

Szűcs Balázs Árpád

A tőzsdei forgalom napon belüli előrejelzése

Befektetések és Vállalati Pénzügy Tanszék

Témavezető: Makara Tamás, PhD

Copyright © Szűcs Balázs Árpád

Budapesti Corvinus Egyetem

Általános és Kvantitatív Közgazdaságtan Doktori Iskola

A tőzsdei forgalom napon belüli előrejelzése

Doktori értekezés

Szűcs Balázs Árpád

Budapest, 2015

Tartalomjegyzék

| | |
|--|-----------|
| 1. Bevezetés | 11 |
| I. Elméleti háttér | 15 |
| 2. A kutatás elhelyezése | 16 |
| 2.1. Piaci mikrostruktúra | 16 |
| 2.2. Likviditás | 18 |
| 2.3. Árhatás | 20 |
| 2.4. Optimális kereskedési stratégiák | 21 |
| 2.5. Kereskedési stratégiák a gyakorlatban | 22 |
| 2.6. VWAP kereskedés | 27 |
| 2.7. Áttekintés | 29 |
| 3. A forgalom előrejelzés irodalma | 31 |
| 3.1. A volumen/forgalom jellemzése | 31 |
| 3.1.1. A kereskedési aktivitás mérőszámai | 31 |
| 3.1.2. Stilizált tények | 34 |
| 3.2. Naiv előrejelző módszerek | 36 |
| 3.3. Fejlettebb előrejelző módszerek | 37 |
| 3.3.1. Kaastra-Boyd | 37 |
| 3.3.2. Lux-Kaizoji | 38 |
| 3.3.3. Bialkowski-Darolles-Le Fol | 40 |
| 3.3.4. Brownlees-Cipollini-Gallo | 46 |
| 3.3.5. Áttekintés | 52 |

| | |
|---|-----------|
| TARTALOMJEGYZÉK | 2 |
| II. Adatok és hipotézisek | 54 |
| 4. Adatok bemutatása | 55 |
| 4.1. Nyers adatok | 55 |
| 4.2. Felhasznált adatok | 56 |
| 5. Kutatási kérdések, hipotézisek | 60 |
| III. Benchmark kiválasztása | 63 |
| 6. Előkészítés | 64 |
| 6.1. Hibamérték választása | 64 |
| 6.2. A becslés és előrejelzés közös részletei | 65 |
| 6.3. Az U-módszer | 67 |
| 7. A BDF modell | 68 |
| 8. A BCG modell | 70 |
| 8.1. A becslés cikkben is közölt részletei | 70 |
| 8.1.1. A becslés menete | 70 |
| 8.1.2. Kezdő értékek | 71 |
| 8.2. A becslés további kifejtése | 72 |
| 8.2.1. A bővített egyenletek felírása | 72 |
| 8.2.2. További kezdő értékek | 74 |
| 8.2.3. Deriváltak | 74 |
| 8.2.4. Célfüggvény | 75 |
| 8.3. Becslések | 76 |
| 8.3.1. A θ induló értékei | 77 |
| 8.3.2. Különböző változatok | 78 |
| 9. Benchmark választás | 85 |
| IV. Saját modellek | 87 |
| 10. Modell keresés | 88 |
| 10.1. U dekompozíció nélküli modellek | 88 |

| | |
|--|------------|
| 10.1.1. Egyszerű AR modellek | 89 |
| 10.1.2. A hét napja hatás | 90 |
| 10.1.3. Ármozgás mutatók | 92 |
| 10.1.4. Áttekintés | 94 |
| 10.2. Az U alak leválasztásának különböző módjai | 95 |
| 10.2.1. U-módszer | 95 |
| 10.2.2. Polinom | 97 |
| 10.2.3. Polinom exponenciális súlyozással | 103 |
| 10.2.4. Spline függvény | 107 |
| 10.2.5. Áttekintés | 114 |
| 10.3. Az egyedi rész előrejelzésének különböző módjai | 116 |
| 10.3.1. Késleltetett ármozgás használata | 117 |
| 10.3.2. Egyidejű ármozgás használata (gondolat kísérlet) | 124 |
| 10.3.3. Előrejelzett ármozgás használata | 126 |
| 10.3.4. ARMA változatok | 136 |
| 10.3.5. GARCH változatok | 138 |
| 10.3.6. Áttekintés | 140 |
| 10.4. Egyéb lehetőségek | 142 |
| 10.4.1. U dekompozíció - egyedi rész nélkül | 142 |
| 10.4.2. Egy korrekciós modell | 144 |
| 10.4.3. Egy multiplikatív modell | 145 |
| 10.4.4. Áttekintés | 150 |
| 10.5. A modellkeresés eredménye | 151 |
| 11. Az értékelés további szempontjai | 156 |
| 11.1. További lehetőségek a modellek értékelésére | 156 |
| 11.1.1. Az értékelés tárgya | 157 |
| 11.1.2. Az értékelés módja | 159 |
| 11.2. Egész napos előrejelzés | 162 |
| 11.2.1. Forgalom értékek | 163 |
| 11.2.2. Statikus stratégia | 165 |
| 11.3. Dinamikus stratégia | 166 |
| 11.4. Áttekintés | 168 |
| 12. Eredmények áttekintése | 172 |

| | |
|--|------------|
| V. Záró gondolatok | 178 |
| 13.Összefoglalás | 179 |
| 13.1. Elméleti háttér (I. rész) | 179 |
| 13.2. Adatok és hipotézisek (II. rész) | 181 |
| 13.3. Benchmark kiválasztása (III. rész) | 182 |
| 13.4. Saját modellek (IV. rész) | 185 |
| 14.A dolgozat főbb eredményei röviden | 193 |
| 15.További kutatási kérdések | 195 |
| Irodalomjegyzék | 197 |
| Melléklet | 206 |
| A témakörrel kapcsolatos saját publikációk | 220 |

Ábrák jegyzéke

| | | |
|-----|--|-----|
| 1. | Átváltás a likviditás három dimenziója között | 20 |
| 2. | A klasszikus ajánlat feldarabolási probléma | 22 |
| 3. | A volumen napon belüli előrejelzésének relevanciája | 29 |
| 4. | A dinamikus SETAR alkalmazásának javító hatása | 45 |
| 5. | Húsz napos becslési és egy napos | 66 |
| 6. | Az U-módszer eredménye | 98 |
| 7. | Az U-módszer, a BDF modell U alakja és a negyedfokú polinom | 99 |
| 8. | Minimális MSE értéket eredményező n fokszámok | 100 |
| 9. | Poli(n)_AR modellből számolt $MSE_{arány}$ | 101 |
| 10. | A (78) súlyfüggvény értékei $J = 20$ nappal számolva | 104 |
| 11. | Minimális MSE értéket eredményező n fokszámok részvényenként | 105 |
| 12. | A Kraft Foods Inc. forgalmának különböző | 108 |
| 13. | Az U alak dekomponálásának különböző módjai | 115 |
| 14. | Az egyedi rész előrejelzésének különböző módjai | 140 |
| 15. | Az U alak dekomponálásának különböző módjai | 185 |
| 16. | Az U-módszer, a BDF modell U alakja és a negyedfokú polinom | 186 |
| 17. | Az egyedi rész előrejelzésének különböző módjai | 188 |

Táblázatok jegyzéke

| | | |
|-----|--|----|
| 1. | A kereskedési aktivitás lehetséges mérőszámai | 33 |
| 2. | A volumen napon belüli U alakja különböző piacokon | 35 |
| 3. | Bialkowski et al. (2008) eredményei portfólióra | 45 |
| 4. | Brownlees et al. (2011) eredményei | 51 |
| 5. | Volumen előrejelző modellek | 52 |
| 6. | Az adatok szerkezete (Boeing Co. példája) | 56 |
| 7. | A többségnél rövidebb adatsorú részvények | 56 |
| 8. | Rendellenes napok a mintában | 58 |
| 9. | A kereskedés felfüggesztése miatt kihagyott napok | 58 |
| 10. | Hiányzó 15 perces adatpontot tartalmazó napok | 59 |
| 11. | A felhasznált adatbázis néhány jellemzője | 59 |
| 12. | A BDF_AR modell és az U-módszer előrejelzésének összevetése . | 68 |
| 13. | A BDF_SETAR modell és az U-módszer előrejelzésének összevetése | 69 |
| 14. | A BDF_AR és BDF_SETAR modellek előrejelzésének összevetése | 69 |
| 15. | A BDF_AR és a BCG_D előrejelzésének összevetése | 78 |
| 16. | A BDF_AR és a BCG_egybecsles előrejelzésének összevetése . . | 80 |
| 17. | A BDF_SETAR és a BCG_egybecsles előrejelzésének összevetése | 80 |
| 18. | Az U-módszer és a BCG_egybecsles előrejelzésének összevetése . | 80 |
| 19. | A BCG_egybecsles és a BCG_check előrejelzésének összevetése . | 83 |
| 20. | A BDF_AR és a BCG_check előrejelzésének összevetése | 83 |
| 21. | A BDF_SETAR és a BCG_check előrejelzésének összevetése . . . | 83 |
| 22. | Az U-módszer és a BCG_check előrejelzésének összevetése | 84 |
| 23. | Az AR_26 és az U-módszer előrejelzésének összevetése | 89 |
| 24. | Az AR_1.26 és az U-módszer előrejelzésének összevetése | 90 |

| | | |
|-----|---|-----|
| 25. | Az AR_1.26 és a BDF_AR előrejelzésének összevetése | 90 |
| 26. | Hét napja hatás a mintában 33 részvényre | 91 |
| 27. | Az AR_1.26 és az ARH_1.26 előrejelzésének összevetése | 92 |
| 28. | Az ármozgás és a forgalom kapcsolata | 93 |
| 29. | Az ARX_1.26 modellek értékelése | 94 |
| 30. | Az U_AR és U_SETAR modellek összevetése a BDF_AR modellel | 96 |
| 31. | Az U_AR és U_SETAR modellek összevetése a BDF_SETAR mo- dellel | 97 |
| 32. | A Poli(14)_AR és Poli(14)_SETAR modellek | 102 |
| 33. | A Poli(14)_AR és Poli(14)_SETAR modellek | 102 |
| 34. | A Poli_expw(7)_AR és a Poli(14)_AR modell összevetése | 106 |
| 35. | A Poli_expw(7)_SETAR és a Poli(14)_SETAR modell összevetése | 106 |
| 36. | A Poli_expw(14)_AR és a Poli(14)_AR modell összevetése | 107 |
| 37. | A Poli_expw(14)_SETAR és a Poli(14)_SETAR modell összevetése | 107 |
| 38. | Becslendő paraméterek száma (M) a különböző spline változatokban | 111 |
| 39. | Átlagos MSE értékek a Spline(NnKk)_AR változatokra (bal), | 112 |
| 40. | Átlagos MSE értékek a Spline(NnKk)_SETAR változatokra (bal), | 112 |
| 41. | A Spline(N4K6)_AR és a Poli(14)_AR modell összevetése | 113 |
| 42. | A Spline(N4K6)_SETAR és a Poli(14)_SETAR modell összevetése | 113 |
| 43. | A (81) módon számolt hiba arányok | 114 |
| 44. | A Poli(14)_ARX modellek értékelése | 118 |
| 45. | Szignifikáns Granger-okságot mutató 20 napos időszakok | 119 |
| 46. | A Poli(14)_ARX_Gr modellek értékelése | 121 |
| 47. | A késleltetett ármozgás és a forgalom közötti 20 napos korrelációk | 122 |
| 48. | A Poli(14)_ARX_korr modellek értékelése | 123 |
| 49. | A Poli(14)_ARX(!) modellek értékelése | 126 |
| 50. | A Poli(14)_f(AR) modellek értékelése | 127 |
| 51. | A Poli(14)_f(AR_D) modellek értékelése | 129 |
| 52. | A Poli(14)_f(AR_1.26) modellek értékelése | 130 |
| 53. | A Poli(14)_f(U) modellek értékelése | 131 |
| 54. | A Poli(14)_f(Um) modellek értékelése | 132 |
| 55. | A Poli(14)_f(ARMA) modellek értékelése | 132 |
| 56. | A Poli(14)_f(ARMA_D) modellek értékelése | 133 |
| 57. | A Poli(14)_f(GARCH) modellek értékelése | 135 |
| 58. | A Poli(14)_ARMA és a Poli(14)_AR modell összevetése | 137 |

| | | |
|-----|---|-----|
| 59. | A Poli(14)_ARMA_AIC és a Poli(14)_ARMA | 138 |
| 60. | A Poli(14)_ARMA_SIC és a Poli(14)_ARMA | 138 |
| 61. | A Poli(14)_GARCH(AR) és a Poli(14)_ARMA | 139 |
| 62. | A Poli(14)_GARCH(ARMA) és a Poli(14)_ARMA | 140 |
| 63. | Dekomponált U alakok összevetése | 142 |
| 64. | Dekomponált U alakok értékelése egyedi rész használata nélkül . . | 143 |
| 65. | Dekomponált U alakok összevetése | 143 |
| 66. | A korrekciós modellek értékelése | 146 |
| 67. | A multiplikatív modellek értékelése ARMA egyedi résszel | 147 |
| 68. | A multiplikatív modellek értékelése ARMA egyedi résszel | 147 |
| 69. | Az egyes modellek hibái az Altria Group Inc. részvényére | 148 |
| 70. | Az egyes részvények átlagos 15 perces forgalma a teljes időszakra . | 149 |
| 71. | A multiplikatív modellek összevetése | 150 |
| 72. | A Poli(14)_ARMA és a Poli(14)_mult | 154 |
| 73. | A Poli(14)_ARMA és a BDF_(SET)AR | 154 |
| 74. | A Poli(14)_mult és a BDF_(SET)AR | 155 |
| 75. | VWAP értékelések MSE és MAPE alapon | 161 |
| 76. | VWAP kereskedés szempontjából a két eset pontosan egyforma . . | 162 |
| 77. | A Poli(14)_ARMA modell értékelése | 164 |
| 78. | A Poli(14)_mult modell értékelése | 164 |
| 79. | Az U-módszer értékelése | 165 |
| 80. | A Poli(14)_ARMA modell értékelése | 166 |
| 81. | A Poli(14)_mult modell értékelése | 167 |
| 82. | A Poli(14)_U modell értékelése | 167 |
| 83. | A Poli(14)_ARMA modell értékelése | 168 |
| 84. | A Poli(14)_mult modell értékelése | 169 |
| 85. | Az U-módszer értékelése | 169 |
| 86. | Az egyéb értékelési szempontok eredményeinek áttekintése | 170 |
| 87. | A Poli(14)_mult és BDF_(SET)AR modellek összevetése | 174 |
| 88. | A Poli(14)_mult és BDF_(SET)AR modellek összevetése | 174 |
| 89. | A Poli(14)_mult és BDF_(SET)AR modellek összevetése | 175 |
| 90. | A Poli(14)_mult és BDF_(SET)AR modellek összevetése | 175 |
| 91. | Adatbázisok méreteinek összevetése | 177 |

| | | |
|------|---|-----|
| 92. | Forgalom előrejelző modellek | 190 |
| 93. | Bialkowski et al. (2008) eredményei az egyes részvényekre | 208 |
| 94. | Bialkowski et al. (2008) eredményei az egyes részvényekre (folyt) . | 209 |
| 95. | Bialkowski et al. (2008) eredményei az egyes részvényekre (folyt.(2)) | 209 |
| 96. | Felhasznált adatok | 210 |
| 97. | Felhasznált adatok (folyt.) | 211 |
| 98. | Felhasznált adatok (folyt.(2)) | 211 |
| 99. | A Poli(14)_ARX_Gr modellek értékelése | 214 |
| 100. | A Poli(14)_ARX_Gr modellek értékelése | 214 |
| 101. | A Poli(14)_ARX_korr modellek értékelése | 215 |
| 102. | A Poli(14)_ARX_korr modellek értékelése | 215 |
| 103. | A Poli(14)_ARX_korr modellek értékelése | 216 |
| 104. | A Poli(14)_ARX_korr modellek értékelése | 216 |
| 105. | A Poli(14)_ARX_korr modellek értékelése | 217 |
| 106. | A Poli(14)_ARX_korr modellek értékelése | 217 |
| 107. | A Poli(14)_ARX_korr modellek értékelése | 218 |
| 108. | A Poli(14)_ARX_korr modellek értékelése | 218 |
| 109. | A részvények átlagos ára az Altria Group Inc. arányában (Arány), | 219 |

Köszönetnyilvánítás

Ezúton is szeretnék köszönetet mondani azoknak, akiknek a legnagyobb szerepe volt abban, hogy elkészítettem ezt a munkát. A nélkülözhetetlen szakmai segítségükért Makara Tamásnak, Berlinger Edinának, Barra Istvánnak és Kisvarga Józsefnek. A töretlen támogatásukért Berlinger Edinának és a szüleimnek.

1. fejezet

Bevezetés

A tőzsdei forgalom *rejtély*. 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 feladarozás 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 Weild (2015) alapján érdemes egy harmadik szempontot is megemlíteni. A NASDAQ korábbi alelnöke szerint az elsődleges részvénykibocsátások (*Initial Public Offering*, IPO) egyértelműen hozzájárulnak a gazdasági növekedéshez. Ennek 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. Weild szerint ennek magyarázata az automatikus kereskedés megjelenésében kere-

sendő. A magyarázat szerint az automatikus kereskedésnek 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.

Jelen dolgozat 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.

A dolgozat 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 adaton való reprodukálását követően olyan modelleket keresni, amelyek azoknál jobb eredményre vezetnek. A dolgozat eredményei között olyan modell javaslatok szerepelnek, amelyek a felhasznált harminchárom darab NYSE és NASDAQ részvény közel tizenegy éves, percenkénti sűrűségű adatbázisára alkalmazva, több szempont

szerint vizsgálva határozott javulást jelentenek az irodalom által javasolt legjobb modellekkel szemben.

A dolgozat 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.

I. rész

Elméleti háttér

2. fejezet

A kutatás elhelyezése

Ebben a fejezetben kontextusba helyezem a kutatást, egyúttal visszatérve arra a kérdésre, hogy miért releváns a volumen napon belüli előrejelzésével foglalkozni. A választ legvilágosabban az alábbi fogalom sorozat ismertetésén keresztül kaphatjuk meg.

2.1. Piaci mikrostruktúra

A volumen előrejelzése alapvetően valamilyen szervezett piacon, tőzsdén belül értelmezhető. A szereplők viselkedését alapjaiban meghatározza az a szabályrendszer, amely keretein belül a piac működik, és amelyet piaci mikrostruktúrának nevezhetünk. A különböző piacok természetesen nem egyeznek meg a legapróbb részletekig a mikrostruktúrájukban, azonban a legtöbb értéktőzsde a tiszta ajánlati könyves piac némileg személyre szabott változatát használja.

Ajánlati könyves piac

Egy korábbi kutatásban foglalkoztunk a piaci mikrostruktúrákkal (l. Havran et al. (2011)), jelen dolgozat szempontjából az alábbiakat érdemes kiemelni a témakörből.

Technikai jellemzés

Ajánlati könyves piacon (limit order market) a szereplők megbízásai az ajánlati könyvbe futnak be, feldolgozásuk ezen keresztül történik. A megbízásoknak két

alaptípusát különböztetjük meg:

1. Limit áras megbízás vagy limit ajánlat (limit order). A limit ajánlatok a könyvben várakoznak arra, hogy teljesüljenek, vagy törlődjenek. Négy fontos információt tartalmaznak: egy árat, egy mennyiséget, egy előjelet (vétel vagy eladás), és a beadás pontos időpontját. Opcionálisan tartalmazhatnak lejárati időpontot is, aminek elérésekor a törlésük automatikus, nem kell a beadónak külön kezdeményeznie. Az előjelüktől függően kerülnek be a könyv bid (vételi limit ajánlatok) vagy ask (eladási limit ajánlatok) oldalára. A limit ajánlatokra, illetve az azokat beadó szereplőkre tekinthetünk úgy, mint a piaci likviditás biztosítóira (liquidity provider).
2. Piaci áras megbízás vagy piaci ajánlat (market order). A piaci ajánlatok csak egy mennyiséget és egy előjelet határoznak meg. A beadás pillanatában azonnal teljesítésre kerülnek, azáltal, hogy összepárosítják őket a várakozó, ellentétes előjelű limit ajánlatokkal. A piaci ajánlatokat, illetve azok beadóit a likviditás felhasználóinak (liquidity taker) tekinthetjük.

Minden egyes tranzakció tehát limit és piaci ajánlatok párosításával születik. A párosítási mechanizmus többféle lehet, de abban mindegyik megegyezik, hogy a limit ajánlatokat elsősorban ár szerint rendezi sorba, és a piaci ajánlat számára automatikusan a legkedvezőbb limit ajánlato(ka)t választja ki. Eltérés abban lehet, hogy az azonos áron lévő limit ajánlatokból milyen sorrend szerint választ. A leggyakoribb a FIFO (First-In-First-Out) párosítási mechanizmus, amely esetén elsőbbséget élvez a régebb óta várakozó limit ajánlat.

Bid-ask spread

A bid és ask ár sosem egyezik pontosan meg, a különbségüket bid-ask spreadnek nevezik. A spread létezése miatt ajánlati könyves piacon nem alakulhat ki egyetlen piactisztító ár, amin bárki vehet vagy eladhat. Aktuális árnak azt szokás tekinteni, amin a legutolsó tranzakció végbement.

Parlour és Seppi (2008) észrevétele szerint a likviditást biztosító szereplők, miután ők ex-ante specifikálják az ajánlat paramétereit, információs hátrányban vannak a likviditást felhasználó szereplőkkel szemben, akiknek az ajánlata azonnal teljesül. Ezt az egyének korlátozott információ feldolgozó képességével magyarázzák. Azt ugyan nincs okunk feltételezni, hogy a limit ajánlatot beadók információ

feldolgozó képessége gyengébb lenne a piaci ajánlatot beadókénál, ugyanakkor a limitet beadó szereplővel szemben az összes többi potenciális piaci ajánlatot adó szereplő megjelenik, és ők együttesen több információval rendelkezhetnek (már csak azért is, mert frissebb információ birtokában vannak), és vélhetően többet is tudnak feldolgozni, mint bármelyik egyedi szereplő.

Ez az információs hátrány két kedvezőtlen hatással is fenyeget a likviditás biztosítókra nézve. Az egyik, hogy egy új beérkező információ hatására a limit ajánlat sokkal kedvezőtlenebbé válik a beadója számára, de még mielőtt törölhetné, lecsap rá egy informált piaci ajánlat. Erre a jelenségre az angol néven szoktak hivatkozni (picking-off risk), fordíthatjuk pl. learatási kockázatnak. Az említett információs hátrány másik lehetséges költsége a teljesülés bizonytalanságából adódik, amit interpretálhatunk a pénz időértékeként is. Ugyanis nem tudni, hogy mikor teljesül az ajánlat, sőt, azt sem, hogy teljesül-e egyáltalán (a piac könnyen elmozdulhat olyan irányba, hogy a limit ajánlatnak gyakorlatilag semmi esélye sorra kerülni).

Ezen információs hátrány egy lehetséges magyarázattal szolgál a bid-ask spread jelenségére: ez a likviditás biztosítók jutalma a piac számára nyújtott *szolgáltatásukért*.

2.2. Likviditás

A likviditás széles körben használatos pénzügyi fogalom, a pontos jelentése azonban némileg változhat a konkrét felhasználás függvényében. Az alábbi áttekintés egy olyan korábbi kutatás felhasználásával készült, melyben magam is részt vettem (Berlinger et al. (2011)).

Brunnermeier és Pedersen (2008) megkülönböztet finanszírozási (funding) és piaci (market) likviditást. Úgy tekint a befektetőre, mint egy mérleggel rendelkező vállalatra, és a mérleg mindkét oldalára definiálja a likviditást. Ilyen szóhasználat mellett a finanszírozási likviditás a finanszírozás (pénz szerzés) *könnyűségét* méri, ami a forrás oldal likviditása. A piaci likviditás ezzel szemben az eszköz oldal likviditása, azt méri, hogy az eszközökkel milyen *könnyű* kereskedni. A kettő között természetesen szoros kapcsolat van, hiszen egyrészt, ha likvidek az eszközök, könnyebb az eladásukból pénzhez jutni, másrészt viszont olyan szereplő képes piaci likviditást biztosítani egy adott eszközben, amelynek megfelelő finanszírozás áll a rendelkezésére.

Jelen dolgozat szempontjából természetesen a piaci likviditás a fontosabb, mely önmagában is több értelmezési lehetőséget kínál.

1. Piacképesség: egy likvid eszköz minimális veszteség mellett gyorsan pénzre váltható. Fontos a *minimális veszteség* kitétel, hiszen megfelelően alacsony ár mellett a legtöbb eszköz viszonylag gyorsan eladható.
2. Bouchaud et al (2008) megközelítése szerint a piaci likviditás azt méri, hogy az ár mennyire mozdul el egy fix méretű kötés eredményeképpen. Likvid piacon azok a kötések, melyek nem szokatlanul nagyok, csak kicsit, vagy egyáltalán nem mozdítják el az eszköz árát.
3. Tekintheünk a piaci likviditásra úgy is, mint a kötések gyakorisága. Ez alapján az adott idő alatt több kötetést produkáló piac likvidebbnek tekinthető.

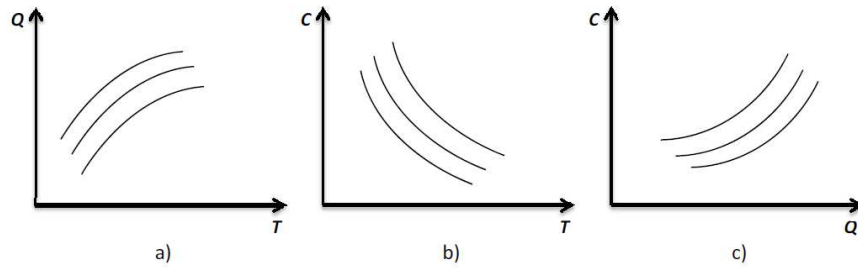
Mindhárom értelmezésben közös, hogy az eszközt (annak piacát) akkor tekinti likvidnek, ha az könnyen kereskedhető. De pontosan mit nevezünk könnyűnek? Alapvetően két aspektusra gondolunk ilyenkor.

- **Idő.** Minél rövidebb a várakozási idő az ajánlat beadása és (minimális veszteséggel történő) teljesülése között, annál likvidebb az eszköz.
- **Árhatás.** Minél nagyobb volumennel tudunk adott idő alatt kereskedni anélkül, hogy elmozdítanánk az árat, vagy másként: minél kevésbé mozdítja el az árat egy adott volumenű kötés, annál magasabb a likviditás.

Ez utóbbi szemléletet tovább finomítva, ha egy piaci szereplő ügyletet kíván kötni, három dimenzió mentén kell gondolkoznia, melyek az alábbiak:

- Q : a megkötni kívánt ügylet mérete (volumene),
- T : a maximális időtáv, ami alatt a teljes ügyletet le kell bonyolítani,
- C : rögzített Q és T mellett az esetleges árhatás okozta veszteség.

Ilyen megközelítésben a piaci likviditást a fenti változók közötti átváltáson (trade-off) keresztül is megadhatjuk. Képletszerű összefüggések nélkül végiggondolva, egy-egy változót rögzítve megállapíthatjuk a másik kettő kapcsolatának az irányát, melyet az 1. ábra érzékeltet. Példaként az 1/b. ábra azt mutatja, hogy



1. ábra. Átváltás a likviditás három dimenziója között
Rögzített a) árhatás, b) volumen, és c) időtáv esetén.
Forrás: Berlinger et al. (2011) módosított változata

rögzített volumen mellett minél tovább tudunk várni, annál alacsonyabb árelmozdító hatásra számíthatunk. (Fontos megjegyezni, hogy az ábra csak illusztráció, ezért a sugallt konvexitási/konkavitási tulajdonságokat mindhárom esetben külön bizonyítani kellene.)

2.3. Árhatás

Az árhatásról mint a szűkös piaci likviditás következményéről fentebb már esett szó, de jelentősége miatt érdemes külön is kiemelni. Ajánlati könyves piacokon csak piaci ajánlat beadása okozhat elmozdulást az árban, melyet az árhatás függvényekkel írhatunk le. Meg szokás különböztetni modellezési szempontból érdekes virtuális árhatás függvényt, és gyakorlati jelentőséggel bíró empirikus (valós) árhatás függvényt. (Bouchaud et al (2008))

A virtuális árhatás függvény az alábbi általános alakot ölti:

$$S(v) = Av \tag{1}$$

ahol A konstans, v pedig a volumen. Ez a megközelítés egy v méretű piaci ajánlat beadásának várható árelmozdító hatását mutatja, azzal a feltevessel, hogy statikus az ajánlati könyv, tehát más szereplők nem lépnek, amíg a tranzakció teljesen le nem zárul. A rugalmasság érdekében meg szokás engedni az aszimmetriát is, mely esetben vételi és eladási ajánlat esetén más A paraméterrel számolhatunk.

Az empirikus árhatás függvényt, mely ex-post megfigyelhető a piacon, gyakran formalizálják az alábbi alakban (a virtuális árhatás függvény esetén említett aszimmetriát itt is célszerű lenne beépíteni A és ψ tekintetében, de a tömörség

érdekében ezt itt is elhagyom):

$$S_e(v) = Av^\psi \quad (2)$$

ahol a ψ paraméter tipikus értéke 0,3-0,4 körüli (Bouchaud et al (2008)), a függvény tehát konkáv, vagyis a nagyobb ajánlatok relatíve kisebb elmozdulást okoznak. Alapvetően két összetevő határozza meg az értékét:

1. különböző volumenek esetén várható árváltozás,
2. annak a valószínűsége, hogy az ár egyáltalán megváltozik.

Formálisan megfogalmazva:

$$S_e(v) = \Pr(+ | v) \cdot E[r | v] \quad (3)$$

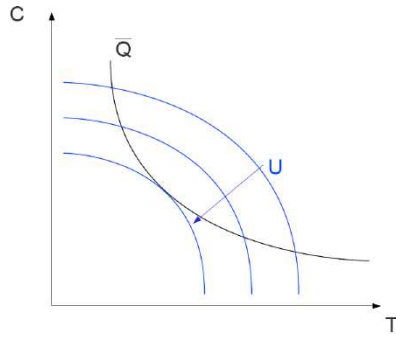
ahol $\Pr(+)$ annak a valószínűsége (konkáv eloszlás függvényként értelmezhető), hogy nullától különböző r mértékben elmozdul az ár.

2.4. Optimális kereskedési stratégiák

Mivel a likviditás egy piacon sem végtelen, mindig létezik egy olyan minimális ajánlat méret, ami felett már árhatásra kell számítani. Ezt a tényt a piaci szereplők nem hagyhatják, és a gyakorlatban nem is hagyják figyelmen kívül. Ennek oka, hogy jelentős megtakarításra tehetnek szert vétel esetén, és jelentős veszteségtől eshetnek el eladás esetén, ha a kereskedési stratégiájukban figyelembe veszik az ajánlatuk árelmozdító hatását, ami minden esetben az ajánlat beadója ellen irányul.

A fejezetben bevezetett jelöléseket használva Berlinger et al. (2011) alapján a következő módon ragadhatjuk meg ezt a problémát.

Az 1. ábrán szemléltetett átváltásokat inputként használva megadható egy Q , T és C változókat felhasználó $f(Q, T, C) = 0$ felület, amiből a piaci szereplő kiválaszthat egy megvalósítandó pontot. Ez a felírás preferenciáktól független kereskedési scenáriókat jelöl, melyekből a piaci szereplő megkeresheti az egyéni $U = g(Q, T, C)$ hasznossági függvénye segítségével a számára legkedvezőbbet. Ezek alapján az alábbi feladatot kell megoldania:



2. ábra. A klasszikus ajánlat feldarabolási probléma
 Forrás: Havran et al. (2011)

$$U = g(Q, T, C) \rightarrow \max_{Q, T, C} \quad (4)$$

$$f(Q, T, C) = 0$$

Az optimális kereskedési stratégia ezen elméleti modell segítségével úgy vizsgálható, hogy külön modellezzük a piaci likviditás kérdését (amit az $f(Q, T, C) = 0$ felület ír le) az egyéni viselkedést eredményező preferenciák kérdésétől (amit az $U = g(Q, T, C)$ hasznossági függvény ragad meg).

Szemléltetésként tegyük fel, hogy a piaci szereplő számára külső adottság a \bar{Q} volumen, és így kell döntenie C és T értékéről (ez a helyzet könnyen előállhat, különösen brókerek vagy alapkezelők munkája során). Ebben az esetben a (4) feladat az alábbi formára egyszerűsödik:

$$U = g(\bar{Q}, T, C) \rightarrow \max_{T, C} \quad (5)$$

$$f(\bar{Q}, T, C) = 0$$

Ezen korlátozott eset egy lehetséges reprezentációját láthatjuk a 2. ábrán, melyet a klasszikus ajánlat feldarabolási (order splitting) problémának tekinthetünk.

2.5. Kereskedési stratégiák a gyakorlatban

A fentebb bemutatott elméleti modell jól megragadja a piaci szereplők problémáját, azonban az U hasznosság függvényt, az $f(Q, T, C)$ felületet, valamint az 1. ábrán illusztrált összefüggéseket konkrét formára hozni komoly kihívást jelenthet.

Egy (zömében magyarországi székhelyű) piaci szereplőkkel folytatott interjú sorozatban (Szűcs és Váradi (2014)) többek között azt is vizsgáltuk, hogy a fentiekből mi az, amit valóban figyelembe vesznek a piaci szereplők, és ez alapján melyek azok a kereskedési stratégiák, amelyeket a gyakorlatban alkalmaznak.

Alapvetően az (5) problémára kell leggyakrabban valamilyen választ adniuk, ami gyakorlatilag időzítési problémaként értelmezhető, előbb azonban még azt is el kell dönteniük, hogy milyen megbízás típust választanak.

Megbízás típusok

A legtöbb piaci szereplő kihasználja az adott piaci mikrostruktúra adta lehetőségeket, ennek ellenére megfigyelhető néhány jellegzetesség a különböző típusú piaci szereplők tekintetében.

Brókerek

A brókerek gyakran kapnak olyan megbízást, amely a napi kereskedett volumen felső határát határozza meg. Ezt általában a teljes piac arányában szokták megadni. Egy $\frac{1}{3}$ -os megbízás esetén például az a feladat, hogy minden piaci tranzakcióra egy újabbal reagáljon a bróker, melynek volumene a megfigyelt tranzakció fele. Az ilyen típusú megbízás könnyen extrém árakhoz vezethet, ha egyszerre több (példánkban legalább három) jelenik meg az adott napon. Az ilyen esetek elkerülése érdekében az arány mellett egy árat is meg szoktak határozni, amit ha elér a piac, a kötések fel kell függeszteni, ezáltal elkerülve, hogy nagyon kedvezőtlen tranzakciók köttessenek. Az ilyen típusú megbízások viszonylag illikvidebb piacokon és OTC (over-the-counter, vagyis tőzsdén kívüli) piacokon gyakoribbak.

Ezzel szemben likvidebb piacokon gyakori a VWAP (Volume Weighted Average Price, azaz volumennel súlyozott átlagár) megbízás, ami azt kéri a brókertől, hogy adott időszak alatt a piacon érvényes VWAP-ot érje el. A megadott időszak lehet például 10h és 12h között, vagy gyakran a megbízástól a nap végéig. Az ilyen megbízást általában $VWAP + x$ bázispontért vállalják el, tehát az ügyfél fix árat fizet, a bróker pedig saját eredményességétől függően akár x bázispontnál többet is kereshet.

Alapkezelők

A fenti két megbízás típust az alapkezelők is használják, azonban van még néhány másik, ami a brókerekre kevésbé jellemző.

Az egyik ilyen a *stop-loss* megbízás, amely inaktív mindaddig, amíg a megadott árat el nem éri a piac, viszont ha ez megtörténik, azonnal zárja a pozíciót.

A másik a *market-on-close* megbízás, amely piaci ajánlatot jelent a napi zárás előtti pillanatban, vagyis gyakorlatilag napi záróáron. Többféle motiváció is állhat egy ilyen megbízás mögött. Egyrészt lehet a záróárra vonatkozó manipulatív célja, másrészt viszont a benchmark követő alapok ezáltal tudják legegyszerűbben biztosítani, hogy az aznap beérkezett tőke követési hiba (tracking error) nélkül kerüljön be az alapba (ugyanis általában záróáron értékelik az ilyen alapokat).

A harmadik a *jéghegy* (iceberg) megbízás, amely akkor érdekes, ha olyan nagy a kereskedési igény, hogy azt már érdemes elrejteni a piac többi szereplője elől. Ha az alapkezelő például nagy mennyiségben szeretne vásárolni, jéghegy megbízást adhat, ilyenkor a bid oldalon (limit áras vételi ajánlat) csak a volumen egy kisebb része jelenik meg. Ha ezt elviszik, egy következő rész kerül a helyére. Ha nagy az eladói érdeklődés, és kiürül a legjobb árszint, az algoritmus rögtön betesz egy bidet a következő árszintre. Ha pedig nagy volumen található az ask oldalon, az algoritmus *felnyúl érte* piaci ajánlattal.

A negyedik a blokk megbízás, amelyet tőzsdén kívül kötnek. Ha az alapkezelőnek nagy a kereskedési igénye, és nem szeretné jelentősen elmozdítani az árat, ugyanakkor sürgős az ügylet, megpróbálhat a tőzsdén kívül találni valakit, aki éppen ellentétes ügyletben érdekelt. Ha ez sikerül, megegyeznek az árról és a mennyiségről.

Időzítés és egyéb kérdések

Brókerek

Az egyik lehetséges stratégia a piac napi életciklusát követi, ahol a kereskedett volumen napi eloszlásának megfelelően osztják el a saját tranzakcióikat is. Ez a megközelítés illikvid piacokon népszerűbb, ahol kevesebb korrekciós lehetőség marad arra az esetre, ha a bróker várakozásától eltérően alakul a piac.

Likvidebb piacon a bróker számára kevesebb kockázattal jár, ha megpróbál eltérni a piaci aránytól a saját várakozásai tükrében. Ennek egyik változata, ha

nagy volumennel indítja a napot. Ilyenkor, ha sikerül megverni a piacot, akkor az akár jelentős mértékű is lehet, ha mégsem, akkor pedig rendelkezésre áll a teljes nap, hogy megpróbálja visszatalálni a VWAP-hoz.

A kereskedett volumen napon belüli U alakja hangsúlyos jelenség (különösen részvényt piacon), amit figyelembe kell venni a kereskedés során is. Az U alak arra a megfigyelésre utal, hogy a volumen magas nyitás után, nap közepén leesik, majd zárás előtt ismét megugrik. Némelyek ezt az emberek életritmusával magyarázzák: az új információra kereskedéssel reagálnak a nap elején, intézik az egyéb feladataikat nap közben, majd nap végén zárják a pozíciókat. Természetesen, ha lényeges információra derül fény nap közben, az módosíthat ezen, illetve más magyarázatok is elképzelhetőek. Például, ha valaki nagy megbízást darabol, nem egyenletesen teszi ezt a nap folyamán, hanem néhány nagyobb szeletet betesz a piacra nyitáskor, aztán vár, hogy avuljon el az az információ, hogy valaki adagol (ugyanis azt ki lehet használni az adagoló kárára), majd nap végén újabb nagy szelettel kereskedik. Végül bizonyos piacokon (pl. Törökország) ebédidőben hivatalosan is szünetel a kereskedés.

Több lehetséges stratégia is van a már említett VWAP elérésére (vagy megverésére), de a teljes volumen egyszeri piaci megbízásként való beadása általában nincs köztük. Általánosságban a limit áras megbízás olcsóbb, de az ügyfél nem mindig tudja megvárni, hogy ezek teljes mértékben teljesüljenek. Ilyenkor a bróker limit áras megbízással indít, és amikor már kezd kifutni az időből, átvált piaci megbízásra, hogy lezárja az ügyletet. A tapasztalat szerint ez a stratégia kb. a spread felébe kerül.

Kellően likvid piacokon az optimális kereskedési stratégia kialakítására felhasználhatóak az árhatás függvények. Az algoritmusnak azonban érdemes figyelembe vennie a többi szereplő reakcióját is, hiszen azok észrevehetik az adagolást, és ezt kihasználhatják, amint arról már esett szó. Az algoritmusnak ezért törekednie kell arra is, hogy ne legyen könnyen felismerhető.

Illikvid piacokon a kereskedés könnyen egy póker játszamához válhat hasonlónvá, ahol erősen korlátozott számú szereplő van egyszerre jelen. Ilyenkor a brókernek (vagy gyakrabban a saját számlára kereskedő spekulánsnak) információs előnye lehet, és ezt kihasználhatja. Ugyanakkor fontos, hogy saját stratégiáját leplezze, a többiekét pedig megpróbálja kitalálni, hiszen az alacsony résztvevőszám miatt ezen múlhat a siker. Erősen illikvid piacokon ezért akár csak amiatt is beadhatnak bizonyos ajánlatokat, hogy megtévesszék a többi szereplőt. Amint korábban

említettem, részben ez is magyarázat lehet a napon belüli U alakra.

Az időtáv és a hatékonyság összefüggését az egyik megkérdezett az alábbi módon fogalmazta meg:

„Minél sűrűbben kereskedik valaki, annál inkább előtérbe kerül a matematika és a statisztika. Emiatt az időskála nagyon fontos. Aki például egy másodperces időközönként kereskedik, azt a hatékonyságot használja ki, amit az követ el, aki mondjuk csak egy napos időközönként köt ügyletet. Az viszont annak a hibáját használja ki, amit az követ el, aki egy hónapos időközönként kereskedik, stb. Minél hosszabb az időtáv, annál inkább a fundamentális elemzés kerül előtérbe.”

Alapkezelők

A technikai alapon működő alapkezelők gyakran azért kezelik a likviditási kockázatot, hogy a leglikvidebb eszközökben kereskednek. Ezáltal általában nem kell az árhatás miatt aggódniuk a megbízásaik teljes piaci volumenhez való alacsony aránya miatt. A saját ajánlat méretükhöz képest illikvid piacokon jellemzően a limit áras ajánlatokat preferálják.

A fundamentális alapon működő alapkezelők gyakran választanak illikvidebb piacokat, ahol nagyobb a potenciális félreárazás mértéke. Ilyen piacokon nem szokatlan az olyan kereskedési terv, amely jelentősen elmozdítaná az árat, ha egyben zúdítaná a piacra. Erre egy bevett megoldás, hogy akkora darabokra osztják fel a megbízást, amekkora egyébként jellemző a piacon, ezáltal próbálva elrejtetni a nagy kereskedési igényüket. Hajlandóak akár hetekig gyűjtögetni egy-egy papírt, ha csak így tudják elkerülni, hogy az árat jelentősen elmozdítsák. Ugyanakkor arra is törekszenek, hogy feleslegesen ne ijesszék meg (vagy vezessék félre) a piacot nagy tranzakciókkal. A már említett blokk ügyletet is használják, illetve olyan algoritmust, amely folyamatosan a legjobb áron tartja az ajánlatukat a könyben. Végül említettek egy időzítési heurisztikát, miszerint, ha egy „*small cap*”-ben (alacsony kapitalizációjú cég részvénye) hirtelen megjelenik egy nagy volumenű limit áras ajánlat, az általában azt jelenti, hogy valakinek elfogyott a türelme, és berakta a maradékot egyben, amit emiatt érdemes elvinni, mert nem valószínű, hogy lesz folytatása.

A benchmark követő alapkezelőknek is szüksége lehet a megbízások feldarabolására, ha adott napon nagy mennyiségű pénz érkezik be. Ilyenkor előfordul,

hogy az alapkezelő tapasztalata alapján úgymond *ad hoc* darabolnak, vagy rábízák a brókere ezt a feladatot. Utóbbi esetben a VWAP-pal megelégszenek, de gyakran ennél jobbat sikerül elérni. Máskor a kereskedési platformjuk által felajánlott módszerrel darabolnak, amely adott (pl száz darabos) csomagokban adja be különböző piacokra a megbízásokat. Végül elő szokott fordulni, hogy bizonyos okokból sürgősen ki kell igazítani a portfólió összetételét, és nincs idő megvárni, hogy teljesüljenek a limit áras megbízások, ilyenkor a piaci ajánlatot kell választaniuk. Egyikük humorosan hozzátette:

„Meg egyébként sem akarunk estig bent maradni, arra várva, hogy az USA [piaca] bezárjon.”

2.6. VWAP kereskedés

Ha eltekintünk attól, hogy a befektetési döntések meghozatalához valamilyen elképzelést érdemes kialakítani az ár várható alakulásáról (melyre különböző iskolák épülnek, pl. technikai és fundamentális elemzők), és ezzel nem foglalkozva explicit módon adottnak vesszük a kereskedési igényt, a fenti elméleti és gyakorlati kereskedési stratégiák megfelelő végrehajtásához még így is további információ szükséges.

Ez az információ alapvetően a piaci likviditás helyzetéről és várható alakulásáról szól, amely több módon is testet ölthet, nem csak a fent említett különböző árhatás függvények, de számos egyéb likviditási mérőszám formájában is (l. Szűcs és Váradi (2014)), melyek bemutatása meghaladja e dolgozat kereteit. Kiemelten a napon belül előrejelzett forgalom is fontos input lehet a legtöbb stratégia számára, amely sok mérőszámhoz valamilyen módon kapcsolható is.

Berkowitz et al. (1988) megmutatja, hogy egy passzív kereskedő (passive trader) optimális teljesítményének értékelésére az általuk bevezetett VWAP egy megfelelő benchmarkot jelent (ők napi VWAP-ot javasoltak, azóta szokás rövidebb időszakra is értelmezni). Éppen emiatt a VWAP kereskedés, melyről (pontosabban a VWAP megbízásról, amely az ügyfél szempontjából közelíti meg a jelenséget) korábban már esett szó, nagy népszerűségnek örvend a kereskedők körében. Bialkowski et al. (2008) szerint manapság az intézményi befektetők kereskedésének kb. fele VWAP kereskedésben nyilvánul meg, tehát valóban elterjedt stratégiáról van szó.

Brownlees et al. (2011) definíciója szerint a VWAP kereskedés az a folyamat, melynek keretében egy nagyobb megbízást kisebb darabokra osztanak fel, mely megbízás darabokat a nap folyamán különböző árakon teljesítik oly módon, hogy az átlagos ár lehetőleg minél közelebb essen a VWAP-hoz. A stratégia az egyszerűsége mellett további előnyökkel is jár, hiszen a viszonylag türelmesebb befektetők némi időbeli kockázatot vállalva jelentősen növelhetik az ügyletkötés nyereségességét az árhatás elkerülésén keresztül.

A forgalom előrejelzés szempontjából a VWAP kereskedés azért bír kiemelt jelentőséggel, mert ez az a stratégia, amely pusztán a forgalom előrejelzésre határozva tökéletesen végrehajtható, semmilyen más inputra (pl. az áralakulás vagy az ajánlati könyv modellezésére) nincsen szükség. Tökéletes forgalom előrejelzés esetén a VWAP is teljes pontossággal elérhető.

A fenti, elsőre talán meglepő állítás minden piaci szereplőre érvényes, a piachoz viszonyítva kicsikre és nagyokra is. Annyi finomítás fűzhető az állításhoz, hogy ha egy szereplő képes a napi forgalomhoz mérhető méretű tranzakciót bonyolítani, akkor gyakorlatilag maga képes alakítani a VWAP-ot. Egy ekkora szereplőnek még a forgalom előrejelzésre mint egyetlen inputra sincsen igazán szüksége a VWAP kereskedéshez, a szereplők döntő többsége azonban nem esik ebbe a kategóriába.

Megbízható napi forgalom előrejelzés birtokában tehát egy (a piacnál lényegesen kisebb) kereskedőnek nincs más dolga, mint a saját megbízás méretét ugyanolyan arányban felosztani a nap folyamán, mint amilyen eloszlásra a teljes piac forgalmával kapcsolatban számít. Ebben az esetben az áralakulástól teljesen függetlenül nap végére éppen a piaci VWAP lesz a saját tranzakcióinak volumennel súlyozott átlagos ára (persze feltéve, hogy a várakozása tökéletesen megvalósult).

Ennek illusztrálására álljon itt az alábbi, Bialkowski et al. (2008, pp. 3) által hozott rövid példa.

„Vegyünk egy olyan egyszerű pénzügyi piacot, amelyen napjában csak három alkalommal lehet tranzakció. A nap végén azt látjuk, hogy 5000 részvéennyel kereskedtek. Az első időpontban 2500 darabbal 100-as áron, a második időpontban 1000 darabbal 101-es áron, a harmadik időpontban pedig 1500 darabbal 102-es áron. A napi VWAP ekkor:

$$\frac{100 \cdot 2500 + 101 \cdot 1000 + 102 \cdot 1500}{5000} = 100,8.$$

Tegyük fel, hogy az Y kereskedőnek a napi volumen $\frac{1}{100}$ részének megfelelő volumenben kellett kötnie aznap, és tudta előre, hogy a vo-



3. ábra. A volumen napon belüli előrejelzésének relevanciája

Forrás: Saját szerkesztés

lumen eloszlása 2500 (50%), 1000 (20%) és 1500 (30%) lesz.

Ha feltesszük, hogy a kereskedésének nincsen árhatása, akkor egyszerűen rendre 25, 10 és 15 darabbal kell kereskednie. Ekkor az általa elért átlagos ár: $\frac{100 \cdot 25 + 101 \cdot 10 + 102 \cdot 15}{50} = 100,8$, ami megegyezik a VWAP-pal.

Most tegyük fel, hogy nem tud árhatás nélkül kereskedni, és az ő tevékenysége nyomán 101-re, 101,5-re és 103-ra változik az ár. Ekkor a saját átlagára: $\frac{101 \cdot 25 + 101,5 \cdot 10 + 103 \cdot 15}{50} = 101,7$, ami magasabb annál a VWAP-nál, ami akkor áll elő, ha ő nem lép piacra. Ugyanakkor a VWAP-ot ex-post számolják, és minden tranzakció benne van, még-hozzá a megfigyelt árakon értékelve. Vagyis a VWAP ebben az esetben a következő: $\frac{101 \cdot 2500 + 101,5 \cdot 1000 + 103 \cdot 1500}{5000} = 101,7$, vagyis Y-nak árhatás mellett is sikerült elérnie a VWAP-ot.”

Ezek a számítások azt feltételezik, hogy az előrejelzett volumenben az Y kereskedő saját tranzakciója is benne van, de ha a sajátja nem lenne benne, megfelelően módosítva a számítást természetesen ugyanerre az eredményre jutnánk.

2.7. Áttekintés

A fejezet a releváns fogalmak sorra vételén keresztül áttekintette a forgalom napon belüli előrejelzésének elhelyezkedését és jelentőségét az elméletben és a gyakorlatban. A gondolatmenet vázát a 3. ábra is szemlélteti.

Alapvetően a forgalom előrejelzésének kérdése szervezett piacokon merül fel, vagyis tőzsdéken, melyeknek a szabályrendszere (mikrostruktúrája) meghatározza a piac jellegzetességeit és a szereplők viselkedését is. Minden piacon fontos kérdés a (piaci) likviditás, melynek jelenléte *könnyen* kereskedhető termékeket eredményez,

hiánya esetén azonban a tranzakciókhoz árhatás kapcsolódik, amely kedvezőtlenül hat a megbízások beadóira.

Miután a likviditás soha nem lehet végtelen, azaz bármely piacon meghatározható egy olyan tranzakció méret, amely felett már árhatásra kell számítani, ezért a szereplőknek érdemes a kereskedési stratégiáikban a likviditási szempontot is figyelembe venni. Ezt a feladatot megragadhatjuk az elmélet eszközeivel, és formalizálhatjuk optimalizálási feladatként, amely segít a probléma jobb megértésében (természetesen az itt bemutatottaknál sokkal összetettebb modellek is építhetők, melyek több szempontot képesek megvilágítani). A piaci szereplőknek ugyanezt a problémát a gyakorlatban is kezelniük kell, tehát olyan stratégiákra van szükségük, melyek inputja általánosságoktól mentes, és az akcióterv is lebontható megbízásokra.

A gyakorlat sokféle válasza közül az egyik népszerű a VWAP kereskedés, mely a VWAP jó benchmark szerepe miatt igen elterjedt stratégia (az intézményi tranzakciók kb. fele ilyen indíttatású, de nem csak az intézményi befektetők kedvelik). A dolgozat szempontjából az adja a jelentőségét, hogy habár a volumen előrejelzés a legkülönbözőbb stratégiáknak fontos eleme lehet, a VWAP kereskedésnek ez az egyetlen inputja (értelemszerűen a kereskedési igény mellett), az előrejelzés pontossága éppen ezért döntő jelentőségű ezen elterjedt stratégia sikeressége szempontjából.

3. fejezet

A forgalom előrejelzés irodalma

A tőzsdei részvények árfolyama, hozama, és hozamának szórása (volatilitás) régóta a tudományos és gyakorlati élet szakemberei érdeklődésének a középpontjában áll. Ennek megfelelően a jellemzésükkel, modellezésükkel és előrejelzésükkel foglalkozó szakirodalom is meglehetősen kiterjedt.

A forgalom esetében ugyanez már kevésbé mondható el. A 3.1. pontban foglalkozunk a forgalom lehetséges mérőszámaival, jellemző tulajdonságaival, stilizált tényeivel, amihez még viszonylag bőségesen állt rendelkezésre forrás. A modellezésnél viszont nagyon gyakran annyiban számolnak csak a forgalommal, amennyiben az hozzájárul más változók, elsősorban a hozam vagy a volatilitás előrejelzéséhez. A 3.2 és 3.3.pontokban áttekintjük a szakirodalom azon (bőségesnek nem nevezhető) részét, amely saját jogon vizsgálja a forgalom előrejelzését.

3.1. A volumen/forgalom jellemzése

Mivel a dolgozat fő fogalmai a volumen és a forgalom, ebben a pontban áttekintjük ezek kapcsolatát, fontosabb mérőszámaikat és stilizált tényeiket.

3.1.1. A kereskedési aktivitás mérőszámai

Volumen alatt az adott időszakban kereskedett részvény darabszámot értjük, míg a forgalom a volumen százalékos változatát jelenti.

Ez alapján egyedi részvényre az alábbi módon számolhatunk (Darolles és Le

Fol, 2003):

$$x_{it} = \frac{V_{it}}{TSO_{it}} \quad (6)$$

ahol x a forgalom, V a volumen, TSO pedig a teljes forgalomban lévő részvény darabszám (Total Shares Outstanding). Az egyes részvényeket $i = 1, 2, \dots, N$ jelöli, a vizsgált időszak (pl. egy nap vagy fél óra) indexe pedig t .

Az adott t időszakra kiszámolhatjuk a volumennel súlyozott átlagos árat is (Volume Weighted Average Price, VWAP):

$$P_{it} = \frac{\sum_n P_{itn} V_{itn}}{\sum_n V_{itn}} \quad (7)$$

ahol n az adott időszak alatt létrejött tranzakciók indexe. Látható, hogy (6) esetén, ha darabszám helyett értékkel számolnánk, vagyis bővítenénk a törtet P_{it} -vel, az eredmény számszerűen és mértékegységében is változatlan maradna, emiatt a forgalom egyedi részvény esetén értékben mérve is értelmezhető.

Portfólió esetén kicsit más a helyzet. Ott pusztán a darabszám alapján számolni félrevezető lenne, ugyanis ki kell szűrnünk annak a hatását, hogy az egyes részvények értéke esetleg jelentősen eltérő lehet. Ezért portfólió esetén a forgalom számítása a VWAP figyelembe vételével célszerű, tehát érték alapon, vagyis az alábbi módon (Darolles és Le Fol, 2003):

$$x_{it}^p = \frac{\sum_i P_{it} V_{it}}{\sum_i P_{it} TSO_{it}} \quad (8)$$

ahol a p index a portfólióra utal. Természetesen elképzelhető, hogy adott t időszakban nincsen minden részvényben kötés, ilyenkor a megfelelő V -k értéke nulla.

Számunkra a fenti definíciók lesznek hasznosak, ugyanakkor meg kell jegyezni, hogy egyéb javaslatok is születtek a kereskedési aktivitás mérésére, melyeknek egy rövid áttekintése következik az alábbiakban Lo és Wang (2009) alapján, megtartva a fenti jelöléseket.

1. **Egyedi volumen.** A fenti tárgyalásban használt V_{it} .

| # | Mérőszám | Tanulmány |
|---|-------------------------------------|---|
| 1 | Egyedi volumen | Epps és Epps (1976), James és Edmister (1983) Lamoureux és Lastrapes (1990, 1994) Andersen (1996) |
| 2 | Aggregált volumen | Ying (1966), Gallant et al (1992) Hiemstra és Jones (1994) |
| 3 | Egyedi érték | James és Edmister (1983) Lakonishok és Vermaelen (1986) |
| 4 | Aggregált érték | - |
| 5 | Relatív egyedi érték | Tkac (1999) |
| 6 | Egyedi forgalom | Morse (1980), Lakonishok és Smidt (1986), Richardson et al (1986), Bamber (1986) Stickel és Verrechia (1994), Hu (1997) |
| 7 | Aggregált forgalom | Smidt (1990), LeBaron (1992), Campbell et al (1993) |
| 8 | Kötések száma | Conrad et al (1994) |
| 9 | Kereskedési napok száma (évente) | James és Edmister (1983) |

1. táblázat. A kereskedési aktivitás lehetséges mérőszámai
Forrás: Lo és Wang (2009, pp. 4.)

2. **Aggregált volumen.** A vizsgált részvények egyedi volumenjeinek aggregálásával kapható:

$$\sum_i V_{it} \quad (9)$$

3. **Egyedi érték.** Árral súlyozott volumen:

$$V_{it}P_{it} \quad (10)$$

4. **Aggregált érték.** Árral súlyozott volumen, a vizsgált részvényekre aggregálva:

$$\sum_i V_{it}P_{it} \quad (11)$$

5. **Relatív egyedi érték.** Az egyedi érték osztva az aggregált piaci értékkel:

$$\frac{V_{it}P_{it}}{\sum_{j=1}^k V_{jt}P_{jt}} \quad (12)$$

ahol k a piacon lévő részvények száma.

6. **Egyedi forgalom.** A fenti tárgyalásban használt x_{it} .

7. **Aggregált forgalom.** A vizsgált részvények egyedi forgalmainak aggregálásaként adódik:

$$\sum_i x_{it} \quad (13)$$

8. **Kötések száma.** Az adott időszakban létrejött tranzakciók száma.

9. **Kereskedési napok száma.** Éves viszonylatban számolva.

Az 1. összefoglaló táblázatban az egyes mutatókkal dolgozó szerzők szerepelnek.

3.1.2. Stilizált tények

„A stilizált tény kifejezést a közgazdaságtanban olyan empirikus jellegzetességekre használják, amelyek olyannyira konzisztensen megfigyelhetők (például eszközök, piacok, időszakok széles skáláján), hogy már tényként kezelik őket. Általánosságukból eredően gyakran kvalitatív természetűek.” (Sewell, 2011 pp. 2.). Másként úgy is fogalmazhatunk, hogy a stilizált tények statisztikailag igazak, de ettől még előfordulhatnak olyan adatok is, amelyeken nem figyelhetők meg. Ebben a fejezetben a kereskedési volumen stilizált tényeit tekintjük át.

A volumen napon belüli U alakja

A volumen stilizált tényei közül leggyakrabban a napon belüli U alakot emelik ki, melyet sokan megfigyeltek már az NYSE-n, valamint különböző egyéb tőzsdéken is (lásd 2. táblázat). Megfigyeléseik szerint a volumen közvetlenül nyitás után és zárás előtt lényegesen magasabb, mint a nap többi szakaszában, melynek oka részben a magasabb átlagos tranzakciónkénti volumen, részben pedig a sűrűbb kötések (Hmaied et al., 2006).

A hét napja hatás

A hét különböző napjain eltér a megfigyelhető átlagos volumen, vagyis hétfőn a legalacsonyabb, a hét közepén pedig magasabb, mint pénteken (Jain és Joh (1988), Foster és Viswanathan (1993), idézi Hmaied et al. (2006)). Ezt a jelenséget Lo és

| Tőzsde | Tanulmány |
|--------------------------|---|
| NYSE | Wood et al (1985), Harris (1986), Jain és Joh (1988) McInish és Wood (1992), Brock és Kleidon (1992) Foster és Viswanathan (1993) |
| Európa | Biais et al (1995), Abhyankar et al (1997), Bildik (2001) |
| Ázsia | Choe és Shin (1993), Lee et al (2001), Ding és Lau (2001) |
| Afrika | Hmaied et al. (2006) |
| Észak-Amerika (egyéb) | McInish és Wood (1990) |

2. táblázat. A volumen napon belüli U alakja különböző piacokon
Forrás: Hmaied et al. (2006), Hussain (2011)

Wang (2009) is dokumentálja, de szerintük nem jelentős a napok közötti eltérés. Chordia et al. (2001, idézi Sewell (2011)) a keddi magas és pénteki alacsony értéket emeli ki. Lakonishok és Maberly (1990, idézi Sewell (2011)) csak a hétfői alacsony értékről számol be.

Hosszú memória

Hosszú memóriájúnak (long memory) nevezünk egy folyamatot, ha benne a késleltetés növelésével hatványfüggvény-szerűen, tehát viszonylag lassan csökken az autokorreláció. Másként szólva az autokorreláció magas késleltetésre is szignifikáns, ezáltal egy-egy sokk sokáig érezteti a hatását (innen az elnevezés). Sewell (2011) az alábbi eredményeket emeli ki.

Lobato és Velasco (2000) a Dow Jones Iparági Átlag harminc részvényét elemezve a kereskedési volumenben hosszú memóriát figyel meg. Hasonló eredményre jut Plerou et al. (2000), Gopikrishnan et al. (2000), Plerou et al. (2001), Lo és Wang (2009), valamint Qiu et al. (2009) is. Eisler és Kertész (2007) szerint az NYSE részvényeinél a tranzakcióknak nem csak a volumene, de a gyakorisága is hosszú memóriájúnak tekinthető.

Hatványszerű eloszlás

Több forrás is dokumentálja, hogy a kereskedési volumen eloszlásának széle hatványszerűen esik, azaz a kiugró értékek is viszonylag gyakoriak (Plerou et al. (2000), Gopikrishnan et al. (2000), Plerou et al. (2004), idézi Sewell (2011)). Plerou et al. (2001) szerint aszimptotikusan hatványszerű az eloszlás széle.

Egyéb

Lo és Wang (2009) az NYSE és AMEX piacok 1962-1996 közötti forgalom adatainak vizsgálatából további észrevételeket is tesz.

Idősoros jellemzők indexek alapján:

- A forgalom az átlagához képest kevésbé volatilis, mint a hozamok (relatív szórása kisebb).
- Negyedévekre nézve stabilnak mondható a forgalom. A harmadik negyedévben a legalacsonyabb, de az eltérés nem jelentős.

Keresztmetszeti jellemzők egyedi részvények alapján.

- Harris és Gurel (1986) megfigyelése szerint egy részvény S&P 500 indexbe való bekerülése azonnal növeli a forgalmát. Ennek részben oka lehet az index arbitrázs (annak kihasználása, hogy az index futures ára és az alkotóelemek spot árai között az elméleti összefüggés nem tökéletesen teljesül), illetve az is, hogy az indexben való részvétel széleskörűbb tulajdonosi kört eredményez.
- A kisebb részvényeknek kisebb a forgalma (tehát nem csak a volumene), mint a nagyobb kapitalizációjúaknak, ami a mérethatékonyság miatt lehet, vagyis, hogy a nagy alapok nem aprózzák el a kereskedést.

Nemnegativitás. Végül említsük meg a volumen nemnegativitását is. Ez valószínűleg nem stilizált, hanem valódi tény, ami a piac szempontjából a volumen természetéből adódóan mindig teljesül, és ezért érdemes szem előtt tartani. A piac helyett egy-egy piaci szereplő szempontjából vizsgálva a volument, bizonyos alkalmazásokban szokás előjeles volument definiálni, de ez csak egy kézenfekvő összevonása két információnak, ugyanis így egy számban tudják megadni a volument, és azt, hogy vételi vagy eladási volt a szereplő szempontjából az ügylet. Erre a kényelmes összevonásra azonban éppen a valódi volumen nemnegativitása miatt nyílik lehetőség, jelen dolgozat szempontjából pedig semmi előnye nincsen.

3.2. Naiv előrejelző módszerek

Bevezetésként következzen néhány módszer, melyeknek egyszerűségükből adódóan külön cikket nem szenteltek. Ezeket a módszereket a 3.3. fejezetben bemutatott

modellek benchmarkként használják teljesítményük értékelésekor, utalva rá, hogy a gyakorlati életben előfordulnak a volumen előrejelzésére önállóan alkalmazva is. Itt röviden kiemelem őket az áttekinthetőség kedvéért.

Előző érték. A következő időszak előrejelzésének a legutóbbi értéket tekinti, vagyis: $\hat{y}_{t+1} = y_t$. Kaastra és Boyd (1995) használja ezt benchmarkként árutőzsdén havi adatokra.

Egyszerű átlag. A mintaátlagot tekinti előrejelzésnek a következő időszakra. Lux és Kaizoji (2007) használja értéktőzsdén napi adatokra.

U-módszer. Rekeszekre (egyenlő hosszúságú időszakokra, pl. 15 perc) osztva vizsgálva a napot, minden rekeszre a korábbi időszak megfelelő rekeszeinek átlagát várja, ezáltal reprodukálva a napon belüli U alakot, vagyis:

$$\hat{y}_{t+1} = \frac{1}{L} \sum_{l=1}^L y_{t+1-k \cdot l} \quad (14)$$

ahol L az átlagoláshoz használt megelőző napok száma, k pedig az egy napban lévő rekeszek (megfigyelések) száma. Ezt a módszert használja Bialkowski et al. (2008) és Brownlees et al. (2011) értéktőzsdén napon belüli adatokra.

3.3. Fejlettebb előrejelző módszerek

Fejlettebb módszerekként hivatkozva tekintsük át a fentieknél összetettebb, publikált modelleket is.

3.3.1. Kaastra-Boyd

Cél és adatok. Az első általam ismert munka, amely kizárólagosan a volumennel foglalkozik Kaastra és Boyd (1995). A szerzők 1977-1993 közötti havi aggregáltságú árutőzsdei adatokból (Winnipeg Commodity Exchange) készítenek havi előrejelzést a futures volumenre 9 hónapra előre. A vizsgált hat termék: árpa, repce, len, zab, rozs és búza 20 tonnás kontraktusai.

Módszertan. Az előrejelzést a neurális hálók módszertanának segítségével készítik, és az átlagos négyzetes hiba gyöke (Root Mean Squared Error, RMSE) illetve az átlagos abszolút százalékos eltérés (Mean Absolute Percentage Error, MAPE) használatával értékelik. Benchmark előrejelzésként két lehetőséget is figyelembe vesznek: egyrészt a legutóbbi értéket ($\hat{y}_t = y_{t-1}$), másrészt az ARIMA modellből nyert előrejelzést használják.

Eredmények. Az előrejelzési eredményeik változóak. A hat termékből két esetben a naiv előrejelzés (vagyis az előző érték) jobbnak bizonyult a neurális hálónál kilenc hónapos időtávra, és jobbnak minden esetben egy hónaposra. Az ARIMA modell inkább rövidebb előrejelzésre mutatkozott hatékonyának. A neurális hálók hosszabb időszakra pontosabbak voltak a maradék négy termékre, amit a szerzők sikerként értékelnek, figyelembe véve, hogy a két rosszul teljesítő termék a teljes kereskedett érték kb 10%-át adja.

3.3.2. Lux-Kaizoji

Cél és adatok. Lux és Kaizoji (2007) a volatilitás mellett külön foglalkozik a volumen előrejelzésével is. Az általuk használt adatbázis 1975.01.01. és 2001.12.31. közötti napi volumen adatokat tartalmaz a tokiói tőzsde (Tokyo Stock Exchange) több, mint 1000 különböző első szekciós részvényére. Munkájukban napi előrejelzéseket készítenek 1, 10, 20, \dots , 100 napra előre. A mintát időpontok szerint két részre osztják. Az első 1975-1985 időszakot használják a paraméterek becslésére, a második 1986-2001 időszakot pedig előrejelzésre. A minta kezelhetősége érdekében nem használnak fel minden részvényt, hanem két darab száz elemű kórsarat képeznek. Az egyikbe véletlenszerűen választanak részvényeket, a másikba a száz leglikvidebbet (legnagyobb átlagos kereskedett volumennel bíró részvények) teszik.

Módszertan. Abból a megfigyelésből indulnak ki, amiről már volt szó a 3.1.2. fejezetben, miszerint a volumen idősora hosszú memóriával bír. Elemzésük ennek megfelelően a hosszú memóriát figyelembe vevő modellekre fókuszál.

Benchmarknak egy rövid memóriájú modellt tekintenek, nevezetesen az ARMA(p, q) modellből nyert előrejelzést, $p \leq 5$ és $q \leq 5$ választás mellett. A megfelelő késleltetést minden részvényre az AIC kritérium alapján határozzák meg. Kiemelik, hogy

a választás szándékos a SIC kritériummal szemben, mert ez utóbbi szigorúbban bünteti a magas becslendő paraméter számot, márpedig a hosszú memória miatt érdemes lehet engedékenyebbnek lenni ez ügyben. Ennek ellenére megállapítják, hogy a két kritérium hasonló választást javasol a legtöbb esetben, amit a hosszú memóriájú modellek indokoltságaként értelmeznek.

Másrészt benchmarkként használják a becslési időszak mintaátlagát is (naiv előrejelzés).

Két alternatív hosszú memóriájú modellt javasolnak. Az egyik az ARFIMA(p,q,d) modell előrejelzése. Egyrészt a számításigény miatt, másrészt pedig azért, mert a magasabb rendű késleltetéseket a frakcionális differencia útján kívánják kezelni, $p \leq 1$ és $q \leq 1$ választással dolgoznak. A d paramétert Geweke és Porter-Hudak (1983) módszerével becslik.

A másik javasolt modell a Calvet és Fisher (2001, 2002) által közölt multifraktál (MF) modell Lux (2003) módosításával, ami lényegében egy sztochasztikus volatilitás modell. A volatilitás folyamat:

$$\theta_t = 2^k \prod_{i=1}^k m_t^{(i)} \quad (15)$$

ahol $m_t^{(i)}$ adott paraméterekkel rendelkező lognormális eloszlásból származó véletlen változó. Maga a volumen ekkor:

$$vol_t = \theta_t \chi_i \quad (16)$$

ahol χ_i egy skálázási paraméter, amire azért van szükség, mert az egyes részvények átlagos volumene eltér.

Az MF modell előnye a másik kettővel szemben, hogy nem engedi nulla alá a volument, míg ARMA és ARFIMA esetén kellett egy külön kikötés, miszerint, ha negatív lenne az előrejelzett volumen, akkor nullával helyettesítik azt. Ugyanakkor hozzáteszik, hogy ez a korlátozás nagyon ritkán volt effektív.

Az előrejelzéseket relatív MSE és relatív átlagos abszolút eltérés (Mean Absolute Error, MAE) alapján értékelik, ahol a relatív azt jelenti, hogy a mutatót osztják a naiv előrejelzésre kiszámolt mutatóval a részvények közti összehasonlíthatóság érdekében.

Végül megvizsgálják azt is, hogy az egyéni részvényekből nyert paraméterek

átlagával számolt modell milyen előrejelzést ad az egyes részvényekre vonatkozóan (pooled estimates).

Eredmények. Az MF modellt találták legjobbnak a legtöbb időtávon. A naiv előrejelzéshez képest (minta átlag) egy napra átlagban 53% illetve 33% javulást éreredményez relatív MSE és MAE alapon, 100 napra előre már csak 6-8% az előnye (egy korábbi változatban ennél némileg gyengébb javulást mutattak ki minden időtávra: Lux és Kaizoji (2004)). Az ARFIMA modell rövidebb távon a második legjobb, hosszabb időtávon viszont az ARMA jobban teljesít nála. De 1 napos időtávon az ARMA is jobb a naiv előrejelzésnél (hosszabb távon már nem).

Nem csak az átlagos hibát, de annak szórását is érdemes figyelni. Ez is MF esetén a legalacsonyabb, míg ARFIMA és ARMA esetben nagy kilengések, tehát nagy alkalmi tévedések is előfordulnak.

Végül a szerzők szükségét érzik magyarázattal szolgálni az ARMA modell viszonylag gyenge teljesítményére, mivel több korábbi tanulmány arra jutott (pl. Basak et al. (2001) és Man (2003)), hogy megfelelően illesztett ARMA modellek hasonló előrejelzéssel szolgálhatnak, mint amit a *valódi* mögöttes ARFIMA modellből kapnánk, tehát hosszú memóriájú folyamatoknál sem indokolt ez a rossz teljesítmény. A fő magyarázat az lehet, hogy a d paraméter magasnak tekinthető, vagyis a folyamat memóriája *meglehetősen hosszú*, amit már nehéz ARMA modellel lekövetni, mint ahogyan több cikk is kitért rá, hogy csak alacsonyabb d értéknél működhet jól az ARMA előrejelzés (Brodsky and Hurvich (1999), Crato and Ray (1996)).

Az átlagolt paraméterekkel számolt előrejelzés ARMA és ARFIMA esetén jelentett némi javulást, elsősorban nem az átlagban, hanem a nagy tévedések némileg alacsonyabb száma tekintetében, ugyanakkor a javulás összességében nem jelentős (a módszer inkább volatilitásnál hasznos). A javulás jelentősebb a likviditás alapján választott százas mintában.

3.3.3. Bialkowski-Darolles-Le Fol

Cél és adatok. Bialkowski et al. (2008) a VWAP kereskedés szolgálatában készít napon belüli volumen előrejelzést, melynek fő jellemzője, hogy az U alakot igyekszik megragadni mint a leghangsúlyosabb napon belüli stilizált ténynt. A vizsgálatot a CAC40 által 2004 szeptemberében tartalmazott minden részvényre

elvégzi. A felhasznált minta 2003. szeptember - 2004. augusztus közötti napon belüli adatokat tartalmaz 20 perces rekeszekbe aggregálva, így naponta 25 adatponttal dolgozik (nem használja a nyitás előtti és zárás utáni kötéseket). Volumen helyett a forgalommal dolgozik.

Módszertan. Több korábbi munka úgy kezeli a napon belüli szezonalitást, hogy alkalmas módon megszabadul tőle, pl. naptári idő helyett tranzakciós idő (event time) használatával (Engle (2000), Gouriéroux és Le Fol (1998)). Bialkowski et al. (2008) célja ezzel szemben modellbe foglalni ezt a stilizált ténytet, melyet az adatsorok szezonális és dinamikus komponensre bontásával tesz meg, kihasználva, hogy minden részvényre hasonló módon megfigyelhető az U alak. A szezonális (vagy piaci) komponens a forgalom várt szintjét hivatott megjeleníteni egy átlagos napon (minden részvényre egyedileg), a dinamikus (vagy egyedi) komponens pedig az ettől való várható eltérést mutatja. A dekompozíció additív módon történik, tehát a két komponens összege adja az előrejelzett forgalmat.

A dekompozíciót a mellékletben bemutatott magas dimenziójú faktorelemzés segítségével végzik, ahol a feladat felírását a (117) mutatja, azaz:

$$X = K + e \quad (17)$$

ahol $X_{(TxN)}$ a megfigyelt forgalom adatokat tartalmazza T megfigyelésre és N részvényre, K a közös vagy piaci komponens, e pedig az egyedi rész.

A közös komponens becslését a (121) által megadott módon kapjuk:

$$\tilde{K} = \tilde{F}\tilde{\Lambda}' \quad (18)$$

ahol \tilde{F} a faktorok, $\tilde{\Lambda}$ pedig a faktorsúlyok becsült mátrixa.

A \tilde{K} közös komponens X -hez hasonlóan (TxN) -es, és minden részvényre megfigyelhető benne a jellegzetes U alak. Azonban ez még csak a becslési időszakra vonatkozó szezonális rész. Az előrejelzéshez felteszik, hogy habár némi ingadozás van a piacon, a közös komponens alapvetően részvényenként stabil. Ennek megfelelően rekeszenkénti átlagot számolnak a becslési időszak közös komponenséből,

és ezt tekintik a közös rész előrejelzésének:

$$\tilde{K}_{t+1,i} = \frac{1}{L} \sum_{l=1}^L K_{t+1-25 \cdot l,i} \quad (19)$$

ahol L az átlagolt napok száma, 25 pedig a rekeszek száma naponta.

Az egyedi rész a becslési időszakra a (17) összefüggés alapján egy egyszerű kivonással kapható:

$$\tilde{\varepsilon} = X - \tilde{K} \quad (20)$$

melyből aztán két módszerrel is készítenek részvényenkénti előrejelzést.

Egyrészt az alábbi modell segítségével:

$$\tilde{\varepsilon}_{t,i} = c + \phi_1 \tilde{\varepsilon}_{t-1,i} + \varepsilon_{t,i} \quad (21)$$

ahol ε fehér zaj. Láthatjuk, hogy ez egy ARMA(1,0), azaz egy AR(1) folyamat.

A cikkben a (21) egyenletre annak felírásakor ARMA(1,1)-ként hivatkoznak. Később egyszerűen ARMA-ként emlegetik (ami ARMA(1,0)-ként értve már konzisztens), és az egyszerű ARMA(1,1) megnevezésen kívül semmi nem utal arra később sem, hogy használtak volna mozgóátlagolású tagot. A cikknek megtaláltam egy korábbi, kiadatlan (working paper) verzióját (Bialkowski et al. (2006)), amely jelen szempontból mindenben megegyezik a 2008-as publikált verzióval, így nem segített feloldani az ellentmondást. Mindezek miatt azt feltételezem, hogy az ARMA(1,1) elnevezésben van egy karakter elütés, az egyenletet pedig a megadott formában becsülték, és nem pedig azt, hogy az egyenlet felírásából hiányzik a $\theta_1 \varepsilon_{t-1,i}$ tag, aminek a paraméterét aztán szisztematikusan meg is becsülték, és helyesen hivatkoztak rá ARMA(1,1)-ként. Ez azért is tűnik jó feltevésnek, mert az alternatív modell (az alábbi SETAR) az AR(1)-nek közvetlenebb továbbfejlesztése, mint az ARMA(1,1)-nek (természetesen önmagában ez nem indokol semmit).

Másrészt az alábbi alakban felírt SETAR modellel:

$$\tilde{\varepsilon}_{t,i} = (c_{1,1} + \phi_{1,2} \tilde{\varepsilon}_{t-1,i}) I(\tilde{\varepsilon}_{t-1,i}) + (c_{2,1} + \phi_{2,2} \tilde{\varepsilon}_{t-1,i}) (1 - I(\tilde{\varepsilon}_{t-1,i})) + \varepsilon_{t,i} \quad (22)$$

ahol

$$I(x) = \begin{cases} 1 & \text{ha } x \leq \tau \\ 0 & \text{egyébként} \end{cases} \quad (23)$$

A forgalom előrejelzés végül a két előrejelzett komponens összege lesz:

$$\tilde{X}_{t+1,i} = \tilde{K}_{t+1,i} + \tilde{e}_{t+1,i} \quad (24)$$

Benchmarknak azt a stratégiát tekintik, amely a forgalom adatot az előző időszak megfelelő rekeszeinek az átlagaként jelzi előre, azaz:

$$\tilde{x}_{t+1,i} = \frac{1}{L} \sum_{l=1}^L x_{t+1-25 \cdot l,i} \quad (25)$$

módon. Ez láthatóan hasonló logikát követ, mint a (19) egyenlet esetén a közös rész előrejelzése.

Az előrejelzést mindig egy napra előre készítik el a megelőző húsz nap (ami egy naptári hónap) adatai alapján, és minden nap egyvel eltolják a becslésre használt ablakot, hogy elkészítsék a következő napi előrejelzést. Összesen ötven napra jeleznek ilyen módon előre.

Mivel kifejezetten a VWAP kereskedés szempontjából értékelik az előrejelzéseket, az alábbi három alternatív stratégiát vizsgálják.

1. Elméleti. Itt mindig egy rekeszre jeleznek előre, mindig felhasználva a beérkező legújabb információt is. Azért nevezik elméletinek, mert a napi összes volumen csak a nap végére derül így ki, ezért gyakorlatban nem kivitelezhető szerintük ez alapján kereskedni, mert az előrejelzett rekeszenkénti forgalmat csak nap végén tudják napi arányban kifejezni.
2. Statikus. Itt nap elején előrejelzik az egész napot, kizárólag a közös részt tekintve előrejelzésnek. (Az AR és SETAR előrejelzését nem használják fel, mert gyengén teljesítőnek találják hosszabb távra.)
3. Dinamikus. Hasonló az elméletihez, csak kiküszöböli annak hibáját, ami miatt az nem valósítható meg. Ennek érdekében előrejelzi az egész napot a nap elején, majd a beérkező új információk fényében lépésenként mindig frissíti az előrejelzést, míg a nap utolsó lépésében már valóban megegyezik azzal, amit az elméleti modell csinálna.

A pontosságot a VWAP és a kereskedés által elért volumennel súlyozott átlagár eltéréseként értelmezik a korábban már megismert átlagos abszolút százalékos eltérés (MAPE) segítségével. Rekeszenkénti árnak az adott 20 perc alatt teljesített kötések súlyozatlan számtani átlagát tekintik.

Eredmények. Az előrejelzésben elért eredmények értékelésénél a statikus módszerről nem sok szót ejtenek, mert gyenge teljesítményt vártak tőle, és ez be is igazolódott. Az egyedi részt előrejelző két modell közül a SETAR meggyőzően jobban teljesített az AR modellnél, így az előbbit elemzik részletesebben. Vizsgálják külön az egyéni részvények esetén elért teljesítményt, illetve azon portfólió teljesítményét is, amely az összes mintában lévő részvényt tartalmazza.

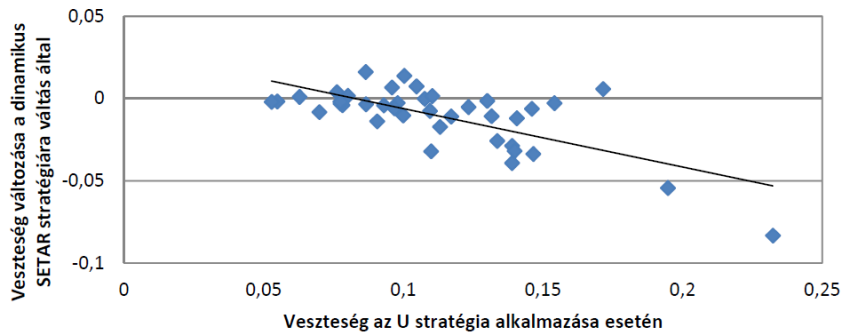
A melléklet 93-95. táblázata részvényenkénti bontásban mutatja a százalékban kifejezett MAPE értékeket három módszerre: a benchmark U-módszerre, valamint az elméleti és dinamikus SETAR módszerekre. Elsőként az elméleti módszert összevetve a benchmarkkal látható, hogy mind a harminckilenc esetben sikerült javulást elérni (az ötödik oszlop végig negatív értékeket mutat), ebből harminc esetben 1bp feletti mértékben. A maximális javulás 8,75bp (9. részvény).

A dinamikus SETAR módszer használata a benchmarkhoz képest harminc esetben eredményezett csökkenést a kereskedési költségben, melyből tizennégy esetben haladja meg a javulás az 1bp-ot, a legnagyobb fejlődés pedig ismét a 9. részvény esetén jelentkező 8,32bp. Abból a kilenc esetből, amikor romlást tapasztalunk, csak kettő mutatott 1bp feletti eltérést.

A gyengébben teljesítő, de megvalósítható dinamikus esetre megvizsgálták, hogy a benchmark által elkövetett hiba mértéke és a dinamikus módszer által elért javulás között milyen összefüggés áll fenn. Arra jutottak, hogy tendenciájában a nagyobb hiba esetén nagyobb javulást ér el az általuk javasolt modell, a két legnagyobb javulást pedig valóban a két legnagyobb hiba esetén figyelhetjük meg. Ez látható grafikusán az 4. ábrán.

Egyéni részvény szinten tehát összességében érdemes a javasolt modellre váltani, mert várható értékben mindenképpen jobb eredményt kapunk, hiszen a javulás mind esetszámban mind átlagos mértékben meghaladja a romlásokat, emellett pedig a nagy tévedések valószínűsége is alacsonyabb.

Végül megvizsgálták egy olyan portfólió esetén is az előrejelzésben elért teljesítményt, amelyben a CAC40 indexnek megfelelő arányban szerepelnek a részvények. Az aggregált eredményeknél az AR változatot is megjelenítik, amint a 3.



4. ábra. A dinamikus SETAR alkalmazásának javító hatása a klasszikus U stratégiával szemben
 Forrás: Bialkowski et al. (2008)

| Módszer | Átlag | Szórás |
|-----------------|--------|--------|
| Elméleti SETAR | 0.0770 | 0.0942 |
| Elméleti AR | 0.0833 | 0.0956 |
| Dinamikus SETAR | 0.0898 | 0.0954 |
| Dinamikus AR | 0.0922 | 0.0994 |
| U stratégia | 0.1006 | 0.1171 |

3. táblázat. Bialkowski et al. (2008) eredményei portfólióra Százalékban értendő MAPE értékek
 Forrás: Bialkowski et al. (2008)

táblázat is mutatja. Ebből is látható, hogy a klasszikus megközelítés 10bp körüli hibájával szemben a dinamikus 9bp, az elméleti 8bp körüli hibát eredményez, ami megközelítőleg 10% és 20% javulás. Megfigyelhetjük továbbá, hogy a SETAR jobb eredményt ad az AR modellnél mind a hiba átlaga, mind a szórása tekintetében.

Utolsó megjegyzésként megemlítik, hogy ha nem a MAPE értékeket átlagolnák, hanem előbb átlagolnák a hibákat, és azután néznék meg a MAPE értéket, még kedvezőbb eredményt kapnának, hiszen az előjeles hibák kiolthatják egymást, ami a MAPE értékekről nem mondható el.

3.3.4. Brownlees-Cipollini-Gallo

Cél és adatok. Brownlees et al. (2011) napon belüli volumen előrejelzést készít, kiemelten a VWAP kereskedésben való felhasználás céljából. Napon belül a leghangsúlyosabb stilizált tény az U alak, amelyet a modellnek is meg kell ragadnia valamilyen módon. Az empirikus vizsgálatot 2002. január - 2006. december közötti napon belüli adatokon végzik három fő USA-beli részvényindex likvidnek tekintett ETF-jére (Exchange Traded Fund, magyarul is ezzel a rövidítéssel utalnak a termékre, amely egyetlen tőzsdei tranzakcióvá egyszerűsíti egy olyan termék megvásárlását, amely egy választott részvénykosárnak megfelelő hozamot hoz), melyek az alábbiak:

- SPDR S&P 500, más néven SPY, az S&P 500 indexet követi,
- Dow Diamonds, más néven DIA, a Dow Jones iparági átlag indexet követi,
- PowerShares QQQ, más néven QQQQ (azóta QQQ-ra változott), a NASDAQ 100 indexet követi.

Egy kényelmes egyszerűsítés motiválta az ETF-ek használatát az egyedi részvényekkel szemben, ugyanis az előbbi esetben nem kell számolni a részvények egyedi jellegzetességeiből adódó kilengésekkel, amelyeket külön kellene modellezni, hiszen ETF-ek esetén a követett kosárban lévő részvények magas száma miatt az egyedi részvények szintjén jelentkező sokkok aggregált szinten eltűnnek.

A modellezés során volumen helyett forgalom adatokkal dolgoznak. A napon belüli adatok 15 perces sűrűségűek, így naponta 26 adatpont van. Azokat a napokat, ahol találtak üres 15 perces rekeszt (vagyis 26-nál kevesebb adatpont

volt), kihagyták. Az előrejelzés tesztelésére a 2005. január - 2006 december közötti időszakot használják, az előrejelzéshez használt modell paramétereit hetente újrabecsülvé.

Módszertan. A napon belüli forgalmat multiplikatív komponensekre bontva modellezik. A három komponens: (1) napi (2) napon belüli periodikus (3) napon belüli nem-periodikus. A napi komponens az alacsonyabb frekvenciájú piaci tendenciát ragadja meg, hiszen a volumen ingadozhat a stilizált tények által megragadott szezonálisitáson túl is. A szerzők ezt úgy fogalmazzák meg, hogy klasztereződik a kereskedési aktivitás. A második, napon belüli periodikus komponens az U alakot hivatott követni, míg a harmadik, napon belüli nem-periodikus komponens egyfajta *maradék*.

Az alkalmazott modell a multiplikatív hiba modellek (Multiplicative Error Model, MEM) kiterjesztése, egy komponens multiplikatív hiba modell (Component MEM, CMEM). Utalnak rá, hogy az általuk használt CMEM modellben felfedezhető némi hasonlóság Engle et al. (2007) komponens GARCH modelljével, illetve Bollerslev és Ghysels (1996) periodikus GARCH modelljével is, mindkettő a napon belüli volatilitást modellezi.

A modellt az alábbi formában írják fel:

$$x_{ti} = \eta_t \phi_i \mu_{ti} \varepsilon_{ti} \quad (26)$$

ahol $t \in \{1, \dots, T\}$ a megfigyelt napok száma, $i \in \{1, \dots, I\}$ a napon belüli rekeszek száma, x a forgalom, η a napi komponens, ϕ a napon belüli periodikus komponens, μ a napon belüli nem-periodikus komponens, ε pedig független, azonos eloszlású hiba, amely nemnegatív, átlaga egységnyi, szórása konstans σ .

A megelőző rekesz időpontjáig rendelkezésre álló információra nézve x_{ti} feltételes várható értéke a három komponens szorzata:

$$x_{ti} = \eta_t \phi_i \mu_{ti} \quad (27)$$

A komponenseket az alábbi módon formalizálják.

A napi komponens:

$$\eta_t = \alpha_0^{(n)} + \beta_1^{(n)} \eta_{t-1} + \alpha_1^{(n)} x_{t-1}^{(n)} \quad (28)$$

ahol az $x^{(\eta)}$ sztenderdizált napi forgalmat

$$x_t^{(\eta)} = \frac{1}{I} \sum_{i=1}^I \frac{x_{ti}}{\phi_i \mu_{ti}} \quad (29)$$

módon számolják, ami a napon belüli forgalmak átlaga pontonként normálva a másik két komponenssel. Ennek a sztenderdizált tagnak az az értelme, hogy így bekerül az ε hiba, vagyis gyakorlatilag felhasználjuk a legfrissebb tény adatot is az előrejelzéshez. A várható értéke pedig éppen η_t lesz. Az átlagolásra azért van szükség, mert a jobb oldalon 15 percenként frissül az adat, míg a napi átlag csak naponta egyszer.

A ϕ napon belüli periodikus komponens (U alak) Fourier reprezentációval ragadják meg:

$$\phi_{i+1} = \exp \left\{ \sum_{k=1}^K [\delta_{1k} \cos(f k i) + \delta_{2k} \sin(f k i)] \right\} \quad (30)$$

ahol

- $f = \frac{2\pi}{I}$
- $K = \text{int}(\frac{I}{2})$
- $i = \{0, \dots, I - 1\}$
- $\delta_{2K} = 0$, ha I páros.

A választást azzal indokolják, hogy a több lehetséges felírás közül ez viszonylag takarékosan bánik a paraméterekkel, és ezáltal célszerűnek tekinthető.

A napon belüli nem-periodikus komponens pedig:

$$\mu_{ti} = \alpha_0^{(\mu)} + \beta_1^{(\mu)} \mu_{ti-1} + \alpha_1^{(\mu)} x_{ti-1}^{(\mu)} \quad (31)$$

ahol $x^{(\mu)}$ a sztenderdizált napon belüli forgalom:

$$x_{ti}^{(\mu)} = \frac{x_{ti}}{\eta_t \phi_i} \quad (32)$$

Ennek a sztenderdizált tagnak szintén az az értelme, hogy így bekerül az ε hiba is az előrejelzésbe. Ez a (26) egyenlet átrendezésével itt közvetlenebbül látszik, mint $x_i^{(\eta)}$ esetén, mert nincs átlagolás.

A feltétel nélküli várható értéke a napon belüli nem-periodikus komponensnek egységnyi. Ebből az alábbi következnek a paraméterekre:

$$\alpha_0^{(\mu)} = 1 - \beta_1^{(\mu)} - \alpha_1^{(\mu)} \quad (33)$$

Emiatt eggyel csökken a becslendő paraméterek száma.

Az általánosított momentumok módszerét használva (Generalized Method of Moments, GMM; Wooldridge (1994) illetve Newey és McFadden (1994)) együtt becslik a CMEM paramétereit. Mielőtt azonban ezt megtennék, néhány további jellemőt is hozzáadnak a fenti modellhez, hogy az realisztikusabb legyen (elmondásuk szerint a jelölés bonyolítását elkerülendő hagyták ki a korábbi részletes leírásból ezeket). Ezek az alábbiak:

- A zárás és a másnapi nyitás között érkező új információ, ami tehát a záró volumenben még nem tükröződött, így az nem feltétlenül magyarázza megfelelően a nyitáskori értéket. Ezt figyelembe veendő egy dummy változóval bővítik a (31) egyenletet.
- A (28) és (31) egyenleteket bővítik egy-egy olyan változóval, amely aszimmetrikusan figyelembe veszi az előző rekesz hozamának az előjelét. A motiváció ezen lépés mögött, hogy egyrészt a volatilitásra ismert a rossz hírek erősebb hatása a jó hírekkel szemben (leverage effect: Nelson (1991), Glosten et al. (1993), Rabemananjara és Zakoian (1993)), másrészt kimutattak közös látens komponenst a volatilitás és a volumen között (melyet nevezhetünk információ áramnak: Andersen (1996), Hautsch (2008)).
- Egy második késleltetés a (31) egyenletben.

Mindezeket figyelembe véve négy verzióban becslik a modellt:

- alap, dummyval (A),
- alap dummyval és aszimmetrikus hatással (B),
- alap dummyval és másodrendű késleltetéssel (C),
- alap dummyval, aszimmetrikus hatással és másodrendű késleltetéssel (D).

Az előrejelzés értékelésére több mutatót is fontolóra vesznek. Az egyik az általánosan jól használható MSE kritérium. Ugyanakkor, mivel a cikk a volumen előrejelzést a VWAP kereskedés alkalmazásában vizsgálja, a további lehetséges kritériumokat ennek fényében állítják fel.

A következő kritérium szintén az MSE logikáját követi, de nem az aktuális és előrejelzett volumen eltérésére, hanem a megfelelő időszak VWAP-ja és az ügyleten elért volumennel súlyozott átlagos ár eltérésére vonatkozóan számítják. A számításokban árként mindig az adott rekeszben utoljára megfigyelt árat használják. Mivel azonban ez torzítja a volumen előrejelzésének értékelését, hiszen az arra vonatkozó várakozás hibáját is tartalmazza, ezt a mutatót másodlagosnak tekintik.

A szerzők által kreált új kritérium, a feldarabolási veszteség (Slicing loss) függvény az alábbi alakba írható:

$$L^{slicing} = -\sum_{t=1}^T \sum_{i=1}^I w_{ti} \log \hat{w}_{ti} \quad (34)$$

ahol

$$w_{ti} = \frac{x_{ti}}{\sum_{i=1}^I x_{ti}} \quad (35)$$

valamint

$$\hat{w}_{ti|t-1} = \frac{\hat{x}_{ti|t-1}}{\sum_{i=1}^I \hat{x}_{ti|t-1}} \quad (36)$$

Érdeemes észrevenni, hogy noha itt is a kisebb érték a kedvező, a tökéletes előrejelzést nem a zérus érték jelenti.

Javasolnak még további két mutatót, melyeket részletesen nem mutatok be, ugyanis azonos sorrendet adnak minden esetben a feldarabolási veszteség függvényvel (gyakorlatilag tartalmazzák azt egyéb, adott időtávú előrejelzés esetén konstans tagok mellett), ezért a szerzők is végül ez utóbbit választják.

A (36) alapján is látható, hogy az előrejelzést mindig nap elején készítik, azonban kétféleképpen használják fel a rendelkezésre álló információt. Statikus esetben előrejeleznek minden értéket, és ezt véglegesnek tekintik, dinamikus esetben pedig minden új információ (új rekesz) beérkezésekor újraszámítják a hátralévő értéket.

| | SPY | | | DIA | | | QQQ | | |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | MSE | Slicing | VWAP | MSE | Slicing | VWAP | MSE | Slicing | VWAP |
| | Statikus | | | | | | | | |
| BM | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| (A) | 0,964 | 0,998 | 0,998 | 0,977 | 0,996 | 1,012 | 0,971 | 0,999 | 0,992 |
| (B) | 0,962 | 0,998 | 0,998 | 0,976 | 0,996 | 1,013 | 0,968 | 0,999 | 0,991 |
| (C) | 0,964 | 0,998 | 0,987 | 0,977 | 0,996 | 1,001 | 0,971 | 0,998 | 0,992 |
| (D) | 0,962 | 0,998 | 0,983 | 0,976 | 0,996 | 1,001 | 0,968 | 0,998 | 0,992 |
| | Dinamikus | | | | | | | | |
| (A) | 0,856 | 0,997 | 0,987 | 0,906 | 0,995 | 0,916 | 0,873 | 0,997 | 0,985 |
| (B) | 0,855 | 0,997 | 0,985 | 0,906 | 0,995 | 0,912 | 0,872 | 0,997 | 0,982 |
| (C) | 0,855 | 0,996 | 0,962 | 0,903 | 0,993 | 0,916 | 0,867 | 0,996 | 0,967 |
| (D) | 0,855 | 0,996 | 0,961 | 0,903 | 0,993 | 0,914 | 0,866 | 0,996 | 0,966 |

4. táblázat. Brownlees et al. (2011) eredményei
 Forrás: Brownlees et al. (2011)

Benchmark stratégiaként kiszámolják az utolsó 40 nap átlagos értékét minden rekeszre, és ezt a 26 értéket tekintik előrejelzésnek a megfelelő rekeszekre. Több napos előrejelzés esetén mindig a legutóbbi 40 nappal számolnak.

Eredmények. A szerzők interpretációja az eredményekről a következő. A CMEM előnye dinamikus előrejelzés esetén domborodik ki, amikor is mindhárom termékre határozottan megveri a benchmark előrejelzést. A különböző modellverziók összevetéséből látszik, hogy habár a legjobb a (D), tehát a legszofisztikáltabb változat, a második késleltetés hozzáadása jelentett jelentősebb javulást (C), szemben az aszimmetrikus komponens hatásával (B).

Statikus esetben a javulás kevésbé jelentős a benchmarkhoz képest, DIA esetén VWAP MSE alapon pedig romlás tapasztalható (a VWAP kereskedésben történő alkalmazás miatt őket elsősorban a feldarabolási veszteség függvény és a VWAP alapú MSE érdekli).

A szerzők által közölt eredménytáblázatot a kritériumok eltérő nagyságrendje miatt normáltam a benchmark hibájával a könnyebb összevethetőség érdekében, az így módosított változat látható az 4. táblázatban, ezrelékre kerekített értékekben megjelenítve. Ez alapján annyiban árnyalnám az interpretációjukat, hogy a feldarabolási veszteség függvény alapján (aminek a használatát ők vezetik be) a dinamikus esetben sem éri el a javulás az 1%-ot, sőt, néhány kivételtől eltekintve a 0,5%-ot sem haladja meg. Annyit meg kell azonban jegyezni, amire már korábban

| # | Szerzők | Év | Adat | Adat-sűrűség | Módszertan |
|---|---------------------------------|------|----------|--------------|---|
| 1 | Kaastra, Boyd | 1995 | áru | 1 hónap | neurális hálók |
| 2 | Lux, Kaizoji | 2007 | részvény | 1 nap | ARFIMA, MF |
| 3 | Bialkowski, Darolles, Le Fol | 2008 | részvény | 20 perc | Sokdimenziós faktor- modell, AR, SETAR |
| 4 | Brownlees, Cipollini, Gallo | 2011 | ETF | 15 perc | CMEM |

5. táblázat. Volumen előrejelző modellek

Forrás: Saját szerkesztés

is utaltam, hogy a feldarabolási veszteség függvény esetén a tökéletes előrejelzést nullánál magasabb érték jelzi, vagyis eleve kisebb javulásra számíthatunk százalékosan is, mint az MSE alapú esetekben.

3.3.5. Áttekintés

A 5. táblázat segít röviden áttekinteni a fejezetben megismert volumen előrejelző modelleket.

Amint látható, a négy modellből a Kaastra és Boyd által javasolt megoldás jelen dolgozat szempontjából erősen korlátozott relevanciával bír, egyrészt az áru és értéktőzsde eltérő tulajdonságai miatt, másrészt az alacsony frekvenciájú (havi) adatok miatt, amin belül is még inkább a hosszab távú előrejelzés esetén volt sikeresnek tekinthető az általuk javasolt modell.

Lux és Kaizoji modellje már közelebb áll dolgozatom témájához, hiszen részvényekkel dolgozik, és a frekvencia is magasabb. Azonban a modell elsősorban a hosszú memória stilizált tényére koncentrál, amit ugyan napon belül is lehet értelmezni, ugyanakkor a napi adatok még mindig nem teszik szükségessé a jellegzetes napon belüli U alak jelenségének modellezését, aminek a megragadása döntő jelentőségűnek tekinthető a napon belüli előrejelzésben.

A következő két modell viszont már ez utóbbi probléma kezelését helyezi a középpontba, ennek megfelelően napon belüli adatokkal dolgozik, aminek következtében közvetlen relevanciája van számomra is. Ez utóbbi következményeként az első két modellhez képest részletesebben mutattam be őket.

Brownlees, Cipollini és Gallo modellje, amely multiplikatív módon bontja komponensekre a napon belüli forgalmat, megragadja az U alakot is. A felhasznált

adatok szándékosan ETF-ek, és nem részvények, hogy az egyedi ingadozások hatásának modellezése ne legyen külön feladat.

Mivel engem elsősorban az egyedi részvények forgalmának előrejelzése érdekel, számomra legfontosabb Bialkowski, Darolles és Le Fol modellje. Ennek köszönhető, hogy ezen relevancia szerinti áttekintésben az utolsó helyen szerepel ez az additív komponenseket használó módszer. Amint láthattuk, az U alak megragadását a modell tartalmazza, és egyedi részvények esetén is hatékonynak mondható a forgalom előrejelzése az alternatívákkal szemben. Ennek egyik oka az lehet, hogy viszonylag sok információt használ fel, hiszen a piaci komponenst CAC40 index összes részvényét felhasználva állítja elő, emellett pedig az egyedi részvények idősorát külön is figyelembe veszi.

II. rész

Adatok és hipotézisek

4. fejezet

Adatok bemutatása

A kutatás empirikus részéhez az elmélet mellett szükség lesz adatokra is. Ebben a fejezetben bemutatom a rendelkezésemre álló adatbázist, illetve annak a kutatáshoz felhasznált szeletét.

4.1. Nyers adatok

Az ár és volumen adatokat a kibot.com oldalon vásároltam. 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. (A Kraft Foods Inc. NYSE részvényként szerepel, azonban az adatbázis által lefedett időszak kb. utolsó három hetében, 2012.06.26. kezdettel átvezették a NASDAQ-ra.). A mellékletben található 96-98. táblázat részletes információt ad a részvényekről.

Az adatok szerkezetét a 6. táblázat szemlélteti a Boeing Co. példáján. A megfigyelések percenkénti gyakoriságúak. Abban az esetben, ha egy adott percben nem volt kereskedés, a vonatkozó sor hiányzik. A kereskedés mind a NYSE-n mind a NASDAQ-on minden munkanapon 09:30-16:00 között zajlik. Az adatbázis ezen felül tartalmazza a nyitás előtti és zárás utáni kötések is, szintén percenkénti bontásban. Az oszlopokban a naptári dátum, az idő (perc pontossággal), a nyitó ár, a legmagasabb és legalacsonyabb ár, a záróár, végül pedig a volumen található

| Date | Time | Open | High | Low | Close | Volume |
|-----------|-------|--------|--------|--------|--------|--------|
| 7/13/2012 | 09:30 | 71.930 | 71.980 | 71.850 | 71.950 | 46 130 |
| 7/13/2012 | 09:31 | 71.970 | 72.080 | 71.910 | 72.070 | 3 950 |
| 7/13/2012 | 09:32 | 72.050 | 72.090 | 71.900 | 71.910 | 7 552 |
| 7/13/2012 | 09:33 | 71.985 | 72.200 | 71.985 | 72.160 | 4 186 |

6. táblázat. Az adatok szerkezete (Boeing Co. példája)

Forrás: Saját szerkesztés

| # | Részvény | Első megfigyelés |
|---|---------------------------------|------------------|
| 1 | Exxon Mobil Corporation | 1999.12.01. |
| 2 | Verizon Communications Inc. | 2000.07.03. |
| 3 | Kraft Foods Inc. | 2001.06.14. |
| 4 | Chevron Corporation | 2001.10.10. |
| 5 | Pimco Global Stocksplus & Incom | 2005.05.26. |
| 6 | The Travelers Companies, Inc. | 2007.02.27. |
| 7 | General Motors Company | 2010.11.18. |

7. táblázat. A többségnél rövidebb adatsorú részvények

A többi huszonkilenc részvény első megfigyelése 1998.01.02.

Forrás: Saját szerkesztés

ilyen sorrendben. Az árak USD-ben, a volumen darabban szerepelnek.

Az első adatnap 1998.01.02. minden részvénytől, kivéve, ha az adott tőzsdére később vezették be a papírt. Hét ilyen részvény van a mintában, ezeket a 7. táblázatban tekinthetjük át. Az utolsó adatnap minden részvénytől 2012.07.13.

A fentiek mellett szükségem volt a forgalomban lévő részvény darabszám (TSO) értékeire is (l. (6) egyenlet). Ezeket az adatokat a teljes időszakra Bloomberg terminálról töltöttem le részvénykénti negyedéves bontásban.

4.2. Felhasznált adatok

Először is ki kell választani a részvények közül, hogy melyek szerepeljenek a mintában. Alapvetően érdemes lehetőség szerint minél többet megtartani, ezért potenciális kiesőket kerestem. Mivel túlnyomó többségben vannak az NYSE részvények, felmerülhet, hogy a három NASDAQ-os maradjon ki. Ez a három azonban (Cisco Systems Inc., Intel Corporation, Microsoft Corporation) kifejezetten likvid részvények, a kobot.com szerint az Amerikai Egyesült Államokban ezek rendre a tizenötödik, kilencedik és hatodik leglikvidebb részvények. Továbbá, mint ar-

ról már esett szó, a két tőzsde nyitvatartási ideje is megegyezik, ezért ebből a szempontból sem jelent problémát a NASDAQ részvények megtartása. Hasonló megfontolásokból a Kraft Foods Inc. már említett, 2012 június végi NASDAQ-ra történő átvezetését sem tekintettem kizáró érvnek (noha az a kbot.com likviditási listájában *csak* a százhatodik).

Összességében is elmondható, hogy a legtöbb részvény kifejezetten likvidnek tekinthető. Likviditás szempontjából egyedül a Pimco Global Stocksplus & Incom maradt el a többitől, ezért ezt az egyet kihagytam a mintából, és 35 részvénnel dolgoztam tovább.

Következő lépésként meg kell határozni, hogy időben mettől meddig tartson a használt minta. Bialkowski et al. (2008) modelljéhez (l. 3.3.3. alpont) szükséges, hogy a választott időtávon minden megtartott részvényre legyen adat, ezért a választott időtávot az a részvény fogja meghatározni, amelynek a megtartottak közül legrövidebb az adatsora. Korábban láttuk, hogy hét olyan részvény van, amelynek adatsora később kezdődik, mint a többi részvényé. A 7. táblázat szerint az utolsó két részvény (The Travelers Companies, Inc., illetve General Motors Company) adatsora lényegesen rövidebbnek tekinthető a többinél, ezért ezeket elhagyom. Megfontolandó lehet még a következő legrövidebb, a Pimco Global Stocksplus & Incom, de azt már likviditási szempontok miatt kihagytuk korábban is. Tehát ezt a három részvényt kihagyva végül harminchárom részvénnel dolgozom. Az első megfigyelésem így 2001.10.10. lesz, ami a megtartott részvények közül a legrövidebb, a Chevron Corporation első adatpontja.

Meg kell vizsgálni azt is, hogy indokolt-e bizonyos rendellenes napokat kihagyni a mintából. Az ünnepnapokon, amint hétvégén is, alapvetően szünetel a kereskedelem, tehát ezek nem jelentenek problémát, azonban bizonyos ünnepeket környező napokon rövidebb nyitvatartással működnek a tőzsdék. Ilyenkor a volumen jellegzetes, napon belüli U alakja *összenyomódik* a rövidebb nyitvatartási időre, ami torzítást okoz. Ilyen napok karácsony előestéje, a hálaadás másnapja, valamint a függetlenség napja előtti nap. Ezek a napokon, ha munkanapra esnek, 13:00-kor zárnak a tőzsdék. A vizsgált időszakban huszonnégy ilyen rendellenes nap volt, ezért ezeket a napokat kihagytam a mintából. A kihagyott napokat a 8. táblázat mutatja.

Ezen felül az is előfordulhat, hogy egy-egy részvény (gyakran egyszerre több) kereskedését felfüggesztik napon belül hosszabb-rövidebb időre. A mintában négy napon fordult ilyen elő, ezeket is kihagytam a további vizsgálatból (9. táblázat).

| | | | |
|-------------|-------------|-------------|-------------|
| 2001.11.23. | 2003.11.28. | 2006.11.24. | 2008.12.24. |
| 2001.12.24. | 2003.12.24. | 2007.07.03. | 2009.11.27. |
| 2002.07.05. | 2003.12.26. | 2007.11.23. | 2009.12.24. |
| 2002.11.29. | 2004.11.26. | 2007.12.24. | 2010.11.26. |
| 2002.12.24. | 2005.11.25. | 2008.07.03. | 2011.11.25. |
| 2003.07.03. | 2006.07.03. | 2008.11.28. | 2012.07.03. |

8. táblázat. Rendellenes napok a mintában

Forrás: Saját szerkesztés

| | |
|-------------|-------------|
| 2002.09.11. | 2005.02.28. |
| 2003.01.02. | 2005.12.14. |

9. táblázat. A kereskedés felfüggesztése miatt kihagyott napok

Forrás: Saját szerkesztés

Az adatbázis, mint már említettem, nem csak a folyamatos kereskedés adatait, hanem a nyitás (09:30) előtti és zárás (16:00) utáni kötések adatait is tartalmazza. Mivel célom a folyamatos kereskedés modellezése, illetve előrejelzése, az ezen kívül eső adatokat töröltem.

Az adatokat 15 perces tömbökbe aggregáltam, így minden napra maximum 26 megfigyelés jut¹. Ennek előnye abból ered, hogy a módszertan bizonyos pontokon megkívánja, hogy azonos megfigyelés szám legyen a különböző részvényekben. Mivel egy kivételtől eltekintve a legtöbb részvény kellően likvid ahhoz, hogy 15 percnként legalább egy kötés legyen, és ezáltal pontosan 26 aggregált adatpont legyen minden nap, ezért az egyenlő megfigyelés szám feltétele automatikusan teljesül a 15 perces aggregálás esetén. Az egy kivétel a már említett Pimco Global Stocksplus & Incom, amely esetén rengeteg hiányzó 15 perces tömb maradt, ezért ezt a részvényt (a korábban írtaknak megfelelően) kihagytam a vizsgálatból. Ha feltétlenül bent akartam volna tartani ezt a részvényt is, mivel volumenről van

¹A 15 perces választás elsőre önkényesnek tűnhet, ezért meg kell indokolni. Mivel az irodalomban található modelleket is meg fogom becsülni, célszerű olyan aggregálást választani, amely közel esik azokhoz, hiszen egy teljesen eltérő aggregálási szinten esetleg más határfokú lehet egy-egy modell, és ennek tesztelése nem elsődleges célom. Bialkowski et al. (2008) 20 perces aggregálást használ, és így maximum 25 megfigyelést kap naponta. Brownlees et al. (2011) 15 percnként aggregál, ami maximum 26 megfigyelést eredményez egy nap. Kézenfekvő, hogy elsősorban ezt a két változatot vizsgálom, de mivel az én mintámban 1 nap egyenlő részekre nem osztható fel 20 percnként (9:30 és 16:00 között 390 perc van), a 15 perc ésszerű választásnak tűnik. Ez azért is jó, mert így pontosan egyezik a választásom Brownlees et al. (2011) modelljével, és nagyon közel esem a Bialkowski et al. (2008) által választott napi rekesz számhoz is.

| | | | |
|-------------|-------------|-------------|-------------|
| 2001.10.17. | 2002.01.30. | 2006.09.18. | 2011.01.31. |
| 2001.11.12. | 2002.12.03. | 2007.02.08. | - |
| 2001.11.15. | 2003.04.24. | 2007.06.28. | - |
| 2001.12.11. | 2006.06.05. | 2010.12.08. | - |

10. táblázat. Hiányzó 15 perces adatpontot tartalmazó napok
Forrás: Saját szerkesztés

| | |
|------------------------------|--------|
| Részvények | 33 |
| Napok | 2 668 |
| Megfigyelés naponta | 26 |
| Idősor hossza részvényenként | 69 368 |

11. táblázat. A felhasznált adatbázis néhány jellemzője
Forrás: Saját szerkesztés

szó, nulla értékkel pótolni lehetett volna a hiányzó adatpontokat, de nem ezt a megoldást választottam (a kihagyása mellett szól a 7. táblázat alapján a rövidebb adatsor is).

A fenti gondolatot folytatva megnéztem azt is, hogy a megtartott részvények esetén milyen gyakori a hiányzó 15 perces adatpont. Összesen tizenhárom olyan napot találtam, amelyen valamelyik részvélynél volt hiányzó adatpont a rendellenes napok törlése után is. Ahogy fentebb említettem, az ilyen napok alapvetően nem okoznak problémát, de mivel a teljes mintához képest nagyon ritkák, inkább töröltem őket (10. táblázat).

Az eddigiek eredményeként felhasználható adatbázis méreteit a 11. táblázat mutatja. Az adattisztítás és a 15 perces aggregálás után tehát 2,29 millió volumen/forgalom megfigyelés maradt az összes megtartott részvény tekintetében, az empirikus részben ezekkel dolgoztam.

5. fejezet

Kutatási kérdések, hipotézisek

A kutatás empirikus részének megkezdése előtt célszerű megfogalmazni a kutatási kérdéseket, illetve az ezekhez kapcsolódó hipotéziseket is.

Mivel céлом a szakirodalom legjobbjánál jobb modellt alkotni, az első lépés értelemszerűen az irodalom legjobb modelljének azonosítása kell, hogy legyen. Ez azonban pusztán az irodalom feldolgozásán keresztül nem érhető el, ezért empirikus kutatást igényel. Ennek oka, hogy a két releváns cikk² teljesen különböző adatbázison alapul, hiszen eltérőek a piacok (39 részvény vs. 3 ETF) és a vizsgált időszak is (2003.09.–2004.08. vs. 2002.01.–2006.12.), noha ez utóbbiban, mint látjuk, átfedés van. Ezen felül pedig a hibamérés módja sem egyezik meg a két esetben. Mivel tehát a későbbi cikk nem veti össze a saját teljesítményét a korábbi cikkével, ezt nekem kell megtennem.

A modelleket ismerve azt gondolom, hogy Bialkowski et al. (2008) modellje mellett szól, hogy azt részvény adatbázison fejlesztették ki. Mivel a kutatásomban magam is részvények forgalmát próbálom előrejelezni, ez mindenképp előny, hiszen a Brownlees et al. (2011) által használt ETF-ek mentesek az egyedi részvényekben megfigyelhető ingadozásoktól, ezért az ilyen adatbázis alapján fejlesztett modell elképzelhető, hogy kevésbé jól ragadja meg ezt az aspektust, ha a szerzőkkel ellentétben részvényekre próbálom meg alkalmazni a modellt. Másrészt, míg Brownlees et al. (2011) modellje egy adott eszköz forgalmának előrejelzéséhez csak az adott eszköz adatait használja, addig Bialkowski et al. (2008) modelljében minden részvényhez felhasználjuk a teljes piaci információt. Habár ez előre nem tudható, alapvetően van esély rá, hogy több információból pontosabb előrejel-

²Emlékeztetőül: Bialkowski et al. (2008), valamint Brownlees et al. (2011).

zést sikerül készíteni. Végül pedig előzetesen szerencsésebbnek tűnik a Bialkowski et al. (2008) által használt additív struktúra a Brownlees et al. (2011) modelljében található multiplikatív változatnál, ugyanis multiplikatív esetben ha az egyik (esetleg önmagában kisebb jelentőségű) tényezőben nagyot tévedünk, az nagyobb mértékben öröklődhet át a teljes előrejelzésre, mint additív esetben.

H1 Hipotézis: Benchmark. Azonos adatbázison, azonos módon értékelve Bialkowski et al. (2008) modellje jobban szerepel, mint Brownlees et al. (2011) modellje, és ebből eredően a szakirodalomból vett benchmarknak az előbbit érdemes tekinteni. Máshogy fogalmazva, hipotézisem szerint részvényekre alkalmazva Bialkowski et al. (2008) modellje a szakirodalomban megtalálható legjobb napon belüli forgalom előrejelző modell³.

A második kutatási kérdésem arra irányul, hogy hogyan lehetne egy jobb modellt alkotni a benchmarknál. A modell kereséshez kiindulásként érdemes visszagondolni a forgalommal kapcsolatban megismert stilizált tényekre. A legfontosabb stilizált tény a szakirodalom alapján egyértelműen a napon belüli U alak, ezért ezt mindenképp érdemes bevonni a vizsgálatba. Található ugyanakkor olyan kutatás is (l. Chiang et al. (2010)), amely a forgalom és az ármozgás közötti szignifikáns kapcsolatra hívja fel a figyelmet. Ezért ezt az összefüggést is megpróbálhatjuk modellbe foglalni egy pontosabb előrejelzés érdekében.

H2 Hipotézis: Jobb modell. Található az irodalom alapján kijelölt benchmarknál jobb modell a részvények napon belüli forgalom előrejelzésére. Ezen hipotézis igazolása érdekében megkísérlek felírni legalább egy olyan modellt, amely a rendelkezésemre álló adatbázison jobban teljesít a benchmarknál.

A fentiek alapján a második hipotézishez kapcsolódóan megfogalmazható néhány további kiegészítés arra vonatkozóan, hogy a jobb modellt milyen szempontok modellbe foglalása mentén keresem. Mint láttuk, ezek elsősorban a napon belüli U alak, valamint az ármozgással való kapcsolat. Ezek alapján felírhatjuk az alábbi kiegészítő hipotéziseket is.

³A kutatási tervem, amely a szakirodalom feldolgozását már tartalmazta, 2013 márciusában került elfogadásra. Az empirikus részben ezért az addig megismert szakirodalmat vizsgáltam részletesen.

H2.1 Hipotézis: Az U alak. A forgalom napon belüli U alakjának modellezése hozzájárul a benchmarknál jobb modell felírásához.

H2.2 Hipotézis: Ármozgás mutatók. A forgalom és az ármozgás mutatók kapcsolatának figyelembe vétele hozzájárul a benchmarknál jobb modell felírásához.

A H1 hipotézist a III. részben, a H2 hipotézist és kiegészítéseit pedig a IV. részben vizsgálom részletesen.

III. rész

Benchmark kiválasztása

6. fejezet

Előkészítés

Mielőtt nekilátnék egy olyan modellt keresni, amely pontosabb előrejelzést ad a szakirodalomban található legjobbnál, néhány alapvető kérdést meg kell válaszolni. Az egyik, hogy mi alapján tekintek egy modellt jobbnak, mint egy másikat. A másik kérdés, hogy az irodalomban melyik a legjobb modell, aminél jobbat keresek. Ezek közül egyik sem egyértelmű, hiszen egyrészt számtalan hibamérték szóba jöhet, másrészt az irodalomban található modellek nem reflektálnak egymásra, hanem különböző adatokat különböző időszakon vizsgálva egy egyszerű benchmarkot, a (14) egyenletben bemutatott U-módszert veszik alapul, amelynél aztán jobbnak bizonyulnak. A dolgozatom egyik eredménye, hogy a szakirodalomban megtalálható, napon belüli forgalom előrejelzésre javasolt modelleket összevetem egymással azonos adatokon és azonos időszakra becsülve, hogy megtudjuk, melyiket érdemes előnyben részesíteni, ha részvények napon belüli forgalmát szeretnénk előrejelezni.

Ebben a fejezetben kijelölöm, hogy milyen szempontok szerint tekintem jobbnak egyik modellt a másiknál, és még néhány további előkészületet is teszek. A III. rész többi fejezetében megbecslem a választott modelleket, és kiválasztom közülük azt, amelyet később benchmarknak fogok tekinteni a szakirodalomban találhatóaknál jobb modell keresése közben.

6.1. Hibamérték választása

Sokféle hibamértéket lehetne választani. Első megközelítésben egyszerűen az érdekel, hogy a becslés mennyire esik közel a később megfigyelt tényhez, és ha közelebb

van, azt jobbnak értékelem. A konkrét céltól függően egyéb lehetőségek is szóba jöhetnek, melyekre a 11. fejezetben vissza is térek.

Az átlagos négyzetes hiba (Mean Squared Error, MSE) egy a szokásos hibamértékek közül, amelyet modellek előrejelző képességének összehasonlítására használnak (Hamilton (1994)):

$$MSE = \frac{\sum_{t=1}^N (Y_t - Y_t^f)^2}{N} \quad (37)$$

ahol az Y_t előrejelzése Y_t^f . Tetszőleges előrejelzett lépés esetén használható, az előrejelzés és a realizáció négyzetes eltéréseinek egyszerű átlagát számolja. Bizonyos szerzők ennek a négyzetgyökével dolgoznak (Root Mean Squared Error, RMSE), de mivel a gyökvonás önmagában semmit nem változtat a sorrenden, ezért ezzel külön nem foglalkozom.

Egy másik gyakran használt mérőszám az átlagos abszolút százalékos eltérés (Mean Absolute Percentage Error, MAPE):

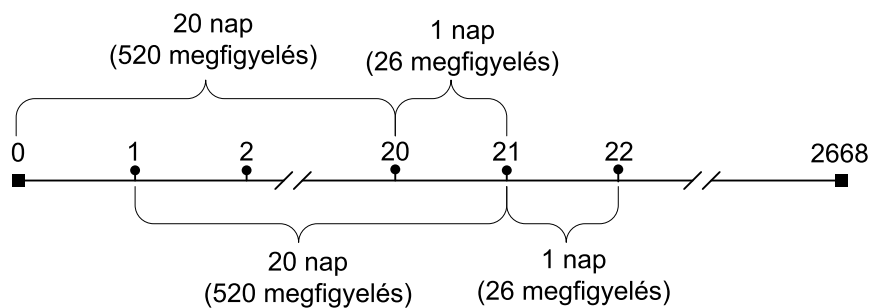
$$MAPE = \frac{\sum_{t=1}^N \left| \frac{Y_t - Y_t^f}{Y_t} \right|}{N} \quad (38)$$

Az MSE és MAPE közti választás azon múlhat, hogy mennyire szeretném büntetni a nagy tévedéseket (a négyzetre emelés jobban bünteti ezt). Ebben a tekintetben talán előnyben részesíthető az MSE, másrészt viszont meggyőzően jobb egy modell, ha ettől a választástól függetlenül, tehát mindkét hibamérték szerint kevesebb hibát vét az előrejelzés során.

A fenti két hibamértéket kiszámolom minden részvényre, és a részvények átlagában is. Egy modell akkor meggyőzően jobb a másiknál, ha több egyedi részvényre, és az összes részvény átlagában is alacsonyabb értéket ad az adott hibamérték szerint.

6.2. A becslés és előrejelzés közös részletei

A modellek előrejelzéseinek összehasonlíthatósága szempontjából fontos, hogy néhány alapvető választás összhangban legyen az egyes modellek tesztelése során. Ezért az ebben a pontban leírtakat nem csak a benchmark keresésénél, hanem



5. ábra. Húsz napos becslési és egy napos előrejelzési ablakok naponkénti csúsztatása
Forrás: Saját szerkesztés

később a saját modellek tesztelése közben is változatlan formában fogom alkalmazni.

Egy ilyen szempont természetesen a tesztelésre használt adatok egyezősége. Ennek megfelelően a 4.2. pontban részletezettek szerinti 33 részvény 2668 napját fogom használni (l. 11. táblázat).

A becslésre minden esetben 20 napot fogok felhasználni, ami közelítőleg egy naptári hónapnak felel meg (20 kereskedési nap). Mivel naponta 26 megfigyelésem van az aggregálás után, ez 520 megfigyelést jelent részvényenként. Ez a 20 napos választás összhangban áll a Bialkowski et al. (2008) cikkben alkalmazott becslési időszakokkal.

Az előrejelzés értékelésére a becslési időszakot követő egy napot fogom használni, ami 26 megfigyelés. Ez szintén a Bialkowski et al. (2008) cikkben alkalmazott választással egyezik meg. Ennek megfelelően a modellek paramétereit a 20 napos ablak naponkénti csúsztatása mellett naponta újra megbecslem, vagyis minden részvényre 2648 becslést végzek, és ugyanennyi előrejelzett napot értékelek ki. A becslési és előrejelzési időszakok kezelését az 5. ábra illusztrálja.

A 6.1. pontban említettem, hogy az itt használtakon kívül egyéb hibamértékek is szóba jöhetnek. Ugyanilyen megfontolásból az is kérdés lehet, hogy az előrejelzéseket az adott napra hogyan végzem el. A többféle lehetőségből egyelőre úgy jelzek előre, hogy minden lépésben (tehát napon belül is) mindig frissítem az információshalmazt, azaz folyamatosan figyelembe veszem az idő előrehaladtával megismert újabb tény adatokat.

A fentieket úgy lehet összefoglalni, hogy a becslés paramétereit naponta, az információshalmazt pedig 15 percenként frissítem, és mindig a következő 15 percet

jelzem előre. A későbbiekben egyéb lehetőségeket is vizsgálok majd.

6.3. Az U-módszer

A dolgozat szempontjából közvetlen relevanciával bíró két cikk tesztelése előtt szükségesnek látom elkészíteni a mindkét cikkben benchmarknak tekintett U-módszer előrejelzését (l. (14)), melyet értékelek is a 6.1. pontban bemutatott módon. Az eredményeket ugyanakkor nem közlöm itt külön, mert önmagában nehezen, inkább csak összehasonlításban értelmezhetőek.

7. fejezet

A BDF modell

A Bialkowski et al. (2008) által javasolt modellre a továbbiakban az egyszerűség kedvéért BDF modellként fogok hivatkozni (a szerzők vezetékneveinek kezdőbetűiből). Amint a 3.3.3. pontban láttuk, két különböző módon jelzik előre az egyedi részt, ezért gyakorlatilag két különböző modellt kell vizsgálnunk, melyekre a továbbiakban BDF_AR és BDF_SETAR modellként utalok, jelezve, hogy az egyedi részt AR vagy SETAR modellel jelezték-e előre.

A modell a 3.3.3. pontban leírtak alapján megbecsülhető, ezért azonnal rátérhetünk az eredmények U-módszerrel szembeni összevetésére.

A 12. táblázat segítségével először hasonlítsuk össze az U-módszer teljesítményét a BDF_AR modellel⁴. Láthatjuk, hogy a BDF_AR modell átlagban, és az egyedi részvényekre is jobb mind MSE, mint MAPE alapon. Egyetlen részvény van, ahol az U-módszer MAPE alapon némileg jobb eredményt mutat, de MSE alapon már nem találunk ilyen esetet.

Ugyanezt az összevetést láthatjuk 13. táblázatban, ezúttal a BDF_SETAR modell tekintetében. Az értelmezés nagyon hasonló, mint AR esetben.

⁴Jelölés: $aE+b \equiv a \cdot 10^b$, illetve $aE-b \equiv a \cdot 10^{-b}$

| | MSE | | MAPE | |
|------------------------|----------|----------|-------|--------|
| | U | BDF_AR | U | BDF_AR |
| Hány részvélynél nyert | 0 | 33 | 1 | 32 |
| Átlagos érték | 1,02E-03 | 6,49E-04 | 50,3% | 40,3% |

12. táblázat. A BDF_AR modell és az U-módszer előrejelzésének összevetése
Forrás: Saját szerkesztés

| | MSE | | MAPE | |
|----------------------|----------|-----------|-------|-----------|
| | U | BDF_SETAR | U | BDF_SETAR |
| Hány részvényt nyert | 0 | 33 | 1 | 32 |
| Átlagos érték | 1,02E-03 | 6,60E-04 | 50,3% | 39,9% |

13. táblázat. A BDF_SETAR modell és az U-módszer előrejelzésének összevetése

Forrás: Saját szerkesztés

| | MSE | | MAPE | |
|----------------------|----------|-----------|--------|-----------|
| | BDF_AR | BDF_SETAR | BDF_AR | BDF_SETAR |
| Hány részvényt nyert | 27 | 6 | 3 | 30 |
| Átlagos érték | 6,49E-04 | 6,60E-04 | 40,3% | 39,9% |

14. táblázat. A BDF_AR és BDF_SETAR modellek előrejelzésének összevetése

Forrás: Saját szerkesztés

Érdekes lehet még esetleg az, hogy ezen minta alapján kijelenthetjük-e, hogy egyik BDF modell jobb lenne a másikonál. A 12. és 13. táblázatokból ez nem derült ki, hiszen csak azt láttuk, hogy az U alaknál mindettől jobb volt. A 14. táblázat a két BDF modellt veti össze. Ez alapján megállapíthatjuk, hogy MSE alapon a BDF_AR tűnik jobbnak, míg MAPE szerint a BDF_SETAR. A választás tehát attól függ, melyik hibamértéket tartjuk jobbnak, vagyis nincs egyértelmű döntés, ahogyan Bialkowski et al. (2008) sem dönt egyértelműen, ezért is mutatja be mindkettőt. A cikkben MAPE alapon a SETAR verziót találják jobbnak (MSE-t nem számolnak), ezt az én mintám is megerősíti, ahogyan az U-módszerrel szembeni egyértelmű javítást is.

8. fejezet

A BCG modell

A Brownlees et al. (2011) által javasolt modellre a továbbiakban az egyszerűség kedvéért BCG modellként fogok hivatkozni (a szerzők vezetékneveinek kezdőbetűiből). Mint látni fogjuk, a cikk alapján a modell nem reprodukálható *tökéletesen*.

8.1. A becslés cikkben is közölt részletei

A 3.3.4. pontban részletesen bemutattam a modell specifikációját, de a becslésről ott csak annyi szerepelt, hogy az egy lépésben történik GMM segítségével. A következőkben ezért néhány további, a cikkben is szereplő részlet következik a becsléssel kapcsolatban.

8.1.1. A becslés menete

A becslést tehát GMM módszerrel végzik egy lépésben, vagyis együttesen becslik az összes paramétert. A keresett paramétereket gyűjtjük a következő oszlopvektorba:

$$\theta = (\theta^{(\eta)}; \theta^{(\mu)}; \theta^{(\phi)}) \quad (39)$$

ahol a pontosvessző azt jelenti, hogy egymás alá írjuk az általa elválasztott vektorokat.

Legyen:

$$u_\tau = \frac{x_\tau}{\eta_t \phi_i \mu_\tau} - 1 \quad (40)$$

ahol

$$\tau = I \cdot (t - 1) + i \quad (41)$$

Olyan paramétereket szeretnénk kapni, hogy u_τ várható értéke 0 legyen, ugyanis a (26) egyenlet felhasználásával láthatjuk, hogy $u_\tau = \varepsilon_\tau - 1$.

Brownlees et al. (2011) az 500-502 oldalon található levezetésben megadja, hogy az instrumentális változó milyen választása lesz az, ami hatékony a fenti feladat szempontjából. Ezek alapján az alábbi egyenlet megoldását jelentő θ értéke lesz a becslés:

$$\frac{1}{N} \sum_{\tau=1}^N a_\tau u_\tau = 0 \quad (42)$$

ahol

$$a_\tau = \eta_t^{-1} \nabla_\theta \eta_t + \mu_\tau^{-1} \nabla_\theta \mu_\tau + \phi_i^{-1} \nabla_\theta \phi_i \quad (43)$$

N pedig a megfigyelések száma.

8.1.2. Kezdő értékek

Mivel a modell rekurzív, meg kell adni néhány kezdő értéket.

A napi átlag kezdő értéke az első 5 nap átlaga, vagyis a cikk szerint:

$$\eta_0 = x_0^{(\eta)} = \frac{1}{5} \sum_{t=1}^5 \sum_{i=1}^I x_{ti} \quad (44)$$

A képlet ebben a formájában valószínűleg elütést tartalmaz, ugyanis $5I$ tagot ad össze, de csak 5-tel oszt. Ebben a felírásban az η napi átlagos teljes forgalom lenne, míg a korábbi felírásokban egyértelműen aznapra érvényes átlagos rekesz forgalom volt, mely legjobban talán a (29) egyenletből, a $x_t^{(\eta)}$ felírásából látszik. Emiatt ezt a kezdő értéket az alábbi formában értelmezem:

$$\eta_0 = x_0^{(\eta)} = \frac{1}{5I} \sum_{t=1}^5 \sum_{i=1}^I x_{ti} \quad (45)$$

A napon belüli nem-periodikus komponens feltétel nélküli várható értéke egy-

ségnyi, ezért a kezdő értéke is ennyi lesz a minta elején, vagyis:

$$\mu_{10} = x_{10}^{(\mu)} = 1 \quad (46)$$

Egyébként pedig az előző nap utolsó értéke lesz minden nap a kezdő érték, azaz:

$$\mu_{t0} = \mu_{t-1I} \quad (47)$$

és

$$x_{t0}^{(\mu)} = x_{t-1I}^{(\mu)} \quad (48)$$

Ez egyszerűen azt jelenti, hogy AR(1) tagként értelmezhetjük mindkettőt napon belül, és napváltáskor is.

Brownlees et al. (2011) mindezt úgy foglalja össze, hogy váltakozva frissül a napi átlag és a napon belüli nem-periodikus komponens.

8.2. A becslés további kifejtése

A becslés menetének ismertetésével Brownlees et al. (2011) megáll ott, ameddig a 8.1. pontban eljutottunk, azonban a konkrét implementációhoz további részletekre lesz szükségem, melyek itt következnek.

Amint azt a 3.3.4. pontban láttuk, a modellt négy különböző verzióban becslik, melyek közül mindegyik bővebb annál, mint amit a modell ismertetésekor a cikk részletez. Az alábbiakban a legbővebb, (D)-vel jelölt specifikációt fejtem ki. Egyrészt azért, mert a szerzők ennek a teljesítményét találják a legjobbnak, másrészt pedig azért is, mert a megfelelő tagok elhagyásával ebből megkapható a másik három változat (A-B-C).

A (D) esetben tehát, mint láttuk, a cikkben explicit módon részletezett modell annyival bővül, hogy egy dummy kerül a (31) egyenletbe, valamint aszimmetrikus hatás kerül a (28) és a (31) egyenletbe, végül pedig második késleltetés kerül a (31) egyenletbe.

8.2.1. A bővített egyenletek felírása

Elsőként egészítsük ki a (26) által megadott modell tényezőit, hogy azok a (D) specifikációnak feleljenek meg.

A napi komponens:

$$\eta_t = \alpha_0^{(\eta)} + \beta_1^{(\eta)} \eta_{t-1} + \alpha_1^{(\eta)} x_{t-1}^{(\eta)} + \gamma_1^{(\eta)} x_{t-1}^{-(\eta)} \quad (49)$$

ahol

$$x_t^{-(\eta)} = x_t^{(\eta)} I(r_t < 0) \quad (50)$$

vagyis az aszimmetrikus hatás megragadásakor a sztenderdizált változót szorozzuk egy indikátor függvénnyel (a napi hozam előjele szerint). A többi tag változatlan a (28) egyenlethez képest.

A napon belüli nem-periodikus komponens:

$$\mu_{ti} = \alpha_0^{(\mu)} + \beta_1^{(\mu)} \mu_{ti-1} + \alpha_1^{(\mu)} x_{ti-1}^{(\mu)} + \nu_1^{(\mu)} D_{ti} + \gamma_1^{(\mu)} x_{ti-1}^{-(\mu)} + \alpha_2^{(\mu)} x_{ti-2}^{(\mu)} \quad (51)$$

ahol

$$x_{ti}^{-(\mu)} = x_{ti}^{(\mu)} I(r_{ti} < 0) \quad (52)$$

vagyis az aszimmetrikus hatás a sztenderdizált változót szorozza egy indikátor függvénnyel (a rekesz hozam előjele szerint). Bekerült egy második késleltetés is a sztenderdizált változóra, valamint egy dummy-t tartalmazó tag is, amire:

$$D_{ti} = \begin{cases} 1 & \text{ha } i = 1 \\ 0 & \text{egyébként} \end{cases} \quad (53)$$

vagyis a nap első megfigyelését kezeljük kiemelten. A többi tag változatlan a (31) egyenlethez képest.

Végül a napon belüli periodikus komponens:

$$\phi_{i+1} = \exp \left\{ \sum_{k=1}^K [\delta_{1k} \cos(fki) + \delta_{2k} \sin(fki)] \right\} \quad (54)$$

ami tehát változatlan. Mivel a mintámban $I = 26$, ezért $\delta_{213} = 0$, és így csak 25 db tag, és ugyanennyi becslendő paraméter lesz ennél az egyenletnél. Az eddigiektől eltérő indexelés ($i = \{0, \dots, I-1\}$) azzal magyarázható, hogy technikai okokból kényelmesebb nulláról indítani a sin és cos argumentumok miatt.

8.2.2. További kezdő értékek

A 8.1.2 fejezetben sorra vettük a Brownlees et al. (2011) által megadott kezdőérték választásokat. Azonban, mint a 8.2.1. pontban láttuk, a (D) modell változat bevezet három új tagot, amiknek szintén meg kell adni a kezdő értékét. Ezeket a cikk iránymutatása híján az alábbi módon választottam meg.

- A hozamtól függő tagokra (l. (50) és (52)) azt feltételeztem, hogy induláskor nulla az értékük, vagyis kezdetben nem módosítok a hozam alapján.
- A második késleltetés (l. (51)) kapcsán azt feltételeztem, hogy kezdő értéke megegyezik az első késleltetés kezdő értékével (l. (46))

8.2.3. Deriváltak

A (43) megadásához szükségünk lesz a megfelelő deriváltakra, amihez először meg kell határoznunk a változót, ami szerint deriválni kell. Ez nem más, mint a (39)-ben látott θ , ami a (49), (51) és (54) egyenletek paramétereit rakja egymás alá, vagyis:

$$\theta = (\theta^{(\eta)}; \theta^{(\mu)}; \theta^{(\phi)}) \quad (55)$$

ahol

$$\theta^{(\eta)} = (\alpha_0^{(\eta)}, \beta_1^{(\eta)}, \alpha_1^{(\eta)}, \gamma_1^{(\eta)})^T \quad (56)$$

$$\theta^{(\mu)} = (\alpha_0^{(\mu)}, \beta_1^{(\mu)}, \alpha_1^{(\mu)}, \nu_1^{(\mu)}, \gamma_1^{(\mu)}, \alpha_2^{(\mu)})^T \quad (57)$$

$$\theta^{(\phi)} = (\delta_{11}, \delta_{21}, \delta_{12}, \delta_{22}, \dots, \delta_{112}, \delta_{212}, \delta_{113})^T \quad (58)$$

Vagyis θ egy (35x1)-es oszlopvektor.

A deriváltak tehát (1x35)-ös sorvektorok lesznek, *sok nullával kitöltve*:

$$\nabla_{\theta} \eta_t = [1, \eta_{t-1}, x_{t-1}^{(\eta)}, x_{t-1}^{-(\eta)}, 0, \dots, 0] \quad (59)$$

$$\nabla_{\theta} \mu_{\tau} = [0, 0, 0, 0, 1, \mu_{t-1}^{(\mu)}, x_{t-1}^{(\mu)}, D_{\tau}, x_{t-1}^{-(\mu)}, x_{t-2}^{(\mu)}, 0, \dots, 0] \quad (60)$$

Mint a (41)-ben láttuk, a (t, i) indexet lehet helyettesíteni τ -val. Végül:

$$\nabla_{\theta}\phi_i = [0, \dots, 0, A \cos(f1i), A \sin(f1i), \dots, A \sin(f12i), A \cos(f13i)] \quad (61)$$

ahol

$$A \equiv \exp \left\{ \sum_{k=1}^K [\delta_{1k} \cos(fki) + \delta_{2k} \sin(fki)] \right\} \quad (62)$$

Ezzel meghatároztuk a deriváltakat.

8.2.4. Célfüggvény

A (42)-ben felírt célfüggvény inputként tartalmazza a keresett θ vektort és az x forgalom adatsor becslési időszakra vonatkozó szeletét. A függvényérték megadásához először előregöngyölítés szükséges, hogy megkapjuk η , μ és ezek sztenderdizáltjainak, valamint ϕ -nek az értékét a teljes becslési időszakra. A 8.1.2. pontban látott kezdőértékek segítségével ez végrehajtható.

Mindezek után

$$u_{\tau} = \frac{x_{\tau}}{\eta_t \phi_i \mu_{\tau}} - 1 \quad (63)$$

egy skalár, ami kiszámolható minden megfigyelésre.

Ezek után (43) egyszerűen a skalárral osztott deriváltak összege. Vagyis ez is egy (1×35) -ös sorvektor lesz, és mivel a deriváltak *máshol voltak nullák*, a_{τ} gyakorlatilag *egymás mellé* gyűjti az eddigi információkat.

A paraméterek becsléséhez a (42) egyenletet kell megoldani, melyet itt megismételtek:

$$(B \equiv) \frac{1}{N} \sum_{\tau=1}^N a_{\tau} u_{\tau} = 0 \quad (64)$$

ahol N a megfigyelések száma. A megoldást numerikus módszerrel keresem, méghozzá a következő módon:

$$BB^T \rightarrow \min_{\theta} \quad (65)$$

Néhány paraméter megkötést tennünk kell. Egyrészt fent kell, hogy álljon a 3.3.4. pontban említett

$$\alpha_0^{(\mu)} = 1 - \beta_1^{(\mu)} - \alpha_1^{(\mu)} \quad (66)$$

Másrészt, habár a cikk ezt külön nem emeli ki, természetesen fontos az is, hogy az AR paraméterek $(\beta_1^{(\eta)}, \alpha_1^{(\eta)}, \gamma_1^{(\eta)}, \beta_1^{(\mu)}, \alpha_1^{(\mu)}, \gamma_1^{(\mu)}, \alpha_2^{(\mu)})$ abszolút értékben ne haladják meg az egységet.

Miután (65) megoldásával megkaptuk a θ értékét, (27) szerint előregöngyölítés segítségével már megadható a forgalom előrejelzése is.

8.3. Becslések

Brownlees et al. (2011) a négy modell változat közül (A-B-C-D) egyértelműen a (D) verziót találja legjobbnak, ezért csak ezt fogom tesztelni.

Amint azt a 8.2.3. pontban láttuk, θ harmincöt elemű, ezért (33) figyelembe vételével alap esetben harmincnégy paramétert kell becsülni. Ugyanakkor a szerzők is megjegyzik, hogy az (54) egyenletben csökkenthető a tagok száma, nincs szükség mind a huszonötre. Ez érthető is, hiszen a periodikus komponens, vagyis a 26 pontból álló U alakot próbáljuk leírni vele.

Mivel nem adják meg pontosan, hogy hány taggal dolgoznak, a becslések során erről döntenem kellett. Ebben az egyik szempont, hogy elég paraméter legyen ahhoz, hogy megfeleljen a célnak, vagyis, egyrészt lehessen U alakja, másrészt pedig engedjünk meg eltéréseket a szabályos U alaktól, hiszen az csak stilizált tény, nem törvényszerűség. Ugyanakkor túl sokat sem érdemes megtartani, hiszen amellett, hogy nehezítené az optimalizálást, esetleg fennállna annak a veszélye, hogy a túlzott rugalmassága miatt az idősrnak olyan tulajdonságait is magára veszi, amit nem célunk a periodikus komponensben megragadni.

Mindezek és célirányos tesztelések alapján végül négy tagot tartottam meg, és a továbbiakban végig ennyivel dolgoztam. Ezzel a választással tizenháromra, vagyis huszonegyre csökken a becslendő paraméterek száma.

8.3.1. A θ induló értékei

Már csak egy kérdés maradt mielőtt ténylegesen meg tudnánk becsülni a modellt. Ez pedig nem más, mint az, hogy milyen θ értékről indítjuk az optimumkeresést. Ennek eldöntésére végeztem előzetes tesztekkel különböző időszakokra és részvényekre, amiből az látszik, hogy ennek a kérdésnek döntő jelentősége van a becslés sikere szempontjából. Ez persze nem meglepő, hiszen ez a legtöbb, lokális szélsőértékekkel rendelkező függvény numerikus optimalizálására igaz.

Hosszas tesztelés eredményeként sikerült odaáig eljutnom, hogy a véletlenszerűen kiválasztott napok és részvények egy részén sikerüljön értelmes eredményt elérnem a modellel⁵, aminek köszönhetően lett egy előzetes képem a megtalálni kívánt θ értékéről.

Az a probléma, hogy a numerikus solverek alapvetően lokális szélsőértékek megtalálására alkalmasak, és az általam vizsgált 13 paraméteres feladat megoldása közben könnyen beragadnak egy-egy ilyenbe. Ennek áthidalására egy gyakran jól működő megoldást nyújt a MATLAB GlobalSearch algoritmus, de ennél a feladatnál ez sem volt elég. Ezért minden optimumkeresést megelőzően végeztem egy előzetes vizsgálatot is, aminek keretében a tesztelés során nyert információ alapján szisztematikusan kialakított pontháló pontjaira ellenőrzöm a célfüggvény értékét, és a minimális eredményt adó értékről indítom a (globális) solver⁶.

Ez utóbbi előzetes vizsgálat nagyon sikeres tud lenni a végső eredmény szempontjából, azonban jelentősen meghosszabbíthatja a futási időt. Mivel minden részvény minden napjára új optimalizálást kell futtatnom, ez az adatbázisom tekintetében (l. 11. táblázat) azt jelenti, hogy ha egy adott részvény adott napjára egy másodperc alatt eredményre jutok, akkor a teljes futás 24,3 órát vesz igénybe. Ezért úgy alakítottam ki a ponthálót, hogy nagyságrendileg 60 másodperc alatt lefusson egy nap egy részvényére a becslés és az előrejelzés együtt. Ily módon nagyjából 60 nap alatt fut le a program.⁷ A rövidebb futási idő reményében próbálkoztam kisebb ponthálóval is, de úgy nem kaptam elfogadható eredményt. A pontháló további növelésével természetesen javítható lenne az eredményes becs-

⁵Akkor tekintek egy eredményt „értelmes”-nek, ha az általa megadott előrejelzés legalább nagyságrendileg közel esik a tényekhez.

⁶Ezt az eljárást szokás „scatter search”-nek is nevezni.

⁷Ez természetesen gépidőben értendő. Mivel részvényenként és naponként is független a feladat, több számítógép párhuzamos használata esetén a tényleges várakozási idő csökkenthető. Az említett 60 másodperc a rendelkezésemre álló hat darab, nem teljesen azonos teljesítményű számítógép átlagában értendő.

| | MSE | | MAPE | |
|------------------------|----------|-----------|--------|----------|
| | BDF_AR | BCG_D | BDF_AR | BCG_D |
| Hány részvénynél nyert | 33 | 0 | 33 | 0 |
| Átlagos érték | 6,49E-04 | 1,53E+133 | 40,3% | 1,61E+65 |

15. táblázat. A BDF_AR és a BCG_D előrejelzésének összevetése
Forrás: Saját szerkesztés

lés esélye, de az alternatív módszerekkel összehasonlítva ez már így is rendkívül lassan vezet eredményre.

8.3.2. Különböző változatok

A 8.3.1. pontban leírt módon eljárva sem garantált, hogy minden részvény minden napjára *értelmes* eredményt kapok. Mivel egy-egy igazán nagy tévedés is jelentősen ronthatja a végén a kiértékeléskor mutatott teljesítményt, ennek áthidalására további módosításokat is teszteltem.

8.3.2.1. A (D) változat

Elsőként mindenben az eddig leírt módon futtattam le a modellt, melyre a továbbiakban BCG_D változatként utalok. Habár (egyedi ellenőrzések alapján) gyakran elfogadható eredményt kaptam, sajnos így becsülve olyan gyakori volt a hibás nap⁸, hogy összességében már egy részvénynél sem kaptam elfogadható eredményt. Emiatt a BDF modellel egyelőre nincs is értelme összevetni az előrejelzést, de az áttekinthetőség kedvéért mégis megteszem. A BDF_AR modellt választva benchmarknak a 15. táblázatban találjuk az eredményeket. Az átlagos értékekből nyilvánvalóan látható, hogy tovább kell próbálkozni a modell becsülésével.

8.3.2.2. Egyéb beállítások

A fentiek után megróbáltam néhány beállításon változtatni az alábbiak szerint:

- különböző solver algoritmusok használata,
- a kiértékelhető függvények maximális számának nagyságrendekkel való növelése,

⁸Az előrejelzés tekintetében hibásnak tekintek egy napot, ha teljesen nyilvánvalóan rossz az érték nagyságrendje (példaként l. a 15. táblázat átlagos értékeit).

- az iterációs szám nagyságrendekkel való növelése,
- a becslésre használt 20 napos időszak 200-ra növelése (520 helyett 5200 megfigyelés).

A fentiek egyike sem oldotta meg az instabil becslés problémáját (külön és együtt alkalmazva sem). Mivel egyik sem hozott kimutatható javulást, a futási időt viszont jelentősen (tovább) növelik, ezért a későbbiekben nem alkalmaztam ezeket a módosításokat.

8.3.2.3. Az „egybecsles” beállítás

A cikkben említik, hogy a becsült paraméterek viszonylag stabilnak tekinthetők a különböző részvények és időszakok tekintetében is. Ebből kiindulva módosítottam a becselő programon az alábbiak szerint.

Első lépésben minden egyes részvényre megkeresem az első olyan napot, ahol *értelmes* az eredmény⁹. Ez 22 részvélynél rögtön az első nap, és csak 3 részvényenél későbbi, mint az ötödik nap. Ezáltal minden részvényre kaptam egy-egy jól használható θ értéket. Második lépésben minden részvényre az így kapott, hozzá tartozó θ értékből jeleztem előre a teljes időszakon.

Erre a megoldásra a továbbiakban BCG_egybecsles változatként utalok, mivel minden részvényre csak egy becslést használok a teljes időszakra. A korábbiakkal ellentétben az így kapott eredményeket már érdemes elemezni. A 16. táblázatból láthatjuk, hogy habár a BDF_AR modell MSE alapon egyértelműen jobb, és MAPE alapon is jobbnak tűnik (átlagban jobb és többször nyer), ez utóbbi már korántsem olyan fölényes győzelem. Mivel a két BDF modell közül MAPE alapon a BDF_SETAR volt jobb, a fentiek után érdemes ezzel is összevetni a BCG_egybecsles modellt, ezt mutatja a 17. táblázat. Láthatóan MAPE alapon nem változott az arány az egyedi részvények tekintetében (a BDF jobb továbbra is), az MSE szerint pedig a BDF_SETAR is egyértelműen jobban teljesít a BCG_egybecsles változatnál.

Ezek után érdekes lehet megnézni, hogy az U-módszerrel szemben hogyan teljesít a BCG_egybecsles. Ebben nyújt segítséget a 18. táblázat, mely alapján megállapítható, hogy ebben az összevetésben az U-módszer határozottan alul marad.

⁹Mint már említettem, és ahogyan a 15. táblázat is sugallja, ennek eldöntése minden esetben teljesen egyértelmű, mert ha hiba van, akkor az sok nagyságrendnyi.

| | MSE | | MAPE | |
|----------------------|------------|--------------------|------------|--------------------|
| | BDF_ AR | BCG_ egybecsles | BDF_ AR | BCG_ egybecsles |
| Hány részvényt nyert | 32 | 1 | 17 | 16 |
| Átlagos érték | 6,49E-04 | 7,81E-04 | 40,3% | 41,4% |

16. táblázat. A BDF_AR és a BCG_egybecsles előrejelzésének összevetése
Forrás: Saját szerkesztés

| | MSE | | MAPE | |
|----------------------|---------------|--------------------|---------------|--------------------|
| | BDF_ SETAR | BCG_ egybecsles | BDF_ SETAR | BCG_ egybecsles |
| Hány részvényt nyert | 29 | 4 | 17 | 16 |
| Átlagos érték | 6,60E-04 | 7,81E-04 | 39,9% | 41,4% |

17. táblázat. A BDF_SETAR és a BCG_egybecsles előrejelzésének összevetése
Forrás: Saját szerkesztés

| | MSE | | MAPE | |
|----------------------|----------|--------------------|-------|--------------------|
| | U | BCG_ egybecsles | U | BCG_ egybecsles |
| Hány részvényt nyert | 8 | 25 | 3 | 30 |
| Átlagos érték | 1,02E-03 | 7,81E-04 | 50,3% | 41,4% |

18. táblázat. Az U-módszer és a BCG_egybecsles előrejelzésének összevetése
Forrás: Saját szerkesztés

Mindezek alapján kijelenthetjük, hogy a BCG modell alapvetően jó is lehet, sőt, a cikk eredményét, miszerint az U-módszernél jobb, ebben a pontban a választott hibamértékek alapján sikerült alátámasztani. Ugyanakkor minden tesztelés és finomhangolás ellenére nagyon esetleges a paraméterbecslés sikeressége. Ebben a pontban ezt úgy hidaltam át, hogy az első jó eredmény elmentése után nem becsültem újat (így a futási idő is jelentősen rövidült). Azonban hosszabb időszakon (napon belüli adatoknál ilyen az általam vizsgált 11 éves időszak is) a változatlan paraméterekkel való számolás mellett nehéz érvelni, ezért még néhány további kísérletet teszek ennek a problémának az áthidalására.

8.3.2.4. Az „elozokezdo” beállítás

Kiindulva ismét abból, hogy időben stabilnak tekinthető a θ értéke (melyet határozottan alátámasztott a BCG_egybecsles változat is), egy másik ötlet az lehet, hogy a ponthálón való keresést helyettesítem azzal, hogy mindig az előző naphoz megtalált θ értékről indítom a globális solvert (ezzel némi futási idő is megspórolható). Az első naphoz minden részvénynél azt a θ értéket használom, amit az előző pontban megtaláltam a BCG_egybecsles esetén, ezzel elkerülendő, hogy egy rossz eredményt adó kezdő értékről indulva végig hordozzam annak hatását a teljes időszakra.

Ez a megközelítés azért is tűnhet jó ötletnek, mert két egymás utáni nap húsz napos becslési időszakából tizenkilenc nap teljesen megegyezik, és csak egy nap tér el, vagyis vélhetően jó kiindulási alap az előző θ . Az ilyen beállításokkal futtatott becslésre a továbbiakban BCG_elozokezdo néven utalok, mivel az előző becslés eredményét tekintem kezdő értéknek a következő becslés során.

Sajnos az így kapott eredmények ismét használhatatlanok. Az eredmény még rosszabb, mint amit a 15. táblázatban a BCG_D változatnál láttunk, ezért ezt felesleges is táblázatba foglalni.

A rossz eredménynek az lehet az oka, hogy ha egyszer félremegy az optimalizálás, akkor az így kapott rossz θ értékből indulva a következő napon feltehetően még kevésbé valószínű, hogy jó eredmény születik. Ugyanakkor az egymást követő becslési időszakok közti 95%-os egyezés, illetve a BCG_egybecsles eredményének tükrében mégis meglepő, hogy egyáltalán előfordulhat rossz nap.

8.3.2.5. A „check” beállítás

Tesztek még egy kísérletet a BCG_egybecsles azon problémájának orvoslására, hogy nehezen védhető ilyen hosszú időszakra a változatlan paraméterek használata.

Egy lehetséges megoldás lehetne, hogy időszakonként újra megkeresem a legközelebbi jó napot, és egy ideig az így kapott θ értéket használom. Ez a módszer feltehetően hozna némi javulást a BCG_egybecsles-hez képest, de túlságosan önkényes, hogy milyen gyakran és hol keresem a jó eredményt, ezért ezzel inkább nem foglalkozom.

A BCG_elozokezdo sikertelenségét látva az említett 95%-os átfedés ellenére elvettem ezt az irányt, és visszatérek a ponthálós módszerhez. Emellett érdemesnek tartom naponta újra keresni a paramétereket, de megpróbálom kihasználni a korábban már megtalált jó θ értékeket is.

A fentiek ötvözeteként a továbbiakban BCG_check változatként fogok utalni az alábbi módszerre.

Az első napra előrejelzek ugyanabból a θ értékből, amit a BCG_egybecsles esetén használtam (ez a mintában megtalált első jó eredmény volt, ami 22 esetben az első naphoz köthető). Ezt láttuk, hogy működik. A későbbiekben az optimumkeresést mindig a már látott ponthálós módszerrel indítom. Ezen a ponton két eset lehetséges:

1. Ha az így kapott optimalizálás értelmes előrejelzést ad (mint említettem, ez teljesen egyértelműen eldönthető), akkor ezt tekintem előrejelzésnek.
2. Amennyiben nem értelmes az így kapott előrejelzés, visszatérek az előző napi θ értékéhez, és ebből jelzek előre.

Ez tehát a BCG_egybecsles láthatóan működő módszerét annyiban egészíti ki, hogy minden nap esélyt adok a θ frissítésére azáltal, hogy megbecslem a paramétereket, de berakok egy ellenőrzést, és ha nem jó az eredmény, a legutóbbi működő θ értéket használom. Ily módon azonban a BCG_egybecsles után ismét visszatértünk a nagyságrendileg 60 napos futási időhöz.

Elsőként nézzük meg, sikerült-e javítani a BCG_egybecsles teljesítményén. A 19. táblázat szerint határozottan igen, noha MAPE alapon a javulás szerényebb.

| | MSE | | MAPE | |
|------------------------|-----------------|------------|-----------------|------------|
| | BCG_ egybecsles | BCG_ check | BCG_ egybecsles | BCG_ check |
| Hány részvénynél nyert | 6 | 27 | 15 | 18 |
| Átlagos érték | 7,81E-04 | 6,77E-04 | 41,4% | 40,2% |

19. táblázat. A BCG_ egybecsles és a BCG_ check előrejelzésének összevetése
Forrás: Saját szerkesztés

| | MSE | | MAPE | |
|------------------------|----------|------------|---------|------------|
| | BDF_ AR | BCG_ check | BDF_ AR | BCG_ check |
| Hány részvénynél nyert | 31 | 2 | 21 | 12 |
| Átlagos érték | 6,49E-04 | 6,77E-04 | 40,3% | 40,2% |

20. táblázat. A BDF_ AR és a BCG_ check előrejelzésének összevetése
Forrás: Saját szerkesztés

A fentiek alapján ez az eddigi legjobb BCG becslés. A BDF_ AR eredményével való összevetése látható a 20. táblázatban, ahol azt látjuk, hogy MSE alapon egyértelműen jobb a BDF_ AR, és MAPE alapon is jobbnak tűnik, noha összességében minimálisan a BCG_ check ad kisebb hibát.

A 21. táblázatban a BDF_ SETAR-ral való összevetés látható, amely minden mutató szerint jobban szerepel, mint a BCG_ check.

Végül a 22. táblázatban hasonlítsuk össze a BCG_ check változatot az U-módszerrel is. Természetesen, mivel már a BCG_ egybecsles-t is jobbnak találtuk az U-módszernél, az ahhoz képest javulást hozó BCG_ check esetén is ugyanez mondható el.

Ezzel tehát sikerült megbecsülni a BCG modellt. Több próbálkozás után a BCG_ check beállítással sikerült elfogadható eredményre jutni úgy is, hogy a paraméterek rendszeresen frissülnek (ha nem is feltétlenül naponta). Mivel

| | MSE | | MAPE | |
|------------------------|------------|------------|------------|------------|
| | BDF_ SETAR | BCG_ check | BDF_ SETAR | BCG_ check |
| Hány részvénynél nyert | 28 | 5 | 26 | 7 |
| Átlagos érték | 6,60E-04 | 6,77E-04 | 39,9% | 40,2% |

21. táblázat. A BDF_ SETAR és a BCG_ check előrejelzésének összevetése
Forrás: Saját szerkesztés

| | MSE | | MAPE | |
|------------------------|----------|-----------|-------|-----------|
| | U | BCG_check | U | BCG_check |
| Hány részvénynél nyert | 5 | 28 | 0 | 33 |
| Átlagos érték | 1,02E-03 | 6,77E-04 | 50,3% | 40,2% |

22. táblázat. Az U-módszer és a BCG_check előrejelzésének összevetése

Forrás: Saját szerkesztés

így jobb eredményt kaptunk minden egyéb változathoz képest, a továbbiakban a BCG_check verziót tekintem a BCG modell becslésének.

9. fejezet

Benchmark választás

A III. rész végére érve nem maradt más hátra, mint kiválasztani az irodalom legjobb modelljét, ezáltal kijelölve a benchmarkot a későbbi modell kereséshez.

Amint az előző fejezetekben láthattuk, habár az U módszert mindkét modell megverte az általam használt adatbázison is, a választott hibamértékek alapján egyértelműen a BDF modell tekinthető jobbnak a BCG-vel szemben. A BDF_SETAR minden szempont szerint jobban teljesít a BCG modellnél, és a BDF_AR is csak hajszálnyival marad el a MAPE átlagban, egyéb szempontok szerint pedig szintén jobb. Annyi bizonytalanság maradt csak ez ügyben, hogy a BDF_AR és BDF_SETAR változatok között nem tudunk dönteni (ahogyan a cikk szerzői sem tudtak ezen általuk javasolt verziók között), hiszen MSE alapján az előbbi, MAPE alapján pedig az utóbbi adott jobb eredményt.

Mint azt már korábban említettem, az itt alkalmazottaktól eltérő módon is mérhető a teljesítmény, és a későbbiekben ebből a szempontból is tovább fogunk lépni. Ennek ellenére a BDF és BCG modellek összehasonlítását nem látom szükségesnek további módokon is elvégezni. Ennek elsősorban nem a BDF modell itt alkalmazott hibamérték szerinti meggyőzően jobb teljesítménye az oka, sokkal inkább a BCG modell becslésének a nehézsége és a kimenet meglehetősen esetleges volta.

A BCG modell saját adatokon történő reprodukálását nehezítette, hogy két kezdőérték esetében saját feltevással kellett élnem, illetve a Fourier reprezentáció esetén sem volt egyértelmű, hogy hány taggal dolgoztak. Ezen felül a θ kezdő értékek meghatározásához sem nyújt támpontot a cikk. Igaz ugyan, hogy ez utóbbi pusztán az optimalizálás során felmerülő technikai kérdés, ugyanakkor mint

láttuk, az eredmények szempontjából ennek döntő jelentősége van.

A BCG modell további hátránya a reprodukálás nehézségén túl, hogy rendkívül lassan vezet eredményre (az általam használt adatbázison nagyságrendileg 60 nap alatt, szemben a BDF modell változatok együttesen is csak néhány óras futási idejével). Ráadásul az eredmény még így is nagyon esetleges maradt, csak különböző *trükkök* árán sikerült egyáltalán értelmezhető eredményre jutni¹⁰. A lényegesen egyszerűbb és transzparensőbb BDF modell pedig még mindezek után is jobban szerepelt.

Összességében az adatbázison végzett összehasonlítás alapján a szakirodalomban található modellek közül a BDF modellt találtam részvények napon belüli forgalom előrejelzésére a legalkalmasabbnak, ezért a modell keresés során a IV. részben ezt (pontosabban ennek két változatát) tekintem benchmarknak.

¹⁰Véleményem szerint kétséges, hogy mindez pusztán abból eredne, hogy ETF-ek helyett részvényekre alkalmazom a BCG modellt. Ezt a kérdést azonban jelen dolgozatban külön nem vizsgálom.

IV. rész

Saját modellek

10. fejezet

Modell keresés

Miután a III. részben sikerült azonosítani a benchmarkot, a következő céloom olyan modellt találni, amely az általam használt adatbázison a benchmarknál jobban teljesít. Az egyes modellek teljesítményét ebben a fejezetben is a 6.1. pontban bemutatott módon, tehát az eddigiekhez hasonlóan mérem. Azokat a modelleket, amelyek ez alapján a sztenderdnek tekinthető összevetés alapján versenyben maradnak, a IV. rész későbbi fejezeteiben további szempontok mentén is tesztelni fogom.

10.1. U dekompozíció nélküli modellek

Elsőként olyan modelleket tesztelek, amelyek nem veszik figyelembe a 3.1.2. pontban a stilizált tények közül legfontosabbként megjelölt napon belüli U alakot, vagyis nem bontják külön U alak és maradék részre az idősort. Nyilvánvalóan nem véletlen, hogy a napon belüli forgalom előrejelzéssel foglalkozó modellek közül az irodalomban kivétel nélkül mindegyik valamilyen módon beépíti a modellbe az U alak leválasztását. Emiatt nagy reményeket nem is fűzök azokhoz a modellekhez, amelyek ezt nem teszik meg, ugyanakkor talán mégsem felesleges ilyen irányban is vizsgálni. Egyrészt elvileg lehet valami marginális esélye annak is, hogy akár egy ilyen modell is jól szerepeljen (érhet meglepetés), másrészt pedig általánosságban is érdemes lehet az egyszerűbb modellektől haladni a bonyolultabbak felé, mert közben esetleg a vizsgált folyamatról is szerezhethetünk ismereteket, és kiderülhet, hogy milyen aspektusok modellezését érdemes finomítani vagy elhagyni.

| | MSE | | MAPE | |
|----------------------|----------|----------|-------|-------|
| | AR_26 | U | AR_26 | U |
| Hány részvényt nyert | 2 | 31 | 1 | 32 |
| Átlagos érték | 1,08E-03 | 1,02E-03 | 55,4% | 50,3% |

23. táblázat. Az AR_26 és az U-módszer előrejelzésének összevetése
 Forrás: Saját szerkesztés

10.1.1. Egyszerű AR modellek

Első lépésben ellenőriztem a legegyszerűbb, AR(1) modellt is, de valójában ettől nem igazán várhattuk, hogy jó legyen, és nem is lett az.

Az U alakot egyelőre nem modellezzük külön, de az U-módszer (l. (14)) logikájából kiindulva érdemes lehet megnézni egy olyan AR modellt, ahol a késleltetés éppen egy napos (ami esetemben 26 megfigyelés), vagyis az előző nap azonos időszakával magyarázom a következő megfigyelést.

$$\tilde{y}_{t,i} = c_i + \beta_1 y_{t-26,i} + \varepsilon_{t,i} \quad (67)$$

ahol t az időszakot, i a részvényt jelölő index, ε pedig fehér zaj. Erre a továbbiakban AR_26 modellként utalok. Ennek az értékelését láthatjuk a 23. táblázatban, az U-módszerrel összevetésben. Láthatóan egyértelműen rosszabb, mint az U-módszer, ezért az effektív benchmarkkal (BDF modell) nem is érdemes külön összehasonlítani. Ugyanakkor azt is érdemes megjegyezni, hogy az eredménye (a hibák átlagos értéke alapján) nagyságrendileg nem esik távol az U-módszer eredményétől.

Ez utóbbi megfigyelésből kiindulva egészítsük ki az AR_26 modellt egy olyan taggal, ami figyelembe veszi a legfrissebb információt is, azaz:

$$\tilde{y}_{t,i} = c_i + \beta_1 y_{t-1,i} + \beta_2 y_{t-26,i} + \varepsilon_{t,i} \quad (68)$$

Erre a továbbiakban AR_1.26 modellként utalok. Vessük össze ezt is az U-módszerrel, ebben segít a 24. táblázat. Láthatóan ez már lényegesen jobb az U-módszernél, ezért érdemes lesz a BDF modellel is összehasonlítani. A 25. táblázat alapján láthatjuk, hogy a BDF_AR modell egyértelműen jobban szerepel, de az átlagos hibák értéke hasonló nagyságrendű.

Számomra jelentett némi meglepetést, hogy a gyakorlatban elterjedt U-módszert ilyen egyszerűen sikerült megverni, ugyanis a korábban részletesen megismert BDF

| | MSE | | MAPE | |
|----------------------|----------|----------|---------|-------|
| | AR_1.26 | U | AR_1.26 | U |
| Hány részvényt nyert | 32 | 1 | 33 | 0 |
| Átlagos érték | 6,74E-04 | 1,02E-03 | 42,4% | 50,3% |

24. táblázat. Az AR_1.26 és az U-módszer előrejelzésének összevetése
Forrás: Saját szerkesztés

| | MSE | | MAPE | |
|----------------------|----------|----------|---------|--------|
| | AR_1.26 | BDF_AR | AR_1.26 | BDF_AR |
| Hány részvényt nyert | 2 | 31 | 1 | 32 |
| Átlagos érték | 6,74E-04 | 6,49E-04 | 42,4% | 40,3% |

25. táblázat. Az AR_1.26 és a BDF_AR előrejelzésének összevetése
Forrás: Saját szerkesztés

és BCG modell eredménye is kizárólag ennek a modellnek a megverése volt. Tétel-összehasonlítást felesleges lenne végezni egy olyan modellel, amit nem tekintek benchmarknak, de megjegyzem, hogy a BCG_check modell átlagos MSE értéke (6,77E-04, l. 22. táblázat) minimálisan ugyan, de magasabb volt, mint a lényegesen egyszerűbb AR_1.26 modell megfelelő értéke (6,74E-04).

10.1.2. A hét napja hatás

A 3.1.2. pont alapján a stilizált tények közül az U alak mellett a hét napja hatás is érdemesnek tűnik a vizsgálatra. Mivel az irodalomban nincs konszenzus abban, hogy a hét napja hatás pontosan miben áll (csak abban, hogy jelen van), röviden megvizsgáltam az adatbázisomat ilyen szempontból. Mivel az adatbázis egy 11 éves időszakot ölel fel, egy részletes vizsgálat keretében esetleg rövidebb időszakokra bontva érdemes lenne almintákat is vizsgálni. Egyelőre azonban időben nem, csak részvényenként bontom meg az adatbázist.

Minden részvényre kiszámoltam az egyes munkanapokra eső átlagos forgalmat, és részvényenként növekvő sorrendbe rendeztem a napokat ez alapján. Ennek az eredménye látható összefoglalva a 26. táblázatban. Alapvetően a határozottan alacsony, vagy a határozottan magas forgalmú nap az, ami segíthet az előrejelzésben. Ezzel szemben az olyan nap, ami jellemzőem átlagos forgalmú, nem sokat segít.

Azt láthatjuk, hogy a hétfő egy kivétellel minden részvényt a minimális

| | min | 2. | 3. | 4. | max |
|-----------|-----|----|----|----|-----|
| hétfő | 32 | 0 | 1 | 0 | 0 |
| kedd | 0 | 16 | 6 | 5 | 6 |
| szerda | 0 | 4 | 9 | 10 | 10 |
| csütörtök | 0 | 1 | 6 | 15 | 11 |
| péntek | 1 | 12 | 11 | 3 | 6 |

26. táblázat. Hét napja hatás a mintában 33 részvényre

Forrás: Saját szerkesztés

forgalmú nap volt, annál az egynél pedig a harmadik legkisebb. A kedd előfordult, hogy maximális volt, de jellemzően inkább átlagos. A szerda talán inkább magas, mint alacsony forgalmú, a csütörtökre ugyanez még inkább elmondható. A péntek az egyedüli nap, amelyik minden kategóriában szerepel.

Ezek alapján a hétfő minimális forgalmúnak tekinthető, azonban a teljes mintára más egyértelmű megállapítást nem tudunk tenni.

10.1.2.1. Az AR_1.26 modell kiegészítése

Az adatbázison végzett rövid előzetes vizsgálat alapján tehát csak a hétfő modellbe foglalásával érdemes próbálkozni. Egészítsük ki az AR_1.26 modellt a következő módon:

$$\tilde{y}_{t,i} = c_i + \beta_1 y_{t-1,i} + \beta_2 y_{t-26,i} + \beta_3 H + \varepsilon_{t,i} \quad (69)$$

ahol H az alábbi dummy változó:

$$H = \begin{cases} 1 & \text{hétfőn} \\ 0 & \text{egyébként} \end{cases} \quad (70)$$

Nevezzük ezt ARH_1.26 modellnek. Ennek az eredményét láthatjuk a 27. táblázatban a dummy nélküli változattal történő összevetésben. Ha nem is jelentősen, de a legtöbb részvénynél, és átlagban is romlott a teljesítmény mindkét hibamérték szerint. A hét napja hatás figyelembe vétele ez alapján nem tűnik hasznosnak az előrejelzésben, noha természetesen a modellbe foglalásnak nem ez az egyetlen lehetséges módja.

| | MSE | | MAPE | |
|----------------------|----------|----------|---------|----------|
| | AR_1.26 | ARH_1.26 | AR_1.26 | ARH_1.26 |
| Hány részvényt nyert | 30 | 3 | 33 | 0 |
| Átlagos érték | 6,74E-04 | 6,76E-04 | 42,4% | 42,8% |

27. táblázat. Az AR_1.26 és az ARH_1.26 előrejelzésének összevetése

Forrás: Saját szerkesztés

10.1.3. Ármozgás mutatók

A technikai elemzők körében köztudottnak tekinthető, hogy az ármozgást nem önmagában, hanem a forgalommal együtt érdemes figyelni és értelmezni. A szakirodalomban is találhatunk olyan vizsgálatokat, melyek megerősítik, hogy részvények esetén a forgalom és az árváltozás Granger-oksági viszonyba hozható, habár az eredmények a hét napja hatáshoz hasonlóan nem egybehangzóak. A tesztelés módjától függően Chiang et al. (2010) mindkét irányú okságot ki tudja mutatni.

Amint a 6. táblázatból láttuk, az adatbázisom a volumen mellett tartalmazza a szokásos ár adatokat is (nyitó, legmagasabb, legalacsonyabb, záró) percenkénti bontásban, ezekből pedig többféle ármozgás mutató számítására is lehetőség van. Néhányat ezek közül ki is számoltam.

Loghozam. A loghozam maga is tekinthető árváltozás mutatónak. A 15 perces loghozamot az adott időszakhoz tartozó záróárakkal számoltam.

Volatilitás. A loghozam szórását szokás volatilitásnak nevezni, amely egy gyakran használt mutató a pénzügyekben. Az általam számolt volatilitás 15 perces időszakra vonatkozik (tehát nem egy évre, ami a szokásos használata).

Rés. Rés (gap) alatt egy adott időszak nyitó, és a megelőző időszak záró értéke közötti eltérést értjük. Ezt általában napok között szokás figyelni, de ugyanúgy értelmezhető 15 perces lépésenként is.

Ársáv. Ársáv (range) alatt az adott időszak (jelen esetben 15 perc) legmagasabb és legalacsonyabb értéke közötti különbséget értjük. Értelmezhető százalékos formában is, ez esetben a legmagasabb érték százalékaként határoztam meg a százalékos ársávot.

| | Szignifikancia szint | | |
|---------------------|----------------------|----|-----|
| | 1% | 5% | 10% |
| Loghozam | 26 | 32 | 32 |
| Volatilitás | 33 | 33 | 33 |
| Rés | 33 | 33 | 33 |
| Ársáv | 33 | 33 | 33 |
| Ársáv (%) | 33 | 33 | 33 |
| Tényleges ársáv | 33 | 33 | 33 |
| Tényleges ársáv (%) | 33 | 33 | 33 |

28. táblázat. Az ármozgás és a forgalom kapcsolata

Nullhipotézis: Az ármozgás nem Granger-okozza a forgalmat

A táblázat azon részvények számát mutatja az adott szignifikancia szinteken, amelyeknél a nullhipotézis elvethető

Forrás: Saját szerkesztés

Tényleges ársáv. Tényleges ársáv (true range) alatt a következő kifejezést értjük:

$$TR_t = \max(H_t, C_{t-1}) - \min(L_t, C_{t-1}) \quad (71)$$

ahol H az időszak legmagasabb, L a legalacsonyabb ára, C pedig a zárár. A százalékos tényleges ársávot $\max(H_t, C_{t-1})$, vagyis ismét a magasabb érték arányában számoltam.

Az adatbázison minden részvényre ellenőriztem az összes mutatót, az eredmények összefoglalását a 28. táblázat mutatja. A vizsgált nullhipotézis az, hogy az ármozgás nem Granger-okozza a forgalmat. Ezt az esetek zömében el lehet vetni, egyedül a loghozamra vonatkozó eredmény különbözik némileg a többitől. Az, hogy a legtöbb mutatóra hasonló eredményt kapunk, nem olyan meglepő, hiszen lényegében ugyanazt a jelenséget, vagyis az ármozgást mérik.

Természetesen a teljesebb kép érdekében lehetne rövidebb időszakokra bontva is tesztelni, de ennyi alapján is úgy gondolom, érdemes megvizsgálni, hogy ezen ármozgás mutatók valamelyike nyújthat-e érdemi segítséget a forgalom előrejelzésben. Az irodalom és a fenti vizsgálat alapján erre elvileg van esély.

10.1.3.1. Az AR_1.26 modell kiegészítése

Egészítsük ki az AR_1.26 modellt az alábbi módon:

| ARX_1.26 változatok | Hány részvénynél nyert | | Átlagos érték | |
|------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 5 | 2 | 6,75E-04 | 42,5% |
| Volatilitás | 8 | 7 | 6,71E-04 | 42,5% |
| Rés | 3 | 16 | 6,76E-04 | 42,4% |
| Ársáv | 6 | 5 | 6,73E-04 | 42,6% |
| Ársáv (%) | 5 | 4 | 6,73E-04 | 42,6% |
| Tényleges ársáv | 10 | 7 | 6,69E-04 | 42,5% |
| Tényleges ársáv (%) | 8 | 5 | 6,70E-04 | 42,6% |
| AR_1.26 | - | - | 6,74E-04 | 42,4% |

29. táblázat. Az ARX_1.26 modellek értékelése különböző ármozgás mutatók használata esetén
Összevetés az ármozgást nem tartalmazó AR_1.26 modellel
Forrás: Saját szerkesztés

$$\tilde{y}_{t,i} = c_i + \beta_1 y_{t-1,i} + \beta_2 y_{t-26,i} + \beta_3 X_{t-1,i} + \varepsilon_{t,i} \quad (72)$$

ahol X legyen a fent felsorolt ármozgás mutatók egyike, tehát ennek a késleltetésével bővítem a modellt. Nevezzük ezt ARX_1.26 modellnek.

Az így számolt modellek eredményeit láthatjuk a 29. táblázatban, ahol az összehasonlítás mindig az ármozgást nem tartalmazó AR_1.26 modellel szemben értendő, melynek átlagos értékei a legalsó sorban szerepelnek.

Az átlagos MAPE értéket egyik mutató használatával sem sikerült javítani, és az esetek többségében egyedileg is ugyanez mondható el. Egyedül a rés használata közelít ahhoz, hogy a részvények felénél javulást jelentsen. MSE alapon a loghozam és a rés kivételével elértünk némi javulást az átlagos értékben, azonban ez a javulás nem jelentős, egyedileg pedig csak a részvények kisebb részénél figyelhető meg ugyanez.

Ezek alapján nem tudjuk megerősíteni, hogy az ármozgás mutatók modellbe foglalása hasznos az előrejelzés szempontjából, de természetesen itt is vannak további lehetőségek a felhasználására.

10.1.4. Áttekintés

A 10.1. pontban az U alak tételes megragadása nélkül vizsgáltam néhány modellt, és amint az várható volt, a benchmarknak tekintett BDF modellnél nem sikerült

jobb eredményt elérni (az U-módszernél viszont igen). Az U alak helyett azonban két másik stilizált tény, a hét napja hatás, valamint a forgalom és ármozgás kapcsolatából kiindulva az ármozgás modellbe foglalásával is próbálkoztam. A hét napja hatás esetén csak a hétfő figyelembe vétele tűnt biztató lehetőségnek, de egyértelmű romlást okozott az előrejelzésben. A különböző ármozgás mutatók modellbe foglalása az itt látott formában szintén inkább rontott az eredményen, ugyanakkor egyrészt ez kevésbé volt egyértelmű (átlagban láttunk némi javulást), másrészt a modellbe foglalás az ármozgás mutatók esetén lényegesen különböző módon is megtehető. Ezek miatt erre később, egy jobb alapmodell kiegészítéseként még érdemes lehet visszatérni.

10.2. Az U alak leválasztásának különböző módjai

A továbbiakban a legfontosabbnak tartott stilizált tény, az U alak modellbe foglalása mellett keresem a benchmarknál jobb modelleket. Ennek során a III. részben látottak tanulsága alapján alapvetően a BDF modell logikájából indulok ki, vagyis a dekompozíciót additív módon közelítem meg (szemben a BCG modell multiplikatív felírásával). Ennek oka, hogy a vizsgált két modell közül az additív bizonyult jobbnak, amely ráadásul lényegesen egyszerűbb is, és a paraméterbecslése is sokkal stabilabb.

Az additív logika alapján két dolgot érdemes vizsgálni. Egyrészt kereshető módszer a leválasztandó U alak becslésére, valamint a dekomponálás után létrejövő *maradék* rész előrejelzésére is. Ebben a pontban az U alak dekomponálására keresek lehetőségeket. Az U alak leválasztása után a *maradék* rész előrejelzésére egyelőre a BDF modell által is használt AR(1) és SETAR modelleket használom annak érdekében, hogy kizárólag a dekompozíció hatását tudjam vizsgálni. Később, a 10.3. pontban külön vizsgálom majd a választott dekompozíció utáni *maradék* rész előrejelzési lehetőségeit is.

10.2.1. U-módszer

A BDF modell, mint láttuk a 3.3.3. pontban, az U alak leválasztására magas dimenziójú faktorelemzést végez. Mivel Bialkowski et al. (2008) benchmarknak az U-módszert tekinti, szinte kézenfekvően felvetődik, hogy azt ne önmagában használjuk mint előrejelzést, hanem az így kapott U alak leválasztása után a ma-

| | Hány részvényt nyert | | Átlagos érték | |
|---------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| U_AR | 27 | 17 | 6,18E-04 | 39,5% |
| U_SETAR | 19 | 30 | 6,17E-04 | 38,9% |
| BDF_AR | - | - | 6,49E-04 | 40,3% |

30. táblázat. Az U_AR és U_SETAR modellek összevetése a BDF_AR modellel
Forrás: Saját szerkesztés

radékot is jelezzük előre (egyelőre tehát kizárólag AR(1) és SETAR módszerrel), és a két rész összegét tekintjük előrejelzésnek.

A modell tehát a következő:

$$\tilde{y}_{t,i} = \tilde{u}_{t,i} + \tilde{e}_{t,i} \quad (73)$$

ahol \tilde{u} az U-módszerből kapott előrejelzés (l. (14)), \tilde{e} pedig megegyezik a BDF modellben használt egyedi részekkel (l. (20), (21) és (22)).

Az így kapott modelleket nevezzük U_AR és U_SETAR modellnek attól függően, hogy az egyedi részt hogyan modelleztük. Mivel mindkettő egyértelműen jobb a *sim*a U-módszernél, ezért az effektív benchmarkkal, vagyis a BDF modellel történő összevetést mutatom csak be.

A 30. táblázat a BDF_AR verzióval való összevetést mutatja. Azt látjuk, hogy mindkét U alapú modell egyértelműen jobb, hiszen az átlagos hibájuk alacsonyabb, és több részvényt nyertek, mint ahánynál veszítettek a 33-ból. Mivel a 10.2. pontban a cél a dekompozíció értékelése, alapvetően az U_AR modell összevetése az érdekes a 30. táblázat alapján. Az látszik, hogy átlagosan az MSE érték 4,8%-kal, a MAPE érték 2,2%-kal javult a BDF_AR-hoz képest.

A 31. táblázat a BDF_SETAR verzióval szembeállítja a két U alapú modell verziót. Itt MAPE alapon az U_AR egyedileg kevesebb részvényt nyert, a többi mutató szerint azonban jobbak az U alapú modellek. A dekompozíció értékelése miatt ebből a táblázatból az U_SETAR összevetése a fontosabb, amely MSE alapon átlagosan 3,2%-kal, MAPE alapon 2,5%-kal jobb, mint a BDF_SETAR modell.

Összességében számomra igen meglepő eredmény született. A dekompozíciót a szakirodalom ismeretében kézenfekvő U-módszerrel végezve (aminél egyszerűbbet nehéz találni) szisztematikusan jobb eredményt kaptunk, mint a benchmark

| | Hány részvélynél nyert | | Átlagos érték | |
|-----------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| U_AR | 31 | 8 | 6,18E-04 | 39,5% |
| U_SETAR | 30 | 25 | 6,17E-04 | 38,9% |
| BDF_SETAR | - | - | 6,60E-04 | 39,9% |

31. táblázat. Az U_AR és U_SETAR modellek összevetése a BDF_SETAR modellel

Forrás: Saját szerkesztés

megfelelő verziója, bármelyik módszert használtuk is az egyedi rész előrejelzésére. A javulás a használt egyedi rész modelltől és a választott hibamértéktől függően átlagosan 2,2% és 4,8% közé esik.

Mindehhez még azt is érdemes hozzátenni, hogy az ebben a pontban bemutatott módszer adatigénye töredéke a BDF modellének, utóbbi ugyanis a teljes piac, jelen esetben 33 részvény adatát használja fel minden egyedi részvény forgalmának előrejelzéséhez, míg az U_AR és U_SETAR modellek kizárólag az előrejelezni kívánt részvény adatával dolgoznak. Ugyanez az összes később vizsgált modellre is igaz lesz, ami szerintem *ceteris paribus* előnynek tekinthető.

10.2.2. Polinom

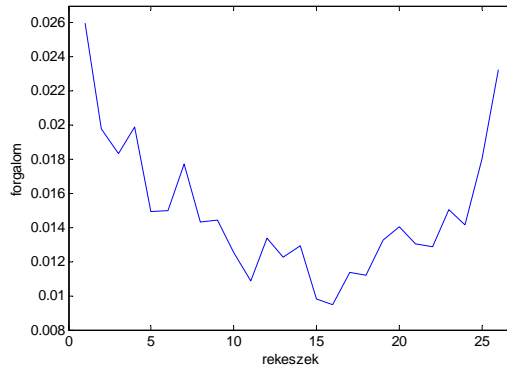
Az előző pontban tapasztalt javulás biztató arra nézve, hogy érdemes az U alak leválasztásával foglalkozni. Ebben a pontban ezért továbbra is a (73) szerint keresem a következő modellt, csak az \tilde{u} értékét máshogy próbálom meghatározni.

Az U-módszer, miután gyakorlatilag átlagolásról van szó, természetes, hogy *rücskös*, vagy úgy is mondhatjuk, hogy *zajos* U alakot eredményez, amint azt az 6. ábra is szemlélteti¹¹. Érdekes lenne 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 foksámot azonnal rögzíteni. Tekintsük az alábbi n -edfokú polinomot:

$$p_t = \sum_{i=0}^n \beta_i x_t^i \quad (74)$$

¹¹Noha enyhébb mértékben, de a BDF által dekomponált U alak is hasonló képet mutat.



6. ábra. Az U-módszer eredménye az Alcoa, Inc. mintabeli első 20 napjára illesztve
Forrás: Saját szerkesztés

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:

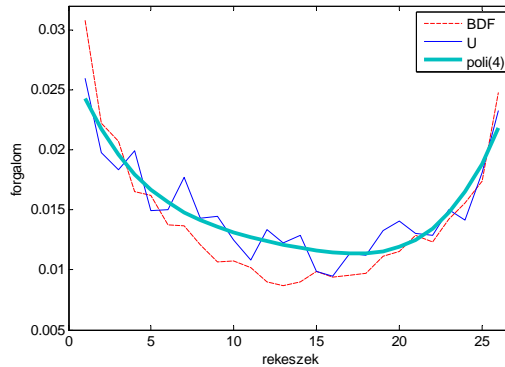
$$\sum_{j=1}^J \sum_{t=1}^T (p_t - y_{j,t})^2 \rightarrow \min_{\beta_i} \quad (75)$$

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

Illusztrációként (az előző példánál maradva) tekintsük a 7. ábrát, amelyen a 6. ábrán is látott, U-módszer által megadott görbe mellett szerepel a BDF modell által dekomponált görbe, valamint a negyedfokú polinom által megadott U alak is. Ezen a példán is láthatjuk, hogy míg a BDF határozottan más lefutású (és az U-nál kevésbé ugyan, de szintén zajos), a polinom ránézésre is tekinthető az U-módszer által megadott görbe simított változatának.

10.2.2.1. A fokszám megválasztása

A (74)-(75) felírásban az n fokszámot az optimalizálás előtt kell megválasztani, és azután fix marad. Alapvetően az $n \in \{1, 2, \dots, 26\}$ halmazon érdemes a fokszámot keresni, de természetesen a lineáristól nem sokat várunk, illetve valószínűleg



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

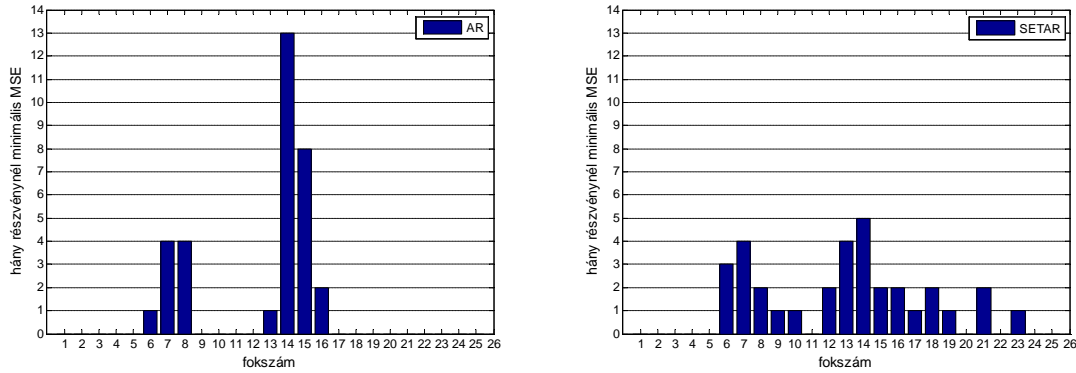
kevesebb paraméter is elég, mint ahány pontot keresünk. Ebben a pontban arra keresem a választ, hogy n milyen értékével célszerű dolgozni.

Érdeemes megjegyezni, hogy nem feltétlenül a legjobban illeszkedő polinom lesz a legjobb választás (ami jó eséllyel a legmagasabb fokszámú lenne), vagyis n megválasztásakor nem ez a fő szempont, hanem annak a modellnek az előrejelző képessége, amelynek keretében alkalmazom a polinom alapú dekompozíciót.

Mint már említettem, a 10.2. pontban a modelleket a (73) alakban keresem, ahol $\tilde{\epsilon}$ mindig ugyanaz a BDF-ből ismert két modell. A polinom alapú verzióban \tilde{u} -nak pedig az illesztett polinom pontjai felelnek meg, vagyis az U-módszerhez hasonlóan a megtalált polinom pontjait tekintem az U alak előrejelzésének (és a becslési időszakban is ez alapján számolom az egyedi részt). Jelölje az így becslött modelleket az egyedi rész számításától függően $\text{Poli}(n)\text{-AR}$ és $\text{Poli}(n)\text{-SETAR}$, ahol n a választott fokszámot mutatja.

Megbecsültem $n \in \{1, 2, \dots, 26\}$ minden elemére mindkét változatot, vagyis a $\text{Poli}(n)\text{-AR}$ -t és a $\text{Poli}(n)\text{-SETAR}$ -t is, tehát összesen 52 modellt, és összevetettem a különböző fokszámú változatok előrejelzéseit annak érdekében, hogy képet kapjak arról, milyen n értékkel érdemes dolgozni. Az egyszerűség kedvéért ebben a lépésben csak MSE alapon vizsgáltam az előrejelzéseket. A választott fokszámmal becslött modellt később majd a szokásos módon vetem össze a benchmarkkal.

Lássuk először a $\text{Poli}(n)\text{-AR}$ verziót. Minden részényre megkerestem, hogy melyik fokszám eredményezte a legkisebb MSE értéket az előrejelzésben. A 8. ábra bal oldali grafikonján ábrázoltam ezek megoszlását, ahol látszik, hogy $n = 14$



8. ábra. Minimális MSE értéket eredményező n fokszámok részvényenként a Poli(n)_AR (bal) és Poli(n)_SETAR (jobb) modellből történő előrejelzés esetén

Forrás: Saját szerkesztés

kiemelkedik, hiszen 13 részvény esetén ez volt a legpontosabb az előrejelzés során. Második helyen szerepel $n = 15$, illetve $n = 7$ és $n = 8$ szintén több részvénynél minimális MSE értéket adott.

A 8. ábra jobb oldali grafikonján ugyanezt megvizsgálhatjuk a Poli(n)_SETAR verzióra. A legmagasabb érték itt is az $n = 14$ értékhez kapcsolódik 5 részvényvel, de a helyzet kevésbé koncentrált. Közel ugyanennyi, pontosan 4 részvénynél nyert $n = 7$ és $n = 13$ is.

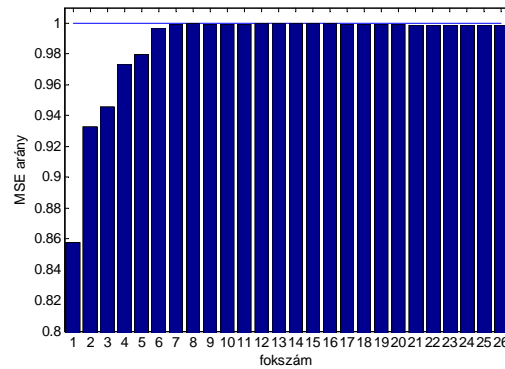
Ezek alapján az $n = 14$ tűnik a jó választásnak, mert mindkét modell verziónál ez eredményezte leggyakrabban a legpontosabb előrejelzést, noha a Poli(n)_SETAR esetén ez kevésbé kiugróan volt igaz.

A végső döntés előtt végeztem egy másik típusú vizsgálatot is, ugyanis a fenti nem tájékoztat arról, hogy mennyivel nyert a győztes fokszám az egyes esetekben. Elképzelhető, hogy olyan minimális a különbség, hogy a kérdés eleve nem is releváns, hanem véletlenszerűen bármilyen fokszám nyugodtan választható.

Ennek érdekében minden részvényre kiszámoltam az alábbi értéket:

$$MSE_{arány}(i, n) = \frac{MSE_{min}(i)}{MSE_{fokszám}(i, n)} \quad (76)$$

ahol $i \in \{1, 2, \dots, 33\}$ index a részvényt, n továbbra is a fokszámot jelöli, továbbá MSE_{min} az adott részvényre minimális MSE érték (a különböző fokszámmal futtatott előrejelzések közül), $MSE_{fokszám}$ pedig az adott részvényre az adott fokszám mellett kapott MSE érték. Ez alapján tehát az $MSE_{arány}$ értéke minden



9. ábra. Poli(n)_AR modellből számolt $MSE_{arány}$ különböző fokszámokra az Alcoa, Inc. esetén
Forrás: Saját szerkesztés

részvényre egyszer felvesz 100%-ot, a többi fokszámra pedig ennél kisebb értéket.

A 9. ábrán minden fokszámra ábrázoltam az $MSE_{arány}$ értékeit az Alcoa, Inc. példáján a Poli(n)_AR verzióból számolva. Azt látjuk, hogy a fokszám növelésével eleinte határozottan növekszik az $MSE_{arány}$ (vagyis javul az előrejelzés), de az $n = 7$ érték már nagyon közel esik 100%-hoz, ami a fokszám további növelésével már nemigen változik (az ábrán emiatt nem is látszik, hogy a 100%-ot az $n = 14$ esetén éri el). Talán a 20 feletti fokszámoknál látható szabad szemmel is némi csökkenés, de az sem jelentős.

A következő lépésben megnéztem a többi 32 részvényre is ugyanezt az ábrát, ezeket nem mellékelem. Néhány kivételtől eltekintve (ahol már kisebb fokszám is közel esik a 100%-hoz) nagyon hasonló képet mutatnak, vagyis először a hetedfok közelít a 100%-hoz, és a fokszám növelése ezen nem nagyon változtat.

Szintén nem ábrázolom a Poli(n)_SETAR esetén kirajzolt 33 grafikont sem. Itt már kicsit változatosabb a kép, de hetedfoknál kisebbet a SETAR alapú modellnél sem javasolhatunk.

Mivel jó lenne választanom egy fokszámot, amivel később dolgozom, hogy ne kelljen feleslegesen sokat számolni, a 8. ábra és az $MSE_{arány}$ vizsgálatok fényében végül $n = 14$ mellett döntöttem. Az AR alapú modellnél az esetek zömében valószínűleg elég lett volna $n = 7$, de a SETAR miatt inkább egy nagyobb értéket választok. Hozzá kell tenni, hogy a fentiekből is látszik, hogy a választásnak korlátozott jelentősége van csak, ezért nem érzem nagy kockázatnak egyet kiválasztani, és végig azzal dolgozni tovább. Persze később részletesebben is tesztelhető a kér-

| | Hány részvénynél nyert | | Átlagos érték | |
|----------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli(14)_AR | 28 | 30 | 6,07E-04 | 39,0% |
| Poli(14)_SETAR | 20 | 33 | 6,06E-04 | 38,1% |
| BDF_AR | - | - | 6,49E-04 | 40,3% |

32. táblázat. A Poli(14)_AR és Poli(14)_SETAR modellek összevetése a BDF_AR modellel

Forrás: Saját szerkesztés

| | Hány részvénynél nyert | | Átlagos érték | |
|----------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli(14)_AR | 32 | 18 | 6,07E-04 | 39,0% |
| Poli(14)_SETAR | 30 | 33 | 6,06E-04 | 38,1% |
| BDF_SETAR | - | - | 6,60E-04 | 39,9% |

33. táblázat. A Poli(14)_AR és Poli(14)_SETAR modellek összevetése a BDF_SETAR modellel

Forrás: Saját szerkesztés

dés, akár különböző hibamértékek alapján is. Egyelőre azonban érdekesebb azt megvizsgálni, hogy az így becsült modell hogyan szerepel a benchmarkoz képest.

10.2.2.2. Értékelés

Vessük össze a Poli(14)_AR és Poli(14)_SETAR modelleket a benchmarkkal. A 32. táblázat a BDF_AR modellel történő összevetést mutatja. Látható, hogy a polinom alapú modellek közül mindkettő egyértelműen jobb minden szempont szerint, mint a BDF_AR modell. A Poli(14)_AR átlagosan 6,5%-os javulást hozott MSE alapon, illetve 3,3%-osat MAPE alapon.

A 33. táblázatban a BDF_SETAR modell fényében értékelem a modelleket. A Poli(14)_AR némileg több, mint a részvények felénél nyert, a többi mutatónál pedig még egyértelműbb javulás tapasztalható mindkét polinom verziónál. A Poli(14)_SETAR javulása az átlagos MSE értékben 8,2%, az átlagos MAPE értékben pedig 4,5%.

Tehát a javasolt polinom alapú dekompozícióval a választott egyedi rész modellettől függően átlagosan MSE alapon 6,5% és 8,2% közötti, MAPE alapon pedig 3,3% és 4,5% közötti javulást sikerült elérni a megfelelő benchmark változathoz képest, illetve nem csak átlagban, de a részvények döntő többségénél is javulás

volt tapasztalható (MSE alapon 28 illetve 30 esetben, MAPE alapon pedig 30 illetve 33 részvénynél).

Annyit érdemes még hozzátenni, hogy az MSE alapú mérésben elért nagyobb javulás valószínűleg annak köszönhető, hogy a megfelelő fokszámot MSE alapon választottam ki. Ha egy adott alkalmazáshoz a MAPE érték a fontosabb, érdemes lehet az alapján is megnézni, hogy milyen fokszám a legjobb választás. Ugyanakkor, mivel mindkét hibamérték szerint egyértelmű javulást értem el, ezzel a kérdéssel most nem foglalkozom.

10.2.3. Polinom exponenciális súlyozással

A 10.2.2. pont alapján az U alak polinom illesztéssel történő dekompozíciója jó iránynak tűnik, hiszen a kapott előrejelzés minden szempontból jobbnak bizonyult a benchmarknál. Érdemes lehet ezért finomítani a módszeren, hiszen egy célszerű módosítás esetleg további javulást hozhat.

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.

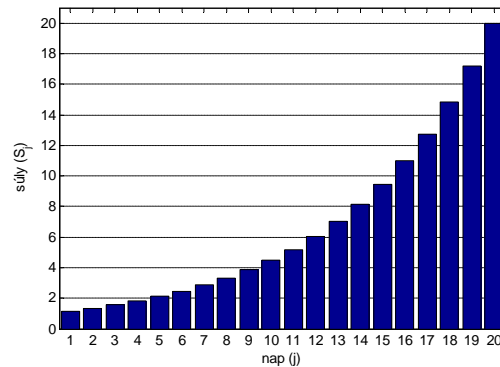
A 10.2.2. pontban bemutatott polinom illesztés ekkor annyiban módosul, hogy a (75) feladatba bekerül egy súly is. Ezen módosítás után a célfüggvény az alábbi formát ölti:

$$\sum_{j=1}^J \sum_{t=1}^T S_j \cdot (p_t - y_{j,t})^2 \rightarrow \min_{\beta_i} \quad (77)$$

ahol

$$S_j = e^{-\frac{j}{J} \cdot \ln(1/J)} \quad (78)$$

Tehát naponta exponenciálisan nő a hibához rendelt súly, és a jelenhez legközelebbi napon lesz a legnagyobb. A súlyfüggvény értékeit a 10. ábrán láthatjuk a becsléseknél használt $J = 20$ napos időszakra ábrázolva.



10. ábra. A (78) súlyfüggvény értékei $J = 20$ nappal számolva
 Forrás: Saját szerkesztés

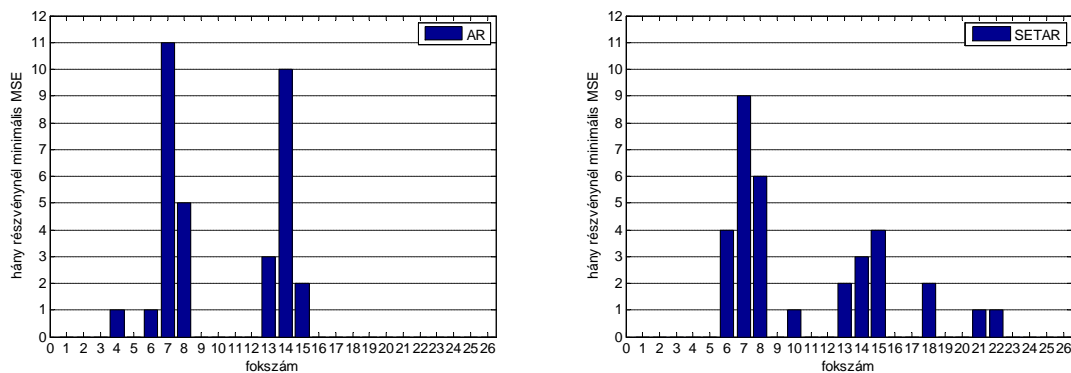
10.2.3.1. A fokszám megválasztása

A korábbiakhoz hasonlóan itt is dönteni kell arról, hogy milyen n fokszám mellett becsüljük a polinomot. Az exponenciális súlyozással illesztett n -edfokú polinom alapú modelleket a korábbiak mintájára attól függően, hogy hogyan becsültem az egyedi részt, Poli_expw(n)_AR és Poli_expw(n)_SETAR modellnek nevezem.

Exponenciális súlyok mellett is megbecsültem $n \in \{1, 2, \dots, 26\}$ minden elemére mindkét modellt, és kigyűjtöttem, hogy az egyes részvényeknél mely fokszámok eredményezték a minimális MSE értéket az előrejelzés értékelésekor. A 11. ábra bal oldali grafikonján a Poli_expw(n)_AR modell látható, amely esetben a legtöbb, összesen 11 részvénynél az $n = 7$ választás adott minimális MSE értéket. Közel ugyanennyi, 10 esetben az $n = 14$ volt minimális hibájú.

Ahogy korábban a 8. ábrán, úgy itt a 11. ábra jobb oldali grafikonján is azt látjuk, hogy habár a Poli_expw(n)_SETAR használatakor kevésbé koncentrált a helyzet, a kiemelkedő értékek hasonlóak, mint az AR verziónál, hiszen $n = 7$ esetén láthatjuk a maximális, 9 értéket, és $n = 8$ után $n = 15$ a következő magas érték, vagyis ugyanaz a két körzet rajzolódik ki az alkalmasnak ígérkező fokszámok közül. Mivel csak egyet szeretnék kiválasztani, ezek alapján leginkább az $n = 7$ fokszám használata javasolható.

Ismét érdemes megnézni a (76) által definiált MSE arányokat is, amely alapján újabb $2 \cdot 33$ darab ábrát kellett elemeznem. Az egyenlően súlyozott polinom esetén a 9. ábrát tipikusnak találtam, ezért csak azt az egyet ábrázoltam. Az ott látott jelenség az exponenciálisan súlyozott verzióra is nagyon hasonlóan jellemző,



11. ábra. Minimális MSE értéket eredményező n fokszámok részvényenként a Poli_expw(n)_AR (bal) és Poli_expw(n)_SETAR (jobb) modellből történő előrejelzés esetén
Forrás: Saját szerkesztés

ezért nem mellékelek újabb ábrát, mert szabad szemmel nehéz lenne észrevenni a különbséget. A következtetés tehát itt is hasonló, nevezetesen, hogy az alacsony fokszámok esetén jelentős javulást hoz az MSE értékben a fokszám növelése, de az $n = 7$ fokszámot elérve a javulás már szerény, és a fokszám 15-20 fölé növelésével némileg romlásba is fordul. Tehát hetedfok alá nem érdemes menni, de a választás ezen túl korlátozott kockázattal jár, mert nem okoz jelentős eltérést. Ezért a 11. ábra és az MSE arányok alapján is az exponenciálisan súlyozott polinom alapú modellek fokszámául jó választás lehet az $n = 7$, a továbbiakban ezért ezzel dolgozom.

10.2.3.2. Értékelés

A következő lépés a kiválasztott Poli_expw(7)_AR és Poli_expw(7)_SETAR modellek értékelése a szokásos módokon. Elsőként vessük össze a Poli(14)_(SET)AR modellekkel, hiszen alapvetően azok továbbfejlesztése volt a célom az exponenciális súlyozással. Ha az egyenlő súlyozású változatnál jobban szerepel, akkor érdemes lesz közvetlenül a benchmarkkal is összevetni.

Mivel az egyedi rész hatását most nem figyelem, azokat a modelleket érdemes összevetni, amelyek azonos módon számolják az egyedi részt, így kizárólag a polinom illesztés eredményességére koncentrálhatunk. Nézzük elsőként a 34. táblázatot, amelyben az AR alapú modelleket vetem össze.

MSE alapon az egyedi részvényeknél a javítások számát figyelve az egyenlő

| | Hány részvénynél nyert | | Átlagos érték | |
|-----------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli_expw(7)_AR | 6 | 0 | 6,07E-04 | 39,5% |
| Poli(14)_AR | 27 | 33 | 6,07E-04 | 39,0% |

34. táblázat. A Poli_expw(7)_AR és a Poli(14)_AR modell összevetése

Forrás: Saját szerkesztés

| | Hány részvénynél nyert | | Átlagos érték | |
|--------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli_expw(7)_SETAR | 7 | 0 | 6,08E-04 | 38,5% |
| Poli(14)_SETAR | 26 | 33 | 6,06E-04 | 38,1% |

35. táblázat. A Poli_expw(7)_SETAR és a Poli(14)_SETAR modell összevetése

Forrás: Saját szerkesztés

súlyozás sokkal jobban szerepel, de meg kell jegyezni, hogy a szokásos pontosság mellett az átlagos értékben nem látunk eltérést a két modell között a kerekítés miatt (a megjelenített tizedesjegyek számának növelésével az exponenciális minimálisan jobbnak mutatkozik). MAPE alapon már látható az exponenciális súlyozás okozta visszaesés az átlagban, amely azzal párosul, hogy egyetlen egyedi részvény esetén sem tudott jobb lenni ez a verzió.

Rátérve a SETAR alapú modellekre, a 35. táblázatban is hasonló kép tárul elénk. Itt már mind MSE, mind MAPE alapon látható az exponenciális súlyozású változat romló tendenciája mind az egyedi részvények, mind az átlagos érték tekintetében, noha az átlagos értékek továbbra sem térnek el igazán jelentősen a két modell esetén.

Ezek alapján az exponenciális súlyozás ötlete úgy tűnik, nem váltotta be a hozzá fűzött reményeket, hiszen inkább rontott, mintsem javított volna. Hozzá kell tenni, hogy hacsak nincs egyértelmű javítás, érdemesebb az egyszerűbb modellt preferálni, ezért továbbra is határozottan a Poli(14)_ (SET)AR modell a legjobb, amit találtunk eddig.

Ezek után már nem is érdemes részletesen bemutatni, hogy a benchmarkot (ami a BDF_(SET)AR volt) az exponenciális változat is egyértelműen megveri.

Ugyanakkor azt még célszerű ellenőrizni, hogy esetleg nem a fokszám rossz megválasztása okozza-e a romlást, ugyanis $n = 14$ helyett exponenciális esetben $n = 7$ értékkel dolgoztunk. Ennek eldöntésére tekintsük a 36. táblázatot, amelyben az egyenlően és az exponenciálisan súlyozott változatokból is a 14-edfokú

| | Hány részvényt nyert | | Átlagos érték | |
|------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli_expw(14)_AR | 7 | 0 | 6,07E-04 | 39,4% |
| Poli(14)_AR | 26 | 33 | 6,07E-04 | 39,0% |

36. táblázat. A Poli_expw(14)_AR és a Poli(14)_AR modell összevetése

Forrás: Saját szerkesztés

| | Hány részvényt nyert | | Átlagos érték | |
|---------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli_expw(14)_SETAR | 7 | 0 | 6,09E-04 | 38,5% |
| Poli(14)_SETAR | 26 | 33 | 6,06E-04 | 38,1% |

37. táblázat. A Poli_expw(14)_SETAR és a Poli(14)_SETAR modell összevetése

Forrás: Saját szerkesztés

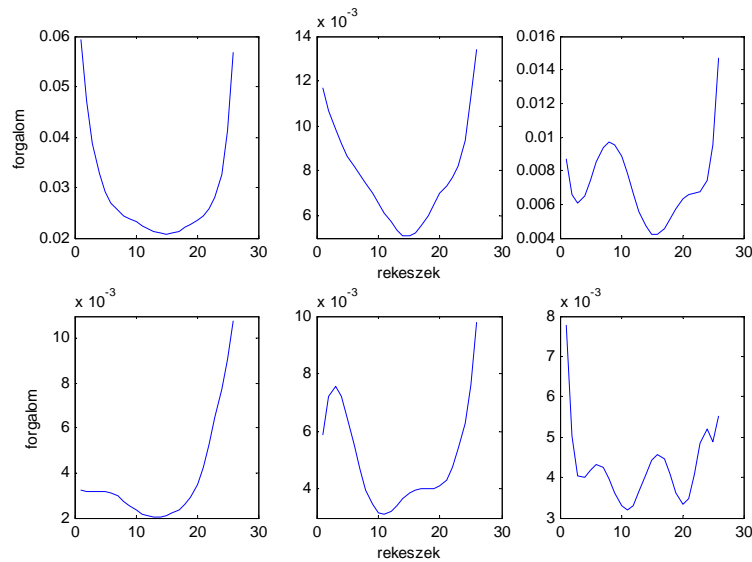
becslést szerepeltetem. Az MSE arány elemzések után nem meglepő, hogy alig látunk eltérést a 34. táblázathoz képest. Mivel a foksám megválasztása $n = 7$ felett nem bír nagy jelentőséggel, ugyanúgy rosszabb ez az exponenciális súlyozású modell is (ez esetben a megjelenített tizedesjegyek számának növelésével MSE alapon már átlagban is alul marad, szemben a hetedfokúval, ami így még minimálisan jobb volt).

Végül rátérve a SETAR alapú modellek összevetésére, a 37. táblázatban láthatjuk, hogy nem sok változás van a 35. táblázathoz képest, ami megfelel a fent elmondottaknak. Összességében a legjobbnak ítélt hetedfokú, és a kevesebb változtatás miatt tisztábban összevethető 14-edfokú, exponenciális súlyokkal illesztett polinom is rosszabbul teljesít, mint a modell egyenletes súlyozású megfelelője.

Nem változtat a következtetésen, de meg kell jegyezni, hogy a fentiek alapján az exponenciális súlyozás látványosabban ronott MAPE, mint MSE alapon. Ennek a fő oka valószínűleg az, hogy a használt foksámot kizárólag MSE alapon választottam ki. Ugyanakkor MSE alapon sem sikerült javítani, ezért ez a torzítás megítélésem szerint nem indokol további vizsgálatokat.

10.2.4. Spline függvény

Habár U alakként szokás rá utalni, a szabályos U betűhöz képest igen változatos alakot ölthet a napon belüli szezonális adott részvény adott időszakában. Erre



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

láthatunk néhány példát a 12. ábrán a Kraft Foods Inc. forgalmának különböző 20 napos időszakaira illesztett 14-edfokú polinomokat szemlélve¹². 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 hullámvázis is. Részben az ebből eredő rugalmasságra való igény is magyarázat lehet arra, hogy viszonylag magasabb fokszám mellett volt érdemes dönteni a polinom illesztésnél (az egyértelmű volt, hogy $n = 7$ alá nem érdemes menni).

Ugyanakkor, habár a rugalmasság előny, a túlzottan jó illeszkedés (még magasabb fokszám) mégsem volt célravezető, hiszen összességében rontott 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. 12. ábra), ugyanakkor a zajt el szeretnénk kerülni. Erre a problémára potenciálisan megoldást kínálhatnak a hozamgörbe becslésben is hasonló motiváció alapján

¹²A Poli(n)_AR modellnél 14 volt a minimális MSE-t adó fokszám erre a részvényre.

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.” (Makara (2000), 33-34. o.)

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 fokszám 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:

$$p_t = \sum_{m=1}^M \beta_m f_m(x_t) \quad (79)$$

Ez hasonlít a (74)-ban 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 De Boor (1978) által javasolt B-spline bázisfüggvények használata (Makara (2000)), melyek értékei rekurzív módon határozhatóak meg a fokszám és a csomópontok ismeretében (1. Melléklet).

A paraméterbecslés az egyenlően súlyozott polinomoknál látott (75) feladathoz hasonlóan történik.

Modellezői döntés kérdése mind a csomópontok K száma és azok elhelyezése, mind pedig az N fokszám megválasztása. Megmutatható (1. Makara (2000)), hogy a szükséges bázisfüggvények, és ezáltal a becslendő paraméterek M száma a következő módon határozható meg:

$$M = N + K - 1 \quad (80)$$

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.

10.2.4.1. A paraméterek megválasztása

A következőkben a fokszámot az $N \in \{2, 3, \dots, 7\}$, míg a csomópontok számát a $K \in \{3, 4, 5, 6\}$ értékekre fogom tesztelni. A fokszám választás mögött az a motiváció, hogy egyrészt alacsonyról induljon (a lineárisnak megfelelő $N = 1$ esettel felesleges külön foglalkozni), másrészt viszont a polinomoknál vízválasztónak nevezhető 7 fölé nem érdemes menni, hiszen éppen a fokszám alacsonyan tartása lenne a cél a zaj kiszűrése érdekében. Hasonlóan, K értékét azért indítom 3-ról, mert ez a legkisebb érték, ami már eltér a normál polinom illesztéstől, illetve túl sok részre valószínűleg nem érdemes osztani a 26 megfigyelést, különben nagyon kevés megfigyelés kerülne 1-1 intervallumba.

Lássuk elsőként a $K = 3$ választást, ami két intervallumnak felel meg. Mivel nincs elvi indok aszimmetriát generálni, ezért minden intervallumba hasonló számú megfigyelést teszek, ami a 26 rekeszre osztott nap esetén 13+13 megfigyelést jelent. Tehát így lesz egy nap eleji, és egy nap végi intervallum. A csomópontok a következők: $t_1 = 1$, $t_2 = 13, 5$, $t_3 = 26$.

$K = 4$ esetén napi 3 intervallumot kapunk: nap eleje, nap közepe és nap vége. A hasonló megfigyelés szám jegyében az egyes intervallumokba 9+8+9 megfigyelés esik. A csomópontok: $t_1 = 1$, $t_2 = 9, 5$, $t_3 = 17, 5$, $t_4 = 26$.

$K = 5$ napi 4 intervallumnak felel meg, melyekbe rendre 7+6+6+7 megfigyelés kerül. A csomópontok: $t_1 = 1$, $t_2 = 7, 5$, $t_3 = 13, 5$, $t_4 = 19, 5$, $t_5 = 26$.

Végül pedig $K = 6$ választással 5 intervallumot kapunk, naponta rendre 6 + 5 + 5 + 5 + 5 megfigyeléssel. A csomópontok ebben az esetben a következők:

| | K | | | |
|---|---|----|----|----|
| N | 3 | 4 | 5 | 6 |
| 2 | 4 | 5 | 6 | 7 |
| 3 | 5 | 6 | 7 | 8 |
| 4 | 6 | 7 | 8 | 9 |
| 5 | 7 | 8 | 9 | 10 |
| 6 | 8 | 9 | 10 | 11 |
| 7 | 9 | 10 | 11 | 12 |

38. táblázat. Becslendő paraméterek száma (M) a különböző spline változatokban
 Forrás: Saját szerkesztés

$t_1 = 1, t_2 = 6, 5, t_3 = 11, 5, t_4 = 16, 5, t_5 = 21, 5, t_6 = 26.$

A különböző modell változatokban becslendő paraméterek számát a 38. táblázat mutatja.

A továbbiakban azt a modellt, melyben az U alakot n -ed fokú, k csomóponttal számolt spline függvény illesztésével dekomponáltam, Spline(NnKk)_AR illetve Spline(NnKk)_SETAR modellnek fogom nevezni attól függően, hogy az egyedi részt milyen módon határoztam meg.

10.2.4.2. Értékelés

A polinomoknál használt eljáráshoz hasonlóan először MSE alapon megkeresem a megbecsült 24-ből a legjobbnak tűnő spline változatot, ebben segít a 39. és 40. táblázat. Ezután csak a legjobbnak ítélt konfigurációt vetem össze részletesen a benchmarkkal.

A 39. táblázat mutatja azokat a modelleket, melyek AR módon számolják az egyedi részt. A táblázat jobb oldali blokkjában azt látjuk, hogy minimális hibát a legtöbb részvényenél (28) egyértelműen a Spline(N4K6)_AR verzió eredményezett, a maradék néhány részvényen pedig 3 különböző paraméterezés osztozik. A táblázat bal oldali blokkjában megfigyelhető, hogy szintén a Spline(N4K6)_AR adta átlagosan is a minimális MSE értéket (a megjelenített pontosság mellett az N6K6 és az N5K4 paraméterezés is ugyanekkora hibát eredményez, minden más pedig ennél nagyobb).

A 40. táblázatban ugyanezt vizsgálhatjuk meg a SETAR alapon számolt egyedi részekkel dolgozó modellekre. A következtetés hasonló. A jobb oldali blokkban a Spline(N4K6)_SETAR verzió az, ami a részvények zöménél (18) a

| | | K | | | | | | K | | | |
|---|--|----------|----------|----------|-----------------|---|---|----------|---|-----------|--|
| N | | 3 | 4 | 5 | 6 | N | 3 | 4 | 5 | 6 | |
| 2 | | 6,32E-04 | 6,24E-04 | 6,20E-04 | 6,14E-04 | 2 | 0 | 0 | 0 | 0 | |
| 3 | | 6,19E-04 | 6,16E-04 | 6,11E-04 | 6,08E-04 | 3 | 0 | 0 | 0 | 0 | |
| 4 | | 6,15E-04 | 6,09E-04 | 6,08E-04 | 6,07E-04 | 4 | 0 | 0 | 0 | 28 | |
| 5 | | 6,08E-04 | 6,07E-04 | 6,08E-04 | 6,09E-04 | 5 | 0 | 3 | 0 | 0 | |
| 6 | | 6,12E-04 | 6,08E-04 | 6,08E-04 | 6,07E-04 | 6 | 0 | 1 | 0 | 1 | |
| 7 | | 6,08E-04 | 6,08E-04 | 6,09E-04 | 6,09E-04 | 7 | 0 | 0 | 0 | 0 | |

39. táblázat. Átlagos MSE értékek a Spline(NnKk)_AR változatokra (bal), illetve annak a száma, hogy a 33-ból hány részvénynél adott minimális MSE értéket az adott konfiguráció (jobb)

Forrás: Saját szerkesztés

| | | K | | | | | | K | | | |
|---|--|----------|----------|----------|-----------------|---|---|----------|----------|-----------|--|
| N | | 3 | 4 | 5 | 6 | N | 3 | 4 | 5 | 6 | |
| 2 | | 6,31E-04 | 6,26E-04 | 6,22E-04 | 6,15E-04 | 2 | 0 | 0 | 0 | 0 | |
| 3 | | 6,21E-04 | 6,19E-04 | 6,11E-04 | 6,09E-04 | 3 | 0 | 0 | 0 | 0 | |
| 4 | | 6,18E-04 | 6,09E-04 | 6,08E-04 | 6,06E-04 | 4 | 0 | 0 | 2 | 18 | |
| 5 | | 6,08E-04 | 6,08E-04 | 6,08E-04 | 6,08E-04 | 5 | 0 | 2 | 0 | 0 | |
| 6 | | 6,12E-04 | 6,09E-04 | 6,09E-04 | 6,06E-04 | 6 | 0 | 3 | 2 | 5 | |
| 7 | | 6,08E-04 | 6,09E-04 | 6,09E-04 | 6,09E-04 | 7 | 0 | 1 | 0 | 0 | |

40. táblázat. Átlagos MSE értékek a Spline(NnKk)_SETAR változatokra (bal), illetve annak a száma, hogy a 33-ból hány részvénynél adott minimális MSE értéket az adott konfiguráció (jobb)

Forrás: Saját szerkesztés

| | Hány részvényt nyert | | Átlagos érték | |
|-----------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Spline(N4K6)_AR | 10 | 1 | 6,07E-04 | 39,1% |
| Poli(14)_AR | 23 | 32 | 6,07E-04 | 39,0% |

41. táblázat. A Spline(N4K6)_AR és a Poli(14)_AR modell összevetése
Forrás: Saját szerkesztés

| | Hány részvényt nyert | | Átlagos érték | |
|--------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Spline(N4K6)_SETAR | 9 | 0 | 6,06E-04 | 38,2% |
| Poli(14)_SETAR | 24 | 33 | 6,06E-04 | 38,1% |

42. táblázat. A Spline(N4K6)_SETAR és a Poli(14)_SETAR modell összevetése
Forrás: Saját szerkesztés

legjobb volt, valamint 6 egyéb konfiguráció tudott még egyedi részvényt (maximum 5-nél) minimális lenni. Átlagosan az N6K6 paraméterezés adta a legkisebb MSE értéket, de a megjelenített pontosság mellett az N4K6 is ugyanekkora hibát jelez, a különbség tehát nem jelentős (0,05% alatti).

Ezek alapján a Spline(N4K6)_ (SET)AR modellt tekintem a legjobbnak a tesztelt változatok közül, ezért a következő lépés ennek a Poli(14)_ (SET)AR modellel történő összevetése, hiszen szándékom szerint ennek a továbbfejlesztése lenne a spline-os változat.

A 41. táblázatban az AR alapú verziókat vetem össze. Az egyedi részvényekre nézve egyértelmű, hogy mind MSE, mind MAPE alapon lényegesen gyakrabban volt a Poli(14)_AR modell a minimális hibájú. Emellett átlagban is mindkét hibamérték szerint a polinom alapú volt a jobb, noha az adott pontosság mellett csak a MAPE értékekben látunk különbséget, és az sem jelentős.

Ugyanez látható a SETAR alapú verziókra a 42. táblázatban. Az egyedi részvényekre sokkal gyakrabban nyert a polinom alapú dekompozíció, és mindkét hibamérték szerint átlagban is jobb volt, noha itt is csak MAPE alapon látható a megjelenített pontosság mellett a (minimális) különbség.

Hogy jobban érzékelhető legyen az eltérés/hasonlóság a két modell között, tekintsük az alábbi hiba arányt:

$$HA = \frac{Hiba_{polinom}}{Hiba_{spline}} \quad (81)$$

| | AR | | SETAR | |
|-------|---------|---------|---------|--------|
| | MSE | MAPE | MSE | MAPE |
| min | 99,78% | 99,39% | 99,39% | 99,51% |
| max | 100,34% | 100,00% | 101,68% | 99,98% |
| átlag | 99,95% | 99,82% | 100,00% | 99,74% |

43. táblázat. A (81) módon számolt hiba arányok a különböző hibamértékek és egyedi rész modellek szerint a Spline(N4K6) és a Poli(14) alapú dekompozíció viszonylatában
Forrás: Saját szerkesztés

ahol a *Hiba* jelölheti az MSE és a MAPE értékeket is. A 43. táblázatba kigyűjtöttem, hogy minden egyedi részvényt figyelembe véve milyen értékek között mozgott a *HA* hiba arány az egyes hiba mértékek esetén. Ebből is látható (mivel 100% körüliek az értékek), hogy lényegében azonos pontosságú a két modell, hiszen a köztük lévő eltérés minden részvényre csekély korlátok között marad.

Ezek alapján tehát úgy tűnik, nem érdemes a spline-os változatra áttérni, hiszen az (ha minimális mértékben is, de) egyértelmű romlást mutat, és ráadásul bonyolultabb is az egyszerű polinom illesztésnél, ami *ceteris paribus* szintén hátránynak tekinthető.

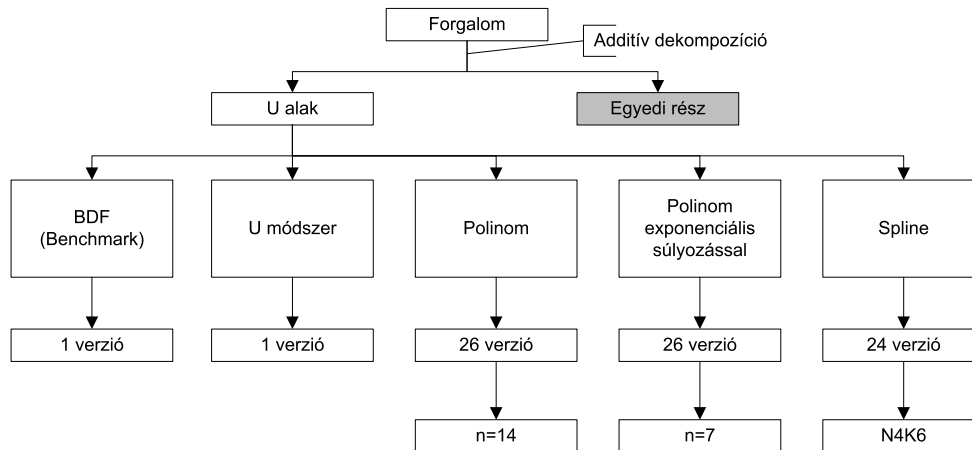
Mivel egyrészt végül elvetem, másrészt pedig lényegében azonos teljesítményt mutat a Poli(14)_(SET)AR modellel, ezért tételesen nem mutatom be, hogy az effektív benchmarknak tekintett BDF_(SET)AR modellt a Spline(N4K6)_(SET)AR is egyértelműen megveri.

10.2.5. Áttekintés

A 13. ábra segítségével tekintsük át a 10.2. pont eredményeit. A forgalmat additív módon dekomponálva két feladatunk van: modellezni az *U* alakot, és modellezni a maradékot, vagyis az egyedi részt. A 10.2. pont az előbbivel foglalkozik.

A benchmarknak tekintett BDF modell magas dimenziójú faktorelemzéssel dekomponálja az *U* alakot. A BDF modell által használt egyedi rész modellezését egyelőre változatlanul hagyva csak a dekompozíció hatását vizsgáltam ebben a pontban.

Az első megközelítés az *U*-módszer használata volt, vagyis ezzel helyettesítettem a magas dimenziójú faktorelemzést. Meglepetésemre egyértelműen javult így az előrejelzés akár MSE, akár MAPE alapon mértem a hibát. A részvények



13. ábra. Az U alak dekomponálásának különböző módjai

Forrás: Saját szerkesztés

többségénél, és átlagosan is csökkent az előrejelzés hibája.

A második megközelítés az U-módszer által adott görbe simítása volt, melyet polinom illesztéssel értem el. A vizsgálatok alapján 26 különböző modell közül a fokszámot $n = 14$ értéknek választva a modell (ugyanazon egyedi rész modellekkel kiegészítve) egyértelmű javítást eredményezett nem csak a benchmark BDF modellhez képest, de a fent említett, egyedi résszel kibővített U-módszerhez képest is.

A harmadik megközelítésben a polinom illesztést finomítottam oly módon, hogy a közelebbi napokhoz jobban illeszkedjen, mint a távolabbiakhoz. Ennek elérésére exponenciális súlyozást használtam a célfüggvényben. Az egyenlő súlyozású polinomnál használt eljáráshoz hasonlóan megállapítottam, hogy a 26 különböző modell közül az $n = 7$ fokszám mellett becsült változat használata javasolt. Ez a megoldás (szintén kiegészítve ugyanazon egyedi rész modellekkel) továbbra is jobb a BDF benchmarknál, de az egyenlő súlyozású polinom illesztésnél némileg gyengébb.

Negyedik megközelítésben spline függvények illesztésével kísérleteztem, szintén a polinom illesztés finomításaként. A fentiekhez hasonló módszerrel arra jutottam, hogy a 24 vizsgált változat közül a negyedfokú, 6 csomóponttal tagolt (N4K6) modellt érdemes használni. Ez a megoldás, a szokásos egyedi rész modellekkel kiegészítve szintén egyértelműen jobb a BDF benchmarknál, és némileg rosszabb az egyenlő súlyozású polinom illesztésnél.

Összegzésként tehát azt mondhatjuk, hogy az általam javasolt módszerek közül mind a négy jobban teljesít, mint a BDF modell által használt magas dimenziójú faktorelemzés, és ezáltal megverik a szakirodalomban legjobbnak tekintett modellt (1. III. rész). A négy közül a legjobb az egyenlő súlyozással illesztett polinom. A fokszámot $n = 14$ értéknek választottam, de mint láttuk, $n \geq 7$ választás mellett nem változik lényegesen az eredmény. Ellenben vizsgálataim alapján az $n < 7$ választás nem célszerű.

Végül érdemes még azt is megjegyezni, hogy ezen modellek mindegyikének előnye a benchmarkhoz képest, hogy annál lényegesen kevesebb adatból dolgozik, ugyanis kizárólag az előrejelezni kívánt részvény forgalom adatait használja fel, míg a benchmark a teljes piac, esetemben 33 részvény forgalom adatát igényli. A teljes piaccal való számolás szükségességének két hátrányát is látom.

Az egyik, hogy az eredmény értelemszerűen nagyban függ az elemzésbe bevont (piacnak tekintett) részvényektől. Attól függően, hogy mit tekintek piacnak, a modell eredménye számottevően változhat. Márpedig a kérdés közel sem egyértelmű. Az USA piacán dolgozva ezúttal a DJIA index részvényeit használtam, de ugyanilyen jogos lehetne bármelyik másik index, vagy az N leglikvidebb részvény, vagy számtalan más szűrés is. Ebből kifolyólag a modell tartalmaz egy ilyen irányú bizonytalanságot.

A másik, hogy bizonyos piacokon, például a hazain sokkal kevesebb olyan részvény van, ami elég likvid ahhoz, hogy bekerülhessen az elemzésbe, a részvények alacsony száma pedig erősen torzíthatja a végeredményt (valószínűleg a nagyobb hiba irányába), hiszen kevesebb információból nehezebb meghatározni a piac egésze által indokolt U alakot. Ezáltal pedig egy kisebb piacon elképzelhető, hogy drasztikusan romlik az előrejelzés minősége akár még a kifejezetten likvid részvényekre is, ahol egyébként kizárólag az adott részvény adataiból dolgozva ez nem lenne indokolt.

10.3. Az egyedi rész előrejelzésének különböző módjai

A 10.2. pontban az U alak (additív) leválasztásának különböző módjait vizsgáltam, ezért az egyedi rész (az U alak leválasztása utáni maradék) előrejelzésére

végig ugyanazt a két modellt használtam, amit a BDF modellből megismertünk (AR és SETAR). Arra a megállapításra jutottam, hogy a vizsgált lehetőségek közül az $n = 14$ paraméter mellett illesztett polinom (Poli(14)) adta a legjobb eredményt.

A 10.3. pontban az egyedi rész előrejelzési lehetőségeivel foglalkozom (l. 13. ábra). Annak érdekében, hogy kizárólag az egyedi rész modellje okozta különbségeket tudjam értékelni, az U alak leválasztását azonos módon fogom végezni minden vizsgált esetben. Mivel a Poli(14) modellt találtam a legjobbnak az U alak leválasztására, ezért ebben a pontban végig ezt fogom használni.

10.3.1. Késleltetett ármozgás használata

A 10.1. pontban az U dekompozíció nélküli modellek vizsgálatakor arra jutottam, hogy a hét napja hatás nem tűnik hasznosnak az előrejelzésben. Az ármozgás mutatók modellbe foglalása az ott látott módon szintén nem volt célravezető, azonban egyrészt ez már kevésbé volt határozott (esetenként láttunk némi javulást az átlagban), másrészt pedig, mint említettem, ezek modellbe foglalása lényegileg más módokon is elképzelhető.

Ezért ebben a pontban külön az egyedi rész előrejelzéséhez használva is tesztelem az ármozgás mutatókat, még hozzá ARX modell keretében. Mivel ez a Poli(14)_AR modell kiterjesztésének tekinthető, ezért azt fogom vizsgálni, hogy ezen modell előrejelzésén sikerül-e javítani.

Korábban a 10.1.3. pontban a következő ármozgás mutatókat definiáltam: loghozam, volatilitás, rés, normál és százalékos ársáv, valamint normál és százalékos tényleges ársáv.

10.3.1.1. Egyszerű késleltetett

A modellbe foglalásnak egy kézenfekvő lehetősége az, hogy a választott ármozgás mutató késleltetett értékét használom magyarázó változóként. Az egyedi rész modellje ebben az esetben az alábbi lesz:

$$\tilde{e}_{t,i} = c_i + \beta_1 e_{t-1,i} + \beta_2 X_{t-1,i} + \varepsilon_{t,i} \quad (82)$$

ahol X az ármozgás mutatók egyike, e pedig az egyedi rész. Nevezzük az így kapott változatokat Poli(14)_ARX modelleknek.

| Poli(14)_ARX változatok | Hány részvénynél nyert | | Átlagos érték | |
|----------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 1 | 1 | 6,09E-04 | 39,1% |
| Volatilitás | 10 | 0 | 6,06E-04 | 39,3% |
| Rés | 1 | 3 | 6,10E-04 | 39,1% |
| Ársáv | 7 | 9 | 6,03E-04 | 39,1% |
| Ársáv (%) | 7 | 9 | 6,03E-04 | 39,1% |
| Tényleges ársáv | 7 | 3 | 6,03E-04 | 39,2% |
| Tényleges ársáv (%) | 7 | 3 | 6,04E-04 | 39,2% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

44. táblázat. A Poli(14)_ARX modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

A 44. táblázat segítségével hasonlítsuk össze az így kapott modellek előrejelzéseit a Poli(14)_AR modell előrejelzésével. A loghozam és a rés használata minden szempont szerint romlást okozott. Az MSE kritérium alapján azt mondhatjuk, hogy a többi ármozgás mutató használata az összes részvény átlagában némi javulást eredményezett, azonban egyedileg mindegyik egyértelműen többször okozott romlást, mint javulást. Ugyanakkor érdemes megjegyezni, hogy a volatilitás esetén átlagban is minimális a javulás, az ársáv alapú mutatóknál pedig kizárólag egy részvénynek (Altria Group Inc.) köszönhető minden esetben, ezt kivéve az elemzésből már átlagban is romlást tapasztalnánk mind a négy ársáv alapú mutató használatakor. A MAPE kritérium szerint még rosszabb a helyzet, ugyanis ott átlagban is némi romlást figyelhetünk meg minden esetben, bármelyik ármozgás mutatót használjuk is.

Összességében tehát kizárólag MSE átlag alapon sikerült némi javulást elérni bizonyos ármozgás mutatóknál, de az is jobbára egyetlen részvénynek tudható be, tehát semmiképp nem általános eredmény. Ezek alapján az ármozgás mutatók késleltetett értéke nem tűnik hasznos segítségnek a forgalom előrejelzésben. Lehetne még esetleg más késleltetéssel is kísérletezni, de ezt az irányt a fentiek alapján nem tartok ígéretesnek.

| | Szignifikancia szint | | |
|---------------------|----------------------|-------|-------|
| | 0,01 | 0,05 | 0,10 |
| Loghozam | 9,3% | 18,2% | 25,7% |
| Volatilitás | 34,3% | 49,5% | 58,4% |
| Rés | 11,6% | 20,7% | 27,6% |
| Ársáv | 19,6% | 33,3% | 42,1% |
| Ársáv (%) | 19,7% | 33,4% | 42,1% |
| Tényleges ársáv | 26,8% | 40,4% | 49,0% |
| Tényleges ársáv (%) | 26,4% | 40,2% | 48,8% |

45. táblázat. Szignifikáns Granger-oktságot mutató 20 napos időszakok számának az aránya aggregáltan az összes részvényre nézve
Nullhipotézis: Az ármozgás nem Granger-okozza a forgalmat

Forrás: Saját szerkesztés

10.3.1.2. Feltételes modell: Granger-oktság

Korábban azt találtam (l. 28. táblázat), hogy a teljes mintát tekintve a loghozam kivételével minden más ármozgás mutató minden részvélynél és minden szokásos szignifikancia szinten Granger-okozza a forgalmat, és 5%-os szignifikancia szinten a loghozamnál is csak egyetlen részvényre nem sikerült ugyanezt kimutatni, ezért érdemesnek tűnt az ármozgás mutatók késleltetettjeivel dolgozni. Ugyanakkor az eddig tesztelt módokon nem sikerült valódi javítást elérni az ármozgás mutatók késleltetettjének modellbe foglalásával.

A fentiek alapján felmerül, hogy nem csak a teljes mintára kellene tesztelni a Granger-oktságot, hanem mindig az éppen használni kívánt 20 napos almintára (becslési időszakra) is. Mivel részvényenként hasonló képet mutat az eredmény, a 45. táblázatban az összes részvényre nézve aggregáltan mutatom be az eredményeket. Azt látjuk, hogy egy kivételtől eltekintve (volatilitás 10%-os szignifikancia szint mellett) az esetek többségében Granger értelemben nem előzi meg szignifikánsan az ármozgás a forgalmat. A korábban is már gyengébbnek talált loghozamtól és réstől eltekintve a többi ármozgás mutatónál 10%-os szignifikancia szint mellett 42-58%, míg 5%-on 33-49%, 1%-os szignifikancia szint mellett pedig 19-34% körül találjuk a szignifikáns esetek arányát. Ezt látva már kevésbé meglepő, hogy az ármozgást minden 20 napos időszakra a modellben hagyva nem sikerült érdemi javítást elérni.

Érdekes lehet ezek után tesztelni egy olyan modellt is, ahol figyelembe vesszük, hogy van-e szignifikáns Granger-oktság az adott becslési időszakban az adott rész-

vényre nézve. Ekkor az eredménytől függően más modellt használnánk: amennyiben kimutatható szignifikáns Granger-okság az adott részvény becslési időszakában, bekerül a modellbe az ármozgás mutató, ellenkező esetben pedig kimarad belőle.

Legyen az egyedi rész modellje a következő:

$$\begin{aligned} \tilde{e}_{t,i} = & (c_{1,i} + \beta_1 e_{t-1,i} + \beta_2 X_{t-1,i}) I(p_{gr,i}) + \\ & + (c_{2,i} + \beta_3 e_{t-1,i}) (1 - I(p_{gr,i})) + \varepsilon_{t,i} \end{aligned} \quad (83)$$

ahol

$$I(p_{gr,i}) = \begin{cases} 1 & \text{ha } p_{gr,i} < \textit{küszöb} \\ 0 & \text{egyébként} \end{cases} \quad (84)$$

X a korábbiakhoz hasonlóan az egyes ármozgás mutatókat jelöli, e pedig az egyedi részt. Az adott időszakra és részvényre vonatkozó Granger-okság p -értéke p_{gr} , a *küszöb* értéke pedig a választott szignifikancia szintnek felel meg. Tehát, ha a választott szignifikancia szinten kimutatható Granger-okság az ármozgás mutató és a forgalom között (az ármozgás Granger értelemben megelőzi a forgalmat), akkor használjuk az ármozgás mutató késleltetettjét, egyébként viszont kihagyjuk a modelltől, és maradunk az egyszerű AR(1) modellnél. Nevezzük az így kapott modelleket Poli(14)_ARX_Gr modelleknek.

A *küszöb* értékét 10%-nak választva, az egyes ármozgás mutatókra kapott eredményeket találjuk a 46. táblázatban. Látható, hogy az egyedi részvényeket tekintve minden esetben többször értünk el romlást, mint javulást MSE és MAPE alapon egyaránt (a legjobb eredmény 8 részvénynél elért javulás volt a 33 részvényből). Az átlagot tekintve is romlás tapasztalható minden ármozgás mutatóra MSE és MAPE alapon is. Az adott megjelenítési pontosság mellett a volatilitást tartalmazó modell átlagos MSE értéke, illetve a rést tartalmazó modell átlagos MAPE értéke azonosnak tűnik a Poli(14)_AR benchmarkéval, de valójában némileg ezek is rosszabbak.

A *küszöb* csökkentésével elérhető, hogy csak a fentnél erősebbnek tekintett Granger okság esetén használjunk ármozgás mutatót. A mellékletben a 99 - 100. táblázatokban találhatóak a *küszöb* = {0,01; 0,05} választásnak megfe-

| Poli(14)_ARX_Gr küszöb: 0,1 | Hány részvényt nyert | | Átlagos érték | |
|--------------------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 2 | 0 | 6,09E-04 | 39,1% |
| Volatilitás | 7 | 0 | 6,07E-04 | 39,2% |
| Rés | 2 | 8 | 6,10E-04 | 39,0% |
| Ársáv | 5 | 3 | 6,08E-04 | 39,1% |
| Ársáv (%) | 6 | 4 | 6,08E-04 | 39,1% |
| Tényleges ársáv | 6 | 0 | 6,08E-04 | 39,2% |
| Tényleges ársáv (%) | 7 | 0 | 6,08E-04 | 39,2% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

46. táblázat. A Poli(14)_ARX_Gr modellek értékelése különböző ármozgás mutatók használata esetén (küszöb: 0,1)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

elő eredmények. Megállapítható, hogy az eredmények szinte alig változnak a küszöb csökkentésével. Átlagban továbbra sincs javulás MSE és MAPE alapon sem. A 10%-os szignifikancia szinthez képest néhány további egyedi részvényt is sikerült javulást elérni, de a legmagasabb érték még mindig csak 11, ami az összes részvénynek csupán egyharmada. Ezek alapján elmondható, hogy az ebben a pontban bemutatott Granger-okság alapú feltételes modell sem volt elég jó ahhoz, hogy alátámassza az ármozgás mutatók késleltetettjének használatát.

10.3.1.3. Feltételes modell: Korreláció

A 10.3.1.2. pontban bemutatott modellben a Granger-okság helyett esetleg használható a forgalom és a késleltetett ármozgás közötti korreláció is, hiszen ha ez viszonylag magas, bízhatunk abban, hogy ez a Granger-okságnál esetleg jobb választóvonalnak bizonyul, és érdemes a késleltetett ármozgást feltételesen megtartani magyarázó változóként egy korreláción alapuló kritérium szerint.

A 47. táblázatban a késleltetett ármozgások és a forgalom közötti, 20 napos almintákra számolt korreláció abszolút értékének (ritkán ugyan, de előfordul negatív korreláció is) statisztikáit látjuk az egyes ármozgás mutatók szerint, ezúttal is aggregáltan az összes részvényre nézve. A loghozam és a rés (késleltetettje) továbbra is egyértelműen kevésbé mutat szoros kapcsolatot a forgalommal, mint a többi ármozgás mutató, melyeknél az átlagos korreláció 24-29% körül alakul, a maximális pedig eléri a 79-83%-os értéket is. Ez utóbbi biztató jel lehet arra nézve, hogy

| | Min | Átlag | Max |
|---------------------|------|-------|-------|
| Loghozam | 0,0% | 7,3% | 61,5% |
| Volatilitás | 0,0% | 24,2% | 79,1% |
| Rés | 0,0% | 7,0% | 72,1% |
| Ársáv | 0,0% | 29,2% | 82,5% |
| Ársáv (%) | 0,0% | 29,4% | 83,6% |
| Tényleges ársáv | 0,0% | 29,0% | 82,1% |
| Tényleges ársáv (%) | 0,0% | 29,2% | 83,1% |

47. táblázat. A késleltetett ármozgás és a forgalom közötti 20 napos korrelációk abszolút értékének statisztikái az összes részvény tekintetében

Forrás: Saját szerkesztés

egyek esetekben érdemes lehet ez alapján bevonni a modellbe az ármozgás mutatók késleltetettjét. A minimális értékek ugyanakkor minden ármozgás mutatónál nulla körül alakulnak, amely alátámasztja azt a korábbi megállapítást, miszerint nem érdemes minden esetben felhasználni az ármozgás mutató késleltetettjét.

Módosítsuk az előző pontban látott modellt ezek alapján. Az egyedi részmodellje legyen tehát a következő:

$$\begin{aligned} \tilde{e}_{t,i} = & (c_{1,i} + \beta_1 e_{t-1,i} + \beta_2 X_{t-1,i}) I(\rho_i) + \\ & + (c_{2,i} + \beta_3 e_{t-1,i}) (1 - I(\rho_i)) + \varepsilon_{t,i} \end{aligned} \quad (85)$$

ahol

$$I(\rho_i) = \begin{cases} 1 & \text{ha } |\rho_i| > \textit{küszöb} \\ 0 & \text{egyébként} \end{cases} \quad (86)$$

Az adott 20 napos időszakra, adott részvényre vonatkozó, késleltetett ármozgás és forgalom közötti korreláció értékét ρ_i jelöli. A *küszöb* ezúttal tetszőleges érték lehet a $[0, 1]$ intervallumon, de 0 vagy 1 értéket nem érdemes választani, mert ilyenkor visszkapunk egy olyan modellt, amelyet korábban már teszteltünk. Minden más jelölés megegyezik a (83) - (84) egyenletben látottakkal.

Tehát ha adott *küszöb* felett van a korreláció, felhasználjuk az ármozgás mutató késleltetettjét, egyébként viszont kihagyjuk a modelltől. Az így kapott változatot nevezzük Poli(14)_ARX_korr modellnek.

| Poli(14)_ARX_korr küszöb: 0,3 | Hány részvényt nyert | | Átlagos érték | |
|----------------------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 3 | 2 | 6,07E-04 | 39,0% |
| Volatilitás | 8 | 4 | 6,08E-04 | 39,1% |
| Rés | 10 | 11 | 6,07E-04 | 39,0% |
| Ársáv | 10 | 15 | 6,09E-04 | 39,0% |
| Ársáv (%) | 9 | 16 | 6,09E-04 | 39,0% |
| Tényleges ársáv | 8 | 8 | 6,08E-04 | 39,1% |
| Tényleges ársáv (%) | 8 | 8 | 6,09E-04 | 39,1% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

48. táblázat. A Poli(14)_ARX_korr modellek értékelése különböző ármozgás mutatók használata esetén (küszöb: 0,3)
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

Mivel a 47. táblázat szerint az átlagos korreláció több ármozgás mutatóra is közel 30%-os, elsőként nézzük a $küszöb = 0,3$ esetet, tehát az átlagosnál erősebb korreláció esetén használjuk az ARX modellt. Az így kapott eredményeket mutatja a 48. táblázat, melyben az látható, hogy átlagban sem MSE, sem pedig MAPE alapon nem sikerült javítani egyik ármozgás mutató esetén sem. A megadott pontosság mellett több ármozgás mutató használata is egyenlő értéket adott a Poli(14)_AR modellel, de valójában némileg minden esetben romlott az eredmény a benchmarkhoz képest. Az egyedi részvények tekintetében sem látunk javulást. Az ársáv és a százalékos ársáv MAPE alapon 15-ös, illetve 16-os értékkel az egyedi részvények közel felénél adott jobb eredményt, minden más esetben ennél gyengébb teljesítmény figyelhető meg.

Ezek alapján tehát a $küszöb = 0,3$ mellett nem sikerült javítani a Poli(14)_AR modellen. Természetesen nem biztos, hogy éppen ez a legjobb választás, ezért megvizsgáltam azt is, hogy a $küszöb = \{0,1; 0,2; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9\}$ értékek mellett mit mondhatunk a modellről. Az ilyen paraméterek melletti eredményeket a mellékletben a 101 - 108. táblázatokban találhatjuk. Úgy tűnik ezek alapján, hogy az eredmények érdemben nem változnak, bármilyen $küszöb$ értéket használunk is. Alacsony $küszöb$ esetén megtaláljuk a tisztán ARX modellnél látott minimális javulást bizonyos ármozgás mutatókkal számolt átlagos MSE értékekben, de ezek továbbra is egyetlen részvénynek köszönhetőek. A $küszöb$ emelésével fokozatosan elmozdulunk a tisztán ARX modelltől a tisztán AR(1) modell felé.

A *küszöb* = {0,5; 0,6; 0,7} értékeire a rés használata minimális átlagos MAPE javulást eredményez, ezt külön jelöltem a táblázatokban, mert a megadott pontosság mellett nem is látszana. Minden egyéb esetben romlik az átlagos hibamérték értéke a Poli(14)_AR modellhez képest. Az egyedi részvények tekintetében a benchmarknál jobb esetek száma eleinte nő (de nem nő az esetek fele fölé), utána pedig csökken. Ez utóbbinak az az oka, hogy magasabb *küszöb* értékeknél már gyakorlatilag megegyezik a modell a sima AR(1) modellel. Azért csökken tehát a nyerő esetek száma, mert egyre több részvénynél lesz teljesen megegyező az előrejelzés a Poli(14)_AR modellével, amelyet viszont nem tekinthetünk javulásnak. A 47. táblázatban látott maximális értékekkel összhangban ez loghozam esetén 0,7-es értéktől, volatilitás és rés esetén 0,8-as értéktől, az ármozgás alapú mutatók esetén pedig a *küszöb* 0,9-es értékétől minden részvénynél így alakul, ezért ezen határokon túl már minden részvényre végig az AR(1) modellt alkalmazzuk, és így az átlagos hibák is megegyeznek a benchmarkéval.

Összességében tehát a korreláció alapú feltételes modellel sem sikerült javítani a sima AR(1) módon modellezett egyedi rész előrejelzését.

Az eddigiek alapján tehát az a megállapítás rajzolódik ki, hogy az ármozgás mutatók késleltetettjét nem tudjuk érdemben felhasználni a forgalom előrejelzésben. Természetesen lennének még egyéb lehetséges felhasználási módok is, de az eddigiek alapján ezt az irányt nem tartom érdemesnek folytatni.

10.3.2. Egyidejű ármozgás használata (gondolat kísérlet)

Ugyanakkor felmerülhet, hogy késleltetett helyett esetleg az egyidejű ármozgás mutatót használjuk fel magyarázó változóként. Ennek az eredményességében két feltétel teljesülése esetén bízhatunk.

1. Az egyik természetesen az, hogy az egyidejű ármozgás jól magyarázza a forgalmat, és ezért annak a modellbe foglalása javít a forgalom előrejelzésén is. Ez azonban nem elég önmagában, hiszen az egyidejű ármozgást csak akkor tudjuk megfigyelni, amikor az előrejelezni kívánt forgalmat is, tehát az előrejelzés pillanatában ez az információ nem áll rendelkezésre.
2. Ezért a másik feltétel, hogy jól előre tudjuk jelezni magát az ármozgás mutatót. Az egyidejű értéket az előrejelzett értékkel helyettesítve már elvégezhető a becslés.

Ebben az alponban az első feltételt vizsgálom meg, ha ugyanis az nem teljesül, akkor felesleges az ármozgás előrejelzésével bajlódni. Ezt úgy fogom megtenni, hogy az előrejelzés pillanatában ismertnek kezelem a következő időszak ármozgás mutatóját. Ez természetesen nem reális, az ismeretlen jövőbe nem tudok így előrejelezni. Ezért ezt nevezhetjük gondolat kísérletnek is, aminek az eredményét nem tekinthetjük előrejelző modellnek, hiszen csak utólag megvalósítható. Ugyanakkor mégis hasznos lehet, mert ha azt kapjuk, hogy egyértelműen javul így az előrejelzés a Poli(14)_AR modellhez képest, ez azt jelenti, hogy a következő lépésben érdemes előrejelzett ármozgás mutatóval próbálkozni, illetve különböző (ármozgás) előrejelző módszereket is tesztelni. Ellenkező esetben, ha tehát azt találnánk, hogy így sem javul az előrejelzés, akkor semmi esetre sem várhatjuk, hogy előrejelzett ármozgás mutatóval jobb eredményre juthatunk, mint a Poli(14)_AR modellel.

Az egyedi rész modellje tehát legyen az alábbi:

$$\tilde{e}_{t,i} = c_i + \beta_1 e_{t-1,i} + \beta_2 X_{t,i} + \varepsilon_{t,i} \quad (87)$$

ahol X az ármozgás mutatók egyike, e pedig az egyedi rész. Nevezzük az így kapott változatokat Poli(14)_ARX(!) modelleknek, azaz felkiáltójellel jelezzük, hogy ez valójában nem egy megvalósítható specifikáció.

Tekintsük át az így kapott eredményeket a 49. táblázat segítségével. Az összehasonlítás alapja továbbra is a Poli(14)_AR modell. A korábbiakhoz hasonlóan minden szempont szerint romlást okozott a loghozam és a rés használata is.

Ugyanakkor MSE alapon 32, MAPE alapon pedig az összes egyedi részvénynél javulást értünk el az összes többi ármozgás mutatóval, illetve az átlagos értéket nézve is határozottan javítottunk mindegyik esetben. Az eredmények alapján MSE alapon a legnagyobb (13,5%-os) javulást átlagban a százalékos tényleges ársáv használata eredményezte, míg MAPE alapon a legnagyobb (10,3%-os) átlagos javulást a százalékos ársáv hozta, noha ez utóbbi a megadott pontosság mellett nem tér el a normál ársáv okozta javítástól.

Ezek alapján úgy tűnik, egyértelmű javulást értünk el az egyidejű ármozgás mutatók modellbe foglalásával (a rés és a loghozam kivételével). Mivel azonban ez nem megvalósítható módszer, a következő lépésben érdemes az ármozgás mutatókat valamilyen módon előrejelezni, és az így kapott előrejelzést használni magyarázó változóként. Vélhetően így romlani fog a modell teljesítménye az egy-

| Poli(14)_ARX(!) változatok | Hány részvénynél nyert | | Átlagos érték | |
|-------------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 1 | 0 | 6,16E-04 | 39,5% |
| Volatilitás | 32 | 33 | 5,59E-04 | 36,7% |
| Rés | 0 | 4 | 6,23E-04 | 39,1% |
| Ársáv | 32 | 33 | 5,35E-04 | 35,0% |
| Ársáv (%) | 32 | 33 | 5,29E-04 | 35,0% |
| Tényleges ársáv | 32 | 33 | 5,34E-04 | 35,2% |
| Tényleges ársáv (%) | 32 | 33 | 5,25E-04 | 35,1% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

49. táblázat. A Poli(14)_ARX(!) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

idejű ármozgás mutató használatához képest, de a 10-13%-os átlagos javulásból még így is elképzelhető, hogy meg tudunk tartani valamennyit.

10.3.3. Előrejelzett ármozgás használata

Az előző pont gondolatmenetét folytatva tehát ebben a pontban az ármozgás mutatók különböző előrejelzéseit fogom vizsgálni. A modell specifikációja csak annyiban változik a gondolat kísérlethez képest, hogy tény helyett előrejelzett ármozgás mutatót használok. Ily módon már egy valóban megvalósítható előrejelző modellt kapunk.

Az U alakot tehát továbbra is a Poli(14) modellel dekomponálom, az egyedi rész modellje pedig az alábbi lesz:

$$\tilde{\varepsilon}_{t,i} = c_i + \beta_{1,i}e_{t-1,i} + \beta_{2,i}\tilde{x}_{t,i} + \varepsilon_{t,i} \quad (88)$$

ahol \tilde{x} az előrejelzett ármozgás mutató. Az alább tárgyalt különböző modell változatok tehát pusztán abban térnek el egymástól, hogy az \tilde{x} előrejelzést hogyan állítom elő.

Amint a gondolat kísérletben láttuk, a loghozam és a rés használata akkor sem javítana a forgalom előrejelzésén, ha tökéletesen előre látnánk az alakulásukat (l. 49. táblázat). Emiatt a továbbiakban ezek előrejelzésének a modellbe foglalásától eltekintek, és csak a többi öt ármozgás mutatót vizsgálom.

| Poli(14)_f(AR) változatok | Hány részvényt nyert | | Átlagos érték | |
|------------------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Volatilitás | 2 | 3 | 6,15E-04 | 39,1% |
| Ársáv | 4 | 25 | 6,15E-04 | 38,9% |
| Ársáv (%) | 4 | 24 | 6,18E-04 | 38,9% |
| Tényleges ársáv | 3 | 14 | 6,14E-04 | 39,0% |
| Tényleges ársáv (%) | 7 | 20 | 6,15E-04 | 39,0% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

50. táblázat. A Poli(14)_f(AR) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

10.3.3.1. Egyszerű AR modell

Első megközelítésben használjunk egy egyszerű AR(1) modellt az ármozgás előrejelzésére:

$$\tilde{x}_{t,i} = c_i + \beta_{1,i}x_{t-1,i} + \varepsilon_{t,i} \quad (89)$$

ahol x a (88)-hez hasonlóan a választott ármozgás mutatót jelöli. Nevezzük Poli(14)_f(AR) modellnek az így előrejelzett ármozgás mutatót használó forgalom előrejelző modellt.

Az 50. táblázat segítségével hasonlítsuk össze az így becsült modelleket a javítani szándékolt Poli(14)_AR modellel¹³. Láthatóan MSE alapon mindegyik ármozgás mutató esetén a részvények töredékénél tapasztalunk csak javulást, és ezzel összhangban az átlagos értékek is romlanak minden esetben. MAPE alapon a normál és a százalékos ársáv, valamint a százalékos tényleges ársáv is a részvények több, mint felénél hoz javulást, ugyanakkor az átlagból is láthatjuk, hogy ez a javulás nem jelentős.

10.3.3.2. AR modell dummy-val

Az áralakulás, és ezen keresztül az ármozgás mutatók esetén is kiemelt szerepe van a nap első megfigyelésének. Ennek oka, hogy az általam vizsgált tőzsdék délutánonként bezárnak, és csak másnap reggel nyitnak ki. Ezáltal a nap utolsó

¹³Azt, hogy az ármozgás előrejelzés önmagában mennyire jó, külön nem vizsgálom, csak a forgalom előrejelzéshez való hozzájárulásán keresztül, tehát a szokásos módon.

megfigyelése és a következő nap első megfigyelése között lényegesen több idő telik el, mint egyéb szomszédos megfigyelések között. Ebből következően potenciálisan több olyan hír érkezik ebben az időszakban, amelynek a tőzsde nyitáskor be kell épülnie az árba. Ez még inkább igaz, ha a zárás és a következő nyitás között hétvége, vagy egyéb szünnap is van. Részben ez a jelenség indokolja a már említett rés mutató létezését is, melyet elsősorban napváltáskor szokás figyelni.

A fenti észrevétel modellbe foglalható egy alkalmasan megválasztott dummy változó használatával. Egészítsük ki az előző pontban használt modellt az alábbi módon:

$$\tilde{x}_{t,i} = c_i + \beta_{1,i}x_{t-1,i} + \beta_{2,i}D_{\tau,j} + \varepsilon_{t,i} \quad (90)$$

ahol j napon belül az adott rekesz sorszám, ezért $j = \{1, 2, \dots, 25, 26\}$. Az adott napot τ jelöli, vagyis $\tau = \frac{t-j}{26} + 1$ teljesül¹⁴. Végül pedig D egy dummy az alábbiak szerint:

$$D_{\tau,j} = \begin{cases} 1 & \text{ha } j = 1 \\ 0 & \text{egyébként} \end{cases} \quad (91)$$

Nevezzük ezt modellt Poli(14)_f(AR_D) verziónak, és az 51. táblázat segítségével vizsgáljuk meg a teljesítményét.

MSE alapon romlást tapasztalunk mind a nyerő részvény darabszám, mind az átlagos érték tekintetében az összes ármozgás mutatónál. Ugyanez elmondható nem csak a benchmarkkal összehasonlítva, de az előző pontban vizsgált modellhez képest is, ami kizárólag a dummy hiányában különbözik ettől. MAPE alapon a részvények zöménél javulást értünk el minden ármozgás mutatóval, az átlagok azonban kivétel nélkül romlottak (ha nem is jelentős mértékben). Ez úgy fordulhatott elő, hogy ahol javulás volt, az csak minimális, ezzel szemben a JPMorgan Chase & Co. részvényénél a romlás jelentősebb. Hasonló megállapítás tehető a dummy nélküli változattal összehasonlítva is: a nyerő részvények száma némileg nőtt, de az átlagok romlottak.

Ezek alapján megállapítható, hogy a dummy ebben a specifikációban nem volt segítségünkre, hiszen jobb eredményt kaptunk a dummy nélküli, minden másban azonos modellel (és az ármozgást nem tartalmazó benchmarkkal is).

¹⁴Ugyanez t -re rendezve talán többet mond: $t = 26(\tau - 1) + j$

| Poli(14)_f(AR_D) változatok | Hány részvénynél nyert | | Átlagos érték | |
|--------------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Volatilitás | 1 | 29 | 6,38E-04 | 39,3% |
| Ársáv | 2 | 30 | 6,29E-04 | 39,2% |
| Ársáv (%) | 3 | 30 | 6,31E-04 | 39,2% |
| Tényleges ársáv | 1 | 30 | 6,41E-04 | 39,3% |
| Tényleges ársáv (%) | 1 | 30 | 6,43E-04 | 39,3% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

51. táblázat. A Poli(14)_f(AR_D) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

10.3.3.3. AR modell egy napos késleltetéssel bővítve

A forgalomhoz hasonlóan a volatilitás stilizált tényei közt is szerepel a napon belüli U alak (l. Wood et al. (1985), Harris (1986), Andersen és Bollerslev (1997)). Korábban (l. 10.1.1 pont) az U alak dekomponálása nélküli modelleket vizsgálva egyszerűségéhez képest jól teljesített az a modell, amelyben szerepelt egy egy napos késleltetésű tag is. Mivel ennek oka éppen az U alakban kereshető, ebből kiindulva az ármozgás előrejelzésében is helyettesítsük az előző pontban látott dummy-t egy ilyen taggal:

$$\tilde{x}_{t,i} = c_i + \beta_{1,i}x_{t-1,i} + \beta_{2,i}x_{t-26,i} + \varepsilon_{t,i} \quad (92)$$

Nevezzük az így felírt modellt Poli(14)_f(AR_1.26) verzióknak. A volatilitás mellett a többi ármozgás mutatóra is vizsgáljuk meg az így felírt modell teljesítményét az 52. táblázat segítségével.

Gyakorlatilag minden ármozgás mutatóra nagyon hasonló kép rajzolódik ki, ezért tárgyalhatjuk őket közösen. MSE alapon a benchmarkhoz képest mindkét szempont szerint romlást tapasztalunk, és ugyanez mondható el az egy napos késleltetést nem tartalmazó egyszerűbb változattal szemben is (a dummy-zott modellnél viszont némileg jobb). MAPE alapon a részvények döntő többségénél javulást látunk a benchmarkhoz képest, és ez minimális mértékben az átlagos értékekben is megjelenik. (Ez a változat némileg MAPE alapon is jobb az egyszerű AR verziónál, és a dummy-t tartalmazónál is).

Összességében nem sikerült érdemi javulást elérni az egy napos késleltetéssel

| Poli(14)_f(AR_1.26) változatok | Hány részvénynél nyert | | Átlagos érték | |
|-----------------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Volatilitás | 1 | 31 | 6,26E-04 | 39,0% |
| Ársáv | 3 | 30 | 6,18E-04 | 38,8% |
| Ársáv (%) | 4 | 32 | 6,19E-04 | 38,8% |
| Tényleges ársáv | 2 | 31 | 6,24E-04 | 39,0% |
| Tényleges ársáv (%) | 3 | 31 | 6,23E-04 | 38,9% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

52. táblázat. A Poli(14)_f(AR_1.26) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

sem, ugyanis a benchmarkhoz képest a MAPE szerint látott javulás nem jelentős, hiszen habár sok részvénynél, de csak minimális javulás látható, MSE alapon pedig egyértelműen romlott a helyzet.

10.3.3.4. U-módszer

A volatilitás napon belüli U alakjának stilizált tényéből kiindulva ezúttal vizsgáljuk meg azt is, hogyan teljesít a forgalom előrejelzésben elterjedt U-módszer (l. 3.2. alpont) az ármozgás mutatókra alkalmazva. Alap esetben tehát legyen a modell a következő:

$$\tilde{x}_{t,i} = \frac{1}{L} \sum_{l=1}^L x_{t-26-l,i} \quad (93)$$

ahol $L = 20$ az átlagoláshoz használt megelőző napok száma. Nevezzük a fenti módon előrejelzett ármozgás mutatóval dolgozó modellt Poli(14)_f(U) verziónak. Az 53. táblázatban ezen változat benchmarkkal való összevetését láthatjuk. Úgy tűnik, minden ármozgás mutatóra, valamint MSE és MAPE alapon egyaránt romlott a teljesítmény, akár a nyerő részvény darabszámot, akár az átlagos értékeket figyeljük. Egyetlen kivétel a MAPE alapon értékelt volatilitás, amely 23 részvénynél ért el javulást, de az átlagos MAPE érték itt is romlott.

Az U-módszer által megadott előrejelzés *felhasználásának* azonban van egy hatékonyabb módja is. Eddig az előző modellekhez hasonlóan (88)-ben $\beta_{2,i}$ becsléséhez a tény x értékeket használtam, és csak az előrejelzéshez használtam az \tilde{x}

| Poli(14)_f(U) változatok | Hány részvénynél nyert | | Átlagos érték | |
|-----------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Volatilitás | 1 | 23 | 6,38E-04 | 39,4% |
| Ársáv | 1 | 5 | 6,46E-04 | 39,8% |
| Ársáv (%) | 1 | 4 | 6,56E-04 | 39,8% |
| Tényleges ársáv | 1 | 10 | 6,49E-04 | 39,7% |
| Tényleges ársáv (%) | 1 | 10 | 6,55E-04 | 39,8% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

53. táblázat. A Poli(14)_f(U) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

ármozgás értékeket. Mivel azonban az U-módszer olyan átlagos értéket ad előrejelzésként, amely a becslési időszakra jellemző, ezért a (88)-ben $\beta_{2,i}$ becslésekor a becslési időszakban helyettesíthetjük az x tény értékeket a megfelelő \tilde{x} átlagos értékekkel. Az így módosított módon becsült modellt nevezzük Poli(14)_f(Um) verziónak.

Az 54. táblázat szerint így dolgozva minden ármozgás mutatóra nézve hasonló eredményt kaptunk. MSE alapon csak 2 részvénynél sikerült javítani, de lényegében ugyanarra az eredményre vezetett ez a módszer is, mint maga a benchmark, hiszen az átlagos értékek a megjelenített pontosság mellett nem térnek el. MAPE alapon 31 részvénynél sikerült javítani, de az átlagos értékekből látszik, hogy ez a javítás minimális, a volatilitásnál a megjelenített pontosság mellett nem is mutatkozik különbség.

Összességében azt mondhatjuk, hogy az U-módszert érdemes volt a módosított módon felhasználni, mert ez önmagában jelentősen javította az eredményeket a normál módon történő felhasználáshoz képest. Ugyanakkor a Poli(14)_f(Um) lényegében ugyanarra az eredményre vezetett, mint maga a benchmark, viszont nem csak bonyolultabb annál, de az adatigénye is kétszer akkora, ezért nem javasolható az erre való áttérés sem.

10.3.3.5. ARMA modell

Mivel az U-módszer nem bizonyult kellően sikeresnek, térjünk vissza az AR modellhez, és bővítsük ezúttal egy mozgóátlagolású taggal. Legyen tehát a modell az alábbi ARMA(1,1) specifikáció:

| Poli(14)_f(Um) változatok | Hány részvénynél nyert | | Átlagos érték | |
|------------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Volatilitás | 2 | 31 | 6,07E-04 | 39,0% |
| Ársáv | 2 | 31 | 6,07E-04 | 38,9% |
| Ársáv (%) | 2 | 31 | 6,07E-04 | 38,9% |
| Tényleges ársáv | 2 | 31 | 6,07E-04 | 38,9% |
| Tényleges ársáv (%) | 2 | 31 | 6,07E-04 | 38,9% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

54. táblázat. A Poli(14)_f(Um) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

| Poli(14)_f(ARMA) változatok | Hány részvénynél nyert | | Átlagos érték | |
|--------------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Volatilitás | 18 | 30 | 6,08E-04 | 38,8% |
| Ársáv | 12 | 33 | 6,09E-04 | 38,6% |
| Ársáv (%) | 14 | 33 | 6,09E-04 | 38,5% |
| Tényleges ársáv | 18 | 33 | 6,07E-04 | 38,6% |
| Tényleges ársáv (%) | 19 | 33 | 6,04E-04 | 38,6% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

55. táblázat. A Poli(14)_f(ARMA) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

$$\tilde{x}_{t,i} = c_i + \beta_{1,i}x_{t-1,i} + \beta_{2,i}\varepsilon_{t-1,i} + \varepsilon_{t,i} \quad (94)$$

Nevezzünk ezt a felírást Poli(14)_f(ARMA) verziónak. Az 55. táblázat alapján MAPE alapon javulás látszik minden ármozgás mutató szerint. A volatilitással 30 részvénynél, a többi ármozgás mutatóval pedig az összesnél javulást értünk el, amely kisebb-nagyobb mértékben az átlagokban is megmutatkozik (a legalacsonyabb átlagot a százalékos ársávval sikerült elérni). MSE alapon már vegyesebb a kép. A nyerő részvények száma csak a volatilitás és a tényleges ársáv két változata mellett haladja meg az összes részvény felét. Az átlagos értékek tekintetében pedig egyedül a százalékos tényleges ársáv esetén mutatkozik némi javulás.

Összességében azt mondhatjuk, hogy a Poli(14)_f(ARMA) változat a szá-

| Poli(14)_f(ARMA_D) változatok | Hány részvénynél nyert | | Átlagos érték | |
|----------------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Volatilitás | 3 | 31 | 6,25E-04 | 39,0% |
| Ársáv | 4 | 31 | 6,14E-04 | 38,8% |
| Ársáv (%) | 8 | 31 | 6,12E-04 | 38,8% |
| Tényleges ársáv | 3 | 31 | 6,23E-04 | 38,9% |
| Tényleges ársáv (%) | 3 | 31 | 6,21E-04 | 38,8% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

56. táblázat. A Poli(14)_f(ARMA_D) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

zalékos tényleges ársáv felhasználásával minden szempont szerint némileg jobb eredményre vezetett, mint a benchmarknak tekintett Poli(14)_AR.

10.3.3.6. ARMA modell dummy-val

A 10.3.3.2. alpontban elméleti indítatásból az ármozgás mutató modelljében dummy változót használtunk a nap első megfigyelésére. Habár ott inkább romlást okozott a dummy változó, érdemes lehet egy másik specifikációban is tesztelni. Egészítsük ki az előző pontban vizsgált ARMA modellt az alábbi módon:

$$\tilde{x}_{t,i} = c_i + \beta_{1,i}x_{t-1,i} + \beta_{2,i}\varepsilon_{t-1,i} + \beta_{3,i}D_{\tau,j} + \varepsilon_{t,i} \quad (95)$$

ahol a dummy-t a (91)-hez hasonlóan definiáljuk. Legyen ez a modell a Poli(14)_f(ARMA_D) verzió.

Az 56. táblázat szerint MSE alapon egyértelműen rosszabb a modell, mint a benchmark, vagy akár a dummy nélküli változat. MAPE alapon a legtöbb részvénynél javulás látható a benchmarkhoz képest, de a volatilitás kivételével a dummy nélküli változathoz képest keveseb esetben. Az átlagok a benchmarkénál némileg alacsonyabbak, de a dummy nélküli változat felett maradnak.

Összességében megállapítható, hogy a dummy ez esetben sem segített javítani az előrejelzésen.

10.3.3.7. GARCH modell

Általánosan elfogadott stilizált tény, hogy a hozamok volatilitása nem konstans időben, hanem magas és alacsony volatilitású időszakok váltják egymást. Ezt a jelenséget gyakran a volatilitás tömörülésének (volatility clustering) is nevezik. Elsősorban ezen jelenség figyelembe vételére alkották meg az ARCH típusú modelleket, melyeknek egy, a gyakorlatban is széles körben elterjedt változata a GARCH(1,1) modell. (Petrimán és Tulassay (2005))

Ebben a pontban egy ilyen modellt illeszttek az ármozgás mutatókra. Az alábbi specifikációban tehát a várható érték egyenlet AR(1), a variancia egyenlet pedig GARCH(1,1) szerint alakul:

$$\begin{aligned}\tilde{x}_{t,i} &= c_i + \beta_{1,i}x_{t-1,i} + \eta_{t,i} \\ \eta_{t,i} &= \sigma_{t,i}\varepsilon_{t,i} & \varepsilon_{t,i} &\sim N(0,1) \\ \sigma_{t,i}^2 &= a_{0,i} + a_{1,i}\eta_{t-1,i}^2 + b_{1,i}\sigma_{t-1,i}^2\end{aligned}\tag{96}$$

Legyen az így kapott modell neve Poli(14)_f(GARCH) verzió. A klasszikus felírásban a modell várható érték egyenlete a loghozamra vonatkozik, de mint korábban láttuk, esetünkben ez tökéletes előrejelzés esetén sem lenne hasznos. Ezért az 57. táblázatban továbbra is a szokásos ármozgás mutatókra alkalmazott modell eredményét tekinthetjük meg. Megállapítható, hogy MSE alapon mindkét szempont szerint rosszabb eredményt kapunk a benchmarknál az összes ármozgás mutató mellett. MAPE alapon a részvények többségénél javult az eredmény, de az átlagos értékek most sem csökkentek a korábban látottaknál nagyobb mértékben. A megállapítások akkor sem változnának sokat, ha a benchmark helyett a *sima* AR(1) módon előrejelzett ármozgás mutatókat használó modellel (l. 50. táblázat) vetjük össze a GARCH hatást is tartalmazó jelenlegi modellünket: MSE alapon a GARCH modell előrejelzése némileg gyengébben teljesít, MAPE alapon pedig minimálisan jobban.

Összességében ezek alapján nem érdemes az ebben a pontban bemutatott modellekre áttérni a benchmarkról.

10.3.3.8. Áttekintés

Korábban (l. 10.3.2.) az egyidejű ármozgások felhasználásával sikerült javulást elérnem a forgalom előrejelzésben, az egyidejű ármozgás azonban az előrejelzés pillanatában még nem áll rendelkezésre, ezért ennek a megoldásnak nincs gyakorlati

| Poli(14)_f(GARCH) változatok | Hány részvényt nyert | | Átlagos érték | |
|---------------------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Volatilitás | 1 | 24 | 6,23E-04 | 38,9% |
| Ársáv | 3 | 33 | 6,20E-04 | 38,6% |
| Ársáv (%) | 3 | 33 | 6,21E-04 | 38,6% |
| Tényleges ársáv | 2 | 32 | 6,22E-04 | 38,8% |
| Tényleges ársáv (%) | 4 | 33 | 6,20E-04 | 38,7% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

57. táblázat. A Poli(14)_f(GARCH) modellek értékelése
különböző ármozgás mutatók használata esetén
Összevetés a Poli(14)_AR modellel
Forrás: Saját szerkesztés

relevanciája. Ebből kiindulva a 10.3.3. pontban azt vizsgáltam, hogy az egyidejű ármozgás helyett különböző ármozgás előrejelzések felhasználásával is sikerül-e javítani a forgalom előrejelzésén. Az itt bemutatott modellek annyiban térnek el a benchmarknak tekintett Poli(14)_AR modelltől, hogy az egyedi rész egyenletében egy előrejelzett ármozgás mutató is szerepel. A 10.3.3. egyes alpontjaiban ötféle ármozgás mutató különböző előrejelző modelljei mellett vizsgáltam a kérdést.

A tesztelt ármozgás előrejelző modellekben elsősorban olyan jellemzőket vettem figyelembe, mint a volatilitás U alakja, a volatilitás tömörülése, illetve az a jelenség, hogy a nap végi zárás és a következő nyitás között hosszabb idő telik el, mint az egyéb szomszédos megfigyelések között. A modellek eredményességét továbbra is a forgalom előrejelzéshez való hozzájárulásukon keresztül értékeltem.

Bármelyik ármozgás előrejelzést is tartalmazó modelltől elmondható, hogy legalább kétszer annyi adatot igényel¹⁵, mint a Poli(14)_AR modell, és az ármozgás előrejelzés miatt bonyolultabb is, mint ez utóbbi, benchmarknak tekintett változat. Ezek pedig ceteris paribus mindenképp hátrányt jelentenek.

Összességében egyik itt vizsgált modell sem bizonyult elég jónak ahhoz, hogy érdemes legyen azt az új benchmarknak tekinteni. Egyetlen esetben, nevezetesen a Poli(14)_f(ARMA) százalékos tényeges ársávra történő alkalmazása esetén láttuk azt, hogy mind MSE, mind MAPE alapon javulást értünk el a nyerő részvénytársam és az átlagos hibaérték tekintetében is, de ezt a javulást sem értékelem kellően jelentősnek a fent említett hátrányokkal szembeállítva.

¹⁵A forgalom mellett az ármozgás mutató adatsorát is, de pl. a tényeges ársáv előállításához a legmagasabb, legalacsonyabb, és záró árakra is szükség van (l. (71)).

Meg kell azonban jegyezni, hogy a forgalommal ellentétben az ármozgás mutatók (különösen a volatilitás) modellezésének az irodalma olyannyira kiterjedt, hogy akár a sztochasztikus volatilitás modellek, vagy csak az ARCH típusú modellek kiterjesztéseinek a feltérképezése és megbecslése önmagában is egy olyan nagy lélegzetű kutatást igényelne, amely túlmutat jelen értekezés egyetlen alpontjának a keretein. Ebben az irányban tehát később érdemes lehet további kutatásokat végezni.

Ugyanakkor az itt vizsgált modellek egyöntetű sikertelensége azt sejteti, hogy a 10.3.2. pontban látott, egyidejű ármozgás mutatókkal elért javulás nagyobb része talán mégsem tartható meg, ha előrejelzéssel helyettesítjük az egyidejű értéket. Ennek egy lehetséges oka, hogy a forgalom és az ármozgás mutatók együtt mozgása annak tudható be, hogy hasonló hatások mozgatják őket, és így a rájuk ható váratlan sokkok is ugyanazok. Emiatt az egyidejű ármozgás azért lehet hasznos, mert abban már megjelenik azon előrejelezhetetlen sokkok (például egy előzmény nélküli, szokatlanul nagy vételi ajánlat) hatása, ami a forgalmat is eltéríti ugyanabban az időszakban. Az ilyen váratlan sokkokat azonban feltehetően a fejlettebb ármozgás mutató előrejelző modellek sem fogják tudni megfelelően reprodukálni.

10.3.4. ARMA változatok

Az ármozgás mutatókat magunk mögött hagyva vizsgáljuk meg a Poli(14) modell által dekomponált U alak után fennmaradó egyedi rész néhány egyéb, ármozgás mutatók használata nélküli előrejelzési lehetőségét is.

10.3.4.1. ARMA(1,1)

Mivel eddig az egyedi részre az AR(1) modell szerepelt a legjobban, kézenfekvőnek tűnik tesztelni az ARMA(1,1) modellt is:

$$\tilde{e}_{t,i} = c_i + \beta_{1,i}e_{t-1,i} + \beta_{2,i}\varepsilon_{t-1,i} + \varepsilon_{t,i} \quad (97)$$

Az 58. táblázatban ezen specifikáció (nevezzük Poli(14)_ARMA modellnek) teljesítményét vethetjük össze a Poli(14)_AR modellel. Megállapítható, hogy a Poli(14)_ARMA minden vizsgált szempont szerint egyértelműen jobban teljesít. Ez alapján a Poli(14)_ARMA az eddig vizsgált modellek közül a legjobb, ezért a további modellek értékelésekor ezentúl ennél próbálunk meg jobbat találni.

| | Hány részvényt nyert | | Átlagos érték | |
|---------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli(14)_ARMA | 33 | 33 | 5,72E-04 | 36,7% |
| Poli(14)_AR | 0 | 0 | 6,07E-04 | 39.0% |

58. táblázat. A Poli(14)_ARMA és a Poli(14)_AR modell összevetése
 Forrás: Saját szerkesztés

10.3.4.2. ARMA - modellszelekciós kritériumok alapján

Az előző pontban látott ARMA(1,1) sikerét látva felmerül, hogy más késleltetéseket választva esetleg tovább javítható az eredmény. A lehetséges specifikációkat ARMA(P,Q) módon jelölve tehát az alábbiak szerint írhatjuk fel a lehetőségeket.

Az eseteket vizsgáljuk $0 \leq P \leq 5$ és $0 \leq Q \leq 5$, illetve $P + Q > 0$ mellett a következő alakban:

$$\tilde{\varepsilon}_{t,i} = \begin{cases} c_i + \sum_{p=1}^P \alpha_{p,i} e_{t-p,i} + \sum_{q=1}^Q \beta_{q,i} \varepsilon_{t-q,i} + \varepsilon_{t,i} & \text{ha } P \in A; Q \in A \\ c_i + \sum_{p=1}^P \alpha_{p,i} e_{t-p,i} + \varepsilon_{t,i} & \text{ha } P \in A; Q = 0 \\ c_i + \sum_{q=1}^Q \beta_{q,i} \varepsilon_{t-q,i} + \varepsilon_{t,i} & \text{ha } P = 0; Q \in A \end{cases} \quad (98)$$

ahol $A = \{1, 2, 3, 4, 5\}$. Tehát a fenti 35 specifikációt megbecsüljük minden részvény minden egyes napjára, és mindig a legjobbal készítjük el az egyedi rész előrejelzését. Már csak azt kell eldönteni, hogy mi számít legjobbnak. Mivel az előrejelzés során a tényértékek ismerete nélkül kell döntenünk a modell jóságáról, a szokásos MSE és MAPE hibamértékek nem jöhetnek szóba.

A modellbeli változók számának növelésével nem nő (általában csökken) a reziduumok négyzetösszege (SSR), ugyanakkor csökken a modell szabadságfoka (ami csökkenti az együtthatókon végzett próbák erejét). A két hatás együttes figyelembe vétele érdekében modellszelekciós kritériumokat szokás alkalmazni, melyek közül a két leggyakoribb az Akaike (AIC) és a Schwartz (SIC) információs kritérium. A különböző kritériumok természetesen nem feltétlenül rangsorolják azonos módon az egyes modelleket, (Ramanathan (2003)).

Használjuk elsőként az Akaike kritériumot. Az ennek a segítségével készített modellt nevezzük Poli(14)_ARMA_AIC modellnek, melynek a teljesítményét az 59. táblázatban vethetjük össze az eddigi legjobb modellel. Ez alapján a Poli(14)_ARMA egyértelműen jobban teljesít, vagyis a magasabb késleltetése-

| | Hány részvénynél nyert | | Átlagos érték | |
|-------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli(14)_ARMA_AIC | 0 | 1 | 6,06E-04 | 37,7% |
| Poli(14)_ARMA | 33 | 32 | 5,72E-04 | 36,7% |

59. táblázat. A Poli(14)_ARMA_AIC és a Poli(14)_ARMA modell összevetése

Forrás: Saját szerkesztés

| | Hány részvénynél nyert | | Átlagos érték | |
|-------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli(14)_ARMA_SIC | 2 | 5 | 5,85E-04 | 36,8% |
| Poli(14)_ARMA | 31 | 28 | 5,72E-04 | 36,7% |

60. táblázat. A Poli(14)_ARMA_SIC és a Poli(14)_ARMA modell összevetése

Forrás: Saját szerkesztés

sek rontottak az eredményen, hiába tűntek jobbnak előzetesen az információs kritérium alapján.

Következő lépésben vizsgáljuk meg a modellt a SIC kritérium használata mellett is, amely erősebben bünteti a szabadságfok csökkenését, és ezért alacsonyabb késleltetéseket eredményez. Nevezzük Poli(14)_ARMA_SIC verziónak az így becsült modellt, melynek teljesítményét a 60. táblázat mutatja. A táblázat alapján azt látjuk, hogy az átlagokban, és a nyerő részvények számában is javulás mutatkozik az AIC esethez képest, de még mindig a Poli(14)_ARMA a jobb, noha MAPE alapon az átlagos értékek már közel kerültek egymáshoz.

Mindezek alapján továbbra is az egyszerűbb, Poli(14)_ARMA modell maradt az eddigi legjobb változat.

10.3.5. GARCH változatok

Végezetül vizsgáljunk meg két GARCH(1,1) változatot is az egyedi rész előrejelzésére.

Az első eset legyen a már korábban is látott GARCH(1,1) modell AR(1) várható érték egyenlettel:

| | Hány részvényt nyert | | Átlagos érték | |
|--------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli(14)_GARCH(AR) | 0 | 3 | 6,47E-04 | 37,1% |
| Poli(14)_ARMA | 33 | 30 | 5,72E-04 | 36,7% |

61. táblázat. A Poli(14)_GARCH(AR) és a Poli(14)_ARMA modell összevetése

Forrás: Saját szerkesztés

$$\begin{aligned}
\tilde{\varepsilon}_{t,i} &= c_i + \beta_{1,i}e_{t-1,i} + \eta_{t,i} \\
\eta_{t,i} &= \sigma_{t,i}\varepsilon_{t,i} & \varepsilon_{t,i} &\sim N(0,1) \\
\sigma_{t,i}^2 &= a_{0,i} + a_{1,i}\eta_{t-1,i}^2 + b_{1,i}\sigma_{t-1,i}^2
\end{aligned} \tag{99}$$

Nevezzük ezt a verziót Poli(14)_GARCH(AR) modellnek, melynek az értékelését a 61. táblázatban láthatjuk. A táblázat alapján a Poli(14)_ARMA egyértelműen jobb ennél.

Mivel továbbra is az ARMA(1,1) modell az eddigi legjobb az egyedi részre, második esetként becsüljük meg ugyanezt GARCH varianciával is a következő módon:

$$\begin{aligned}
\tilde{\varepsilon}_{t,i} &= c_i + \beta_{1,i}e_{t-1,i} + \beta_{2,i}\eta_{t-1,i} + \eta_{t,i} \\
\eta_{t,i} &= \sigma_{t,i}\varepsilon_{t,i} & \varepsilon_{t,i} &\sim N(0,1) \\
\sigma_{t,i}^2 &= a_{0,i} + a_{1,i}\eta_{t-1,i}^2 + b_{1,i}\sigma_{t-1,i}^2
\end{aligned} \tag{100}$$

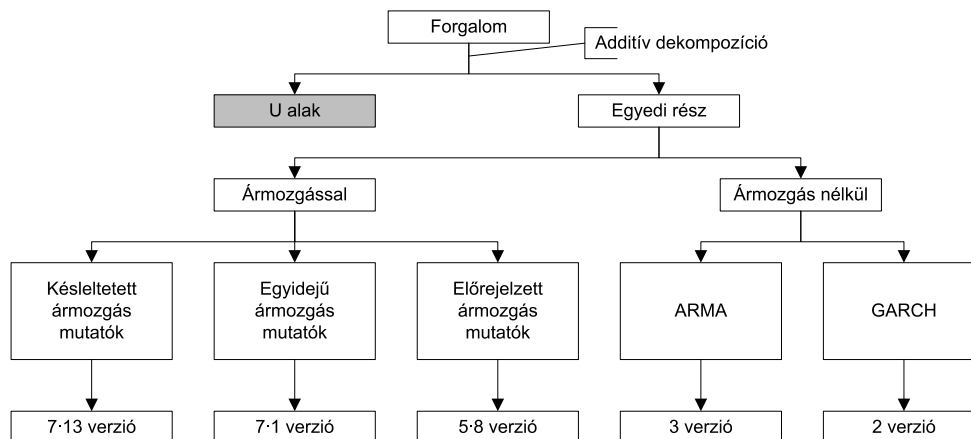
Legyen ez a Poli(14)_GARCH(ARMA) változat. A 62. táblázat alapján ez esetben már érdekesebb képet látunk. MSE alapon továbbra is határozottan a Poli(14)_ARMA a jobb mindkét szempont szerint. MAPE alapon azonban a részvények döntő többségénél a GARCH változat adott kisebb hibát, és az átlagos érték is ennél lett alacsonyabb. A döntés itt tehát attól függ, hogy az MSE vagy a MAPE kritériumot tartjuk-e fontosabbnak az adott helyzetben.

Ha mégis választani kell egy modellt, amelyet a legjobbnak nevezhetünk, akkor, véleményem szerint, az továbbra is a Poli(14)_ARMA lenne a következők miatt. Egyrészt mellette szól, hogy ez az egyszerűbb specifikáció. Másrészt az MSE átlagos értékek aránya a jobb modell szempontjából $\frac{5,72E-04}{6,46E-04} = 88,5\%$, míg a MAPE értékre ugyanez $\frac{36,2\%}{36,7\%} = 98,6\%$, vagyis az ARMA modell választása esetén többet javítunk MSE alapon, mint amit feláldozunk MAPE alapján.

| | Hány részvényt nyert | | Átlagos érték | |
|----------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli(14)_GARCH(ARMA) | 5 | 28 | 6,46E-04 | 36,2% |
| Poli(14)_ARMA | 28 | 5 | 5,72E-04 | 36,7% |

62. táblázat. A Poli(14)_GARCH(ARMA) és a Poli(14)_ARMA modell összevetése

Forrás: Saját szerkesztés



14. ábra. Az egyedi rész előrejelzésének különböző módjai

Az U alak minden esetben Poli(14) alapján dekomponálva

Forrás: Saját szerkesztés

10.3.6. Áttekintés

A forgalmat additív módon dekomponálva két feladatunk van: modellezni az U alakot, és modellezni az egyedi részt. Az U alak modellezésével a 10.2. pontban foglalkoztam (l. 13. ábra), ami alapján a Poli(14) modellt találtam a legjobbnak a vizsgált lehetőségek közül. Ezután a 10.3. pontban rátértem a Poli(14) dekompozíció után visszamaradt egyedi rész előrejelzési lehetőségeinek vizsgálatára. A 14. ábra szemlélteti ennek a pontnak a felépítését.

Az első három alpontban hétféle ármozgás mutató (loghozam, volatilitás, rés, normál és százalékos ársáv, valamint normál és százalékos tényleges ársáv) különböző felhasználási lehetőségeit vizsgáltam. A késleltetett ármozgás mutatók használata a korábbi vizsgálatok alapján ígéretes volt, azonban mégsem sikerült javítást elérni a segítségükkel. Teszteltem olyan feltételes modelleket is, ahol a késleltetett ármozgás és a forgalom közötti kapcsolat különböző módokon definiált szorosságának függvényében használok vagy hagyom ki a modelltől az ármozgás

mutatókat, de ez sem vezetett eredményre.

A következő lépés az egyidejű ármozgás mutatók használata volt. Ez alapján a loghozam és a rés kivételével a többi ármozgás mutató alkalmazása esetén határozottan javultak az eredmények. Ez természetesen csak egy elméleti lehetőség, a gyakorlatban nem lehet így előrejelezni. Azt ugyanakkor megmutatja ez az eredmény is, hogy ha tökéletesen előre tudnánk jelezni az ármozgás mutatókat, akkor mekkora javulást érhetnénk el ilyen módon. Mivel az így elért javulás egyértelmű (10-13%-os) volt, ezért érdemes az előrejelzett ármozgás mutatók modellbe foglalását is vizsgálni.

Ezután következett az előrejelzett ármozgás mutatók tesztelése (az előzőek alapján a loghozam és a rés kivételével). Az itt vizsgált ármozgás előrejelző modellekben elsősorban olyan jellemzőket vettem figyelembe, mint a volatilitás U alakja, a volatilitás tömörülése, illetve az a jelenség, hogy a nap végi zárás és a következő nyitás között hosszabb idő telik el, mint az egyéb szomszédos megfigyelések között. Az itt vizsgált modellek az ármozgás előrejelzés miatt bonyolultabbak, és több adattal dolgoznak a benchmarknál, ami *ceteris paribus* hátránynak tekinthető. Egyetlen verziót találtam ebben a pontban¹⁶, ami minimális javulást eredményezett MSE és MAPE alapon is mindkét szempont szerint, de ezt a javulást nem tekintettem kellően jelentősnek a fent említett hátrányokkal szembeállítva.

Az ármozgás mutatók felhasználásának sikertelensége után a negyedik alpontban már ezek nélkül, pusztán a forgalom egyedi részére vizsgáltam különböző ARMA változatokat. Ezek közül az egyszerű ARMA(1,1) jobbnak bizonyult nem csak a szintén ebben a pontban vizsgált, információs kritériumokon alapuló modellekhez képest, de az eddigi legjobb modellhez képest is. Ezek alapján a továbbiakban a Poli(14)_ARMA modellnél keresünk még jobbakat.

Végül megbecsültem két GARCH változatot is. Ezek közül az ARMA várható érték egyenlettel rendelkező verzió MAPE alapon némileg jobb volt a GARCH variancia nélküli ARMA modellnél, ugyanakkor ennél nagyobb mértékben romlott a teljesítménye MSE alapon, ezért összességében nem tartottam meg ezt a változatot sem.

Az egyedi rész vizsgálata során tehát sikerült felülmúlni a korábbi modelleket egy jobbal, nevezetesen a Poli(14)_ARMA modellel.

¹⁶A Poli(14)_f(ARMA) modell a százalékos tényleges ársávra alkalmazva.

| | Hány részvénynél nyert | | Átlagos érték | |
|----------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| BDF_U | 0 | 15 | 1,05E-03 | 48,9% |
| U-módszer | 0 | 0 | 1,02E-03 | 50,3% |
| Poli(14)_U | 0 | 0 | 1,01E-03 | 50,1% |
| Spline(N4K6)_U | 0 | 0 | 1,01E-03 | 50,2% |
| Poli_expw(7)_U | - | - | 9,33E-04 | 47,7% |

63. táblázat. Dekomponált U alakok összevetése
egyedi rész használata nélkül
Forrás: Saját szerkesztés

10.4. Egyéb lehetőségek

A 10.1. pontban U dekompozíció nélküli modelleket vizsgáltunk. A 10.2-ben az U dekompozíció különböző lehetőségeit teszteltük, mindvégig ugyanazon (a BDF modell által is használt) egyedi részek mellett. Ezután a 10.3. pontban a korábban legjobbnak ítélt dekompozíció mellé kerestünk additív módon különböző egyedi részeket.

A 10. fejezet végéhez közeledve vizsgáljunk meg néhány, a fentiekbe nem tartozó lehetőséget is.

10.4.1. U dekompozíció - egyedi rész nélkül

Elsőként ellenőrizzük, hogy az egyes U dekompozíciók hogyan teljesítenek önmagukban, tehát az egyedi rész nélkül. A vizsgált módszerek legyenek a BDF módszer (Bai (2003) alapján), a gyakorlatban is használt U-módszer, valamint a 10.2. pontban bevezetett módszerek a legjobbnak ítélt paraméterezés mellett, tehát a Poli(14), a Poli_expw(7), valamint a Spline(N4K6) modellek. Jelöljük a módszer neve mögé írt U karakterrel (pl. Poli(14)_U), hogy a modell csak a dekomponált U alakot használja fel.

A 63. táblázatban a Poli_expw(7)_U modellel összevetésben figyelhetjük meg a többit, amelyek jól láthatóan alulmaradnak minden szempont szerint. A BDF_U az egyetlen, amely MAPE alapon felveszi a versenyt az exponenciális súlyozású polinom illesztéssel, de összességében a nyerő részvénytársam és az átlagos érték tekintetében is alul marad annak ellenére, hogy a polinom és spline alapú modellek paramétereit MSE, és nem MAPE alapon választottam ki.

| | Hány részvényt nyert | | Átlagos érték | |
|----------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| BDF_U | 0 | 0 | 1,05E-03 | 48,9% |
| U-módszer | 0 | 0 | 1,02E-03 | 50,3% |
| Poli(14)_U | 0 | 0 | 1,01E-03 | 50,1% |
| Poli_expw(7)_U | 0 | 0 | 9,33E-04 | 47,7% |
| Spline(N4K6)_U | 0 | 0 | 1,01E-03 | 50,2% |
| Poli(14)_ARMA | - | - | 5,72E-04 | 36,7% |

64. táblázat. Dekomponált U alakok értékelése egyedi rész használata nélkül
Összevetés a Poli(14)_ARMA modellel
Forrás: Saját szerkesztés

| | Hány részvényt nyert | | Átlagos érték | |
|----------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Poli_expw(7)_U | 33 | 18 | 9,33E-04 | 47,7% |
| U-módszer | 33 | 3 | 1,02E-03 | 50,3% |
| Poli(14)_U | 33 | 4 | 1,01E-03 | 50,1% |
| Spline(N4K6)_U | 33 | 4 | 1,01E-03 | 50,2% |
| BDF_U | - | - | 1,05E-03 | 48,9% |

65. táblázat. Dekomponált U alakok összevetése
egyedi rész használata nélkül
Forrás: Saját szerkesztés

A 64. táblázatban ugyanezen modelleket az eddigi legjobbhoz viszonyítva azt látjuk, hogy meg sem közelítik annak a teljesítményét.

Mindezek alapján, noha javítani nem sikerült az eddigi eredményeken, néhány észrevételt érdemes tenni.

Egyrészt az irodalomból vett benchmark, vagyis a BDF által használt dekompozíció MSE alapon egyedi résszel (l. korábban), és anélkül is rosszabb az általam javasolt modelleknél. MAPE alapon ugyanakkor egyedi rész nélkül jobb¹⁷, egyedi résszel együtt viszont már szintén rosszabb. Ez utóbbi megállapítás is megerősíti a simítás hatékonyságát, aminek köszönhetően jobban előrejelezhető egyedi részt kapunk.

Másrészt, mivel egyedi rész nélkül minden szempontból az exponenciális súlyozású polinom illesztés a legjobb, csak a dekompozícióval számolva érdemes nagyobb súllyal figyelembe venni a frissebb megfigyeléseket. Ugyanakkor az egyedi

¹⁷Mármint az exponenciális súlyozású kivételével (l. 65. táblázat).

rész előrejelezhetőségéből ezzel a módszerrel annyit veszítünk, hogy egyedi résszel együtt már nem ez a legjobb modell (l. korábban).

Úgy tűnik tehát, hogy az U dekompozíció önmagában nem kellően jó, hiszen sokat javít a teljesítményen, ha valamilyen módon figyelembe vesszük a 15 percnként beérkező új információt is, és nem csak naponta egyszer figyeljük meg az újabb adatokat. Ennek egy lehetséges módja az additív egyedi rész előrejelzése, de elképzelhető egyéb megoldás is. Erre látunk példát a következő pontokban.

10.4.2. Egy korrekciós modell

Induljunk ki ismét a 10.4.1. pontban látott dekomponált U alakok egyikéből. Az ebből kapott előrejelzés legyen $\tilde{u}_{\tau,j}$, ahol $j = \{1, 2, \dots, 25, 26\}$ minden elemére, azaz egész napra megadjuk az előrejelzést, és a korábbiakhoz hasonlóan τ az adott napot jelöli.

Ezután alakítsuk ezt aránnyá a következő módon:

$$u_{\tau,j}^p = \frac{\tilde{u}_{\tau,j}}{\sum_{j=1}^{26} \tilde{u}_{\tau,j}} \quad (101)$$

Az $u_{\tau,j}^p$ megmutatja, hogy az adott τ napra várható teljes forgalom mekkora része várható az egyes (15 perces) j időszakok során.

A következő lépésben jelezzük előre a forgalom napi idősorából, hogy a τ napra összesen mekkora forgalomra számítunk. Tegyük ezt az alábbi ARMA modell segítségével:

$$\tilde{z}_\tau = c + \beta_1 z_{\tau-1} + \beta_2 \varepsilon_{\tau-1} + \varepsilon_\tau \quad (102)$$

ahol z az egy napra eső teljes forgalom. Ahol ez negatív értéket adott (56 eset a 87 ezerből¹⁸), ott az előző tényre cseréltem az előrejelzést.

Az eddigieket felhasználva tehát $u_{\tau,j}^p \cdot \tilde{z}_\tau$ megmutatja, hogy az adott nap j -edik megfigyelésére milyen forgalom értéket várunk a nap elején. Ez eddig lényegében nem sokban különbözik az egyedi részt nem tartalmazó, U dekompozíción alapuló modellektől.

Bővítsük most az eddigieket úgy, hogy beépítünk egy olyan elemet, amely segítségével minden tényadat beérkezése után értékeljük az addigi (aznapi) tel-

¹⁸2648 · 33 = 87384

jesítményt, és ennek megfelelően felülvizsgáljuk a következő 15 percre készített előrejelzést.

Legyen k egy korrekciós tényező az alábbiak szerint:

$$k_{\tau,j} = \begin{cases} 1 & \text{ha } j = 1 \\ \frac{\sum_{i=1}^{j-1} y_{\tau,i}}{\sum_{i=1}^{j-1} u_{\tau,i}^p \cdot \tilde{z}_{\tau}} & \text{egyébként} \end{cases} \quad (103)$$

ahol y a forgalmat jelöli. Ha tehát $j = 1$, akkor a nap első megfigyelését próbáljuk előrejelezni. Ilyenkor nincsen információnk az aznapi forgalomról, ezért nem korrigálunk ($k = 1$). Minden egyéb esetben megnézzük, hogy hogyan viszonyul az aznap megismert teljes forgalom ahhoz, amit addig a pontig előrejeleztünk, és ha a tény nagyobb ($k > 1$), akkor a következő időszakra megnöveljük az eredeti előrejelzésünket. Ha kisebb, akkor pedig csökkentjük.

Az így kapott modellben tehát a következő lesz az előrejelzés:

$$\tilde{y}_{\tau,j} = k_{\tau,j} \cdot u_{\tau,j}^p \cdot \tilde{z}_{\tau} \quad (104)$$

Jelöljük ezt a multiplikatív modellt a használt dekomponáló eljárás neve mögé írt *korr* karaktersorral (pl. Poli(14)_korr), ezzel utalva a korrekciós tényezőre. Gyakorlatilag a 10.4.1. pontban megismert modelleket fejlesztettük tovább oly módon, hogy 15 percenként frissítjük az információs halmazt, vagyis beépítjük az időközben beérkezett új információt is az előrejelzésbe.

A 66. táblázat tanúsága szerint az additív egyedi rész módszerével hatékonyabban be tudjuk építeni az új információt, mint az itt bemutatott modellel, hiszen a Poli(14)_ARMA jobban szerepelt a korrekciós modellnél, bármelyik dekomponálási eljárást használtuk is. Ezért, habár a 10.4.1-ben bemutatott, tisztán U dekompozícióval készült modelleken egyértelműen sikerült javítani az itt leírt korrekcióval¹⁹, összességében nem érdemes ezek egyikére sem áttérni a Poli(14)_ARMA modellről.

10.4.3. Egy multiplikatív modell

Az előző pontban azt láttuk, hogy az információs halmaz frissítésében az ott leírt multiplikatív korrekciónál jobban teljesített az egyedi rész alapú additív megoldás.

¹⁹Ez az átlagos értékekből is jól látszik, ezért nem mellékelek külön összehasonlító táblázatot.

| | Hány részvénynél nyert | | Átlagos érték | |
|-------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| BDF_korr | 0 | 1 | 7,28E-04 | 39,4% |
| U-módszer_korr | 0 | 0 | 7,26E-04 | 39,0% |
| Poli(14)_korr | 0 | 0 | 7,21E-04 | 38,6% |
| Poli_expw(7)_korr | 0 | 0 | 7,74E-04 | 39,0% |
| Spline(N4K6)_korr | 0 | 0 | 7,19E-04 | 38,7% |
| Poli(14)_ARMA | - | - | 5,72E-04 | 36,7% |

66. táblázat. A korrekciós modellek értékelése
 Összevetés a Poli(14)_ARMA modellel
 Forrás: Saját szerkesztés

Az egyedi részt azonban szintén figyelembe vehetjük additív helyett multiplikatív módon is.

Ehhez az U alakot *szorozzuk* egy egyedi rész figyelembe vételéért felelős tényezővel:

$$y_t = u_t \cdot e_t \quad (105)$$

ahol y a forgalom, u az U alak, e pedig az egyedi rész. Az U alakot a már megismert dekompozíciós eljárások egyikével becsüljük, az így kapott előrejelzést jelölje \tilde{u} . Az egyedi tényező a becslési időszakban az alábbi lesz:

$$e_t = \frac{y_t}{u_t} \quad (106)$$

Az egyedi rész előrejelzése \tilde{e} , melyet valamilyen egyedi rész modellel állítunk elő. Ekkor a forgalom előrejelzés:

$$\tilde{y}_t = \tilde{u}_t \cdot \tilde{e}_t \quad (107)$$

Legyen az egyedi rész modellje a korábban legjobbnak talált ARMA modell:

$$\tilde{e}_t = c + \beta_1 e_{t-1} + \beta_2 \varepsilon_{t-1} + \varepsilon_t \quad (108)$$

Az így kapott modelleket jelöljük a dekompozíciós eljárás mögé írt *mult* karakterrel (pl. Poli(14)_mult), jelezve, hogy multiplikatív modellről van szó. Az eredményeket a 67. táblázat mutatja.

Megállapítható, hogy a BDF alapú multiplikatív modell egyértelműen rosszab-

| | Hány részvényt nyert | | Átlagos érték | |
|-------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| BDF_mult | 1 | 3 | 5,94E-04 | 38,0% |
| U-módszer_mult | 29 | 31 | 5,81E-04 | 36,4% |
| Poli(14)_mult | 30 | 32 | 6,62E-04 | 36,1% |
| Poli_expw(7)_mult | 28 | 30 | 6,03E-03 | 37,2% |
| Spline(N4K6)_mult | 30 | 31 | 6,03E-04 | 36,1% |
| Poli(14)_ARMA | - | - | 5,72E-04 | 36,7% |

67. táblázat. A multiplikatív modellek értékelése ARMA egyedi résszel
Összevetés a Poli(14)_ARMA modellel
Forrás: Saját szerkesztés

| 32 részvényre | Hány részvényt nyert | | Átlagos érték | |
|-------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| BDF_mult | 1 | 2 | 3,93E-04 | 38,0% |
| U-módszer_mult | 29 | 30 | 3,73E-04 | 36,3% |
| Poli(14)_mult | 30 | 32 | 3,66E-04 | 35,9% |
| Poli_expw(7)_mult | 28 | 30 | 3,70E-04 | 36,5% |
| Spline(N4K6)_mult | 30 | 31 | 3,67E-04 | 36,0% |
| Poli(14)_ARMA | - | - | 3,75E-04 | 36,6% |

68. táblázat. A multiplikatív modellek értékelése ARMA egyedi résszel
az Altria Group Inc. részvényének figyelmen kívül hagyása mellett
Forrás: Saját szerkesztés

bul szerepel az eddigi legjobbnál. A többi verziónál azonban érdekes képet látunk, hiszen mind MSE, mind MAPE alapon a részvények döntő többségére mindegyik kisebb hibát eredményez, mint a Poli(14)_ARMA modell. Ugyanakkor MAPE alapon az exponenciális súlyozású szorzatos változat átlagos értéke magasabb, MSE alapon pedig minden verzió magasabb átlagos értéket ad a Poli(14)_ARMA modellnél.

Mindez úgy fordulhat elő, hogy a szorzatos modellek az Altria Group Inc. részvényére gyengébb előrejelzést produkálnak, és ez az egy részvény elrontja a 33 részvény átlagában mért értéket is. A 68. táblázatban láthatjuk az értékelést az Altria Group Inc. figyelmen kívül hagyása mellett is. A BDF alapú modell továbbra is minden szempont szerint rosszabb, a többi azonban így már egyértelműen jobb a Poli(14)_ARMA modellnél.

Lehetséges-e, hogy ennél az egyetlen részvénytől olyan kiemelkedően gyenge a

| | Altria Group Inc. | |
|-------------------|-------------------|--------|
| | MSE | MAPE |
| BDF_mult | 102,3% | 99,4% |
| U-módszer_mult | 105,4% | 99,3% |
| Poli(14)_mult | 147,2% | 106,7% |
| Poli_expw(7)_mult | 2721,0% | 152,3% |
| Spline(N4K6)_mult | 118,5% | 102,4% |
| Poli(14)_ARMA | 100,0% | 100,0% |

69. táblázat. Az egyes modellek hibái az Altria Group Inc. részvényére a Poli(14)_ARMA modell hibájának arányában
Forrás: Saját szerkesztés

szorzatos specifikáció, hogy egyedül elrontja az átlagot? Vizsgáljuk meg az egyes modellek hibáit külön az Altria Group Inc. részvénye esetén is. A 69. táblázatban a Poli(14)_ARMA modell hibájának arányában tüntetem fel az adott értékeket. Jól látszik, hogy MSE alapon minden modell rosszabbul szerepel az additív verziónál, a polinom és spline alapú modellek emellett MAPE alapon is gyengébbek a Poli(14)_ARMA változatnál. Ugyanakkor az exponenciális súlyozású kivételével nem végletesen rossz az arányok.

Észrevehetjük, hogy az átlagos MSE érték jelentősen csökkent az Altria Group Inc. kihagyásával, míg a MAPE ehhez képest alig változott. Ennek az az oka, hogy a MAPE százalékos mutató, míg az MSE nem az. Ezért az MSE esetén az egyszerű átlagolás torzít, ugyanis nagyobb súlyt kapnak a nagyobb átlagos forgalmú papíroknak vétett hibák. A 70. táblázatban láthatjuk az egyes részvények 15 perces időszakra vetített átlagos forgalmát a teljes vizsgált időszakban²⁰. Mind közül a legnagyobb forgalommal az Altria Group Inc. (MO) rendelkezik, melynek 9,3-szor magasabb az átlagos értéke, mint a legkisebb forgalmú részvénynek. Ennek fényében már jobban érthető, hogy egyetlen részvény hibája hogyan ronthatta el a teljes MSE átlagot a multiplikatív modellek esetén, miközben egyedileg csak néhány részvényre adott rosszabb eredményt, mint az additív változat.

Ejtsünk pár szót arról a kiemelkedően nagy hibáról is, amelyet a 69. táblázatban tapasztaltunk a Poli_expw(7) dekompozíció használata során. Amint korábban már láttuk, az exponenciális súlyozás önmagában alkalmazva jól teljesít, azonban az ebből nyert egyedi rész előrejelezhetősége rosszabb az alternatíváknál.

²⁰A mellékletben a 96-98. táblázatban megtalálhatóak a kódokhoz tartozó vállalatok nevei is.

| Kód | Arány | Kód | Arány | Kód | Arány |
|------|-------|-----|-------|-----|-------|
| MO | 9,3 | AXP | 4,0 | VZ | 2,8 |
| AA | 9,2 | BA | 3,9 | GE | 2,8 |
| C | 7,2 | MCD | 3,7 | IBM | 2,7 |
| BAC | 7,0 | DD | 3,6 | CVX | 2,6 |
| INTC | 6,4 | HPQ | 3,5 | T | 2,6 |
| CAT | 6,3 | HON | 3,4 | KFT | 2,3 |
| AIG | 5,2 | MRK | 3,3 | JNJ | 2,2 |
| CSCO | 5,1 | MMM | 3,3 | XOM | 2,1 |
| JPM | 4,5 | PFE | 3,1 | PG | 2,0 |
| HD | 4,2 | DIS | 3,0 | WMT | 2,0 |
| MSFT | 4,0 | UTX | 2,9 | KO | 1,0 |

70. táblázat. Az egyes részvények átlagos 15 perces forgalma a teljes időszakra a The Coca-Cola Company értékének arányában

Forrás: Saját szerkesztés

Az additív esetben ez kevésbé volt probléma, mert egy jó dekompozíció után az egyedi rész aránya alacsony a teljes forgalomhoz képest, ezért annak az előrejelzésében vétett hiba sem lesz olyan jelentős a teljes forgalom szempontjából. Ezzel szemben a multiplikatív specifikációban az egyedi részben vétett hiba teljes egészében átöröklődik a forgalom előrejelzésére. Ha például kétszeresére becsüljük az egyedi részt, akkor hiába sikerült megragadni a dekompozícióval a teljes forgalom jelentős részét, mégis a teljes előrejelzett forgalom is duplázódik. Ez az aspektus, véleményem szerint, hátránya a szorzatos modelleknek az additívakkal szemben.

Végül pedig vessük össze a különböző dekompozícióval becsült modelleket. Ebben segít a 71. táblázat. Látható, hogy a nyerő részvények száma alapján MSE és MAPE alapon is a Poli(14)_mult a legjobb. Emellett az átlagos értéke MAPE alapon is ennek a legkisebb, noha az adott kerekítés mellett a spline-os változat megegyezik vele. A Poli(14)_mult átlagos MSE értékét, mint láttuk, egyetlen részvény viszi a többi verzió fölé, ezt a részvényt kihagyva a MAPE mellett az MSE átlagos értéke is a Poli(14)_mult verzióra a legkisebb (l. 68. táblázat).

Az MSE értékek átlagolása okozta torzítástól eltekintve tehát a vizsgáltak közül a Poli(14)_mult a legjobb multiplikatív modell.

| | Hány részvényt nyert | | Átlagos érték | |
|-------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| BDF_mult | 2 | 1 | 5,94E-04 | 38,0% |
| U-módszer_mult | 9 | 1 | 5,81E-04 | 36,4% |
| Poli_expw(7)_mult | 3 | 0 | 6,03E-03 | 37,2% |
| Spline(N4K6)_mult | 8 | 1 | 6,03E-04 | 36,1% |
| Poli(14)_mult | - | - | 6,62E-04 | 36,1% |

71. táblázat. A multiplikatív modellek összevetése

Forrás: Saját szerkesztés

10.4.4. Áttekintés

A 10.4. alpontban néhány, a korábbi additív dekompozíciótól (U alak + egyedi rész) eltérő logikájú modellt vizsgáltunk meg.

Elsőként az U dekompozíciókat értékeltük egyedi rész figyelembe vétele nélkül. Ezek közül az exponenciális súlyokkal illesztett polinom szerepelt legjobban, ugyanakkor az egyedi részt is figyelembe vevő modelltől jelentősen elmaradtak a mért teljesítmények.

A következő lépésben az U dekompozíciót kiegészítettük egy multiplikatív korrekcióval, melynek lényege, hogy az additív egyedi rész helyett egy alternatív módon veszi figyelembe a 15 percenként beérkező új információt. Ezzel határozottan javult az előrejelzés a tisztán U dekompozícióval dolgozó modellekhez képest, de az additív egyedi rész használata még jobb eredményre vezet.

Végül a multiplikatív dekompozícióval kapcsolatos tesztek következtek. A fentiekben csak ARMA(1,1) egyedi résszel becsültem meg a modelleket, mivel az additív esetben a 10.3. pont alapján ez volt a legjobb. Ezen kívül megbecsültem még AR(1) egyedi résszel is, de minden szempontból hasonló képet mutatott az eredmény, mint ARMA módszerrel, csak gyengébb volt annál. Ezért ezeket az eredményeket nem is részletezem a dolgozatban. Természetesen mindez nem jelenti azt, hogy a multiplikatív változatra esetleg ne lenne érdemes jobban teljesítő egyedi rész modellt keresni. Ennek ellenére ebbe az irányba ezúttal nem indultam el, ez egy későbbi kutatás tárgya lehet.

A multiplikatív dekompozícióval becsült modellek közül a Poli(14)_mult modell bizonyult a legjobbnak, eltekintve az MSE átlagolása okozta torzítástól. Ez a modell ugyanúgy Poli(14) módszerrel választja le az U alakot, mint az eddigi legjobb, a Poli(14)_ARMA, és az egyedi részt is ahhoz hasonlóan, tehát ARMA(1,1)

modellel jelzi előre. Annyi a különbség, hogy mindezt additív helyett multiplikatív dekompozícióval végzi.

A Poli(14)_ARMA és a Poli(14)_mult összevetésében megállapítható, hogy az Altria Group Inc. részvényét kihagyva az elemzésből a multiplikatív modell minden vizsgált szempont szerint jobban szerepel (l. 68. táblázat). Mind a 33 részvény figyelembe vétele mellett pedig kizárólag az MSE átlagos értékében marad el az additív változattól (l. 67. táblázat).

10.5. A modellkeresés eredménye

A 10. fejezetben a forgalom előrejelzésének különböző lehetőségeit teszteltem azzal a céllal, hogy az MSE és MAPE értékek alapján jobbat találjak a benchmarknak tekintett BDF modellnél.

A 10.1. pontban U dekompozíció nélküli modelleket vizsgáltam. Az egyszerű U-módszernél (meglepő módon²¹) sikerült jobb modellt találni ezen specifikációk között is, de a BDF modellnél már nem. A forgalom két stilizált tényét, nevezetesen a hét napja hatást (a tesztek alapján csak a hétfőt érdemes kiemelten kezelni), valamint a forgalommal való kapcsolatot is modellbe foglaltam ebben a pontban, de egyik sem bizonyult sikeres iránynak.

A 10.2. pontban az U alak dekomponálásának különböző lehetőségeit vizsgáltam oly módon, hogy additív egyedi részként végig a BDF modell által is alkalmazott változatokat használtam. Ez utóbbi tette lehetővé, hogy a dekompozíció hatékonyságát külön tudjam értékelni. A dekomponáló eljárások keretében a következőket teszteltem: a korábbról ismert U-módszert, polinom illesztést egyenlő súlyokkal, polinom illesztést exponenciális súlyokkal, valamint spline illesztést. Az egyenlően súlyozott polinomok közül 26 lehetőséget elemezve az $n = 14$ paraméterezést választottam. Az exponenciálisan súlyozott polinomok közül szintén 26 változattól az $n = 7$ fokszám mellett döntöttem. A spline illesztésnél 24 változat közül az N4K6 specifikációt találtam legjobbnak. Ezek közül (meglepő módon az U-módszert is beleértve) mindegyik határozottan jobb eredményre vezet, mint a BDF modellben használt dekompozíció. A legjobb mind közül az egyenlően súlyozott polinom volt, tehát a Poli(14)_(SET)AR modellek.

²¹Azért meglepő, mert a BDF és BCG modellek eredménye is ennek a megverése volt, hiszen mindkét cikkben az U-módszert tekintették benchmarknak.

A 10.3. pontban az előzőekben megtalált Poli(14) dekompozíció mellé kerestem a BDF modell által használt (SET)AR egyedi rész helyett egyéb alternatív lehetőségeket. Végig ezzel a dekompozíciós eljárással dolgozva a tesztelt additív egyedi rész modellek hatékonyságát külön tudtam értékelni. Elsőként különböző ármozgás mutatók²² felhasználási lehetőségeit vizsgáltam. A késleltetett ármozgás mutatók a vizsgált 13 specifikáció egyikében sem vezettek kedvező eredményre. Az egyidejű ármozgás mutatók használata nem megvalósítható irány, azt azonban megmutatta, hogy az ármozgások tökéletes előrejelzése esetén egyértelmű javulást érhetnénk el a forgalom előrejelzésben is. Ezért a következő lépésben ármozgás előrejelző modelleket vizsgáltam 8 különböző specifikációban, és az általuk előrejelzett egyidejű értéket foglaltam modellbe az egyedi rész előrejelzésekor. Az így kapott előrejelzések sem voltak meggyőzően jobbak a Poli(14)_AR modellnél. Végül teszteltem három ARMA és két GARCH változatot is, ezúttal tisztán az egyedi rész előrejelzésére, tehát ármozgás használata nélkül. Ezek közül az ARMA(1,1) volt a legjobb, amely határozottan jobb az eddigi legjobb modellnél is, ezért a továbbiakban már a Poli(14)_ARMA modell megverése lett a következő cél. Ezzel tehát a BDF által alkalmazott egyedi rész modellen is sikerült javítani, noha a várakozásaimmal ellentétben az ármozgás mutatók használata nélkül.

Végül a 10.4-ben teszteltem néhány, a 10.1-10.3. pontokba logikailag nem tartozó modellezési lehetőséget is, elsőként a különböző U dekompozíciókat egyedi rész figyelembe vétele nélkül. Ezek közül az exponenciális súlyozású polinom illesztés szerepelt legjobban, de a Poli(14)_ARMA modelltől jelentősen elmaradt. Ezután megvizsgáltam egy korrekciós modellt, amely a tisztán U dekompozícióval dolgozó modellhez hasonló, de multiplikatív specifikációban figyelembe veszi a 15 percenként beérkező új információt is. Ezzel javult az egyedi rész nélküli dekompozíció eredménye, de az addigi legjobb modellt ezzel a változattal sem sikerült megverni. A következő lépésben egy olyan multiplikatív modellt teszteltem, amelyben a már megismert U dekompozíciós eljárások kiegészítéseként az egyedi részt szorzatos formában vettem figyelembe a korábbi additív alakokkal szemben. Ezúttal az egyedi részt kizárólag ARMA(1,1) modellel becsültem. A különböző U dekompozíciók összevetése alapján az egyenlően súlyozott polinom illesztés használata, tehát a Poli(14)_mult modell bizonyult a legjobbnak a szorzatos verziók közül. Erről a modelltől megállapítható, hogy mind MSE, mind MAPE alapon a

²²Ezek a loghozam, a volatilitás, a rés, a normál és a százalékos ársáv, valamint a normál és a százalékos tényleges ársáv.

részvények zömére jobban szerepel, mint a Poli(14)_ARMA modell. Ezen felül MAPE alapon átlagban is jobb, illetve az Altria Group Inc. vizsgálatból történő kihagyása után a MAPE mellett az MSE átlag is a multiplikatív modellnél lesz kisebb.

A dolgozatban eddig végig az MSE és MAPE kritériumok szerint hasonlítottuk össze a modelleket. Egyrészt vizsgáltuk, hogy a 33 részvényből hány esetben nyer egyik vagy másik modell. Ezen felül azonban az átlagos MSE és MAPE értéket is kiszámoltuk, hogy lássuk, van-e valós különbség a modellek hibái között. Ez a módszer a legtöbb esetben megfelelő segítséget biztosított a döntéshez. A multiplikatív modellek azonban az Altria Group Inc. esetén keresztül felhívták a figyelmet a választott hibamérés egy gyenge pontjára, nevezetesen, hogy az MSE esetén az átlagolás torzításhoz vezet. Ennek oka, hogy az egyes részvények forgalma jelentősen eltér, ezért így számolva olyan, mintha a nagyobb forgalmú részvények hibája nagyobb súlyt kapna az átlagolás során.

Ennek kiküszöbölése céljából számolhatunk egy olyan átlagos MSE értéket is, amely nem torzít a részvények eltérő átlagos forgalma miatt. Erre egy lehetséges megoldás az alábbi:

$$MSE^* = \frac{1}{N} \sum_{i=1}^N \frac{MSE_i}{A_i} \quad (109)$$

ahol MSE_i az i . részvény átlagos MSE értéke, az A_i arány pedig a 70. táblázatban látott arányokat jelenti a megfelelő részvényre. Ily módon a legkisebb átlagos forgalmú részvény szintjére normáljuk az átlagolandó értékeket. Hasonlítsuk össze a két legjobb modellt ezen mutató segítségével is. A 72. táblázatban tehát továbbra is az összes részvényt figyelembe véve láthatjuk a Poli(14)_ARMA és a Poli(14)_mult modellek összevetését. Ezek alapján (a két modell kapcsán korábban tett megállapításokat meg nem ismételve) azt érdemes kiemelni, hogy a Poli(14)_mult modellnek az MSE^* értéke is magasabb, vagyis nem csak az Altria Group Inc. magas forgalma vitte félre a korrigálatlan átlagot, hanem ezen részvényre valóban, tehát arányaiban is olyan nagyot téved a szorzatos modell, hogy a korrigált MSE átlag is elromlik. Ez megerősíti azt a Poli_expw(7) modell kapcsán tett korábbi megállapításunkat, miszerint a szorzatos modellek hátránya az additívakkal szemben, hogy az egyedi részben vétett hiba arányosan átöröklődik a teljes forgalom előrejelzésre. Ez pedig jelentősen félreviheti azt olyan esetekben is,

| | Hány részvény- nél nyert | | Átlagos érték | | Korrigált átlagos érték |
|---------------|-----------------------------|------|------------------|-------|----------------------------|
| | MSE | MAPE | MSE | MAPE | MSE* |
| Poli(14)_ARMA | 3 | 1 | 5,72E-04 | 36,7% | 9,66E-05 |
| Poli(14)_mult | 30 | 32 | 6,62E-04 | 36,1% | 1,05E-04 |

72. táblázat. A Poli(14)_ARMA és a Poli(14)_mult
modellek összevetése
Forrás: Saját szerkesztés

| | Hány részvény- nél nyert | | Átlagos érték | | Korrigált átlagos érték |
|---------------|-----------------------------|------|------------------|-------|----------------------------|
| | MSE | MAPE | MSE | MAPE | MSE* |
| BDF_AR | 0 | 0 | 6,49E-04 | 40,3% | 1,09E-04 |
| BDF_SETAR | 0 | 0 | 6,60E-04 | 39,9% | 1,12E-04 |
| Poli(14)_ARMA | - | - | 5,72E-04 | 36,7% | 9,66E-05 |

73. táblázat. A Poli(14)_ARMA és a BDF_(SET)AR
modellek összevetése
Forrás: Saját szerkesztés

amikor egyébként az U dekompozíció jól sikerül, és az egyedi rész a teljes forgalom kis hányadát adja, tehát önmagában nem lenne jelentős.

Ezek alapján tehát nem egyértelmű, hogy a kettő közül melyik modellt tekintjük legjobbnak, hiszen az Altria Group Inc. részvényt sem hagyhatjuk figyelmen kívül. A 10. fejezet zárásaként hasonlítsuk össze mindkét talált modellt az irodalomból vett benchmarkkal, vagyis a BDF_(SET)AR modellekkel.

A 73. táblázatból látszik, hogy a Poli(14)_ARMA mindkét BDF verziónál egyértelműen jobban szerepel, bármelyik szempont szerint hasonlítjuk is össze őket. A 74. táblázat alapján látható, hogy az Altria Group Inc. előrejelzésében a BDF modellek MSE alapon, illetve a BDF_SETAR esetén MAPE alapon is jobbak a Poli(14)_mult modellnél, az összes többi részvény előrejelzésében azonban rosszabbak nála. Ennek ellenére a korrigálatlan MSE átlag (amely azonban torzít) még a BDF modelleknél is rosszabb a Poli(14)_mult esetén, hiszen az Altria Group Inc. előrejelzésében vétett nagy hibát tetézi, hogy éppen ez a legnagyobb átlagos forgalmú részvény. Az átlagos MAPE és MSE* értéke azonban már a multiplikatív modellnek az alacsonyabb.

Mindezek alapján mindkét javasolt modell, tehát a Poli(14)_ARMA, valamint a Poli(14)_mult is jobbnak tekinthető a BDF benchmarknál, amit az irodalom

| | Hány részvény- nél nyert | | Átlagos érték | | Korrigált átlagos érték |
|---------------|-----------------------------|------|------------------|-------|----------------------------|
| | MSE | MAPE | MSE | MAPE | MSE* |
| BDF_AR | 1 | 0 | 6,49E-04 | 40,3% | 1,09E-04 |
| BDF_SETAR | 1 | 1 | 6,60E-04 | 39,9% | 1,12E-04 |
| Poli(14)_mult | - | - | 6,62E-04 | 36,1% | 1,05E-04 |

74. táblázat. A Poli(14)_mult és a BDF_(SET)AR
modellek összevetése

Forrás: Saját szerkesztés

legjobbjaként azonosítottunk.

11. fejezet

Az értékelés további szempontjai

A dolgozatban mindvégig ugyanúgy értékeltük az egyes modellek teljesítményét. Mindig egy lépésre, tehát 15 percre jeleztünk előre, és felhasználtuk a rendelkezésre álló legfrissebb adatokat. Az így kapott előrejelzéseket aztán MSE és MAPE, illetve a legjobbakat a 10. fejezet végén MSE* (azaz korrigált MSE) alapján is összevetettük egymással.

Ebben a fejezetben megvizsgálunk néhány egyéb lehetőséget is a modellek értékelésére, és ezek alapján is összevetjük a két legjobb²³ modellünket az irodalomból vett benchmarkkal. Új modellt ez a fejezet már nem tartalmaz.

11.1. További lehetőségek a modellek értékelésére

A modellek teljesítményének mérési lehetőségeit több szempont szerint is vizsgálhatjuk:

1. Az értékelés tárgya szerint (Mit értékelünk?)
2. Az értékelés módja szerint (Hogyan értékeljük?)

Az első kérdés, hogy *mit* értékeljünk, arra vonatkozik, hogy hogyan használjuk fel a modellt az előrejelzés elkészítésére. Eddig egy lépésre jeleztünk előre, és lépésenként frissítettük az információs halmazt, majd pedig az így előrejelzett forgalom *értékeket* értékeltük.

²³Poli(14)_ARMA és Poli(14)_mult

A második kérdés, hogy *hogyan* értékeljük, a hiba mérésének a módjára utal. Eddig egyszerűen az előrejelzett értékeket értékeltük minden további, tehát például az áralakulás figyelembe vétele nélkül. A használt hibamértékek az MSE és a MAPE voltak.

Ebben a fejezetben a fenti két kérdés mentén vizsgáljuk tovább a forgalom előrejelzését.

11.1.1. Az értékelés tárgya

Az egy lépéses előrejelzés mellett kézenfekvő kérdés, hogy hogyan teljesít egy adott modell egynél több lépéses előrejelzés esetén. Noha csak napon belüli forgalom előrejelzéssel foglalkozunk, ez így is igen sok változatot jelenthet²⁴. Az egyszerűség kedvéért tekintsük csak az egész napos előrejelzéseket, tehát amikor a tőzsde nyitása előtt próbáljuk meg előrejelezni a teljes napot, ami esetünkben 26 pontot jelent. Ebben az esetben is az előrejelzett forgalom *értékeket* fogjuk értékelni.

Korábban a napon belüli forgalom előrejelzés elméleti és gyakorlati elhelyezésekor már érintettük a VWAP kereskedés jelentőségét (l. 2.6. alfejezet). Ez a kereskedés típus tehát a tőzsdei forgalom számottevő hányadát teszi ki, és pusztán a forgalom előrejelzés segítségével megvalósítható. Éppen emiatt a szakirodalommal összhangban érdemes kifejezetten a VWAP kereskedés szolgáltatában is értékelni a forgalom előrejelző modelleket. Az egyszerűség kedvéért emeljük ki ismét azt az esetet, ahol az egész nap rendelkezésre áll, tehát a tőzsde nyitása előtt ismert a megbízás, melyet a nap végéig kell teljesíteni, és ezáltal a teljes napra érvényes VWAP lesz a cél²⁵.

A VWAP stratégiához lényegében azt kell előrejeleznünk, hogy a teljes napi forgalom hány százaléka várható az egyes 15 perces időszakokban. Ezek tehát a korábbiakkal ellentétben nem forgalom *értékek*, hanem forgalom *arányok*. Ezen forgalom arányok meghatározásához szükségünk van arra, hogy a teljes nap lefutásáról is legyen valamilyen elképzelésünk. Épp ezért a dolgozat korábbi fejezeteiben

²⁴Attól függően, hogy melyik pontban kezdjük az előrejelzést, és milyen időtávra végezzük el, napi 26 megfigyelés esetén is $\sum_{i=1}^{26} i = 351$ különböző előrejelzés készíthető el minden napra (ebben természetesen már benne van az összes egy lépéses, és az egész napos is).

²⁵Elvileg ugyanaz a 351 lehetőség adott a VWAP kereskedés esetén is, mint fentebb, noha az, amikor csak 15 percünk van, valójában nem releváns, mert akkor be kell adni egyben az egész megbízást, és nincs jelentősége az előrejelzésnek. Ugyanakkor minden további nélkül előfordulhat például az, hogy 12:00 és 15:00 között kell megvenni vagy eladni az adott részvény csomagot.

látott egy lépéses előrejelzés ilyen kereskedésre önmagában nem alkalmas, mert nem ad információt az egész napra várt teljes forgalom értékről. Ugyanakkor a fent említett egész napos előrejelzésre már lehet VWAP stratégiát alapozni, melyet statikus stratégiának is nevezhetünk.

Az egész napos előrejelzésre alapozott VWAP kereskedésnek azonban megvan az a hátránya, hogy nem használja ki teljes mértékben a modellek előnyeit. Ennek oka, hogy a nap elején elkészített egyetlen előrejelzésből dolgozik egész nap, és nem veszi figyelembe a nap során folyamatosan beérkező információkat, amelyeket a modellek egyébként fel tudnának használni. Épp ezért az egész napos mellett érdemes értékelni egy olyan stratégiát is, amely figyelembe tudja venni a folyamatosan frissülő információs halmazt. A folyamatosan frissülő információs halmaz segítségével végzett VWAP kereskedés azonban számtalan különböző stratégia követésével végezhető. Ezek közül az alábbiakban azt emeljük ki, amelyet a benchmarkul szolgáló cikk is használ (Bialkowski et al. (2008)).

Nevezzük tehát dinamikus stratégiának a következőt. Minden lépésben két előrejelzést készítünk. Egyet a szokásos egy lépéses módon a következő 15 percre, egyet pedig a nap hátralévő részére, tehát a nap végéig minden további 15 perces időszakra. Mindkét esetben vegyük figyelembe a rendelkezésre álló legfrissebb adatokat. Mivel forgalom arányokat kívánunk előrejelezni, az FA_t forgalom arány adott t időszakra a következő lesz:

$$FA_t = \frac{\hat{x}_t}{\sum_{l=t}^{26} \hat{x}_l} \quad (110)$$

ahol $t = \{1, 2, 3, \dots, 26\}$, valamint \hat{x}_t a t -edik időszakra előrejelzett forgalom értéket jelöli. A számláló tehát a következő 15 perces időszakra várt forgalom érték, a nevező pedig a nap hátralévő részére előrejelzett teljes forgalom értéke. Mindkettőt lépésenként újra előrejelezzük, vagyis minden lépésben újragondoljuk, hogy milyen megoszlást várunk a nap hátralévő részére, és ez alapján a legközelebbi ponban kereskedünk is. Az így kapott FA_t érték megmutatja, hogy a *hátralévő* kereskedési igényünk mekkora részét kell beadni a következő 15 percben. Értékelni ezek után a beadott ajánlatokat kell, pontosabban ezek arányát a teljes kereskedési igényhez. Ezeket a beadásra felhasznált arányokat jelölje BA_t :

$$BA_t = H_t \cdot FA_t \quad (111)$$

ahol tudjuk, hogy

$$H_t = \begin{cases} 1 & \text{ha } t=1 \\ 1 - \sum_{i=1}^{t-1} BA_i & \text{egyébként} \end{cases} \quad (112)$$

vagyis a H_t hátralévő kereskedési igény 100%-ról indul, és folyamatosan csökken a beadott arányokkal. Ily módon, mivel az $FA_{26} = 1$ mindig teljesül, az utolsó lépésben a maradék kereskedési igényünket fogjuk beadni megbízásként. Ezáltal teljesül az a minden VWAP stratégiával szemben elvárt összefüggés, hogy nap végére a kereskedési igényünk pontosan 100%-ával kereskedjünk, vagyis:

$$\sum_{i=1}^{26} BA_i = 1 \quad (113)$$

Áttekintés. Ebben az alpontban az értékelés tárgyaként három lehetőséget jeleltünk ki. Egyrészt az egész napos előrejelzésből nyert forgalom értékeket, másrészt pedig szintén az egész napos előrejelzés alapján a forgalom arányokat is. Ez utóbbi megközelítést *statikus stratégiaként* is értelmezhetjük. Mivel a statikus stratégiában nem vizsgáljuk felül sem a teljes hátralévő napra, sem a következő 15 percre várt forgalom értékeket, a beadott arányok már a nap kezdetekor a teljes napra meghatározhatóak, és az a feltétel is természetes módon teljesül, hogy az előzetes kereskedési igényünk 100%-át adjuk be megbízásként a nap folyamán. Harmadikként pedig a benchmarkul szolgáló cikket követve bemutattunk egy lehetséges *dinamikus stratégiát*, amely a beadott forgalom arányokat határozza meg.

11.1.2. Az értékelés módja

Miután kijelöltük, hogy az egyes modellek kapcsán mit szeretnénk értékelni, a következő kérdés az, hogy hogyan is tegyük ezt meg.

Az egész napos forgalom előrejelzésből nyert forgalom értékeket ugyanúgy értékelhetjük MSE és MAPE alapon, ahogy a dolgozatban korábban is tettük a forgalom értékekkel. Mivel ezúttal nem frissül lépésenként az információs halmaz, várhatóan rosszabb eredményt fogunk így kapni, mint a korábbi fejezetekben, ahol frissült.

A következő lépés a forgalom arányok értékelése statikus és dinamikus stratégia használata esetén. Mivel ezt kifejezetten a VWAP kereskedés apropóján vizsgáljuk, és a stratégiákat is ez alapján határoztuk meg, a legcélszerűbb mérőszám az lehet, hogy milyen jól sikerült eltalálni a napi VWAP értéket. Az értékeléskor ezért a tényleges (tehát az adott nap végén a piacon meghatározható) VWAP, és a kereskedés során elért forgalommal súlyozott átlagos ár (mintegy *saját* VWAP) eltérését fogjuk vizsgálni, ahol árnak minden 15 perces időszakra az abban az időszakban elért átlagos árat tekintjük. Fontos megjegyezni, hogy az eltérés iránya továbbra sem számít, hiszen eladás illetve vétel esetén ellentétes irányú hiba lenne kedvező (eladni drágábban, venni olcsóbban jobb), a stratégiákat azonban mindkettőre szeretnénk alkalmazni. Cél tehát a kereskedés eredményeként minél pontosabban eltalálni a VWAP-ot, ezért az ettől való eltérést próbáljuk meg mérni előjeltől függetlenül.

Ezen eltérések mérésére pedig adódnak a korábban is használt MSE és MAPE hibamértékek. Vegyük azonban észre, hogy az MSE ezúttal nem lesz használható. Ennek oka elsősorban az adatbázis által lefedett időszak meglehetősen hosszú volta, ugyanis ezen időszak alatt egy-egy részvény értéke igen nagy mértékben ingadozott. Ezen felül az egyes részvények átlagos árszintje is jelentős eltéréseket mutat. A mellékletben a 109. táblázatban látható, hogy egy adott részvényre a teljes időszak alatt felvett maximális és minimális ár aránya minden esetben legalább 2 volt, az American International Goup Inc. esetén pedig elérte 244,9-es értéket is. Ugyanezen részvény átlagos árszintje 46,1-szeres a legkisebb átlagos árszintű részvényhez képest.

Az átlagos árak közti eltérésből eredő torzítás rokon a korábban a forgalommal kapcsolatban tett hasonló megállapítással, amelynek nyomán bevezettük a korrigált MSE hibamértéket²⁶, hogy a részvények eredményeinek átlagolásakor csökkentjük a torzítást. Az egy-egy részvényen belüli áringadozások azonban már egy részvényre nézve is jelentős torzítást okoznak az MSE használata esetén.

A 75. táblázat segítségével lássunk egy példát arra, hogy hogyan jelentkezik az ár bevonása okozta torzítás a VWAP stratégia MSE alapú értékelésében. Az átláthatóság kedvéért tegyük fel, hogy egy adott részvény két különböző napját kívánjuk összevontan értékelni, és minden napot csak három időszakra osztunk. Mindkét napon pontosan ugyanúgy alakultak a forgalom arányok, és pontosan

²⁶MSE*, l. (109).

| Ár | Forgalom arány | |
|------|----------------|-----------|
| | Tény | Stratégia |
| 100 | 0,5 | 0,6 |
| 105 | 0,2 | 0,3 |
| 110 | 0,3 | 0,1 |
| VWAP | | |
| | Tény | Stratégia |
| | 104 | 102,5 |
| MSE | 2,25 | |
| MAPE | 1,44% | |

| Ár | Forgalom arány | |
|------|----------------|-----------|
| | Tény | Stratégia |
| 1000 | 0,5 | 0,6 |
| 1050 | 0,2 | 0,3 |
| 1100 | 0,3 | 0,1 |
| VWAP | | |
| | Tény | Stratégia |
| | 1040 | 1025 |
| MSE | 225 | |
| MAPE | 1,44% | |

75. táblázat. VWAP értékelések MSE és MAPE alapon különböző árszintek mellett
Forrás: Saját szerkesztés

ugyanazt az előrejelzést is adtuk mindkét napra. Az egyetlen eltérés, hogy az egyik esetben az ár éppen 10-szerese volt a másik esethez képest. Ekkor a VWAP is 10-szeres lesz, ahogyan a stratégiával elért átlagos ár is. Egy ilyen helyzetben a MAPE értékek megegyeznek, az MSE érték azonban 100-szoros lesz a magasabb árú esetben. Ha tehát ezt a két napot szeretnénk átlagolni, a 10-szeres árú nap 100-szoros súllyal számítana az MSE alkalmazása mellett, tehát dominálná az átlagot, miközben mindkét nap egyforma fontos számunkra.

Ezen túlmenően azonban a fenti két eset nem csak egyforma fontos, de a VWAP kereskedés szempontjából pontosan egyforma jó is. Gondoljuk tovább az előző példát a 76. táblázat segítségével. A kereskedési igényünk nap elején rögzített, legyen egy adott részvény vétele 1000 értékben (ennyi pénzt fektetünk be). Mivel a tény és a stratégia által elért átlagos ár eltér, ezért különböző lesz a két esetben a megvásárolt darabszám is. Jelen példában alacsonyabb árat értünk el a VWAP-nál, és így több részvényt tudunk megvenni ugyanakkora befektetéssel. A többletet záróáron értékelve megkapjuk, hogy nap végére pénzben kifejezve hogyan érintett minket a VWAP-tól való eltérés. Láthatóan pontosan ugyanannyit nyertünk mindkét esetben. (Eladás esetén ugyanez veszteség lenne, hiszen ugyanannyi pénzért több részvénytől kellett volna megválnunk.)

Úgy célszerű tehát mérnünk a hibát, hogy pusztán az árszintek eltérése miatt a hibamértékek ne változzanak. A fenti példa is érzékelteti, hogy az MSE nem

| | | | |
|-------------------------|--------|-------------------------|--------|
| Elköltendő tőke | 1000 | Elköltendő tőke | 1000 |
| Tény (darab): | 9,615 | Tény (darab): | 0,9615 |
| Stratégia (darab): | 9,756 | Stratégia (darab): | 0,9756 |
| Eltérés (darab): | 0,141 | Eltérés (darab): | 0,0141 |
| Eltérés értéke záróáron | 15,478 | Eltérés értéke záróáron | 15,478 |

76. táblázat. VWAP kereskedés szempontjából a két eset pontosan egyforma
Forrás: Saját szerkesztés

felel meg ennek a kritériumnak, a MAPE viszont a százalékos logikája miatt már igen.

Áttekintés. Az értékelés tárgyának meghatározása után ebben az alponban az értékelés módjáról volt szó. Az egész napos forgalom előrejelzésből meghatározott forgalom értékeket értékelhetjük a már megszokott módon MSE és MAPE alapon. A statikus és dinamikus stratégiák által megadott forgalom arányokat pedig a VWAP és az adott stratégia segítségével elért átlagos ár eltérése alapján értékeljük MAPE mutatóval.

11.2. Egész napos előrejelzés

Hasonlítsuk össze a modellek teljesítményét egész napos előrejelzés esetén forgalom értékekre, és forgalom arányokra (vagyis a statikus stratégiára) is. Az értékelésbe bevont modellek legyenek a következők.

Alapvetően a benchmark a BDF modell, melynek két változata a BDF_AR és a BDF_SETAR, ezeket tehát mindenképpen meg kell vizsgálnunk. A 10. fejezetben a saját modellek közül a Poli(14)_ARMA és a Poli(14)_mult bizonyult a két legjobbnak, ezért ezekre is szükség lesz. Mindezek mellett érdemesnek látom megvizsgálni az U-módszer teljesítményét is, hiszen Bialkowski et al. (2008) eredménye ennek a megverése volt. Emiatt érdekes lehet ellenőrizni, hogy a korábbiakhoz hasonlóan az ebben a fejezetben alkalmazott teljesítmény értékelés alapján is meg tudjuk-e erősíteni ezt az eredményt a saját adatainkon. Végül pedig, mint látni fogjuk, az utóbbi mellett néhány esetben érdekes lesz az U-dekompozíciót egyedi rész nélkül alkalmazó egyéb modellek (l. 10.4.1. alpont) vizsgálata is.

11.2.1. Forgalom értékek

Kezdjük tehát az egész napos előrejelzéssel, azon belül is a forgalom értékek előrejelzésének értékelésével.

Elsőként vizsgáljuk meg a Poli(14)_ARMA modellt, melyben a 77. táblázat lesz segítségünkre²⁷. Láthatóan MSE alapon csak a Poli(14)_mult jobb egyértelműen a részvények többségére, illetve a Poli_expw(7)_U egy hajszállal. MAPE alapon ez utóbbi előnye el is tűnik, és csak a Poli(14)_mult veri meg a részvények többségénél a vizsgált modell változatot. MAPE alapon átlagban is ez a két modell jobb csak a Poli(14)_ARMA modellnél. A korrigált MSE átlag azonban a BDF modellektől elektve az összes többi esetben alacsonyabb lett, mint a Poli(14)_ARMA értéke, melyet két részvény előrejelzése során vétett jelentősebb hiba okoz.

A siker tehát nem teljes, az azonban egyértelműen látszik, hogy a vizsgált Poli(14)_ARMA minden szempont szerint lényegesen jobb bármelyik BDF változatnál. Különösen igaz ez a BDF_SETAR modellre, melynek átlagos hibája mindkét mérték szerint gyakorlatilag értelmezhetetlenül nagy. Ez legnagyobb részben öt részvénynek köszönhető, de ezek elhagyásával is határozottan rosszabb lenne a BDF_SETAR a vizsgált modellünknél, csak a nagyságrendi különbségek nem lennének ennyire zavaróan nagyok.

Bialkowski et al. (2008) védelmében el kell mondani, hogy ők is tisztában vannak modelljük gyenge teljesítményével egész napos előrejelzés esetén. Épp ezért ilyen előrejelzésre az egyedi rész nélküli, tehát jelölésünkben a BDF_U modell változatot ajánlják, amely azonban a korrigált átlagos MSE kivételével szintén mindenben rosszabb a Poli(14)_ARMA modellnél.

Következőként térjünk rá a Poli(14)_mult modellre. A 78. táblázat alapján láthatjuk, hogy az interpretáció tekintetében most sokkal könnyebb helyzetben vagyunk, hiszen ez a modell minden szempont szerint egyértelműen a legjobb mind közül.

Végül vizsgáljuk meg, hogy Bialkowski et al. (2008) eredményét, tehát az U-módszer megverését alá tudjuk-e támasztani az itt vizsgált szempont szerint a saját adatainkon. Ebben segít a 79. táblázat. A BDF_AR és BDF_SETAR modellek az U-módszernél is határozottan gyengébben teljesítenek. A BDF_U

²⁷Az MSE átlagok tekintetében már csak a korrigált értékeket tüntetem fel.

| Egész napos előrejelzés | Hány részvény- nél nyert | | Átlagos érték | Korrigált átlagos érték |
|----------------------------|-----------------------------|------|------------------|----------------------------|
| | MSE | MAPE | MAPE | MSE* |
| BDF_AR | 6 | 0 | 67,0% | 4,53E+00 |
| BDF_SETAR | 2 | 0 | 6 887 517,7% | 4,56E+12 |
| Poli(14)_mult | 29 | 31 | 47,1% | 1,52E-04 |
| U-módszer | 7 | 2 | 50,3% | 1,76E-04 |
| BDF_U | 7 | 14 | 48,9% | 1,83E-04 |
| Poli(14)_U | 7 | 2 | 50,1% | 1,75E-04 |
| Poli_expw(7)_U | 17 | 15 | 47,7% | 1,61E-04 |
| Spline(N4K6)_U | 7 | 2 | 50,2% | 1,75E-04 |
| Poli(14)_ARMA | - | - | 48,2% | 3,83E-03 |

77. táblázat. A Poli(14)_ARMA modell értékelése
egész napos előrejelzés esetén
Forrás: Saját szerkesztés

| Egész napos előrejelzés | Hány részvény- nél nyert | | Átlagos érték | Korrigált átlagos érték |
|----------------------------|-----------------------------|------|------------------|----------------------------|
| | MSE | MAPE | MAPE | MSE* |
| BDF_AR | 2 | 0 | 67,0% | 4,53E+00 |
| BDF_SETAR | 2 | 0 | 6 887 517,7% | 4,56E+12 |
| Poli(14)_ARMA | 4 | 2 | 48,2% | 3,83E-03 |
| U-módszer | 3 | 0 | 50,3% | 1,76E-04 |
| BDF_U | 2 | 10 | 48,9% | 1,83E-04 |
| Poli(14)_U | 3 | 0 | 50,1% | 1,75E-04 |
| Poli_expw(7)_U | 3 | 4 | 47,7% | 1,61E-04 |
| Spline(N4K6)_U | 3 | 0 | 50,2% | 1,75E-04 |
| Poli(14)_mult | - | - | 47,1% | 1,52E-04 |

78. táblázat. A Poli(14)_mult modell értékelése
egész napos előrejelzés esetén
Forrás: Saját szerkesztés

| Egész napos előrejelzés | Hány részvény- nél nyert | | Átlagos érték | Korrigált átlagos érték |
|----------------------------|-----------------------------|------|------------------|----------------------------|
| | MSE | MAPE | MAPE | MSE* |
| BDF_AR | 12 | 2 | 67,0% | 4,53E+00 |
| BDF_SETAR | 3 | 0 | 6 887 517,7% | 4,56E+12 |
| Poli(14)_mult | 30 | 33 | 47,1% | 1,52E-04 |
| Poli(14)_ARMA | 26 | 31 | 48,2% | 3,83E-03 |
| Poli(14)_U | 31 | 33 | 50,1% | 1,75E-04 |
| BDF_U | 0 | 30 | 48,9% | 1,83E-04 |
| Poli_expw(7)_U | 33 | 33 | 47,7% | 1,61E-04 |
| Spline(N4K6)_U | 33 | 29 | 50,2% | 1,75E-04 |
| U-módszer | - | - | 50,3% | 1,76E-04 |

79. táblázat. Az U-módszer értékelése
egész napos előrejelzés esetén

Forrás: Saját szerkesztés

az MSE alapján szintén határozottan alul marad, csak a MAPE mérték szerint jobb az U-módszernél²⁸. Az általam javasolt modellek (a táblázatban összesen öt) ezzel szemben minden szempont szerint jobbak az U-módszernél, egyetlen kivétel a Poli(14)_ARMA átlagos korrigált MSE értéke.

11.2.2. Statikus stratégia

Folytassuk a vizsgálatot a statikus stratégiával, amely szintén az egész napos előrejelzésen alapul, de abból forgalom arányokat készít. A korábban elmondottaknak megfelelően ezt a VWAP kereskedés sikeressége szempontjából fogjuk értékelni. Előzetesen elmondható, hogy ez egy olyan stratégia, amelyet valószínűleg nem érdemes alkalmazni, hiszen nem használja fel a nap során beérkező információkat, hanem nap elején minden ajánlat beadást előre rögzít. Nehéz is elképzelni, hogy egy piaci szereplő teljesen figyelmen kívül hagyja a kereskedés közben megfigyelt friss adatokat.

Ennek ellenére azonban érdekes lehet látni az eredményét, ezért elsőként vizsgáljuk meg a Poli(14)_ARMA modellt ilyen szempontból a 80. táblázat segítségével. A látottakat röviden úgy lehet összefoglalni, hogy a BDF változatok (a BDF_U verzióval együtt már három különböző) mindkét szempont szerint rosszabbul szerepelnek, minden más modell azonban jobban.

²⁸Bialkowski et al. (2008) csak a MAPE mértéket használta, az MSE-t nem.

| VWAP statikus stratégiával | Hány részvénynél nyert | Átlagos érték |
|-------------------------------|------------------------|---------------|
| | MAPE | MAPE |
| BDF_AR | 3 | 7,12E-04 |
| BDF_SETAR | 1 | 7,13E-04 |
| Poli(14)_mult | 32 | 6,13E-04 |
| U-módszer | 32 | 6,08E-04 |
| BDF_U | 6 | 6,76E-04 |
| Poli(14)_U | 32 | 6,08E-04 |
| Poli_expw(7)_U | 29 | 6,15E-04 |
| Spline(N4K6)_U | 32 | 6,08E-04 |
| Poli(14)_ARMA | - | 6,26E-04 |

80. táblázat. A Poli(14)_ARMA modell értékelése
statikus stratégia használata esetén
Forrás: Saját szerkesztés

Ezután a 81. táblázatot elemezve térjünk rá a Poli(14)_mult modellre. Itt már jobb a helyzet, de az U-módszer, az egyszerű polinom dekompozíció, és a spline dekompozíció ennél a változatnál is jobban teljesít. Tehát a statikus stratégia használata esetén az U-módszer nem csak a BDF változatokat verte meg (ez az eddigiekből is kiolvasható, ezért erről külön táblázatot nem készítek), hanem az az általam javasolt két modellt is.

A 82. táblázatból látható, hogy statikus stratégiára mind közül a Poli(14)_U modell a legjobb, tehát az általam használt modellek U alakja egyedi rész nélkül. Ugyanakkor meg kell jegyezni, hogy az U-módszer és a spline dekompozíció a részvények alig kevesebb, mint felénél nyert, és az átlagos MAPE értékekről is megállapítható, hogy lényegében megegyeznek. Tehát ez a három modell a statikus stratégia szempontjából gyakorlatilag egyforma jónak tekinthető.

11.3. Dinamikus stratégia

Végül térjünk rá a dinamikus stratégiára, amely alapján már valóban kiválasztható a VWAP kereskedés szempontjából legjobb modell. Az értékelésbe ugyanazon modelleket vonjuk be, mint az előző alfejezetben az egész napos előrejelzés esetén. Mivel a modellekben az U alak változatlan volt a nap során, csak az egyedi rész frissült a beérkező információ felhasználásával, ezért a dinamikus stratégia csak az egyedi részt is tartalmazó modellekre értelmezhető. Az egyedi részt nem

| VWAP statikus stratégiával | Hány részvénynél nyert | Átlagos érték |
|-------------------------------|------------------------|---------------|
| | MAPE | MAPE |
| BDF_AR | 2 | 7,12E-04 |
| BDF_SETAR | 1 | 7,13E-04 |
| Poli(14)_ARMA | 1 | 6,26E-04 |
| U-módszer | 27 | 6,08E-04 |
| BDF_U | 4 | 6,76E-04 |
| Poli(14)_U | 28 | 6,08E-04 |
| Poli_expw(7)_U | 15 | 6,15E-04 |
| Spline(N4K6)_U | 27 | 6,08E-04 |
| Poli(14)_mult | - | 6,13E-04 |

81. táblázat. A Poli(14)_mult modell értékelése
statikus stratégia használata esetén
Forrás: Saját szerkesztés

| VWAP statikus stratégiával | Hány részvénynél nyert | Átlagos érték |
|-------------------------------|------------------------|---------------|
| | MAPE | MAPE |
| BDF_AR | 1 | 7,12E-04 |
| BDF_SETAR | 1 | 7,13E-04 |
| Poli(14)_mult | 5 | 6,13E-04 |
| Poli(14)_ARMA | 1 | 6,26E-04 |
| U-módszer | 16 | 6,08E-04 |
| BDF_U | 4 | 6,76E-04 |
| Poli_expw(7)_U | 2 | 6,15E-04 |
| Spline(N4K6)_U | 14 | 6,08E-04 |
| Poli(14)_U | - | 6,08E-04 |

82. táblázat. A Poli(14)_U modell értékelése
statikus stratégia használata esetén
Forrás: Saját szerkesztés

| VWAP dinamikus stratégiával | Hány részvényenél nyert | Átlagos érték |
|--------------------------------|-------------------------|---------------|
| | MAPE | MAPE |
| BDF_AR | 0 | 6,44E-04 |
| BDF_SETAR | 0 | 9,18E-04 |
| Poli(14)_mult | 26 | 5,67E-04 |
| U-módszer | 1 | 6,08E-04 |
| BDF_U | 3 | 6,76E-04 |
| Poli(14)_U | 1 | 6,08E-04 |
| Poli_expw(7)_U | 1 | 6,15E-04 |
| Spline(N4K6)_U | 1 | 6,08E-04 |
| Poli(14)_ARMA | - | 5,76E-04 |

83. táblázat. A Poli(14)_ARMA modell értékelése
dinamikus stratégia használata esetén
Forrás: Saját szerkesztés

tartalmazó modellek esetén pedig megtartjuk a statikus stratégia eredményét, és ezt használjuk fel az összehasonlításra.

Kezdjük ismét a Poli(14)_ARMA modellel a vizsgálatot. A 83. táblázat alapján megállapítható, hogy a Poli(14)_mult modell kivételével minden másnál egyértelműen jobb ez a változat. Ennek megfelelően mind közül a Poli(14)_mult modell teljesít legjobban, amint az a 84. táblázatból is látszik.

Zárásként, mivel ez az értékelési mód az, amelyet Bialkowski et al. (2008) döntőnek tekint a maga szempontjából, vizsgáljuk meg, hogy az U-módszer hogyan viszonyul a BDF változatokhoz. A 85. táblázat alapján a saját adatainkon nem tudjuk megerősíteni, hogy a BDF modell három változata közül (vagyis a BDF_U verziót is beleértve) bármelyik jobb lenne az egyszerű U-módszernél akár a részvények többsége tekintetében, akár pedig az átlagos hibát figyelve.

11.4. Áttekintés

Míg a dolgozat korábbi fejezeteiben a modellek egy lépésre előrejelzett forgalom értékeit értékeltük, a 11. fejezetben az értékelés egyéb lehetőségeit vizsgáltuk a korábban azonosított modellekre alkalmazva. Az elemzésbe bevont modellek a benchmarknak tekintett BDF modell egyes változatai, a Poli(14)_ARMA és a Poli(14)_mult modellek, valamint az U dekompozíciót egyedi rész nélkül alkalmazó modellek, köztük a gyakorlatban is elterjedt U-módszerrel, amely egyúttal

| VWAP dinamikus stratégiával | Hány részvénynél nyert | Átlagos érték |
|--------------------------------|------------------------|---------------|
| | MAPE | MAPE |
| BDF_AR | 0 | 6,44E-04 |
| BDF_SETAR | 0 | 9,18E-04 |
| Poli(14)_ARMA | 7 | 5,76E-04 |
| U-módszer | 1 | 6,08E-04 |
| BDF_U | 2 | 6,76E-04 |
| Poli(14)_U | 1 | 6,08E-04 |
| Poli_expw(7)_U | 0 | 6,15E-04 |
| Spline(N4K6)_U | 1 | 6,08E-04 |
| Poli(14)_mult | - | 5,67E-04 |

84. táblázat. A Poli(14)_mult modell értékelése
dinamikus stratégia használata esetén
Forrás: Saját szerkesztés

| VWAP dinamikus stratégiával | Hány részvénynél nyert | Átlagos érték |
|--------------------------------|------------------------|---------------|
| | MAPE | MAPE |
| BDF_AR | 9 | 6,44E-04 |
| BDF_SETAR | 0 | 9,18E-04 |
| Poli(14)_mult | 32 | 5,67E-04 |
| Poli(14)_ARMA | 32 | 5,76E-04 |
| BDF_U | 4 | 6,76E-04 |
| Poli(14)_U | 17 | 6,08E-04 |
| Poli_expw(7)_U | 1 | 6,15E-04 |
| Spline(N4K6)_U | 16 | 6,08E-04 |
| U-módszer | - | 6,08E-04 |

85. táblázat. Az U-módszer értékelése
dinamikus stratégia használata esetén
Forrás: Saját szerkesztés

| Értékelési szempont | Legjobb modell | Melyik BDF változat jobb az U-módszernél |
|-------------------------|----------------|--|
| Egész napos előrejelzés | Poli(14)_mult | BDF_U (csak MAPE szerint) |
| Statikus stratégia | Poli(14)_U | Egyik sem |
| Dinamikus stratégia | Poli(14)_mult | Egyik sem |

86. táblázat. Az egyéb értékelési szempontok eredményeinek áttekintése
Forrás: Saját szerkesztés

Bialkowski et al. (2008) benchmarkja is volt.

Az értékelés egyéb lehetőségeiként három esetet emeltünk ki. Egyrészt az egész napos előrejelzést, melynek vizsgálata kézenfekvően adódik az egy lépéses előrejelzés után. Az egész napra előrejelzett forgalom értékeket ugyanúgy értékelhetjük, mint az egy lépéses előrejelzéseket. Másrészt, mivel a napon belüli forgalom előrejelzés talán legközvetlenebbül megragadható alkalmazási területe a VWAP kereskedés, a szakirodalommal összhangban érdemes ennek fényében is tesztelni a modelleket. Ezt megtehetjük a nap során érkező információ felhasználása nélkül, mely elvi lehetőségre statikus stratégiaként hivatkoztunk. Emellett azonban megtehetjük úgy is, hogy felhasználjuk a nap folyamán fokozatosan megismert adatokat is. Ez utóbbi módon végrehajtott módszerre, amely nyilvánvalóan közelebb áll a realitáshoz, dinamikus stratégiaként utaltunk. A stratégiák forgalom arányokat határoznak meg, értékelésük pedig az alapján történik, hogy a stratégia által meghatározott módon végezve a kereskedést mennyire sikerült megközelíteni a piacon megfigyelhető VWAP értéket.

A benchmarknak tekintett Bialkowski et al. (2008) is vizsgálja a teljesítményt mind statikus, mind dinamikus stratégiára, de a statikusát csak a teljesség kedvéért értékeli, valójában a dinamikusát tekinti relevánsnak.

Az értékelések elvégzése után a 86. táblázat segít áttekinteni a fejezet főbb megállapításait. Az egész napos előrejelzés terén a Poli(14)_mult modell teljesített legjobban. Ugyanez elmondható a dinamikus stratégiára is, amely a VWAP kereskedés szempontjából a legfontosabbnak tekinthető. A statikus stratégiára a Poli(14)_U modell volt a legjobb, amely egyúttal a Poli(14)_mult modell U-alak komponense is, egyedi részt azonban nem tartalmaz. A statikus stratégia értékelésének azonban, mint láttuk, igen korlátozott jelentősége van csak.

Végül, habár a fejezetben a fő célunk nem ez volt, ellenőriztük azt is, hogy Bialkowski et al. (2008) eredményeit meg tudjuk-e erősíteni a saját adatainkon is,

vagyis, hogy az itt vizsgált szempontok szerint valóban jobb-e a BDF modell az U-módszernél. Amint a táblázat is mutatja, sem statikus, sem dinamikus stratégiára nem tudtuk ezt megtenni, ugyanis az U-módszer a BDF modell minden változatánál jobbnak mutatkozott ezen az adatbázison mindkét stratégiára. Egyedül az egész napos előrejelzés esetén találtuk a BDF_U modellt jobbnak, mint az U-módszert, de ezt is csak az egyik hibamérték, a MAPE szerint.

12. fejezet

Eredmények áttekintése

A IV. rész utolsó fejezetéhez érve tekintsük át ennek a résznek az eredményeit.

Miután korábban már azonosítottuk a benchmarkot, a 10. fejezet a modell keresésről szólt. Ebben a fejezetben egyszerűen az egy lépéses előrejelzéseket értékeltem MSE és MAPE alapon annak érdekében, hogy olyan modellt találjak, amelyet érdemes később egyéb hibamértékek alapján is tesztelni. Elsőként U dekompozíció nélküli modelleket vizsgáltam, ami alapján úgy tűnt, egy jól teljesítő modellben az U alak stilizált ténye nem hagyható figyelmen kívül.

A következő lépésben ezért az U alak dekompozíciójára kerestem különböző, a szakirodalomban nem szereplő új lehetőségeket. Itt meglepetésemre már az egyszerű U-módszer is jobban szerepelt a benchmarknál, amennyiben a szokásostól eltérően egyedi rész használata mellett alkalmaztam. Mivel azonban az U-módszer (a BDF dekompozícióhoz hasonlóan) igen zajos U alakot eredményez, felmerült, hogy érdemes lenne inkább egy simított U dekompozícióval dolgozni. Az erre vonatkozó ötletem alapvetően polinom illesztése valamilyen formában. Ezt a lehetőséget megvizsgáltam egyenlő és exponenciális súlyozás mellett, valamint spline függvény használatán keresztül is, és mindegyik esetben határozottan jobb eredményt kaptam a benchmarknál, illetve a fent említett U-módszer egyedi résszel történő alkalmazásánál is.

Az U alak additív módon történő leválasztása után rátértem az egyedi rész előrejelzési lehetőségeinek vizsgálatára. Az ármozgás mutatók egyszerű késleltettségének bevonása nem segített, az egyidejű ármozgás mutatók felhasználásának gondolat kísérlete alapján azonban érdemesnek tűnt különböző ármozgás előrejelzések modellbe foglalásával próbálkozni. Az adott terület fejlettségéhez képest

viszonylag egyszerűnek számító modellekkel nem sikerült olyan ármozgás előrejelzést készítenem, amelyet felhasználva érdemben javulna a forgalom előrejelzése. Elképzelhető ugyan, hogy komplexebb ármozgás előrejelző modellekkel jobb eredmény is elérhető, ugyanakkor véleményem szerint az egyidejű ármozgás használata éppen azért javított látványosan a forgalom előrejelzésen, mert megjelennek benne a váratlan (vagyis sokk jellegű), alap esetben előrejelezhetetlen hatások, amelyeket azonban vélhetően semmilyen előrejelző modell nem fog tudni jól megragadni.

Ezek után, az ármozgás mutatókat hátrahagyva teszteltem még néhány modellt az egyedi rész előrejelzésére, melyek közül az egyszerű ARMA(1,1) bizonyult a legjobbnak.

Végül megvizsgáltam néhány egyéb, a korábbi logikától némileg eltérő lehetőséget is a forgalom előrejelzésére. Ezek között szerepelt az egyedi rész figyelmen kívül hagyása, valamint egy korrekciós modell is, de ezek nem teljesítettek kellően jól. A harmadik vizsgált modell azonban már igen. Ez a modell az eddigi legjobb dekompozíciót és egyedi rész előrejelzést használja fel, de a korábbi additív helyett multiplikatív logikával teszi ezt. Mivel egyetlen részvény, nevezetesen az Altria Group Inc. miatt nem volt teljesen egyértelmű, hogy a multiplikatív modell az eddigi legjobb, az additív megfelelőjével együtt mindkettőt megtartottam a későbbi vizsgálatokra.

A 11. fejezetben a fent említett két modellt egyéb hibamérési lehetőségek mellett is megvizsgáltam. Értékeltem egyrészt az egész napos forgalom értékek előrejelzését az eddigiekhez hasonló módszerrel. Másrészt, mivel a forgalom előrejelzés egy jelentős felhasználási területe a VWAP kereskedés, ilyen szempontból is megvizsgáltam a modelleket. Ehhez a szakirodalommal összhangban napon belüli forgalom arányokat jeleztem előre statikus és dinamikus stratégiával, és ezen előrejelzéseket értékeltem aszerint, hogy mennyire sikerülne eltalálni a VWAP-ot abban az esetben, ha ezek alapján kereskednénk. Az eredmények alapján az egész napos előrejelzés, illetve a dinamikus stratégiával történő kereskedés esetén egyaránt a multiplikatív modell volt a legjobb. A statikus stratégiának a szakirodalom szerint is csak korlátozott jelentősége van, hiszen egyértelmű, hogy nem érdemes ez alapján kereskedni a gyakorlatban. Ennek ellenére erre is elvégeztem az értékelést, ami alapján az egyedi rész nélkül alkalmazott, egyenlő súlyokat használó polinom dekompozíció szerepelt legjobban (amely egyébként a multiplikatív modell U dekompozíciós eljárása is egyben).

| Egy lépéses előrejelzés | Hány részvény- nél nyert | | Átlagos érték | Korrigált átlagos érték | Elért átlagos javítás | |
|----------------------------|-----------------------------|-----|------------------|----------------------------|--------------------------|------|
| | MAPE | MSE | MAPE | MSE* | MAPE | MSE* |
| BDF_AR | 0 | 1 | 40,3% | 1,09E-04 | 11,6% | 3,8% |
| BDF_SETAR | 1 | 1 | 39,9% | 1,12E-04 | 10,5% | 6,7% |
| Poli(14)_mult | - | - | 36,1% | 1,05E-04 | - | - |

87. táblázat. A Poli(14)_mult és BDF_(SET)AR modellek összevetése egy lépéses előrejelzésre a különböző hibamértékek szerint

Forrás: Saját szerkesztés

| Egész napos előrejelzés | Hány részvény- nél nyert | | Átlagos érték | Korrigált átlagos érték | Elért átlagos javítás | |
|----------------------------|-----------------------------|-----|------------------|----------------------------|--------------------------|----------|
| | MAPE | MSE | MAPE | MSE* | MAPE | MSE* |
| BDF_AR | 0 | 2 | 67,0% | 4,53E+00 | 42,3% | 2,98E+04 |
| BDF_SETAR | 0 | 2 | 6,89E+04 | 4,56E+12 | 1,46E+05 | 3,00E+16 |
| Poli(14)_mult | - | - | 47,1% | 1,52E-04 | - | - |

88. táblázat. A Poli(14)_mult és BDF_(SET)AR modellek összevetése egész napos előrejelzésre a különböző hibamértékek szerint

Forrás: Saját szerkesztés

Végül érdekességképpen megvizsgáltam, hogy az általam használt adatbázison meg tudom-e erősíteni, hogy a benchmarknak tekintett BDF modell az itt vizsgált hibamérés szerint is jobb az U-módszernél. Az egész napos előrejelzés esetén azt találtam, hogy a BDF dekompozíció egyedi rész nélkül jobb, mint az U módszer, de csak MAPE alapon, MSE alapon már nem az (egyedi résszel együtt pedig minden hibamérték szerint rosszabb). A statikus és dinamikus stratégia esetén az U-módszer bármelyik BDF változatnál jobban teljesített. Mindez (különösen a dinamikus stratégia miatt) ellentmond Bialkowski et al. (2008) megállapításainak, miszerint az ő modelljüket érdemes VWAP kereskedésre használni az egyszerű U-módszer helyett.

Összegzésként tehát megállapítható, hogy a IV. részben a modell keresés eredményeképpen sikerült a szakirodalom legjobbjaként azonosított BDF modellnél jobb modellt találnom, hiszen a Poli(14)_mult modell a szakirodalom alapján meghatározott hibamérési lehetőségek mindegyike alapján jobban teljesít, mint a BDF benchmark. Ennek áttekintését találjuk a 87-90. táblázatokban.

A 87. táblázatban azt az esetet vizsgálhatjuk meg, amely az egy lépéses elő-

| Statikus stratégia | Hány részvény-nél nyert | Átlagos érték | Elért átlagos javítás |
|--------------------|-------------------------|---------------|-----------------------|
| | MAPE | MAPE | MAPE |
| BDF_AR | 2 | 7,12E-04 | 16,2% |
| BDF_SETAR | 1 | 7,13E-04 | 16,3% |
| Poli(14)_mult | - | 6,13E-04 | - |

89. táblázat. A Poli(14)_mult és BDF_(SET)AR modellek összevetése statikus stratégiára a különböző hibamértékek szerint
Forrás: Saját szerkesztés

| Dinamikus stratégia | Hány részvény-nél nyert | Átlagos érték | Elért átlagos javítás |
|---------------------|-------------------------|---------------|-----------------------|
| | MAPE | MAPE | MAPE |
| BDF_AR | 0 | 6,44E-04 | 13,6% |
| BDF_SETAR | 0 | 9,18E-04 | 61,9% |
| Poli(14)_mult | - | 5,67E-04 | - |

90. táblázat. A Poli(14)_mult és BDF_(SET)AR modellek összevetése dinamikus stratégiára a különböző hibamértékek szerint
Forrás: Saját szerkesztés

rejelzéseket értékeli. A BDF változatok MSE és MAPE alapon is maximum egy részvényre adtak pontosabb előrejelzést (a vizsgált 33-ból) a Poli(14)_mult modellnél, és az átlagos hiba értékeik is magasabbak. A benchmarkhoz képest elért javítást értelmezzük az alábbi módon:

$$Javitás(Benchmark, Saját) = \frac{Hiba(Benchmark)}{Hiba(Saját)} - 1 \quad (114)$$

Ez tehát egy hozamhoz hasonlatos mutató, amely azt fejezi ki százalékos formában, hogy milyen mértékben sikerült javítani a benchmark hibáján. Értelmeszerűen a kisebb hiba jobb, ezért a javítás pozitív értéke kedvező, negatív értéke pedig romlást jelez. Ezen mutató alapján a Poli(14)_mult MAPE alapon 10,5% illetve 11,6% javulást hozott, korrigált MSE (vagyis MSE*) alapon pedig 3,8% és 6,7% javulást eredményezett az egyes BDF változatokhoz képest.

A 88. táblázatban az egész napos előrejelzések összevetését láthatjuk. A BDF modellek egy részvényre sem jobbak MAPE alapon, és két részvényre jobbak MSE alapon a 33-ból a Poli(14)_mult modellnél. A BDF modellek hiba értékei rendkívül rosszak egész napos előrejelzés esetére. A BDF_AR változathoz képest

a Poli(14)_mult 42,3%-os javulást eredményez, az összes többi esetben pedig ennél is lényegesen nagyobb, *sok nagyságrendnyi* javítást figyelhetünk meg az átlagos hiba értékekben.

A 89. táblázat a statikus stratégiák összevetését mutatja. A BDF változatok 33 részvényből mindössze 1 illetve 2 esetben szerepeltek jobban a Poli(14)_mult modellnél (mint korábban bemutattam, a VWAP alapú értékelések esetén csak a MAPE mérték megfelelő, mivel az MSE torzít). A Poli(14)_mult modellre való áttérés 16,2% illetve 16,3% javulást eredményez.

Végül a dinamikus stratégiákat a 90. táblázat segítségével hasonlíthatjuk össze. A 33 részvényből egy esetben sem tudtak jobbak lenni a BDF változatok a Poli(14)_mult modellnél. Utóbbira való áttérés a BDF_AR esetén 13,6%-os, míg BDF SETAR esetén 61,9%-os javulást eredményez.

Természetesen minden megállapításhoz hozzá értendő, hogy mindez az általam vizsgált adatbázison érvényes. Ugyanakkor azt is ki kell emelni, hogy az általam vizsgált adatbázis a szakirodalomban találhatóakhoz képest hosszú időszakot ölel át, és az eszközök (részvények) darabszáma is aránylag magas. Az adatbázisom közel 11 éves hosszával szemben a BDF cikk²⁹ 1 éves, a BCG cikk³⁰ pedig 5 éves adatbázisból dolgozik. A vizsgált eszközök tekintetében az adatbázisom 33 részvénye nagyságrendileg hasonló, habár némileg kevesebb a BDF cikk 39 részvényénél. Azonban 10-szer több a BCG cikk által vizsgált három ETF-hez képest. A 91. táblázatból látható, hogy ha a minta mérete alatt az alábbi kifejezést értjük:

$$Minta_méret = Hónapok_száma \cdot Eszközök_száma \quad (115)$$

akkor a mintám a BDF cikkéhez képest 9,2-szer, a BCG cikkben használthoz képest pedig 23,8-szor nagyobb. Az általam végzett vizsgálat így pusztán az adatbázis méretéből adódóan megbízhatóbbnak illetve robusztusabbnak tekinthető.

Végezetül érdemes kiemelni, hogy a benchmarknak tekintett BDF modell mindkét változata lényegesen több adatot használ, mint a Poli(14)_mult modell. Ennek oka, hogy a BDF modelleknek a piacnak tekintett minden (esetünkben 33 darab) részvény adatára szükségük van 1-1 egyedi részvény forgalmának az előrejelzéséhez. Ezzel szemben a Poli(14)_mult modell csak az adott részvény adatából

²⁹Bialkowski et al. (2008)

³⁰Brownlees et al. (2011)

| | Saját adatbázis | BDF modell | BCG modell |
|---------------------------------|-----------------|------------|------------|
| Hónapok száma | 130 | 12 | 60 |
| Eszközök száma | 33 | 39 | 3 |
| Minta mérete (fentiek szorzata) | 4290 | 468 | 180 |

91. táblázat. Adatbázisok méreteinek összevetése

Forrás: Saját szerkesztés

dolgozik, tehát esetünkben 33-szor kevesebb adatból. Ez véleményem szerint ceteris paribus előnynek tekinthető, annál is inkább, mivel annak eldöntése, hogy mit tekintünk releváns piacnak 1-1 instrumentum esetén, közel sem egyértelmű. Ez bizonytalanságot visz a BDF modell alkalmazásába, hiszen a relevánsnak tekintett piac megválasztásától függően esetlegesen jelentősen eltérő eredmények jöhetnek ki. Ez a probléma hatványozottan jöhet elő olyan kisebb piacokon, mint például a hazai, ahol csak néhány likvid papír van, amit fel lehet használni mint piacot.

Ezen felül a BDF cikk nem foglal egyértelműen állást arról, hogy az egyedi részt AR vagy SETAR módon érdemes-e előrejelezni, amely további bizonytalanságot jelent a modell alkalmazása során.

V. rész

Záró gondolatok

13. fejezet

Összefoglalás

Kutatásom témája a tőzsdei forgalom napon belüli előrejelzése, melyet jelen dolgozatban négy részre bontva tárgyaltam: Elméleti háttér (I.), Adatok és hipotézisek (II.), Benchmark kiválasztása (III.), Saját modellek (IV). Az alábbiakban ennek a négy résznek az áttekintése következik.

13.1. Elméleti háttér (I. rész)

Az I. részben a kutatás elméleti hátterével foglalkoztam. A tőzsdei forgalom napon belüli előrejelzése egy olyan terület, amely az utóbbi években kezdett kibontakozni. Az első releváns cikk a témában 2008-ban jelent meg. A kérdés vizsgálata elméleti szempontból igen érdekes, hiszen magáról a forgalomról keveset tudunk. A legtöbb elméleti modell elsősorban az árba, illetve annak alakulására fókuszál. Ugyanakkor világos, hogy az árfeltárás kereskedés nélkül nem valósulhatna meg, ez pedig azt jelenti, hogy a piacokon az információ gyakorlatilag a forgalmon keresztül épül be az árba. Habár a forgalom előrejelzése elsősorban empirikus kutatási kérdés, ezen keresztül a forgalom viselkedésének az elméleti megértéséhez is közelebb kerülhetünk.

A forgalom napon belüli előrejelzésének az elméleti jelentősége mellett kézzelfogható (pénzben mérhető) gyakorlati relevanciája is van, amely minden kereskedéssel foglalkozó piaci szereplő számára nyilvánvaló. A tőzsdék ajánlati könyves piacán (mint minden egyéb piacon) korlátos a likviditás, ezért bizonyos megbízás méret felett számolni kell az ajánlatok áreltérítő hatásával (röviden árhatásával). Ha egy piaci szereplő erre nincs tekintettel, jelentős veszteségeket szenvedhet el

pusztán a kereskedésből adódóan. Emiatt elméletileg is érdemes optimális kereskedési stratégiákon gondolkodni, emellett azonban a gyakorlatban használt kereskedési stratégiákban is figyelemmel kell lenni az árhatás jelenségére. A legtöbb kereskedési stratégiának éppen ezért bemenő paraméterként valamilyen módon tartalmaznia kell a forgalmat. Egy igen gyakori, és egyúttal a teljes piaci forgalom jelentős hányadát adó stratégia a VWAP kereskedés, melynek keretében olyan áron próbálunk venni vagy eladni, amely lehetőség szerint minél közelebb esik az utólag megismert napi átlagárhoz. Ez megfelel sok hosszú távú befektető számára, akik nem napon belül spekulálnak. A VWAP kereskedés érdekessége, hogy kizárólag a forgalom alapján végezhető, az árat nem kell figyelni hozzá. A forgalom minél pontosabb előrejelzése mint egyetlen bemenő változó azonban elengedhetetlen. Mindezek miatt a forgalom előrejelzését szokás kifejezetten a VWAP stratégia fényében is értékelni, ahogyan ezt a szakirodalomhoz hasonlóan a dolgozatban is megtettem.

A téma jelentőségének, illetve az elméletben és gyakorlatban elfoglalt helyének bemutatása után rátértem a forgalom előrejelzés irodalmának áttekintésére. Elsőként választanunk kellett egy mérőszámot, amivel a kereskedési aktivitást mérjük. A dolgozatban felsoroltak közül leggyakrabban a kereskedett darabszámot (volumen), illetve ennek forgalomban lévő darabszámhoz viszonyított arányát, a forgalmat szokás használni. A későbbi modellezéshez érdemes megfigyelni, hogy a szakirodalom milyen stilizált tényeket emel ki a forgalom kapcsán. A legfontosabb ezek közül egyértelműen a napon belüli U alak, de dokumentálták még ezen kívül a hét napja hatást, a hosszú memóriát, a hatványszerű eloszlást, illetve az ármozgással való kapcsolatot is.

Az előrejelzés kapcsán a naiv módszerek közül az U-módszert érdemes kiemelni, amely az előző L nap megfelelő napon belüli rekeszeinek átlagaként jelzi előre a következő rekesz értékét. Ennek a módszernek a jelentősége abban áll, hogy egyrészt a gyakorlatban is gyakran jelzik ilyen módon előre a forgalmat, másrészt éppen emiatt a napon belüli előrejelzéssel foglalkozó cikkek is ezt a módszert tekintik benchmarknak a saját elemzésük során. A fejlettebb módszerek közül fontos megemlíteni Kaastra és Boyd (1995), valamint Lux és Kaizoji (2007) cikkeiket, amelyek havi, illetve napi adatokkal dolgoznak, és ezért csak közvetett előzménynek tekinthetőek. A közvetlenül releváns cikkek a dolgozat szempontjából Bialkowski et al. (2008) és Brownlees et al. (2011), mivel ezek már napon belüli adatokat

használnak.

13.2. Adatok és hipotézisek (II. rész)

A II. részben bemutatam a rendelkezésemre álló adatbázist. Az adattisztítás után a 2001.10.10 és 2012.07.13 közötti 130 hónapos, azaz közel 11 éves időszakot tartottam meg, illetve a Dow Jones Industrial Average indexben szereplő harminchárom (nagy részt NYSE, és néhány NASDAQ) 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.

A II. részben fogalmaztam meg a kutatási kérdéseket is, amelyeket a későbbi empirikus kutatásban vizsgáltam. A fő kérdésem alapvetően az, hogy hogyan lehet a szakirodalomban található modelleknél jobb előrejelző modellt készíteni. Ehhez első lépésben meg kellett határoznom, hogy melyik a legjobb modell a szakirodalomban, ugyanis a releváns cikkek eltérő adatokon, és eltérő hibamértékek szerint számolnak, ezért közvetlenül nem lehet őket összehasonlítani. Az első kutatási kérdésem a fentiekre irányul.

H1 Hipotézis: Benchmark. Az első hipotézisem szerint azonos adatbázison azonos módon értékelve Bialkowski et al. (2008) modellje jobban szerepel, mint Brownlees et al. (2011) modellje, és ezért a szakirodalomból vett benchmarknak az előbbit érdemes tekinteni. Máshogy fogalmazva hipotézisem szerint részvényekre alkalmazva Bialkowski et al. (2008) modellje a tudomány jelenlegi állása szerint a legjobb napon belüli forgalom előrejelző modell.

A benchmark azonosítása után következhet egy annál jobb modell keresése. A második kutatási kérdésem arra irányul, hogy találjak egy ilyen modellt.

H2 Hipotézis: Jobb modell. Az tehát a második hipotézisem, hogy található az irodalom alapján kijelölt benchmarknál jobb modell a részvények napon belüli

forgalom előrejelzésére. Ezen hipotézis igazolása érdekében megkísérlek felírni legalább egy olyan modellt, amely a rendelkezésemre álló adatbázison jobban teljesít a benchmarknál.

A második hipotézishez kapcsolódóan megfogalmazható néhány további kiegészítés arra vonatkozóan, hogy a jobb modellt milyen szempontok modellbe foglalása mentén keresem. Ezek elsősorban a napon belüli U alak, valamint az ármozgással való kapcsolat. Ezek alapján felírhatjuk az alábbi kiegészítő hipotéziseket is.

H2.1 Hipotézis: Az U alak. A forgalom napon belüli U alakjának modellezése hozzájárul a benchmarknál jobb modell felírásához.

H2.2 Hipotézis: Ármozgás mutatók. A forgalom és az ármozgás mutatók kapcsolatának figyelembe vétele hozzájárul a benchmarknál jobb modell felírásához.

A H1 hipotézist a III. részben, a H2 hipotézist és kiegészítéseit pedig a IV. részben vizsgáltam részletesen.

13.3. Benchmark kiválasztása (III. rész)

Az első kutatási kérdést tehát 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 összevetettség 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 használok, 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)³¹ és Brownlees et al. (2011)³² 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 szerepelt nála.

A BCG modell becslése során azonban a cikk ismeretében is felmerültek bizonytalanságok. A modell ismételt részletes bemutatása nélkül, csak utalásszerűen ezek 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

³¹Röviden: BDF modell

³²Röviden: BCG modell

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 becslés³³, 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 modelltől, a BCG modelltől 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.

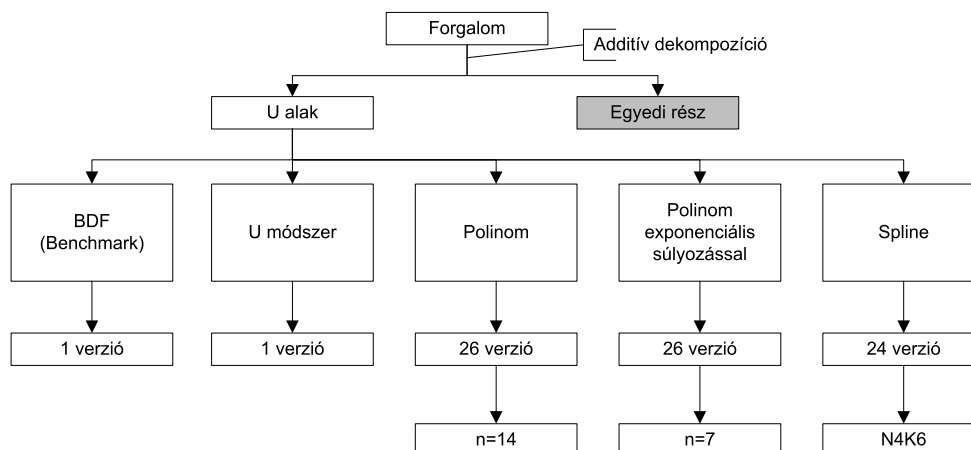
Hozzá kell tenni, hogy Brownlees et al. (2011) ETF adatokkal dolgozik, nem pedig részvény adatokkal, de ez a különbség véleményem szerint kétséges, hogy önmagában indokolhatná a fenti problémákat.

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

³³Ez 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.



15. ábra. Az U alak dekomponálásának különböző módjai

Forrás: Saját szerkesztés

tekinthető a szakirodalomból megismerhető legjobb napon belüli részvény forgalom előrejelző modellnek.

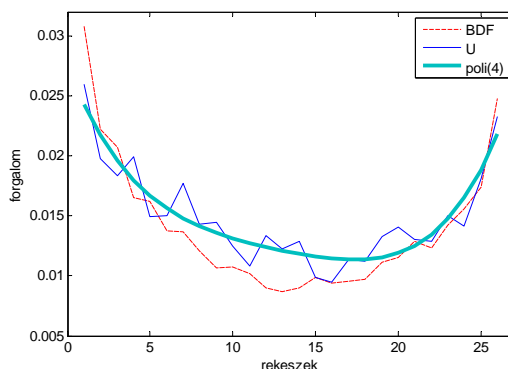
Ezek alapján a IV. részben a modell keresés során a BDF modellt tekintettem benchmarknak.

13.4. Saját modellek (IV. rész)

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 vizsgálgódnia. 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 a 15. ábra.

Elsőként az egyszerű U-módszert egészítettem ki egyedi résszel, és érdekes módon már az így kapott modell is jobban szerepelt a benchmarknál. Ezután észre-



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

vettem, 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.

Az első egy n -edfokú polinom illesztése az adatokra, pontosabban az előrejelzeni kívánt részvény forgalom idősorára. A 16. 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 $n = 7$ mel-

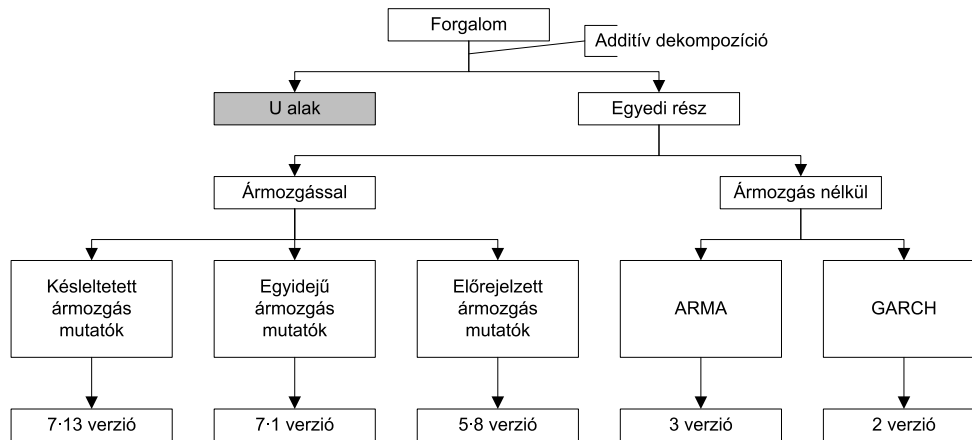
lett döntöttem a korábbiakhoz hasonló vizsgálatok után, szintén hozzátéve, hogy $n \geq 7$ választás mellett elenyésző eltérést tapasztalhatunk a modellek között. Az alacsony, tehát 7-nél kisebb fokszámok használatát azonban sem az egyenlő, sem az exponenciális súlyozás mellett nem javaslom. Az exponenciálisan súlyozott polinom is határozottan jobban szerepel a benchmarknál, de az egyenlően súlyozott esetenél már nem.

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 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 vizsgál-



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

tam különböző egyedi rész előrejelzési lehetőségeket. Az egyes modell változatok áttekintésében a 17. 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 (1. 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 változókra 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 tesztet öltének, 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.

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.

| Fejezet | Mit jelzünk előre | Mit figyelünk | Hiba-mérték | 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 |

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

A IV. rész zárásaként a legjobb saját modelleket az eddigiek mellett további szempontok mentén is összevetettem a benchmarkkal. Tekintsük át ezen szempontokat a 92. 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.

A 92. 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 jobbnak³⁴.

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 módszertől, de ez az előzőekhez képest kevésbé jelentős eltérés.

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 változa-

³⁴Az egyetlen kivétel az egy lépéses forgalom érték előrejelzésének átlagos MSE* értéke, amely az additív esetben volt alacsonyabb.

tától 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.

14. fejezet

A dolgozat főbb eredményei röviden

Tekintsük át még egyszer röviden a dolgozat főbb eredményeit.

1. Azonos adatokon (időszak és eszközök) azonos módon értékelve összevetetem 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.

15. fejezet

További kutatási kérdések

A tőzsdei forgalom napon belüli előrejelzésének témájában az első publikáció csak néhány évvel ezelőtt, 2008-ban jelent meg. Ez tehát még mindig egy friss, épp kibontakozóban lévő kutatási terület, ezért természetesen számtalan szempontból érdemes vele a továbbiakban is foglalkozni. Az alábbiakban kiemelek néhány szempontot, amelyek jelen dolgozathoz is közvetlenül kapcsolódnak.

Adatkezelés. Érdemes lehet megvizsgálni, hogy az általam használt 15 perces-től eltérő aggregálási szinteken hogyan változik a modellek hatékonysága. Előfordulhat, hogy bizonyos aggregálási szinteket célszerűbb választani, mint másokat. Szintén az adatokkal kapcsolatos felvetés, hogy az általam használtaktól lényegesen kevésbé likvid eszközökre hogyan működnek a modellek. Ha gyakran nincs kereskedés egy adott rekeszben, akkor az adott hiányzó megfigyelés nulla forgalommal minden további nélkül létrehozható ugyan, de az egyes modellek érzékenysége az ilyen esetekre külön vizsgálatot érdemel. Feltehetően ebbe a kategóriába esne a magyar részvények jelentős része. Végül pedig szintén az adatok felhasználásának kérdése az is, hogy milyen hosszú becslési időszakot választunk. A dolgozatban egységesen 20 kereskedési nappal számoltam, de érdekes lehet megvizsgálni, hogy ennek változtatása hogyan hat a modellek előrejelzésére.

Modellezés. A dolgozatban egy multiplikatív felírású modellt találtam a legjobbnak. Ugyanakkor ennek mind az U dekompozíciós, mint pedig az egyedi részét az additív modellezés alapján választottam ki. Érdekes lenne megvizsgálni, hogy multiplikatív esetben melyik a legjobb U dekompozíció az általam javasoltak közül (lehet, hogy ez esetben nem az egyenlő súlyozású polinom lenne az). Az sem

biztos, hogy a választott foksám megegyezne multiplikatív esetben. Hasonlóan, érdemes lenne a multiplikatív specifikációban is tesztelni egyéb egyedi rész modelleket is. Végül pedig, amint azt már korábban is megjegyeztem, az ármozgás mutatók előrejelzésének a dolgozatban vizsgálnál léteznek lényegesen fejlettebb modelljei is, melyeket szintén érdemes lenne bevonni a vizsgálatba.

Értékelés. Az előrejelzések értékelése kapcsán is felmerülnek nyitott kérdések. Mivel igen hosszú, 130 hónapos a vizsgált időszak, érdemes lehet nem csak egészében, de rövidebb időszakokra felbontva is értékelést végezni. A felbontás történhetne a gyakran választott *válságok előtt/alatt/után*, vagy egyéb logika szerint is. Szintén az értékeléshez kapcsolódik, hogy a vizsgált 12 szempont mellett egyéb módon is lehetne esetleg értékelni a modellek teljesítményét. Kiemelt jelentősége lehet például az itt látottaktól eltérő kereskedési stratégiák értékelésének. Ehhez természetesen előbb magukat a kereskedési stratégiákat kellene definiálni, amely már csak közvetve kapcsolódik az előrejelzés témájához, de ettől még érdekes iránynak tekinthető. Végül pedig természetesen érdemes figyelni a szakirodalomban később megjelenő újabb modell javaslatokat, és azokkal is összevetni a dolgozatban tárgyalt modelleket.

Irodalomjegyzék

- Abhyankar, A., Ghosh, D., Levin, E., and Limmack, R. J. (1997). Bid-ask spreads, trading volume and volatility: Intraday evidence from the London Stock Exchange. *Journal of Business Finance and Accounting*, 24:343–362. DOI: 10.1111/1468–5957.00108.
- Andersen, T. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance*, 51:169–204. DOI: 10.1111/j.1540–6261.1996.tb05206.x.
- Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4:115–158 DOI: 10.1016/S0927–5398(97)00004–2.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71:135–171. DOI: 10.1111/1468–0262.00392.
- Bamber, L. (1986). The information content of annual earnings releases: A trading volume approach. *Journal of Accounting Research*, 24:40–56. DOI: 10.2307/2490803.
- Basak, G., Chan, N., and Palma, W. (2001). The approximation of long-memory processes by an ARMA model. *Journal of Forecasting*, 20:367–389. DOI: 10.1002/for.799.
- Berkowitz, S., Logue, D., and Noser, E. (1988). The total cost of transactions on the NYSE. *Journal of Finance*, 41:97–112. DOI: 10.1111/j.1540–6261.1988.tb02591.x.
- Berlinger, E., Gerencsér, L., Havran, D., Margitai, I., and Szűcs, B. A. (2011).

- Trading strategies on limit order markets. Working paper, Morgan Stanley, University Research & Development Projects.
- Biais, B., Hillion, P., and Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *The Journal of Finance*, 50:1655–1689. DOI: 10.1111/j.1540-6261.1995.tb05192.x.
- 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.
- Bildik, R. (2001). Intra-day seasonalities on stock returns: Evidence from the Turkish Stock Market. *Emerging Markets Review*, 2:387–417. DOI: 10.1016/s1566-0141(01)00026-7.
- Bollerslev, T. and Ghysels, E. (1996). Periodic autoregressive conditional heteroskedasticity. *Journal of Business and Economic Statistics*, 14:139–151. DOI: 10.1080/07350015.1996.10524640.
- Bouchaud, J.-P., Farmer, J. D., and Lillo., F. (2008). How markets slowly digest changes in supply and demand. Working paper, <http://ssrn.com/abstract=1266681>.
- Brock, W. A. and Kleidon, A. W. (1992). Periodic market closure and trading volume: A model of intraday bid and asks. *Journal of Economic Dynamics and Control*, 16:451–489. DOI: 10.1016/0165-1889(92)90045-g.
- Brodsky, J. and Hurvich, C. (1999). Multi-step forecasting for long-memory processes. *Journal of Forecasting*, 18:59–75. DOI: 10.1002/(sici)1099-131x(199901)18:1<59::aid-for711>3.0.co;2-v.
- 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.

- Brunnermeier, M. K. and Pedersen, L. H. (2008). Market liquidity and funding liquidity. *The Review of Financial Studies*, 22:2201–2238. DOI: 10.1093/rfs/hhn098.
- Calvet, L. and Fisher, A. (2001). Forecasting multifractal volatility. *Journal of Econometrics*, 105:27–58. DOI: 10.1016/s0304-4076(01)00069-0.
- Calvet, L. and Fisher, A. (2002). Multi-fractality in asset returns: Theory and evidence. *Review of Economics & Statistics*, 84:381–406. DOI: 10.1162/003465302320259420.
- Campbell, J., Grossman, S., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics*, 108:905–939. DOI: 10.2307/2118454.
- Chiang, T. C., Qiao, Z., and Wong, W.-K. (2010). New evidence on the relation between return volatility and trading volume. *Journal of Forecasting*, 29:502–515. DOI: 10.1002/for.1151.
- Choe, H. and Shin, H. (1993). An analysis of interday and intraday return volatility, evidence from the Korean Stock Exchange. *Pacific Basin Finance Journal*, 1:175–188. DOI: 10.1016/0927-538x(93)90007-5.
- Chordia, T., Roll, R., and Subrahmanyam, A. (2001). Market liquidity and trading activity. *The Journal of Finance*, 56:501–530. DOI: 10.1111/0022-1082.00335.
- Conrad, J., Hameed, A., and Niden, C. (1994). Volume and autocovariances in short-horizon individual security returns. *Journal of Finance*, 49:1305–1329. DOI: 10.1111/j.1540-6261.1994.tb02455.x.
- Crato, N. and Ray, B. (1996). Model selection and forecasting for long-range dependent processes. *Journal of Forecasting*, 15:107–125. DOI: 10.1002/(sici)1099-131x(199603)15:2<107::aid-for612>3.0.co;2-d.
- Darolles, S. and Le Fol, G. (2003). Trading volume and arbitrage. Working Paper, Institut Nationale de la Statistique et des Etudes Economiques, CREST.
- De Boor, C. (1978). *A Practical Guide to Splines*. Springer Verlag. DOI: 10.1007/978-1-4612-6333-3.

- Ding, D. K. and Lau, S. (2001). An analysis of transactions data for the Stock Exchange of Singapore: Patterns, absolute price change, trade size and number of transactions. *Journal of Business Finance and Accounting*, 28:151–174. DOI: 10.1111/1468–5957.00369.
- Eisler, Z. and Kertész, J. (2007). Liquidity and the multiscaling properties of the volume traded on the stock market. *EPL (Europhysics Letters)*, 77:28001 p1–p5. DOI: 10.1209/0295–5075/77/28001.
- Engle, R. (2000). The econometrics of ultra high frequency data. *Econometrica*, 68:1–22. DOI: 10.1111/1468–0262.00091.
- Engle, R. F., Sokalska, M., and Chanda, A. (2007). Forecasting intraday volatility in the US equity market. Multiplicative component GARCH. Technical report, North American Winter Meetings of the Econometric Society.
- Epps, T. and Epps, M. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture of distribution hypothesis. *Econometrica*, 44:305–321. DOI: 10.2307/1912726.
- Foster, F. D. and Viswanathan, S. (1993). Variations in trading volume, return volatility and trading costs: Evidence on recent price formation models. *The Journal of Finance*, 48:187–211. DOI: 10.1111/j.1540–6261.1993.tb04706.x.
- Gallant, R., Rossi, P., and Tauchen, G. (1992). Stock prices and volume. *Review of Financial Studies*, 5:199–242. DOI: 10.1093/rfs/5.2.199.
- Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221–238. DOI: 10.1111/j.1467–9892.1983.tb00371.x.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48:1779–1801. DOI: 10.2307/2329067.
- Gopikrishnan, P., Plerou, V., Liu, Y., Amaral, L. A. N., Gabaix, X., and Stanley, H. E. (2000). Scaling and correlation in financial time series. *Physica A: Statistical Mechanics and its Applications*, 287:362–373. DOI: 10.1016/s0378–4371(00)00375–7.

- Gouriéroux, C. and Le Fol, G. (1998). Effet des modes de négociation sur les échanges. *Revue Economique*, 49:795–808. DOI: 10.2307/3502810.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.
- Harris, L. (1986). A transaction data survey of weekly and intraday patterns on stock returns. *Journal of Financial Economics*, 16:99–117. DOI: 10.1016/0304-405x(86)90044-9.
- Harris, L. and Gurel, E. (1986). Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures. *Journal of Finance*, 41:815–829. DOI: 10.2307/2328230.
- Hautsch, N. (2008). Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model. *Journal of Economic Dynamics and Control*, 32:3978–4015. DOI: 10.1016/j.jedc.2008.01.009.
- Havran, D., Margitai, I., and Szűcs, B. A. (2011). Kereskedési stratégiák ajánlati könyves piacokon. In *Budapesti Corvinus Egyetem Közgazdasági Doktori Iskola VII. Éves Konferenciája*.
- Hiemstra, C. and Jones, J. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, 49:1639–1664. DOI: 10.1111/j.1540-6261.1994.tb04776.x.
- Hmaied, D. M., Sioud, O. B., and Grar, A. (2006). Intradaily and weekly patterns of bid-ask spreads, trading volume and volatility on the Tunisian Stock Exchange. *Banque & Marchés*, 84:35–44.
- Hu, S. (1997). Trading turnover and expected stock returns: Does it matter and why? Working paper, National Taiwan University.
- Hussain, S. M. (2011). The intraday behaviour of bid-ask spreads, trading volume and return volatility: Evidence from DAX30. *International Journal of Economics and Finance*, 3:23–34. DOI: 10.5539/ijef.v3n1p23.

- Jain, P. and Joh, G. (1988). The dependence between hourly prices and trading volume. *Journal of Financial and Quantitative Analysis*, 23:269–284. DOI: 10.2307/2331067.
- James, C. and Edmister, R. (1983). The relation between common stock returns, trading activity and market value. *Journal of Finance*, 38:1075–1086. DOI: 10.2307/2328012.
- 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.
- Kovács, E. (2006). *Pénzügyi adatok statisztikai elemzése*. Tanszék Pénzügyi Tanácsadó és Szolgáltató Kft., Budapest.
- Lakonishok, J. and Maberly, E. (1990). The weekend effect: Trading patterns of individual and institutional investors. *The Journal of Finance*, 45:231–243. DOI: 10.2307/2328818.
- Lakonishok, J. and Smidt, S. (1986). Volume for winners and losers: Taxation and other motives for stock trading. *Journal of Finance*, 41:951–974. DOI: 10.1111/j.1540–6261.1986.tb04559.x.
- Lakonishok, J. and Vermaelen, T. (1986). Tax-induced trading around ex-dividend days. *Journal of Financial Economics*, 16:287–319. DOI: 10.1016/0304–405x(86)90032–2.
- Lamoureux, C. and Lastrapes, W. (1990). Heteroskedasticity in stock return data: Volume vs. GARCH effects. *Journal of Finance*, 45:221–229. DOI: 10.1111/j.1540–6261.1990.tb05088.x.
- Lamoureux, C. and Lastrapes, W. (1994). Endogenous trading volume and momentum in stock-return volatility. *Journal of Business and Economic Statistics*, 12:253–260. DOI: 10.2307/1391488.
- LeBaron, B. (1992). Persistence of the Dow Jones index on rising volume. Working paper, University of Wisconsin.

- Lee, Y. T., Fok, R., and Liu, Y. (2001). Explaining intraday pattern of trading volume from the order flow data. *Journal of Business Finance and Accounting*, 28:199–230. DOI: 10.1111/1468–5957.00371.
- Lo, A. W. and Wang, J. (2009). Stock market trading volume. Draft to appear in Y. Ait-Sahalia and L. Hansen, eds., 2009, *The Handbook of Financial Econometrics*. New York: North-Holland.
- Lobato, I. N. and Velasco, C. (2000). Long memory in stock-market trading volume. *Journal of Business & Economic Statistics*, 18:410–427. DOI: 10.2307/1392223.
- Lux, T. (2003). The multi-fractal model of asset returns: Its estimation via GMM and its use for volatility forecasting. Working paper, University of Kiel.
- 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.
- Makara, T. (2000). A hozamgörbe becslési módszerei. *Ph.D. értekezés, Budapesti Közgazdaságtudományi és Államigazgatási Egyetem*.
- Man, K. (2003). Long memory time series and short term forecasts. *International Journal of Forecasting*, 19:477–491. DOI: 10.1016/s0169–2070(02)00060–2.
- McInish, T. H. and Wood, R. A. (1990). An analysis of transactions data for the Toronto Stock Exchange: Return patterns and end-of-the-day effect. *Journal of Banking and Finance*, 14:441–458. DOI: 10.1016/0378–4266(90)90058–a.
- McInish, T. H. and Wood, R. A. (1992). An analysis of intraday patterns in bid/ask spreads for NYSE stocks. *The Journal of Finance*, 47:753–764. DOI: 10.1111/j.1540–6261.1992.tb04408.x.
- Morse, D. (1980). Asymmetric information in securities markets and trading volume. *Journal of Financial and Quantitative Analysis*, 15:1129–1148. DOI: 10.2307/2330176.

- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370. DOI: 10.2307/2938260.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier.
- Parlour, C. A. and Seppi, D. J. (2008). Limit order markets: A survey. In: A. V. Thakor and A. W. A. Boot, editors, *Handbook of Financial Intermediation and Banking*, Elsevier DOI: 10.1016/b978-044451558-2.50007-6.
- Petrimán, Z. and Tulassay, Z. (2005). Bepillantás az ARCH modellek világába. *Hitelintézeti Szemle*, 2:74–79.
- Plerou, V., Gopikrishnan, P., Amaral, L. A. N., Gabaix, X., and Stanley, H. (2000). Economic fluctuations and anomalous diffusion. *Physical Review E*, 62:R3023–R3026. DOI: 10.1103/physreve.62.r3023.
- Plerou, V., Gopikrishnan, P., Gabaix, X., Amaral, L. A. N., and Stanley, H. E. (2001). Price fluctuations, market activity and trading volume. *Quantitative Finance*, 1:262–269. DOI: 10.1088/1469–7688/1/2/308.
- Plerou, V., Gopikrishnan, P., Gabaix, X., and Stanley, H. E. (2004). On the origin of power-law fluctuations in stock prices. *Quantitative Finance*, 4:C11–C15. DOI: 10.1088/1469–7688/4/1/c02.
- Qiu, T., Zhong, L. X., Chen, G., and Wu, X. R. (2009). Statistical properties of trading volume of Chinese stocks. *Physica A: Statistical Mechanics and its Applications*, 388:2427–2434. DOI: 10.1016/j.physa.2009.02.038.
- Rabemananjara, R. and Zakoian, J. M. (1993). Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics*, 8:31–49. DOI: 10.1002/jae.3950080104.
- Ramanathan, R. (2003). *Bevezetés az ökonometriába alkalmazásokkal*. Panem Könyvkiadó, Budapest.
- Richardson, G., Sefcik, S., and Thompson, R. (1986). A test of dividend irrelevance using volume reaction to a change in dividend policy. *Journal of Financial Economics*, 17:313–333. DOI: 10.1016/0304–405x(86)90068–1.

- Sewell, M. (2011). Characterization of financial time series. Research note, UCL Department of Computer Science.
- Smidt, S. (1990). Long-run trends in equity turnover. *Journal of Portfolio Management*, Fall:66–73. DOI: 10.3905/jpm.1990.409300.
- Stickel, S. and Verrecchia, R. (1994). Evidence that volume sustains price changes. *Financial Analysts Journal*, 6:56–57. DOI: 10.2469/faj.v50.n6.57.
- 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.
- Tkac, P. (1999). A trading volume benchmark: Theory and evidence. *Journal of Financial and Quantitative Analysis*, 34:89–114. DOI: 10.2307/2676247.
- Weild, D. (2015). Economic development and the IPO market. *Előadás, Budapesti Corvinus Egyetem, 2015. szeptember 30.*
- Wood, R. A., Mc Inish, T. H., and Ord, J. K. (1985). An investigation of transactions data for NYSE stocks. *The Journal of Finance*, 40:723–739. DOI: 10.1111/j.1540–6261.1985.tb04996.x.
- Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 45, pages 2639–2738. Elsevier. DOI: 10.1016/s1573-4412(05)80014-5.
- Ying, C. C. (1966). Stock market prices and volume of sales. *Econometrica*, 34:676–686. DOI: 10.2307/1909776.

Melléklet

Magas dimenziójú faktorelemzés

A faktorelemzés az alábbi egyenletből indul ki (Kovács (2006)):

$$\underset{(TxN)}{X} = \underset{(Txr)}{F} \underset{(rxN)}{\Lambda'} + \underset{(TxN)}{e} \quad (116)$$

ahol

- X a kiinduló adatokat tartalmazó mátrix
- F jelöli a faktorok mátrixát
- Λ a faktorsúlyok mátrixa
- e a hibatag mátrixa
- T a megfigyelések száma, N a változók száma, r a faktorok száma.

Az egyenlet jobb oldalán található változók nem megfigyelhetők.

A klasszikus faktorelemzés legfontosabb feltevései az alábbiak (Bai (2003) és Kovács (2006)):

1. N „sokkal kisebb”, mint T
2. A hibák időben és változónként is függetlenek
3. A faktorok szintén függetlenek
4. A faktorok a hibáktól is függetlenek

Ezek a feltevések bizonyos közgazdasági alkalmazásokban túlságosan szigorúnak bizonyulnak, melynek oka részben a magas dimenziószám mind N mind pedig T tekintetében. Bai (2003) a következőket emeli ki.

1. A figyelembe venni kívánt változók száma (N) könnyen meghaladhatja a megfigyelések számát (T), amennyiben éves vagy akár negyedéves adatokkal dolgozunk (ami például a makroökonómiában gyakori). Amennyiben néhány változót választunk csak ki az N és T relációjának helyreállítása érdekében, jelentős mennyiségű információt veszíthetünk.
2. A hibatagok változónkénti függetlensége idősoros adatoknál gyakran nem elfogadható feltevés.
3. A klasszikus faktorelemzés konzisztens becslést ad a súlyvektorra, de a faktor értékére nem feltétlenül. Közgazdasági alkalmazásokban azonban gyakran a faktor értéke az, ami közvetlenül érdekli az elemzőt.

Bai (2003) a következő módszert javasolja a fenti problémák enyhítésére. A kiinduló feladat:

$$X = F\Lambda' + e = K + e \quad (117)$$

ahol K a közös komponens jelölő ($T \times N$) méretű mátrix, és minden más meg egyezik a (116) egyenlettel.

A faktormodellek becsülhetőek maximum likelihood módszerrel, főkomponens módszerével, valamint állapotter reprezentációban is. Magas dimenziók esetén a főkomponens módszer a leghatékonyabb, amely a következő feladat megoldását jelenti:

$$V(r) = \min_{\Lambda, F} (N, T)^{-1} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda_i' F_t)^2 \quad (118)$$

Az $F'F/T = E$ normalizálással élve (E egységmátrix) a fentivel ekvivalens feladat a $\text{tr}(F'(XX')F)$ kifejezés maximalizálása. Tekintsük az XX' kifejezést, melynek dimenziója ($T \times T$). Keressük meg ennek a sajátértékeit és sajátvektorait, és rendezzük a sajátértékeket csökkenő sorrendbe. Jelölje Eig az r legnagyobb (Txr) sajátértékhez tartozó sajátvektorok mátrixát (az első oszlop tartozik a legnagyobb sajátértékhez, a második a második legnagyobbhoz, stb.). Ekkor a következő eredményeket kapjuk.

A becsült faktor mátrix:

$$\tilde{F} = \sqrt{T}Eig \quad (119)$$

a becstült súlyvektor transzponáltja:

$$\tilde{\Lambda}' = (\tilde{F}'\tilde{F})^{-1}\tilde{F}X = \tilde{F}X/T \quad (120)$$

a közös komponens becslése pedig:

$$\tilde{K} = \tilde{F}\tilde{\Lambda}' \quad (121)$$

Bialkowski et al. (2008) eredményei

Bialkowski et al. (2008) egyéni részvényekre vonatkozó eredményei láthatóak a 93-95. táblázatban.

| # | U (1) | Elméleti SETAR (2) | Dinamikus SETAR (3) | (2)-(1) | (3)-(1) |
|-----|--------|-----------------------|------------------------|---------|---------|
| 1 | 0,1047 | 0,0906 | 0,1121 | -0,0141 | 0,0074 |
| 2 | 0,1316 | 0,1023 | 0,1209 | -0,0293 | -0,0107 |
| 3 | 0,0801 | 0,0726 | 0,0818 | -0,0075 | 0,0017 |
| 4 | 0,1336 | 0,0845 | 0,1079 | -0,0491 | -0,0257 |
| 5 | 0,1171 | 0,0665 | 0,1062 | -0,0506 | -0,0109 |
| 6 | 0,093 | 0,072 | 0,0889 | -0,021 | -0,0041 |
| 7 | 0,0782 | 0,071 | 0,0742 | -0,0072 | -0,004 |
| 8 | 0,1715 | 0,1623 | 0,1773 | -0,0092 | 0,0058 |
| 9 | 0,2323 | 0,1448 | 0,1491 | -0,0875 | -0,0832 |
| 10 | 0,0628 | 0,0537 | 0,0638 | -0,0091 | 0,001 |
| 11 | 0,1465 | 0,1054 | 0,1129 | -0,0411 | -0,0336 |
| 12 | 0,1389 | 0,0902 | 0,1102 | -0,0487 | -0,0287 |
| 13 | 0,0548 | 0,0459 | 0,0531 | -0,0089 | -0,0017 |
| ... | | | | | |

93. táblázat. Bialkowski et al. (2008) eredményei az egyes részvényekre
Százalékban értendő MAPE értékek
Forrás: Bialkowski et al. (2008)

| # | U (1) | Elméleti SETAR (2) | Dinamikus SETAR (3) | (2)-(1) | (3)-(1) |
|-----|--------|--------------------|---------------------|---------|---------|
| 14 | 0,1099 | 0,0848 | 0,0779 | -0,0251 | -0,032 |
| 15 | 0,1947 | 0,1434 | 0,1404 | -0,0513 | -0,0543 |
| 16 | 0,1398 | 0,1006 | 0,108 | -0,0392 | -0,0318 |
| 17 | 0,0866 | 0,0698 | 0,0832 | -0,0168 | -0,0034 |
| 18 | 0,1076 | 0,0964 | 0,1075 | -0,0112 | -0,0001 |
| 19 | 0,1003 | 0,0816 | 0,1141 | -0,0187 | 0,0138 |
| 20 | 0,1131 | 0,0913 | 0,0959 | -0,0218 | -0,0172 |
| 21 | 0,1541 | 0,138 | 0,1513 | -0,0161 | -0,0028 |
| 22 | 0,0775 | 0,0532 | 0,0745 | -0,0243 | -0,003 |
| 23 | 0,0762 | 0,059 | 0,0801 | -0,0172 | 0,0039 |
| 24 | 0,1389 | 0,0778 | 0,0998 | -0,0611 | -0,0391 |
| 25 | 0,1406 | 0,1076 | 0,1287 | -0,033 | -0,0119 |
| 26 | 0,0979 | 0,0895 | 0,0952 | -0,0084 | -0,0027 |
| ... | | | | | |

94. táblázat. Bialkowski et al. (2008) eredményei az egyes részvényekre (folyt)
Százalékban értendő MAPE értékek
Forrás: Bialkowski et al. (2008)

| # | U (1) | Elméleti SETAR (2) | Dinamikus SETAR (3) | (2)-(1) | (3)-(1) |
|----|--------|--------------------|---------------------|---------|---------|
| 27 | 0,0999 | 0,0707 | 0,0897 | -0,0292 | -0,0102 |
| 28 | 0,0865 | 0,0788 | 0,1027 | -0,0077 | 0,0162 |
| 29 | 0,0699 | 0,0653 | 0,0617 | -0,0046 | -0,0082 |
| 30 | 0,1233 | 0,0806 | 0,1182 | -0,0427 | -0,0051 |
| 31 | 0,0906 | 0,0802 | 0,0768 | -0,0104 | -0,0138 |
| 32 | 0,0968 | 0,0725 | 0,0908 | -0,0243 | -0,006 |
| 33 | 0,1103 | 0,0899 | 0,1118 | -0,0204 | 0,0015 |
| 34 | 0,0959 | 0,0782 | 0,1027 | -0,0177 | 0,0068 |
| 35 | 0,146 | 0,0784 | 0,1398 | -0,0676 | -0,0062 |
| 36 | 0,0528 | 0,0496 | 0,0508 | -0,0032 | -0,002 |
| 37 | 0,13 | 0,0899 | 0,1286 | -0,0401 | -0,0014 |
| 38 | 0,0774 | 0,0559 | 0,0755 | -0,0215 | -0,0019 |
| 39 | 0,1095 | 0,0746 | 0,102 | -0,0349 | -0,0075 |

95. táblázat. Bialkowski et al. (2008) eredményei az egyes részvényekre (folyt.(2))
Százalékban értendő MAPE értékek
Forrás: Bialkowski et al. (2008)

Adatok

A [kibot.com](http://www.kibot.com)³⁵ oldalon részletes információ található az általam is használt adatokról, melynek a fontosabb elemei a 96-98. táblázatban szerepelnek. A Kraft Foods Inc. a táblázatban NYSE részvényként jelenik meg, azonban az adatbázis által lefedett időszak kb. utolsó három hetében (2012.06.26. kezdettel) átvezették a NASDAQ-ra.

| # | Kód | Név | Szektor | Jegyzés |
|-----|------|--------------------------------------|----------------|---------|
| 1 | AA | Alcoa, Inc. | Alapanyagok | NYSE |
| 2 | AIG | American International Group, Inc. | Pénzügy | NYSE |
| 3 | AXP | American Express Company | Pénzügy | NYSE |
| 4 | BA | Boeing Co. | Ipari termékek | NYSE |
| 5 | BAC | Bank of America Corporation | Pénzügy | NYSE |
| 6 | C | Citigroup, Inc. | Pénzügy | NYSE |
| 7 | CAT | Caterpillar Inc. | Ipari termékek | NYSE |
| 8 | CSCO | Cisco Systems, Inc. | Technológia | NASDAQ |
| 9 | CVX | Chevron Corporation | Alapanyagok | NYSE |
| 10 | DD | E. I. du Pont de Nemours and Company | Alapanyagok | NYSE |
| 11 | DIS | Walt Disney Co. | Szolgáltatás | NYSE |
| 12 | GE | General Electric Company | Ipari termékek | NYSE |
| ... | | | | |

96. táblázat. Felhasznált adatok

Forrás: Kibot.com

³⁵http://www.kibot.com/Historical_Data/Dow_30_Historical_Intraday_Data.aspx

| # | Kód | Név | Szektor | Jegyzés |
|-----|------|---------------------------------------|---------------------|---------|
| 13 | GM | General Motors Company | Fogyasztói termékek | NYSE |
| 14 | HD | The Home Depot, Inc. | Szolgáltatás | NYSE |
| 15 | HON | Honeywell International Inc. | Ipari termékek | NYSE |
| 16 | HPQ | Hewlett-Packard Company | Technológia | NYSE |
| 17 | IBM | International Business Machines Corp. | Technológia | NYSE |
| 18 | INTC | Intel Corporation | Technológia | NASDAQ |
| 19 | JNJ | Johnson & Johnson | Egészség | NYSE |
| 20 | JPM | JPMorgan Chase & Co. | Pénzügy | NYSE |
| 21 | KFT | Kraft Foods Inc. | Fogyasztói termékek | NYSE |
| 22 | KO | The Coca-Cola Company | Fogyasztói termékek | NYSE |
| 23 | MCD | McDonald's Corp. | Szolgáltatás | NYSE |
| 24 | MMM | 3M Co. | Konglomerátum | NYSE |
| ... | | | | |

97. táblázat. Felhasznált adatok (folyt.)

Forrás: Kibot.com

| # | Kód | Név | Szektor | Jegyzés |
|----|------|---------------------------------|---------------------|---------|
| 25 | MO | Altria Group Inc. | Fogyasztói termékek | NYSE |
| 26 | MRK | Merck & Co. Inc. | Egészség | NYSE |
| 27 | MSFT | Microsoft Corporation | Technológia | NASDAQ |
| 28 | PFE | Pfizer Inc. | Egészség | NYSE |
| 29 | PG | Procter & Gamble Co. | Fogyasztói termékek | NYSE |
| 30 | PGP | Pimco Global Stocksplus & Incom | - | NYSE |
| 31 | T | AT&T, Inc. | Technológia | NYSE |
| 32 | TRV | The Travelers Companies, Inc. | Pénzügy | NYSE |
| 33 | UTX | United Technologies Corp. | Konglomerátum | NYSE |
| 34 | VZ | Verizon Communications Inc. | Technológia | NYSE |
| 35 | WMT | Wal-Mart Stores Inc. | Szolgáltatás | NYSE |
| 36 | XOM | Exxon Mobil Corporation | Alapanyagok | NYSE |

98. táblázat. Felhasznált adatok (folyt.(2))

Forrás: Kibot.com

A B-spline bázis

A 10.2.4. pont utal a B-spline bázisra, de a főszövegben nem éreztem indokoltnak a számításmódot részletesen bemutatni. Ugyanakkor az átláthatóság érdekében

itt a mellékletben kitekerek rá.

Makara (2000) a 36-37. oldalon részletesen bemutatja, hogyan lehet a B-spline bázist rekurzív módon meghatározni. Az alábbiakban ennek a leírásnak a szó szerinti idézése következik.

„Kényelmi és számítási okokból célszerű tehát egy olyan bázist meghatározni, ahol a bázisfüggvények ortogonálisak. Az ortogonális bázis azt jelentené, hogy a $[t_1, t_K]$ intervallum minden egyes pontjában csak egyetlen bázisfüggvény értéke tér el a nullától. Sajnos ortogonális spline bázis nem határozható meg, de közel ortogonális igen. Egy ilyen bázist jelentenek a B-spline bázisfüggvények (De Boor (1978)).

Legyen a csomópontok szigorúan monoton növekvő listája t_1, t_2, \dots, t_K . A csomópontokon értelmezett r -ed fokú spline függvénytér B-spline bázisát kívánjuk meghatározni. Jelölje a k . bázisfüggvény értékét a $t \in [t_1, t_K]$ pontban $(B_k^r(t), k = 1, 2, \dots, K + r)$. A $B_k^r(t)$ értékek kiszámítására De Boor (1978) a következő algoritmust javasolja. Hozzuk létre a csomópontok $d_1, d_2, \dots, d_{K+2r}$ bővített listáját a következőképpen:

$$d_1 = d_2 = \dots = d_r = t_1$$

$$d_{j+r} = t_j, \quad j = 1, 2, \dots, K$$

$$d_{K+r+1}, d_{K+r+2}, \dots, d_{K+r+r} = t_K$$

Mindössze annyit csináltunk, hogy a csomópontok elé raktuk r -szer a t_1 értéket, és r -szer a csomópontok mögé helyeztük a t_K értéket.

A $B_k^r(t) = \Phi_k^r(t)$ érték kiszámításához a következő rekurzív összefüggést használhatjuk:

$$\Phi_i^l(t) = \frac{t-d(i)}{d(i+l)-d(i)} \Phi_i^{l-1}(t) + \frac{d(i+l+1)-t}{d(i+l+1)-d(i+1)} \Phi_{i+1}^{l-1}(t)$$

$$\Phi_i^0(t) = \begin{cases} 1, & \text{ha } d_i \leq t < d_{i+1} \\ 0, & \text{egyébként} \end{cases}$$

$$\Phi_{r+K}^0(t_K) = 1$$

Az algoritmus alkalmazása során a $0 \div 0 = 0$ konvenció alkalmazandó.

Az alábbi ábra a rekurzió menetét ábrázolja. A rekurzió során balról jobbra haladunk, minden elem a tőle balra lévő oszlopban, a vele azonos sorban és az alatta lévő sorban található elem lineáris kombinációjaként számítható ki.

$$\begin{array}{ccccccc}
 \Phi_k^0(t) & \cdots & \Phi_k^{r-2}(t) & \Phi_k^{r-1}(t) & \Phi_k^r(t) & & \\
 \Phi_{k+1}^0(t) & \cdots & \Phi_{k+1}^{r-2}(t) & \Phi_{k+1}^{r-1}(t) & & & \\
 \Phi_{k+2}^0(t) & \cdots & \Phi_{k+2}^{r-2}(t) & & & & \\
 \vdots & \cdots & & & & & \\
 \Phi_{k+r}^0(t) & & & & & &
 \end{array}$$

” (Makara (2000), 36-37. o.)

Feltételes modellek további eredményei

A 10.3.1.2. és a 10.3.1.3. pontokban csak 1-1 választott küszöbérték mellett mutattam be az eredményeket: a Granger-okság használata esetén a *küszöb* = 0,1 értékkel, míg a korreláció használatakor a *küszöb* = 0,3 értékkel számolt modellek eredményeit találjuk a főszövegben. Itt a mellékletben a 99 - 100. táblázatok mutatják a *küszöb* = {0,01; 0,05} értékekkel számolt modellek eredményeit Granger-okság esetére, illetve a 101 - 108. táblázatokban találjuk a *küszöb* = {0,1; 0,2; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9} értékekkel számolt modellek eredményeit korreláció esetére.

| Poli(14)_ARX_Gr küszöb: 0,01 | Hány részvényt nyert | | Átlagos érték | |
|---------------------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 2 | 0 | 6,09E-04 | 39,1% |
| Volatilitás | 8 | 0 | 6,07E-04 | 39,2% |
| Rés | 5 | 11 | 6,10E-04 | 39,0% |
| Ársáv | 6 | 3 | 6,08E-04 | 39,1% |
| Ársáv (%) | 7 | 3 | 6,09E-04 | 39,1% |
| Tényleges ársáv | 6 | 0 | 6,08E-04 | 39,1% |
| Tényleges ársáv (%) | 8 | 0 | 6,08E-04 | 39,2% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

99. táblázat. A Poli(14)_ARX_Gr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,01)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_Gr küszöb: 0,05 | Hány részvényt nyert | | Átlagos érték | |
|---------------------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 4 | 0 | 6,09E-04 | 39,1% |
| Volatilitás | 7 | 0 | 6,07E-04 | 39,2% |
| Rés | 3 | 9 | 6,10E-04 | 39,0% |
| Ársáv | 6 | 4 | 6,08E-04 | 39,1% |
| Ársáv (%) | 7 | 4 | 6,08E-04 | 39,1% |
| Tényleges ársáv | 5 | 0 | 6,08E-04 | 39,2% |
| Tényleges ársáv (%) | 7 | 0 | 6,09E-04 | 39,2% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

100. táblázat. A Poli(14)_ARX_Gr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,05)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_korr küszöb: 0,1 | Hány részvénynél nyert | | Átlagos érték | |
|----------------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 3 | 0 | 6,08E-04 | 39,1% |
| Volatilitás | 9 | 0 | 6,05E-04 | 39,2% |
| Rés | 2 | 8 | 6,10E-04 | 39,0% |
| Ársáv | 8 | 10 | 6,03E-04 | 39,1% |
| Ársáv (%) | 7 | 11 | 6,03E-04 | 39,1% |
| Tényleges ársáv | 7 | 4 | 6,03E-04 | 39,1% |
| Tényleges ársáv (%) | 8 | 4 | 6,04E-04 | 39,1% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

101. táblázat. A Poli(14)_ARX_korr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,1)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_korr küszöb: 0,2 | Hány részvénynél nyert | | Átlagos érték | |
|----------------------------------|------------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 7 | 1 | 6,08E-04 | 39,1% |
| Volatilitás | 14 | 0 | 6,05E-04 | 39,2% |
| Rés | 5 | 7 | 6,10E-04 | 39,0% |
| Ársáv | 8 | 14 | 6,03E-04 | 39,0% |
| Ársáv (%) | 10 | 14 | 6,03E-04 | 39,0% |
| Tényleges ársáv | 9 | 7 | 6,03E-04 | 39,1% |
| Tényleges ársáv (%) | 10 | 8 | 6,04E-04 | 39,1% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

102. táblázat. A Poli(14)_ARX_korr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,2)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_korr küszöb: 0,4 | Hány részvényt nyert | | Átlagos érték | |
|----------------------------------|----------------------|------|---------------|-------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 2 | 1 | 6,07E-04 | 39,0% |
| Volatilitás | 6 | 6 | 6,08E-04 | 39,0% |
| Rés | 3 | 4 | 6,07E-04 | 39,0% |
| Ársáv | 7 | 13 | 6,09E-04 | 39,0% |
| Ársáv (%) | 9 | 13 | 6,09E-04 | 39,0% |
| Tényleges ársáv | 7 | 9 | 6,09E-04 | 39,1% |
| Tényleges ársáv (%) | 8 | 9 | 6,09E-04 | 39,1% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

103. táblázat. A Poli(14)_ARX_korr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,4)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_korr küszöb: 0,5 | Hány részvényt nyert | | Átlagos érték | |
|----------------------------------|----------------------|------|---------------|--------------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 1 | 0 | 6,07E-04 | 39,0% |
| Volatilitás | 7 | 5 | 6,08E-04 | 39,0% |
| Rés | 0 | 2 | 6,07E-04 | 39,0% (jobb) |
| Ársáv | 9 | 13 | 6,09E-04 | 39,0% |
| Ársáv (%) | 10 | 15 | 6,09E-04 | 39,0% |
| Tényleges ársáv | 4 | 8 | 6,09E-04 | 39,0% |
| Tényleges ársáv (%) | 8 | 7 | 6,09E-04 | 39,0% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

104. táblázat. A Poli(14)_ARX_korr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,5)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_korr küszöb: 0,6 | Hány részvényt nyert | | Átlagos érték | |
|----------------------------------|----------------------|------|---------------|--------------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 0 | 0 | 6,07E-04 | 39,0% |
| Volatilitás | 4 | 4 | 6,07E-04 | 39,0% |
| Rés | 0 | 1 | 6,07E-04 | 39,0% (jobb) |
| Ársáv | 7 | 7 | 6,09E-04 | 39,0% |
| Ársáv (%) | 8 | 7 | 6,08E-04 | 39,0% |
| Tényleges ársáv | 4 | 5 | 6,08E-04 | 39,0% |
| Tényleges ársáv (%) | 17 | 6 | 6,09E-04 | 39,0% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

105. táblázat. A Poli(14)_ARX_korr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,6)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_korr küszöb: 0,7 | Hány részvényt nyert | | Átlagos érték | |
|----------------------------------|----------------------|------|---------------|--------------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 0 | 0 | megegyezik | megegyezik |
| Volatilitás | 0 | 1 | 6,07E-04 | 39,0% |
| Rés | 0 | 1 | 6,07E-04 | 39,0% (jobb) |
| Ársáv | 0 | 0 | 6,07E-04 | 39,0% |
| Ársáv (%) | 1 | 3 | 6,08E-04 | 39,0% |
| Tényleges ársáv | 1 | 1 | 6,07E-04 | 39,0% |
| Tényleges ársáv (%) | 2 | 3 | 6,07E-04 | 39,0% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

106. táblázat. A Poli(14)_ARX_korr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,7)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_korr küszöb: 0,8 | Hány részvénynél nyert | | Átlagos érték | |
|----------------------------------|------------------------|------|---------------|------------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 0 | 0 | megegyezik | megegyezik |
| Volatilitás | 0 | 0 | megegyezik | megegyezik |
| Rés | 0 | 0 | megegyezik | megegyezik |
| Ársáv | 0 | 0 | 6,07E-04 | 39,0% |
| Ársáv (%) | 0 | 0 | 6,07E-04 | 39,0% |
| Tényleges ársáv | 0 | 0 | 6,07E-04 | 39,0% |
| Tényleges ársáv (%) | 0 | 0 | 6,07E-04 | 39,0% |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

107. táblázat. A Poli(14)_ARX_korr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,8)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

| Poli(14)_ARX_korr küszöb: 0,9 | Hány részvénynél nyert | | Átlagos érték | |
|----------------------------------|------------------------|------|---------------|------------|
| | MSE | MAPE | MSE | MAPE |
| Loghozam | 0 | 0 | megegyezik | megegyezik |
| Volatilitás | 0 | 0 | megegyezik | megegyezik |
| Rés | 0 | 0 | megegyezik | megegyezik |
| Ársáv | 0 | 0 | megegyezik | megegyezik |
| Ársáv (%) | 0 | 0 | megegyezik | megegyezik |
| Tényleges ársáv | 0 | 0 | megegyezik | megegyezik |
| Tényleges ársáv (%) | 0 | 0 | megegyezik | megegyezik |
| Poli(14)_AR | - | - | 6,07E-04 | 39,0% |

108. táblázat. A Poli(14)_ARX_korr modellek értékelése
különböző ármozgás mutatók használata esetén (küszöb: 0,9)

Összevetés a Poli(14)_AR modellel

Forrás: Saját szerkesztés

Az elemzésbe bevont részvények árszintje

A 109. táblázatban a felhasznált részvények áráiról találunk összehasonlító információt a teljes időszakra vonatkoztatva. A mellékletben a 96-98. táblázatban megtalálhatóak a kódokhoz tartozó vállalatok nevei is.

| Kód | Arány | max/min | Kód | Arány | max/min | Kód | Arány | max/min |
|-----|-------|---------|-----|-------|---------|------|-------|---------|
| AIG | 46,1 | 244,9 | WMT | 3,4 | 2,0 | DIS | 1,9 | 4,1 |
| C | 17,6 | 52,0 | KO | 3,2 | 2,8 | KFT | 1,9 | 2,2 |
| IBM | 7,3 | 4,4 | MCD | 3,1 | 10,5 | BAC | 1,9 | 16,9 |
| MMM | 4,7 | 2,6 | AXP | 2,8 | 6,7 | MSFT | 1,7 | 2,4 |
| CVX | 4,2 | 5,0 | HON | 2,6 | 4,2 | GE | 1,6 | 6,4 |
| XOM | 4,0 | 3,6 | DD | 2,5 | 3,9 | T | 1,6 | 3,2 |
| BA | 3,8 | 4,7 | JPM | 2,4 | 4,1 | AA | 1,5 | 9,1 |
| UTX | 3,7 | 4,4 | MRK | 2,3 | 2,8 | CSCO | 1,4 | 4,2 |
| CAT | 3,7 | 8,6 | HPQ | 2,2 | 5,4 | INTC | 1,4 | 2,8 |
| JNJ | 3,7 | 2,1 | HD | 2,1 | 3,4 | PFE | 1,4 | 2,9 |
| PG | 3,5 | 2,5 | VZ | 1,9 | 2,9 | MO | 1,0 | 9,2 |

109. táblázat. A részvények átlagos ára az Altria Group Inc. arányában (Arány), valamint az egyes részvények maximális és minimális árának az aránya (max/min)

Mindkét mutató a teljes vizsgált időszakra számolva

Forrás: Saját szerkesztés

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.