



**Általános és Kvantitatív
Közgazdaságtan
Doktori Iskola**

TÉZISGYŰJTEMÉNY

Szűcs Balázs Árpád

A tőzsdei forgalom napon belüli előrejelzése

című Ph.D. értekezéséhez

Témavezető:

Makara Tamás, Ph.D
egyetemi docens

Budapest, 2015

Befektetések és Vállalati Pénzügy Tanszék

TÉZISGYŰJTEMÉNY

Szűcs Balázs Árpád

A tőzsdei forgalom napon belüli előrejelzése

című Ph.D. értekezéséhez

Témavezető:

Makara Tamás, Ph.D

egyetemi docens

© Szűcs Balázs Árpád

Tartalomjegyzék

1. Kutatási előzmények és a téma indoklása	2
2. Az értekezés felépítése, felhasznált adatok és módszerek	5
2.1. Az értekezés felépítése	5
2.2. Felhasznált adatok	5
2.3. Felhasznált módszerek	6
3. Az értekezés eredményei	10
3.1. A szakirodalom legjobb modelljének azonosítása	10
3.2. A szakirodalomnál jobb saját modell	12
3.3. Áttekintés	19
4. Főbb hivatkozások	21
5. A témakörrel kapcsolatos saját publikációk	22

1. Kutatási előzmények és a téma indoklása

A tőzsdével foglalkozó kutatások jellemzően inkább az árat vizsgálják, a forgalomra sokkal kevesebb figyelem jut, és ennek megfelelően sokkal kevesebbet is tudunk róla. Az árat modellező elméletek gyakran teljesen figyelmen kívül hagyják magát a forgalmat. Az ár modellezésének a dominanciája a forgalommal szemben feltehetően abból a természetes igényből ered, hogy a befektetők pénzt szeretnének keresni a tőzsdén, és az ár alakulásának megértése, és ideális esetben a többi szereplőnél jobb előrejelzése ehhez a lehető legkézenfekvőbben járul hozzá. Látnunk kell azonban, hogy a forgalom jobb ismerete, illetve a többiekénél jobb előrejelzése ugyanúgy hozzájárul a piaci szereplők vagyonának gyarapodásához.

Induljunk ki abból a tényből, hogy a likviditás korlátos, ami azt jelenti, hogy egy bizonyos tranzakcióméret felett már nem lehet *büntetlenül* kereskedni. Ez minden piacra kivétel nélkül igaz, a kérdés csak annyi, hogy ez a korlát hol helyezkedik el. Ennek oka, hogy egy átlagosnál lényegesen nagyobb megbízás nem teljesülhet teljes egészében azon az áron, ahol a piac a megbízás beadása előtt állt. Egy nagy megbízással ugyanis az árat magunk ellen mozdítjuk el, és az ügylet előjelétől függően átlagosan drágábban fogunk tudni venni, vagy olcsóbban fogunk tudni eladni, mint egy lényegesen kisebb megbízás esetén. Ezt a hatást nevezzük árhatásnak, amely esetenként igen jelentős méreteket ölthet, ezzel veszteséget okozva a megbízást adó befektetőnek. Az árhatás ugyanakkor hatékonyan csökkenthető, vagy akár elkerülhető is lehet, ha a forgalomról megfelelő előrejelzéssel rendelkezünk, hiszen ilyen esetben feldarabolható a megbízás kisebb, árhatást nem okozó szeletekre. Erre a gyakorlatban minden piaci szereplő oda is figyel.

Az NYSE-n 2015 szeptemberében a legmagasabb forgalmú napon 118 milliárd USD összértékben cseréltek gazdát részvények. Ez tehát egyetlen tőzsde egyetlen napjának az adata, miközben összehasonlításként Magyarország 2014-es (egész éves) GDP értéke 137 milliárd USD volt. Gondoljunk bele, hogy csak ezen az egy napon mennyit kereshettek az egyéni befektetők azzal, ha odafigyeltek az ajánlat feladarobás problémájára, akár csak egy szerény, 1%-os átlagos elkerült árhatást feltételezve is. Látható, hogy az egyéni befektető szintjén egy jó forgalom előrejelzés jelentős vagyongyarapodásra fordítható le.

A forgalom ugyanakkor az egyéni befektetőn túl a piac egésze szempontjából is fontos fogalom, melynek belátásához elég a piaci hatékonyságra gondolni. Annál hatékonyabb egy piac, minél több információ, és minél gyorsabban épül be az árakba. A piaci hatékonyság lehető legnagyobb mértéke ezért nyilvánvalóan kívánatos. Az információ az

árakba azonban csak egyetlen módon tud beépülni, még hozzá kereskedés útján. Minél nagyobb forgalommal zajlik a kereskedés, az árfeltárás annál jobban működhet. Úgy is fogalmazhatunk, hogy az információ beépülése az árakba kizárólag a forgalmon keresztül lehetséges.

A forgalom előrejelzésének a jelentősége ebben a folyamatban is felismerhető. Mint korábban láttuk, a korlátos likviditásból következő árhatás elkerülése érdekében a szereplők feldarabolják a megbízásaikat, és ezáltal lassítják a kereskedést. Egy jobb forgalom előrejelző modellel csökkenthető a bizonytalanság azzal kapcsolatban, hogy mekkora megbízás adható be árhatás nélkül, és így a darabolás során nagyobb szeleteket adhatnak be egyszerre, amiből következően hamarabb teljesül a teljes megbízás. Ezáltal gyorsabbá válik a kereskedés, ami közvetlenül hozzájárulhat a piaci hatékonyság növekedéséhez. Ez pedig azt jelenti, hogy a piac egésze jobban működik.

Végül a forgalom előrejelzés jelentősége kapcsán az egyéni befektető, illetve a piac egésze mellett érdemes egy harmadik szempontot is megemlíteni, amely az elsődleges részvénykibocsátásokkal kapcsolatos. Az elsődleges részvénykibocsátások (*Initial Public Offering*, IPO) egyértelműen hozzájárulnak a gazdasági növekedéshez, melynek oka, hogy ha egy cég sikeresen tőzsdére lép, azzal új finanszírozási forráshoz jut, amelyből növekedni tud, és ennek következtében egyúttal új munkahelyeket is teremt, így pedig az egész gazdaságra jótékony hatással van. Megfigyelhető azonban, hogy a 2000-es évek kezdete óta az IPO-k száma jelentősen visszaesett az USA-ban, vagyis kevesebb cég lép új szereplőként a tőzsdére. Ennek magyarázata az automatikus kereskedés megjelenésében is kereshető, aminek köszönhetően néhány részvény szuper likviddé vált, ennek következtében azonban a többi már kevésbé likvidnek számít. Ezért ez utóbbiak elvárt hozamában megjelenik egy illikviditási prémium (az árukban pedig egy illikviditási diszkont), amelyet nem tudnak minden esetben kitermelni. Épp ez a helyzet egy új tőzsdei belépő esetén is, amely kezdetben mindenképp illikvidként indulna, de az említett illikviditási prémium miatt megemelkedett hozamelvárásnak már nem tud megfelelni, és ezért eleve be sem lép a tőzsdére, mert kudarcra van ítélve.

Az említett likviditási prémium tehát a likviditási kockázat miatt jelenik meg a hozamban, hiszen nehéz jó áron megvenni vagy eladni a részvény nagyobb pakettjeit. Ez a likviditási kockázat azonban éppen az árhatás miatt jelentkezik. Ahogyan korábban láttuk, az árhatás csökkentésének hatékony módja a forgalom előrejelzés, ami ezen keresztül a likviditási kockázat kezelésére is alkalmas. Ha pedig sikerül csökkenteni a likviditási kockázatot, azzal csökken az illikviditási prémium, aminek következtében több cég képes sikeresen tőzsdére lépni.

A tőzsdei forgalom jobb előrejelzése tehát úgy az egyéni befektetőre, mint magára a piac egészére, illetve a teljes gazdaságra nézve is kedvező hatást gyakorol.

Az értekezés a tőzsdei kereskedési volumen, illetve annak százalékos alakban megadott formája, a forgalom napon belüli előrejelzéséről szól. Ez a terület még fejlődőben van, és meglehetősen friss. Az értéktőzsdei volumen előrejelzésével kevés kutatás foglalkozott eddig, a nyilvánosan elérhetőek közül legalábbis mindenképpen. Az első ilyen kutatást bemutató cikk 2007-ben jelent meg, de az még napi sűrűségű adatok felhasználásával készült, ami a volumen jellegzetes napon belüli stilizált tényei miatt csak közvetett előzménynek tekinthető. A volumen napon belüli előrejelzéséről 2008-ban jelent meg az első cikk, a következő pedig 2011-ben.

Az értekezés célja egyrészt áttekinteni a tőzsdei forgalom előrejelzésében eddig elért tudományos eredményeket az ezekhez vezető elméleti és módszertani előzményekkel együtt, másrészt pedig a legpontosabb előrejelzések saját adatokon való reprodukálását követően olyan modelleket keresni, amelyek azoknál jobb eredményre vezetnek.

2. Az értekezés felépítése, felhasznált adatok és módszerek

2.1. Az értekezés felépítése

Az értekezés felépítését tekintve öt részre osztható.

Az I. részben a kutatás előzményeit ismertetem a következők szerint. A 2. fejezetben keretbe helyezem a kutatást a legfontosabb fogalmak bemutatásán keresztül, egyúttal megvilágítva a kutatási téma relevanciáját is. A 3. fejezetben már az empirikus kutatás közvetlen előzményeit, vagyis a forgalom előrejelzés szakirodalom által lefedett eredményeit mutatom be.

A II. rész az empirikus kutatás megalapozását célozza. A 4. fejezetben bemutatom a rendelkezésemre álló adatbázist, illetve annak a kutatáshoz felhasznált szeletét is. Az 5. fejezetben a konkrét kutatási kérdéseket ismertetem, valamint megfogalmazom a vizsgálandó hipotéziseket.

A III. rész a szakirodalmi eredmények saját adatokon történő reprodukálását tartalmazza annak érdekében, hogy eldönthessük, ezeken az adatokon a szakirodalom melyik modelljét tekinthetjük a legjobbnak. A 6. fejezetben a becslés közös részleteit rögzítem, míg a 7. és 8. fejezetekben a szakirodalom két releváns modelljének a becslését találjuk. Ezek alapján a 9. fejezetben azonosítom a benchmarkot, amelynél később megpróbálok jobb saját modellt készíteni.

A IV. rész a saját modell keresését tartalmazza. A 10. fejezetben sztenderd hibamértékek alapján keresek a korábban azonosított benchmarknál jobb modelleket. A 11. fejezetben a legjobbnak vélt saját modelleket egyéb, a szakirodalom által relevánsnak ítélt szempontok alapján is értékelem. A 12. fejezetben összefoglalom a negyedik rész, azaz a modellezés eredményeit.

Az V. rész a dolgozat zárásául szolgál. Itt találjuk a 13. fejezetben az összefoglalást, a 14. fejezetben a főbb eredmények rövid áttekintését, illetve a 15. fejezetben a további kutatási kérdések felvetését.

2.2. Felhasznált adatok

Az adatbázis a Dow Jones Iparági Átlag (röviden DJIA vagy Dow 30) index részvényeit tartalmazza, amely az Amerikai Egyesült Államok tőzsdéinek harminc jelentős vállalatát foglalja magában. Az indexet 1896 óta számítják, ezalatt az idő alatt némileg

változott a benne szereplő részvények listája. Ennek köszönhető, hogy az adatbázisban nem harminc, hanem harminchat részvény szerepel, melyek közül harminchárom az NYSE-n, a maradék három pedig a NASDAQ-on jegyzett. Az első adatnap 1998.01.02. minden részvénytől, kivéve, ha az adott tőzsdére később vezették be a papírt, mely esetben a bevezetés napja az első adatnap. Az utolsó adatnap minden részvénytől 2012.07.13.

Az adattisztítás után a 2001.10.10 és 2012.07.13 közötti 130 hónapos, azaz közel 11 éves időszakot tartottam meg, illetve 33 részvényt. Az eredeti adatok percenkénti sűrűségűek voltak, de a szakirodalommal való jobb összevethetőség érdekében 15 perces rekeszekbe aggregáltam az adatokat, így naponta 26 megfigyelést kaptam. A felhasznált adatbázisban szereplő részvények kellően likvidek voltak ahhoz, hogy minden részvény minden 15 perces rekeszében legyen kötés, és ezáltal nullától különböző forgalom adat is. Így összességében a felhasznált forgalom adatbázisom 2,29 millió megfigyelésből áll.

2.3. Felhasznált módszerek

Az értekezésben felhasznált módszerek közül az általam javasolt, szakirodalomhoz képest új módszereket érdemes kiemelni. Ezek elsősorban az U dekompozícióra vonatkozó különböző lehetőségek.

2.3.1. Polinom illesztés

Az U-módszer, miután gyakorlatilag átlagolásról van szó, természetes, hogy *rücskös*, vagy úgy is mondhatjuk, hogy *zajos* U alakot eredményez. Érdemes megnézni, hogy mit eredményez egy olyan dekompozíció, amely mentes az efféle zajtól, vagyis egy simább lefutású, *simított* U alakra vezet.

A simaság és az U alak együtteséről eszünkbe juthat a másodfokú polinom, de természetesen nem érdemes a fokszámot azonnal rögzíteni. Tekintsük az alábbi n -edfokú polinomot:

$$p_t = \sum_{i=0}^n \beta_i x_t^i \quad (1)$$

ahol $t \in \{1, 2, \dots, T\}$ és $x_t = t/T$, valamint T jelöli az egy napon belüli rekeszek számát. Vagyis az U-módszerhez hasonlóan $T = 26$ darab p_t pontot keresünk, amelyek egy polinomra fekszenek fel, és a legjobban illeszkednek a becslési időszakban megfigyelt forgalom idősor $J = 20$ napjára. Ehhez az alábbi feladatot kell megoldani:

$$\sum_{j=1}^J \sum_{t=1}^T (p_t - y_{j,t})^2 \rightarrow \min_{\beta_i} \quad (2)$$

ahol t tehát a rekeszek indexe, $j \in \{1, 2, \dots, J\}$ pedig a napoké, és y jelöli a megfigyelt forgalom adatokat.

2.3.2. Exponenciális súlyozású polinom illesztés

Elméletileg elfogadható feltevésnek tűnik, hogy a frissebb megfigyelések jobban magyarázzák a holnapi forgalmat, mint a régebbiek. Ebből kiindulva érdemes tesztelni a modell egy olyan módosítását, amelyben minél aktuálisabb egy adott megfigyelés, annál nagyobb súllyal vesszük figyelembe az előrejelzésben. Annak érdekében, hogy az egyes napokhoz rendelt súlyok érdemben eltérjenek, olyan megoldást választok, ahol a jelenhez közeledve exponenciálisan nő az adott nap súlya.

A fentebb bemutatott polinom illesztés ekkor annyiban módosul, hogy a (2) feladatba bekerül egy súly is. Ezen módosítás után a célfüggvény az alábbi formát ölti:

$$\sum_{j=1}^J \sum_{t=1}^T S_j \cdot (p_t - y_{j,t})^2 \rightarrow \min_{\beta_i} \quad (3)$$

ahol

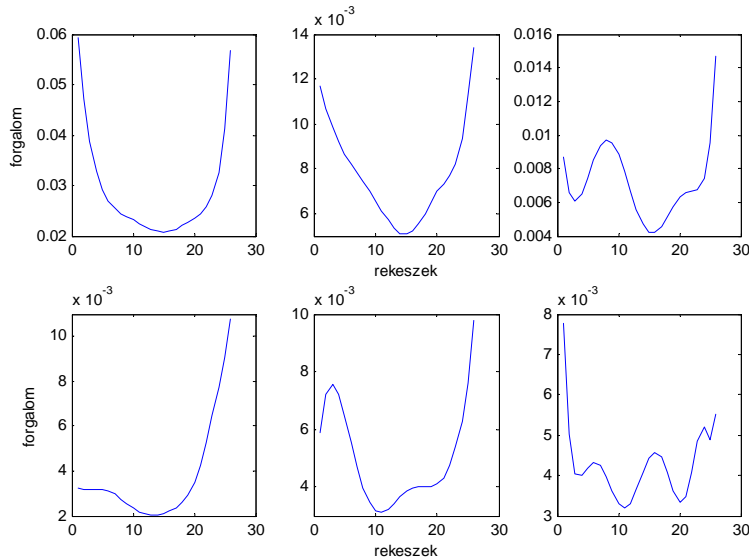
$$S_j = e^{-\frac{j}{J} \cdot \ln(1/J)} \quad (4)$$

Tehát naponta exponenciálisan nő a hibához rendelt súly, és a jelenhez legközelebbi napon lesz a legnagyobb.

2.3.3. Spline illesztés

Habár U alakként szokás rá utalni, a szabályos U betűhöz képest igen változatos alakot ölthet a napon belüli szezonális adottság adott időszakában. Erre láthatunk néhány példát az 1. ábrán a Kraft Foods Inc. forgalmának különböző 20 napos időszakaira illesztett 14-edfokú polinomokat szemlélve. A szabályos U alakon kívül találunk olyat is, ami inkább hasonlít V-re, W-re vagy J-re, illetve előfordul, hogy a forgalom nap elején felfelé indul el a kezdeti értékről, valamint néha jelentősebb hullámmás is. Részben az ebből eredő rugalmasságra való igény is magyarázat lehet arra, hogy viszonylag magasabb fokszám mellett érdemes dönteni a polinom illesztésnél.

Ugyanakkor, habár a rugalmasság előny, a túlzottan jó illeszkedés (még magasabb



1. ábra. A Kraft Foods Inc. forgalmának különböző 20 napos időszakaira illesztett 14-edfokú polinomok
 Forrás: Saját szerkesztés

fokszám) mégsem célravezető, hiszen összességében ront az előrejelzésen. Felmerülhet, hogy ennek az az oka, hogy ilyenkor esetleg olyan kilengéseket is lekövet az illesztett polinom, ami már inkább csak zaj, márpedig éppen azért merült fel a polinom mint simított forma illesztése, hogy elkerüljük a zajos U alakot, amit például az U-módszer eredményezett.

Tehát egyrészt szeretnénk, hogy meglegyen a rugalmasság lehetősége (l. 1. ábra), ugyanakkor a zajt el szeretnénk kerülni. Erre a problémára potenciálisan megoldást kínálhatnak a hozamgörbe becslésben is hasonló motiváció alapján használt spline függvények. Az alábbiakban a spline függvény definíciója következik.

Definíció: Legyen adott K valós szám, $t_1 < t_2 < \dots < t_K$, melyeket csomópontoknak nevezünk. Az adott csomópontokon értelmezett N -ed fokú spline függvénynek nevezünk minden olyan $[t_1, t_K] \rightarrow \mathbb{R}$ folytonos függvényt, amelyik bármely két szomszédos csomópont között egy N -ed fokú polinom értékeit veszi fel, és amelyik $(N - 1)$ -szer folytonosan differenciálható.

A fenti definíció ebben a formában $N \geq 2$ esetén érvényes, de külön értelmezhetünk $N = 1$ és $N = 0$ értékekre is spline függvényt. Az adott feladat tükrében számunkra azonban ez a kibővítés nem járna érdemi előnnyel.

A spline függvények használatával tehát úgy illeszthetünk polinomot, hogy a fok-

szám alacsony marad, ezáltal a zajt kevésbé veszi magára, de mivel a polinom paraméterei intervallumonként változnak, mégis lényegesen rugalmasabb lehet az illesztett spline függvény, mint egy azonos fokszámú polinom.

A fentiek alapján az illesztett függvényt ezúttal az alábbi formában keressük:

$$p_t = \sum_{m=1}^M \beta_m f_m(x_t) \quad (5)$$

Ez hasonlít az (1)-ben felírt polinom illesztésre, tehát ugyanúgy $t \in \{1, 2, \dots, T\}$, azaz $T = 26$ pontot keresünk, valamint $x_t = t/T$. Annyi a különbség, hogy most az f függvény más lesz, ezért ezt még specifikálni kell.

Az f függvények (bázis függvények) megválasztásánál fontos szempont, hogy elkerüljük a multikollinearitást. Erre egy alkalmas megoldás a B-spline bázisfüggvények használata, melyek értékei rekurzív módon határozhatóak meg a fokszám és a csomópontok ismeretében.

A paraméterbecslés az egyenlően súlyozott polinomoknál látott (2) feladathoz hasonlóan történik.

Modellezői döntés kérdése mind a csomópontok K száma és azok elhelyezése, mind pedig az N fokszám megválasztása. Megmutatható, hogy a szükséges bázisfüggvények, és ezáltal a becslendő paraméterek M száma a következő módon határozható meg:

$$M = N + K - 1 \quad (6)$$

Láthatóan ez a szokásos polinom illesztésre is alkalmazható. Ha például egy egyszerű harmadfokú polinomot ($N = 3$) szeretnénk illeszteni, akkor csak egy intervallum lesz, melynek két végpontja a két csomópont ($K = 2$), így $M = 3 + 2 - 1 = 4$, ami a konstans figyelembe vételével valóban a becslendő paraméterek száma.

Ugyanakkor a spline függvényeket épp azért vezettük be, hogy eltérjünk a normál polinom illesztéstől, ezért legalább két intervallumot érdemes vizsgálni, ami $K \geq 3$ választásnak felel meg.

3. Az értekezés eredményei

3.1. A szakirodalom legjobb modelljének azonosítása

Az első kutatási kérdést a III. részben vizsgáltam. Célom a szakirodalom két releváns modelljének (ezek Bialkowski et al. (2008) és Brownlees et al. (2011)) az összehasonlítása azonos adatokon és azonos módon értékelve annak érdekében, hogy eldönthessem, melyik jelenleg az irodalomból megismerhető legjobb napon belüli forgalom előrejelző modell, amelyet a későbbiek során benchmarknak tekinthetek egy jobb modell keresése során.

Ehhez néhány szempontot előzetesen rögzítenem kellett. Egyrészt meghatároztam, hogy hogyan mérem a hibát, és mi alapján tekintek egy modellt jobbnak egy másiknál. Ezen a ponton egyelőre egyszerűen az MSE és MAPE hibamértékeket vizsgáltam az előrejelzett értékre vonatkoztatva, ahol minél kisebb a hiba, tehát minél közelebb van az előrejelzés a később megfigyelt értékhez, annál jobb. Ennek kapcsán jobbnak tekintek egy modellt, ha több részvényre ad kisebb hiba értéket, illetve a részvények átlagában kisebb hibát mutat. Utóbbira elsősorban azért van szükség, mivel ha az átlagos értékek közel egyezők, az általában azt jelzi, hogy nincs lényeges eltérés az egyedi részvények hibáiban sem.

Másrészt hasonló adatokon, és hasonló módon kell előrejelezni az összevethetőség érdekében, ezért ezeket a szempontokat is rögzítenem kellett. A használt adatok természetesen a korábban meghatározott adatbázis adatai, tehát 33 részvény 130 hónapja, 15 perces aggregálással, ami naponta 26 megfigyelést eredményez minden részvényben. A becsléshez 20 napot használok minden esetben, amely nagyjából egy naptári hónapnak felel meg (20 kereskedési nap). Ez alapján egy napot jelzek előre, tehát az előrejelzés értékeléséhez a becslési időszakot követő 1 napot használom, ami azt is jelenti, hogy a paramétereket naponta frissítem. Így 2648 előrejelzett napot kapok minden részvényre. Az információ minden esetben 15 percenként frissül, vagyis minden lépésben beépítem a beérkező új információt (ez esetben forgalom adatot).

Ezek után minden készen áll Bialkowski et al. (2008)¹ és Brownlees et al. (2011)² modelljeinek megbecsléséhez. A BDF modell becslése a cikk alapján minden további nélkül elvégezhető volt. A gyakorlatban használt U-módszerrel összehasonlítva egyértelműen megállapítható, hogy a BDF modell az itt vizsgált hibamértékek szerint jobban

¹Röviden: BDF modell

²Röviden: BCG modell

szerepelt nála.

A BCG modell becslése során azonban a cikk ismeretében is felmerültek bizonytalanságok, melyek az alábbiak voltak. Egyrészt két esetben a szerzők nem adták meg, hogy milyen kezdő értékről indítják a változót a rekurzió során. Másrészt a napon belüli periodikus komponens egyenletében jelzik, hogy csökkentik az eredetileg 25 tagból álló kifejezésben a tagok számát, de pontosan nem derül ki, hogy hogyan. Harmadrészt, habár ez már csak technikai kérdés, nem derül ki, hogy hogyan állítják be a keresett θ kezdő értékeit az optimalizálás során, amely, mint később kiderült, kulcskérdés a becslés sikere szempontjából. Ezekben az esetekben tehát saját feltevésekkel kellett élnem. Mindezek önmagában apróságnak tűnhetnek, de a cikk *tökéletes* reprodukcióját összességében lehetetlenné teszik. Ezek mellett az, hogy a konkrét modell specifikációt nem abban a formában írták fel, amit meg is becsültek később, hanem egy jelentősen egyszerűbb változatban, már csak egy áthidalható nehézséget jelentett, hiszen a bővítésre vonatkozóan megadták a kellő útmutatást.

Magával a becsléssel azonban a fentiekén túl is adódtak problémák, nevezetesen az, hogy a GMM becsléshez megadott célfüggvény úgy tűnik, nem kellően sima ahhoz, hogy *elfogadható időn* belül *elfogadható megoldásra* lehessen jutni belőle, ugyanis túl sok a lokális szélső értéke, és ezért a solver gyakran olyan pontban áll meg, amely nem ad értelmes előrejelzést. A nem elfogadható idő alatt azt értem, hogy az adatbázison az alapbeállítással 60 napig futott a becslés³, amely lényegesen hosszabb, mint bármelyik másik vizsgált modell esetén (ami maximum 1 nap volt). Az elfogadható megoldás pedig azt jelenti, hogy legalább nagyságrendileg megközelíti az előrejelzés a később megfigyelt tényeket, amely elmondható bármely más általam vizsgált modellről, a BCG modellről azonban még a fent említett futási idő után sem.

Végül a becslés menetében történő bizonyos változtatások árán sikerült elfogadható eredményeket kapnom a BCG modelltől is. Ugyanakkor ezen változtatások szükségessége összességében egyértelművé teszi, hogy a modell a cikkben szereplő formában, az általam használt adatbázison legalábbis mindenképpen instabil, és az eredményessége emiatt meglehetősen esetleges.

Hozzá kell tenni, hogy Brownlees et al. (2011) ETF adatokkal dolgozik, nem pedig részvény adatokkal, de ez a különbség véleményem szerint kétséges, hogy önmagában indokolhatná a fenti problémákat.

³Ez gépidőben értendő, vagyis ha a 6 rendelkezésemre álló, némileg eltérő teljesítményű számítógép helyett egy átlagos gépet használtam volna. A tényleges várakozási idő a hat számítógép használata miatt rövidebb volt 60 napnál.

Az alkalmazott módosítások után kapott, értelmesnek tekinthető eredményeket összevettem az U-módszerrel és a BDF modellel is. Ezek alapján az U-módszer előrejelzése rosszabbnak, a BDF modell előrejelzése azonban az általam vizsgált szempontok szerint határozottan jobbnak tekinthető (különösen SETAR módon becsült egyedi rész mellett) a BCG modellnél.

A BCG modell fent említett instabil előrejelzése miatt nem tartottam szükségesnek egyéb értékelési szempontok szerint is összevetni a két modellt. Ezek alapján tehát választ kaptunk az első kutatási kérdésre.

H1 Hipotézis: Benchmark. Az első hipotézisemet a fentiek alapján elfogadom, miszerint a BDF modell jobb a BCG modellnél, és ezáltal a BDF modell tekinthető a szakirodalomból megismerhető legjobb napon belüli részvény forgalom előrejelző modellnek.

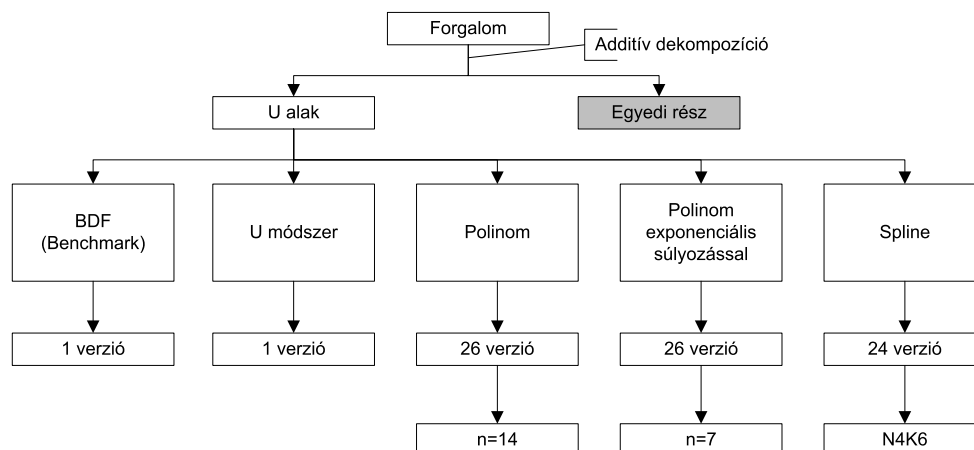
Ezek alapján a továbbiakban a BDF modellt tekintetem benchmarknak.

3.2. A szakirodalomnál jobb saját modell

A IV. részben a második kutatási kérdést és kiegészítéseit vizsgáltam. Ez a rész tehát egy, a benchmarknál jobb modell kereséséről szólt, ahol a hiba mérését a korábbiakkal megegyező módon végeztem.

Első lépésben teszteltem néhány egyszerű, U dekompozíciót nem tartalmazó modellt, de ezek a várakozásaimnak megfelelően nem szerepeltek jól, ami alapján nem tartottam érdemesnek ebben az irányban tovább vizsgálódni. Ezután az U alak leválasztásának a szakirodalomhoz képest új lehetőségeit vizsgáltam, első megközelítésben Bialkowski et al. (2008) modelljéhez hasonlóan additív struktúrában (U alak + egyedi rész). Ehhez az egyedi részt mindvégig változatlanul tartottam (a jobb összehasonlíthatóság érdekében a benchmark BDF modell által használt verzióban), így kizárólag az U alak modellezésének hatását tudtam vizsgálni. A tesztelt modellek áttekintésében segít az 2. ábra.

Elsőként az egyszerű U-módszert egészítettem ki egyedi résszel, és érdekes módon már az így kapott modell is jobban szerepelt a benchmarknál. Ezután észrevettem, hogy a BDF modell által dekomponált U alakhoz hasonlóan az U-módszer által meghatározott U alak is *zajos*, tehát nem kifejezetten sima lefutású. Ezért érdemesnek láttam ennek az U alaknak egy simított változatával próbálkozni, melyre több lehetőséget is találtam.



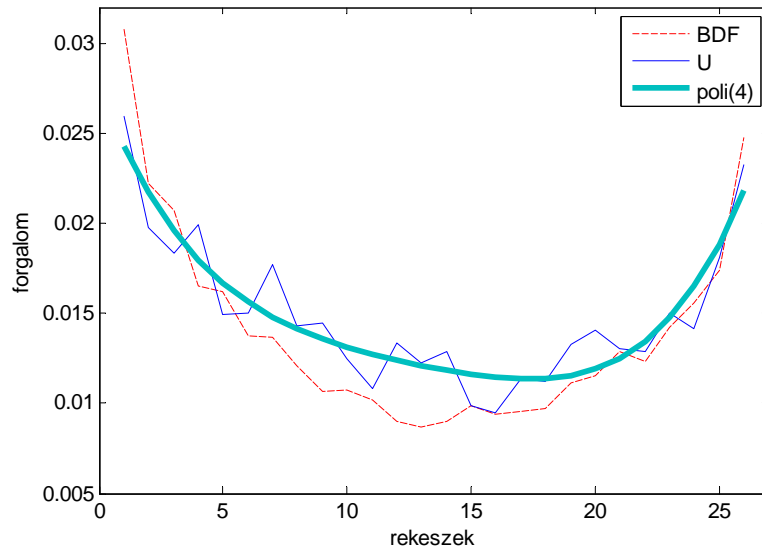
2. ábra. Az U alak dekomponálásának különböző módjai
 Forrás: Saját szerkesztés

Az első egy n -edfokú polinom illesztése az adatokra, pontosabban az előrejelezni kívánt részvény forgalom idősorára. A 3. ábra egy szemléltető példa arra, hogy egy tetszőlegesen választott fokszám ($n = 4$) esetén hogyan alakulnak a modellek által meghatározott U alakok. Látható, hogy a BDF modell U alakja határozottan eltér a másik kettőtől, illetve az is megfigyelhető, hogy a polinom pusztán vizuálisan értékelve az U-módszer simított változataként is felfogható.

Mivel az adatbázisomban naponta 26 megfigyelés van részvényenként, ezért az n fokszámra 1 és 26 között kerestem a legjobban használható értéket. Végül az $n = 14$ választást találtam legjobbnak az adatbázisomon, de kiemeltem, hogy $n \geq 7$ választás esetén már lényeges különbség nem látszik az egyes hiba értékek között.

A 14-edfokú polinommal mint U alakkal végzett előrejelzés határozottan jobban szerepelt a benchmarknál, ezért tovább vizsgáltam ebben az irányban esetleges további javítások után kutatva.

A következő modellem az U alak leválasztására a fentiek alapján egy exponenciálisan súlyozott polinom illesztése volt. A fenti egyenlően súlyozott változathoz képest ez abban tér el, hogy nagyobb súllyal veszi figyelembe az időben közelebbi napok adatait. A fokszámot itt is 1 és 26 között kerestem, és ezúttal $n = 7$ mellett döntöttem a korábbiakhoz hasonló vizsgálatok után, szintén hozzátéve, hogy $n \geq 7$ választás mellett elenyésző eltérést tapasztalhatunk a modellek között. Az alacsony, tehát 7-nél kisebb fokszámok használatát azonban sem az egyenlő, sem az exponenciális súlyozás mellett nem javaslom. Az exponenciálisan súlyozott polinom is határozottan jobban szerepel a benchmarknál, de az egyenlően súlyozott esetről már nem.

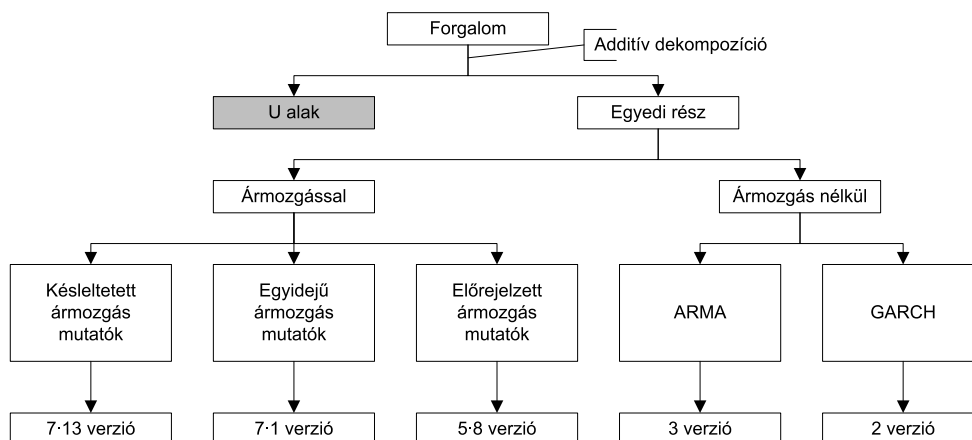


3. ábra. Az U-módszer, a BDF modell U alakja és a negyedfokú polinom az Alcoa, Inc. mintabeli első 20 napjára illetve
 Forrás: Saját szerkesztés

Végül megvizsgáltam a spline illesztés lehetőségét is, amelynek a polinom illesztéssel szemben vannak bizonyos előnyei, ezért elképzelhetőnek tartottam, hogy a használatával javulhat az előrejelzés. Spline esetén a fokszám mellett a csomópontok számáról is dönteni kell. Ez esetben 24 változatból a negyedfokú, hat csomóponttal felírt verziót választottam (N4K6). Ez a modell is határozottan jobb a benchmarknál, de az egyenlő súlyozású polinomnál már nem. Az eltérés a kettő közt ugyan minimális, de mivel a polinom illesztés lényegesen egyszerűbbnek tekinthető, azt tartottam meg preferált modellként.

Az U alak modellezésén keresztül tehát a szakirodalomhoz képest új módszereket használva minden bemutatott esetben sikerült a benchmarknál jobb előrejelzést produkálnom. A vizsgáltak közül a legjobb a 14-edfokú polinom illesztéssel leválasztott U alakot használó modell volt, ezért a későbbi vizsgálatokban ezt tekintettem preferált változatnak. A fentiek alapján a második kutatási kérdéssel kapcsolatos részkérdésre sikerült választ találnom.

H2.1 Hipotézis: Az U alak. A forgalom napon belüli U alakjának modellezése a vizsgálataim alapján valóban hozzájárul a benchmarknál jobb modell felírásához. Az U alak modellezését mellőző modell változatokban nem sikerült megfelelő eredményt



4. ábra. Az egyedi rész előrejelzésének különböző módjai
 Az U alak minden esetben Poli(14) alapján dekomponálva
 Forrás: Saját szerkesztés

elérnem, az U alak saját módszerekkel történő modellezésével viszont önmagában is sikerült megvernem a benchmarknak tekintett BDF modellt. Az általam javasolt U dekompozíciós modell változatok közül (U-módszer, egyenlően súlyozott polinom illesztés, exponenciálisan súlyozott polinom illesztés, spline illesztés) mindegyik jobb eredményre vezet, mint a benchmark. Mind közül az egyszerű polinom illesztést találtam legjobbnak.

Az additív dekompozíció logikájából következik, hogy az U alak modellezése után az egyedi rész modellezésére is érdemes külön figyelmet fordítani. Ezért a továbbiakban ezzel folytattam, melynek során a korábban legjobbnak talált U modell (vagyis az egyenlően súlyozott polinom illesztés) használata mellett vizsgáltam különböző egyedi rész előrejelzési lehetőségeket. Az egyes modell változatok áttekintésében a 4. ábra segít.

Két részre oszthatóak az egyedi résszel kapcsolatos vizsgálataim: az elsőben szerepel valamilyen ármozgás mutató, a másodikban pedig nem. Az első esetben tehát a forgalom mellett különböző ármozgás mutatók modellbe foglalásának lehetőségeivel kísérleteztem, mivel a forgalom leíró jellemzésével foglalkozó szakirodalom alapján ez az irány ígéretesnek tűnt. Az alábbi hét ármozgás mutatót vontam be a vizsgálatba: loghozam, volatilitás, rés, ársáv, százalékos ársáv, tényleges ársáv, tényleges százalékos ársáv.

Az ármozgás mutatók késleltetettjeinek a használatát több különböző módon is teszteltem, az egyszerű késleltetettek mellett Granger-okságon illetve korreláción alapuló

feltételes modellek formájában is, de egyik megoldás sem vezetett kielégítő eredményre.

Ezután megvizsgáltam egy elméleti lehetőséget, mely szerint az előrejelezni kívánt forgalom adattal megegyező időpontbeli (vagy röviden egyidejű) ármozgás mutatót használtam fel az előrejelzéshez. Ez a megoldás valós helyzetben nem alkalmazható, hiszen az előrejelzés pillanatában még nem ismert a következő megfigyelés. Arra azonban alkalmas ez a vizsgálat, hogy eldöntsük, érdemes-e az ármozgás mutatók előrejelzésével foglalkozni, hiszen ha még az egyidejű tény érték sem járul hozzá érdemben a forgalom előrejelzéshez, még kevésbé várhatjuk ezt egy előrejelzett értéktől. Ezzel kapcsolatban azt találtam, hogy a vizsgált egyidejű ármozgás mutatók a loghozam és a rés kivételével mind egyértelműen javítanak a forgalom előrejelzésen, ezért a következő lépésben ezen gondolat kísérlet után a maradék öt ármozgás mutató előrejelzett értékeit használtam fel a forgalom előrejelzéshez.

Az ármozgás mutatók előrejelzésével az adott terület (l. volatilitás előrejelzés) fejlettségéhez képest viszonylag egyszerűbb modellek használata mellett nem sikerült kedvező eredményre jutnom. Ezzel kapcsolatban két megjegyzést szükséges tenni. Egyrészt a volatilitás előrejelzés rendkívül kiterjedt irodalommal rendelkezik (elég csak a sztochasztikus volatilitás modellekre, vagy a GARCH változatokra gondolni), melynek feltérképezése meghaladja jelen dolgozat kereteit, ezért természetesen az itt elvégzett néhány modell értékelése alapján nem jelenthető ki, hogy ez az irány nem járható út. Másrészt viszont véleményem szerint az egyidejű ármozgás mutatók éppen azért nyújthattak hathatós segítséget a forgalom előrejelzésében, mert megjelennek bennük ugyanazon sokkszerű, tehát definíció szerint előrejelezhetetlen hatások, amelyek az azonos időszakban is testet öltenek, és amik nem láthatóak előre a forgalom korábbi értékeiből. Éppen ezért azonban könnyen megeshet, hogy ugyanezek a hatások az ármozgás mutatók korábbi értékeiből sem következnek, hiába alkalmaznánk esetleg összetettebb előrejelző modelleket.

H2.2 Hipotézis: Ármozgás mutatók. Vizsgálataim alapján nem sikerült alátámasztanom, hogy az ármozgás mutatók modellbe foglalásával jobb forgalom előrejelző modellt lehetne készíteni.

Az egyedi résszel kapcsolatos vizsgálataim második részében ármozgás mutatók nélkül, tehát csak a forgalomra támaszkodva specifikáltam modelleket. Ennek keretében ARMA és GARCH változatokat teszteltem, és arra jutottam, hogy az ARMA(1,1) modell a legjobb egyedi részt előrejelző változat az összes általam vizsgált közül.

Fejezet	Mit jelzünk előre	Mit figyelünk	Hibamérték	Összevetés
10.	Egy lépéses forgalom érték (<i>One-step-ahead</i>)	Egyszerű eltérés	MSE	Nyerő darab
				Átlag
			MAPE	Nyerő darab
				Átlag
11.	Egész napos forgalom érték (<i>Multiple-step-ahead</i>)	Egyszerű eltérés	MSE	Nyerő darab
				Átlag
			MAPE	Nyerő darab
				Átlag
	Egész napos forgalom arány (Statikus stratégia)	VWAP eltérése	MAPE	Nyerő darab
				Átlag
	Egy lépéses forgalom arány (Dinamikus stratégia)	VWAP eltérése	MAPE	Nyerő darab
				Átlag

1. táblázat. Forgalom előrejelző modellek értékelésének különböző szempontjai
Forrás: Saját szerkesztés

Ezek után megvizsgáltam még néhány olyan modellt, amelynek nem a fentiekben követett additív logika adja a vázát. Az első ilyen irány keretében tisztán U alak modelleket vizsgáltam, amelyek tehát egyedi részt nem tartalmaznak. A második egy korrekciós modell volt, a harmadik pedig az eddigi legjobb U alak modell és egyedi rész modell additív helyett multiplikatív felírásban. Ezek közül egyedül az utóbbi, vagyis a multiplikatív modell bizonyult ígéretesnek.

A IV. rész zárásaként a legjobb saját modelleket az eddigiek mellett további szempontok mentén is összevetettem a benchmarkkal. Tekintsük át ezen szempontokat az 1. táblázat segítségével.

A modell keresés során, tehát mindeddig egy lépéses (*one-step-ahead*) előrejelzést végeztem lépésenkénti információ frissítéssel, az értékelést pedig MSE és MAPE alapon tettem meg az előrejelzett és a tény értékek eltérésére vonatkozóan. Amit figyeltem, hogy hány részvényre adott kisebb hibát az adott modell, illetve, hogy a részvények átlagában hogyan alakult a hiba. Ez utóbbi elsősorban annak eldöntésében segített, hogy ki tudjam szűrni a lényegében nem különböző hibájú előrejelzéseket. Első körben tehát ezek alapján szelektáltam a modellek között, és két modell jutott tovább a későbbi vizsgálatokra.

1. Additív modell polinom U alakkal és ARMA egyedi résszel

2. Multiplikatív modell polinom U alakkal és ARMA egyedi résszel

A második szempont egész napos, tehát 26 lépéses (*multiple-step-ahead*) előrejelzés, melynek értékelése a fentiekhez hasonlóan történt.

Mivel a forgalom előrejelzésnek kiemelt szerepe van a VWAP kereskedésben, Bialkowski et al. (2008) alapján az előzőek mellett további két szempontot is megvizsgáltam. Mindkét esetben napon belüli forgalom arányokat jelzünk előre, és az a kérdés, hogy ha a megadott stratégia alapján kereskednénk, akkor milyen mértékben sikerülne megközelíteni az utólag megfigyelt VWAP értéket. Ezt a hibát mérjük MAPE alapon. Az MSE mérték ezúttal nem használható, mert az árszint bevonása miatt torzítást okozna.

Az egyik vizsgált stratégia a statikus stratégia, amely esetén nap elején a teljes napra előrejelezzük a forgalom arányokat, és ezeket később nem vizsgáljuk felül. Ez nem reális választás a valóságban, ezért ennek a jelentősége kisebb. A másik a dinamikus stratégia, amely keretében mindig 15 percre jelezzük előre a forgalom arányt, és folyamatosan frissítjük az információs bázist. Bialkowski et al. (2008) szerint egy forgalom előrejelző modellnek az utóbbi, tehát a dinamikus stratégia VWAP-tól való eltérése a legfőbb mércéje.

Az 1. táblázat jobb oldali oszlopából láthatjuk, hogy a két legjobb saját modellt összességében 12 szempont szerint hasonlítottam össze a benchmarkkal, melyek tartalmazzák a benchmark cikk által felállított szempontokat is. Megállapítható, hogy mindkét általam javasolt modell mind a 12 szempont szerint jobban teljesít, mint a benchmark BDF modell bármelyik változata. A két saját modell közül is inkább a multiplikatív változat preferálható, hiszen a 12 szempontból 11-ben az bizonyul jobbnak⁴.

H2 Hipotézis: Jobb modell. A modell keresés eredményeként sikerült olyan modellt találnom, amely a vizsgált 12 szempont mindegyike szerint egyértelműen jobban szerepel, mint a szakirodalom alapján kiválasztott benchmark. Ez a modell a benchmarkkal ellentétben multiplikatív logikára épül. Az U alakot mások által eddig nem használt módon, polinom illesztéssel választja le. A multiplikatív struktúra mellett ez a dekompozíciós eljárás adja a saját modellem lényegi újítását, amely akár additív módon alkalmazva is egyértelmű javítást jelent a szakirodalomban látott dekompozíciókhoz képest. Az egyedi rész modellje szintén különbözik a benchmark módszerétől, de ez az előzőekhez képest kevésbé jelentős eltérés.

⁴Az egyetlen kivétel az egy lépéses forgalom érték előrejelzésének átlagos MSE* értéke, amely az additív esetben volt alacsonyabb.

A javasolt multiplikatív modell a benchmark cikk által legfontosabbnak tartott dinamikus stratégia értékelése alapján minden egyes részvényre jobb eredményre vezet a szakirodalom legjobb modelljénél. A benchmark cikk egyedi rész változatától függően az általam javasolt multiplikatív modellre történő áttérés átlagosan 13,6%-os, illetve 61,9%-os javulást eredményez. Mindezt a benchmark cikkhez képest 33-szor kevesebb adat felhasználásával sikerült elérni, hiszen a teljes piaccal szemben a modellem csak az adott részvény forgalmából számol. A benchmark cikkben használt adatbázishoz viszonyítva az általam használt adatbázis több, mint 9-szer nagyobb.

3.3. Áttekintés

Tekintsük át még egyszer röviden a dolgozat főbb eredményeit.

1. Azonos adatokon (időszak és eszközök) azonos módon értékelve összevetettem a szakirodalomban megtalálható napon belüli forgalom előrejelző modelleket annak érdekében, hogy kiderüljön, a tudomány jelenlegi állása szerint melyik a legjobb modell. Vizsgálataim alapján Bialkowski et al. (2008) modellje a szakirodalomból megismerhető legjobb napon belüli részvény forgalom előrejelző modell.
2. Mások által eddig nem alkalmazott új eljárást (polinom illesztés) javasoltam a forgalom U alakjának dekomponálására, amely a szakirodalomban található dekompozíciós eljárásokhoz képest jobban teljesít a forgalom előrejelző modellek részeként.
3. Az ármozgás mutatók vizsgálataim alapján nem használhatóak fel érdemben egy jobb forgalom előrejelző modell kialakításában, amely ellentétes a forgalom leíró jellemzését adó szakirodalom alapján kialakított előzetes várakozásaimmal.
4. Javasoltam egy, a szakirodalomban látottaktól jelentősen eltérő saját modellt a forgalom napon belüli előrejelzésére, amely a fent említett új dekompozíciós eljárást multiplikatív struktúrában ($U_alak \cdot egyedi_rész$) alkalmazza. Az így felírt modell egyértelműen jobban szerepel a benchmarknál a vizsgált 12 szempont mindegyike szerint, melyek tartalmazzák a benchmark cikk által felállított szempontokat is. Az általam javasolt modell mindezek mellett lényegesen kevesebb adat felhasználásával dolgozik, hiszen a teljes piac minden részvénye helyett csak az adott részvény forgalmát használja.

A rendelkezésemre álló adatbázis 33 részvény 130 hónapos időszakának napon belüli adatait tartalmazza. Ez lényegesen nagyobb a szakirodalomban korábban használt adatbázisoknál, amely kedvez az eredmények robusztusságának.

4. Főbb hivatkozások

A dolgozatban 89 hivatkozás szerepel. Ezek közül a forgalom előrejelzéssel közvetlenül foglalkozókat tartom a legfontosabbnak, melyek az alábbiak.

Bialkowski, J., Darolles, S., and Le Fol, G. (2006). How to reduce the risk of executing VWAP orders? - New approach to modeling intraday volume. Working paper.

Bialkowski, J., Darolles, S., and Le Fol, G. (2008). Improving VWAP strategies: A dynamic volume approach. *Journal of Banking & Finance*, 32:1709–1722. DOI: 10.1016/j.jbankfin.2007.09.023.

Brownlees, C. T., Cipollini, F., and Gallo, G. M. (2011). Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics*, 9:489–518. DOI: 10.1093/jjfinec/nbq024.

Kaastra, I. and Boyd, M. S. (1995). Forecasting futures trading volume using neural networks. *The Journal of Futures Markets*, Vol. 15, No. 8,:953–970. DOI: 10.1002/fut.3990150806.

Lux, T. and Kaizoji, T. (2004). Forecasting volatility and volume in the Tokyo stock market: The advantage of long memory models. Economics working paper, Christian-Albrechts-Universität Kiel, Department of Economics.

Lux, T. and Kaizoji, T. (2007). Forecasting volatility and volume in the Tokyo Stock Market: Long memory, fractality and regime switching. *Journal of Economic Dynamics & Control*, 31:1808–1843. DOI: 10.1016/j.jedc.2007.01.010.

5. A témakörrel kapcsolatos saját publikációk

Szűcs, B. A. (2010). Liquidity in practice: Interview series with market players. In *Annual Financial Market Liquidity Conference, Budapest*.

Szűcs, B. A. (2012). A tőzsdei forgalom napon belüli előrejelzése. In *Budapesti Corvinus Egyetem Közgazdasági Doktori Iskola VIII. Éves Konferenciája*.

Szűcs, B. A. (2013). Versengő modellek a tőzsdei forgalom napon belüli előrejelzésére. In *Budapesti Corvinus Egyetem Közgazdasági Doktori Iskola IX. Éves Konferenciája*.

Szűcs, B. A. and Illés, F. (2015). *Mastering R for quantitative finance*, chapter 3. Forecasting volume, pages 59–75. PACKT Publishing, Birmingham, UK.

Szűcs, B. A. and Váradi, K. (2014). Measuring and managing liquidity risk in the Hungarian practice. *Society and Economy*, Volume 36, Number 4:543–563. DOI: 10.1556/socec.36.2014.4.6.