemző, hogy a hatalmas nem-strukturált adathalmaz (Big Data) minél nagyobb szelete a cél, és utólagos vizsgálatok képezik utána az értéket hordozó részhalmaz leválasztását.

3.2. A hírek világa, a szentiment forrása

Nem lenne célszerű a hírekre leszűkíteni a tárgyalást. De mégis ez a szó, ami történetileg az (új) információ hordozó médiuma. Mostantól, hacsak külön nem szűkítem le, a *hír* kifejezés az összes számunkra releváns tartalmat lefedi: statisztikai bejelentések, twittek, elemzések, újságcikkek.

3.2.1. A szentiment másodlagos forrásai

Mint már a bevezetőben is említettem, a dolgozatban az *elsődleges szentiment forrásokra* koncentrálok: azokra a már kvantitatív faktorokra, amiket a hírekből nyerhetünk ki – jelen esetben gépi módszerekkel. Ezek alapvetően különböznek a hagyományos mutatóktól, melyeket a szentiment hagyományos, *másodlagos forrásainak* tekintek. Ezek a másodlagos mutatók ugyanis már kész, leképeződött véleményeket tartalmaznak. Pl. „Mi a véleménye: a következő hónapban esni vagy emelkedni fog a BUX index?” Ez a vélemény valamiknek a hatására már kialakult, emiatt tartalmazza a véleményalkotó személynek a torzítását is¹⁷. Élesen nem elkülöníthető a két forrás, mert aki a híreket írja vagy szerkeszti, rá is bizonyosan hatással vannak a másodlagos mutatók is. Azonban mégis, ha elfogadjuk, hogy a hírek világából érkező információk hatnak a pénzügyi világban tevékenykedőkre, akkor talán jobb, de mindenképpen más jellegű adathoz jutunk az elsőt vizsgálva, mintha ennek a lenyomatát a másodikon vizsgálnánk.

A másodlagos forrásoknak felsorolás szintjén¹⁸ tekinthetjük a **felméréseket** (például Bull-Bear Survey, Merill Lynch Survey, Fund Manager Survey, portfolio.hu havi alapkezelői felmérés), **bizalmi indexeket** (például kiskereskedelmi bizalmi indexek), **pozicionáltsági indikátorokat** (put-call arány, ahol érdemes megkülönböztetni indexre vonatkozót, és az egyedi részvényekre vonatkozót), **befektetési alapok mutatóit** (például készpénzarány, long és short pozícióinak aránya vagy a nettó pénzáramlás az alapokba, esetleg a bullish és bearish alapokba áramló tőke aránya), **CFTC¹⁹ kimutatásokat** (például a nettó fedezők és a nettó spekulánsok aránya), **implicit volatilitás mérőszámait** (például VIX, VXO, VXD...²⁰) és egyéb **gazdasági, gazdaságpolitikai előrejelzéseket, célszámokat** (például inflációs várakozás, GDP várakozás).

¹⁷ Lásd második fejezet, különösen például a megerősítéses torzítás.

¹⁸ Nagyban támaszkodva Zsiday Viktor Tőzsdepszichológia című előadására, mely az Egyetemen került megtartásra a 2011/2012-es tanévben.

¹⁹ Commodity Futures Trading Commission

²⁰ CBOE volatilitás indexei. A teljes lista itt: <http://www.cboe.com/micro/volatility/introduction.aspx>

3.2.2. A hírek: a szentiment elsődleges forrásai

A csoportosítás nem egzakt feladat, függ a projekt céljától, alanyától is. (Mittra & Mittra , 2011) 6. fejezetében Leinweber és Sisk az alábbi csoportosítást tette azzal a kiegészítéssel, hogy a csoportokon belül érdemes még elkülöníteni az előre várt (szabályos időközönként bejelentett) és nem várt híreket.

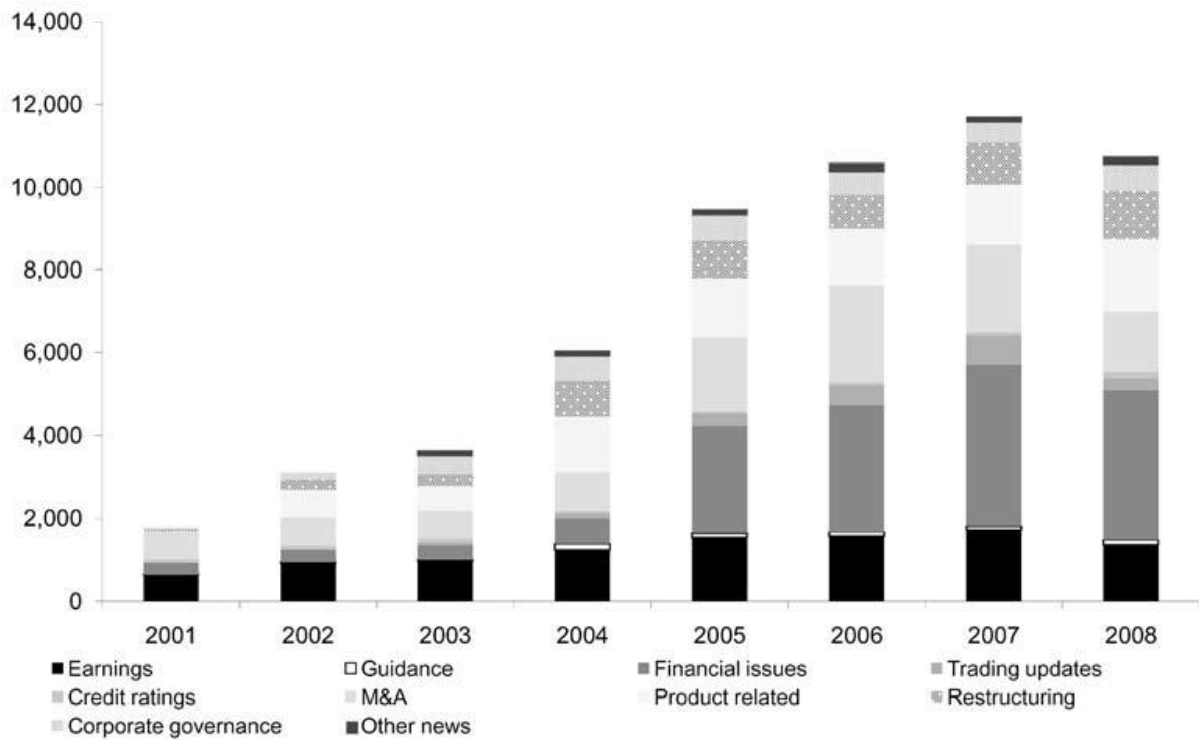
Pre-Hírek: Azok az anyagok, melyeket az újságírók a cikkeikhez felhasználnak. Ezek a legelsődlegesebb forrásokból származnak, például a tőzsdedefelügyelet adatai, bírósági felvételek, leiratok vagy pletykák, szóbeszéddek. Ez az első csoport nyersanyaga, feldolgozatlan, nincs is feltétlenül publikálásra kész állapotban, akár ellenőrizetlen is lehet.

Hírek: Régebben papír alapon, mára már többnyire elektronikus formában megjelenő, információk és közlemények. Olyan céllal születnek, hogy nyilvánosságra kerüljenek, hivatásos személyek által.

Social Media: Pletykák a Facebookon, Twitteren, Amazon értékelések, blogok, stock message bordok. Mivel nincsenek belépési korlátok, az itt található információ megbízhatósága igen alacsony.

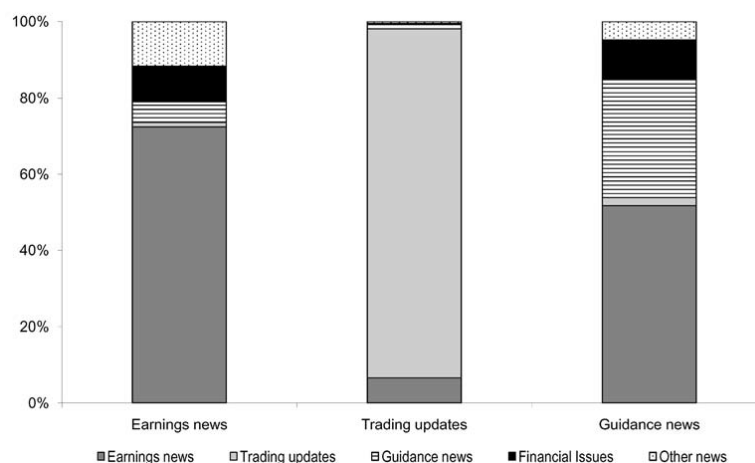
Egy másik lehetőség a "hard", azaz könyveléssel kapcsolatos és a "soft", azaz inkább stratégiai jellegű hírek megkülönböztetése. Ezt teszi (Moniz, et al., 2011) is. Ebben a tanulmányukban az S&P Large Cap Europe, körülbelül 450 vállalatáról megjelent hírek szerepelnek az általuk önkényesen megválasztott forrásokból²¹. A 2. ábra az előző folytatásaként azt mutatja, hogy a különböző kategóriákban hogyan alakult a hírek mennyisége, ami – ahogy várható volt – szinte folyamatosan nőtt az évek során. Látható, hogy a bevételekről szóló hírek mindig stabil hányadát tették ki az összesnek, de érdekes, ahogy 2005-ben milyen erőteljesen előretör a pénzügyi nehézségek kategóriája. A 2001-es évi kétezer hír nagyon kevés, és véleményem szerint még 2007-es csúcson az évi 12000 is. Ez ugyanis 250 munkanappal számolva naponta vállalatonként 0,12 hír, de ez így azért csalóka. Átlagosan ugyanis csak a vállalatok feléről van valamilyen hír minden hónapban(!), az új vállalatok esetén ez az arány magasabb (60%). Továbbá megfigyelték, hogy kevesebb a hír hétfőnként és péntekenként, illetve egyes tanulmányok arra jutottak, hogy a cégek hajlamosabbak a rossz híreket pénteken zárás után közzétenni a lankadtabb figyelem miatt.

²¹ Ezek a szöveg szerint: a legfontosabb hírszolgáltatók, tőzsdei jelentések, sajtóközlemények a vállalati weblapokról, nemzeti (USA), üzleti folyóiratok cikkei. Részletesebben nem osztja meg.



2. ábra A hírek mennyiségének növekedése az egyes hírkategóriákban, forrás: RavenPack, Factiva Factset, Macquarie Quant Research²²

Egy adott vállalatról egy napon belül többféle hír is megjelenhet és az is érdekes lehet, hogy milyen híreket jelentetnek meg együtt, egy napon belül. Az 3. ábra. ábra 3. oszlopából például kiolvasható, hogy ha egy vállalat guidance-et²³ jelent be, akkor az esetek közel felében még aznap a bevételekkel kapcsolatos bejelentést is tesz. Ez önmagában nem túl meglepő, de fontos tisztában lennünk vele, ha nagy mennyiségű hírt aggregálva, statisztikai szinten kezelünk és elemzünk.



3. ábra A hírek közötti napi átfedés hírkategóriánként, forrás: RavenPack, Factiva Factset, Macquarie Quant Research

²² Sajnos csak 2008-ig tartalmaz adatokat, de a növekedés folytatása a követő időszakban is feltehető.

²³ A management által kiadott, a valamivel kapcsolatos várakozásait tartalmazó jelentés. Ide tartozik a profit warning is, ami a profitvárakozások romlásának kötelező bejelentése.

Azt is találták, hogy a vállalat hírbeli megítélése szempontból viszonylag alacsony az autokorreláció: egy vállalat éven belül többször is változtathatja, hogy „hír-győztes” vagy „hír-vesztes”, azaz, hogy átlagosan pozitív, vagy negatív színben tűnik fel a hírekben. Továbbá, körülbelül ugyanakkora a valószínűsége, hogy győztesből vált vesztesbe, mint annak, hogy fordítva. Ez a két tényező az előzőekkel szemben nem utal manipulációra.

Azt, hogy pontosan mennyi „hírt” veszünk figyelembe, milyen forrásokat és milyen további megkötésekkel, az nagyon szubjektív, csakúgy, mint ezek végső osztályozása. Egy másik szerzőpárosnak egy másik eszközzel, egy másik részvényindexre más számok adódnak. Ők az S&P 1500 univerzumra a Thompson Reuters News Scope Sentiment Engine adatbázisát használták a 2003-2009 közötti időszakban. Havonta(!) 2003-ban körülbelül 30000 cikk, 10000 riasztás és 10000 melléklet (append) adódott, ami viszonylag monoton növekedéssel 2009-re rendre 110000-re, 60000-re és 20000-re módosult²⁴. Az 1500 vállalat lefedettsége azonban közel sem egyenletes itt sem. Azon vállalatok aránya, amelyekről az adott hónapban legalább 1 hír megjelent, a 2003-as 40%-45%-ról 2008-ra 50-55%-ra nő, majd 2009-re 60% fölé. További tanulságként levonhatjuk, hogy még manapság is vannak rejtőzködő vállalatok, vagy ami ezzel ekvivalens, a hírek többsége néhány vállalatról szó²⁵.

3.3. A piaci hírek és a piaci szentiment, mint alfa²⁶ forrás

Azt, hogy az (új) hírek hatással vannak az árfolyamokra, külön nem kívánom bemutatni, mert ezt már eddig is alapfeltevésnek tekintettem. A két érdekes kérdés az, hogy

(1) a tartalmi elemzést automatizálva képesek vagyunk-e kapcsolatot kimutatni ennek eredménye és az árfolyammozgások között, illetve, hogy

(2) a tartalmi elemzés egyik ága²⁷, a szentimentelemzés szintén releváns eredményeket szolgáltat-e ugyanerre.

Fontos tisztázni a különbséget az (1) és (2) között, annak ellenére, hogy mint már többször ebben a dolgozatban, most is fel kell hívnom a figyelmet az átfedésekre. A tartalmi elemzés azt jelenti, hogy egy strukturálatlan szövegből kinyerjük egy cég árbevételét, X dollár,

²⁴ (Leinweber & Sisk, 2011) A szolgáltató Reuterssel kapcsolatban felhívják a figyelmet az időközbeni technológiaváltásra. 2003-tól kezdődően a cégnél a Mumbai központjuk vezetésével a webes „információ kinyerés” (information retrieval) és „intelligencia felerősítés” (intelligence amplification) térhódítása jegyében „elektronikus riporterek” a kezdtek böngészni a webet gépi támogatással, majd az onnan kinyert morzsákat gyorsan hírré formázták, végül továbbították. A dinamikus növekedés ennek a gyakorlatnak is köszönhető, főleg 2006-tól kezdődően.

²⁵ Erről még később konkrétan is lesz szó.

²⁶ Jensen alfájára értendő.

²⁷ Egy másik szemlélet, ha tartalmat tekinthetjük az (inkább) objektív komponensnek, a szentimentet pedig az (inkább) szubjektívnek, azaz annak, ahogy a szöveg szerzője a tartalmat szubjektíven értékelte.

esetleg, hogy ez az X kevesebb vagy több-e, mint az előzetes várakozási konszenzus. A szentiment elemzéskor viszont arra koncentrálunk, hogy az adott szövegben hogyan értékelik az X dollárt. Leegyszerűsítve negatív, semleges vagy pozitív módon. Az egész mögött ugyanis ki nem mondva az a gondolat, hogy beállíthatják X -et magasnak vagy alacsonynak, úgyis az a döntő, hogy a piaci szereplő körében végül melyik érzet terjed el és rögzül X -szel kapcsolatban: jó vagy rossz az X ? Tartalmi elemzéskor tehát számokon, eseményeken van a hangsúly, szentiment elemzéskor pedig az író ezekre vonatkozó személyes, szubjektív értékítéletén. A dolgozatban igyekeztem elválasztani a kettőt, azonban ez nem volt teljesen lehetséges. Valószínűnek tűnik, hogy van átfedés, mert alacsony eladásokhoz, elbocsátásokhoz, elmarasztaló bírósági döntésekhez nehéz pozitívan hozzáállni, de máshol van a két esetben a fókusz.

(1)-re először Niederhoffer próbált meg válaszolni 1971-ben, melyet talán az első ilyen irányú kutatásnak is tekinthetünk. (Niederhoffer, 1971) A New York Times címlaphíreit kategóriákba sorolva azt találta, hogy a hozamok és világhírek között képes kapcsolatot kimutatni, de a konkrét kategóriák és a hozamok között nem. Ezzel szemben (Lavrenko, et al., 2000), (Koppel & Shtirberg, 2004), (Gidofalvi, 2001) és (Mittermayer, 2004) már bemutatja a mérhető nagyságú és nem véletlenszerű kapcsolatot a hírkategóriák és az árfolyam elmozdulás között, ahol a kategorizálást gépi algoritmusok végezték minden esetben. De nem hibáztathatjuk Niederhoffert, mert a *cégspecifikus*, azaz főleg egy adott vállalatra vonatkozó hírek hatásainak a vizsgálata kifejezetten nehéz feladat. Egyrészt más, általánosabb, esetleg nem is gazdasági hírek párhuzamosan nagyobb hatást fejthetnek ki, másrészt természetesen a hírek értékelése is mindig különbözhet, de még azonos értékelés sem esetén egyértelmű a válaszreakció, ugyanis például az úgynevezett kontrárius stratégiát folytató kereskedők a negatív hírekre vásárolnak, és a pozitívokra eladnak. Végző esetben, nincs is feltétlenül szükség hírekre, mert nélkülük is adható racionális magyarázat a nagy árfolyamváltozásokra, érvel Romer, rögtön két magyarázattal is szolgálva (Romer, 1993).

A tartalmi elemzés a speciális kockázatok felismerésén keresztül is támogathatja a tőkemenedzsmentet. (Cecchini, et al., 2010) kombinált, hagyományos pénzügyi mutatószámokon és automatizált szövegelemzésen alapuló rendszere 81% és 83% fölötti pontossággal jelezte előre a csalást és a csődeseményeket. (Ma, et al., 2010) pedig az egy cikken belül említett vállalatnevek alapján felépített vállalati kapcsolati háló segítségével 80% fölötti pontossággal tudták egyes vállalatok bevételeit meghatározni a többieké ismeretében – egy hasznos lehetőség arra az esetre is, ha valami miatt pontos pénzügyi adatok nem elérhetőek vagy nem bízunk bennük.

Az előre megadott időpontban megjelenő hírek vagy statisztikák beépülése az árfolyamokba milliszekundumok alatt megtörténik, de a nem várt²⁸ híreknek ennél jóval lassabban²⁹, percek, órák, hetek alatt³⁰ történik meg. Gidofalvi egy naív Bayes-i osztályozóval kategorizált pénzügyi híreket, és arra jutott, hogy a hírek beépülési sebessége és a piac alkalmazkodása, azaz a kritikus időablak a hírek megjelenése előtt és után 20 perc, pontosabban, ebben az intervallumban a benchmarknál szignifikánsan hatékonyabb előrejelzést tudott adni arra, hogy a kiválasztott részvény árfolyama leesik, fölmegegy vagy gyakorlatilag változatlan marad a hozzárendelt indexéhez (benchmark) képest (Gidofalvi, 2001). A hírt megelőző 20 percet a bennfentes kereskedésnek tulajdonították, míg az azt követőt a webes elterjedésnek és a szereplőkhöz való fokozatos eljutásnak (Gidofalvi & Elkan, 2003). Szintén percekben és órákban mérte a beépülést a napon belüli abnormális hozamokon keresztül (Hafez, 2009a). Még lassabb, napokban és hetekben mérhető beépülést talált (Cahan, et al., 2009), illetve 1 éves befektetési horizont esetében is képes volt extraprofitot generálni egy szentimenten alapuló részvényszelekciós stratégia (Kittrel, 2010). Mindenesetre önmagában az a tény, hogy az alacsony médiacoveredtséggel rendelkező vállalatok (kevés hír jelenik meg róluk) felülteljesítik a magassal rendelkezőket, arra utal, hogy a hírek információforrásnak számítanak, ezért hiányukért cserébe a befektetők extrahozamot várnak el. Erre a következtetésre jutott legalábbis (Fang & Peress, 2009)³¹, és nagyon hasonló eredményeket ért el többek között (Easley & O'Hara, 2001) a tőkeköltség illetve a zárt-körben és nyilvánosan terjesztett információk kapcsolatát elemezve.

A dolgozat első két fejezete elsődlegesen az a célt szolgálta, hogy megalapozzuk, hogy a (2) kérdésre is pozitív válasz adható. Itt azonban a félreértések miatt fel kell hívnom a figyelmet, hogy szentiment alapú kereskedés alatt hagyományosan nem a piaccal kapcsolatos hírek hangulata alapján történő, hanem a múlt *hozamai* alapján történő kereskedést ért a szakirodalom. Pozitív feedback stratégiáról beszélhetünk, amikor áremelkedést követően nyitunk vételi pozíciót, azaz a trend, a bikapiac folytatására számítunk. Ehhez hasonló, amikor a múlt nyerteseit vesszük meg (az indexéhez vagy szektorához képesti relatív teljesítményt figyelve), arra számítva, hogy ők a jövőben is felülteljesítik a piacot. Negatív feedback illetve kontrárius stratégia esetén pont ellenkezőleg, medvepiac után veszünk fel long pozíciót, illetve a múlt veszteségeit súlyozzuk felül. A dolgozatban nem a hozamok által

²⁸ Az angol kifejezések a scheduled news és az unscheduled vagy event-driven news.

²⁹ (Mitra & Mitra, 2011) 130. oldal.

³⁰ A két hír-fajta más téren is messzemenőig különbözik egymástól, a kiváltott volatilitásbeli hatásokért lásd például (Abraham & Taylor, 1993), (Ederington & Lee, 1996) vagy (Kwag, et al., 2000).

³¹ Az eredményekből kiszűrve hagyományos faktorok hatását, például a vállalatméretét, mely erősen összefügg a vállalat hírcoveredtségével.

generált múltbeli hangulatot, hanem a hírek hangulatának hatásait vizsgálom, továbbra is szem előtt tartva, hogy az előbbi is hat az utóbbira.

Hogy hogyan nyerik ki géppel a szentimentet a szövegekből azt az 3.4 fejezetben tárgyalom részletesebben, de azért, hogy egy gyors képet is kaphassunk, tekintsük a lenti képleteket. A legegyszerűbb módszer különböző elemek (tipikusan szavak, általánosan karaktorsorozatok³²) összeszámolása, melyekből az általam áttekintett alkalmazások többsége ezt vagy hasonló mutatószámokat számítja egy adott szövegre: szubjektivitás, polaritás, szentiment. Egyszerű formulákkal kifejezve, ahol p a pozitív, n a negatív N pedig az összes elemek száma, ahol az, hogy mi számít negatív vagy pozitív elemnek az határozza meg, hogy miként definiálta a modellező a két halmazt, melyik elemet hova sorolta.

$$polaritás = \frac{p - n}{p + n}$$

$$szubjektivitás = \frac{n + p}{N}$$

$$szentiment = \frac{p - n}{N}$$

(Das, 2010) azt találta, hogy a szentiment hozzáadása árfolyampredikciós modelljéhez növelte annak pontosságát, míg egy másik dolgozatában üzenőfalak postjain keresztül a kisbefektetők szentimentjét elemezve kimutatható kapcsolatot talált a hozammal, a volatiliással és a forgalommal is. (Das & Chen, 2007) Az általános piaci hangulat szerepét vizsgálta (He & Mian, 2007). Eredményeik alapján a cégspecifikus hírek és a kisbefektetők nettó vásárlása között pozitív kapcsolat van, különösen, ha az általános piaci hangulat mindeközben negatív. A hatás erősebb kicsi, fiatal, növekedési vállalatoknál, melyek volatilitása amúgy is magasabb. Ugyanez a hatás nagybefektetőkre sokkal gyengébb. Az alacsony kapitalizációjú, pénzügyi nehézségekkel küzdő, fiatal, nem nyereséges, osztalékot nem fizető, magas volatilitású vállalatokat (Baker & Wurgler, 2007) és (Leinweber & Sisk, 2011) is érzékenyebbnek találta a piaci hangulat hullámozásával szemben. Ezen a téren a kis cégek átlagosnál magasabb kisbefektetői tulajdonhányadát emeli ki a sérülékenység okaként (Tetlock, 2007). A Wall Street Journal egyik hasábján és a piac közötti kapcsolatot vizsgálva jutott arra, hogy negatív hangulat lefelé ható nyomást gyakorol a piacra, mely sokkal tartósabb kis papírok esetén, feltevése szerint az esetükben magasabb zajkereskedői hányad

³² Nem feltétlenül az eredeti formájukban és nem is feltétlenül egy szó. Lásd részletekért az 3.4.3. fejezet.

miatt³³. A kisbefektetők könnyebb befolyásolhatóságát igazolja (Barber & Odean, 2008) is és *attention-driven buying*nak nevezi a jelenséget, hogy a hírekben szereplés, extrém napi hozamok, magas forgalmú napok után ezekben a vállalatokban a kisbefektetői réteg³⁴ nettó vevő, érvelése szerint a keresési probléma miatt, mely vásárlás esetén van erősebben jelen³⁵. Az ismertebb vállalatoknak eleve magasabb a kisbefektetői tulajdoni hányaduk, hiszen ez a réteg nem (sem) szereti azokat a vállalatokat, melyeket nem ismer. Érvelnek azok, akik a kisbefektetők aránya és a vállalat hirdetési aktivitása (kölségvetése) között mutatnak ki pozitív kapcsolatot (Grullon, et al., 2004), (Dong & , 2008) és (Chemmanur & Yan, 2009). Szintén a kisbefektetők ilyen irányú fogékonyságára utal, hogy nettó vevő pozíciót vesznek fel eredményre vonatkozó hírek után – függetlenül attól, hogy a várakozások tekintetében az pozitív vagy negatív (Hirscheifer, et al., 2003), (Lee, 1992). Utóbbi azt sejtí, hogy ilyenkor a brókerházak az eredménytől függetlenül rutinszerűen körbetelefonálják ügyfeleiket, vásárlást ajánlva nekik. Erre a rétegre hasonló hatást gyakorol az is, ha az árfolyam egy bizonyos határt fölfelé átlép, különösen azok körében, akik először vásárolnak az adott papírból, fogalmazza meg (Seasholes & Wu, 2004) a Shanghai Stock Exchange-en végzett kutatásaik alapján. A napon belüli volatilitásra gyakorolt hatásra koncentrált (Ho, et al., 2013) a RavenPack News Analytics - Dow Jones Edition programját alkalmazva 65, a Dow Jones Composite Average indexben szereplő vállalatra. A legtöbb cég esetében szignifikáns kapcsolatot talált a cégspecifikus hírek szentimentjével, még azután is, hogy kiszűrték az aznapi makro adatok hatását. Mindez a negatív hírek esetén volt különösen megfigyelhető³⁶. Szintén a volatilitásra és a forgalomra gyakorolt hatás robosztusságát hangsúlyozza (Gross-Klussman & Hautsch, 2011) és (Mitra, et al., 2008), akik az implicit volatilitáson alapuló volatilitás előrejelző modelljüket javítottá hírek és szentiment, mint exogén változók felvételével (Ser-Huang & Granger, 2005) ajánlását követve. Esetükben a motiváció az volt, hogy a hagyományos modellek nem alkalmazkodnak a piaci kondíciók változásakor, de a hírek által hordozott információval javítani tudtak a hozam-kovarianciamátrix alakulásának előrejelzésén. A szentiment hatását igazolják (Kurov, 2008) eredményei is, aki index futures piacokon pozitív kapcsolatot talált a szentiment és a pozitív feedback kereskedés intenzitása között, azaz pozitívabb hangulat esetén hajlamosabbak voltak a traderek egy áremelkedés után long pozíciót nyitni. Továbbá azt találta, hogy negatív szentimentű időszakokban megnő a

³³ (Kittel, 2010) megjegyzi, hogy a Tetlock tanulmányban használt General-Inquirer nevű szövegelemző szoftware szavak kombinációját nem, csak különálló szavak előfordulását figyelte.

³⁴ Intézményi befektetők esetén nem mértek hasonló reakciót.

³⁵ Hiszen rengeteg részvény közül tud választani vásárlás esetén, míg eladni inkább csak a portfólióban már szereplők közül tud (a short sellingre megszorítások vannak és kevesebb befektető is él vele).

³⁶ Szintén a negatív hírszentimentet találta erősebbnek (Leinweber & Sisk, 2011).

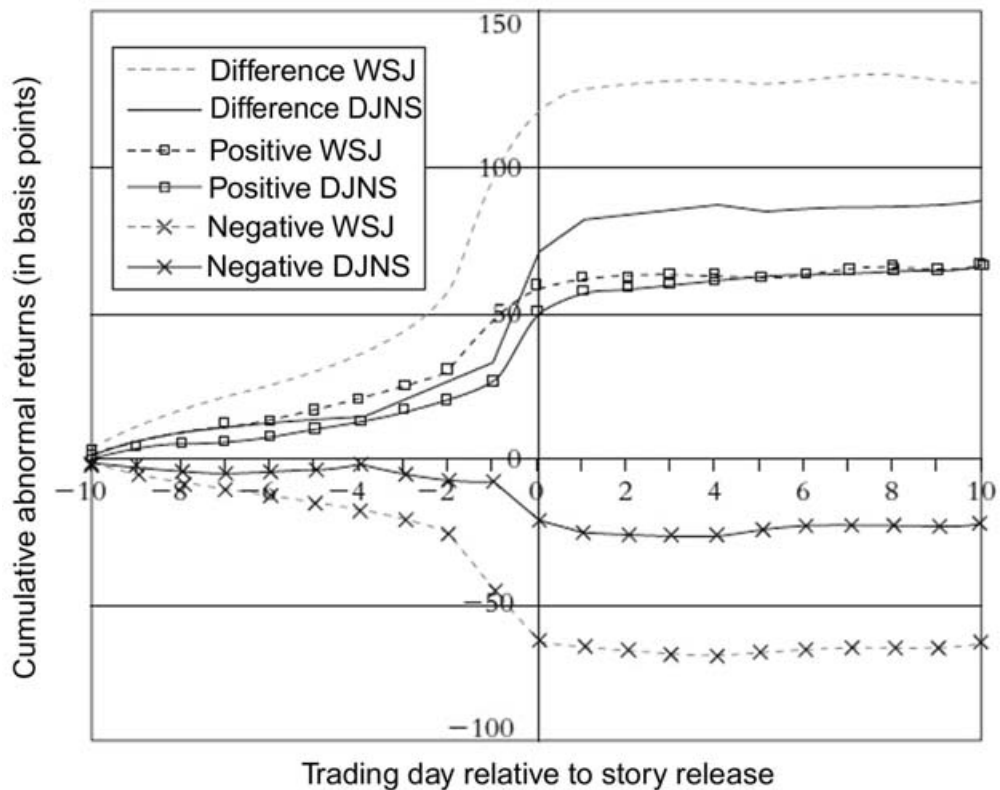
zajkereskedés, és ezzel a napon belüli volatilitás, (pozitív időszakokban kevésbé jellemző) ami kedvezően hat az árfolytonosságra és a likviditásra.

Széleskörű kutatást folytatott a témában (Zhang & Skiena, 2009). Céljuk a teljes információ-spektrum lefedése volt, melynek eredményeképpen 1000 gigabyte-nyi szöveget elemeztek a Twittertől kezdve blogokon át a Wall Street Journalig. A vizsgálat 2005-2009 között 5 évet és több ezer vállalatot ölelt fel. Rövid távú, de szignifikáns kapcsolatokat találtak, különösen a hírek polaritása és a hozamok között illetve a szubjektivitás és a forgalom között. Különbséget találtak a blogok és a szakmai folyóiratok szentimentjének között. A blogok szentimentje stabilabb, azaz lassabban változik és lassabban is épül be az árfolyamokba, 2-3 nap alatt. Az utóbbi gyorsabban változik, de gyorsabban, tipikusan még aznap be is épül, megjegyezve, hogy jobb modell építhető fel a kapcsolatra, ha a blogok, twittek és hírek információit együttesen használjuk fel.

A pénzügyi microblogok bejegyzéseit vizsgálta (Sprenger & Welpé, 2010) is, hasonló eredményeket kapva. Szintén késleltetett kapcsolatot találtak a blogbejegyzések szentimentje és a hozamok, forgalom között. Ők emellett a minőséget emelték ki a mennyiséggel szemben: fontosabb a bejegyzések szentimentje a számosságuknál.

Egy másik, szintén sokat citált és átfogónak tekinthető elemzést mutatott be (Tetlock, et al., 2008). A Dow Jones Industrial Service és a Wall Street Journal adatbázisából az S&P 500 vállalatot érintő 350 ezer hírt elemeztek szentiment szempontjából³⁷ az 1984-2004 közötti időszakra. A két forrás pozitív és negatív cikkei körüli átlagos abnormális hozamot mutatja az alábbi ábra. A középső függőleges vonal a hír megjelenése, egyébként a függőleges tengely az abnormális hozam, a vízszintes pedig az időablak a hír megjelenése körül (nap):

³⁷ Az általuk szentimentelemzésre használt software a General Inquirer volt.



4. ábra CAAR alakulások pozitív és negatív Wall Street Journal (WSJ) és Dow Jones News Service (DJNS) híresemények előtt és után az S&P 500 vállalataira 1984 és 2004 között. Forrás: (Tetlock, et al., 2008)

Az azonnal látszik, hogy van kapcsolat, ennek iránya és nagysága megfelelő, és hogy a Wall Street Journal cikkeihez nagyobb hatás kapcsolódik, de nagy probléma, hogy a CAAR³⁸ 90%-a a hírek megjelenése előtt következik be. A piacok hatékonyságának végleges elfogadása helyett Leinweber és Sisk³⁹ inkább arra hívja fel a figyelmet, hogy a hírek nagy részét nem az eredeti, adott eseményt először bemutató hírek teszik ki, hanem ezek későbbi átírtai, visszahivatkozások. Ha csak ezek a legelső, eredeti írások szereplnének a tanulmányban, akkor sokat javulna az eredmény. Ez egyébként az automatizált hírfeldolgozás egyik kritikus kérdésének tűnik, később is visszatérő probléma. Egy jó hírekre long, rosszakra short pozíció jellegű stratégia (Tetlock, et al., 2008) eredeti tanulmányában csak 9 bázispontos tranzakciós költségek mellett volt profitábilis, holott az intézményi átlagos tranzakciós költség 50+50 bázispontnak tekinthető.

³⁸ Cumulated Average Abnormal Return. Tipikusan a figyelt esemény utáni hozamok egy benchmarkhoz képesti (például Fama-French modelljéből vagy CAPM modellel becsült) eltéréseinek átlaga.

³⁹ (Mitra & Mitra, 2011) 6. fejezetének szerzői. A (Tetlock, et al., 2008) tanulmány bemutatását és interpretációját részben ez alapján tettem meg.

3.4.A szentiment kinyerésének technikai oldala

Csak az „automatizált” módszertan esik most vizsgálat alá, mely két fő lépésre bontható/bontandó, amennyiben egy *service provider*, azaz szolgáltató igénybe vétele helyett a teljes infrastruktúrát magunk akarnánk létrehozni.

Az első lépés a kívánt elsődleges források meghatározása és ezután meghatározott formában való lementése egy merevlemezre, előkészítve a terepet az elemzések lefuttatására. Erre a lépésre *indexelés*ként szokás hivatkozni.

A második lépésben⁴⁰ továbbsszűrjük és előfeldolgozzuk a lementett tartalmakat, majd megtörténik tulajdonképpeni szentimentelemzés vagy opinion-mining, melynek során kiszámításra kerülnek például a már röviden bemutatott szubjektivitás, polaritás mutatószámok is. Később ezekből valamilyen szempontok alapján képezzük az aggregált szentiment értéket vagy szentimentindexet.

3.4.1. Első lépés: a források meghatározása és a tartalmak indexelése

A piaci szentiment szinte minden esetben *nem strukturált adatokként* mai divatos nevén Big Data-ként hivatkozott⁴¹ forrásból kerül kiszámításra. Meg szoktak még különböztetni *félleg strukturált adatokat* is, de nem okoz félreértést, ha a nem strukturáltra egyszerűen a strukturált komplementereként tekintünk. Mivel egyáltalán nem tűnik kritikusnak a fogalom definiálásának pontos módja, ezért ezek olyan adatok lesznek, amik nincsenek abban a formában, hogy a bennük lévő összes releváns információt lekérdezhessük. Ilyenek tehát pl. (az emberek által írt) szövegek, képek, videók⁴². Ezek olyan sokfélék lehetnek, hogy egyértelmű kategorizálásuk, felcímkézésük, rendszerezésük során mindenképpen valamilyen kompromisszumot kell kötnünk az információvesztés terén és a lekérdezhetőség között.

Mára az információ legnagyobb hányada nem strukturált formában van, aránya az IBM és a Gartner szerint 80%-ra tehető és havonta duplázódik a mennyiség. Egy másik aspektusa a Big Data-nak, hogy 90%-a az elmúlt két évben keletkezett. Ha az adatokat a webről gyűjtik, akkor az adatok beindexelése tipikusan a webet URL-enként vagy linkenként bejáró kis robot (crawler) segítségével történik. Ezek képesek kommunikálni a webszerverekkel, majd ha megkapják a kért oldalt, akkor lementik az URL által azonosított tartalmat. Ha mi magunk

⁴⁰ Hogy miért bontjuk a folyamatot két lépésre főleg technikai okai vannak. Egyrészt a második lépést sokkal hatékonyabban tudjuk helyi adatokon futtatni, másrészt, csupán beállítás kérdése, hogy meddig tároljuk az adatokat. Elemzés után akár azonnal törölhetjük is, de mind a verifikációhoz, mind a validációhoz elengedhetetlen, ha a tartalmakat mégis megtartjuk, célszerűen az eredeti formájukkal együtt.

⁴¹ Nem ekvivalens a két fogalom, de a dolgozat céljainak nem mond ellent az együtt kezelésük.

⁴² A dolgozat keretein belül csak az írott adatok elemzésével foglalkozok, képpel, hanggal, videóval nem.

határozzuk meg az URL tartományt, akkor a frissítési gyakoriság kulcstényező, hiszen ezt ritkábbra állítva az új tartalom lassabban kerülhet lementésre. Ha megengedjük, hogy a robotjaink elvándoroljanak a kezdeti oldalakról, akkor pedig ennek a szabályait kell meghatároznunk. Ez persze nem igazán hatékony megközelítés, ha ugyanis nekünk a millisecundumok is számítanak, akkor inkább olyan megoldásban kell gondolkodnunk, ahol a kérdéses forrás magától és azonnal elküldi nekünk az új információt.

3.4.2. Második lépés: szentimentelemzés

Az összegyűjthető adatok minősége azonban megkérdőjelezhető, és ez újfajta kihívások elé állítja azokat, akik az adatokból információt szeretnének kinyerni. Ezen technikák és módszerek összefoglaló elnevezése *adatbányászat* (data mining), ha pedig kifejezetten a tartalmak szubjektív, hangulatot kifejező dimenziója érdekel minket, akkor a *szentimentelemzés* eredményeit használjuk.

A pénzügyi hírek elemzése egyrészt könnyebb, mint egy általános szövegé, mert ezek a szövegek központosítása és nyelvtani helyessége általában kiváló, szókinccse jól körülhatárolható, és maga a szöveg felépítése is céltudatos, logikus – a pénzügyi, üzleti zsargon elemi jelennek meg benne. Másrészt pont emiatt, a szentiment elemzésénél nehézségek is adódnak, mert a hangvétel egyúttal szándékosan pártatlanságra törekvő, ugyanis a független elemzők nagyon ügyelnek arra, nehogy túl határozottan állást foglaljanak, hangulatot szítsanak.

A technikai oldalával részleteiben nem kívánhatok foglalkozni, mert önmagában is nagyon szerteágazó, önálló tudományág, és nyelvenként eltérő a nyelvi sajátosságok miatt. Ráadásul mivel az egyre több adatból információ kinyerése és ennek megakadályozása folyamatos ellentétben áll és fejlődik, etikailag sem mindig egyértelműek a határok⁴³. Széles és interdiszciplináris a paletta, a statisztika, a nyelvészet és a valószínűség-számításon alapuló számtalan megoldással találkozunk. Nincs minden helyzetre legjobb módszer, de a trend, úgy látszik, hogy az algoritmusok direkt betanításától a tanulási képességekkel való felruházás felé mutat. Ekkor ismert tulajdonságú adatokon saját magukat tanítják, optimalizálják az algoritmusok. Ez a szentimentre lefordítva azért jó, mert ekkor nincs szükség a szabályok pontos megfogalmazására, hogy pl. mikor is negatív szentimentű egy szöveg. Elégséges, ha létrehozunk egy sokváltozós, sokparaméteres függvényt (az úgynevezett *osztályozót*), ezután elé teszünk minél több, általunk ismertén negatív vagy pozitív szöveget – ez lesz az

⁴³ Adatbányászat és adathalászat kissé hasonló kapcsolatban állnak, mint az adóptimalizálás és az adókerülés. Etikus dolognak számít-e például egy online póker oldalon egy ellenfelem minden lépését néhány éven keresztül ledokumentálni, majd amikor nagy tételben egymással szemben állunk, a statisztikákat felhasználni ellene?

úgynevezett *tanítóhalmaz* –, és azt mondjuk neki, hogy határozza meg úgy a változókat és paramétereiket, hogy a negatív szövegekhez hozzárendelt számok (skalárok vagy vektorok) a lehető legtávolabb essenek a pozitív szövegekhez rendeltéktől, azaz minél jobban szétválassza a két szövegtípust (*optimalizálás* vagy *tanítás*). Ha elég jók vagyunk az algoritmus, a tanítóhalmaz és a tanulási paraméterek kiválasztásában, akkor nem is kell beszélnünk az adott nyelven, az algoritmusunk képes lesz a szentimentet meghatározni. A trükk az, hogy azok a szabályok, amiket esetleg nem is ismerünk vagy képtelenek vagyunk megfogalmazni, már implicit módon benne lehetnek a tanítóhalmazban. Ennek következtében legalábbis egy pontosan ismert tulajdonságú tanítóhalmazzal mindig rendelkezünk kell.

3.4.3. Néhány alpmódszer és további szempontok

3.4.3.1. Szózsák és társai, osztályozás, tanítás

A természetes nyelvek feldolgozásának (natural language processing) egy standard alapsémája a *szózsák* modell (bag of words). Az elemezni kívánt szöveg szavai közül kiszűrjük a *stopwordöket* (tipikusan önálló jelentéssel nem bíró szavak, névelők, határozószók), kiírjuk a rövidítéseket, levágjuk a ragokat, toldalékokat (stemming). Ezután a szavak, például szófajokként, külön halmazokba kerülnek, ahol a sorrend már nem számít, csak az előfordulás száma, innen a zsák kifejezés. A stopwordök szűrése és a szavak megfelelő zsákokba kategorizálása is úgynevezett *szótárak* segítségével történik, amik valamelyest valóban hasonlítanak a klasszikus szótárakra, valójában szavak és a hozzájuk tartozó utasítások a számítógép számára.

Most, hogy már rendelkezünk a szózsákokkal, meg kell határoznunk, hogy miről is szól az adott szöveg, megtehetjük ezt például a szavak valamilyen fontosság alapú súlyozásával. Ez megtörténhet a fent már említett módon, hasonló és ismert témájú dokumentumok betáplálásával, és valamilyen metrikát rendelve ezekhez és a vizsgált dokumentum szózsákjaihoz. Végül meg kell nézni, hogy melyikhez esik legközelebb. Teljes szózsákok összehasonlítása azonban kevésbé hatékony, egy gyorsabb megközelítés a fontos szavak megkeresése. Megfigyelték, hogy azok a szavak a legfontosabbak statisztikai alapon legalábbis, melyek az adott szövegben sokszor, de más szövegekben kevésszer fordulnak elő. A szózsákhoz képest más megközelítés, hogy a szöveget 2 vagy 3 tagból álló szó és/vagy karakterláncokra bontjuk, úgynevezett *n-gramokra*. Ekkor nem veszik el a szavak sorrendjében rejlő információ teljes egésze. Ez után a hasonló dokumentumok megkeresése hasonlóan történhet. A módszer egyik legfontosabb alkalmazása a beszélt nyelv felismerése, ugyanis megfigyelték, hogy a különböző nyelvekben mások a tipikus hangsorozatok. Úgy

gondolom, hogy a módszer hatékonysága a szótárakon áll vagy bukik, ezek megalkotása rengeteg munka. Szótárakat használunk a stopwords szűrésénél, a stemmingnél, majd a szavak vagy n-gramok kategorizálásához is. A szótárak tehát a folyamat alapkövei, létrehozásukhoz magas szintű nyelvi ismeret szükséges, és a legjobb eredmény elérése érdekében azt is figyelembe kell vennünk, hogy milyen típusú szövegeket fogunk elemezni velük.

Sokféle szöveget sokféle szempont szerint, sok szinten lehet elemezni. Nehéz megítélni a pontosságát ezeknek a módszereknek, de az áttekintett források alapján összességében úgy tűnik, hogy (irodalmi) angol nyelven, (géppel) írott szövegeket – és már a rövid twitteket is - tartalmi szempontból nagy pontossággal képesek vagyunk kiértékelni. (Pang, et al., 2002) megjegyzi, hogy kifejezetten szentimentelemzésre sok tipikus tanítási módszer nem működik olyan hatékonyan, mint a tartalmi elemzésnél. A pontosság növelése érdekében bevett szokás a kétséges szövegek (alacsony megbízhatósági szinten kategorizált) kihagyása vagy a különböző módszerek együttes használata és „szavaztatása”, aszerint, hogy hova sorolnák az adott szöveget a legtöbben vagy a legnagyobb megbízhatósági szinten (Cahan, et al., 2009). Szemléltetésképp a RavenPack News Analytics 5 osztályozó módszert használ, melyek között szerepel a szószákhöz hasonló szavakat és kifejezéseket kereső és elemző, illetve olyan, ami szakértők által összeállított tanítóhalmazok alapján kategorizálja a cikkeket szentiment szerint.

3.4.3.2. Néhány további szempont

Elemi kérdés, hogy a vizsgált tartalom hány szereplőhöz jut el, és hogy ezek a szereplők mennyire relevánsak a piaci folyamatok alakulása szempontjából. Ezen a téren rengeteg kompromisszumot kell kötnünk, mert egy oldal látogatottsága ugyan viszonylag könnyen mérhető és hozzáférhető, de az, hogy ezeket kik olvassák, és hogy közvetve kikhez jut el, legális eszközökkel és hatékonyan nem követhető.

Fontos másrészt, hogy az adott tartalom mekkora hatással van az olvasóra. Nem hibás hozzáállás kezdetben feltenni, hogy ez alapvetően összefügg az oldal olvasottságával, hiszen miért olvasnánk olyat, amit nem veszünk komolyan, de azért erre is van példa. (Kucuntunc & et al., 2012) azt találta, hogy pénzügyi kérdések esetében a Yahoo Answers-en a legjobbnak ítélt válasz szentimentje inkább semleges, mint a más területeken feltett kérdések esetében⁴⁴, ami abba az irányba mozdíthat, hogy a szentimentindex megalkotásakor a semleges

⁴⁴ Ez némileg ellentétes a már korábban hivatkozott (Sprenger & Welp, 2010) eredményével, de nem mond teljesen ellent neki.

válaszokat inkább fölül súlyozzuk. Az eredmény talán nem meglepő abból a szempontból, hogy a valóban hozzáértő, szakmai szereplők a tárgyyszerű, száraz stílust preferálják, döntéseiket kevésbé az érzelem teljes, elfogult írásokra alapozzák, mint esetleg az egyedi befektetők.

A cikkeket éjjel nappal érdemes figyelni, mert mint (French & Roll, 1986) megmutatták a tőzsdék zárva tartásakor az „árnyék volatilitás” jelenléte igazolható és mértéke a tényleges 1/7-ed részére tehető. Az OTC⁴⁵ kereskedés tehát soha nem áll meg, a papírok jelentős részével (vagy rá szóló letéti jeggyel, derivatívákkal) a Föld valamelyik pontján mindig kereskednek, és a manapság már szinte azonnal záródó különbözetek (arbitrázslehetőség) miatt az árfolyam alakulása megfelelő technológiai felkészültséggel folytonosnak tekinthető.

A szóhasználat és szintaxis alapján lehetőség nyílik a szerző beazonosítására, ha az ismeretlen, de rendelkezünk tőle más munkákkal, mint azt (Abbasi & Chen, 2008) is bemutatta. Ez egy érdekes továbbfejlesztési lehetőség, amennyiben kifejezetten nagy befolyással rendelkező szerzőket sikerül beazonosítanunk. A dolgozat keretein belül ezt azonban egyelőre figyelmen kívül hagyjuk.

⁴⁵ Over the counter ügylet, magyar elnevezés jobb híján a Fix ügylet. Ebben az ügyletben az ár és a mennyiség (bizonyos megkötések mellett) a felek megegyezése alapján történik, nem feltétlenül az aktuális piaci áron.

4. Piaci tapasztalatok és néhány konkrét szentimentstratégia

A szentiment – és szövegek más tulajdonságainak – elemzése egyre népszerűbb, de pont az elterjedés miatt az olyan egyszerű stratégiák, mint a „vétel jó hírekre” és „eladás rossz hírekre” nem generálnak jelentős alfat. ... Annak ellenére, hogy a híreken alapuló kereskedési jelek piaci befogadottsága korai szakaszban van, azt tisztán lehet látni, hogy ennek segítségével az algoritmizált stratégiák milyen sokat javíthatóak. ... Csak idő kérdése, hogy ezek a technikák elég mainstreammé váljanak ahhoz, hogy a „könnyű alfa eltűnjön”.⁴⁶

Piaci szentiment és pszichológiai megfontolások alapján kereskedett 2008 és 2010 között a MarketPsych Capital kicsit több mint két év alatt 28%-os hozamot elérve. Az alap - tisztázatlan okokból bekövetkezett - bezárása után a tulajdonos az általa használt szentimentelemző szoftverrel képzett cégspecifikus szentiment indikátorok forgalmazásába kezdett, a vetélytársai ezen a piacon többek között a Thomson Reuters, a Bloomberg és a RavenPack. A (sokak által csak Twitter-Fundként emlegetett) Derwent Absolute Return a Twitter bejegyzések alapján kereskedett, ám mindössze csak egyetlen hónapig 2011-ben. Ezalatt 1,86%-ot ért el, ami nem rossz eredmény, de ennek ellenére szintén átállt a közvetlen felhasználásról a szentiment indikátorok forgalmazására⁴⁷.

Azt, hogy *van* kapcsolat a szentiment és a hozamok között, néhány megszorítással, de bemutattuk. A legtöbb kereskedési stratégiához azonban a pusztán kapcsolat önmagában nem elég. Az a kérdés, hogy a szentiment hatása az árfolyamra elég erős, megfelelő jellegű, és robusztus-e a lekereskedéshez. Most bemutatok három stratégiát, ami mindenesetre arra utal, hogy talán igen.

4.1.1. Első stratégia: Iparág-rotáció⁴⁸

Két iparág-rotációs, illetve egy ezeket megelőző egyszerűbb stratégiának az eredményét mutatja be (Hafez, 2010). Hafez 2005 májusa és 2009 decembere közötti időszakon tesztel, a használt szentimentindexek a RavenPack News Analytics adatbázisából⁴⁹ származnak. A vizsgálat tárgya S&P 500 index és vállalatai. Mindhárom stratégia esetén közös, hogy minden hónap elején az előző hónap különböző szentimentindexeinek alakulása függvényében vesz

46 Richard Brown, a Lexalytics vezetőjének gondolatai (Mitra & Mitra, 2011) 310. oldal, saját fordítás.

47 Forrás: 'Trading on a World of Sentiment' írása a Bloomberg Businessweektől és 'Last tweet for Derwent's Absolute return' a Financial Times cikke.

48 Hasonló szemléletű stratégiák bemutatása itt: Quant 3.0 – Harnessing the Mood of the Web, QWAFEFW, 2012 december 18, a Deutsche Bank munkatársainak prezentációja. Hasonlóság mutatkozik a szentimentindex képzésében, a jel felhasználási időtávjában, illetve a stratégia jellegében (szektor rotáció).

⁴⁹ Az adatbázis nem torzít túlélés szerint, és követi a névváltozásokat, összeolvadásokat. A vállalati akciókkal kiigazított árfolyamadatokat a Commodity Systems Inc. szolgáltatja.

fel egységnyi pozíciót, és egy hónapig tartja azt. Ezek az indexek a teljes *piaci szentiment*⁵⁰ illetve 10 *iparági szentimentindex*⁵¹, minden esetben 90 napos⁵² mozgóátlagai. A középpontban nem az indexek abszolút értéke, hanem a megváltozása (deltája) szerepel. A híreket szűrte relevancia szerint, csak azokat a híreket vette figyelembe, melyek az elérhető 100 relevanciapontból legalább 90-et értek a RavenPack önkényes skáláján, de ami gyakorlatilag azt jelenti, hogy szinte kizárólagosan az adott vállalattal foglalkoznak⁵³. A vállalatok súlyai mindkét indexben azonosak.

1. stratégia: az S&P 500 indexben long pozíció felvétele, ha az előző hónap piaci szentimentjének deltája pozitív, és short, ha az negatív.

2. stratégia: a 10 iparágba sorolva az S&P 500 vállalatait képezi az iparági szentiment indexeket, majd normalizálás útján összehasonlíthatóvá teszi a deltáikat. Ez alapján rangsorolja az iparágakat, majd minden hónapra az előző hónap 5 legjobb iparágában long, 5 legrosszabbjában pedig short pozíciót vesz fel.

3. stratégia: az 5 legjobb iparágban (legjobb olyan értelemben, mint a 2. stratégiában) long pozíciót vesz fel, ha a piaci szentiment index deltája pozitív, és az 5 legrosszabbban shortot, ha negatív.

Mint ismeretes, a S&P 500 a vizsgált időszakban a következőképpen alakult 2007 augusztusáig stabil növekedést, majd 2009 februárjában mélypontra kerülő visszaesést mutat. Az ez utáni időszakban emelkedő trend figyelhető meg.

⁵⁰ A piaci szentiment index esetén a híreket szűrte újdonságérték szempontjából is, csak azokat vette figyelembe melyek egy adott eseményt először dolgoznak fel és 24 órán belül. Ez kritikus lépés volt a kereskedési hatékonyság szempontjából. Azonos, de ezt a lépést nem elhagyó stratégia feleakkora (átlagos) éves hozamot ér csak el a 2005-2009 es időszakban. Ugyanezt a szűrést az iparági szentiment indexek esetén nem teszi meg.

⁵¹ Az iparágak az ICB kategóriái alapján.

⁵² A 90 napra a simítás például a negyedéves jelentések körüli napok felfokozott történései miatt van szükség.

⁵³ A piaci szentiment index esetén csak a 100-as relevanciaértékű híreket.

		1. stratégia	2. stratégia	3.stratégia	S&P 500 tartása
Információs hányados	Teljes	1,75	1,23	1,91	0,00
	2007 október előtt	1,02	0,92	1,03	0,00
	2007 október után	2,47	1,53	2,79	0,00
Éves hozam	Teljes	27%	11%	31%	-2%
	2007 október előtt	8%	5%	9%	5%
	2007 október után	47%	16%	54%	-8%
Éves volatilitás	Teljes	15%	9%	16%	17%
	2007 október előtt	8%	6%	9%	7%
	2007 október után	19%	11%	20%	23%
Nyereséges hónapok aránya	Teljes	70%	66%	68%	61%
	2007 október előtt	62%	62%	59%	72%
	2007 október után	78%	70%	78%	48%
	2005-os hozam	13,1%	9,6%	14,8%	10,8%
	2006-os hozam	-9,2%	3,5%	-8,0%	11,6%
	2007-os hozam	23,0%	7,5%	25,1%	-4,2%
	2008-os hozam	58,3%	-2,0%	56,6%	-51,2%
	2009-os hozam	42,9%	33,7%	60,5%	26,3%

1. táblázat: Iparág-rotációs stratégiák alapadatai

Mindhárom stratégia felülteljesíti az index passzív tartását, és külön kiemelendő, hogy a 2007 októberi csúc előtt és után is, azaz emelkedő és eső piac esetén is. Az egyetlen kakukktojás év a 2006-os. Ekkor igen rosszul működtek a szentiment stratégiák, ugyanis amíg az index 11,6%-ot erősödött, addig kettő majdnem 10%-ot veszített és a harmadik is csak 3,5%-ot tudott hozni⁵⁴. Ezt a rossz évet is figyelembe véve a teljes időszakban a három stratégia átlagosan évente több, mint 23%-kal hozott többet mint a piac, és a legjobb stratégia pedig évente több, mint 31%-kal. Az éves hozamokat egyenként nézve 2005-ben nem teljesítettek sokkal jobban, 2006-ban megzavarodtak, 2007-ben pedig stabil eredményt produkáltak és annak ellenére jól hoztak, hogy a piac eléri és átlendül a csúcspontján. 2008-ban a piac végig lejtmenetben volt, és a 2. stratégiát leszámítva a másik kettő abszolút értékben is többet hozott, mint amennyit az S&P veszített. Különösen igaz ez a harmadik évre, amikor is megindul az emelkedés, de míg a piac csak 26%-ot emelkedik, a három stratégia átlagosan 45,7%-ot, a legjobb pedig több mint 60%-ot. Itt tehát kifejezetten erősnek tűnik a szentiment szelekciós képessége. Összességében a többi mutató alapján is kedvező képet kapunk. A stratégiák egyszerűsége és erősen intuitív volta pozitívum. A hasonló megközelítést alkalmazó – az előző hónap ’’hozam – nyerteseit’’ longoló, veszteseit shortoló (momentum) stratégia mindössze éves 6,8%-ot hozott volna – szemben a 27%, 11% és 31%-kal –, ami azt

⁵⁴ Ez nem szabad, hogy a kedvünket szegje, nem várható el, hogy ilyen egyszerűen felfedezzük a piacok Szent Grálját. (Romer, 1993) is azon az állásponton van, hogy a piacok sok rétege, a szereplők heterogenitása és az ezek között interferencia (interakció) következtében nincs egyszerű, mechanikus kapcsolat a hírek és a hozamok között.

mutatja, hogy ha hasonló jellegű döntéseket, de a múltbeli szentiment értékek alapján hozunk meg, és nem az árfolyam alapján, akkor lényegesen jobb hatásokkal tudunk kereskedni.

Kritikák és észrevételek

A szentiment indexek számításának pontos képletét nem közöltem itt, mert úgy vélem, hogy egy átlagolási folyamat nem olyan izgalmas. Pontosabban mondva nem szabad, hogy bonyolult legyen, ugyanis itt már nem (sem) szabad olyan műveleteket végezni, amik esetleg érdemben torzítják a végeredményt. Részletesebb bontású hozam táblázatot nem közölt a szerző, de a hivatkozott cikkben található kumulált hozam ábrák viszonylag monoton növekedő jellege legalábbis biztató. Az említett normalizálási folyamat függvényeinek a leírása véleményem szerint pontatlan, de a módszer szimmetrikus, és nem tűnik torzítónak annak ellenére, hogy a normalizálás mellett fel is erősíti a jelet azáltal, hogy az átlagtól távoli értékeket tovább növeli. A szerző nem portfólió szimulációt végez, de nem tér ki a tranzakciós költségekre, ezért feltehetjük, hogy nem számol velük, illetve, hogy hozamokat újra befekteti-e, vagy a pozíció méretek végig egységnyiek-e szintén nem tisztázza.

4.1.2. Második stratégia: Extrém szentiment nap (ESZN)⁵⁵

Más megközelítéssel vizsgálja egy szentimenten alapuló stratégia gyakorlati kivitelezhetőségét David Leinweber and Jacob Sisk, akik a különlegesen erős szentimentű napokat, mint vételi/eladási jeleket állították a középpontba (Leinweber & Sisk, 2010) és (Leinweber & Sisk, 2011). A szerzők (Tetlock, et al., 2008) munkájának logikáját követve először több előzetes esettanulmányt végeztek, és végül ezek alapján határozták meg az ESZN stratégia végleges formáját, amire egy portfóliószimulációt mutatnak be. Az esettanulmányok során részben az ESZN paramétereit, illetve a vizsgált vállalatok körét módosítják, de minden esetben akkor veszünk fel pozíciót egy adott vállalatban, ha erre a vállalatra úgynevezett ESZN következett be. A szerzők a Thomson Reuters NewsScope Sentiment Engine (RNSE) adatbázisát használták, az esettanulmányok időablaka 2003-2008, a vállalatok univerzuma pedig az S&P 1500.

Előzetes esettanulmányok

Az alábbi táblázat (1) oszlopában az ESZN egy olyan 15:30-15:30 terjedő 24 órás időszak, amikor legalább kettő olyan, az adott vállalatra legalább 60%-ban releváns, hír⁵⁶ jelenik meg,

⁵⁵ Egy független és nagyon hasonló kutatás (Cahan, et al., 2010).

⁵⁶ Lehet cikk, riasztás vagy megjegyzés (append). Az RNSE hírkategóriái.

ami szentimentjét tekintve az összes előző nap megjelent hír felső vagy alsó 10%-ában van⁵⁷, valamint további a kritérium, hogy ne létezzen a vállalatra vonatkozó hasonló hír az elmúlt hónapban⁵⁸ (újdonságérték). Tehát ESZN, amikor legalább kettő releváns, különböző, új és nagyon negatív vagy pozitív hír jelenik meg 24 órán belül⁵⁹. Az egész S&P 1500 univerzumra pozitív ESZN-t követő 3. npra átlagosan 14 bázispont, negatívra pedig -33bázispont CAAR-t⁶⁰ mértek. A spread tehát 47 bázispont, lásd 2. táblázat (1) oszlop. Mivel ezt nem tartották elégségesnek a tranzakciós költségek ellensúlyozására, megvizsgálták, hogy az egyes iparágak esetén a spread nem magasabb-e⁶¹.

Ezt elvégezve azt találták, hogy különösen a nyersanyag, a ciklikus és nem ciklikus, a pénzügyi, ipari és technológiai szektorok bizonyultak érzékenyek. A hasonló logikájú, de paramétereiben különböző (2), (3), (4) és (5) esettanulmányok CAAR adatait a 2. táblázat rendre megfelelő elnevezésű oszlopai tartalmazzák.

	(1)	(2)	(3)	(4)	(5)
Hírek száma a 24 órás ablakban	min. 2	min. 2	min. 2	min. 2	min. 2
Iparágak	összes	pénzügyi	pénzügyi	nem ciklikus	nem ciklikus
Szentiment pozitív/negatív	top 10%	top 10%	top 5%	top 10%	top 5%
Hír újdonságérték	adott hónapban sincs hasonló	adott hónapban sincs hasonló	adott hónapban sincs hasonló	adott hónapban sincs hasonló	adott hónapban sincs hasonló
ESZN-ek száma összesen	10317	1552	452	615	182
3 napos pozitív - negatív ESZN CAAR spread	47 bp	85bp	147bp	55bp	138bp

2. táblázat A teljes S&P 1500 és a szektoronkénti szentiment CAAR spreadek

A következő lépésben a kapitalizációt vették szemügyre. Azt a – dolgozatomban már bemutatott – jelenséget vizsgálták, hogy a kis kapitalizációjú (ezáltal feltehetően a médiában is kevesebbet szereplő) vállalatok érzékenyebbek a szentimentre. Egy ESZN-hez most is legalább 2 új⁶², releváns, különböző hírt⁶³ megkövetelve, a 5. ábra a CAAR⁶⁴ alakulását mutatja kapitalizáció szerinti különböző csoportokra, az ESZN-t követő 60 napig. Az abnormális hozamok többnyire a várakozásoknak megfelelően alakultak (minél kisebb

⁵⁷ Habár kifejezetten nem írták, feltehetjük, hogy mindekét hírek vagy a negatív vagy a pozitív tartományban kell lennie. Az RNSE nem közvetlenül egy hír pozitívitasát vagy negatívitasát mutatja, hanem, hogy az osztályozó milyen valószínűségekkel osztályozza pozitív vagy negatív hírek.

⁵⁸ Az RNSE adatbázisban a hírek egyik metaadata a „linked count” ami azt mutatja, hogy különböző időszakokra visszamenőleg jelent-e meg hasonló hír, az adott vállalatra vonatkozóan.

⁵⁹ Habár kifejezetten nem írták, feltehetjük, hogy a pozíció felvétele a második hír megjelenésekor történik.

⁶⁰ A SPY hoz képest.

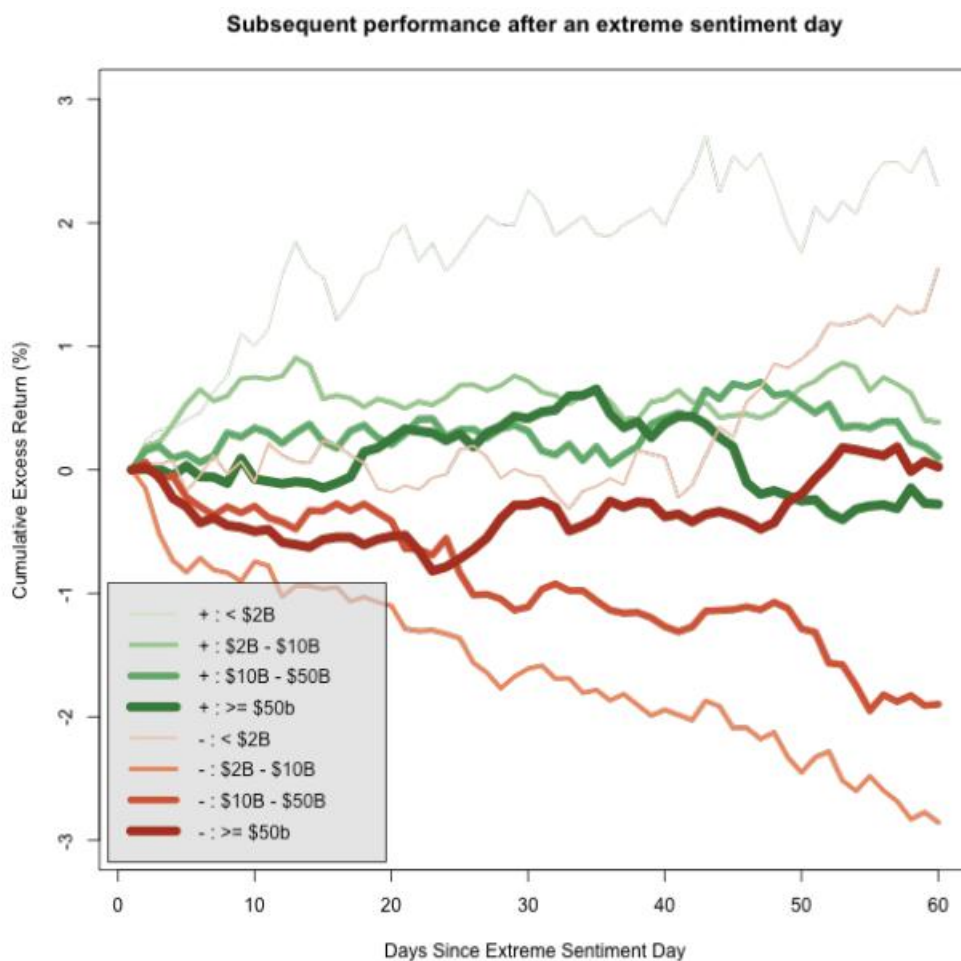
⁶¹ Az iparági bontás a Thomson Reuters Business Classification (TRBC) szerint történt.

⁶² Itt: nincs hasonló hír az adott vállalatra a megelőző 5 napban.

⁶³ A (3) és (5)-ös esettanulmányokhoz hasonlóan az napi szélső 5% -ban szélsőséges szentimentű hírt.

⁶⁴ Az adott vállalat iparága és kapitalizáció szerinti csoportjához képesti.

vállalat, annál nagyobb hatás), de érdekes kivétel a legkisebb vállalatok csoportja (<2 milliárd USD). Itt a negatív ESZN 60 napos CAAR-ja pozitív, több mint 1%. További tanulság, hogy 50 milliárd USD kapitalizáció fölött sem egy pozitív, sem egy negatív ESZN-nek sincs szignifikáns CAAR hatása a 60 napos ablakban, és 10-50 milliárd USD között is inkább csak a negatív esetben. Az előbbi anomália miatt talán azt lehet egyelőre mondani, hogy nem a legkisebb, hanem a közepesen nagy, 2-50 milliárd USD kapitalizációjú vállalatok negatív ESZN-eket és a 2 milliárd USD-nél kisebb vállalatok pozitív ESZN-eket követő árfolyamalakulására lehet stabilan kereskedni.



5. ábra Átlagos ESZN 60 napos CAAR hatások a S&P 1500 vállalatokra kapitalizáció szerint, 2003-2009Q3 között.
 Forrás: (Leinweber & Sisk, 2010)

ESZN stratégia: portfólió szimulációk

A tanulságok alapján a szerzők az ESZN stratégiának a végső formáját az alábbiakban határozták meg, fenntartva, hogy optimális esetben a szentimentre építő döntéstámogatás az eszköztár fontos eleme, de másokkal együtt alkalmazva ajánlják.

A stratégiában részt vevő iparágak az S&P 1500 technológiai, ipari, egészségügyi, pénzügyi, alapanyag vállalatai, mely iparágak a 2003-2005 időszak ESZN reakciói alapján lettek

kiválasztva. Vételi jelzés az ESZN alapján, most az alábbi paraméterekkel: legalább 4(!) új, különböző, releváns hír a new-yorki 15:30-at megelőző 24 órából, melyek az aznapi átlagos hírszentsiment eloszlás szélső 5%-ába tartoznak. ESZN esetén pozíciót veszünk fel az adott vállalatban aznapi záró áron a megfelelő irányban, amit 20 nap után zárunk, szintén záró áron. 25bp tranzakciós költséggel számolnak (vétel+eladás összesen), profit loss 5%, take profit 20% egy pozícióra. Tőke management megkötés volt, hogy egy pozícióban sem lehet több, mint a teljes tőke 15%-a⁶⁵. A szimuláció időablaka 2006 január 1 – 2009 november 1, a szerzők végül nem tettek megkötést a stratégiában szereplő vállalatok kapitalizációjára. A szimuláció alapadatai az alábbiak:

Éves hozam:	21,48%
Sharpe-mutató:	0,72
Maximum drawdown:	60%
Nyereséges hónapok aránya:	52%
Kötések száma:	229

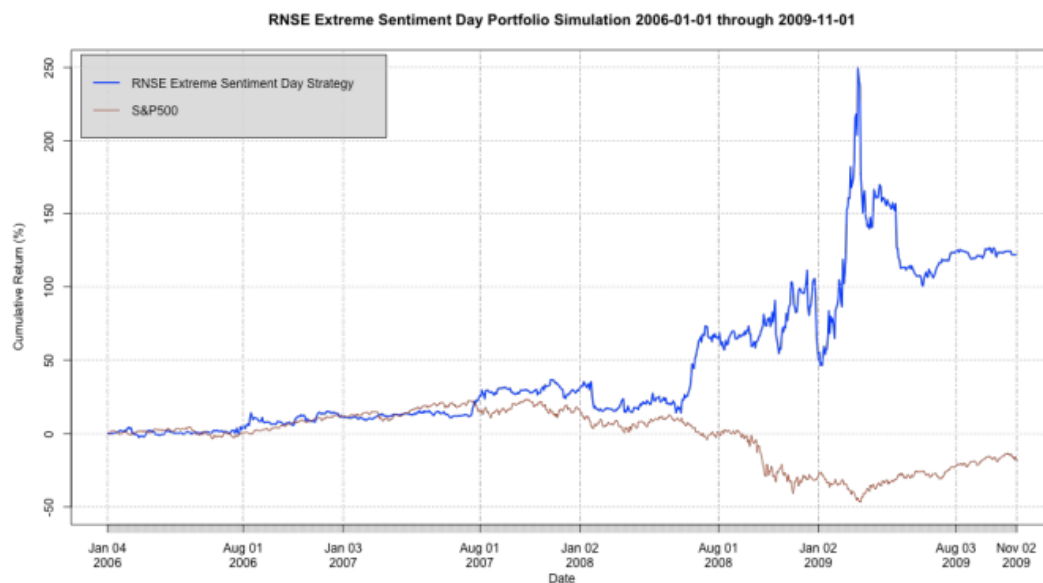
3. táblázat: ESZN stratégia portfóliószimuláció alapadatok

Kritikák és észrevételek

Összefoglalva az eredményeket, az éves 21,48%-os hozam önmagában impozáns (az S&P 500 hozama az időszakra -6% körül volt), de a portfólió időbeli teljesítménye (6. ábra), illetve nyereséges és veszteséges hónapok megoszlása (7. ábra) elgondolkodtató. A teljesítmény ugyanis közel sem mondható egyenletesnek, a szimuláció több mint feléig alig tudja az S&P 500-at túlteljesíteni, és a teljes hozam tulajdonképpen két nagy ugrásból tevődik össze, pontosabban ezeket leszámítva összességében veszteséget könyvelhetnénk el végül. A stratégia az első évben tulajdonképpen igen gyengén teljesít. Annak, hogy ezután miért kezd el javulni, annak a szerzők szerint a RNSE-ben bekövetkező technológiai fejlesztés és az ebből fakadó hírmennyiség növekedés az oka. Habár az ESZN, mint elgondolás, szerintem megfelelő, mert jól megragadja a rövid időn belül érkező intenzív új információt, amit a várakozásaink szerint a befektetők lassabban/rosszul tudnak csak feldolgozni, de véleményem szerint nem használja ki az automatizált szentimentelemzés lehetőségeit. Ugyanis egy ilyen ESZN olyan rendkívüli esemény, hogy feltehetően azok is felfigyelnek rá, akik nem használnak ilyen eszközöket. Hafez stratégiáját azért tartom ennél szerencsésebbnek, mert finomabb szentimenthatásokkal dolgozik, mondhatni statisztikailag próbálja learbitrálni a szentimentet. Továbbá jobbnak tartom, hogy egyedi vállalatok helyett iparágakba fektetett,

⁶⁵ Sajnos ennél több konkrétumot több megkeresésem ellenére sem tudtam a szerzőktől megtudni. Mindazonáltal a két dolgozat hangvétele és a prezentációk során érdekes módon mindig sikeres újszerű befektetési stratégiák iránti erős kritikájuk alapján arra tudok következtetni, hogy igazságos módon határozták meg a pozíciók nagyságát, ezzel nem „csaltak”.

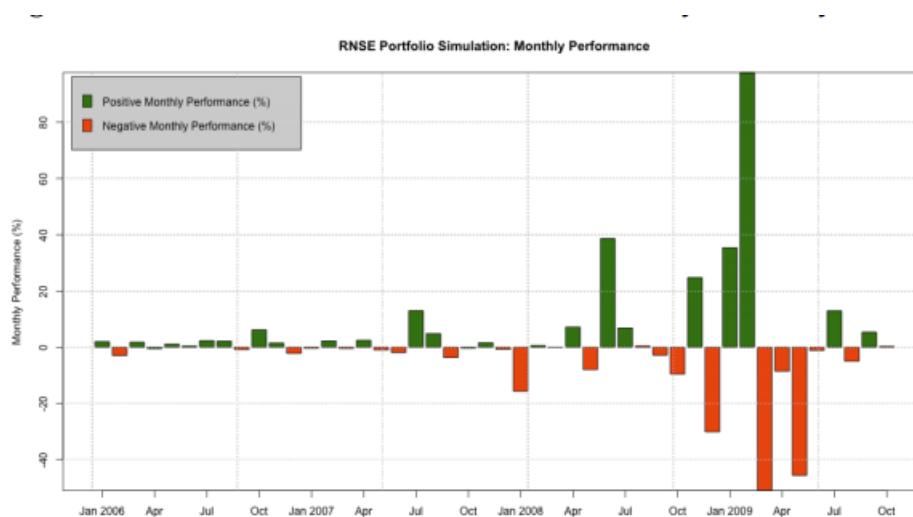
mert az iparágakon belüli jelentős szentiment spillover hatások miatt a hozamból talán kevesebbet veszít, mint amit azon nyer, hogy nem fut bele egy-két igen rossz egyedi befektetésbe. A szentiment hatása ugyanis, mint már láttuk: „nem egyértelmű”, főleg nem egyetlen vállalatra tekintve. Emellett azonban meg kell említeni azt is, hogy a veszteséget is ugyanúgy két-három nagy bukás eredményezi, de a nagy ugrásokat tartalmazó szakaszokat leszámítva a maradék időszakban viszonylag lapos, de inkább emelkedő jellegű a tőke pályája. Hasonló a benyomásom a nyereséges-veszteséges hónapokat bemutató ábra alapján is. A maximális drawdown értéke is kétszerese az ajánlott 30%-nak, illetve az eredményeket eltorzítja a stop loss és take profit alkalmazása⁶⁶, ami a szerzők szerint is nagyon meghatározó volt a portfólió szempontjából⁶⁷. Pozitívumként azonban meg kell említeni, hogy a szerzők futtattak még egy forward tesztet 2010 első 9 hónapjára (ők *ténylegesen* mintán kívülnek nevezik, mert olyan adatsoron végzik a tesztet, ami az ESZN stratégia megalkotásakor még nem is létezett). Itt kiegyensúlyozott teljesítmény mellett 11,5%-kal teljesítette felül a benchmark S&P 500-at (évesített abszolút hozam közel 14%).



6. ábra: ESZN stratégia összehasonlítása az S&P 500-zal. Forrás (Leinweber & Sisk, 2010).

⁶⁶ A választott 5% és 20%-os értékek standardnek tekinthetők egy hasonló jellegű stratégia esetén.

⁶⁷ Ezt az itt nem közölt, de az eredeti tanulmányban megtalálható nyereséges és veszteséges pozíciók eloszlása alapján megerősíthetünk.



7. ábra ESZN stratégia: nyereséges és veszteséges hónapok. Forrás: (Leinweber & Sisk, 2010).

4.1.3. Harmadik stratégia: Szentiment-visszafordulás⁶⁸

A cél ebben az esetben is egy átlagos technológiai és anyagi háttérrel rendelkező befektető számára olyan stratégia megalkotása, mellyel képes kihasználni az automatizált hír és szentimentelemzés lehetőségeit. A vételi jel egy cégspecifikus szentimentindex átmeneti, de tartós negativitásának pozitívításba való (vissza)fordulása. A szerző a Dow Jones News Analytics-et (DJNA) használta, mely a Dow Jones RavenPack-kel történő együttműködésének gyümölcse. Az adatbázisban 1987-ig visszamenően 27000 amerikai vállalat kapcsolódó hírei és azok metaadatai szerepelnek, többek között természetesen a szentiment pontszámok. Ez az adatbázis is követi a vállalatok időközbeni névváltozásait⁶⁹ és nem torzít a túlélés miatt⁷⁰. A stratégia csak vételi jelet alkalmaz, így nem tekinthető piacsemlegesnek. A stratégiába csak az 500 millió USD-nél nagyobb vállalatokat vonták be⁷¹, de más módon nincs leszűkítve a vállalatok köre.

⁶⁸: Sentiment reversion, (Kittel, 2010) alapján.

⁶⁹ Illetve felvásárlások, összeolvadások és az is, hogy az IBM-et időnként Big Blue-nak is hívják.

⁷⁰ Angolul survivalship bias-ként hivatkozott torzítás. Ha vállalatok egy csoportjára, például egy szektorra készítünk valamilyen elemzést, például múltbeli átlagos teljesítményükre vonatkozóan, akkor az eredményeket helyesen kell értelmezni. A kapott átlagos teljesítmény ugyanis nem a szektor átlagos múltbeli teljesítménye, hanem csak azon vállalatoké, akik még mind máig működnek – feltehetően nem a legrosszabb vállalatok. Korrekt válaszhoz azokat is be kell vonni a vizsgálatba, akik idő közben megszűntek.

⁷¹ Kapitalizáció. Praktikussági okai vannak, ugyanis alapkezelők bizonyos kezelt tőke nagyság fölött nem fektetnek kis vállalatokba a forrás álláspontja szerint. Az 500 milliós minimum deflátor a S&P 500 index.

A vételi jel tehát a szentiment visszafordulás. Ez a lenti módon képzett cégspecifikus szentimentindex pozitivitása, tartós átmeneti negativitása végül pozitív tartományba való visszatérése⁷². Erősebb jel konstruálása céljából szigorításokkal alkalmazott az alábbi σ

$$\sigma(C, N_i, P) \rightarrow \mathbb{Z}$$

mozgóátlag jellegű függvényre. σ minden N_i „aktuális hír” megjelenése idején képi azt a mozgóátlagot, melybe N_i és N_i megjelenését megelőző P hónapban megjelenő a C vállalatra vonatkozó releváns hírek⁷³, illetve azok szentimentpontszámai kerülnek bele egyenlő súlyozás mellett. A megszorítások:

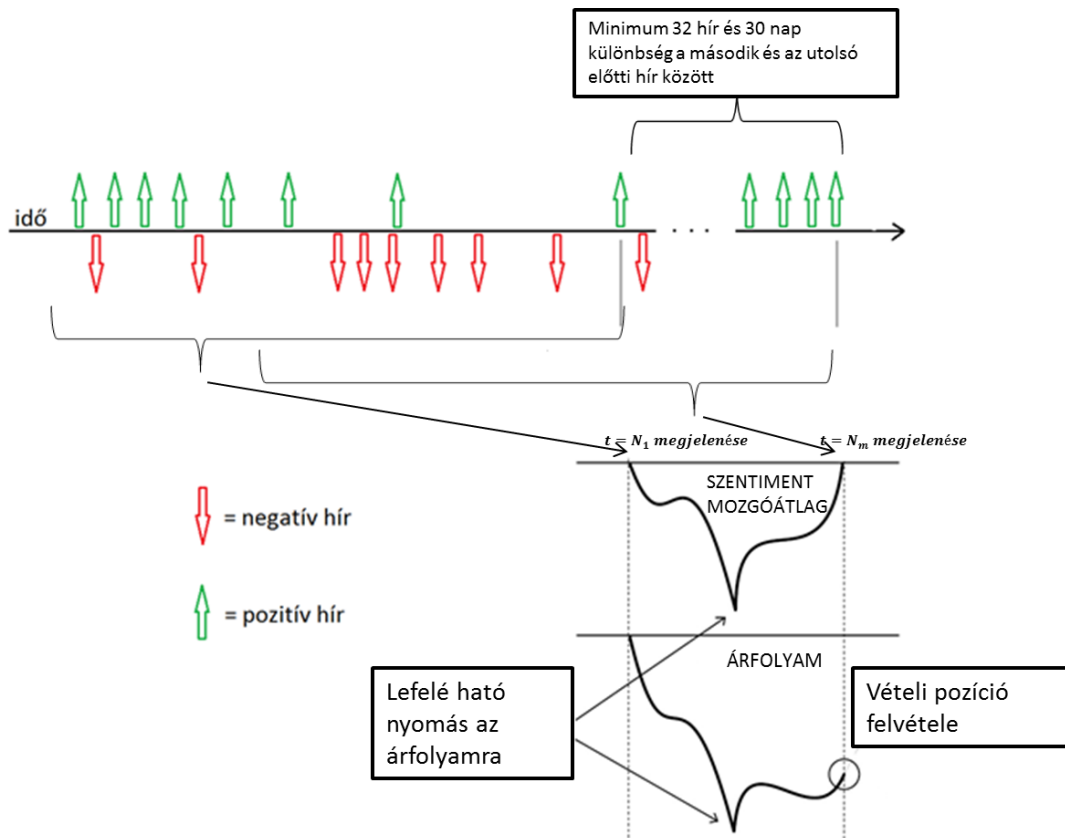
1. $m \geq 32$
2. az N_{m-1} és az N_2 hírek publikálásának ideje között legalább 30 naptári nap van
3. $\sigma(C, N_i, P) < 0$ minden $2 \leq i \leq m - 1$
4. $\sigma(C, N_1, P) \geq 0$
5. $\sigma(C, N_m, P) \geq 0$

Tehát a szentiment visszaforduláshoz legalább $m=32$ hírre van szükség úgy, hogy a második és az utolsó megjelenése között legalább 30 nap különbség szükséges (1 és 2). Továbbá a 3. formálisan azt jelenti, hogy első és az utolsó hír megjelenésével a mozgóátlag nemnegatív, minden más köztes hírre pedig szigorúan negatív. A szentiment visszafordulás, mint vételi jel időpontja az m . hír megjelenési időpontja.

A mozgóátlagra a simítás miatt van szükség, ugyanis egy hír szentimentje önmagában nem bír nagy információértékkel, és túl zajos is lenne az indikátor. A feltevés szerint egy tartós negatív hangulat lefelé ható nyomást gyakorol az árfolyamra, ami a negativitás megfordulásával elmúlik, és az árfolyam visszatér a magasabb szintre. Ez a stilizált ábra bemutatja az időtengelyt, a megjelenő híreket és azt, hogy hogyan csúszik a mozgóátlag ablaka.

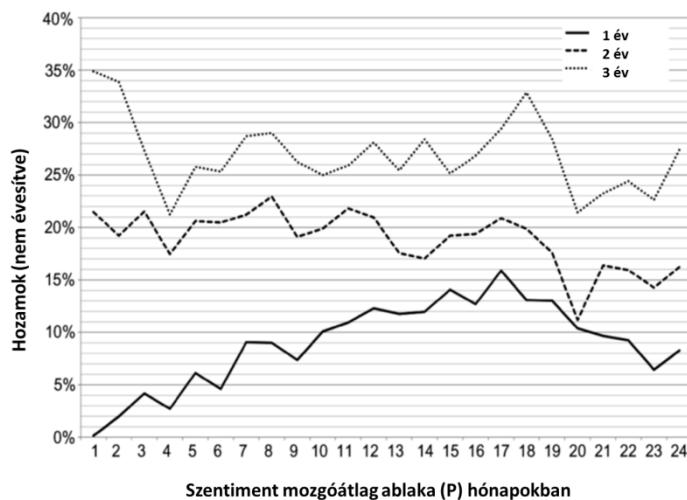
⁷² A formális képletek megtalálhatóak a forrásban, de matematikai szempontból nem olyan érdekesek, hogy értelme lenn itt is szerepeltetni őket.

⁷³ A 90 relevanciaérték fölötti híreket vesszük figyelembe.



8. ábra: Szentiment megfordulás: A szentiment megfordulás és a részvényárfolyam feltételezett kapcsolatának stilizált ábrázolása. (Kittrel, 2010) alapján, részben saját készítés

Utolsó feladatunk annak a meghatározása, hogy mennyire legyen visszatekintő a mozgóátlag. Ez egy önkényes döntés, végül a 9. ábra alapján 17 hónaposnak választotta P értékét a szerző.



9. ábra Szentiment megfordulás: S&P 500 feletti extrahozamok a szentiment megfordulást követően 12, 24 és 36 hónapos tartás esetén, P függvényében⁷⁴, 2003 és 2008 között. Forrás:

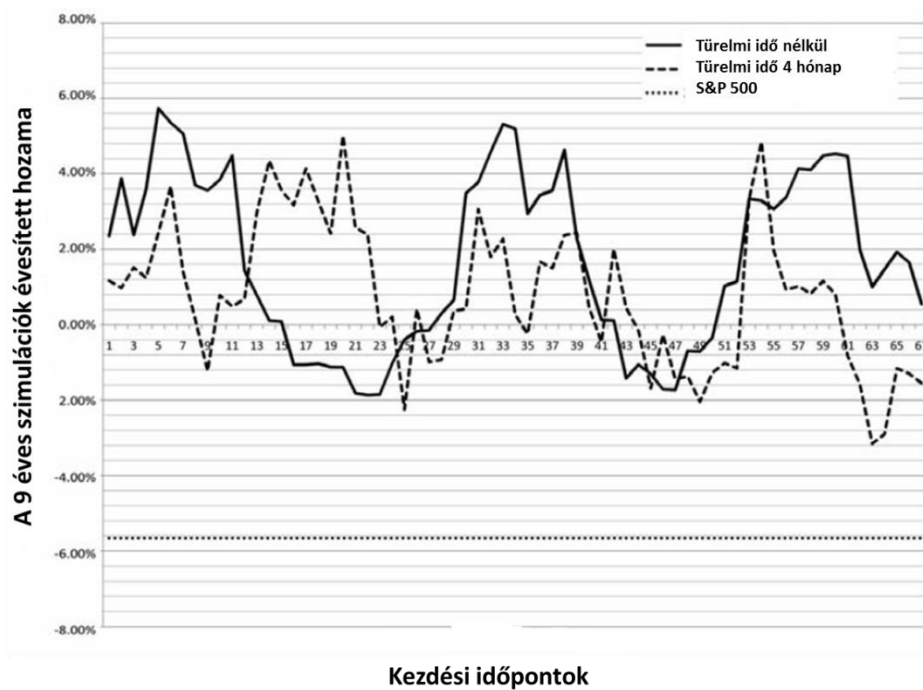
⁷⁴ A hozamokat a jobb összevethetőség kedvéért winzorálták.

A 9. ábra rögtön két kérdésre is választ ad. Mivel 2 illetve 3 éves tartás esetén semmilyen tényleges kapcsolatot sugalló függvényforma nem rajzolódik ki, ezért még talán akkor is az 1 éves tartást választanánk, ha nem ennek lennének a legjobb hozamai (az ábrán a hozamok nem évesítettek). Mindenesetre a tény, hogy mindhárom tartási hossz mellett 17-18 hónapnál van (lokális) maximum, illetve, hogy egy év esetén a függvény alakja nem teljesen esetleges, az arra utal, hogy valóban van kapcsolat a szentiment visszafordulás és azt követő hozamok között, nem pedig csak egy erre az időszakra vonatkozó túltanítás áldozatai vagyunk. A 9. ábra szerint tehát közel 16%-os éves extrahozamra tehetünk volna szert 2003 és 2008 között, ha minden szentiment visszafordulást lekereskedünk (tranzakciós költségek nélkül).

A következő lépésben a szerző szimulálta a kereskedési stratégiát (tranzakciós költségeket nem vett figyelembe) az alábbi befektetési szabállyal. A kezdeti tőkét 10 egyenlő részre osztjuk és sorba állítjuk. Ha vételi jelet kapunk és van be nem fektetett rész, akkor az első szabad részt befektetjük, és amint lejár az 1 év, a pozíciót az aktuális értékén likvidáljuk, és mint szabad részt ezen az értéken a sor végére állítjuk. Egy második változat csak abban különbözik ettől, hogy ha jelet kapunk, de nincs szabad rész, akkor elindul egy 4 hónapos "türelmi idő", melynek lejárta előtt, ha felszabadul egy rész, akkor még beléphetünk vele ebbe a pozícióba (1 évre), így kevesebb vételi jelről maradunk le.

A stratégia jellegéből adódóan nem indifferens, hogy mikor is kezdjük el, ugyanis egy kis változás is eredményezheti, hogy túlnyomórészt végül különböző pozíciókat fogunk birtokolni. A szerző emiatt 67 darab 2000 előtti kezdési dátumot választott véletlenszerűen a stratégia elindítására⁷⁵. A 10. ábra ezeknek a portfólióknak az évesített teljesítményét mutatja a 2003-2008-as időszakra. Emlékeztetőül, ha képesek lennénk a teljes tőkénket állandóan és egyenlő részben minden vételi jelre pozícióban tartani, akkor hozzávetőleg 16% éves extrahozamra tehetnék szert, így ez az elméleti korlátja a stratégiának. Pozitívként meg kell említeni, hogy a stratégia minden egyes kezdési időpont esetén túlteljesíti az S&P 500-at, azonban ez nem egy nagy eredmény, tekintve, hogy az index évesített hozama erre az időszakra -5,65%. A 67 szimuláció átlagos teljesítménye az első stratégiaváltozatra évi 1,67% míg a másodikra évi 0,8%, ami az S&P 500 -hez képest rendre 7,32% és 6,46% éves többlethozamot jelent.

⁷⁵ Hogy miért pont 67, illetve, hogy ezek milyen intervallumon lettek kijelölve nem tér ki a szerző.



10. ábra Szentiment megfordulás: A 67 szimuláció és az S&P 500 évesített hozamai a 2003-2008-as időszakra.

Az eredmények biztatóak, különösen, hogy csak long jellege ellenére képes volt pozitív hozamra egy folyamatosan eső időszak alatt. Ennek ellenére lehetne piac semleges, vagy ha nem, akkor meg kellene indokolni, hogy ellenkező irányban miért nem lehet hasonló (eladási) jelet konstruálni. A szentimentpontszámokat binárisra transzformálja le a szerző, és indoklás hiányában ebben csak az információvesztés látszik. A DIJA adatbázisban ugyanis eredetileg egy széles skálán vannak meghatározva a hírek szentimentpontszámai. A 9. ábra alapján indokolt lenne 1 évnél rövidebb tartásra is megnézni az S&P 500 fölötti többlethozamokat. A leírt stratégia rossz hatásokkal fekteti be a tőkét, sok a készpénz pozíció. Emellett nem veszi figyelembe a tranzakciós költségeket sem. Az alkalmazott 67 szimulációnál többet is alkalmazhatnánk, ráadásul pontosabban meg kellene határozni, hogy hova esnek a lehetséges indítási időpontok.

5. Egy saját stratégia: A globális szentiment és a devizaárfolyamok

Az irodalom és a már bemutatott konkrét stratégiák áttekintése után személyesen is megszerettem volna vizsgálni egy szentimentindex, mint predikciós faktor erejét, biztosítottan olyan motivációktól mentesen, hogy a ténylegesnél kedvezőbb képet alkossak róla. Méréseim pontosan emiatt teljes egészében reprodukálhatóak, a használt VBA kódokat is mellékeltem a 8.2. mellékletben.

Vizsgálatom tehát egy szentimentindex előrejelző erejét méri különböző deviza keresztárfolyamokra vonatkozóan. A devizapiac ugyan a dolgozat alapján tett megállapítások alapján nem feltétlenül ideális terep a szentimentindexre épülő stratégiákhoz (likvid piac, alacsony kisbefektetői hányad), de a devizapiacra nem találtam ilyen irányú kutatást, itt olcsó és folyamatos a kereskedés, illetve könnyen, mindenki számára és alacsony tőkével is hozzáférhető a piac. A hazai fizetőeszköz, illetve általában a Közép-Kelet Európai régió devizái is ki vannak téve a globális befektetői hangulat változásainak az általános vélekedés szerint. A vizsgálat időablaka 2000.1.2 -2013.10.04 és a Magyar Forint (HUF), a Török Lira (TRY), az Orosz Rubel (RUB), a Lengyel Zloty (PLN), a Román Lej (RON), a Cseh Korona (CZK), a Dél-Afrikai Rand (ZAR), a Svéd Korona (SEK), a Brazil Reál (BRL), az Argentin Peso (ARS) és az Indiai Rúpia (INR) Euróval (EUR) és Amerikai Dollárral (USD) szembeni keresztárfolyamai vesznek benne részt⁷⁶, azaz a kereskedett szimbólumok jelen esetben az EURHUF, EURTRY, EURRUB, EURPLN, EURRON, EURCZK EURSEK, EURBRL, EURARS, EURINF, illetve az USDHUF, USDTRY, USDRUB, USDPLN, USDRON, USDCZK, USDSEK, USDBRL, USDARS, USDINR. A használt szentimentindex a Global Sentiment Macro Index, pontosabban ennek 3 hónapos mozgóátlaga (GSMI), melyet a RavenPack bocsátott a rendelkezésemre. A GSMI a világban megjelenő nagymennyiségű makrogazdasági hír algoritmizált módon számított szentiment pontszámait aggregálja egy naponta frissített mutatóban, ezért a változása felfogható a világgazdasági hangulatnak napról napra történő széles körű leképeződéseként. Mind a GSMI, mint a devizakeresztek esetében tehát napi adatokkal dolgoztam, a GSMI esetében annak változását vettem alapul, a devizakereszteknél pedig a loghozamokat. Céлом az objektivitás és a túltanítás elkerülése, ezért a lehető legegyszerűbb stratégiát alkottam: **a hangulat romlását követően (long) pozíciót veszek fel**. Ezt először esettanulmány majd portfólió szimulációk formájában teszteltem.

⁷⁶ Az árfolyam adatok a Bloomberg historikus napi utolsó árai (PX_LAST).

5.1. Esettanulmányok

Az előzetes esettanulmányok célja csupán betekintést nyerni a GSMI és a devizakereszték idősorainak tulajdonságaiba és a köztük lévő kapcsolatok erősségébe, továbbá meghatározni, hogy mi a vizsgálandó tartománya az egyes paramétereknek.

Az *esemény* a GSMI legfeljebb megadott (T_GSMI) nap alatt bekövetkező legalább megadott mértékű (req_chg_GSMI) csökkenése. Ha ez bekövetkezik, akkor egységnyi LONG pozíciót veszünk fel az esemény napjának⁷⁷ utolsó árán a megadott szimbólumokban, majd megadott tartási napot (T_FX) követően szintén aznapi utolsó áron zárjuk. A T_GSMI, req_chg_GSMI és a T_FX a módosítható paraméterek, ezeknek az összes lehetséges kombinációjában elvégeztem a mérést az összes kereszt esetében a teljes időszakra. Minden mérés esetében kiszámításra került (többek között) az események száma, az ezeket követő pozíció felvételek kumulált loghozama, az egy eseményre jutó hozam, illetve az, hogy az esemény után milyen arányban alakultak a két irányba az árfolyamok (a tartási idő alatt). Fontos megjegyezni, hogy az esettanulmány egy puszta mérés, kereskedési szempontból az eredményei nem különösebben relevánsak⁷⁸.

Az adatok alapján az alábbi tartományban vizsgáltam a paramétereket, minden esetben a teljes idősorra, az Euróval és az US Dollárral szemben az összes többi devizára.

T_GSMI			req_chg_GSMI			T_FX		
min.:	max.:	lépésköz:	min.:	max.:	lépésköz:	min.:	max.:	lépésköz:
1	5	1	0,1	3	0,1	1	5	1

Tágabb határok a mérések alapján nem indokoltak, mert a legjobb stratégiák paraméterei tipikusan nem a szélsőértékek közelében vannak, ezek a határok és lépésközök 750 mérést jelentenek devizapáronként.

Már a legelső vizsgálatok alkalmával világossá vált, hogy egy kontrárius jellegű stratégiának van létjogosultsága, azaz a negatív eseményt követően már inkább a korrekcióra érdemes fogadni, vagyis amikor a már a hangulat is elromlott, akkor érdemes a trendfordulásra számítani. Ez konkrét esetben például azt jelenti, hogy az esemény (hangulat elromlása) után

⁷⁷ Naptári nap. A legtöbb devizával minden nap, éjjel nappal lehet kereskedni valamilyen platformon.

⁷⁸ Ennek legalapvetőbb oka, hogy az itt kapott kumulált hozam úgy adódik, hogy minden esemény után pontosan ugyanakkora pozíciót veszünk fel. Ez nem kivitelezhető, ha akkor is esemény következik be, ha éppen nincs szabad tőke (átfedés) vagy nem az egész tőkét fektetem be egy pozícióba, akkor pedig a készpénzállomány - ami nem termel hozamot - rontja az eredményem. Másrészt egy tipikus befektető visszaforgatja a hozamot, így az az álladó pozíció méret sem valóságos.

SHORT pozíciót érdemes felvenni⁷⁹ a megadott szimbólumokban, azaz a kis devizák, (amik az időszakban inkább „risk off” tulajdonságúnak tekinthetők) felértékelődésére érdemes fogadni a nagydevizákkal, a Dollárral és az Euróval szemben. Az EURHUF esetében a 750 mérés majdnem 280000 eseményét figyelembe véve 54,67%-ban negatív, átlagosan -0,03% az elmozdulás az esemény után a tartási időszak alatt. Ezek a számok talán nem tűnnek túl impresszívnek, de figyelembe kell venni, hogy ebbe minden egyes paraméter konfiguráció beletartozik (sok köztük az olyan, ami nyilvánvalóan nem merülne fel éles használat során).

Részben emiatt is, minden szimbólum esetében a méréseket sorba rendeztem aszerint, hogy mekkora az egy eseményre eső átlagos változás (azaz kumulált loghozam a tartási időszak alatt) illetve aszerint, hogy mekkora százalékban a várt irányú (tehát itt negatív) ez a változás. Ez utóbbit találati aráynak neveztem el. A 4. táblázat megfelelő sorai a mérések ezen szempontok szerinti legjobb 20%-nak⁸⁰ néhány összesítő adatát tartalmazzák. Kivétel ez alól a „nem veszteséges stratégiák aránya” sor, ami azt mutatja, hogy az összes 750 beállítás hány százaléka nem volt veszteséges, ami jelen esetben azt jelenti, hogy nem pozitív az egy eseményre eső átlagos változás.

		EURHUF	EURTRY	EURRUB	EURPLN	EURRON	EURCZK	EURZAR	EURSEK	EURBRL	EURARS	EURINR
Korrektions faktor		1,00	0,59	0,97	0,84	1,02	1,34	0,54	1,27	0,53	0,67	0,81
nem veszteséges stratégiák aránya		71%	52%	16%	80%	30%	85%	36%	39%	51%	11%	21%
eseményenkénti változás alapján legjobb 20%	átlagos eseményenkénti változás	-0,33%	-0,21%	-0,01%	-0,19%	-0,13%	-0,37%	-0,10%	-0,18%	-0,31%	0,00%	-0,03%
	átlagos találati arány	61%	65%	54%	61%	85%	60%	71%	63%	57%	54%	53%
	átlagos eseményszám	66	63	244	74	32	76	112	83	39	298	96
	átlagos elméleti éves hozam	1,6%	0,9%	0,2%	1,0%	0,3%	2,0%	0,8%	1,1%	0,9%	0,0%	0,2%
találati arány alapján legjobb 20%	átlagos eseményenkénti változás	-0,12%	-0,09%	0,01%	-0,07%	-0,12%	-0,12%	-0,06%	-0,10%	-0,10%	0,02%	-0,01%
	átlagos találati arány	79%	79%	73%	76%	91%	77%	80%	75%	74%	72%	74%
	átlagos eseményszám	144	171	339	186	23	206	134	135	298	305	227
	átlagos elméleti éves hozam	1,2%	1,1%	-0,2%	1,0%	0,2%	1,8%	0,6%	1,0%	2,1%	-0,4%	0,1%

⁷⁹ A használt kód ezzel szemben LONG pozíciót vesz fel, ezért a kapott hozamok negatívak lesznek. A kérdés, hogy mennyire konzekvensen negatívak, hiszen ez dönti el, hogy mennyire helytálló a stratégia hipotézise. A 4. táblázatban is tehát a negatív hozamok számítanak „jóknak”.

⁸⁰ A 20% önkényes választás és természetesen torzítja az eredményeket, hiszen csak a legjobbakat tartalmazza. A cél azonban itt nem releváns eredmények számítása, hanem csak a devizapárok összehasonlítása.

		USDHUF	USDTRY	USDRUB	USDPLN	USD RON	USDCZK	USDZAR	USDSEK	USDBRL	USDARS	USDINR
Korrektíós faktor		1,00	1,06	2,59	1,07	1,44	1,17	0,87	1,18	1,01	2,80	2,89
nem veszteséges stratégiák aránya		93%	78%	50%	92%	64%	96%	54%	78%	80%	11%	52%
eseményenkénti változás alapján legjobb 20%	átlagos eseményenkénti változás	-0,97%	-0,90%	-0,41%	-0,89%	-1,12%	-1,07%	-0,46%	-0,87%	-1,20%	0,01%	-0,52%
	átlagos találati arány	70%	68%	59%	59%	79%	71%	63%	53%	55%	52%	73%
	átlagos eseményszám	33	36	89	49	24	34	38	45	23	88	25
	átlagos elméleti éves hozam	2,3%	2,3%	2,6%	3,2%	2,0%	2,6%	1,3%	2,9%	2,0%	0,0%	1,0%
találati arány alapján legjobb 20%	átlagos eseményenkénti változás	-0,62%	-0,43%	-0,11%	-0,24%	-0,90%	-0,65%	-0,19%	-0,20%	-0,29%	0,06%	-0,32%
	átlagos találati arány	85%	81%	77%	79%	89%	86%	78%	73%	78%	70%	85%
	átlagos eseményszám	43	161	165	185	27	44	190	370	220	374	87
	átlagos elméleti éves hozam	1,9%	5,0%	1,3%	3,2%	1,8%	2,1%	2,6%	5,3%	4,6%	-1,6%	2,0%

4. táblázat: Az esettanulmányok legjobb 20%-nak összefoglaló táblázata. A felső táblázat az Euróval az alsó az US Dollárral szembeni kereszteket tartalmazza. A Korrektíós faktor az EURHUF szimbólum értékét véve 1-nek korrigálja az átlagos eseményenkénti változásokat a többi szimbólum esetében az alapján, hogy az adott szimbólum és az EURHUF loghozamainak átlagos abszolút eltérése hogyan viszonyul egymáshoz. Tehát összehasonlíthatóvá teszi az eredményeket a szimbólumok volatilitása szerint normálva, ezek volatilitása ugyanis különböző, ami torzítaná az eredményeket. Az US Dollárral szembeni szimbólumok esetében analóg módon. A „nem veszteséges stratégiák aránya” azon stratégiák aránya, ahol az átlagos eseményenkénti hozam 0% vagy alacsonyabb. A 750 beállítás közül pontosan 35 esetben nincs esemény (túl szigorú beállítások), a nyereséges stratégiák aránya így ezekből legalább 4,66%-ot levonva adódik. A többi sor már a leírt módon kiválasztott legjobb 20% adatait tartalmazza. Az „átlagos találati arány” az mutatja, hogy egy esemény után mekkora százalékban mozdul el kedvező irányban az adott keresz. Ennek számításánál nem súlyoztam az esemény számmal, ezt megtéve alacsonyabb érték adódna, hiszen a kevésbé szigorú, több eseménnyel járó beállítások találati aránya alacsonyabb. Az átlagos elméleti éves hozam, pusztán egy elméleti érték: az átlagos eseményszám és az átlagos eseményenkénti változás szorzata elosztva a megfigyelési időszak hosszával (13,83 év).

5.1.1. Következtetések az esettanulmányok alapján

A GSMI esetében is megfigyelhető a volatilitás aszimmetria, azaz gyakrabban romlik el gyorsan a hangulat, mint ahányszor ugyanannyi idő alatt ugyanannyit javul. Minél szigorúbb feltételeket szabunk az eseményhez (azaz legalább mekkora GSMI változás legfeljebb hány nap alatt várunk el), annál nagyobb az események után várható változás és annál jobb az események árfolyammozgást előrejelző képessége, de természetesen az események száma is annál kevesebb.

Az USD esetében minden keresz esetén magasabb a nem veszteséges stratégiák aránya, sőt, szinte majdnem minden keresz esetében az eseményenkénti változás is alacsonyabb, azaz jelen esetben kedvezőbb. A Közép-Kelet Európai régiós devizák hasonlóan viselkednek a nem veszteséges stratégiák arányát tekintve, ez az arány az USD-vel szemben mindhárom esetben 90% fölötti, ami a legjobbnak tekinthető. A HUF, a PLN és a CZK az

eseményenkénti változást tekintve is előkelő helyezést érnek el. Igazán gyengén a RUB, az ARS és az INR teljesítettek, ennek számtalan oka lehet.

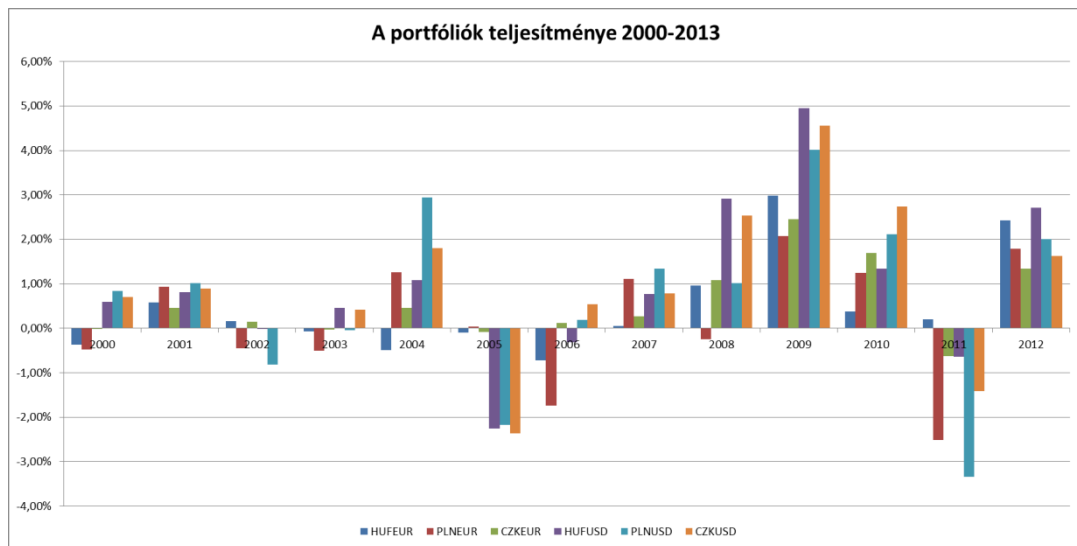
Összességében kijelenthetjük, hogy mérhető nagyságú és nem véletlenszerű kapcsolat van a GSMI és egyes devizakeresztek között. További finomításokat és korrekciókat téve végzem el a portfóliószimulációkat, melyek eredményeit gyakorlati szempontból relevánsabbnak tartok.

5.2. Portfóliószimulációk

Ebben a pontban egy kereskedési stratégiát mutatok be, mely minden szempontban identikus az előző pontban bemutatott esettanulmányokkal, azzal az egyetlen különbséggel, hogy csak akkor lehet pozíciót felvenni, ha az előzőt már lezártuk. Mindig az egész tőke befektetésre kerül egy pozícióba. A szimulációk eredményit itt csak a régiós devizákra (HUF, PLN, CZK) mutatom be. Az pozíció előjele – összhangban az esettanulmányok eredményeivel –korrekciót várva a devizák erősödésére játszik az EUR-ral és az US Dollárral szemben, azaz LONG pozíciót vesz fel negatív eseményt követően a HUFEUR, PLNEUR, CZKEUR, HUFUSD, PLNUSD és CZKUSD instrumentumokban. Tranzakciós költségeket nem vettem figyelembe és az éppen szabad, be nem fektetett tőke hozama nulla. A stratégia paraméterei az alábbi határok között változnak:

T_GSMI			req_chg_GSMI			T_FX		
min.:	max.:	lépésköz:	min.:	max.:	lépésköz:	min.:	max.:	lépésköz:
1	5	1	0,2	3	0,1	1	5	1

Mivel nem minden időpillanatban lehetséges pozíció felvétele, csak ha nincs éppen már pozíció felvéve, ezért befolyásolja az eredményeket, hogy pontosan mikor kezdjük a szimulációt. Emiatt (is) érdemes adott beállítás mellett több szimulációt futtatni. Hogy időben követni tudjuk a stratégiák teljesítményét, az egyes évekre külön kiszámítottam az elért átlagos hozamokat. Ennek eredményeit az alábbi, 11. ábra mutatja.



11. ábra A portfóliószimulációk átlagos eredményei az egyes években

A stratégiák minden instrumentum esetében közel azonos időszakokban működnek igazán jól, összességében azt mondhatjuk, hogy a vizsgált 13 évből mindössze 3 évben teljesítettek igen gyengén (2005,2006,2012) két évben közel nulla hozamot hoztak (2002, 2003) a többi 8 évben pedig jól teljesítettek. Fontos továbbra is hangsúlyozni, hogy az eredmények nem tartalmaznak tranzakciós költséget, azonban a másik oldalon az áll, hogy ugyan éves hozamokat közöltem, de a 13 év átlagában évente mindössze 11 pozíciót⁸¹ veszünk fel (amik átlagosan 3 napig vannak nyitva), a fennmaradó időben a tőke szabadon allokálható. A USD-vel szemben továbbra is jobbak az eredmények.

A 13 évre és a hat keresztre átlagosan 0,64% éves hozam adódik az említett átlagosan évenkénti 11 pozícióból. Az legjobbnak tekinthető CZKUSD instrumentum esetében az átlagos éves hozam éves hozam 0,99%. ehhez azonban átlagosan mindössze 2% max drawdown érték tartozik (egy év hosszúságú szimulációk), ami igen alacsonynak tekinthető, összhangban a veszteséggel zárt pozíciók viszonylag alacsony 25%⁸² körüli arányával. Tehát, ha amennyiben valaki megengedhet magának akár 30% max drawdown értéket, akkor 15-szörös tőkeáttétellel már átlagosan 15%-ot tudott volna elérni⁸³. Fontos még kiemelni, hogy ha szigorúbb követelményeket támasztunk a negatív esemény felé, akkor kevesebb esemény lesz, és az eseményenkénti hozam növekedik. Összességében a stratégiát inkább kiegészítő jelleggel ajánlanám, mert igazán jó eredményeket akkor tudunk elérni, ha inkább szigorú követelményeket támasztunk, ekkor viszont túl kevés lesz belőle egy kizárólagos stratégiához.

⁸¹ Ez minden kereszt esetében azonos, hiszen az esemény és a pozíció felvétele csak a GSMI-től függ, ami közös.

⁸² Nincs súlyozva az adott szimulációban felvett pozíciók számával, minden szimuláció súlya egységes.

⁸³ Továbbra is tranzakciós költségek nélkül.

Ráadásul a stratégia eseményvezérelt, így a tőke az idő túlnyomó részében szabad. A részletes eredmények a 8.1 mellékletben.

5.2.1. Monte-Carlo Portfóliószimulációk

Hibás következtetésekre vezetne, ha az alapján értékelnénk a stratégiát, hogy egy múltbeli időszakon a legjobb paraméterek mellett milyen (optimalizált) eredményt ért el. Emiatt egy olyan módszertant alkottam, ami kombinálja Monte Carlo szimulációk nyújtotta véletlenszerűségét és a mintán kívüli tesztelés objektivitását. (Leinweber & Sisk, 2010) egy „ténylegesen mintán kívüli” (real out of sample) elnevezésű módszertan mellett érvel, mint egyetlen filozófiai szempontból nem kifogásolható módszertan. Ennek lényege az, hogy egy megalkotott befektetési stratégiát a megalkotása után el kell tenni a fiókba, elzárni, és csak egy megfelelően hosszú idő elteltével elővenni. Ezután a közben eltelt időszakon – melyről természetesen semmilyen információval nem rendelkezhattünk a stratégia megalkotása idején – a stratégiát eredeti formájában tesztelni. Elfogadva ennek előnyeit, de amellet érvelve, hogy a mai világban nem lehet ennyit várni egy potenciálisan nyereséges stratégia alkalmazásával a gyors alfa eltűnés miatt javaslom az alábbi módszertant:

1. A lehető legtágabb paraméter beállítások mellett teszteljük a stratégiát egy véletlenszerű (nem a rendelkezésünkre álló teljes) X hosszúságú időszakon (tanítóidőszak)
2. A stratégiákat sorrendbe állítjuk valamilyen kritériumrendszer (elért eredményeik) alapján (optimalizálás a tanítóidőszakon)
3. A legjobb Y% stratégiát teszteljük az 1. pont időszaka után kezdődő Z hosszúságú időszakon. Ezeknek az itt elért eredményeit átlagoljuk (mintán kívüli tesztelés a tesztidőszakon)
4. Megfelelő N számú ismétlést hajtunk végre az 1.-3. pontokon
5. A kapott eredményeket az elvégzett N számú ismétlés fényében értékeljük

A bemutatott lépések annak a várható eredményét szimulálják, hogy várhatóan mekkora hozamot ér el egy befektető, aki az elmúlt időszakban valamikor, az akkor rendelkezésére álló időszak információi alapján optimalizálta a stratégiát és utána ezt Z ideig változatlan formában alkalmazta. A kapott minta tulajdonságai és a mintaelemszám alapján statisztikai precizitással becsülhetjük a stratégia várható hozamát.

Fontos továbbra is hangsúlyozni, hogy mindezek ellenére a kapott eredmény a múltra vonatkozik, még akkor is, ha objektívebb eredményt ad, mint egy stratégia valamilyen múltbeli időszak(ok)on elért átlagos eredménye. Általában a szakirodalomban hiányolom az egységes módszertant a különböző befektetési stratégiák értékelése kapcsán, többször felmerül valamilyen szándékos vagy szándékolatlan torzítása az eredményeknek, aminek

valóság alapja utólag nehezen ellenőrizhető. A 2. Pontban említett kritériumra az éves hozamot fogom alkalmazni, hiszen a legtöbb befektető ez alapján dönt, de figyelemmel kísérhetnénk a maximális drawdown értékét is. Természetesen az általam használt módszertan is torzít valamilyen mértékben például az X, Y és Z számok önkényes megválasztása⁸⁴. X és Z tanító és tesztidőszakokra 1-1 évet állítottam be, illetve az egyszerűség kedvéért mindig a legmagasabb hozamot elért stratégiát választottam a tesztidőszakra. 200 ismétlést végeztem mindegyik instrumentumra, az eredményeket az alábbi táblázat foglalja össze:

		T_GSMI	req_chg_GSMI	T_fx	Teljes hozam	Felvett pozíciók száma	Max drawdown
HUFEUR	átlag	2,47	0,40	6,14	1,07%	30,2	-5,78%
	szórás	1,45	0,40	2,17	5,93%	14,0	5,27%
PLNEUR	átlag	2,27	0,42	6,36	0,14%	31,4	-6,25%
	szórás	1,41	0,51	2,75	5,85%	22,3	5,63%
CZKEUR	átlag	2,62	0,36	6,57	1,09%	30,6	-3,74%
	szórás	1,25	0,26	2,63	3,70%	15,4	3,61%
HUFUSD	átlag	3,07	0,40	6,13	2,56%	30,2	-10,19%
	szórás	1,51	0,38	2,02	8,92%	15,9	9,77%
PLNUSD	átlag	3,36	0,52	7,56	2,42%	22,4	-8,89%
	szórás	1,46	0,49	1,99	9,82%	11,0	9,55%
CZKUSD	átlag	3,25	0,40	7,01	2,99%	28,0	-7,75%
	szórás	1,46	0,37	2,35	8,89%	13,3	6,68%
HUFEUR PLNEUR CZKEUR	átlagos érték	2,45	0,39	6,35	0,77%	30,7	-5,26%
	átlagos szórás	1,37	0,39	2,52	5,16%	17,2	4,84%
HUFUSD PLNUSD CZKUSD	átlagos érték	3,22	0,44	6,90	2,65%	26,8	-8,94%
	átlagos szórás	1,48	0,41	2,12	9,21%	13,4	8,66%
ÖSSZES	átlagos érték	2,84	0,42	6,63	1,71%	28,8	-7,10%
	átlagos szórás	1,42	0,40	2,32	7,19%	15,3	6,75%

5. táblázat: A Monte Carlo portfóliószimulációk eredményei

5.2.2. Következtetések a portfóliószimulációk alapján

A legjobb eredményekkel jelen esetben is a CZKUSD rendelkezik, ezen az instrumentumon kereskedve átlagosan 3% hozamot ért volna el a stratégia -7,75%-os maximális drawdown mellett. A legeredményesebb paraméter beállítás a tanítóhalmazon átlagosan a T_GSMI, req_chg_GSMI és T_fx rendre a 3,25 0,4 és 7 értékei voltak, ez azonban ezeknek a paramétereknek a külön – külön számított átlaga nem pedig a „tipikusan legjobb paraméterkombináció”. Összességében elmondhatjuk, hogy itt az US Dollárral szembeni

⁸⁴ Ha valaki rendelkezik a számítási kapacitással, akkor ennek a három további paraméternek mindenféle elfogadható kombinációjára is elvégezheti a szimulációkat.

keresztek sokkal jobban teljesítettek, hozamot tekintve több mint háromszor olyan jól. Ez, ha mást nem, a módszertanból fakadóan azt jelenti, hogy ezen keresztek árfolyam tulajdonságai jobban öröklődnek egyik időszakról az azt követőre, azaz stabilabbak. Úgy érzem, hogy szignifikáns nagyságú és előrejelelhető irányú kapcsolatot találtunk a szentiment és a devizaárfolyamok között. Ezt még egy ilyen egyszerű stratégiával is be tudtam mutatni, még napi adatokon is. Gyakorlati kereskedési szempontból finomhangolással és további faktorok hozzáadásával egy sikeres stratégia konstruálható, amennyiben bölcsen választjuk meg a kereskedett instrumentumot. További Monte Carlo szimulációkat is elvégeztem más hosszúságú tanító és tesztidőszakokra (X és Z paraméterek), de egyértelmű trendeket nem azonosítottam.

6. Összefoglalás és végkövetkeztetések

A szentiment talán az irracionalitás mérőszáma, André Kosztolányi szerint rövid távon a piac 90% pszichológia és csak 10% racionalitás. Ennek ellenére a legtöbb hagyományos pénzügyi modell nem foglalkozik a szentimenttel mint tényezővel (sem rövid sem hosszú távon). A szentiment betörése a mainstream közgazdasági gondolkodásba mindmáig várat magára, egyáltalán, az olyan számszerűsített gondolkodási keret, mely az emberi irracionalitási tényezőt egyáltalán figyelembe veszi, csak az előző század közepén jelent meg.

A strukturálatlan szövegek tartalmi és szentimentelemzésének lassú megjelenéséért eddig a nehéz számszerűsíthetőséget és a technológiai korlátokat okolhatjuk, a technológia még nem kiforrott, nincs „gold standardje”. Mostanra azonban lehetőség nyílt pontos és az eddiginél jóval szélesebb körű tartalom automatizált kiértékelésére, melyekkel kapcsolatos tanulságokat az alábbiakban foglalhatjuk össze.

Igény van az információ új forrásaira, mert mára a hagyományos faktorok kereskedési hozadéka az elterjedésükkel párhuzamosan jelentősen visszaesett. A piaci szereplők többsége ugyanazokat használja, ráadásul hasonlóképpen. A szentiment hatása zajos, optimális esetben egy kizárólag a hírekre épülő stratégiához komoly infrastruktúrára van szükség, gyors reagálásra (például High Frequency Trading) és alacsony tranzakciós költségekre. Ez belépési korlátként egyben magasan tartja a kereskedés extrahozamát. Lehetséges azonban a szentiment jóval lassabb, hosszabb távon történő kihasználása és beépítése is. Erre láttunk 3+1 konkrét példát is a dolgozatban. Nagy mennyiségű tartalom pontos kiértékelése alapján lehet az átlagnál szisztematikusan jobb kereskedési döntéseket hozni. Már meglévő stratégiák hatékonyságát könnyen javíthatjuk a szentiment mint faktor felvételével.

A hír- és szentiment stratégiák a különböző momentumstratégiákhoz állnak a legközelebb. Ez részben nem is meglepő a hírek és az árfolyam kölcsönös kapcsolata miatt, de azt is megmutatták, hogy a meglévő momentumstratégiákat is képesek vagyunk javítani, illetve egy szentimentindex tipikusan felülmúlja a momentumot mint jelet, azonos felhasználás mellett pl. (Hafez, 2009b).

Kevés alap kereskedik kizárólag szentiment alapon, legalábbis bevallottan, és ezeknek a története sem éppen fényes (lásd 4. fejezet), ezzel szemben rengeteg szentimentindex létezik. Ez a két tény összességében a szentimentindexek másodlagos indikátor jellegére utalhat. A szentimentnek legerősebb empirikus kapcsolata úgy tűnik a volatilitással van, de a volatilitás autoregresszív volta nehezíti a szentiment felhasználhatóságát a volatilitás előrejelzésében. A szentiment a forgalommal is erősen korrelál, de a forgalom - volatilitás szükségszerű

kapcsolata miatt ez alapvetően kevésbé meglepő. Szignifikáns kapcsolatot a hozamokkal is kimutattak (illetve kimutattam), de az irányra és nagyságra nincs általános képlet, ami nehezíti egy stratégia megalkotását.

A hírek újdonságérték szerinti szűrése kritikus pontnak tűnik a szentimentelemzés esetében. Csak adott eseményt elsőnek vagy elsők között feldolgozó híreket van értelme figyelembe venni, ezt követően ugyanis már csak a körbehivatkozások és újrairatok keringenek, amik eltorzíthatják a képet, és másrészt időben késve jelennek meg. Erre hívja fel a figyelmet többek között (Tetlock, et al., 2008) és (Mitra & Mitra , 2011).

Fontos megjegyezni azonban, hogy a forgalom, volatilitás, hozamok iránya és a szentiment kapcsolatrendszerben a hatások iránya tisztázatlan, a kapcsolatrendszer kölcsönös, visszacsatolós. A szakirodalom alapján szentiment főleg kisbefektetőkre, illetve a termékoldalon a kicsi, fiatal cégekre van hatással, de ezen utóbbiak eddig is közismerten magasabb volatilitása csökkenti a felismerés jelentőségét a volatilitás-szentiment kapcsolat miatt. Ha túl sok az alternatív befektetési lehetőség és magasak a keresési költségek, akkor a média-figyelem könnyen felülkerekedhet a racionális értékítéleten és befolyásolhatja a döntést. Ebből a szempontból optimális terepnek tűnik egy újonnan kiépülő szétaprózódott tőkepiac új szereplőkkel és belépőkkel, általánosabban, egy információs hatékonyság szempontjából fejletlenebb piac. Az automatizált tartalom és szentimentelemzés újszerűsége ellenére az extraprofit egy részét talán már lefölozték, de kevésbé „divatos” tőzsdék kevésbé divatos papírai még sok lehetőséget tartogathatnak. Érdeemes kilépni az angol nyelv és a nyelvterületein lévő piacok korlátai közül, hiszen mint utaltam rá, egyik szépsége, hogy alkalmazásához nem is feltétlenül szükséges érteni az adott nyelven.

További felhasználása lehet az automatizált tartalom és szentimentelemzésnek például egy vállalat kapcsolódó híreinek visszamenőleges begyűjtése és elemzése. Ezáltal gyorsan mélységi képet kaphatunk a cégspecifikus szentiment alakulásáról és a vállalat történetéről. Egy másik lehetőség szentimenthez kötődő riasztások alkalmazása. Ez lehet akár egy kis vállalatról megjelenő negatív hír vagy akár blogbejegyzés megjelenése a világhálón. A technológia lehetővé teszi az ilyen események gyors észlelését.

7. Irodalomjegyzék

- Abbasi, A. & Chen, H., 2008. Writeprints: A Stylometric Approach to Identify-level Identification and Similarity Detection in Cyberspace. *ACM Transactions on Information Systems*, p. 26(2).
- Abraham, A. & Taylor, W., 1993. Pricing currency options with scheduled and unscheduled announcements effects on volatility. *Managerial and Decision Science*, pp. 14, 311-326.
- Baker, M. & Wurgler, J., 2007. Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, American Economic Association*, pp. vol. 21(2), pages 129-152.
- Barber, B. & Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, pp. 21, 785–818..
- Barberis, N. –. H. M. –. B. M. J., 2001. Mental Accounting, Loss Aversion, and Individual Stock Returns. *The Journal of Finance*, pp. 56(4), pp. 1247-1292..
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, pp. 31 307-327.
- Cahan, R., Jussa, J. & Lou, Y., 2009. Breaking News: How to Use News Sentiment to Pick Stocks. *Macquarie US Equity Research*.
- Cahan, R., Lou, Y., Jussa, J. & Alvarez, M., 2010. Beyond the Headlines: Using News Flow to Predict Stock Return. *Deutsche Bank Quantitative Strategy*, p. July 10.
- Cecchini, M., Aytug, H., Koehler, G. J. & Pathak, P., 2010. Making words work: Using financial text as a predictor of financial events. *Decision Support Systems*, pp. Volume 50, Issue 1, pp. 164–175.
- Chemmanur, T. & Yan, A., 2009. Advertising, Attention, and Stock Returns. *Working Paper, Boston College and Fordham University*.
- Cont, R., 2005. *Volatility Clustering in Financial Markets: Empirical Facts and Agent Based Models*. to appear in: A. Kirman & G Teyssiere (eds.): *Long memory in economics*,: Springer.
- Das, S., 2010. The Finance Web: Internet Information and Markets. *IEEE Intelligent Systems*, p. 25(2).
- Das, S. & Chen, M., 2007. “Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web,”. *Management Science*, pp. vol. 53, no. 9, pp. 1375–1388..
- Dong, L. & , 2008. Attracting Investor Attention through Advertising. *Working Paper, Yale University*.
- Dzielinski, M., Oliver, M. & Talpsepp, T., 2011. Volatility asymmetry, news, and private investors. In: L. M. Gautam Mitra, szerk. *Handbook of news analytics in finance*. hely nélkül.:Wiley and Sons Ltd., pp. 255-270.
- Easley, D. & O'Hara, M., 2001. Information and the Cost of Capital. *Working Paper, Cornell University*.
- Ederington , E. & Lee, J., 1996. Creation and resolution of market uncertainty: The importance of information releases. *Journal of Financial and Quantitative Analysis*, pp. 31, 513–539.
- Fama, E., 1964. The Behavior of Stock Market Prices. *The Journal of Business*, pp. 38(1): 34-106..

- Fang, L. & Peress, J., 2009. Media coverage and the cross-section of stock returns. *Journal of Finance*, pp. 64 (5) 2023-2051.
- French, K. R. & Roll, R., 1986. Stock Return Variances: The Arrival of information and the reaction of traders. *Journal of Financial Economics*, pp. 17, 5-26.
- Gidofalvi, G., 2001. Using News Articles to Predict Stock Price Movements. *Department of Computer Science and Engineering, University of California, San Diego.*
- Gidofalvi, G. & Elkan, C., 2003. Using News Articles to Predict Stock Price Movements. *Technical Report. Department of Computer Science and Engineering, University of California, San Diego.*
- Gross-Klussman, A. & Hautsch, N., 2011. When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance* , pp. 18 321–340 (a tanulmány 2009-es).
- Grullon, G., Kanatas, G. & Weston, J., 2004. Advertising, breadth of ownership, and liquidity. *Review of Financial Studies*, pp. 17, 439-461.
- Hafez, P., 2009a. Investigation of the Impact of News Sentiment on Abnormal Stock Return, RavenPack Intrnational S.L.. *RavenPack International S.L.*
- Hafez, P., 2009b. Construction of Market Sentiment Indices Using News Sentiment. *RavenPack International S.L.*
- Hafez, P., 2010. How news events impact market sentiment. *RavenPack International S.L.*
- He, W. & Mian, G. M., 2007. *Market Sentiment, Investor Size and Reaction to Firm-Specific News* , hely nélk.: Srinivasan Sankaraguruswamy NUS Business Shool.
- Hirscheifer, D., Myers, J., Myers, L. & Siew, H. T., 2003. Do Individual Investors Drive Post-earnings Announcement Drift? Direct Evidence from Personal Trades. *Working Paper*, p. Available at <http://ssrn.com/abstract=299260>.
- Ho, K.-Y., Shi, Y. & Zhang, Z., 2013. How does news sentiment impact asset volatility? Evidence from long memory and regime-switching approaches. *The North American Journal of Economics and Finance*.
- Kittrel, J., 2010. Sentiment reversals as buy signals. *Knightsbridge Asset Management LLC*, pp. elérhető Gautam Mitra és Leela Mitra (szerkesztők) *Hanbook of News Analytics in Finance 2011*, 9.fejezet.
- Koppel, M. & Shtirtemberg, I., 2004. Good news or bad news? let the market decide. *In AAI Spring Symposium on Exploring Attitude and Affect in Text, Stanford University*, pp. pp. 86-88.
- Kucuntunc & et al., 2012. *A large-scale sentiment analysis for Yahoo! Answers*. Proceedings of the 5th ACM International Conference on Weg Search and Data Mining pp. 633-642: ismeretlen szerző
- Kurov, A., 2008. Investor Sentiment, Trading Behavior and Informational Efficiency in Index Futures Markets. *The Financial Review*, pp. 43 pp. 107-127.
- Kwag, A., Shrieves, R. & Wansley, J., 2000. Partially Anticipated Events: An Application to Dividend Announcements. *Working Paper, University of Tennessee*.
- Lavrenko, V., Schmill, M. & et. al., 2000. Language Models for Financial News Recommendation. *Proceedings of the ninth international conference on Information and knowledge management*.

- Lee, C., 1992. Earnings news and small traders. *Journal of Accounting and Economics*, pp. 15, 265-302.
- Lee, C., Schleifer, A. & Thaler, R., 1991. Investor Sentiment and the Closed-End. *Journal of Finance*.
- Leinweber, D. & Sisk, J., 2011. Event Driven Trading and the 'New News'. *Journal of Portfolio Management*, pp. Vol. 38, No. 1, .
- Leinweber, D. & Sisk, J., 2010. Relating news analytics to stock returns. *The interface of behavioural finance and quantitative finance, conference, 2-3 February 2010, MWB Canada Square, Canary Wharf, London*.
- Markowitz, H., 1952. Portfolio Selection. *The Journal of Finance*, p. 7 (1): 77–91.
- Ma, Z., Sheng, O. R. & Pant, G., 2010. Discovering company revenue relations from news: A network approach. *Decision Support Systems*, pp. Volume 47, Issue 4, Pages 408-414.
- Mitra, G. & Mitra, L., 2011. *The handbook of news analytics in finance*. ISBN 978-0-470-66679-1 szerk. hely nélk.:Wiley.
- Mitra, L., Mitra, G. & diBartolomeo, G., 2008. Equity portfolio risk (volatility) estimation using market information and sentiment. *December 1. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1425624*.
- Mittermayer, M., 2004. *Forecasting Intraday Stock Price Trends with Text Mining Techniques*, hely nélk.: Hawaii International Conference on System Sciences, Kailua-Kona, HI.
- Molnár, M. A., 2006. A magyar tőkepiac vizsgálata pénzügyi viselkedéstani modellel. *Doktori értekezés, Budapesti Corvinus Egyetem, Gazdálkodástani Doktori Iskola*.
- Moniz, A., Davies, C., Brar, G. & Strudwick, A., 2011. The impact of news flow on asset returns: An empirical study. *Handbook of news analytics in finance 8. fejezet*.
- Niederhoffer, V., 1971. The analysis of world events and stock prices. *Journal of Business*, pp. 44, 193-219.
- Pang, B., Lee, L. & Vaithyanathan, S., 2002. Thumbs up? Sentiment classification using machine. *paper presented at Proceedings Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Romer, D., 1993. Rational asset price movement without news. *American Economic Review*, pp. 83(5) 1112-1130.
- Ross, S., 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, pp. 13 (3), 341–360.
- Seasholes, M. & Wu, G., 2004. Profiting from Predictability: Smart Traders, Daily Price Limits, and Investor Attention. *Working Paper*, p. Available at <http://ssrn.com/abstract=527182>.
- Ser-Huang, P. & Granger, C., 2005. Practical issues in forecasting volatility. *Financial Analyst Journal*, p. 61(1) 45ff.
- Shah, A., 2008. Short-term risk from long-term models. *paper presented at Northfield Seminar Proceedings, October*.
- Sprenger, T. & Welpe, I., 2010. Tweets and Trades: The Information Content of Stock Microblogs. *working paper*, pp. Technische Universität München, TUM School of Management, Chair for Strategy and Organization.

Tetlock, P. C., 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance* , pp. Volume 62, Issue 3, pages 1139–1168.

Tetlock, P., Saar-Tsechansky, M. & Macskassy, S., 2008. More than words: Quantifying language (in news) to measure firms' fundamentals. *Journal of Finance*, pp. 63, 1437–1467.

Thaler, R. H., 1988. Anomalies: The Winner's Curse. *The Journal of Economic Perspectives*, pp. Vol. 2, No. 1 (Winter, 1988), pp. 191-202.

Zhang, W. & Skiena, S., 2009. Trading Strategies To Exploit Blog and News Sentiment. *Department of Computer Science, Stony Brook University NY 11794-4400 USA*.

8. Mellékletek

8.1. Első melléklet: Portfóliószimuláció eredmények

	HUFEUR			PLNEUR			CZKEUR			HUFUSD			Eves hozam
	Max Drawdown	Veszteséges pozíciók aránya	Eves hozam	Max Drawdown	Veszteséges pozíciók aránya	Eves hozam	Max Drawdown	Veszteséges pozíciók aránya	Eves hozam	Max Drawdown	Veszteséges pozíciók aránya	Eves hozam	
37%	-0,41%	25,59%	-0,48%	-1,74%	21,07%	-0,02%	-0,54%	19,87%	0,59%	12,52%	-1,21%	0,84%	
39%	-0,66%	13,40%	0,93%	-1,29%	9,46%	0,45%	-0,41%	12,48%	0,82%	18,22%	-0,92%	1,02%	
6%	-0,18%	7,56%	-0,45%	-0,74%	13,71%	0,15%	-0,23%	11,55%	-0,01%	13,42%	-0,50%	-0,82%	
7%	-0,24%	16,84%	-0,50%	-0,62%	15,88%	-0,03%	-0,23%	14,26%	0,46%	5,81%	-0,31%	-0,04%	
48%	-0,99%	23,21%	1,26%	-0,74%	18,01%	0,46%	-0,52%	19,16%	1,08%	18,57%	-1,41%	2,94%	
10%	-0,84%	32,07%	0,04%	-1,65%	24,14%	-0,08%	-0,96%	37,37%	-2,25%	39,67%	-2,78%	-2,17%	
72%	-2,25%	36,31%	-1,74%	-2,55%	28,97%	0,13%	-0,78%	31,70%	-0,31%	38,25%	-2,31%	0,19%	
15%	-1,92%	34,11%	1,11%	-0,67%	26,27%	0,26%	-0,81%	28,20%	0,78%	26,30%	-1,94%	1,33%	
16%	-4,72%	34,49%	-0,24%	-3,92%	33,78%	1,09%	-2,88%	34,31%	2,92%	33,15%	-8,50%	1,01%	
18%	-2,48%	26,89%	2,07%	-3,13%	27,48%	2,46%	-1,87%	22,32%	4,95%	24,95%	-2,96%	4,01%	
18%	-1,78%	44,08%	1,25%	-1,71%	32,80%	1,70%	-0,77%	24,46%	1,34%	44,47%	-3,59%	2,12%	
10%	-2,64%	40,42%	-2,51%	-3,91%	42,95%	-0,63%	-1,97%	51,44%	-0,65%	32,99%	-5,60%	-3,34%	
12%	-1,59%	29,60%	1,79%	-1,29%	17,87%	1,34%	-0,86%	19,67%	2,71%	29,62%	-2,97%	2,01%	
16%	-1,59%	28,04%	0,19%	-1,84%	24,03%	0,56%	-0,99%	25,14%	0,96%	25,99%	-2,69%	0,70%	

6. táblázat: A portfóliószimulációk eredményei

8.2. Második melléklet: A saját stratégia kódjai

A mellékelt kódok egy makróbarát Excel munkafüzetből a benne megadott instrukciókat betartva futtathatók. A változók deklarálása utáni szakaszban szükséges megadni a paramétereket.

Esettanulmányok

Sub casestudy()

'az első munkalapon az első három oszlop rendre a dátumok, GSMI és az adott deviza loghozamai kell, hogy legyenek 'oly módon, hogy a fejléc az első sorban, majd az adatok folyamatosan a második sortól kezdődnek
'az utolsó adatsor sorszámát meg kell adni a last_observation_row változó értékének a makróban

Dim chg_GSMI, T_GSMI, req_chg_GSMI As Variant
Dim chg_fx, T_fx As Variant
Dim eventcount As Integer

Dim T_GSMI_low, req_chg_GSMI_low, T_fx_low As Variant
Dim T_GSMI_high, req_chg_GSMI_high, T_fx_high As Variant '
Dim T_GSMI_stepsize, req_chg_GSMI_stepsize, T_fx_stepsize As Variant
Dim chg_fx_for_this_event, chg_fx_pos, chg_fx_pos_counter, chg_fx_neg, chg_fx_neg_counter As Variant

Dim i, j, k, n As Variant
Dim first_observation_row, last_observation_row As Integer

n = 2

'a paraméterek vizsgálati tartománya:
T_GSMI_low = 1

```

T_GSMI_high = 5
T_GSMI_stepsize = 1
req_chg_GSMI_low = 0.1
req_chg_GSMI_high = 3.1
req_chg_GSMI_stepsize = 0.1
T_fx_low = 1
T_fx_high = 5
T_fx_stepsize = 1
first_observation_row = 2
last_observation_row = 5026

Worksheets(1).Activate

'Csak negatív eseményeket keresünk, long pozíciót veszünk fel

For T_GSMI = T_GSMI_low To T_GSMI_high Step T_GSMI_stepsize
  For T_fx = T_fx_low To T_fx_high Step T_fx_stepsize
    For req_chg_GSMI = req_chg_GSMI_low To req_chg_GSMI_high Step req_chg_GSMI_stepsize

      eventcount = 0
      chg_fx = 0
      chg_fx_pos = 0
      chg_fx_pos_counter = 0
      chg_fx_neg = 0
      chg_fx_neg_counter = 0

      For i = first_observation_row + T_GSMI To last_observation_row - T_fx
        chg_GSMI = 0

        For j = i - T_GSMI + 1 To i
          chg_GSMI = chg_GSMI + Cells(j, 2)
        Next j

        If -1 * chg_GSMI > req_chg_GSMI Then
          eventcount = eventcount + 1

          chg_fx_for_this_event = 0
          For k = i + 1 To i + T_fx
            chg_fx = chg_fx + Cells(k, 3)
            chg_fx_for_this_event = chg_fx_for_this_event + Cells(k, 3)
          Next k

          If chg_fx_for_this_event > 0 Then
            chg_fx_pos = chg_fx_pos + chg_fx_for_this_event
            chg_fx_pos_counter = chg_fx_pos_counter + 1
          Else
            If chg_fx_for_this_event < 0 Then
              chg_fx_neg = chg_fx_neg + chg_fx_for_this_event
              chg_fx_neg_counter = chg_fx_neg_counter + 1
            End If
          End If

        Else
          End If
        Next i

        'eredmények kiírása a 2. munkalapra (a második sortól kezdődően):
        Worksheets(2).Cells(n, 1) = Worksheets(1).Cells(1, 2).Value
        Worksheets(2).Cells(n, 2) = Worksheets(1).Cells(1, 3).Value
        Worksheets(2).Cells(n, 3) = Worksheets(1).Cells(2, 1).Value
        Worksheets(2).Cells(n, 4) = Worksheets(1).Cells(last_observation_row, 1).Value
        Worksheets(2).Cells(n, 5) = T_GSMI
        Worksheets(2).Cells(n, 6) = req_chg_GSMI
        Worksheets(2).Cells(n, 7) = T_fx
        Worksheets(2).Cells(n, 8) = eventcount 'összes esemény száma
        Worksheets(2).Cells(n, 9) = chg_fx / (eventcount + 1E-42) 'átlagos hozam eseményenként
        Worksheets(2).Cells(n, 10) = (eventcount - chg_fx_pos_counter) / (eventcount + 1E-42) 'átlagos találati arány

        n = n + 1

      Next req_chg_GSMI
    Next T_fx
  Next T_GSMI

'fejlécek elnevezése:
Worksheets(2).Cells(1, 1) = "GSMI"
Worksheets(2).Cells(1, 2) = "A vizsgált devizakereszt"
Worksheets(2).Cells(1, 3) = "Kezdő dátum"
Worksheets(2).Cells(1, 4) = "Utolsó dátum"
Worksheets(2).Cells(1, 5) = "T_GSMI"
Worksheets(2).Cells(1, 6) = "req_chg_GSMI"
Worksheets(2).Cells(1, 7) = "T_FX"

Worksheets(2).Cells(1, 8) = "Negatív események száma"
Worksheets(2).Cells(1, 9) = "Átlagos hozam negatív eseményenként"
Worksheets(2).Cells(1, 10) = "Nempozitív hozam a tartás alatt aránya (long esetén)"

End Sub

```


Portfóliószimulációk

Sub portfolsimu()

'az első munkalapon az első három oszlop rendre a dátum, GSMI és az adott deviza loghozamai kell, hogy legyenek

'oly módon, hogy a fejléc az első sorban, majd az adatok folyamatosan a második sortól kezdődnek

'az utolsó adatsor sorszámát meg kell adni a last_observation_row változó értékének a makróban

'csak a negatív eseményekre veszünk fel pozíciót

'FONTOS: az alábbi makró LONG pozíciót vesz fel (ez egy előjel kérdése), jó eredményeket emiatt például nem az EURHUF-on, hanem a HUFEUR-on ér el

Dim T_GSMI, req_chg_GSMI, T_fx As Variant

Dim eventcount, positioncount, neg_pos_counter As Integer

Dim return_total, return_for_this_position, return_yearly, return_total_last_peak, drawdown, drawdown_max, chg_GSMI As Variant

Dim T_GSMI_low, req_chg_GSMI_low, T_fx_low As Variant

Dim T_GSMI_high, req_chg_GSMI_high, T_fx_high As Variant

Dim T_GSMI_stepsize, req_chg_GSMI_stepsize, T_fx_stepsize As Variant

Dim period_length, shift_max As Variant

Dim i, j, n, m As Integer

Dim first_observation_row, last_observation_row, shift As Integer

n = 2

T_GSMI_low = 1

T_GSMI_high = 5

T_GSMI_stepsize = 1

req_chg_GSMI_low = 0.2

req_chg_GSMI_high = 3.1

req_chg_GSMI_stepsize = 0.1

T_fx_low = 1

T_fx_high = 5

T_fx_stepsize = 1

first_observation_row = 2

last_observation_row = 5026

period_length = 365

shift_max = 0

Worksheets(1).Activate

For first_observation_row = 2 To last_observation_row - period_length Step period_length

For T_GSMI = T_GSMI_low To T_GSMI_high Step T_GSMI_stepsize

For T_fx = T_fx_low To T_fx_high Step T_fx_stepsize

For req_chg_GSMI = req_chg_GSMI_low To req_chg_GSMI_high Step req_chg_GSMI_stepsize

For shift = 0 To shift_max

eventcount = 0

positioncount = 0

return_total = 0

neg_pos_counter = 0

i = first_observation_row + T_GSMI + shift

drawdown = 0

drawdown_max = 0

return_total_last_peak = 0

Do

chg_GSMI = 0

For j = i - T_GSMI + 1 To i

chg_GSMI = chg_GSMI + Cells(j, 2)

Next j

'ha esemény van, felvesszük a pozíciót

If -1 * chg_GSMI > req_chg_GSMI Then

positioncount = positioncount + 1

return_for_this_position = 0

'a hozam mérése az adott pozícióra, amíg le nem zárjuk

For i = i + 1 To i + T_fx

return_for_this_position = return_for_this_position + Cells(i, 3)

Next i

'a negatív hozamú pozíciók mérése

If return_for_this_position < 0 Then

neg_pos_counter = neg_pos_counter + 1

Else

End If

return_total = return_total + return_for_this_position

'max drawdown számítás

If return_total > return_total_last_peak Then

return_total_last_peak = return_total

Else

drawdown = return_total - return_total_last_peak

If drawdown < drawdown_max Then

drawdown_max = drawdown

End If

```

End If

Else
i = i + 1
End If
Loop While i <= first_observation_row + period_length - T_fx

'kizárólag az események számolását végző blokk - nem a felvett pozíciókét!
For i = T_GSMI + first_observation_row To first_observation_row + period_length
chg_GSMI = 0
For j = i - T_GSMI + 1 To i
chg_GSMI = chg_GSMI + Cells(j, 2)
Next j
If -1 * chg_GSMI > req_chg_GSMI Then
eventcount = eventcount + 1
End If
Next i

'eredmények kiírása a 2. munkalapra
Worksheets(2).Cells(n, 1) = Worksheets(1).Cells(1, 2).Value
Worksheets(2).Cells(n, 2) = Worksheets(1).Cells(1, 3).Value
Worksheets(2).Cells(n, 3) = Worksheets(1).Cells(first_observation_row, 1).Value
Worksheets(2).Cells(n, 4) = Worksheets(1).Cells(first_observation_row + period_length, 1).Value
Worksheets(2).Cells(n, 5) = T_GSMI
Worksheets(2).Cells(n, 6) = req_chg_GSMI
Worksheets(2).Cells(n, 7) = T_fx

Worksheets(2).Cells(n, 8) = eventcount
Worksheets(2).Cells(n, 9) = positioncount
Worksheets(2).Cells(n, 10) = return_total
Worksheets(2).Cells(n, 11) = return_total / (positioncount + 1E-36)
Worksheets(2).Cells(n, 12) = return_total / (period_length / 365)
Worksheets(2).Cells(n, 13) = neg_pos_counter / (positioncount + 1E-36)
Worksheets(2).Cells(n, 14) = drawdown_max

n = n + 1

Next shift
Next req_chg_GSMI
Next T_fx
Next T_GSMI
Next first_observation_row

'fejlécek elnevezése:
Worksheets(2).Cells(1, 1) = "GSMI"
Worksheets(2).Cells(1, 2) = "A vizsgált devizakereszt"
Worksheets(2).Cells(1, 3) = "Kezdő dátum"
Worksheets(2).Cells(1, 4) = "Utolsó dátum"
Worksheets(2).Cells(1, 5) = "T_GSMI"
Worksheets(2).Cells(1, 6) = "req_chg_GSMI"
Worksheets(2).Cells(1, 7) = "T_fx"

Worksheets(2).Cells(1, 8) = "Események száma"
Worksheets(2).Cells(1, 9) = "Felvett pozíciók száma"
Worksheets(2).Cells(1, 10) = "Teljes hozam"
Worksheets(2).Cells(1, 11) = "Átlagos hozam / felvett pozíció"
Worksheets(2).Cells(1, 12) = "Éves hozam"
Worksheets(2).Cells(1, 13) = "Veszteséges pozíciók aránya"
Worksheets(2).Cells(1, 14) = "Max drawdown"

End Sub

```

Monte Carlo Portfóliószimulációk

Sub MCportfolsimu()

'az első munkalapon az első három oszlop rendre a dátum, GSMI és az adott deviza loghozamai kell, hogy legyenek

'oly módon, hogy a fejléc az első sorban, majd az adatok folyamatosan a második sortól kezdődnek

'az utolsó adatsor sorszámát meg kell adni a last_observation_row változó értékének a makróban

'csak a negatív eseményekre veszünk fel pozíciót

'FONTOS: az alábbi makró LONG pozíciót vesz fel (ez egy előjel kérdése), jó eredményeket emiatt például nem az EURHUF-on, hanem a HUFEUR-on ér el

Dim T_GSMI, req_chg_GSMI, T_fx As Variant

Dim return_total, return_for_this_position, chg_GSMI As Variant

Dim T_GSMI_low, req_chg_GSMI_low, T_fx_low As Variant

Dim T_GSMI_high, req_chg_GSMI_high, T_fx_high As Variant

Dim T_GSMI_stepsize, req_chg_GSMI_stepsize, T_fx_stepsize As Variant

Dim i, j, h As Integer

Dim first_observation_row, last_observation_row As Integer

Dim rep, train_period_length, oos_period_length, MC_number, MC_startpoint As Variant

Dim T_GSMI_best, req_chg_GSMI_best, T_fx_best, return_total_best As Variant

Dim positioncount, drawdown, drawdown_max, return_total_last_peak As Variant

h = 2

T_GSMI_low = 1

T_GSMI_high = 5

T_GSMI_stepsize = 1

req_chg_GSMI_low = 0.1

req_chg_GSMI_high = 3.1

req_chg_GSMI_stepsize = 0.1

T_fx_low = 1

T_fx_high = 10

T_fx_stepsize = 1

first_observation_row = 2 'a rendelkezésünkre álló teljes időablak első sora

last_observation_row = 5026 'a rendelkezésünkre álló teljes időablak utolsó sora

MC_number = 200 'az MC szimulációk száma

train_period_length = 365

oos_period_length = 365

For rep = 1 To MC_number

Worksheets(1).Activate

'annak a beállítás, hogy adott train_period_length és oos_period_length mellett milyen intervallumon lehet kiválasztani az MC szimu kezdetét

MC_startpoint = Int((last_observation_row - train_period_length - oos_period_length - first_observation_row + 1) * Rnd + first_observation_row)

return_total_best = -9999 'a kiinduló értéknek jó alacsonynak kell lenni, hogy mindenképp legyen nála jobb

For T_GSMI = T_GSMI_low To T_GSMI_high Step T_GSMI_stepsize

For T_fx = T_fx_low To T_fx_high Step T_fx_stepsize

For req_chg_GSMI = req_chg_GSMI_low To req_chg_GSMI_high Step req_chg_GSMI_stepsize

return_total = 0

i = MC_startpoint + T_GSMI

Do

chg_GSMI = 0

For j = i - T_GSMI + 1 To i

chg_GSMI = chg_GSMI + Cells(j, 2)

Next j

'ha esemény van, felvesszük a pozíciót

If -1 * chg_GSMI > req_chg_GSMI Then

return_for_this_position = 0

'a hozam mérése az adott pozícióra, amíg le nem zárjuk

For i = i + 1 To i + T_fx

return_for_this_position = return_for_this_position + Cells(i, 3)

Next i

return_total = return_total + return_for_this_position

Else

i = i + 1

End If

Loop While i <= MC_startpoint + train_period_length - T_fx

'legjobb szimuláció paramétereinek mentése

If return_total > return_total_best Then

return_total_best = return_total

T_GSMI_best = T_GSMI

req_chg_GSMI_best = req_chg_GSMI

T_fx_best = T_fx

Else

End If

```
Next req_chg_GSMI
Next T_fx
Next T_GSMI
```

```
return_total = 0
i = MC_startpoint + train_period_length
positioncount = 0
drawdown = 0
drawdown_max = 0
return_total_last_peak = 0
```

```
Do
chg_GSMI = 0
```

```
For j = i - T_GSMI_best + 1 To i
chg_GSMI = chg_GSMI + Cells(j, 2)
Next j
```

```
'ha esemény van, felvesszük a pozíciót
If -1 * chg_GSMI > req_chg_GSMI_best Then
return_for_this_position = 0
positioncount = positioncount + 1
```

```
'a hozam mérése az adott pozícióra, amíg le nem zárjuk
For i = i + 1 To i + T_fx_best
return_for_this_position = return_for_this_position + Cells(i, 3)
Next i
```

```
return_total = return_total + return_for_this_position
```

```
'a drawdown számítása
If return_total > return_total_last_peak Then
return_total_last_peak = return_total
Else
drawdown = return_total - return_total_last_peak
If drawdown < drawdown_max Then
drawdown_max = drawdown
End If
End If
```

```
Else
i = i + 1
```

```
End If
```

```
Loop While i <= MC_startpoint + train_period_length + oos_period_length - T_fx_best
```

```
Worksheets(2).Cells(h, 1) = T_GSMI_best
Worksheets(2).Cells(h, 2) = req_chg_GSMI_best
Worksheets(2).Cells(h, 3) = T_fx_best
Worksheets(2).Cells(h, 4) = return_total
Worksheets(2).Cells(h, 5) = positioncount
Worksheets(2).Cells(h, 6) = drawdown_max
h = h + 1
```

```
Next rep
```

```
Worksheets(2).Cells(1, 1) = "T_GSMI_best"
Worksheets(2).Cells(1, 2) = "req_chg_GSMI_best"
Worksheets(2).Cells(1, 3) = "T_fx_best"
Worksheets(2).Cells(1, 4) = "Teljes hozam"
Worksheets(2).Cells(1, 5) = "Felvett pozíciók száma"
Worksheets(2).Cells(1, 6) = "Max drawdown"
```

```
End Sub
```