

MÁK FRUzsina

**FOGYASZTÁSI KOCKÁZAT
A VILLAMOSENERGIA-PIACON**

Statisztika Tanszék

Témavezetők:

Dr. Oravecz Beatrix, egyetemi adjunktus, Ph.D.

Dr. Sugár András, egyetemi docens, tanszékvezető, Ph.D.

Budapesti Corvinus Egyetem
Gazdálkodástani Doktori Iskola

Fogyasztási kockázat a villamosenergia-piacon

Profilozás a fogyasztási bizonytalanság figyelembevétele mellett

doktori értekezés

Mák Fruzsina

Budapest, 2017

TARTALOMJEGYZÉK

| | |
|--|----|
| BEVEZETÉS | 2 |
| 1. A FOGYASZTÓI PROFILOZÁS FELHASZNÁLÁSI TERÜLETEI ÉS FONTOSABB ALAPFOGALMAI..... | 12 |
| 1.1. Ár- és mennyiségi bizonytalanság a villamosenergia-piacon | 12 |
| 1.2. A profilozás legfontosabb gyakorlati felhasználási területei | 19 |
| 1.2.1. Rövid- és hosszú távú fedezés illetve árazás..... | 19 |
| 1.2.2. Fogyasztói befolyásolás | 22 |
| 1.2.3. Portfólióépítés illetve mérlegkör-összeállítás | 24 |
| 1.3. A profil és profillal kapcsolatos kockázatok | 26 |
| 1.3.1. A fogyasztói profil definíciója | 26 |
| 1.3.2. Fogyasztással kapcsolatos kockázatok..... | 28 |
| 1.4. A fogyasztói idősorok stilizált tényeinek empirikus vizsgálata | 30 |
| 1.4.1. Terhelési alak és szint | 31 |
| 1.4.2. Terhelési értékek napon belüli eloszlása..... | 35 |
| 1.4.3. Időjárásfüggés | 40 |
| 1.4.4. Következtetések | 42 |
| 2. KORÁBBI KUTATÁSI EREDMÉNYEK A FOGYASZTÓI PROFILOZÁS TERÜLETÉRŐL | 44 |
| 2.1. Fogyasztói idősorok kétlépéses klaszterezése | 44 |
| 2.1.1. Idősorok klaszterezése általában | 44 |
| 2.1.2. A profilozás általános keretrendszere | 46 |
| 2.1.3. A profilozás során használt görbetulajdonságok előállítása | 47 |
| 2.1.4. A profilozás során használt klaszterezési algoritmusok..... | 49 |
| 2.2. Az időjárást leíró változók kezelése energetikai idősorok esetében | 51 |
| 2.2.1. Az időjárási változók és a fogyasztás kapcsolata..... | 52 |
| 2.2.2. Az időjárási változók kezelése a profilozás során..... | 54 |
| 2.2.3. A hőmérsékleti véletlen hatás szerepe | 56 |
| 2.2.4. A hazai földgázfogyasztás extrém (véletlen) hőmérsékleti hatásoktól történő megtisztítása és szezonális kiigazítása | 57 |
| 3. AZ ÉRTEKEZÉSBEN HASZNÁLT MÓDSZEREK ÉS PROFILOZÁSI CÉLÚ ALKALMAZÁSAIK BEMUTATÁSA | 68 |
| 3.1. Klasszikus sztochasztikus idősoros regressziós modellek | 68 |

| | | |
|--------|--|-----|
| 3.1.1. | A stacionaritás fogalma és tesztelése | 69 |
| 3.1.2. | A véletlen szerepe integrált idősorok esetében | 69 |
| 3.1.3. | A véletlen szerepe stacioner idősorok esetében | 70 |
| 3.1.4. | SARMA-modell | 72 |
| 3.1.5. | PAR-modell (Periodikus autoregresszív modell) | 73 |
| 3.2. | Keverék-modellek | 76 |
| 3.2.1. | A keverék-modell (MM) és a <i>Gauss</i> -féle keverék-modell (GMM) felírása. | 77 |
| 3.2.2. | Expectation-Maximization (EM) becslési eljárás | 78 |
| 3.2.3. | Empirikus példa Budapest napi földgázfogyasztási adatai példáján | 80 |
| 3.2.4. | A <i>Gauss</i> -féle keverék-moddell kapcsolatos további módszertani kérdések | 84 |
| 3.2.5. | A <i>Gauss</i> -féle keverék-modell regressziós alkalmazása (GMR) | 86 |
| 3.2.6. | A <i>Gauss</i> -féle keverék-regresszió alkalmazása idősorok esetén..... | 90 |
| 3.3. | Keverék-modellek és energetikai alkalmazásai | 92 |
| 3.3.1. | Napi fogyasztási görbék tipizálása..... | 92 |
| 3.3.2. | Fogyasztás eloszlásának modellezése keverék-sűrűségfüggvénnyel..... | 95 |
| 3.3.3. | Fogyasztás eloszlásának modellezése keverék-sűrűségfüggvénnyel és regressziós alkalmazással | 96 |
| 4. | A FOGYASZTÁS BIZONYTALANSÁGÁNAK FIGYELEMBE VÉTELE A PROFILOZÁSBAN – EMPIRIKUS EREDMÉNYEK | 98 |
| 4.1. | Tipikus fogyasztási mintázatok készítése..... | 98 |
| 4.1.1. | Keverék-modell alkalmazása tipikus fogyasztási mintázat készítésére | 99 |
| 4.1.2. | Klasszikus idősoros regresszió alkalmazása tipikus fogyasztási mintázat készítésére..... | 108 |
| 4.1.3. | Profilcsoportok képzése | 112 |
| 4.1.4. | Eredmények, következtetések összegzése | 118 |
| 4.2. | A fogyasztás bizonytalanságának modellezése | 120 |
| 4.2.1. | Volumenkockázat a klasszikus idősoros regressziós modellekben | 120 |
| 4.2.2. | Volumenkockázat modellezése keverék-regresszióval..... | 131 |
| | ÖSSZEFOGLALÁS – AZ ÉRTEKEZÉS FŐBB MEGÁLLAPÍTÁSAI, EREDMÉNYEI | 153 |
| A) | Fogyasztási idősorok stilizált tényeinek vizsgálata..... | 153 |
| B) | Keverék-modell alkalmazása tipikus fogyasztási mintázatok készítésére | 154 |
| C) | Heurisztikus és klasszikus sztochasztikus idősoros módszerek alkalmazása a fogyasztás bizonytalanságának mérésére | 157 |

| | | |
|--|--|-----|
| D) | Keverék-modell alkalmazása a fogyasztás bizonytalanságának mérésére..... | 158 |
| TOVÁBBFEJLESZTÉSI LEHETŐSÉGEK, AZ EREDMÉNYEK HASZNOSÍTHATÓSÁGA | | 161 |
| MELLÉKLETEK | | 165 |
| A) | A számításokhoz használt programcsomagok illetve fontosabb funkciók..... | 165 |
| B) | Empirikus példa Budapest napi földgázfogyasztási adatai példáján – számítási részeredmények | 167 |
| C) | Terhelési idősorok stilizált tényeinek vizsgálata | 168 |
| D) | SI-arányok a hazai földgázfogyasztás szezonális kiigazítása során | 170 |
| E) | Görbék tipikus napi profiljai és heti idősoros ábrái | 171 |
| F) | A volumenkockázat diverzifikálhatósága | 173 |
| IRODALOMJEGYZÉK | | 175 |
| SAJÁT PUBLIKÁCIÓK JEGYZÉKE | | 180 |
| A) | A dolgozat témakörével kapcsolatos magyar nyelvű publikációk | 180 |
| B) | A dolgozat témakörével kapcsolatos idegen nyelvű publikációk..... | 180 |
| C) | A dolgozat témakörével kapcsolatos, nem publikált munkák | 180 |

ÁBRAJEGYZÉK

| | |
|--|-----|
| 1. ábra: Kiegyenlítő energia és másnapi órás árak alakulása | 3 |
| 2. ábra: Egy vállalati portfólió terhelési idősorának szintvonalábrája | 6 |
| 3. ábra: Standard hibák átlaga téli hétköznapiokon egy fogyasztói görbe terhelési idősora esetén | 7 |
| 4. ábra: Keverék-klaszterezés eredménye a napi középhőmérséklet-gázfogyasztás példán | 9 |
| 5. ábra: Kereskedés villamosenergia-piaci termékekkel | 13 |
| 6. ábra: Erőművi <i>merit order</i> ábrázolása | 13 |
| 7. ábra: Magyar másnapi órás villamosenergia-árak historikus alakulása | 14 |
| 8. ábra: A másnapi órás és a kiegyenlítő árak historikus alakulása | 17 |
| 9. ábra: Menetrend megállapításának sematikus szemléltetése | 20 |
| 10. ábra: Fogyasztói igények részleges fedezése zsinór- és csúcsidőszaki termékekkel | 21 |
| 11. ábra: Fogyasztás-megtakarítás szignifikanciájának értékelése | 23 |
| 12. ábra: Fogyasztás-átcsoportosítás szignifikanciájának értékelése | 23 |
| 13. ábra: Portfólióhatás sematikus ábrázolása | 25 |
| 14. ábra: Profillal kapcsolatos kockázatok sematikus ábrázolása | 29 |
| 15. ábra: A terhelési görbék alakját meghatározó tulajdonságok | 31 |
| 16. ábra: A magyar országos rendszerterhelés szintvonalábrája és néhány kiválasztott nap idősoros ábrája | 32 |
| 17. ábra: Portfóliók és egyedi fogyasztók terhelési idősorainak szintvonalábrái | 34 |
| 18. ábra: Portfóliók és egyedi fogyasztók heti terhelési idősorai és a kapcsolódó dobozábrák | 38 |
| 19. ábra: Portfóliók és egyedi fogyasztók terhelése a hőmérséklet függvényében | 41 |
| 20. ábra: Fogyasztói görbék profilozási folyamatábrája | 46 |
| 21. ábra: A magyarországi földgázfogyasztás a hőmérséklet függvényében, 2006–2013 | 57 |
| 22. ábra: A magyarországi földgázfogyasztás és a HDD alakulása, 2006-2013 | 59 |
| 23. ábra: A szezonális kiigazítás végeredménye HDD-eltérések használata nélkül és HDD-eltérések használatával | 61 |
| 24. ábra: Az SI-arányok HDD-eltérések használata nélkül és HDD-eltérések használatával | 62 |
| 25. ábra: A hőmérséklettel kiigazított földgázfogyasztás alakulása, 2006–2013 | 64 |
| 26. ábra: A hőmérséklettel kiigazított földgázfogyasztás alakulása gázév szerint, 2006–2013 | 65 |
| 27. ábra: Reziduumok alakulása egy SARIMA-modellben | 71 |
| 28. ábra: K-középpontú klaszterezés pszeudo-kódja | 80 |
| 29. ábra: Budapesti napi gázfogyasztás és napi középhőmérséklet alakulása illetve középhőmérséklet-gázfogyasztás <i>scatter plot</i> | 81 |
| 30. ábra: K-középpontú klaszterezés eredményei a napi középhőmérséklet-gázfogyasztás példán | 82 |
| 31. ábra: Keverék-klaszterezés eredményei a napi középhőmérséklet-gázfogyasztás példán | 83 |
| 32. ábra: Keverék-sűrűségfüggvény illesztése napi terhelési görbére (V109 görbe) | 94 |
| 33. ábra: Keverék-sűrűségfüggvény illesztése napi terhelési görbére (V148 görbe) | 94 |
| 34. ábra: Keverék-sűrűségfüggvény illesztése terhelési értékek empirikus eloszlására | 96 |
| 35. ábra: Keverék-modell becslése és változó páronkénti illeszkedés ellenőrzése | 96 |
| 36. ábra: Napfelkelte illetve naplemente időpontjától számított előjeles, négyzetes eltérések alakulása | 101 |
| 37. ábra: Keverék-modell becslése és változó páronkénti illeszkedés ellenőrzése a vállalati portfólió esetén | 104 |
| 38. ábra: Komponensek összetétele a vállalati portfólió esetén | 105 |
| 39. ábra: Komponensek közötti távolság alapján készült dendrogram a vállalati portfólió esetén | 105 |
| 40. ábra: Komponensek összetétele és a komponensek közötti távolság alapján készült dendrogramok egyedi görbék esetén | 107 |
| 41. ábra: A hőmérséklet és naplemente hatása a terhelésre a vállalati portfólió esetén | 109 |
| 42. ábra: Vállalati portfólió és V109 görbe napi lefutásai illetve a tipikus napi profilok (TDP-k) alakulása | 111 |
| 43. ábra: Normalizált tipikus napi profilok (TDP-k) alakulása egyedi görbék esetén | 111 |
| 44. ábra: Távolságmátrixok a TDP alapú klaszterezés illetve a keverék-klaszterezés esetén | 114 |
| 45. ábra: Görbék közötti távolság alapján készült dendrogram | 114 |
| 46. ábra: Görbék közötti távolság alapján készült dendrogram (kibővített példa) | 116 |
| 47. ábra: Normalizált tipikus napi profilok (TDP-k) klaszterezésének eredménye | 117 |
| 48. ábra: Kockázati indexek alakulása a vállalati portfólió esetén | 122 |
| 49. ábra: SARMA(1, 0)(1, 0) ₉₆ -modell reziduumainak alakulása a vállalati portfólió esetén | 126 |

| | |
|--|-----|
| 50. ábra: SARMA(1, 0)(1, 0) ₉₆ -modell reziduumainak alakulása a vállalati portfólió esetén néhány kiválasztott hétre | 126 |
| 51. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya a vállalati portfólió esetén (klasszikus regressziós modellek) | 128 |
| 52. ábra: Reziduumok szórása a SARMA(1, 0)(1, 0) ₉₆ -modellben a vállalati portfólió esetén..... | 129 |
| 53. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya a mintában és mintán kívül a vállalati portfólió esetén (keverék-regresszió) | 133 |
| 54. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a vállalati portfólió esetén (hétköznapi)..... | 135 |
| 55. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a fogyasztói portfólió esetén (hétvégi napok) | 135 |
| 56. ábra: Átlagos konfidencia-intervallum SARMA- és keverék-modellek esetén az egyedi görbék esetén | 139 |
| 57. ábra: Átlagos konfidencia-intervallum SARMA- és keverék-modellek esetén az egyedi görbék esetén (kibővített ábra)..... | 139 |
| 58. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya az egyedi görbék esetében (SARMA-modell) | 142 |
| 59. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya az egyedi görbék esetében (keverék-regresszió)..... | 143 |
| 60. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya az egyedi görbék esetében (SARMA-modell) | 144 |
| 61. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya az egyedi görbék esetében (keverék-regresszió)..... | 145 |
| 62. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V25 görbe esetében (hétköznapi)..... | 146 |
| 63. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V25 görbe esetében (hétvégi napok)..... | 146 |
| 64. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V66 görbe esetében (hétköznapi)..... | 147 |
| 65. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V66 görbe esetében (hétvégi napok)..... | 147 |
| 66. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V109 görbe esetében (hétköznapi)..... | 148 |
| 67. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V109 görbe esetében (hétvégi napok)..... | 148 |
| 68. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V108 görbe esetében (hétköznapi)..... | 149 |
| 69. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V108 görbe esetében (hétvégi napok)..... | 149 |
| 70. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V47 görbe esetében (hétköznapi)..... | 150 |
| 71. ábra: Keverék-regresszió reziduumainak szórása és a feltételes szórások átlaga a mintában a V47 görbe esetében (hétvégi napok)..... | 150 |
| 72. ábra: SI-arányok különböző szezonális kiigazítási modellváltozatok esetén | 170 |
| 73. ábra: Egyedi fogyasztói görbék és a vállalati portfólió tipikus napi profiljai | 171 |
| 74. ábra: Egyedi fogyasztói görbék és a vállalati portfólió heti időszaki szezononként | 172 |
| 75. ábra: Lineáris korrelációs együtthatók értéke a V35 egyedi görbe és a vállalati portfólió standardizált reziduumai között..... | 173 |

TÁBLÁZATJEGYZÉK

| | |
|--|-----|
| 1. táblázat: Magyar országos rendszerterhelés és egy egyedi fogyasztói görbe leíró statisztikái | 5 |
| 2. táblázat: A hazai szélenergia-termelésének néhány leíró statisztikája | 15 |
| 3. táblázat: Idősoros klaszterezési algoritmusok lehetséges csoportosításai | 45 |
| 4. táblázat: Profilozás során előállított görbetulajdonságok lehetséges csoportosítása | 47 |
| 5. táblázat: Profilozás során használt klaszterezési algoritmusok | 50 |
| 6. táblázat: A regressziós modell főbb eredményei a hazai földgázfogyasztás példáján | 60 |
| 7. táblázat: A HDD-eltérés paramétereinek értéke és származtatott értelmezése a hazai földgázfogyasztás példáján | 63 |
| 8. táblázat: A regresszió során használt magyarázóváltozók és rövid leírásuk (PAR-modell) | 108 |
| 9. táblázat: Fogyasztói görbék portfóliótól vett távolsága | 115 |
| 10. táblázat: Becsült paraméterek száma a komponensek függvényében a keverék-modell esetén | 118 |
| 11. táblázat: Modellszelekciós kritériumok értéke és becsült paraméterek száma a keverék-modellek esetén | 118 |
| 12. táblázat: A klasszikus idősoros regresszió során használt magyarázóváltozók és rövid leírásuk | 123 |
| 13. táblázat: A klasszikus idősoros regressziós modellek illeszkedését jellemző mutatószámok alakulása a vállalati portfólió esetén | 124 |
| 14. táblázat: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya és átlagos konfidencia-intervallum nagysága a vállalati portfólió esetén (klasszikus regressziós modellek) | 127 |
| 15. táblázat: A modellek illeszkedését jellemző mutatószámok alakulása a vállalati portfólió esetén | 132 |
| 16. táblázat: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya és átlagos konfidencia-intervallum nagysága a vállalati portfólió esetén | 132 |
| 17. táblázat: A modellek illeszkedését jellemző mutatószámok alakulása az egyedi görbék esetén | 136 |
| 18. táblázat: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya és átlagos konfidencia-intervallum nagysága az egyedi görbék esetén | 137 |
| 19. táblázat: BSS/TSS hányadosok alakulása k-középpontú klaszterezés esetén a napi középhőmérséklet-gázfogyasztás példán | 167 |
| 20. táblázat: BIC mutató alakulása keverék-klaszterezés esetén a napi középhőmérséklet-gázfogyasztás példán | 167 |
| 21. táblázat: Klasztercentroidok keverék-klaszterezés esetén a napi középhőmérséklet-gázfogyasztás példán | 167 |
| 22. táblázat: Napok klasztereken belüli megoszlása a különböző hónapok esetén a napi középhőmérséklet-gázfogyasztás példán (keverék-klaszterezés) | 167 |
| 23. táblázat: Napok klasztereken belüli megoszlása a hétköznapi / hétvégi esetén a napi középhőmérséklet-gázfogyasztás példán (keverék-klaszterezés) | 167 |
| 24. táblázat: A magyar rendszerterhelés heti idősorainak leíró statisztikái napi bontásban | 168 |
| 25. táblázat: Szezonális ismérvek által megmagyarázott varianciarányad görbénként | 169 |

Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani a témavezetőimnek, Oravecz Beatrixnak és Sugár Andrásnak, akik fáradhatatlanul segítettek a doktori képzés során, és a dolgozat elkészüléséhez rengeteg bátorítással és tanáccsal járultak hozzá – szakmailag és emberileg egyaránt.

Szeretném megköszönni előopponenseimnek, Fazakas Gergelynek és Mádi-Nagy Gergelynek a dolgozat tervezetének a bírálata során nyújtott értékes észrevételeit, javaslatait. Köszönettel tartozom Jenei Gabriellának, aki rengeteg lelkesítéssel és segítséggel támogatta a dolgozat végső formájának elkészülését.

Hálás vagyok több jelenlegi és egykori Kollégámnak, amiért szakmailag segítettek és biztattak a dolgozat elkészítése során.

Külön köszönöm a Családomnak a rengeteg türelmet és lelki támogatást, amivel sok-sok éve segítik a munkámat.

Köszönettel tartozom a Barátaimnak, akik bátorító-támogató szavaikkal hozzájárultak a dolgozat elkészüléséhez.

BEVEZETÉS

Az energiapiaci szereplők rövid és hosszú távon, az operatív és stratégiai döntések meghozatala során egyaránt rengeteg kockázattal szembesülnek. A kockázatok egy jelentős részének a kezelése, mérése a pénzügyi piacokon is ismert megoldásokkal együtt fejlődött ki, vagy azokra épülve adaptálták az energiapiac sajátosságaihoz.

A **liberalizáció** előrehaladásával párhuzamosan az európai uniós célok között az **energiahatékonyság**, az **energiatakarékosság**, az ezek nyomán elérhető **fogyasztás-csökkentés** és **-megtakarítás** folyamatos, minél eredményesebb realizálása szerepel. Mindeközben az **okos mérés** terjedésén keresztül az energiafelhasználás rendszeres időközönkénti mérését, sok esetben *online* nyomon követését lehetővé tevő különböző feltételek is megvalósulnak. Az ilyen, alapvetően mikroszintű (egyedi fogyasztó szintjén értelmezhető) tendenciák mellett (vagy inkább mögött, azokkal összefüggésben) vannak makroszintűek (rendszerszinten értelmezhetőek) is, amelyek például a rendszer-szintű szabályozási nehézségek kezelésében, vagy éppen a rendszerszintű költségek, veszteségek csökkentésére való törekvésekben manifesztálódnak.

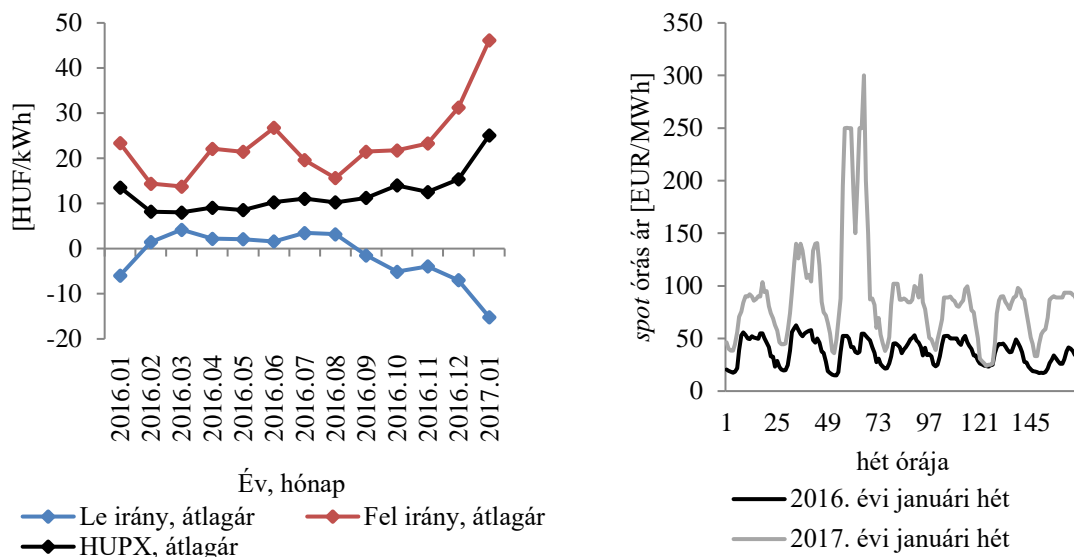
Annak ellenére, hogy az energiapiaci kockázatok egyik legjelentősebb forrása alapvetően az ár, a fentiek nyomán nemcsak a portfóliószintű, hanem az egyedi fogyasztói viselkedés ismerete is egyre fontosabbá válik és egyre nagyobb üzleti értéke van a szolgáltatók, a fogyasztók, de a rendszerüzemeltetők szempontjából is.¹ Az értekezés témájához kapcsolódó **villamosenergia**-piacon – ahogy más energiapiacokon is – egyre több az olyan gyakorlati alkalmazási terület, ahol a **fogyasztás (várható)** alakulásának az ismerete nem elegendő, hanem szükséges annak **bizonytalanságát** is figyelembe venni, és az ebből a bizonytalanságból adódó kockázatot kezelni.

Ilyen terület többek között például a portfóliószintű villamosenergia-igény meghatározása (menetrendkészítés), a portfólió hosszú távú fedezése, vagy az egyedi fogyasztók esetében kalkulált árak megállapítása is. Ezek a példák természetesen összefüggnek, egyrészt keresztmetszetileg (az egyedi fogyasztói görbék összege a portfóliószintű görbe), másrészt idősorosan (a menetrendezés során elkövetett előrejelzési hibák eredményezte kiegyenlítő energia költsége rakódik rá a portfólióra az üzleti év során).

¹ Az, hogy pontosan mekkora, a mindenkor energiapiaci helyzetnek, például az energiapiaci szabályozásnak és politikai döntésnek is függvénye, hiszen alacsony, nyomott árak mellett a fogyasztói takarékos-ság elősegítése, ösztönzése nehéz.

A téma fontosságát **pénzügyi** oldalról is megvilágítva, vizsgáljuk meg a másnapi órás *spot* és a kiegyenlítő energia árak² alakulását Magyarországon 2016 januárja óta.

1. ábra: Kiegyenlítő energia és másnapi órás árak alakulása



Forrás: HUPX Zrt.³ és MAVIR Zrt.⁴ adatok alapján (részben) saját számítások és saját készítésű ábra.

Csak a 2016. és 2017. januári hónapok egy-egy hetének másnapi órás *spot* árait kiemelve (ld. az 1. ábra jobb oldalát) érzékelhető, hogy a kínálati oldali sokkok (például erőműkiesés, olcsó importáram hiánya, stb.) hogyan drágíthatják meg a völgy-, de különösen a csúcsidőszaki fogyasztást. Emellett havi átlagértékeket vizsgálva is egyértelmű, hogy a túl- vagy alulfogyasztásból adódóan a menetrendtől vett tény-terv eltérések utólagos rendezésére szolgáló vételi (fel irány) illetve eladási (le irány) kiegyenlítő átlagárak a *spot* árakhoz képest milyen pozíciót jelentenek (ld. az 1. ábra bal oldalát), és mennyire fontos a fogyasztással kapcsolatos mennyiségi bizonytalanság értékelése is.

A kínálati sokkok között mindenképpen érdemes említést tenni a megújuló energiatermelőkről (a térnyerésük politikai csatározásoktól egyáltalán nem mentes), amelyek az időjárás-függőségből származó termelési bizonytalanság miatt nemcsak az áralakulás változékonyságát növelik, de a villamosenergia-rendszer egyensúlytalansága tekintetében mennyiségi kockázatot is jelentenek.

² A bevezetőben megjelenő fogalmak, definíciók pontos magyarázatát ld. az értekezés későbbi fejezeteiben.

³ Adatok forrása: www.hupx.hu, Hungarian Power Exchange (vagy röviden HUPX) Zrt., a magyar villamosenergia-tőzsdét üzemeltető társaság.

⁴ Adatok forrása: www.mavir.hu, Magyar Villamosenergia-ipari Átviteli Rendszerirányító Zrt., a magyar átviteli rendszerirányítói engedélyes (Transmission System Operator, röviden TSO).

Magas villamosenergia-árat – gondoljunk a fogyasztás viszonylag nagy árugátlanságára – persze kereslet oldali sokkok is okozhatnak⁵, bár azok *ceteris paribus* kevésbé eredményeznek igazán extrém viselkedést az árak vagy a mennyiségek alakulása tekintetében.

Mivel a villamosenergia-rendszer minden időpillanatban fennálló **kereslet-kínálati egyensúly**a sok esetben nehezen és/vagy drágán biztosítható a kínálati (erőművi) oldal szabályozásával, ezért egyre inkább előtérbe kerül nemcsak az egyedi fogyasztói szokásoknak, hanem a fogyasztás bizonytalanságának a kereslet-kínálat egyensúly megvalósításában betöltött lehetséges szerepe; hiszen a fogyasztási szokások valamennyire alakíthatóak, irányíthatóak.

Utóbbi korrekt számszerűsítésének az igénye sokkal hangsúlyosabb a fejlettebb, „előrébb járó” piacokon. Példaként említhetjük többek között a fogyasztói befolyásolással (*demand side management*) kapcsolatos tevékenységeket. Ezeken a piacokon rövidtávon például a tarifális ösztönzéssel elérni kívánt fogyasztáscsökkentés (itt többek között a fogyasztónak a szabályozásban történő részvételéről van szó), vagy a hosszabb távon garantálható fogyasztás-megtakarítás, ahhoz kapcsolódó beruházási döntések vizsgálata kiemelt fontosságú. Ezeken a területeken a bizonytalanság explicit figyelembevétele semmiképpen sem kerülhető meg, hiszen új árazási logikákat is érintő kérdésekről van szó.

A felsorolt gyakorlati feladatokat tekintve nem akarunk egyik esetben sem teljes körű választ adni, hiszen az bőven meghaladná az értekezés kereteit. A cél sokkal inkább a felsorolt kihívásokra adott válaszokhoz történő **hozzájárulás** biztosítása **a fogyasztás bizonytalanságának módszertanilag is jól megalapozott figyelembevételével**.

A fogyasztási bizonytalanság nagyságrendjével kapcsolatban tekintsük az alábbi nagyon egyszerű leíró statisztikai mutatószámokat a magyar országos rendszerterhelés, illetve egy egyedi fogyasztói görbe tekintetében (ld. az 1. táblázatot).

Egyedi görbék esetében természetesen jóval nagyobb a véletlen szerepe, így a szóródást jellemző mutatószámok (a kockázat mérésére sokszor használt szórás, vagy a maximális és minimális értékek különbségeként adódó terjedelem) arányaiban fogyasztónként eltérő mértékben, de jellemzően magasabbak. Az ilyen számítási eredmények

⁵ Nagy nyári melegekben a légkondicionáló készülékek intenzívebb használata végett alakulhatnak ki extrém fogyasztási csúcsok. A 2015-ös év volt például az első olyan év, amikor az éven belüli nyári csúcs meghaladta az éven belüli téli csúcsterhelést (6 456 vs. 6 447 [MW]).

természetesen csak nagyon durva közelítésnek felelnek meg, és semmilyen fundamentális indokolást nem tartalmaznak.

1. táblázat: Magyar országos rendszerterhelés és egy egyedi fogyasztói görbe leíró statisztikái

| Idősor | Átlag | Szórás | Minimum | Maximum |
|-------------------------------------|--------|--------|---------|---------|
| Rendszerterhelés [MW] | 4933,7 | 682,7 | 3047,7 | 6486,6 |
| mutatószámok az átlag arányában [%] | | 13,8 | 61,8 | 131,5 |
| Egyedi görbe [kW] | 8,0 | 5,3 | 2,9 | 25,8 |
| mutatószámok az átlag arányában [%] | | 65,5 | 35,8 | 321,4 |

Forrás: saját számítások és saját készítésű táblázat.

Mint az a különböző kockázatokkal foglalkozó szakirodalmakban megszokott, a kockázat számszerűsítése jellemzően pénzben történik, ami az energiapiacokon is kézenfekvő (hiszen akár egy nagykereskedelmi ügyleten, akár egy fogyasztóval kötött kiskereskedelmi szerződésen – pénzben mérhető – veszteség keletkezhet). Az értekezés témája tekintetében azonban a pénzügyi mellett a fizikai folyamatok önmagukban is értelmezhetők és értékelhetők, így az elemzések során elsősorban a fogyasztási (terhelési) idősorokra koncentrálnak. A külföldi, főként angol nyelvű szakirodalmat tekintve elsősorban a terhelési görbe (*load curve*) fogalom használata terjedt el a fogyasztási görbe (*consumption curve*) kifejezéssel szemben. Mértékegységbeli különbség is van a kettő között, a terhelést az értekezésben jellemzően [kW]-ban, a fogyasztást [kWh]-ban vizsgáljuk, de az eredmények, következtetések a fogalomhasználatától és a mértékegységtől egyaránt függetlenek.^{6,7}

Az értekezés elkészítésével a **célunk** az, hogy a fogyasztás modellezése mellett a fogyasztási bizonytalanság számszerűsítése kapcsán olyan megoldást dolgozzunk ki, amely – a módszertani megalapozottság mellett – jól interpretálható és a gyakorlatban is felhasználható eredményeket ad, hiszen ott sokszor csak *ad-hoc* mérőszámok, hüvelyk-ujj-szabályok léteznek, a klasszikus megoldások a profilozás területén pedig rendkívül kevés lehetőséget biztosítanak főként a bizonytalanság modellezése tekintetében.

A fogyasztói profilok – amelyek egy alapvetően kvantitatív módszertani folyamat eredményeként állnak elő, és leírják, hogy a fogyasztás hogyan függ különböző szezo-

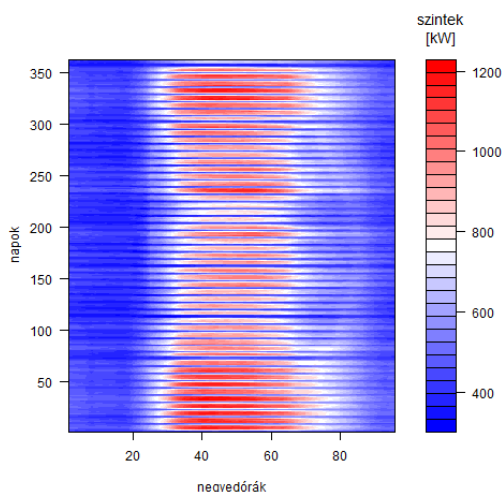
⁶ Negyedórás idősor esetén egy adott negyedórában realizálódott 1 [kW] átlagterhelés $1 [kW] \cdot \frac{1}{4} [h] = \frac{1}{4} [kWh]$ fogyasztásnak felel meg.

⁷ Hasonló mértékegységbeli többféleség a földgáz iparágban is van, ahol a [m³] mellett [TJ]-ban szokás még a fogyasztást számszerűsíteni, de a nominálás (gázigény-bejelentés, a menetrendezés földgázipari megfelelője) a 2015/2016-os gázév óta már [kWh]-ban történik.

nális, naptári vagy egyéb hatásoktól⁸ – vizsgálatával, modellezésével kapcsolatban rendkívül széleskörű, de természetesen korántsem egységes szakirodalom és gyakorlati alkalmazások állnak rendelkezésre.

Annak érzékeltetésére, hogy időben a fogyasztás mennyire különböző lehet, tekintsük a következő, ún. szintvonalábrát (*contour plot*, a részletes ismertetésére a dolgozatban még kitérünk), amely a különböző terhelési szinteket eltérő színnel jelölve mind az éven (y tengely), mind a napon belüli (x tengely) viselkedésről kompakt képet ad. Az értekezésben több ilyen ábra összehasonlításával nagyon szemléletesen látjuk majd azt, hogy a profilozás során egyedileg mennyire különböző mintázatokat kell megragadni – lényegében – már a várható érték szintjén is.

2. ábra: Egy vállalati portfólió terhelési idősorának szintvonalábrája



Forrás: saját készítésű ábra (R).

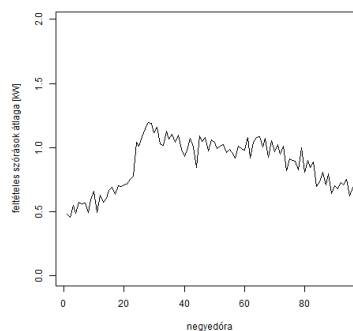
Mivel azonban a fogyasztás maga is **sztochasztikus**, ezért annak bizonytalanságával ugyanúgy számolni kell, mint azt például pénzügyi idősorok esetén is tesszük. Az alapvető különbség ennek kezelésében abból adódik, hogy a fogyasztás (illetve látni fogjuk, hogy annak bizonytalansága is) sokkal inkább modellezhető különböző **fundamentális** változók segítségével, mint maguk a pénzügyi idősorok, és így az alkalmazható módszerek köre is szükségszerűen eltérő, noha bizonyos mértékű analógia, párhuzam azért fennáll. Az értekezésben a fogyasztással kapcsolatos bizonytalanság, az ún. volumenkockázat modellezése alatt a véletlen komponens viselkedésének a leírását értjük.

A saját empirikus eredmények elkészítése során – a klasszikus idősoros regressziós megoldásoknak a hibatagok állandó szóródására vonatkozó feltételének a feloldásá-

⁸ A szakirodalom alapján egyébként nincs egységes, általános érvényű definíció.

val – olyan, ún. feltételes (időfüggő) standard hibákat számolunk, amelyek a magyarázóváltozók konkrét értékének a feltétele mellett kerülnek meghatározásra. A feltételes standard hibák ilyen modell alapú számolása mindenképpen új eredménynek tekinthető, hiszen a szakirodalom ezzel kapcsolatban meglehetősen szegényes és kevés empirikus eredménnyel szolgál. Az értekezésben a 3. ábrán is látható eredményeket egyedi görbék esetében is elkészítjük, és vizsgáljuk, hogy időben mikor nagyobb vagy kisebb a fogyasztás bizonytalansága, illetve ez fundamentálisan hogyan magyarázható.

3. ábra: Standard hibák átlaga téli hétköznapiokon egy fogyasztói görbe terhelési idősora esetén



Forrás: saját számítások (R) és saját készítésű ábra (R).

Az előző két kutatási részeredmény illusztratív célú bemutatási logikájával párhuzamba állítva, az értekezésben az alábbi **területeket** vizsgáltuk:

- a különböző fogyasztási idősorok hogyan jellemezhetőek, melyek azok az ún. **stilizált tények**, amelyeket bármilyen fogyasztást leíró modell, így a profilozás során is rögzíteni kell;
- a különböző fogyasztási idősorok esetén a fogyasztás **bizonytalansága** hogyan alakul, fellelhető-e abban az alapvetően magát a fogyasztási idősort is jellemző (többszintű) **szezonális**, vagy más szabályos tendencia;
- a fentiek hogyan **modellezhetőek**, különös tekintettel például a **linearitás hiányának** – főként az időjárás-függőségnek – illetve az időben változó szóródásnak (a **heteroszkedaszticitásnak**) az együttes kezelésére.

A fenti kérdések, területek vizsgálatára az alábbi **hipotéziseket** fogalmaztuk meg:

- H1: Villamosenergia-fogyasztási görbék esetében jellemzően a napon belüli szezonális a görbék varianciájának legjelentősebb forrása.

- H2: Az ún. klasszikus (tipikus napi profilgörbén alapuló) megoldásokhoz képest realisabb fogyasztási mintázatok kialakítására is lehetőség nyílik a görbe tekintetében egyedileg releváns tulajdonságok kinyerésével.
- H3: A hibák állandó szóródásának feltételezése mellett időszakoktól függően a volumenkockázatot alul- vagy felülbecsüljük.
- H4: A volumenkockázat időben nem állandó, hanem különböző exogén változók, szezonális és naptári hatások függvényében időben változik.

Az **első fejezetben** tehát a fogyasztói profilozás fontosabb gyakorlati felhasználási területeit tekintjük át. Megvizsgáljuk, hogy a profil illetve a profilozással kapcsolatos kockázatok alatt mit értünk, illetve hogy ezek közül melyik az (a volumenkockázat), amelyikkel az értekezés kiemelten foglalkozik.

Ezt követően különböző fogyasztási idősorok esetében megvizsgáljuk az idősorok stilizált tényeit, külön figyelve a többszintű szezonális illetve a véletlen hatás görbénként nagyon különböző megjelenésére. Ezek a gyakorlatban viszonylag ritkán használt módszerek olyan fontos összefüggésekre is felhívják a figyelmet, amelyek egy idősoros ábra alapján kevésbé nyilvánvalóak. Az elemzések során olyan szemléletmódra helyezzük a hangsúlyt, amely a – majd bemutatásra kerülő – korábbi kutatási eredmények tükrében az értekezés rendhagyó nézőpontjának a létjogosultságát, indokoltságát is kiemeli.

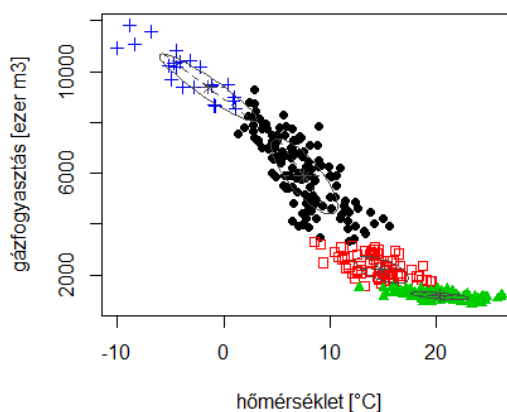
A **második fejezetben** a szakirodalom lényegi összefoglalása szerepel. Röviden áttekintjük a profilozás során leggyakrabban használt módszertani – alapvetően különböző klaszterezési illetve regressziós – megoldásokat. Külön alfejezetet szánunk az időjárási változók kezelésének fontosabb kérdéseire, és egy korábbi saját publikáció bizonyos részének kiemelésével (és kiegészítésével) bemutatjuk, hogy a hőmérséklet véletlen hatásának a szerepe illetve mértéke hogyan számszerűsíthető, a linearitás hiánya hogyan kezelhető elegánsan földgázfogyasztási idősor példáján. Kitérünk arra is, hogy az ilyen megoldások a profilozás kapcsán miért nem olyan előnyösek – akár módszertani (például sok előkészítő lépés szükségessége), akár értelmezésbeli problémák, vagy az eredmények felhasználhatósága (például releváns tulajdonságok elfedése) szempontjából.

A **harmadik fejezetben** az értekezés hátralévő részében használt módszertani eszközöket tárgyaljuk. Mivel ezek egy része a sztochasztikus idősor-elemzésben elterjedt és ismert, ezért a hangsúly inkább a felhasznált módszertan kereteinek rögzítésén

van. Ismertetjük a PAR-modell⁹ lényegét és az általa nyújtott elemzési lehetőségeket. Ezt a gyakorlatban ritkábban használják, de lényegében a klasszikus ARMA-modellek kiterjesztésének tekinthető, és mind integrált, mind stacioner idősorok esetén hasznos tulajdonságai aknázhatóak ki.

A fejezet legnagyobb részét a keverék-modellek (részletesebben a *Gauss*-féle keverék-modell) tárgyalása jelenti. A részletesebb ismertetést többek között az indokolja, hogy ezt a módszertant a hazai közgazdasági gyakorlatban eddig kevésbé alkalmazták. Gyakorlati példán (ismét hazai földgázfogyasztási adatokon) bemutatjuk, hogy hogyan, milyen logika mentén alkalmas ez a változók időben változó kapcsolatának, összefüggésének a modellezésére, leképezésére. Mind a profilozás, mind a fogyasztás bizonytalanságának a mérése tekintetében központi szerepe van a keverék-modell azon feltételezésének (kiindulópontjának), miszerint a mintában a változók kovariancia-struktúrája nem állandó, amely nemcsak a várható érték modellezésében jelentős, hanem a szóródás tekintetében is. Az ábra pontosan erről a logikáról árulkodik: a változók kapcsolata „pontfelhőnként” más, és szemmel láthatóan a „pontfelhők” szóródása is különböző. Előbbi jelenség a linearitás hiányának, utóbbi a heteroszkedaszticitásnak a leképezésében fog manifestálódni.

4. ábra: Keverék-klaszterezés eredménye a napi középhőmérséklet-gázfogyasztás példán



Forrás: saját számítások (R) és saját készítésű ábra (R).

A fejezet több olyan levezetést, értelmezést is tartalmaz, amelyek az idegen nyelvű szakirodalmakban sem jelennek meg, viszont a keverék-modellen alapuló regresszió és a klasszikus többváltozós regresszió kapcsolatát jobban kihangsúlyozzák, így ez a saját eredmények értékelése szempontjából rendkívül hasznos. A fejezet végén külön

⁹ Periodikus autoregresszív modell.

kitérünk a keverék-modellek profilozás területen történő, de az értekezés témájától eltérő fókuszú korábbi alkalmazásaira is, megteremtve ezzel az átmenetet az új empirikus eredmények bemutatásának az irányába.

A **negyedik fejezet** a fogyasztói profilok, illetve a volumenkockázat számszerűsítésével kapcsolatos eredményeket tartalmazza. Mind a profilozással, mind a bizonytalansággal kapcsolatos alfejezetekben szerepelnek ún. klasszikus megoldások is, részben *benchmark*-ként, az új eredményekkel történő szembeállítás, összehasonlítás céljával, részben pedig feltáró jelleggel.

Megvizsgáljuk, hogy a keverék-modellek keretrendszere mennyiben tud többletinformációt nyújtani a hasonló fogyasztási szokásokkal (fogyasztási mintázattal) jellemezhető fogyasztói csoportok, az ún. profilcsoportok készítésében, és a különbség érzékeltetésére az eredményeket összehasonlítjuk egy klasszikus (regressziós) megoldással. A fejezetben azt is tárgyaljuk, hogy az alkalmazott módszertan logikájában gondolkodva mit tekinthetünk „tipikus” fogyasztási mintázatnak.

Hagyományos idősoros regressziós modellek eredményeinek felhasználásával, saját számítások alapján megvizsgáljuk, hogy a hibák milyen – alapvetően heteroszkedasztikus – viselkedést mutatnak, amelynek lefutása, karakterisztikája szintén fogyasztónként különböző lehet. A tapasztalatokat összevetjük egy egyszerűbb, ún. heurisztikus mutatószámmal, amelyeket a gyakorlatban is alkalmaznak a volumenkockázat mérésére. Bemutatjuk, hogy az időben változó kovariancia-struktúra megfelelő kezelésén keresztül az időben változó szóródás jól modellezhető a keverék-regresszió alkalmazásával.

Összességében elmondható, hogy az értekezés eredményei a **hazai és nemzetközi kutatási eredményekhez és gyakorlati alkalmazásokhoz való hozzájárulásnak** tekinthetők – többek között – az alábbiak tekintetében:

- A profilozás témakörében olyan területet vizsgálunk, amely a profil definiálása mellett figyelembe veszi a bizonytalanságot is, ami a gyakorlatban folyamatosan felmerülő igényeknek sokkal inkább megfelel.
- Olyan – a profilozás területén eddig ritkán használt – módszertant (keverék-modellek) alkalmazunk, amely ahogyan magát a fogyasztást, úgy annak a bizonytalanságát is különböző szezonális és naptári hatásokat rögzítő exogén változók függvényeként írja fel, így a kettő modellezése egy lépésben történik meg.

- Bemutatjuk, hogy a – tisztán módszertanilag is sok kedvező tulajdonsággal bíró – keverék-modell sokkal jobban képes a fogyasztás fundamentális okainak a megragadására, amelyre a profilsoportok képzése nagyon jó (közvetett) bizonyítékkal szolgált.
- A keverék-regresszió alkalmas a véletlen heteroszkedasztikus viselkedésének a megragadására, amelynek az alátámasztására ellenőrizzük a hibák illetve a számított ún. feltételes szórások konzisztenciáját, illetve a mögöttes okokat is vizsgáljuk. Bemutatjuk azt is, hogy a keverék-regresszióval átlagosan jóval szűkebb konfidencia-intervallum készíthető, szemben a klasszikus megoldásoknak a gyakorlatban szinte használhatatlan mértékben tág intervallumával szemben.
- Az értekezésben nagyon sok olyan utalás szerepel, amelyekre épülve akár az itt bemutatott eredmények, akár ezeknek más energiapiacra (például földgázpiac) vagy a kínálat oldalon (például erőművek termelése) történő alkalmazásával illetve továbbgondolásával további kutatási utak nyithatóak.
- Az értekezés hangsúlyos és mindenképpen fontos eredménye a *Gauss*-féle keverék-regresszió *R Project* programcsomagban történő implementálása, amely a felsorolt feladatok elvégzése és eredmények értékelése mellett a kutatás egyik fontos alapját képezte.

1. A FOGYASZTÓI PROFILOZÁS FELHASZNÁLÁSI TERÜLETEI ÉS FONTOSABB ALAPFOGALMAI

A (fogyasztói) **profilozás**, a (fogyasztói) **profil** pontos meghatározásai, definíciói tekintetében a szakirodalom és a gyakorlat is meglehetősen heterogén. Ennek lehet az oka az, hogy a profil fogalma sokszor függ a felhasználási céltól, és ennek megfelelően az alkalmazott módszerek is rendkívül sokfélék lehetnek.

A **görbe** (*curve*) kifejezés mind az üzleti alkalmazások, mind a módszertan területén elterjedt fogalom. Az energiapiacra például görbeként tekintünk a különböző terhelési görbékre, vagy árfolyamok esetén az árgörbékre, statisztikai illetve adatbányászati területeken pedig egyre nagyobb érdeklődésre tartanak számot a különböző folyamatokkal kapcsolatos elméleti és empirikus kutatások, ahol a vizsgálat alanyait függvények (*functionals*), görbék (*curves*), idősorok (*time series*) jelentik.¹⁰

A fentiek eredőjeként a profilozás, mint módszertani, modellezési folyamat során azt számszerűsítjük, hogy egy-egy fogyasztási vagy terhelési görbe időbeli lefutása hogyan alakul, hogyan függ többek között különböző szezonális és naptári hatásoktól. A feladattal szemben jogos elvárás az, hogy a profilok a görbéket általánosan, tendenciaszerűen jellemző ún. **stilizált tényeket** jól legyenek képesek rögzíteni.

Az értekezésben kiemelten **fogyasztói** profilozással, fogyasztási vagy terhelési görbék profiljainak a meghatározásával foglalkozunk, így a fontosabb gyakorlati felhasználási területek, fogalmi kérdések mellett a fogyasztói idősorok stilizált tényeinek empirikus vizsgálatával foglalkozunk ebben a fejezetben. A kétféle megfogalmazás (fogyasztási és terhelési görbe) egyébként főként a műszaki és az üzleti-közgazdasági szemléletek különbözőségét tükrözi, mi az értekezésben a két kifejezést párhuzamosan használjuk (fogalmak szintjén inkább a fogyasztás szót részesítjük előnyben, de a számítások terhelési görbékre készülnek), az eredmények, következtetések ettől függetlenek.

1.1. Ár- és mennyiségi bizonytalanság a villamosenergia-piacon

A villamosenergia-piacon a kereslet-kínálat egyensúlynak minden időpontban fenn kell állnia. A kereslet fedezése történhet ún. **standard termékekkel** (például éves,

¹⁰ Az értekezés szempontjából az alábbi kifejezések lehetnek irányadók: *functional data analysis*, *functional clustering*, *curve clustering*, *time series clustering*. Erről ld. bővebben a 2. fejezetet.

negyedéves, havi stb. zsinór- illetve csúcsidőszaki¹¹ (*baseload* illetve *peakload*) termékekkel), majd a szállítási időszakhoz közeledve a következő napi nyitott pozíciót a **másnapi** *dayahead* piac órás termékeivel tudjuk előre lekereskedni (vagy a **napon belüli** *intraday* piacon negyedórás termékekkel), a menetrendtől vett tény-terv eltéréseket pedig utólag a **kiegyenlítő** energia piacon kell rendezni. A különböző villamosenergia-termékekkel történő kereskedés időben változó jellegét mutatja az 5. ábra.

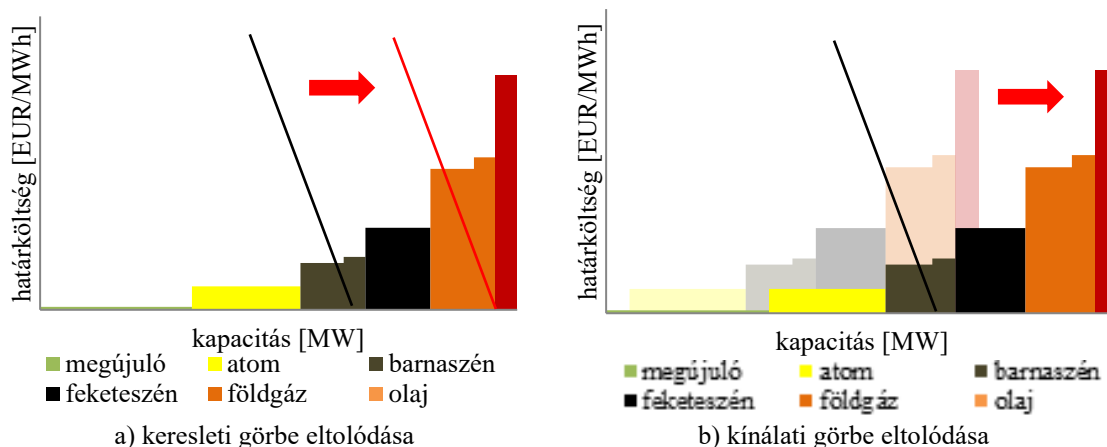
5. ábra: Kereskedés villamosenergia-piaci termékekkel

| Ex-ante | | Ex-post |
|--|--|---------------------|
| standard termékek ¹² (<i>forward-ok, FM, FW ... WE, DA</i> zsinór- és csúcsidőszakra stb.) | másnapi (<i>dayahead, spot</i>) napon belüli (<i>intraday</i>) órás, negyedórás termékek | kiegyenlítő energia |

Forrás: saját készítésű ábra.

A különböző villamosenergia-termékek árai a *merit order*, mint kínálati görbe, illetve az aktuális keresleti viszonyok mellett alakulnak ki. Az 6. ábrán egy *merit order* görbe szerepel, amely a (beépített) kapacitás függvényében ábrázolja az erőművek határköltségét.

6. ábra: Erőművi *merit order* ábrázolása



Forrás: saját készítésű ábra (Excel).

A *merit order* elején a zero határköltséggel üzemelő megújulók állnak, őket követik a nukleáris erőművek, majd a szenes és a gázos (OCGT és CCGT¹³) erőművek kö-

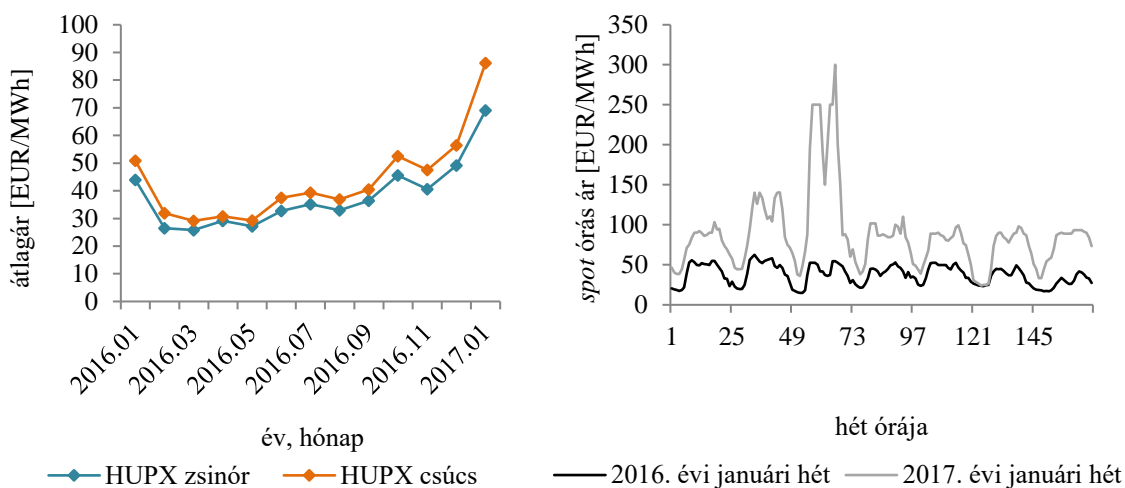
¹¹ A zsinór termék a hét minden napján, minden órában elérhető, a csúcsidőszaki termék a csak hétköznapokon, 8 és 20 óra között elérhető termékeket jelenti.

¹² A forward-ok távolabbi lejáratokra is vonatkozhatnak, a legközelebbi lejáratok a következő hónap (*front month*), következő hét (*front week*), vagy akár a következő nap (*dayahead*), de ebben a kategóriába sorolhatjuk a következő néhány hétvégére, vagy következő néhány napra vonatkozó termékeket is.

vetkeznek. A *merit order* végén pedig a magas határköltséggel működő olajtüzeléses erőművek szerepelnek. Ahogy az ábra is mutatja, a piactisztító ár kialakulásában a keresleti görbe és a kínálati görbe eltolódás is szerepet játszhat.¹⁴

Az áralakulás lehet tehát részben keresletvezérelt: a hétköznapi és csúcsidőszakok, illetve a hőmérsékletfüggő időszakok magasabb villamosenergia-kereslete (a 6. a) ábrán jobbra tolódó keresleti görbe) ilyenkor *ceteris paribus* magasabb villamosenergia-árakat eredményez. Mivel a villamosenergia, mint termék nem igazán helyettesíthető, nem (vagy csak nagyon nehezen) tárolható, és relatíve árrugalmatlan a kereslet, ezért a villamosenergia-árak nagyon volatilisak, könnyen kialakulhatnak magas árak, illetve ún. ártüskék (*spike*-ok) is.

7. ábra: Magyar másnapi órás villamosenergia-árak historikus alakulása



Forrás: HUPX Zrt. adatok alapján saját készítésű ábrák (Excel).

Az utóbbi időben egyre hangsúlyosabb az áralakulás kínálatvezéreltsége (sokszor politikai harcokkal is megfűszerezve). A (hagyományos) erőművi kiesések eredményezhetnek magas villamosenergia-árakat. Erre nagyon jó példa 2017. januári(-februári) időszaka, amikor részben a nagy hidegnek betudható erőművi kiesések¹⁵, részben a tervezett karbantartások, illetve a déli országok felől egyébként sokszor érkező olcsó im-

¹³ Az OCGT (*open cycle gas turbine*) a nyílt ciklusú, a CCGT (*combined cycle gas turbine*) a (gőzturbinának köszönhetően) jobb hatásfokkal rendelkező kombinált ciklusú gázturbinás erőművek szokásos rövid elnevezése.

¹⁴ A 6. b) ábrán például egy olyan kínálati sokkot ábrázolunk, amikor a megnövekvő megújuló kapacitás tolja jobbra a kínálati görbét, erről ld. még bővebben az alfejezetet később.

¹⁵ Ld. a HUPX (*Hungarian Power Exchange*, a magyar villamosenergia-tőzsde rövid neve) ún. (energiapiaci bennfentes információ) közzétételi honlapját: <https://www.insideinformation.hu/hu/pubpages/newslistmain.aspx>.

portáram hiánya miatt a HUPX-en rekordmagas órás árak alakultak ki a másnapi *spot* piacon.¹⁶

A 7. ábrán az eddig leírtak alátámasztására, a HUPX átlagos havi zsinór- és csúcs-időszaki árai szerepelnek. A jobb oldalon 2016. és 2017. januári hónapjaiból választottunk ki egy-egy hetet, az említett ominózus áralakulás nagyságrendjének érzékeltetése céljából.

Természetesen az árak volatilitása nemcsak ilyen, a hagyományos erőművek klasszikusnak is nevezhető kiesési (*outage*) kockázatával függhet össze. A *merit order* elején szereplő megújulók (szél- és naperőművek) termelése ugyanis erősen időjárás-függő, így a kínálat görbe elején egy újfajta kockázatot hordozó termelő van jelen. Itt a kockázatot nemcsak a termelésük rendkívüli variabilitása, hanem az előrejelzés, és így a másnapra lejelentett várható termelés (menetrend) nagymértékű bizonytalansága jelenti. Ennek illusztrálására tekintsük a hazai szélerőmű termelés 2016. év eleje óta történő alakulásának néhány mérőszámát (ld. a 2. táblázatot).

2. táblázat: A hazai szélerőművek 2016. évi termelésének néhány leíró statisztikája

| Széltermelés | Átlag [MW] | Szórás [MW] | Terjedelem ¹⁷ [MW] |
|-------------------|------------|-------------|-------------------------------|
| Tény | 71 | 77 | 311 |
| Terv | 76 | 70 | 298 |
| Tény-terv eltérés | - 4 | 31 | 344 |

Forrás: MAVIR Zrt. adatok alapján saját készítésű táblázat.

A hazai szélerőművek összes beépített kapacitása (kb. 300 MW) nagyságrendekkel kisebb a magyar rendszerterhelés jellemző szintjeinél (időszaktól függően a napi átlagos terhelés 4000 és 5500 MW körül alakul, ld. a C) melléklet 24. táblázatát). Az mindenképpen figyelemre méltó, hogy a tény-terv eltérések terjedelme is elég magas. Azaz nemcsak abban van a kockázat, hogy 0-300 MW közötti megújuló kapacitás ténylegesen ott van-e, elérhető-e a *merit order* elején, hanem abban is, hogy milyen biztonsággal tudjuk megmondani a tényleges termelés értékét, ugyanis az időjárástól való függőség végett utóbbi nem tervezhető úgy, mint ahogy az a hagyományos erőművek esetében történik.

Az értekezésben a kínálat oldali bizonytalansággal, annak számszerűsítésével nem foglalkozunk, habár az eredményeink kínálat oldali továbbgondolása és alkalmazása

¹⁶ Január első felében volt olyan nap, amikor több órás ár is meghaladta a korábbi historikus maximum árat.

¹⁷ A terjedelem a maximális és a minimális termelési értékek különbsége.

mindenképpen megfontolandó és további kutatási irányok megfogalmazására érdemes. A kínálat oldali bizonytalanságra vonatkozó fenti rövid példa arra hívja fel elsősorban a figyelmet, hogy a kereslet oldali bizonytalanság kezelésének – amely a fogyasztói befolyásolással valamilyen mértékben menedzselhető – mennyire fontos szerepe lehet (erről még később ejtünk szót).

Visszatérve a különböző villamosenergia-piaci termékek áaira, a 8. ábra az másnapi órás árak és a kiegyenlítő energia árak, valamint a rendszerszintű szabályozási mennyiségek alakulását mutatja 2016 januárja óta. A másnapi órás árak aukció során határozódnak meg. Azt pedig, hogy a menetrendtől való eltérések miatti szabályozási energiamennyiséget mely erőművek biztosítják, a rendszerirányító vezérli, lényegében *merit order* alapon.

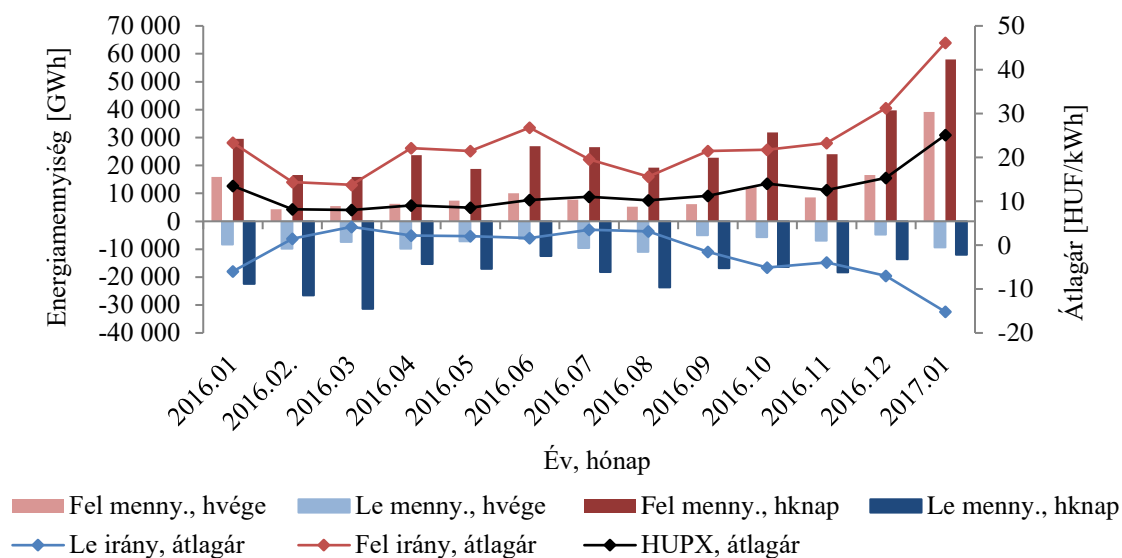
Az ábrát vizsgálva megállapítható, hogy a fel irányú kiegyenlítő energia átlagára nagyságrendekkel meghaladja a másnapi órás árakat, és emellett szorosabban együttmozog az felszabályozási energiamennyiséggel a le irányhoz képest. Az árak szintjén megjelenő havi átlagolás nyilvánvalóan a rövid távú (például napon belüli) tendenciákat elrejt, de az látszódik, hogy a szabályozási mennyiségek évszaktól függetlenül a hétköznapiakon magasabbak. Emögött viszont nemcsak a hétköznapiok nagyobb **fogyasztási bizonytalansága** áll, hanem **strukturális** okok is. A hétköznapiok reggeli felfutása mind volumenét, mind sebességét (gradiensét) tekintve nagyobb, mint a hétvégéké, és itt a kereslet lekövetésében nagy szerepe van a szabályozási piacon szereplő erőműveknek, hiszen a menetrend csak negyedórás lépcsőnkénti konstans terhelést ad meg¹⁸, viszont a kereslet-kínálat egyezőségének minden időpontban fenn kell állnia. A keresleti okok mellett a rendszerszintű szabályozási energiamennyiség nagy hányada az esetleges erőművi kapacitás-**kiesések** (tehát kínálati okok) miatt is adódhat, ez különösen jól látszik 2017 januárjában.

Az utóbbtól eltekintve a rendszerszintű eredmények azért arra engednek következtetni, hogy a fogyasztással kapcsolatos bizonytalanságban is lehet valamiféle (akár szezonális jellegűnek is nevezhető?) szabályosság, ami lényegében a rendszerszintű tény-terv eltérésekben realizálódik; és kisebb portfóliók vagy egyedi fogyasztók szintjén a fogyasztási bizonytalansággal kapcsolatos realizációk kvalitatív értelemben nagyon

¹⁸ A gyakorlatban ez inkább órás lépcsőket jelent, ami a még több szabályozói energia igénybevételét indukálja. Vagyis ha a negyedórás átlagos terhelést pontosan el is találjuk, szabályozási energiára akkor is szükség lesz.

hasznalók lehetnek. A rendszerszintű példafelhozatalnak egy kedvező hozadéka, hogy a probléma makroszintű illetve kínálati oldalt érintő vetületeire is felhívja a figyelmet.

8. ábra: A másnapi órás és a kiegyenlítő árak historikus alakulása¹⁹



Forrás: MAVIR Zrt. adatok alapján (részben saját számítások és) saját készítésű ábra.

Az értekezés egyik központi témája az, hogy a **fogyasztás** alapvetően **sztochasztikus** folyamat, fogyasztási idősortól függően kisebb vagy nagyobb lehet a **véletlen** szerepe. A gyakorlat pedig egyre több olyan követelményt támaszt, ahol a fogyasztói viselkedés mellett annak bizonytalansága is fontos.

Részben az energiaköltségek csökkentése, részben pedig a globális trendek (környezetvédelem, erőforrásokkal történő takarékoság stb.) miatt egyre nagyobb a fogyasztói aktív részvétel szerepe az energiapiacokon. Ezek elsősorban fejlettebb piacokon elterjedtek, kevésbé fejlett piacokon még inkább csak nagyfogyasztóknál. A **fogyasztói befolyásolás** (*demand side management*, röviden DSM) lényege, hogy a kereslet-kínálat minden időpillanatban történő egyensúlyát nemcsak az erőművek irányításával, szabályozásával lehet elérni. A fogyasztói befolyásolásnak nagyon sok, műszaki illetve gazdasági jellegű (mikro- vagy makroszintű) célja lehet, többek között az alábbiak:

- a fogyasztói görbe kisimítása a csúcsidőszaki fogyasztás csökkentésével és völgy- illetve mélyvölgy-időszaki fogyasztás növelésével, ami segíthet a hálózati

¹⁹ MAVIR publikáció alapján az átlagárak az alábbiak: a kiegyenlítő energia átlagára a fel- illetve leszállás negyedórás elszámoló árának az átlaga; a HUPX átlagár a HUPX aktuális órás árak átlaga az elszámolási időszak adott napján érvényes MNB hivatalos devizaárfolyamán átváltva.

ti veszteség csökkentésében, esetleg nagyobb beruházási döntések elhalasztásában;

- a villamosenergia beszerzési költségének csökkentése azáltal, hogy a csúcsidei fogyasztás egy része olyan időszakra tevődik át, amikor az energia ára olcsóbb;
- a kiegyenlítő energia költségének csökkentése azáltal, hogy a menetrendtől vett jelentősebb eltérés esetén a fogyasztás átcsoportosítható;
- a(z esetleg decentralizált) megújuló energiatermelés változékony és nehezen előrejelezhető, és az ebből adódó bizonytalanságnak a mérséklése.

A fenti célok, illetve azoknak részleges megvalósulása is rengeteg jövőbeli kihívással jár valamennyi energiapiaci szereplő szempontjából. Amennyiben például a fogyasztás-átcsoportosítás szokássá válik, az a profilok strukturális átalakulását is jelenti majd, ennek pedig a szabályozói energiapiacra nézve is lehetnek következményei. Ez nemcsak a szabályozás, de infrastruktúra szempontjából is kihívást jelentenek a rendszerirányítók számára.²⁰ A különböző rövid és hosszú távú fogyasztás-átcsoportosítási célok rengeteg árazási kihívást és feladatot jelentenek. Mivel itt gyakran szóba kerül a menetrendtől vett **eltérés**, vagy a fogyasztásban *baseline* scenárióhoz képesti **szignifikáns** (azaz nem a véletlennek betudható) bekövetkező változás **statisztikailag is megalapozott** értékelése, ezért a fogyasztással kapcsolatos bizonytalanság mérése, modellezése a jövőben egyre inkább központi téma lesz.

A bevezetésben eddig inkább a makroszintű kérdésekre helyeztük a hangsúlyt. A folytatásban a profilok, a profilozás gyakorlati felhasználási területei kapcsán elsősorban mikroszintű kérdésekkel foglalkozunk, de látni fogjuk, hogy a kettő egymástól nem teljesen független.

²⁰ Bár a fogyasztói befolyásolás elsődleges céljai között a napi fogyasztás egyenletesebbé tételét, az extrém fogyasztások tompítását szokás emlegetni, az eszköztár ennél jóval szélesebb – és a teljesség igénye nélkül – a napi fogyasztás alakításában többek között az alábbi megoldások jöhetnek szóba: csúcslevágás (*peak clipping*), völgyfeltöltés (*valley filling*), fogyasztás-csökkentés (*strategic conservation*), fogyasztás-növelés (*strategic load growth*), fogyasztás-átcsoportosítás (*load shifting*), ld. bővebben *Macedo et al.* [2015].

1.2. A profilozás legfontosabb gyakorlati felhasználási területei

A fogyasztói vagy keresletoldali profilozás gyakorlati felhasználási területeit illetően – a teljesség igénye nélkül – többek között tehát az alábbiakat tudjuk beazonosítani:

- menetrend-készítés,
- fogyasztói portfólió (hosszú- és rövidtávú) fedezése (*hedge*),
- ún. hagyományos szerződések²¹ árazása,
- fogyasztói befolyásolás,
- portfólióépítés, mérlegkör-összeállítás.

Mivel az értekezés kiemelten foglalkozik a fogyasztás bizonytalanságával is, ezért a lehetséges alkalmazási területeket olyan megvilágításban tárgyaljuk, hogy ennek a bizonytalanságnak a figyelembevétele mennyiben módosíthatja vagy egészítheti ki azt a fajta döntéshozatalt, amely a fogyasztást inkább csak **várható** értékben vizsgálja, a **bizonytalansággal** sokkal kevésbé törődik. Ez a tárgyalási mód egyben rávilágít arra, hogy a profilozással szemben egyre inkább reális elvárás a bizonytalanság korrekt kezelése, ami mind elméleti (módszertani), mind gyakorlati felhasználási szempontból fontos konklúziók levonását eredményezi. Módszertanilag ugyanis egy kevésbé kijárt és vizsgált területről van szó, és a gyakorlatba történő átültetés, alkalmazás sok esetben még a klasszikusnak tekinthető megoldások esetében sem történt meg igazán.

Érdemes még megjegyezni, hogy a profilozás, mint fogalom legtöbbször egyedi fogyasztói görbék lefutásának a vizsgálatakor jelenik meg, ennek ellenére szokás használni a kifejezést más esetekben is, például portfóliószintű görbék elemzésekor.

1.2.1. Rövid- és hosszú távú fedezés illetve árazás

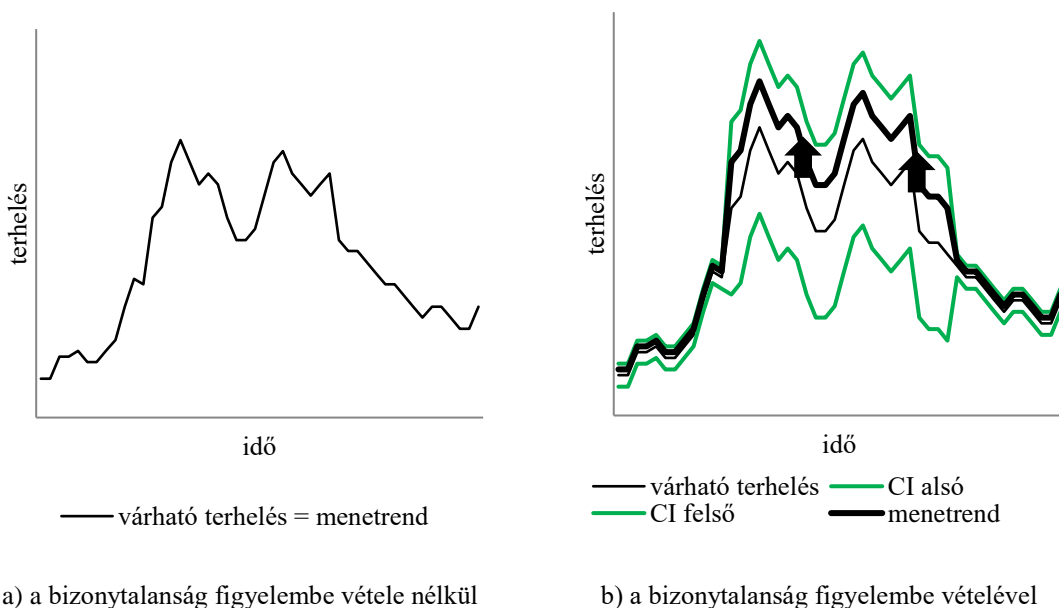
Visszaulva a különböző villamosenergia-piaci termékek árainak egymáshoz képesti alakulására, kockázatkerülő kereskedői magatartás mellett a magas másnapi vagy kiegyenlítő energiaáraktól való félelem **túlfedezést** (*overhedge*) eredményezhet. Példákon keresztül bemutatjuk, hogy a gyakorlatban mit is jelent ez.

²¹ Ilyenek az ún. ellátás- illetve menetrend-alapú szerződések. Ellátás-alapú szerződésben az eladó vállalja a vevő vételezési igényének biztosítását, a vevő pedig a betáplálás átvételét, menetrend-alapú szerződésben a vevő illetve az eladó a szerződésben pontosan meghatározott mennyiség betáplálásra és átvételre vállalnak kötelezettséget.

Magyarországon a másnapi **menetrendet** negyedórás bontásban kell elkészíteni, amely egyetlen egy, 96 negyedórából álló görbe. Első gondolatként a másnapi menetrendnek a másnapi terhelés negyedórás előrejelzését érdemes beadni.²²

Mivel azonban a vételi kiegyenlítő energia (amikor magasabb a fogyasztás, mint a menetrend) jóval drágább, az eladási (amikor alacsonyabb a fogyasztás, mint a menetrend) pedig jóval olcsóbb²³, ezért előfordulhat, hogy az előrejelzéstől felfelé térnek el a kereskedők.²⁴ Az idősorok **bizonytalansága** azonban **időben** nem feltétlenül állandó, ennek megfelelően a bizonytalanabb negyedórákban a várható terhelésnél jóval magasabb értéken kell a menetrendet megállapítani (ilyenkor a kiegyenlítő energia ára is jellemzően magasabb), mint egyébként. Üzletibb nyelvre lefordítva a kockázatkerülő magatartásból adódóan a túlfedezés mértéke a bizonytalanabb időszakokban magasabb, a kevésbé bizonytalan időszakokban alacsonyabb. A túlfedezésből adódó többletköltség úgy is felfogható egyébként, mint a magas kiegyenlítő energiaárak elkerüléséért fizetett kockázati prémium.

9. ábra: Menetrend megállapításának sematikus szemléltetése²⁵



Forrás: saját készítésű ábra (Excel).

²² A másnapi menetrend természetesen napon belül még menetrend-módosítással megváltoztatható, amennyiben van kellő likviditás a napon belüli piacon.

²³ A gyakorlat azonban sokkal bonyolultabb, mivel a kiegyenlítő energia ára nemcsak a mérlegkör, hanem a rendszerállapot függvénye is.

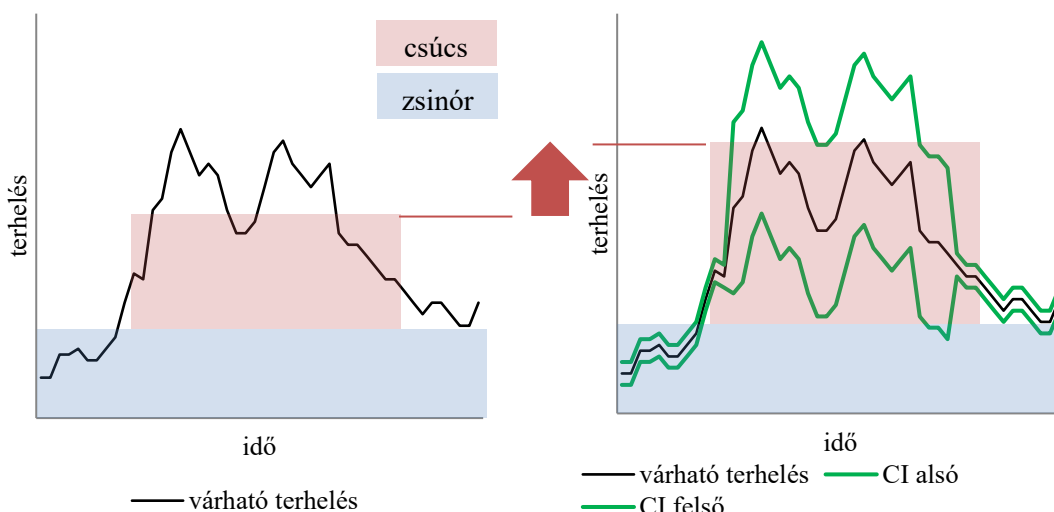
²⁴ Ez történhet az előbb említett kockázatkerülési, de akár spekulációs okok miatt is.

²⁵ A b) a ábrán **menetrend** a kockázatkerülés következtében a várható fogyasztáshoz képest magasabb (felfelé eltérített) menetrendet jelenti. **CI alsó** valamilyen választott megbízhatósági szint mellett a terhelésre készített konfidencia-intervallum alsó, **CI felső** a felső határát jelöli itt és a 10-11-12. ábrákon.

Az 9. ábrán ennek a problémának a szemléltetése szerepel. Természetesen a példa a megfelelő menetrend megállapításához nemcsak a terhelés bizonytalanságának, hanem az áraknak a figyelembevétele is szükséges. Ezzel a kérdéskörrel foglalkozik például *Lo és Wu [2003]* munkája.

A bemutatott kockázatkerülő magatartás természetesen **hosszabb távon** is érvényes lehet (logikailag egyébként meg is előzi az előző menetrendezési példát, azonban a probléma felfeztetése talán így egyszerűbb). Ilyenkor a bizonytalanabb fogyasztású időszakokra optimális lehet túlfedezni magunkat ahhoz képest, mint amit a várható fogyasztás indokol, ami valószínűleg csúcsidőszakban történő nagyobb túlfedezést jelent. Ennek a stratégiának a szemléltetésére szolgál a 10. ábra.

10. ábra: Fogyasztói igények részleges fedezése zsinór- és csúcsidőszaki termékekkel



a) a bizonytalanság figyelembe vétele nélkül

b) a bizonytalanság figyelembe vételével

Forrás: saját készítésű ábra (Excel).

Ez a logika egyébként az egyedi (elsősorban nagyobb) **fogyasztók árazásánál** is megjelenik. A liberalizáció előrehaladásával a szabadpiacra kilépő fogyasztók szabadon választhatják meg, hogy melyik kereskedőtől vásárolnak energiát, a kereskedők pedig maguk szabhatják meg azt, hogy egy-egy fogyasztónak milyen áron értékesítenek. Az egyedi fogyasztási szokások, és így a fogyasztói profilok ismerete különösen az ellátás-alapú szerződéseknél fontos²⁶, hiszen igazságos, *fair* árazás mellett minden egyedi fogyasztónak a saját **fogyasztási szokása** által indokolt árat kell kapnia. Az egyedi fo-

²⁶ Menetrend alapú szerződéseknél az energia mennyisége megfelelő részletezettséggel előre le van bontva, ebben az esetben legfeljebb meghibásodásokból adódó alul- illetve túlvételezések fordulhatnak elő és jelenthetnek kockázatot, de ezeket sok esetben szerződésileg kezelik is.

gyasztói görbére is tekinthetünk azonban úgy, mint olyan fogyasztói igényre, amit nagykereskedelmi termékek vételével (esetleg eladásával) tudunk kielégíteni. Az a fogyasztó, aki csúcsidőben várhatóan többet fogyaszt, magasabb egységáron jut **energiához, mint akinek a csúcsidőszaki fogyasztása alacsonyabb**. Tovább lépve a fogyasztás bizonytalanságának figyelembevétele ügyében, ha a fogyasztás bizonytalanabb a csúcsidőszakban, akkor a bizonytalanabb órák lehetségesen magasabb fogyasztási szintjei miatt több csúcsidőszaki termék allokálásával magasabb árat adunk.²⁷

Az értekezésben természetesen nem célunk sem az előrejelzések, sem különböző kapcsolódó optimalizációs (operációkutatási) logikák kidolgozása, hiszen ezek meghaladnák az értekezés kereteit, a saját kutatási eredményeket tartalmazó fejezet következtetései azonban kitűnő alapot szolgáltatnak az ilyen vizsgálatoknak.

1.2.2. Fogyasztói befolyásolás

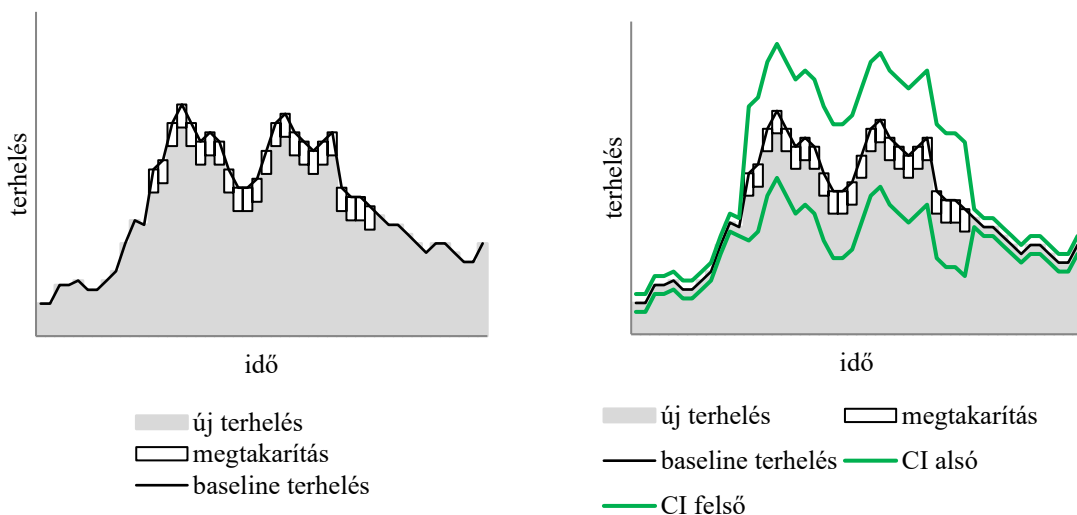
A fogyasztói szokások ismerete természetesen itt sem elhanyagolható – sőt, sokkal inkább központi szerepbe kerül, és természetesen különböző, fogyasztói árazást érintő következmények is vannak. Ilyen, fogyasztói befolyásolást biztosító **tarifális eszközök** például: az ún. zónaidős tarifa (angolul *Time of Use*, ToU), amikor zónaidőszakonként (például csúcs- és völgyidőszakokban) eltérőek az árak; vagy az ún. valós idejű árazás (angolul *Real Time Pricing*, RTP), amikor a beszerzés árkockázatát tulajdonképpen a fogyasztóra hárítják át; de létezhet olyan megoldás is, amikor a profiltól vett eltérést árazzuk csak valós idejűen.

A fogyasztói befolyásolás nemcsak rövidtávon, hanem hosszabb távon is értelmezhető. A különböző **fogyasztás-megtakarítási** projektek esetében például figyelembe kell venni azt, hogy a múltbeli terhelési adatok alapján kalkulált, ún. *benchmark* profilhoz képest elért várható megtakarítás mennyiségében is időszakfüggő a bizonytalanság. Az olyan energia-megtakarításra szakosodott cégeknek (ún. ESCO-k, *energy saving companies* vagy *energy service companies*), akik arra szerződnek fogyasztókkal, hogy bizonyos mennyiségű megtakarítást realizálnak számukra adott időszakon belül, ezt a bizonytalanságot (azaz a megtakarított energiamennyiség nagyságával kapcsolatos bi-

²⁷ Természetesen vannak olyan fogyasztók, **akiket nem mérünk** folyamatosan, és csak rögzített időközönként van mérőóra leolvasás (például diktálás), így az itt bemutatott árazási logika nem alkalmazható; az elszámolás azonban itt is profilok alapján történik. A profilsorozatba sorolás alapja lehet például az, hogy a fogyasztók egy kisebb mintáján – akikről rendelkezünk idősoros mérési adatokkal – kialakítjuk a fogyasztási adatok alapján a profilsorozatokat, majd valamilyen **klasszifikáló, osztályozó** módszertannal megvizsgáljuk, hogy egyéb ún. nem fogyasztási jellegű, de rendelkezésre álló tulajdonságok (pl. tevékenységi kör) hogyan magyarázzák az idősoros mérési adatok alapján kialakított profilsorozatokat (mint osztályokat) és ennek alapján a nem mért fogyasztókat is profilsorozatba tudjuk sorolni.

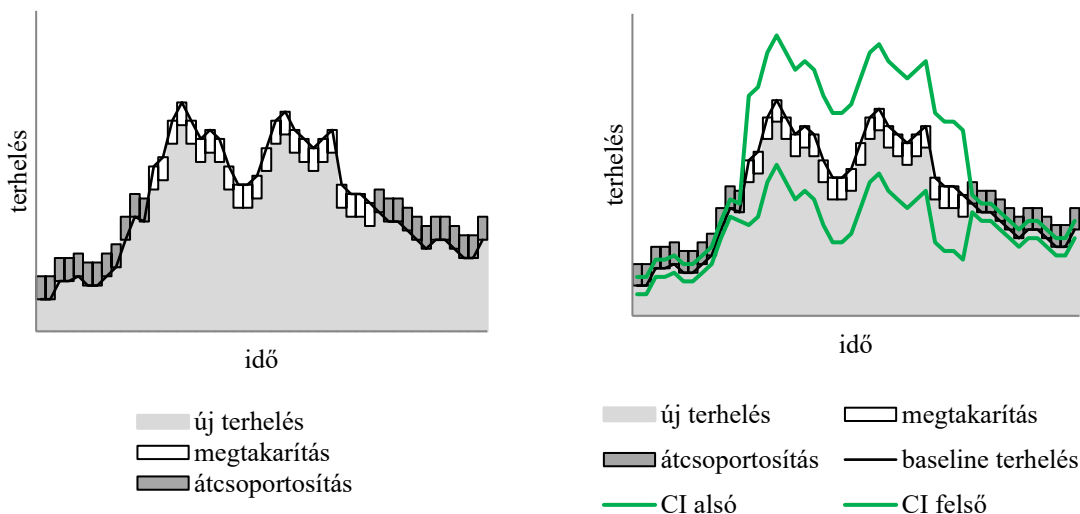
zonytalanságot) mindenképpen figyelembe kell venniük és be kell árazniuk (ld. például *Srivastav et al.* [2013], *Heo et al.* [2012], *Manfren et al.* [2013]).

11. ábra: Fogyasztás-megtakarítás szignifikanciájának értékelése



Forrás: saját készítésű ábra (Excel).

12. ábra: Fogyasztás-átcsoportosítás szignifikanciájának értékelése²⁸



Forrás: saját készítésű ábra (Excel).

A 11-12. ábrák azt szemléltetik, hogy annak eldöntése, hogy például egy fogyasztónál történt-e szignifikáns mértékű megtakarítás vagy fogyasztás-átcsoportosítás, a bizonytalanság explicit kezelése nélkül nem is megválaszolható kérdés.

²⁸ Az ábra jelölései szerint a csúcsidőszakban megtakarított energia átcsoportosításra kerül a völgyidőszakra.

Ez a fajta hosszú távú, akár fenntarthatónak is nevezhető fogyasztói aktivitás nemcsak a megtakarításra törekvésben jelenik meg, hanem például az ún. **háztartási méretű kiserőművi termelés** terjedésében is, amely szintén a fogyasztó oldaláról történő beruházási döntésnek tekinthető. A fogyasztói profil ismerete ilyenkor is fontos, hiszen előnyös az, ha a kiserőmű termelési profilja tud illeszkedni a háztartás fogyasztási profiljához (ld. például *Hino et al.* [2013]).²⁹

Természetesen a fenti alkalmazások megkövetelik, hogy a nagy gyakoriságú, egyedi fogyasztói mérések rögzítése és feldolgozása mind technikailag, mind informatikailag megoldott legyen. Itt ugyanis már nehezen kerülhető meg az, hogy a fogyasztó fogyasztási szokásairól ne álljon rendelkezésünkre a lehető legrészletesebb információ.

1.2.3. Portfólióépítés illetve mérlegkör-összeállítás

Az előző fejezetekben ugyan nem tettünk róla említést, de nem szabad megfeledkezni arról, hogy a fogyasztók jellemzően nem önmagukban, hanem valamilyen portfólió részeként fogyasztanak. Ez a portfólió jelenthet egy kereskedői portfóliót, ahol a fogyasztó egy a kereskedő által leszerződött fogyasztók között, vagy jelenthet egy mérlegkört is, amelyen keresztül majd a menetrendtől vett eltérés kiegyenlítő energiában történő rendezése megtörténik.

Portfólióhatás alatt a gyakorlatban sokszor azt értjük, amikor egy inkább csúcsidőszakban vételező fogyasztót egy inkább csúcsidőszakon kívül vételező fogyasztóval együtt vizsgálva egy zsinór lefutású „portfóliót” kapunk – azaz elsősorban a portfólióhoz a várható értékben történő hozzájárulást.

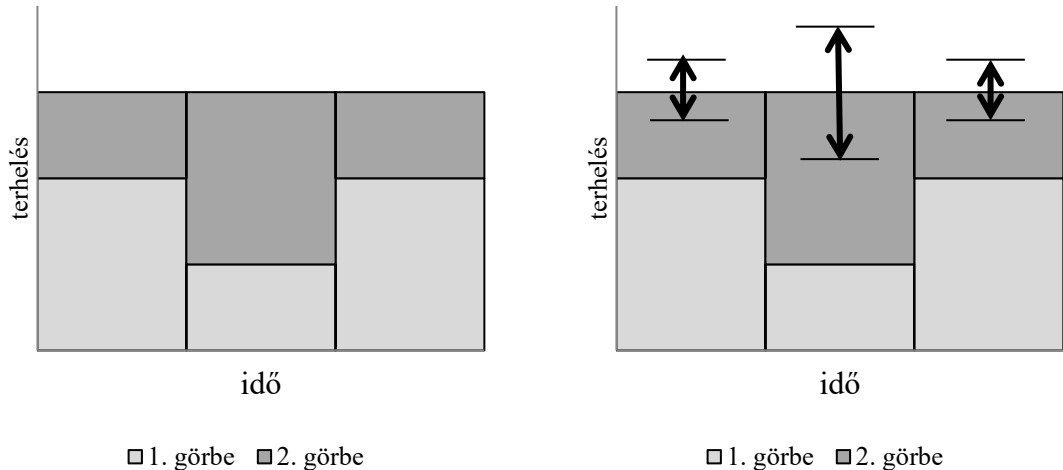
Az 13. ábra bal oldalán ennek egy sematikus ábrázolása látható. Azonban a fogyasztásban minden fogyasztó esetén van bizonytalanság, és ezek fogyasztónként különböző mértékben függhetnek össze (korrelálhatnak) egymással. A portfóliósinten megjelenő bizonytalanságot ezek eredőjeként az ábrán a nyilak és a vízszintes vonalak reprezentálják. Amennyiben a két fogyasztói görbe bizonytalanságai közötti korrelációs együttható értéke kisebb, mint +1, érvényesülni fog az ún. **diverzifikációs hatás**. Azaz a fogyasztással kapcsolatos ilyen jellegű bizonytalanság is diverzifikálható, ugyanúgy, mint például a pénzügyekben sokat vizsgált piaci kockázat (ld. például *Brealey-Myers* [2005]) vagy a likviditási kockázat (ld. például *Váradi* [2012]).³⁰ Ez a diverzifikációs

²⁹ Sokszor előfordul például, hogy akkumulátorokkal biztosítják a megtermelt energia korlátozott tárolhatóságát.

³⁰ Amennyiben egyszerűbb formában, de formálisan szeretnénk felírni, az alábbiról van szó. Tegyük fel, hogy az egyik fogyasztási idősor esetében a véletlen komponens szórása s_1 , a másik esetében s_2 . Amen--

hatás azonban időfüggő, hiszen sem a szórás, sem a korrelációs együttható esetében sem állítható egyértelműen, hogy azok állandóak lennének.

13. ábra: Portfólióhatás sematikus ábrázolása



a) a bizonytalanság figyelembe vétele nélkül

b) a bizonytalanság figyelembe vételével

Forrás: saját készítésű ábra (Excel).

Az itt tárgyaltak a jelentősége abban áll, hogy az egyedi fogyasztókkal kapcsolatos bármilyen **árazási** feladatok elvégzésekor az árban csak a nemdiverzifikálható kockázatnak kell megjelenie, hiszen ez az, ami a portfólióhatás által nem eliminálható. A 9. ábrához visszatérve, a várható terhelés pontos értékével kapcsolatos bizonytalanság pontosan ilyen nemdiverzifikálható kockázatnak tekinthető, amit egyre több görbe portfólióba kapcsolása esetén sem tudunk megszüntetni.

Hasonló, a bizonytalanság és a portfólióhatás kapcsolatát érintő gyakorlati alkalmazás található például Levy [2013] munkájában, amelyben olyan piaci környezetben történő komplex árazási problémával foglalkozik, ahol lehetőség van a fogyasztói befolyásolásra is, azonban a fogyasztással kapcsolatos bizonytalanságot viszonylag egyszerűen, empirikus tapasztalatokkal kevésbé alátámasztottan kezeli.

Természetesen a portfólióhatás vizsgálata sokkal összetettebb, nemcsak a fogyasztás, hanem az árak kockázatával, illetve a nagykereskedelmi termékek elérhetőségével is

nyiben a két fogyasztási idősort az összegét vizsgáljuk, akkor az összeg szórása a varianciák illetve a kétszeres kovariancia összegéből $s_{12} = \sqrt{s_1^2 + s_2^2 + 2 \cdot \rho \cdot s_1 \cdot s_2}$, ahol s_{12} a két fogyasztási idősor összegének a szórása, ρ pedig a két fogyasztási idősor véletlen komponenseinek a lineáris korrelációs együtthatója. Látszik tehát, hogy az összeg szórása csak abban az esetben lesz $s_1 + s_2$, azaz a két szórás összege, amennyiben a véletlenek közötti korrelációs együttható pontosan +1.

számolnunk kell. Azaz, ha egy jellemzően csúcsidőszaki fogyasztót egy jellemzően csúcsidőszakon kívül fogyasztóval egy portfólióba teszünk, a portfólióhatás úgy tud érvényesülni, hogy a portfóliószintű igénygörbe egy zsinórtermékkel lefedhető, a zsinórtermékek pedig általában **likvidebbek**, és a portfólióhatáshoz ez is hozzájárul. Egyébként – a fogyasztási bizonytalanság figyelembe vétele hiányában – a gyakorlatban sokszor csak ezt értik portfólióhatás alatt.

1.3. A profil és profillal kapcsolatos kockázatok

Az alfejezetben definiáljuk, hogy a szakirodalomban milyen profilfogalmakkal találkozhatunk, illetve a fogyasztás alakulásával kapcsolatos kockázatoknak is szánunk egy önálló szakaszt. Az inkább módszertan-hangsúlyos szakirodalom-áttekintésnek külön fejezetben van a helye, itt inkább azzal foglalkozunk, hogy a korábbi munkák fókuszában mi áll.

1.3.1. A fogyasztói profil definíciója

A profil kifejezés nagyon sokféleképpen jelenik meg csak az energetikai szakirodalmat tekintve (is).³¹ Összességében igaz az, hogy profil alatt általában valamilyen tipizálás, tipikus mintázat keresése és megalkotása értendő, amely megjelenhet egyedi fogyasztó szintjén illetve tömegesen is.³²

Barnaby Pitt definícióját felhasználva (Pitt [2000]) profilozás (*load profiling*) alatt annak modellezését értjük, hogy a napi terhelési alak (*load shape*), vagyis a napi terhelési profil (*load profile*) hogyan függ össze különböző olyan tényezőkkel, mint például az időbeliség, időjárás vagy a fogyasztókat jellemző egyéb tulajdonságok. Ezeknek a tényezőknek a kapcsolata a fogyasztással nagyon sok esetben nemlineáris, illetve sok az interakciós kapcsolat (erről ld. részletesen a 2. fejezetet). Az időjárástól függő rész leggyakrabban leválasztásra kerül a fogyasztási idősről (például valamilyen regressziós technikával), és az ettől a hatásától megtisztított időssorral dolgoznak tovább.

A profilozásnak természetesen egy-egy fogyasztói görbe esetében is van értelme, a gyakorlatban azonban nagyon gyakran cél az is, hogy a hasonló profillal, azaz hasonló

³¹ Magának a profilozásnak a célja valószínűleg mindenhol hasonló, de a különböző energianemek területén vannak sajátosságok, különbségek, például a szabályozás, a liberalizáció előrehaladása, de a háttérben zajló fizikai folyamatok tekintetében is. Mivel a dolgozat empirikus részében villamosenergia fogyasztási görbékkel dolgozunk, ezért a legtöbb esetben minden energianemre érvényes illetve villamosenergia specifikus fogalmakkal dolgozunk.

³² Egyébként a **tipikus** kifejezés helyett sokszor az **átlagos** kifejezés jelenik meg a profil mellett, sőt a legtöbb esetben a **módusz**, mint tipikus érték helyett sokszor **átlagolással** kapott értékek szerepelnek – hogy milyen eredmények kapcsán, azt itt szándékosan nem jeleztük.

fogyasztási szokásokkal rendelkező fogyasztókat csoportosítsuk, vagyis fogyasztói profilcsoportokat képezzünk. Ez a feladat is alapvetően módszertani jellegű probléma, hiszen arról szól, hogy idősorokat klaszterezzünk.

Fontos megállapítás, hogy a módszerek nagyon jelentős hányada a mai napig azon alapul, hogy fogyasztónként **egy-egy napi reprezentatív terhelés görbét** (*representative load curve*, RLC, lásd például *Tsekouras et al.* [2008]) készít, amelyekre gyakran szoktunk úgy gondolni, mint tipikus napi fogyasztásokra. Ugyan ezek jól interpretálhatók és a csoportképzés szempontjából is praktikus a használatuk, a legtöbb esetben konstruált, származtatott, a valóságban konkrétan nem realizálódó értékekről beszélünk. A módszerek egy részénél megoldható az, hogy **különböző feltételek mellett** (például nyári, téli, átmeneti időszakai vagy a hét különböző napjaira vonatkozó stb.) **napi profilokat** is előállítsunk.

Espinoza et al. [2005] például a profil (*Typical Daily Profile*, TDP) alatt olyan napi alakot ért, ahol minden exogén változó (szezónális és naptári hatás változók) hatása kiszűrésre kerül, de a regressziós egyenleteknek köszönhetően a különböző feltételek mellett napi profilok természetesen előállíthatóak. A *Carpaneto et al.* [2003] által adott profil definíció (*daily representative load profile*, RLP) megegyezik nagyon sok országban használt profil definícióval. Munkájukban az egyedi fogyasztói profil származtatása különböző exogén változók (például időjárás, tevékenység, rendelkezésre álló elektronikus eszközök stb.) és terhelési feltételek (angolul *loading condition*, például tél/nyár, hétköznap/hétvége stb.) mellett történik. Természetesen ilyenkor több napi terhelési profil is adódik a terhelési feltételeknek megfelelően. *Chicco* [2012] például a fenti definíció szerint úgy állít elő napi profilt (*representative load pattern*), hogy adott terhelési feltételek mellett összegyűjtött néhány nap méréseit átlagolja. *Tsekouras et al.* [2007] és *Tsekouras et al.* [2008] munkáiban az egyedi fogyasztói profil a napi terhelési görbék klaszterezése nyomán áll elő. A legkézenfekvőbb választás profilt a legnagyobb számosságú (leggyakoribb) klaszterből, csoportból választani.

Amennyiben a fókusz nem a hasonló profillal rendelkezők csoportosításán van (sok profilozással foglalkozó munka ezt a célt tűzi maga elé ki), hanem például azon, hogy egy egyedi fogyasztó egy portfólió terhelési profiljához hogyan járul hozzá, már nemcsak a terhelés alakjának, hanem a nagyságának is szerepe van. Ebből adódóan **a dolgozatban profilozás alatt mi annak modellezését értjük, hogy a (napi) terhelési alak illetve (napi) terhelési szint hogyan függ össze különböző tényezőkkel, mint többek között az időbeliség, időjárás vagy esetleg a különböző fogyasztókat jellem-**

zó tulajdonságok. Azaz a Pitt által adott definíciót (Pitt [2000]) kibővítjük még azzal, hogy **a terhelési profilba az alak mellett a szint is beleértendő**, ugyanis bármilyen normálás esetén a fogyasztás szintjével kapcsolatos információ elveszik.

A felsoroltakhoz hasonló, profilozással foglalkozó publikus eredmény magyar viszonylatban nagyon ritka. A VET (2007. évi LXXXVI. törvény a villamos energiáról) értelmében a profil definíciója az alábbi: statisztikai elemzéssel készült normalizált, 1000 [kWh] éves fogyasztásra vonatkoztatott éves felhasználói villamosteljesítmény-igény görbe (ld. 2007. évi LXXXVI. törvény a villamos energiáról). Ez a profil is tipikus napi profilokból áll össze, és ezt a normalizált görbét az adott fogyasztó éves fogyasztásának megfelelően átskálázhatjuk az 1000 [kWh]-ról.

Az előző bekezdésekben leírtakat elsősorban egy konkrét fogyasztói görbe kapcsán tárgyaltuk. Előfordul azonban az is, hogy a profil(ok) definiálása nem egyedi görbe, hanem egy profilsorozatba sorolt görbék összességén kerül kialakításra. Ez főleg akkor hasznos, ha nem rendszeresen mért fogyasztókról beszélünk, illetve ha az egyedi görbék nagyon zajosak, és a jellemző, tipikus tendenciáról kézenfekvő inkább ezek együttese alapján nyilatkozni, ahol az összegzés révén a véletlen hatást – ha nem is a legelegánsabb módon, de – kiszűrtük.

1.3.2. Fogyasztással kapcsolatos kockázatok

Különösen egyedi fogyasztók szintjén a fogyasztással kapcsolatos kockázat nemcsak a (klasszikus értelemben vett) véletlen miatt áll elő, hanem vannak olyan kockázatok is, amelyek modell-alapon jellemzően kevésbé jól leírhatók.

A 14. ábra a különféle kockázatokat jeleníti meg vázlatosan,³³ melyeket természetesen elkülönítve ábrázolunk (és tárgyalunk), a gyakorlatban azonban ezek keverten, akár egymás hatását elmosva jelennek meg.

1.3.2.1. Profilkockázat: alaki és mennyiségi kockázat

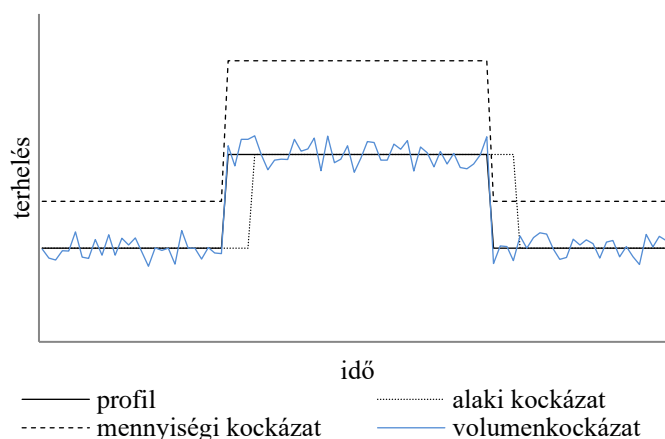
A profilkockázat alatt alapvetően kétféle kockázat érthető (ld. Junghans [2015]): az alaki kockázat, illetve a mennyiségi kockázat.

Alaki kockázat az, amikor a fogyasztó profiljának az alakja változik meg (például üzemrend-változás miatt). Mennyiségi kockázat alatt az értendő, amikor a profil alakja nem változik ugyan, de az összfogyasztás mennyisége igen (például egy gazdasági fel-

³³ Mivel az alaki illetve mennyiségi kockázat sokszor csak manuálisan, *ad-hoc* módon kezelhető, ezért ezek tudományos munkákban kevésbé jelennek meg, a volumen-kockázat viszont sokkal inkább. Elnevezésszerű, fogalmi többértelműségek egyébként vannak, ennek elkerülése érdekében mi az előbb rögzített (egyébként a legtöbb gyakorlati alkalmazással is egybeeső) definíció szerint dolgozunk.

lendüléskor). Olyan mögöttes tényezőkről beszélünk, amelyeknek a hatása – különösen rövidebb fogyasztói görbék esetén – statisztikai vagy hasonló kvantitatív eszközökkel nehezen kimutatható, így ezeket jellemzően egyszerűbb feltételezésekkel élve veszik figyelembe.

14. ábra: Profillal kapcsolatos kockázatok sematikus ábrázolása



Forrás: saját készítésű ábra (Excel).

Profilkockázat alatt tehát olyan kockázatokat értünk elsősorban, amikor maga a fogyasztói profil, tehát a fogyasztásnak az alakja és/vagy a szintje (azaz mennyisége) *strukturálisan* változik meg a különböző külső tényezők függvényében. A profilkockázatba azonban beletartozik az időjárási változók hatása is, hiszen a profilra (alakra, szintre) többek között ezek is hatással vannak, és kockázatot jelentenek. Sztochasztikus változókról van ugyanis szó, ahol a véletlennek szintén fontos szerepe van (erről ld. a 2. fejezetet). Utóbbi viszont már olyan terület, ami sokkal komolyabb módszertani megoldásokat igényel az *ad-hoc*, hüvelykujjszabály-jellegű megközelítésekkel szemben.

1.3.2.2. Volumenkockázat

Amennyiben a profilban nem is történik változás, a fogyasztás természetesen akkor sem mindig a profil által meghatározott pályán történik, attól kisebb vagy nagyobb mértékű eltérések előfordulhatnak. Ennek többféle elnevezése is van, szokás ezt többek között volumenkockázatnak (például *Junghans* [2015]), előrejelzési kockázatnak vagy megbízhatósági kockázatnak (például *Srivastav et al.* [2013]) is nevezni. A volumenkockázatnak tehát alapvetően a (klasszikus értelemben vett) véletlen a forrása, és sokkal inkább modellezhető statisztikai vagy más kvantitatív eszközökkel. Itt a hangsúly tehát a véletlen viselkedésének a helyes leképezésén van.

1.4. A fogyasztói idősorok stilizált tényeinek empirikus vizsgálata

Ahogy a pénzügyi piacokat (ld. *Cont* [2001]) vagy a villamosenergia-árakat (ld. *Marossy* [2010]), vagy akár az energiapiacok és pénzügyi piacok kapcsolatát, egymásra hatását illetően (ld. például *Leng et al.* [2014]), ugyanúgy a fogyasztási idősorok esetében is megfogalmazhatóak ún. stilizált tények (*stylized facts*). Ezek olyan, alapvetően kvalitatív jellegű állítások, amelyek a fogyasztási idősorok többsége esetén igaznak, érvényesnek fogadhatunk el. Bármilyen modellel szemben jogos elvárás az, hogy ezeket a stilizált tényeket minél jobban meg tudja ragadni, amelyek többek között az alábbiak:

- magas időbeli függőség,
- szigorú értelemben vett (erős) stacionaritás hiánya,³⁴
- többszintű (éves, heti, napon belüli) szezonális (*multiple seasonality*),³⁵
- időjárástól való függőség,
- linearitás hiánya és interakciós kapcsolatok megléte,
- heteroszkedaszticitás, azaz időben változó szóródás.

A fentiek klasszikus eszközökkel³⁶ történő, részletes vizsgálata helyett több viszonylag egyszerűbb, de a gyakorlatban mégis ritkán használt ábrát, összefüggést mutatunk be, amelyek nagyon jól reprezentálnak olyan tulajdonságokat, amelyek meglétének vagy hiányának az igazolása sokszor nehéz. Az eredményekben, ábrákban a stilizált tények nyilvánvalóan tetten érhetőek, a tárgyalási mód többek között a saját empirikus kutatási eredmények megalapozását szolgálja úgy, hogy a korábbi kutatási eredmények kielégítő volta, de egyben hiányosságai is jól érzékelhetőek legyenek.

A fejezetben bemutatott ábrák szerveződési logikája a **fogyasztási alak és szint – fogyasztási kockázat – hőmérsékletfüggés** hármass mentén történik, ugyanis az értekezés hozzájárulása a fenti lista utolsó három elemének a klasszikus megoldásokhoz képest rendhagyó kezelésében – az egyszerűség szem előtt tartása mellett – így intuitíve nagyon jól érzékeltethető.

³⁴ Erről ld. a módszertani fejezetet.

³⁵ A villamosenergia-fogyasztást jellemző többszintű szezonális alatt alapvetően az alábbiak értendők:

- az éves szezonális, amelynek alapvetően az időjárás (elsősorban a hőmérséklet) illetve a nap-felkelte-naplemente alakulása a fő mozgatója,
- az elsősorban intézményi, (lényegében) munkarendi és élettani okokra visszavezethető heti szezonális, illetve
- a részben az intézményi, munkarendi és élettani okok, részben az időjárás okozta napi, napon belüli szezonális.

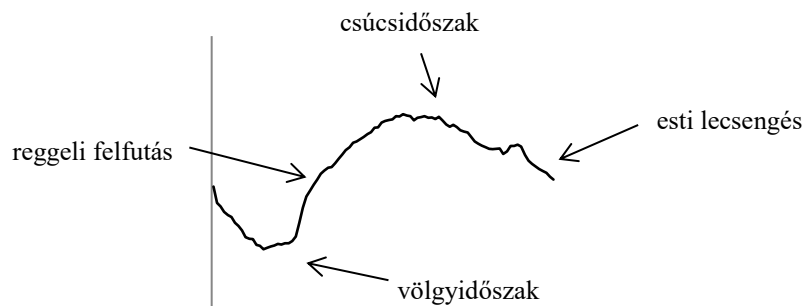
³⁶ Például autokorrelációs együtthatók, stacionaritás tesztek, stb. Ezekről ld. kicsit bővebben például *Hamilton* [1994], *Maddala* [2004], *Ramanathan* [2003] munkáit.

A kutatási kérdések megválaszolásához felhasznált, ebben és a 4. fejezetben is szereplő görbe adatok egy hazai kereskedő cég negyedórás bontású, egyedi fogyasztói terhelési görbe adatai, illetve a MAVIR Zrt. honlapjáról publikusan elérhető magyar országos rendszerterhelés adatok.

1.4.1. Terhelési alak és szint

Nagyon sok profilozással foglalkozó munka a görbék napi alakja tekintetében főként a 15. ábrán látható tulajdonságokat vagy azok valamilyen transzformációját különböztetik meg. Különösen egyedi fogyasztók görbéi esetében azonban könnyen belátható, hogy ezeknek a tulajdonságoknak az egyszerűbb vagy bonyolultabb mutatók formájában történő leképezése sokszor nem elegendő a korrekt leíráshoz. A fejezetben szintvonalábrák felhasználásával mutatjuk be azt, hogy különböző görbék esetén a görbe alakjának illetve szintjének variabilitása milyen jellemzőkkel bírhat.

15. ábra: A terhelési görbék alakját meghatározó tulajdonságok



Forrás: saját készítésű ábra.

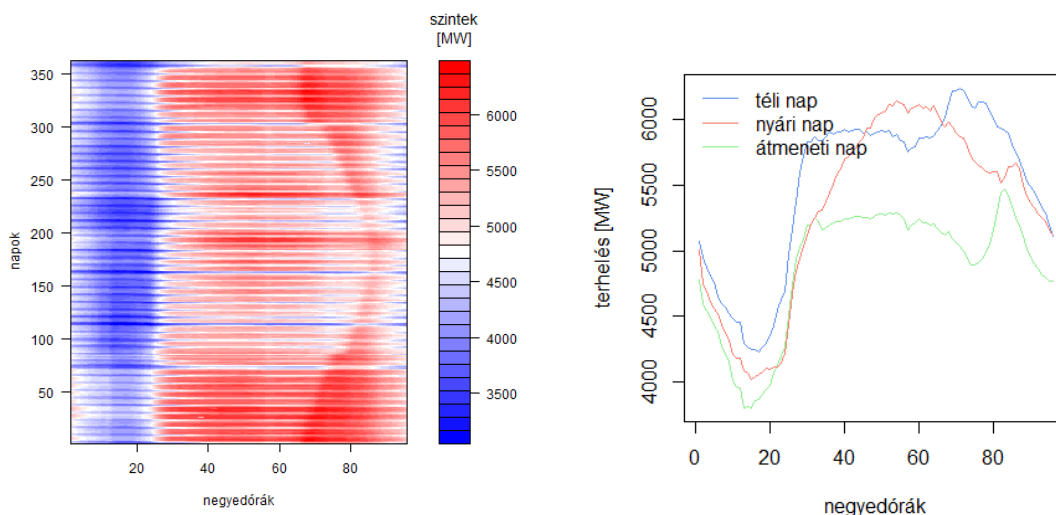
A szintvonalábra lényege, hogy két változó (legyen ez x és y) függvényében egy harmadik változó (legyen ez $z = f(x, y)$) alakulását egy kétdimenziós ábrában úgy képezze le, hogy az azonos z értékeket vonal köti össze. Ezek a vonalak a szintvonalak. A grafikai eszközök illetve vizualizáció fejlődésével elterjedt az, hogy a különböző szinteket különböző színekkel is jelölik, így a harmadik dimenzió (z) térbeli ábrázolása kétdimenziós „lehetőségek” esetén a színekkel praktikusán kiváltható.³⁷ Az itteni ábrákon

³⁷ A fenti rövid leírás alapján már látjuk, hogy a szintvonalábrák több területről is ismerősek lehetnek, használják őket **térképészetben** az adott szélességi illetve hosszúsági körök mellett azonos tengerszint feletti/alatti magasságok kirajzolására, meteorológiában a légköri nyomások ábrázolására stb. Ugyanakkor ilyen szintvonalábrának tekinthetők a mikroökonómiában használt egyenlőtermék-görbék (**izokvant-görbék**) illetve **közömbösségi görbék** is, amely görbék pontjaihoz azonos kibocsátási szint illetve azonos fogyasztói hasznosság tartozik (ld. például Varian [2004]).

x és y változók a nap negyedóráit illetve az év napjait, a z változó pedig az – ezek függvényeként – meghatározott időpontban a terhelés értékét jelölik.³⁸

A magyar országos rendszerterhelés idősorára készített 16. ábra alapján látható, hogy a terhelés legalacsonyabb értékei reggel körülbelül fél 6-6 óráig jellemzőek, majd az azt követő napi felfutás az év egész során szinte pontosan ugyanakkor történik meg (6 és 7 óra között, a nap 20. negyedórája után húzódó fehér sáv). Az ezt követő piros sávok a csúcsideszakokat jelölik, a köztük vízszintesen meghúzódnó fehér sávok pedig a hétfégi napok alacsonyabb terheléseire utalnak.

16. ábra: A magyar országos rendszerterhelés szintvonalábrája és néhány kiválasztott nap idősoros ábrája



Forrás: saját készítésű ábra (R).

Kirajzolódik az ábrán a nap végi második csúcs alakulása, aminek lefutása első sorban a **naplemente** időpontjával függ össze, és főleg hétköznapokon jellemző. Mivel a téli időszakban délután 16-17 óra körül megy le a nap, ezért a napon belüli második csúcs már ekkor elindul, és jóval tovább tart, mint nyáron, amikor a második csúcs nagyságrendileg is és időtartamát (hosszát) tekintve is kisebb. Segítségképpen az 16. ábra jobb oldalán 1-1 kiválasztott téli, nyári illetve átmeneti időszak napi terhelési lefutása is szerepel.

³⁸ A terhelési értékek értéktartományát – 30 osztóponttal – 29 egyenlő hosszúságú (ekvidisztáns) osztályközre osztottuk, és minden egyes terhelési értéket besoroltunk a neki megfelelő osztályközbe. A középső (15.) osztályközbe tartozó értékek fehér színnel jelöltek a szintvonalábrán, és minél magasabb terhelési értékeket tartalmazó osztályközbe tartoznak az értékek, annál sötétebb piros színt, illetve minél alacsonyabb terhelési értékeket tartalmazó osztályközbe tartoznak az értékek, annál sötétebb kék színt kapnak az ábrán. Ennyi osztályköz esetén az ábrák nagyon szép színátmenetes hatást keltenek.

Szembetűnő még a 2011-es nyár két melegebb **hőmérsékletű** időszakának a hatása (a 200-adik illetve a 240-edik napok környékén), amikor a napközbeni csúcs jóval sötétebb, ez az időszak a téli időszakok sötétpiros színéhez közeli szintet kap.

Jól látszik az **óraátállítás** hatása is (ez 2011-ben március 27-én illetve október 30-án volt), amikor a napon belüli második csúcspont indulásában kisebb ugrás van.³⁹

A keskenyebb, jellemzően kék színű sávok a 75., 120., 160., 300. napok környékén illetve az ábra tetején rendre a nagyobb, hosszabb (nem tisztán csak hétvégre eső) **ünnepnapokat** jelölik. A sávok elhelyezkedése az alábbi táblázat alapján könnyen beazonosítható (a táblázat a csak hétvégekre eső ünnepnapokat is tartalmazza):⁴⁰

| Ünnepnap | Az év hányadik napja |
|--|----------------------|
| jan. 1. (szombat) | 1. |
| márc. 15. (kedd, 4 napos hosszú hétvége) | 74. |
| ápr. 24-25. (húsvét) | 114-115. |
| máj. 1. (vasárnap) | 121. |
| jún. 12-13. (pünkösd) | 163-164. |
| aug. 20. (szombat) | 232. |
| okt. 23. (vasárnap) | 296. |
| nov. 1. (kedd, 4 napos hosszú hétvége) | 305. |
| dec. 24-26. (szombat-hétfő, karácsony) | 358-360. |
| dec. 31.-jan. 1. (szilveszter-újév) | 365. és 1. |

Forrás: saját készítésű táblázat.

A 17. ábrán a magyar országos rendszerterhelés mellett egy vállalati portfólió és néhány kiválasztott egyedi görbe szintvonalábrái láthatóak.

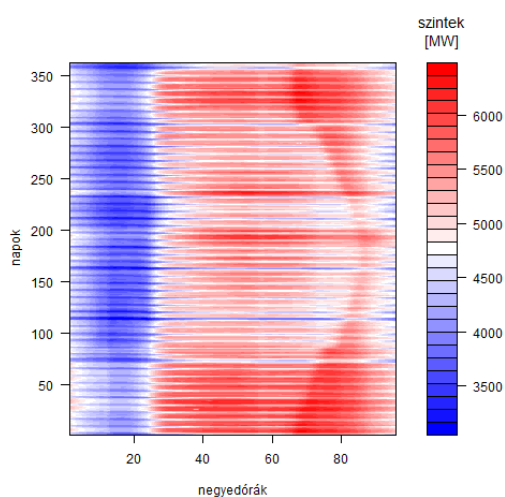
Az ún. **naplemente-hatás** (tehát amikor előbb lemegy a nap, ezért hamarabb fel kell kapcsolnunk a villanyt) a vállalati portfólió szintvonalábráján is megjelenik. Mivel vállalati fogyasztókat tartalmazó portfólióról van szó, ahol a csúcsidezőszak délután 18:00 körül jellemzően véget ér, ezért a hatás valamivel enyhébb. A vállalati portfólió esetén is jól látszódik a **hőmérséklet** hatása a téli, illetve a nyári időszakban, beleértve azt a két nyári időszakot, ami olyan meleg volt.⁴¹ Az augusztus 20-a körüli leállás (amely nagyjából a 200-adik illetve 240-edik napok között van) itt is jól látszik, minthogy a reggeli felfutás is kicsit később történik meg, és a fogyasztás szintje is némileg alacsonyabb.

³⁹ Az ábrán ezek a 100-adik illetve a 300-adik napok környékén keresendők. A 23 illetve 25 órás napok egyébként nem szerepelnek az ábrán, hiszen azok a napok 92 illetve 100 negyedórából állnak.

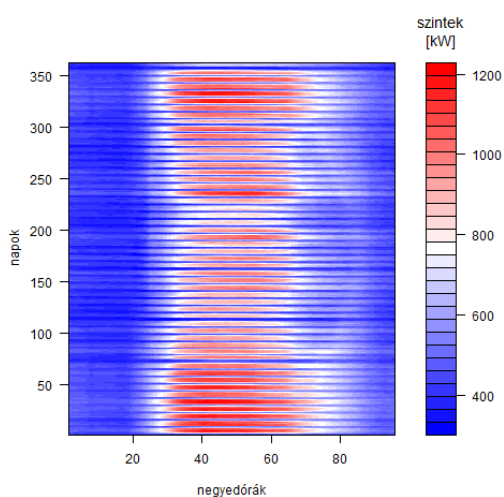
⁴⁰ Augusztus 20-a hétvégi napra esett, de az akörüli (az ún. „két ünnep közötti” időszakhoz hasonló) tipikus leállás természetesen látszódik az ábrán.

⁴¹ Mindkettő kívül esett az augusztus 20-a körüli nagy leálláson.

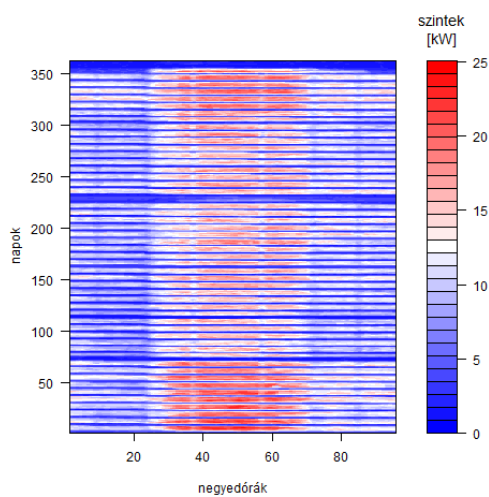
17. ábra: Portfóliók és egyedi fogyasztók terhelési idősorainak szintvonalábrái



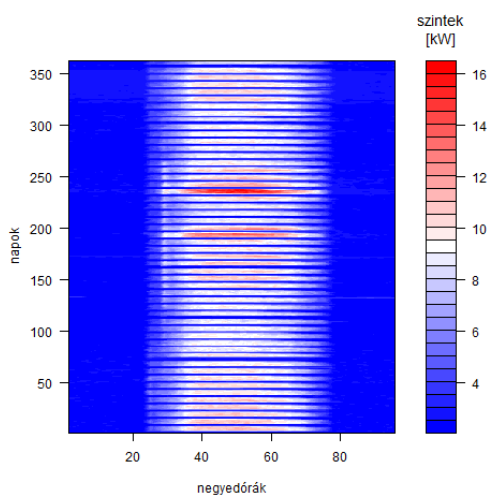
a) Magyar országos rendszerterhelés



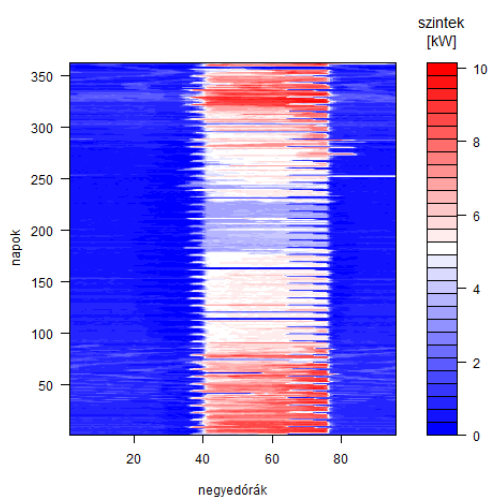
b) Vállalati portfólió



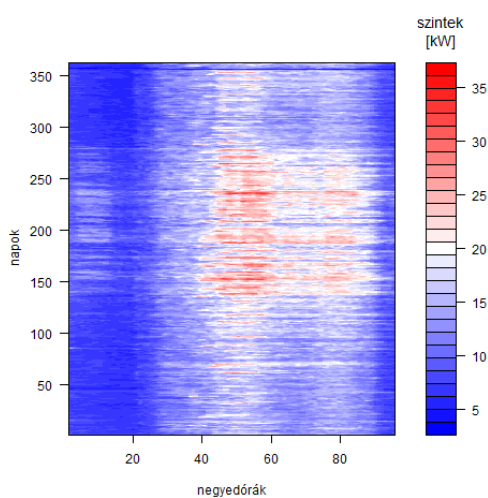
c) V25 görbe



d) V66 görbe



e) V96 görbe



f) V109 görbe

Forrás: saját készítésű ábra (R).

Az eddigiekhez az egyedi fogyasztói görbék szintvonalábrái főleg a fogyasztóspecifikus tulajdonságokat tudják még hozzátenni. A V25 görbe esetében például a téli hőmérséklet hatás erősebb, mint a nyári, és az ünnepi időszakokban gyakorlatilag teljes leállás van. A V66 görbe végig rendkívül szabályos terhelési karakterisztikával bír, és – kicsit érdekes módon – a nyári **hőmérséklet** hatás csak az igazán nagy melegben érvényesül, egyébként nagyon enyhén. A V109 esetében jól megjelenik a nyári hőmérséklet-függés, a V96 esetében pedig a téli. Utóbbinál a 200-adik nap környékén egy hosszabb szinteltolás is megfigyelhető mintegy másfél-két hónapig, valószínűleg ennek valamilyen **strukturális** oka van (pl. egy gyárról lehet szó, ahol néhány üzemet leállítottak).

1.4.2. Terhelési értékek napon belüli eloszlása

A napon belüli eloszlás vizsgálata azért érdemel külön elemzést, mert sokat árulkodik nemcsak a napi profil alakjáról, hanem arról is, hogy a napon belüli eloszlás változása mögött sokszor fundamentális tényezők is beazonosíthatók (pl. sok esetben a hőmérséklet). Ez a fajta vizsgálat azért is előnyös, mert a szakirodalomban a profilok jellemzően valamilyen módon napi bontás mentén készülnek. A napon belüli eloszlás vizsgálata pedig a fogyasztással kapcsolatos bizonytalanság alakulásáról is árulkodik, ami az értekezésben kiemelten kezelendő. Pontosan az eloszlás-központúság módszertani alátámasztása miatt, az alfejezet némileg technikai jellegűnek tűnhet a többihez képest: itt az **eloszlást** jellemző mutatószámok (kvantilisok, terjedelem, minimum, maximum stb.), ezek alakulása, (in)stabilitása legalább olyan fontos, mint a mögöttük meghúzódó **fundamentális** indokok.

Az alfejezet ábráin a következő kiválasztott (természetesen nem reprezentatív) hetek idősorai (illetve a belőlük számolt dobozábrák⁴²) láthatók:

⁴² A **dobozábra** doboz elemét három mennyiség határozza meg: az alsó kvartilis, a medián, valamint a felső kvartilis, így a dobozokon belül egy-egy nap negyedórás terhelési értékeinek a fele szerepel. A dobozon kívül felfelé illetve lefelé nyúló két egyenes a napon belüli legkisebb illetve legnagyobb terhelési értékig húzódik. Amennyiben egy-egy érték a felső kvartilistól legalább az interkvartilis terjedelem másfélszeresével (ún. belső kerítés) nagyobb, vagy az alsó kvartilistól legalább az interkvartilis terjedelem másfélszeresével kisebb, azt az értéket kiugró értéknek tekintjük, és az ábrákon piros ponttal jelöljük (és ebben az esetben a felfelé illetve lefelé nyúló egyenesek is csak az említett másfélszeres távolságig terjednek). Természetesen a belső kerítés megválasztásánál az interkvartilis terjedelem másfélszeresét választani szubjektív döntés, más szorzó is választható, amely nyilvánvalóan hatással lesz az ábrára (alacsonyabb szorzó esetén több megfigyelést tekintünk kiugró értéknek).

Megjegyzés. A kvantilisok a leíró statisztikában használatos mutatószámok, amelyek segítenek abban, hogy valamely ismérv eloszlását – természetesen más mutatószámokkal együtt vizsgálva – jól le tudjuk írni. Az alsó kvartilis az a mennyiség, amelynél a megfigyelések negyede kisebb, háromnegyede nagyobb. A középső kvartilis (medián) az a mennyiség, amelynél a megfigyelések fele kisebb, fele nagyobb. A felső kvartilis az a mennyiség, amelynél a megfigyelések háromnegyede kisebb, negyede pedig nagyobb.

| Időszak (Szezon) | Kezdet | Vég |
|----------------------------|-------------------------------------|-----------------------|
| Téli | 2011-01-17 00:00 CET ⁴³ | 2011-01-23 23:45 CET |
| Nyári | 2011-07-18 00:00 CEST ⁴⁴ | 2011-07-24 23:45 CEST |
| Átmeneti (tavaszi és őszi) | 2011-04-18 00:00 CEST | 2011-04-24 23:45 CEST |

Forrás: saját készítésű táblázat.

Az 18. ábra a) részének bal oldali felén a kiválasztott hetek negyedórás historikus terhelései, a jobb oldali felén pedig a napi bontásban elkészített dobozábrák szerepelnek a magyar országos rendszerterhelésre vonatkozóan.

Az ábráról leszűrhetőek olyan következtetések, amelyeket a szintvonalábrák esetében is láttunk már (például télen közvetve vagy közvetlenül a fűtési hatás (*heating effect*) illetve a világítási célú felhasználás (*illumination effect*) miatt magasabb a fogyasztás szintje, valamint nyáron főként a légkondicionáló készülékek használata miatt az ún. hűtési hatás (*cooling effect*) emeli meg a csúcsidőszaki fogyasztás szintjét).

Az is látható, hogy télen az ún. völgyidőszaki fogyasztás szintje is magasabb (nyilvánvalóan a fűtési hatás miatt), illetve hogy szezononként különböző a reggeli felfutás jellege és a napon belüli csúcsok időbeli és szintbeli elhelyezkedése (például nyáron a hűtési hatás miatti délutáni csúcs magasabb csúcsterhelést eredményez, mint az esti világítási hatás, amikor már a hűtési hatás sem érvényesül annyira).⁴⁵

A terhelési értékek napon belüli eloszlása szezonális időszaktól függetlenül **aszimmetrikus** (jobbra ferde), ugyanis a napon belüli medián terhelés feletti (lényegében csúcsidőszaki) értékek tartománya jóval szűkebb, mint a medián alatti értéké. Hétvégi napokon ez az aszimmetria valamivel gyengébb, hiszen a **hétvégi** terhelés az egyébként csúcsidőszaknak megfelelő órákban alacsonyabb a **hétköznapi**okhoz képest. Szezonális időszaktól függetlenül a hétfői napok terjedelme a legszélesebb, ami azzal magyarázható, hogy a nap eleji felfutás ilyenkor a hétvége által indokolt alacsonyabb szintről kell, hogy megtörténjen.

Úgy tűnik, hogy a napon belüli aszimmetria másik alapvető forrása a **hőmérséklet**, ugyanis a téli napokon illetve a melegebb nyári napokon az eloszlás jóval aszimmet-

⁴³ CET: Central European Time, közép-európai idő (UTC + 1 óra), UTC: egyezményes koordinált világidő

⁴⁴ CEST: Central European Summer Time, a nyári időszámítás alatt a CET helyett ez van érvényben (UTC + 2 óra)

⁴⁵ Esténként a lakossági hűtési célú felhasználás inkább csak nagy melegekben érvényesül, de az nagyságrendekkel elmarad a csúcsidőszaki (munkahelyi) felhasználástól.

rikusabb, mint egyébként⁴⁶, hiszen a hőmérséklet hatás (lehet szó akár hűtési vagy fűtési hatásról) befolyásolja, elhúzza a csúcsidőszaki terhelési szintet. Ezek a napon belüli eloszlást befolyásoló hatások természetesen keveredhetnek. A C) mellékletben szereplő táblázatban a fenti megállapítások számszakilag is jól végigkövethetőek. A lényeg természetesen nem a számok pontos értékén van, de érdemes lehet például akár a relatív szórás, akár a ferdeségi mutató alakulásában lévő tendenciát ott is végigkövetni és visszafejteni.

A dobozábrák alapján elmondható még, hogy – a dobozábra definíciója szerint értelmezett – **kiugró** érték alig szerepel, az is inkább az átmeneti időszakban, a nem csúcsidőszaki órákban. A kiugró értékek kezelésével érdemes azonban vigyázni, hiszen a dobozábrák készítése napi lebontásban történt, így a kiugró érték is az egy adott naphoz tartozó doboz elhelyezkedése és az annak megfelelő interkvartilis terjedelem függvényében kerül értékelésre, sem időbeliség, sem fundamentális okok nem kerülnek figyelembe vételre.

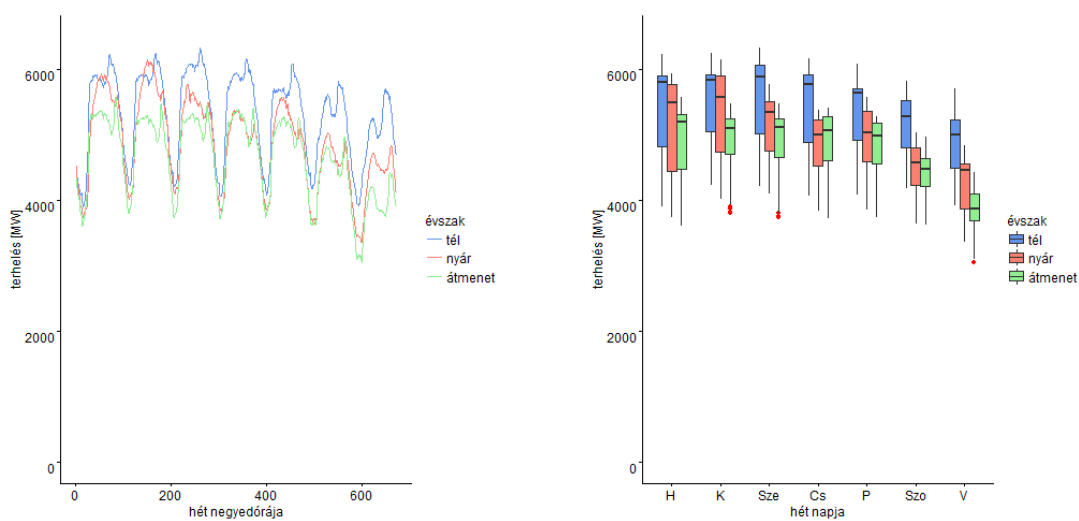
A 18. ábrán az országos rendszerterhelés mellett még a vállalati portfólió és egyedi görbék ábráit is közöljük, így a hasonlóságok illetve különbségek nagyon könnyen felismerhetők.

A téli fűtési és a nyári hűtési hatás szerepe a vállalati portfólió esetében is hasonlóan alakul, mint az országos rendszerterhelés esetében. A hétvégi terhelés **szintje** azonban jóval alacsonyabb a hétvégéken, mint a hétköznapokon, és ilyenkor a fűtési, hűtési hatás is csak nagyon kis mértékben érvényesül. Fontos megállapítás, hogy a magyar rendszerterheléssel szemben hétköznapokon a napon belüli terhelések **terjedelme** illetve a medián elhelyezkedése jóval **stabilabb**. Sokkal markánsabb és egyértelműbb a pénteki délutánok alakulása, a hét végén történő lecsengés: az idősoros ábrán is látható, hogy a pénteki csúcsterhelés jóval rövidebb, és a dobozábrák alapján *ceteris paribus* a terjedelem is kisebb, és a napi maximum is alacsonyabb.

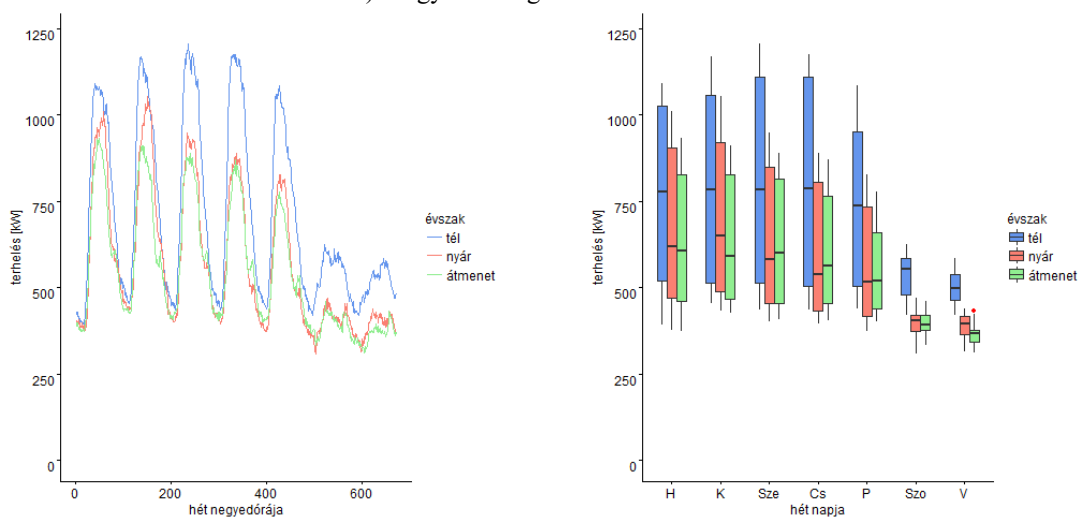
⁴⁶ Július 2. hete, illetve az ábrákon látható júliusi hét első két napjai voltak ezen a nyáron a legmelegebb napok. Ezen kiválasztott három hét napi középhőmérsékleteinek az alakulását mutatja az alábbi táblázat °C-ban:

| Szezon / Hét napja | H | K | Sze | Cs | P | Szo | V |
|--------------------|------|------|------|------|------|------|------|
| Tél | 3,5 | 3,3 | 2,4 | 1,3 | 0,5 | -1,7 | -1,0 |
| Nyár | 13,3 | 15,7 | 14,3 | 16,4 | 17,5 | 16,4 | 13,0 |
| Átmenet | 23,8 | 24,7 | 18,9 | 15,4 | 16,8 | 17,2 | 17,6 |

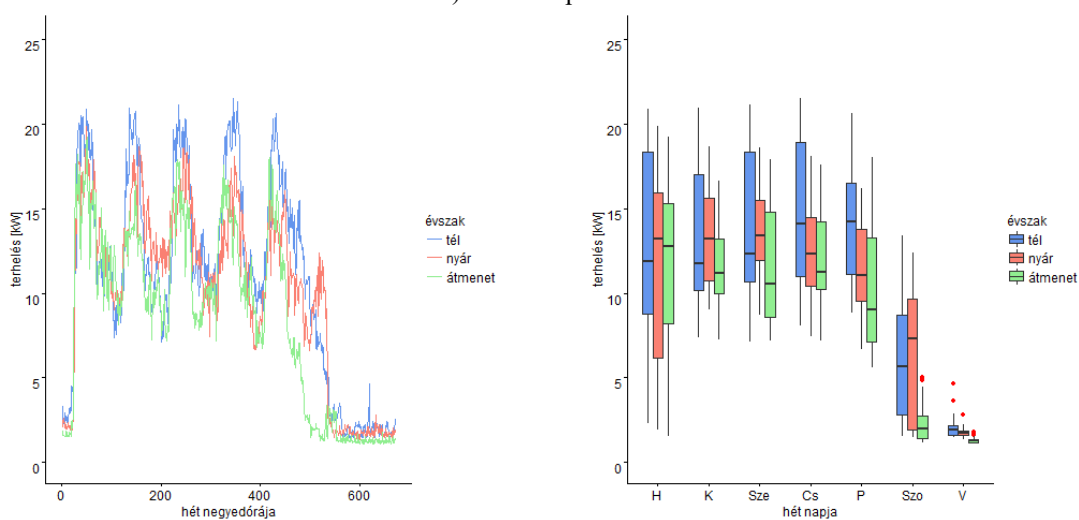
18. ábra: Portfóliók és egyedi fogyasztók heti terhelési idősorai és a kapcsolódó dobozábrák



a) Magyar országos rendszerterhelés



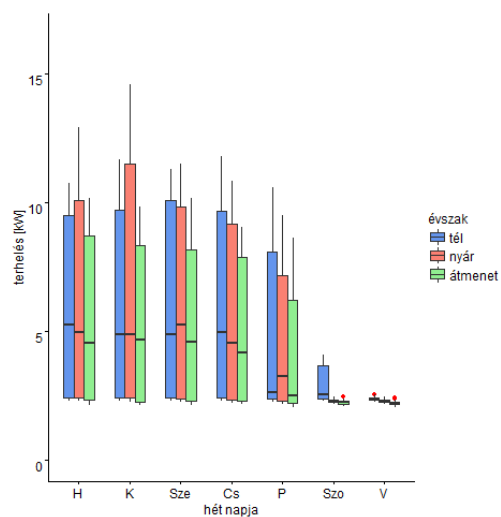
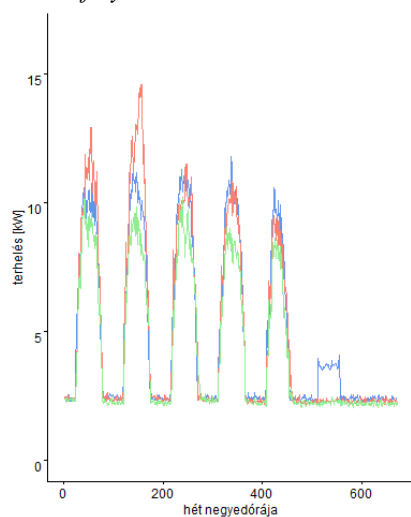
b) Vállalati portfólió



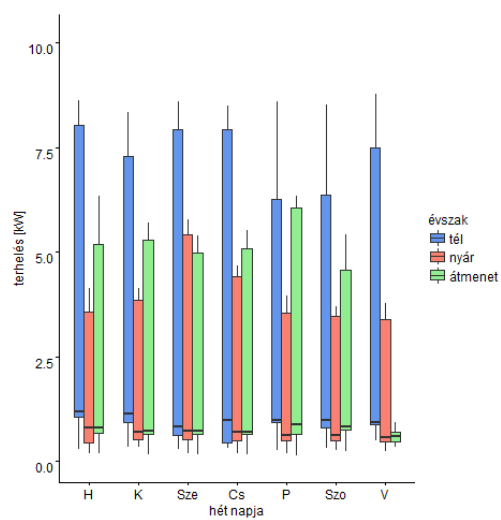
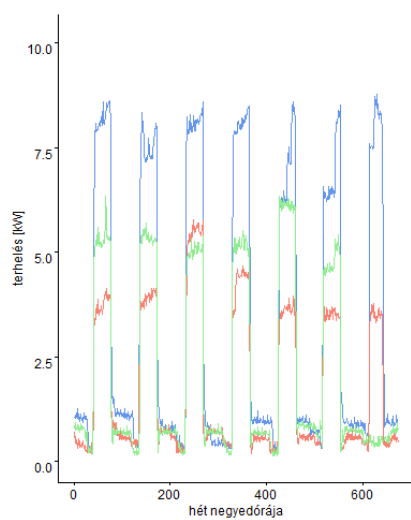
c) V25 görbe

Táblázat folytatása következő oldalon.

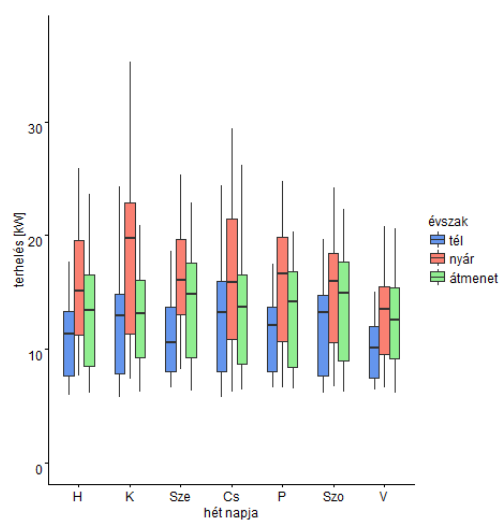
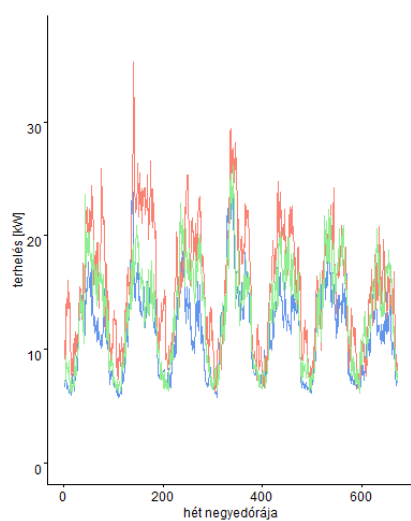
Táblázat folytatása az előző oldalról.



d) V66 görbe



e) V96 görbe



f) V109 görbe

Forrás: saját számítások (R) és saját készítésű ábra (R).

Rátérve az egyedi fogyasztók terhelési görbéire, a V25 görbe esetében a csúcsidőszaki terhelésekben vannak inkább az évszakok között különbségek, és mindkét ábra azt sugallja, hogy inkább csak téli hőmérséklettel kell számolnunk. A terhelési görbe érdekességét inkább a gyors hétfői felfutás, illetve a (szombat közepéig tartó) lassú lecsengés teszi különlegessé, egyébként a hétköznapi egyformának tekinthetők. Így a **görbe lefutását** leginkább úgy tudjuk **jellemezni**, hogy van egy hét eleji felfutási, egy hétvégi lecsengési időszak, hétköznapi időszakok (amit télen a hőmérséklet még eltol), illetve van egy hétvégi, gyakorlatilag konstans terheléssel jellemezhető időszak.

A V66 görbére vonatkozó ábrák azt sugallják, hogy a hétköznapi terhelések **eloszlása** rendkívül **stabil**, és a csúcs időszakon kívül hétköznapiakon illetve egész hétvégén a terhelés gyakorlatilag konstans. Hőmérséklettel csak nyáron van (a hétfői és keddi, már korábban is említett, melegebb napokon magasabb a csúcsidőszaki terhelés értéke), illetve a téli csúcsterhelés szintje is valamivel magasabb, mint az átmeneti időszaké.

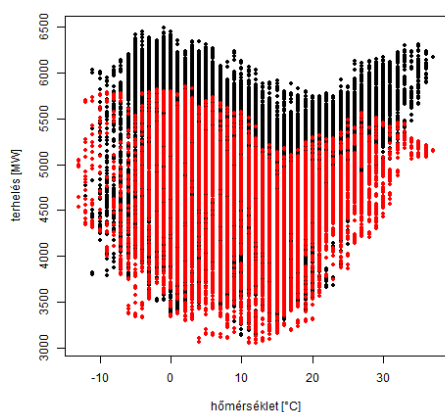
A V96 görbénél csak az 1 [kW] körüli csúcsidőszakon kívüli, és az 5-8 [kW] körüli csúcsidőszaki terhelések **rezsímjei** változtatják egymást, elég **szabályosan**, még a hétvégék és a hétköznapiak sem különböznek. Ezt tükrözendő a napon belüli eloszlások rendkívül stabilak, és végig érvényesül az erős baloldali aszimmetria (hiszen minden nap a csúcsidőszaki értékek aránya jóval alacsonyabb). A hőmérséklet hatását tekintve csak a téli fűtési hatás jelenik meg.

A V109 görbét az előzővel szemben csak nyári hőmérséklettel jellemzi, viszont ez csak a csúcsidőszakokban jellemző, egyébként évszaktól függetlenül a napi minimum szint (néhány melegebb nyári naptól eltekintve) végig konstans.

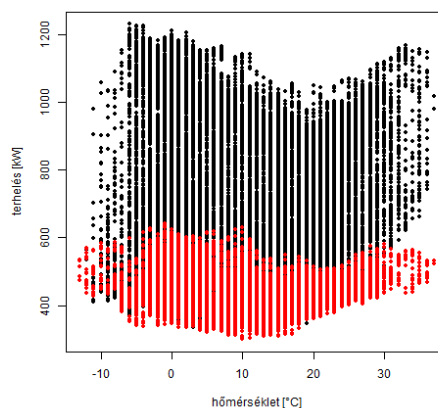
1.4.3. Időjárásfüggés

Szakirodalmi eredményekre (ld. majd a 2. fejezetet) támaszkodva kijelenthető, hogy villamosenergia-fogyasztásra az időjárási tényezők közül a hőmérséklet van a legnagyobb hatással. A 19. ábra alapján azonban jól látható, hogy ez a hatás görbénként nagyon különböző. A vizsgált időszakban az országos rendszerterhelés esetében a fűtési küszöbérték (amely alatt érvényesül az ún. fűtési hatás) kb. 12-13 °C-nál, a hűtési küszöbérték (amely felett érvényesül az ún. hűtési hatás) 21-22 °C-nál van, a kettő között nem beszélhetünk hőmérséklettel. A fogyasztói portfólió esetében nagyjából ugyanezek a küszöbértékek tűnnek érvényesnek, és érdemes megjegyezni, hogy a gya-

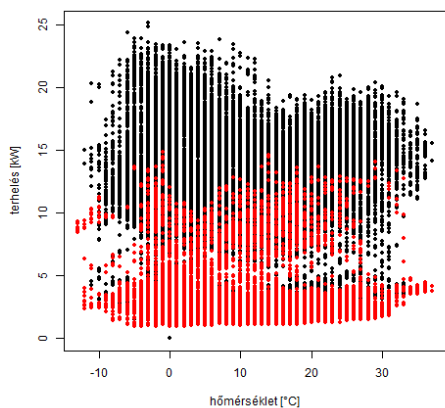
19. ábra: Portfóliók és egyedi fogyasztók terhelése a hőmérséklet függvényében



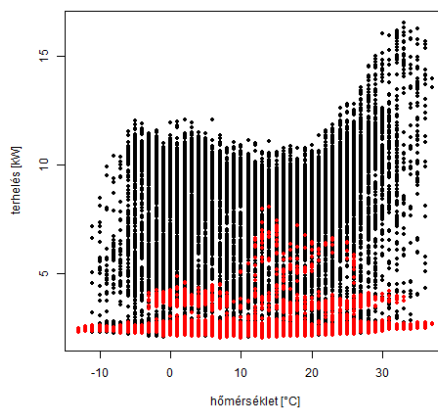
a) Magyar országos rendszerterhelés
(hétköznapok (fekete) – hétvégék (piros))



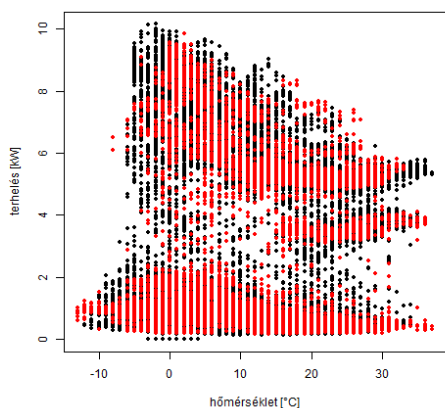
b) Vállalati portfólió
(hétköznapok (fekete) – hétvégék (piros))



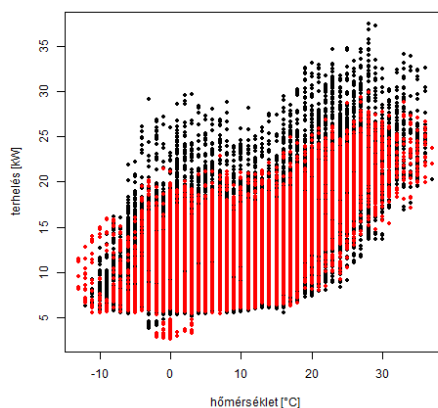
c) V25 görbe
(hétköznapok (fekete) – hétvégék (piros))



d) V66 görbe
(hétköznapok (fekete) – hétvégék (piros))



e) V96 görbe
(hétköznapok (fekete) – hétvégék (piros))



f) V109 görbe
(hétköznapok (fekete) – hétvégék (piros))

Forrás: saját készítésű ábra (R).

korlatban is általában ezeknek a figyelembe vételével számolják az ún. fűtési illetve hűtési napfokokat (*heating degree-day*, röviden *HDD* illetve *cooling degree-day*, röviden *CDD*, ezekről ld. még a 2. illetve 4. fejezeteket is).

Az országos rendszerterhelésnél a hűtési hatás esetében valamivel erősebbnek tűnik $+1\text{ }^{\circ}\text{C}$ hatása, mint a fűtési hatása esetében. Azonban az is látható, hogy ez csak a hétköznapiakra jellemző. Hétvégéken a fűtési hatás viselkedése nagyjából hasonló, mint hétköznapiokon, a hűtési hatás viselkedése kicsit enyhébb, sőt, kevésbé egyértelmű tendenciával jellemezhető.

A vállalati portfólió esetében a hétköznapi fűtési illetve hűtési hatás nagyjából egyforma, hétvégéken azonban mindkettő gyengébb.

Természetesen egyedi görbék esetében már valamivel bonyolultabb a helyzet az ilyen nagyságrendi elemzések tekintetében, ráadásul jóval nagyobb például a véletlen szerepe. Hasonlóan kirajzolódik azonban, hogy hétköznapiokon illetve hétvégéken a hőmérséklet függvényében a terhelési értékek hogy **tömörülnek, csoportosulnak**, mikor kisebb vagy nagyobb a szóródás.⁴⁷ Ez utóbbi szemléletmód az értekezés hátralévő részében is fontos lesz.

1.4.4. Következtetések

A következtetések megfogalmazása kapcsán érdemes visszatérünk ahhoz az állításhoz, hogy a bemutatott eredmények **fogyasztási alak és szint – fogyasztási kockázat – hőmérsékletfüggés** hármass mentén próbálták az értekezés szempontjából fontos jelenségekre felhívni a figyelmet.

A szintvonalábrák – ahogy részben az elnevezés is mutatja – a fogyasztás szintjéről és alakjáról nyújtottak kompakt képet, beleértve a görbénként nagyon különböző mértékű naptári hatás (hétköznapiok, hétvégék, ünnepnapok hatásának) kimutatását, a hőmérsékletfüggést, a világítási (vagy naplemente-) hatást, egyéb **strukturális** változások hatását.

Mivel a legtöbb, terhelési görbék profilozásával foglalkozó tudományos munka a napi profilok kialakításában, képzésében gondolkodik, ezért tartottuk célravezetőnek a napi terhelési értékek eloszlásának vizsgálatát különböző szezonokra. Fundamentálisan lényegében hasonló állításokat fogalmazhatunk meg, mint a szintvonalábrák esetében, azonban itt sokkal jobban ellenőrizhető, hogy mennyire tekinthető **stabilnak** vagy **in-**

⁴⁷ A V25 esetében például a hétvégi napokon gyakorlatilag konstans terhelés van, a 19. c) ábra azért zavaró, mert a hétvégi lecsengés a szombati nap közepéig tart, ebben az időszakban azonban a hőmérsékletnek már – érthető módon – nem nagyon van hatása a terhelésre.

stabilnak egy-egy nap terhelési értékeinek az **eloszlása**, illetve milyen tényezők (pl. hőmérséklet) lehetnek hatással egy-egy nap terhelési értékeinek eloszlására. Így már nemcsak a fogyasztási szintről és alakról, de annak kockázatáról, bizonytalanságáról is van információnk. Természetesen a különféle kockázatok (mint a profil- és volumen-kockázat) ilyen eszközökkel még nem számszerűsíthetők, legfeljebb sejtések fogalmazhatóak meg azok viselkedését illetően.

Az eddigieket kiegészítendő hőmérséklet-terhelés *scatter plot*-okban való gondolkozás megjelenik az értekezésnek majd abban a részében, ahol az ún. keverékmodellekkel dolgozunk. Az ottani eredményeket megelőlegezendő itt megvizsgálhattuk azt, hogy a terhelési görbék negyedőrái a **hőmérséklet-terhelés dimenzióban hogyan csoportosulnak, tömörülnek**.

Igaz ugyan az, hogy **a napi profilok képzése is bizonyos értelemben helyes – és mindenekelőtt jól és könnyen értelmezhető –, ugyanakkor nem minden esetben hatékony, ugyanis a tipikus napi profilok nem feltétlenül a napi bontások mentén képződnek**. Amennyiben a profilok előállítása napi diszkretizálással történik, az feleslegesen jelentheti sok paraméter becslését, de félrevezető következtetések levonását is eredményezheti. Másként megfogalmazva fogyasztási értékek csoportosulását, tömörülését nem feltétlenül napi profilgörbék mentén a leghatékonyabb modellezni, bár például a szintvonalábrákat vertikálisan vizsgálva ez jó megközelítésnek tűnik: annak egy (vagy néhány), valamilyen formában tipikusnak tekinthető sorához (többé-kevésbé) hasonló a többi is, kisebb-nagyobb eltérésekkel ugyanakkor vannak a felfutások, csúcsok, lecsengések. Az itteni eredmények tehát mindenképpen csak tájékoztató jellegűek lehetnek, erre hívja fel a figyelmet például az is, hogy a napi diszkretizálásból adódóan a kiugró értékek itt bemutatott módon történő azonosítása eléggé nehézkes és félrevezető.

A fenti elemzések különböző nézőpontokból ugyan, de a H1 hipotézis elfogadása mellett szólnak, miszerint a villamosenergia fogyasztási görbék esetében jellemzően a napon belüli szezonális a görbék varianciájának legjelentősebb forrása, azaz a H1 hipotézist nem tudjuk elvetni. Ez látszódik egyrészt a naponkénti dobozábrák alakulásából, illetve a szintvonalábrák színskálájának rendeződéséből is. Ezt az eredményt egyébként egy nagyon egyszerű varianciaanalízissel is megkaphattuk volna (ld. a C) mellékletet), de a fejezet **kiemelt célja** az volt, hogy **az idősorok alakulásának fő mozgatórugóit, fontosabb összefüggéseit feltárja** és megismerje, és különösen az **egyedi fogyasztók viselkedésének heterogenitásáról** adjon némi képet, kihangsúlyozza az értekezés által megfogalmazott célok komplexitását.

2. KORÁBBI KUTATÁSI EREDMÉNYEK A FOGYASZTÓI PROFILOZÁS TERÜLETÉRŐL

Az egyedi fogyasztói profilok meghatározásával szorosan együtt jár a hasonló profillal rendelkezőket tömörítő, ún. fogyasztói profilcsoportok létrehozása. Ezért áttekintjük a fejezetben azt is, hogy az idősorok klaszterezésével, illetve szűkebben értelmezve az energiapiaci (fogyasztási) idősorok klaszterezésével kapcsolatban eddig milyen tudományos eredmények születtek. Ehhez kapcsolódóan az ún. görbetulajdonság lényegében a konkrét felhasználás esetén értelmezett tipikus fogyasztási mintázatra vonatkozó tömörített információt jelenti.

Ezek a feladatok szorosan összefüggenek az időjárás hatásának a megfelelő kezelésével, így a fejezet második részében ezzel foglalkozunk, és egy korábbi saját publikációra építve mutatjuk be, hogy a profilozás során leggyakrabban használt megoldások az időjárás hatásának a kezelésére milyen következményekkel járnak mind a profilozás, mind az azzal kapcsolatos bizonytalanság modellezésében és értékelésében.

2.1. Fogyasztói idősorok kétlépéses klaszterezése

Az alfejezetben áttekintjük, hogy a profilozással foglalkozó szakirodalom alapján a fogyasztói profilok illetve profilcsoportok képzését illetően milyen általános keretrendszerben gondolkozunk. Ez praktikusán a kétlépéses vagy kétlépcsős klaszterezés (*two-step* vagy *two-stage clustering*) fogyasztói profilozásra történő adaptálását jelenti, ahol az alkalmazott klaszterezési technikák (2. lépés) mellett a fogyasztói görbéket jellemző ún. görbetulajdonságoknak, mint tömörített információnak az előállítása⁴⁸ (1. lépés) tekinthető inkább hangsúlyosnak és iparág-specifikusnak. Az áttekintést az idősorok klaszterezéséről szóló rövid bevezetővel kezdjük.

2.1.1. Idősorok klaszterezése általában

Mint ismert, a klaszterezés célja alapvetően a megfigyelések csoportokba szervezése olyan módon, hogy azok a csoportokon belül minél inkább hasonlítsanak egymáshoz, viszont a csoportok egymástól minél inkább különbözzenek. Az idősorok vagy görbék klaszterezése (angolul *time series clustering*, *curve clustering*) esetében használt algoritmusok nagyon sok esetben – némi idősoros tulajdonság kihasználása mellett –

⁴⁸ Mivel a fogyasztási idősorok esetében jellemzően idősoronként nagyon sok adatról van szó (pl. egy éves fogyasztási görbe – nem szökőévben – $365 \cdot 24 = 8760$ darab órából vagy $365 \cdot 96 = 35040$ darab negyedórából áll), ezért első lépésként mindenképpen szükséges valamilyen információ-tömörítést végezni.

hasonló megoldásokon alapulnak, mint a klaszterezési megoldások általában (ld. például Liao [2015]).

A témához kapcsolódóan két fontosabb, összefoglaló tanulmány (Liao [2005] illetve Jacques-Preda [2013]) az alkalmazott módszereknek az 3. táblázatban szereplő lehetséges csoportosítását mutatja be.⁴⁹

3. táblázat: Idősoros klaszterezési algoritmusok lehetséges csoportosításai

| Liao csoportosítása | Jacques és Preda csoportosítása |
|--|--|
| 1. alapadatokon történő klaszterezés (<i>raw-data clustering</i>) | 1. alapadatokon történő klaszterezés (<i>raw-data clustering</i>) |
| 2. tulajdonság alapú klaszterezés (<i>feature-based clustering</i>) | 2. kétlépcsős klaszterezés (<i>two-stage clustering</i>) |
| 3. modell alapú klaszterezés (<i>model-based clustering</i>) | 3. nemparaméteres klaszterezés (<i>non-parametric clustering</i>) |
| | 4. modell alapú klaszterezés (<i>model-based clustering</i>) |

Forrás: saját szerkesztésű táblázat.

Az első megoldás alatt mindkét esetben olyan klaszterezési megoldások szerepelnek, amelyek az alapadatokkal (az eredeti idősorral) dolgoznak, vagy az időtartományon vagy a frekvenciatartományon értelmezve azt. A tulajdonság alapú illetve a kétlépcsős klaszterezés lényege, hogy első lépésben megtörténik az alapadatokból valamilyen releváns tulajdonság (*feature*) kinyerése, majd a klaszterezés ebben a térben történik meg. A nemparaméteres megoldások közé sorolhatóak többek között azok a megoldások, amelyek a klaszterezést különböző távolsági vagy hasonlósági mértékek felhasználásával végzik el. A modell alapú klaszterezés előnye, hogy a klaszterezés és a dimenziócsökkentés egy lépésben történik meg az adatokhoz legjobban illeszkedő modell becslése által. Némileg félrevezető módon Liao ide sorolja azt is, amikor (a tulajdonság alapú klaszterezéshez hasonlóan) az idősorokra illesztett modell valamilyen részeredményei (mint idősoros modellből kinyert eredmények) alapján klaszterezünk.

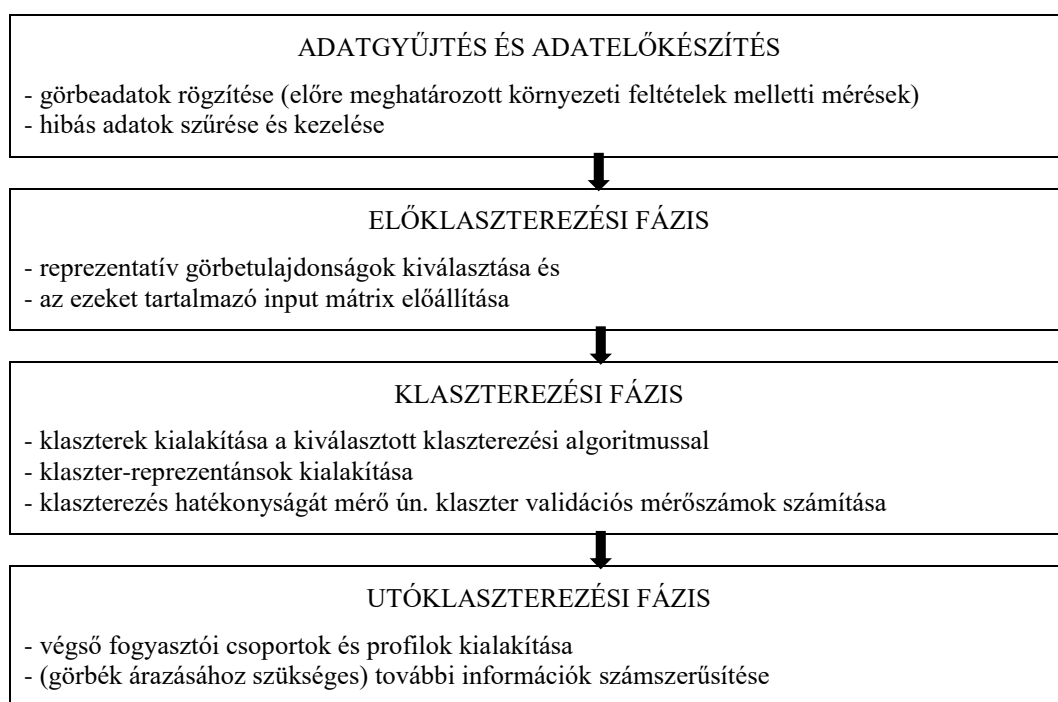
⁴⁹ Jacques és Preda (Jacques-Preda [2013]) cikkükben függvények klaszterezési technikáit (*functional data clustering*) foglalják össze, de mint megjegyzik, a legtöbb esetben az elemezni kívánt adatok az idő – mint folytonos változó – függvényeként jelennek meg. Ez a terület a ma egyre népszerűbb függvény-adatelemzés (*functional data analysis, FDA*) egyik részterületének tekinthető.

2.1.2. A profilozás általános keretrendszere

Az 20. ábrán szereplő, Chicco (ld. *Chicco* [2012]) által használt keretrendszer a kétlépéses vagy kétlépcsős klaszterezés fogyasztói profilozás területére történő alkalmazása, adaptálása.

Az előklaszterezési fázis a klaszterezés során használni kívánt ún. görbetulajdonságok az előállítását jelenti. Ezek azok a jellemzők (paraméterek), amelyek feltételezésünk szerint jól reprezentálják a fogyasztás alakját.⁵⁰ Ezt követi a bemeneti (input) adatok összeállítása.

20. ábra: Fogyasztói görbék profilozási folyamatábrája



Forrás: *Chicco* [2012] alapján saját szerkesztésű ábra.

Maga a klaszterezés nemcsak a klaszterek valamely kiválasztott klaszterezési technika felhasználásával történő kialakítását jelenti, hanem az egyes klasztereket jellemző, ún. klaszter-reprezentánsok kiválasztását, és a klaszterezés jóságát mérő mutatók alapján a klaszterezés megfelelőségének ellenőrzését is.

Az utolsó lépés az ún. utóklaszterezési fázis, ami a végső fogyasztói csoportok és végső fogyasztói profilok megalkotását⁵¹ jelenti. A végső fogyasztói csoportok száma

⁵⁰ A görbetulajdonságok az alkalmazások jelentős részében az egész görbét jól jellemző, egyetlen egy napi profilt jelentenek, ahogy azt korábban a profil definiálása során említettük.

⁵¹ Ezek a végső profilok lényegében azok a különböző feltételek mellett előállított napi profilok, amelyeket szintén említettünk a profilok definiálása kapcsán (tehát például nyári, téli, átmeneti időszak, a hét különböző napjára vonatkozó profilok, esetleg ezek kombinációi stb.). Egy ezzel kapcsolatos nagyon jó osztályozásról ld. *Pitt* [2000] munkáját.

nem feltétlenül kell, hogy azonos legyen klaszterezés által eredményezett klaszterszámmal⁵², hiszen a gyakorlatban egy piaci szereplő sok klasztert nem biztos, hogy tud transzparensen kezelni, így ilyen esetben valamilyen összevonás szükséges.

Természetesen nem célunk valamennyi fenti lépést részletesen ismertetni. Az értékezés szempontjából kiemelt jelentőséggel leginkább a klaszterezés során használni kívánt görbetulajdonságok előállítása bír. Ez magában foglalja a fogyasztási idősor ún. előigazítását is, ami a legtöbb esetben a hőmérséklet hatásától történő megtisztítást jelent (erről ld. bővebben a fejezetet második felét).

2.1.3. A profilozás során használt görbetulajdonságok előállítása

A klaszterezés során az egy-egy fogyasztói görbét jellemző, ún. görbetulajdonságok sokféleképpen képezhetők. Ezeknek a megoldásoknak egy lehetséges csoportosítását tartalmazza a 4. táblázat, amelyet annak alapján állítottunk össze, hogy ezek előállítása a szakirodalomban eddig ez jellemzően hogyan történt meg.

4. táblázat: Profilozás során előállított görbetulajdonságok lehetséges csoportosítása

| Napi reprezentatív terhelési görbe (RLC) esetén | | | | Egyéb módon előállított tulajdonságok |
|---|---|---|---|---|
| alak paraméter- | időtér- alapon előállított tulajdonságok | frekvenciater- | modell- | |
| <i>Chicco et al.</i> [2005] <i>Mathieu et al.</i> [2011] | <i>Chicco</i> [2012] <i>Li et al.</i> [2010] <i>Macedo et al.</i> [2015] <i>Panapakidis et al.</i> [2012] <i>Panapakidis et al.</i> [2014] <i>Tsekouras et al.</i> [2007] <i>Tsekouras et al.</i> [2008] | <i>Carpaneto et al.</i> [2003] <i>Carpaneto et al.</i> [2006] <i>Chicco et al.</i> [2005] <i>Panapakidis et al.</i> [2014] | <i>Espinoza et al.</i> [2005] <i>Hino et al.</i> [2013] <i>McKenna et al.</i> [2014] | <i>Räsänen et al.</i> [2010] <i>Srivastav et al.</i> [2013] <i>Verdú et al.</i> [2006] |

Forrás: saját gyűjtés és saját szerkesztésű táblázat.

Jól látható, hogy a legtöbb esetben a fogyasztói görbét jellemző tulajdonságok előállítása valamilyen napi reprezentatív terhelési görbe alapján történik meg. Ilyen korábban is említett reprezentatív terhelési görbe (*representative load curve*, röviden RLC) készítését foglalja össze például *Tsekouras et al.* [2008]. Chicco pedig (ld. *Chicco*

⁵² A klaszterek számának módszertanilag helyes megválasztására sok mérőszám van, ld. erről bővebben például *Chicco* [2012].

[2012]) abból indul ki, hogy egy egynapos (96 negyedórából álló) terhelési görbe már eleve rendelkezésre áll, amely terhelési görbe átmeneti időszakból származó „reprezentatív” hétköznapi átlagos negyedórás terhelési értékeit jelenti.⁵³ Ezt a két megoldást lényegében összegző, az RLC alapú (kétlépéses) fogyasztói szegmentációt tárgyaló munka például *Tsekouras et al.* [2007] cikke. Ezek a módszerek alapvetően azt feltételezik, hogy a fogyasztói görbék klaszterezése napi órás vagy negyedórás bontású, reprezentatív terhelési görbék, azaz **időtér** alapon előállított tulajdonságok alapján történik meg.

Lehetséges (és egyszerűbb) megoldás azonban az, hogy a rendelkezésre álló napi RLC-ből még további releváns információkat szűrünk ki, amiket az előző táblázatban **alak paramétereknek** nevezünk. *Chicco et al.* [2005] munkájában ezek az ún. alaktényezők (angolul *shape factors*).⁵⁴ *Mathieu et al.* [2011] munkájában pedig az ún. terhelési alak paraméterek (angolul *load shape parameters*).⁵⁵ Ezeknek a módszereknek a hátránya, hogy a mutatószámok nem minden esetben határozhatóak meg egyértelműen, például olyan terhelés görbék esetén, ahol napon belül két csúcsidőszak van (ami háztartási fogyasztók esetén például nagyon gyakori).

Chicco et al. [2005] egy **frekvenciatér** alapú megközelítést (Fourier-transzformációt) használ az egynapos terhelési görbe reprezentálására a korábbi bekezdésekben említett időtér alapú megoldás helyett, aminek az előnye a harmonikusok által elérhető alacsonyabb adattárolási igény is. Ehhez nagyon hasonló megoldást tartalmaz többek között *Carpaneto et al.* [2003], *Carpaneto et al.* [2006] illetve *Panapakidis et al.* [2014] munkája. A lényeg mindegyik munkában az, hogy a rendelkezésre álló napi

⁵³ Mivel átmeneti időszakról van szó, ezért a hőmérséklet hatásának szűrésével a tanulmány nem is foglalkozik.

⁵⁴ Ezek az alaktényezők az alábbiak:

- napi átlagos terhelés / napi maximális terhelés,
- nappali átlagos terhelés / nappali maximális terhelés,
- napi minimális terhelés / napi átlagos terhelés,
- esti, éjszakai átlagos terhelés / napi átlagos terhelés,
- ebédidőben mért átlagos terhelés / nappali átlagos terhelés,
- nappali minimális terhelés / nappali átlagos terhelés.

⁵⁵ Ezek a terhelési alak paraméterek az alábbiak:

- *Near-Base Load* [NBL, kW]: a napi terhelések 2,5-ik percentilise,
- *Near-Peak Load* [NPL, kW]: a napi terhelések 97,5-ik percentilise,
- *High-Load Duration* [HLD, óra]: annak az időszaknak a hossza, amikor a terhelés értéke közelebb van az NPL-hez, mint az NBL-hez,
- *Rise Time* [óra]: annak az időszaknak a hossza, amíg az NBL-ről a terhelés eléri a HLD kezdetét,
- *Fall Time* [óra]: annak az időszaknak a hossza, amíg a HLD végéről a terhelés újra eléri az NBL-t.

RLC-t frekvenciatérbe képezik le, és a klaszterezést az így származtatott eredményeken végzik el.

Az eddig ismertetett megoldásoknál nagyon sok esetben ez az RLC egy átmeneti időszakból származó napi terhelési görbét jelent, amely vagy több nap terhelési értékeinek átlagaként, esetleg napi terhelési értékek klaszterezésével áll elő (ld. *Chicco* [2012], *Tsekouras et al.* [2007]).

Érdekes külön kategóriába sorolni az ún. **modell alapú** megoldásokat, amikor a napi reprezentatív terhelési görbe valamilyen mögöttes modell feltételezése mellett áll elő. Erre jelentenek példát *Hino et al.* [2013] illetve *McKenna et al.* [2014] munkái, ahol a napi terhelési görbét normális eloszlású sűrűségfüggvények keverékeként modellezik, és a napi görbék csoportosítását az így származtatott eredmények alapján végzik el. *Espinoza et al.* [2005] pedig egy teljesen eltérő, idősoros regressziós technikán alapuló megoldást használva származtat olyan napi terhelési görbét, ahol minden exogén változó (például hőmérséklet, hét napjai stb.) hatása kiszűrésre kerül.

Vannak mindezek mellett olyan megoldások is, amikor a napi RLC-k előállítása nem történik meg (legalábbis nem azzal a céllal, hogy a fogyasztói görbéket klaszterezzük). *Räsänen et al.* [2010] munkájában olyan módszer kerül alkalmazásra, ami az idősorokból véletlenszerűen (de azonosan) kiválasztott időpontokra SOM (angolul *Self-organizing maps*, magyarul önszervező hálók) alapján végez dimenziócsökkentést, majd a hasonló karakterisztikával rendelkező görbéket K-középpontú klaszterezéssel (K-Means) vagy hierarchikus klaszterezéssel csoportosítja. *Srivastav et al.* [2013] keverékmodellt alkalmaz a tipikus fogyasztások becslésére, ennek részletes ismertetésével későbbi fejezetben még foglalkozunk. *Verdú et al.* [2006] pedig heti terhelési idősorokon végez dimenziócsökkentést SOM-módszert is felhasználva.

2.1.4. A profilozás során használt klaszterezési algoritmusok

A 5. táblázat – a teljesség igénye nélkül – a profilcsoportok készítésére leggyakrabban alkalmazott klaszterezési algoritmusokat, illetve azok néhány felhasználását mutatja.

A módszerek alkalmazása energetikán túlmutatóan is rendkívül széleskörű, és így az általánosan ismert előnyök és hátrányok természetesen itt is érvényesek. A módszerek közüli választás során a kérdésfeltevés inkább úgy helyes, hogy ezek az előnyök vagy hátrányok a profilozás területén az egyes konkrét esetekben mennyire kritikusak.

5. táblázat: Profilozás során használt klaszterezési algoritmusok

| Klaszterezési algoritmus | Alkalmazás |
|---|---|
| K-középpontú klaszterezés (K-Means) | Chicco <i>et al.</i> [2005] Chicco [2012] Espinoza <i>et al.</i> [2005] McKenna <i>et al.</i> [2014] Panapakidis <i>et al.</i> [2014] Räsänen <i>et al.</i> [2010] Tsekouras <i>et al.</i> [2007] Tsekouras <i>et al.</i> [2008] |
| K-középpontú klaszterezés (K-Means++) | Panapakidis <i>et al.</i> [2014] |
| Fuzzy K-középpontú klaszterezés (fuzzy K-Means) | Chicco <i>et al.</i> [2005] Chicco [2012] Tsekouras <i>et al.</i> [2007] Tsekouras <i>et al.</i> [2008] |
| Hierarchikus klaszterezés | Chicco <i>et al.</i> [2005] Chicco [2012] Hino <i>et al.</i> [2013] Panapakidis <i>et al.</i> [2014] Räsänen <i>et al.</i> [2010] Tsekouras <i>et al.</i> [2007] Tsekouras <i>et al.</i> [2008] |
| K-Medoid klaszterezés | Panapakidis <i>et al.</i> [2012] |
| <i>Follow-the-leader</i> (FDL) és változatai | Carpaneto <i>et al.</i> [2003] Carpaneto <i>et al.</i> [2006] Chicco <i>et al.</i> [2005] Chicco [2012] |
| Mesterséges neurális hálók (ANN ⁵⁶) | Macedo <i>et al.</i> [2015] |
| Önszervező hálók (SOM ⁵⁷) | Chicco <i>et al.</i> [2005] Panapakidis <i>et al.</i> [2014] Verdú <i>et al.</i> [2006] Tsekouras <i>et al.</i> [2008] |
| Kanonikus korreláció elemzés (CVA ⁵⁸) és diszkriminancia-analízis | Li <i>et al.</i> [2010] |

Forrás: saját gyűjtés és saját szerkesztésű táblázat.

Az iteratív K-középpontú (K-Means) klaszterezés és azok különböző variánsainak (K-Means++, fuzzy K-Means) egyszerűsége ellenére hátránya, hogy megköveteli a klaszterek számának *a priori* ismeretét.

A különböző hierarchikus klaszterezési technikák előnyei, hogy rengeteg távolság illetve hasonlósági mérték használható, azonban például agglomeratív esetben⁵⁹ a különböző összevonási eljárások nagyon különböző eredményeket hozhatnak, illetve – alap-

⁵⁶ Angolul *Artificial Neural Network*.

⁵⁷ Angolul *Self-organizing Maps*.

⁵⁸ Angolul *Canonical Variate Analysis*.

⁵⁹ Agglomeratív klaszterezés esetén kiindulásként minden terhelési görbe külön-külön egyelemű klasztereket alkot, majd lépésenként történik meg a klaszterek összevonása.

esetben – nem történhet meg egy-egy terhelési görbe másik klaszterbe történő átsorolása sem.

A K-Medoid klaszterezés során is lehetséges rengeteg távolságmérték használata, a K-középpontú klaszterezéshez képest azonban ez robosztusabb megoldás. A végeredményként kapott klaszterek ún. medoidja az a megfigyelés lesz, amihez a többi klasztertag átlagosan a legközelebb esik, így egyben ez tekinthető a klasztert jellemző reprezentatív megfigyelésnek is (szemben az átlagolásra szintén érzékeny centroiddal).

Jól látható, hogy a különböző mesterséges neurális hálózatos (ANN) megoldások a klaszterezési feladatok területén is terjedőben vannak, és a dimenziócsökkentés mellett csoportképzésre is használatosak. Sokszor nem is önállóan szerepelnek, hiszen például az ún. önszervező hálók (SOM) a vizsgált inputváltozókat – a sokdimenziós skálázáshoz hasonlóan – alacsonyabb, általában kétdimenziós térbe képezik le, és magát a klaszterezést valamelyik korábban is ismertetett klaszterezési technika felhasználásával együtt végzik el.

Hasonlóan önmagában nem alkalmas klaszterezésre a kanonikus korrelációelemzés (CVA). Itt a megfelelő csoportok közötti (*between*) és csoportokon belüli (*within*) kovariancia mátrixok felhasználásával származtatott ún. kanonikus változók alapján történik a klaszterbesorolás (*Li et al.* [2010] munkájában például lineáris diszkriminancia-analízis felhasználásával).

Számtalan további megoldás is létezik még, ezekről többek között a táblázatban is szereplő szakirodalmakban lehet bővebben olvasni.

2.2. Az időjárást leíró változók kezelése energetikai idősorok esetében

A fogyasztói görbék profilozásával kapcsolatban alapvető kérdés a különböző exogén változók, így különösen az időjárást leíró változók hatásának a kezelése, illetve annak eldöntése, hogy ezeknek a változók hatását vizsgálva mit tekintünk a **profil részének**, illetve mit nem tekintünk annak. Mivel az időjárást leíró változók alapvetően ugyanúgy sztochasztikusak, mint maga a fogyasztás, ezért a tipikus (vagy várható) körülményekhez képesti eltérés, illetve ennek a fogyasztásra gyakorolt hatása a profilozás (mint tipikus fogyasztási mintázatok keresése) szempontjából is kiemelt jelentőséggel bír.⁶⁰

⁶⁰ Természetesen az időjárást leíró változók mellett szükséges még többek között a szezonalitást, a kiugró értékeket stb. is valamilyen módon kezelni, módszertanfüggő, hogy ez mikor milyen formában történik meg.

2.2.1. Az időjárási változók és a fogyasztás kapcsolata

Az időjárási változók hatásának kezelése rendkívül komplex feladat. Ez nemcsak a kapcsolat **meglétének** feltárását érinti, hanem azt is, hogy egy-egy változó hatását **hogyan** érdemes, hogyan lehet a leghatékonyabban figyelembe venni. Ebbe beleértendő a mintavételi (megfigyelési) gyakoriság megválasztása, a különböző késleltetési struktúrák használata, vagy az időjárási változókon végzett egyéb transzformációk (például napi átlagos-, vagy napi csúcérték figyelembe vétele, stb.).

Az időjárási változók tekintetében a legtöbb esetben az alábbiakat használjuk: hőmérséklet, páratartalom, szél, felhőtakaró aránya, csapadékmennyiség (ld. Pitt [2000]). Azon túl, hogy a kapcsolat alapvetően nemlineáris, a változók között léteznek különböző interakciós hatások is.

A **hőmérsékletet** jellemző linearitás hiányának a kezelése legtöbb esetben a fűtési illetve a hűtési **napfok** használatával történik (*heating* illetve *cooling degree-day*, HDD illetve CDD, ld. Sugár [2011] vagy Mák [2015]). Espinoza et al. [2005] is ugyanúgy külön-külön definiálja a fűtési illetve a hűtési hatás következtében szükséges energiaigényt (HR illetve CR, azaz *heating* illetve *cooling requirement*).

Interakciós változók szerepeltetése akkor indokolt, amikor a változók egymás hatását erősítik vagy gyengítik. Ilyen például a magas **páratartalom** hatása, ami elsősorban nyáron releváns, amikor a fülledt meleg miatt megnövekszik a légkondicionáló készülékek használata (sok esetben szokás ún. *discomfort*-hatásként is hivatkozni rá). Hasonlóan a **szél** szerepe inkább a téli időszakban szignifikáns, különösen akkor, amikor nagyon hideg van, és a hideg és szeles idő együttes jelenléte nagymértékben növeli a hideg érzetünket (azaz a téli fűtési hatás még erősebb lesz, ez az ún. *wind-chill*-hatás). Az ilyen vagy hasonló hatásoknak a modellben történő szerepeltetése többféleképpen is megtörténhet.

Az egyik megoldás az, hogy a napfokok számításához hasonlóan, empirikus, tapasztalati úton „kikísérletezett” **formulákat** használunk. A különböző meteorológiai intézeteknek sokféle ún. *wind-chill*-formuláik vannak.⁶¹ Hasonlóan létezik az ún. *humidity index* (*humidity index*) is a korábban már említett *discomfort*-hatás modellezésére.⁶²

⁶¹ Ezek közül néhány az alábbiak szerint foglalható össze:

- Kanada esetében:

$$T_{wc} = 13.12 + 0.6215T_a - 11.37V^{+0.16} + 0.3965T_aV^{+0.16},$$

ahol:

- T_{wc} az ún. *wind-chill* index, °C skálán mérve,
- T_a a levegő hőmérséklete,

Természetesen vannak olyan esetek, amikor ezek a mutatók nem a legmegfelelőbbek (mint ahogyan a fűtési illetve a hűtési napfok esetében sem mindig optimális a 12 vagy 21°C körüli küszöbértékek, és az ettől számolt lefelé illetve felfelé vett eltérések használata, amit Sugár [2011] javasol). Mindezek kiküszöbölésére használhatók **exploratív** megoldások is, ilyen pl. az ún. MARS (*Multiple Adaptive Regression Splines*), amely mind a nemlinearitás, mind az interakciók kezelésére alkalmas anélkül, hogy tapasztalati úton kikísérletezett formulára kellene hagyatkoznunk, és mind a küszöbértékek, mind a lehetséges interakciós kapcsolatok feltárása alapvetően exploratív módon, egzakt mérőszámok kiértékelése (pl. modellszelekciós kritériumok) útján történik.

A kutatások során a nehézséget azonban sokszor az okozza, hogy megfelelő minőségű adatok nem álltak rendelkezésre a vizsgálathoz (például az OMSZ⁶³ adatok még kutatási célokra sem érhetően el ingyenesen, az interneten is rendelkezésre álló – egyébként sokszor ingyenes – adatforrások pedig nem olyan megbízhatóak).

A témához kicsit közvetettebb módon kapcsolódik ugyan, de az időjárást leíró változók hatásában is lehet némi késleltetés (például a t -edik időszak fogasztást nemcsak a t -edik időszak, ún. *spot* hőmérséklet befolyásolja). Ennek a kezelésére különböző (jellemzően egyszerűbb) technikák születtek, amelyeket röviden a hőmérséklet esetében legtöbbször előforduló megoldásokon keresztül összegzünk:

- néhány időszakos késleltetések bevezetése, azaz a t -edik időszak hőmérséklet (T_t) mellett korábbi időszakok hőmérsékleteinek (T_{t-1} , T_{t-2} ...) a szerepeltetése,

-
- V a szélsébség 10 méteren (km/h),
 - az Amerikai Egyesült Államok esetében:

$$T_{wc} = 35.74 + 0.6215T_a - 35.75V^{+0.16} + 0.4275T_aV^{+0.16},$$

ahol:

- T_{wc} az ún. *wind-chill* index, °F skálán mérve,
- T_a a levegő hőmérséklete,
- V a szélsébség 10 méteren (mph),

A fenti számítások csak 10°C (50°F) alatt és 4.8 km/h felett történnek meg, ez tehát az a két küszöbérték, amelyek meghaladása esetén a változók együttes, interakciós hatásáról beszélhetünk.

⁶² Ennek a számítási módja az alábbi:

$$Humindex = T_{air} + 0.5555 \left[6.11e^{5417.7530 \left(\frac{1}{273.16} - \frac{1}{T_{dew}} \right)} - 10 \right],$$

ahol:

- T_{air} a levegő hőmérséklete °C-ban,
- T_{dew} a harmatpont K-ban.

Mivel a harmatpont és páratartalom között lehetséges némi közelítő konverzió, ezért a *humindex* valóban a hőmérséklet és a páratartalom interakciós hatásának számszerűsítésére alkalmazható.

⁶³ Országos Meteorológiai Szolgálat.

- exponenciális súlyozás bevezetése, azaz T_t helyett a

$$T_t(\alpha) = \alpha \cdot T_t + (1 - \alpha) \cdot T_{t-1}(\alpha)$$

szerepeltetése ($0 < \alpha < 1$),

- az *Elexon* [2013] által definiált *Noon Effective Temperature* (NET) megoldás, ami az adott napon a fogyasztás szempontjából releváns hőmérsékletet az aznapi, a megelőző napi illetve az azt megelőző napi hőmérsékletek átlagolásából állítja elő rögzített súlyok felhasználásával az alábbi módon:

$$T_t^{NET} = 0.57 \cdot T_t + 0.28 \cdot T_{t-1} + 0.15 \cdot T_{t-2}, \text{ illetve}$$

- ilyen megoldásnak tekinthető a magyarországi földgázpiacon a profilos fogyasztók profiljainak számítása során használt ún. felejtősúlyos hőmérséklet is, amely az adott t-edik napot megelőző napok hőmérsékleteit egyre kisebb súllyal veszi figyelembe.

2.2.2. Az időjárási változók kezelése a profilozás során

A sztochasztikus jellegből adódóan az időjárási változók kezelése tekintetében alapvetően kétféle nézőpont lehetséges:

- magának az időjárási változónak a hatását, vagy
- az időjárási változó, mint sztochasztikus változó esetében természetes módon megjelenő véletlennek, a tipikus (átlagos) tendenciától való eltérésnek a hatását szeretnénk közvetlenül modellezni.

Ennek a két lehetőségnek a profilozás területére történő transzformációja az alábbiakban összegezhető:

- az időjárási változónak a teljes hatásától megtisztítjuk a profilt, azaz az időjárás hatása egyáltalán nem része a profilnak, vagy
- az időjárási változónak csak a véletlen hatásától (azaz az időjárási véletlen hatástól) tisztítjuk meg a profilt, azaz csak a tipikus (átlagos) időjárás hatása része a profilnak.

A szakirodalomban jellemzően az első megoldás fordul elő, és néhány kivételtől eltekintve az időjárást leíró változók közül csak a hőmérséklet hatását veszik figyelembe.

Räsänen et al. [2010] a hőmérséklet hatásától történő megtisztítást egyszerű regressziós technikával végzi el külön-külön a négy évszakra, azért hogy a **linearitás hiá-**

nyát (például téli illetve nyári hőmérsékletváltozást) figyelembe vegye. Ez az évszakhatárok definiálása végett nehézkes és nem is túl elegáns.

Espinoza et al. [2005] külön-külön definiálja a fűtési illetve a hűtési **napfokot** (HR illetve CR, azaz *heating* illetve *cooling requirement*, ezekről lásd még a későbbi fejezeteket), ami elegánsabb megoldás, azonban a napfokokhoz használt küszöbértékek helyes megválasztása okozhat nehézségeket.

Pitt [2000] munkájában több időjárási tényező hatását is figyelembe veszi, és a profilozást úgy végzi el, hogy a fogyasztási idősort egy ún. időjárásfüggő (*weather-dependent*) és időjárásfüggetlen (*weather-independent*) részre osztja. Az időjárási tényezők figyelembe vétele a MARS (*Multiple Adaptive Regression Splines*) módszerrel történik, amely lehetőséget ad a nemlineáris és **interakciós** hatások idősorhoz legjobban illeszkedő kombinációinak a feltárására és becslésére. Érdeemes megjegyezni, hogy az ehhez hasonló, rengeteg időjárási változót figyelembe vevő megoldás nagyon ritka.

Mutanen et al. [2011] valamivel eltérőbb megoldást alkalmaz a korábban említettekhez képest, ugyanis a korrekciót azonos napokra és hónapokra végzi el, és mind a hőmérséklet, mind a fogyasztás esetében a megfelelő **átlagtól vett eltéréssel** dolgozik. Ennek a megoldásnak a hátránya azonban, hogy a hőmérsékletfüggőség napon belül is lehet nagyon különböző, így – ha olyan nagymértékben nem is – de ez befolyásolja az eredményeket, azaz többszintű szezonális esetén nem működik olyan jól. Egy másik, kifejezetten időjárással foglalkozó alfejezetben bemutatjuk, hogy egy hasonló megoldás (*Mák* [2015]) havi adatok esetén jól tud működni. Ott kitérünk a profilozás kapcsán levezethető tanulságokra is.

Az a jelenség, hogy hagyományosan az időjárás teljes hatása kiszűrésre kerül profilozás során, többek között azzal is magyarázható, hogy a profilozás többé-kevésbé már előzetesen csoportosított fogyasztókon történik meg, akiknek a hőmérsékletfüggése hasonló (pl. mindenkinek csak téli, csak nyári hőmérsékletfüggés van, vagy mindkettő, vagy egyik sem). Ez nem is jelent problémát abban az esetben, amikor „csak” a homogén csoportok létrehozása a cél, és az egyedi fogyasztói viselkedés (sem a várható tendenciát, sem a bizonytalanságot tekintve) nem érdekel bennünket. Az időjárásfüggés ugyanis lehet erősebb-gyengébb egy-egy fogyasztó esetében (ami bármilyen egyedi kezelésnél, pl. egyedi árazásnál számíthat), sőt, a bizonytalanság is lehet időjárásfüggő.

2.2.3. A hőmérsékleti véletlen hatás szerepe

A 2.2.3. és 2.3.4. szakaszok egy korábbi saját tanulmányunk a dolgozat szempontjából releváns, kis mértékben módosított részeit tartalmazzák.

Az időjárás alakulásában előforduló többé-kevésbé szélsőséges körülmények nem eseti jelleggel előforduló kilengések, hanem az átlagostól kisebb vagy nagyobb mértékű, állandóan jelen levő eltérések, ezért azok klasszikus *outlier*-ként történő modellezése nehézkes, és nem is igazán megfelelő megoldás.

Az ilyen természetű véletlen hatások kezelése és számszerűsítése rövid és hosszú távon egyaránt fontos (az itt bemutatásra kerülő fejezetben a hosszú távú kérdéseken van inkább a hangsúly). Mivel a bizonytalanság (nem olyanról van szó, amit például egy strukturális törés jelent) folyamatosan fennáll, ezért a kapcsolódó kockázatok kezelése és számszerűsítése is fontos és állandó kontrollt igényel, legyen szó akár szabályozásról, akár hosszú távú portfóliókezelésről beszerzés (a hosszú távon szükséges szerződéses mennyiségek meghatározása) és fogyasztás oldalon (a fogyasztás várható alakulásának előrejelzése) egyaránt.

A szakaszban a szezonális kiigazítási módszertan alkalmazása mellett mutatjuk be, hogy az időjárás (hőmérséklet) okozta sztochasztikus véletlen milyen mértékben képes befolyásolni a modellezési eredményeket (például magát a szezonális kiigazítást, az *outlier*-ek felismerését) és azt, hogy annak explicit kezelése mellett milyen eredménybeli javulásokat érhetünk el.

A véletlen jelleg természetesen (a szezonális kiigazítás módszertanában ismert regARIMA-hoz hasonló) regressziós modell becslésével is kezelhető, hiszen az idősor komponensekre bontása ettől gyakorlatilag független, ugyanis többek között az idősor előigazítására szolgál. A szezonális kiigazítás módszertana azonban – mintegy plusz hozadékként – megadta a lehetőséget egyrészt a „hagyományos” *outlier*-ek⁶⁴ kezelésére, valamint az eredmények validálására (ld. csúszó tartományok (*slidings span*) diagnosztika). Mivel a szezonális kiigazítás módszertana nem kapcsolódik olyan szorosan a dolgozat témájához, ezért itt csak a hőmérséklet hatásának kezelésére koncentrálnunk, és a korábbi tanulmányból (Mák [2015]) csak a fontosabb részeket mutatjuk be.

A korábban már publikált eredmények között tehát többek között egy olyan, jól interpretálható modell megalkotása is szerepel, amely figyelembe veszi a hőmérséklet-fogyasztás nemlineáris kapcsolatát két oldalról is. Egyik oldalról abban a tekintetben,

⁶⁴ Ezek az alábbiak: additív outlier (*additive outlier*, AO), szinteltolás (*level shift*, LS), csillapodó jellegű törés (*transitory change*, TC).

hogy a hőmérséklet hatása csak egy bizonyos küszöbérték alatt releváns, így a fűtési hatás használata indokolt a modellekben. Ez a megoldás ekvivalens az ezen a téren jellemzően alkalmazott megoldásokkal. Másik oldalról abban a tekintetben, hogy a modell havonként az átlagtól vett (fűtési) napfokeltérésekkel számol, így lehetőség van annak figyelembe vételére, hogy a hőmérséklet hatása az említett küszöbérték alatt sem feltétlenül egyforma.

Az itt bemutatott eredmények nagyon sok olyan fontos következménnyel szolgálnak, amelyek a profilozás szempontjából tanulságosak. Ezeket az alfejezet végén összegezzük, a tanulmányt néhány további modell számítással kiegészítve.

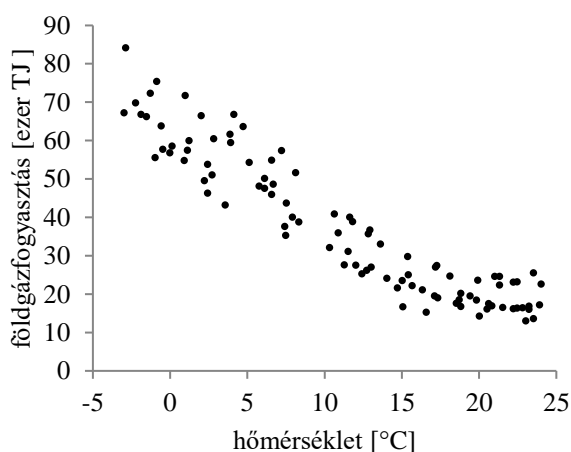
2.2.4. A hazai földgázfogyasztás extrém (véletlen) hőmérsékleti hatásoktól történő megtisztítása és szezonális kiigazítása

A fejezetben tehát egy korábbi tanulmány (Mák [2015]) egy részletét mutatjuk be. A tanulmány ugyan a hazai földgázfogyasztás alakulásával kapcsolatos, a dolgozat szempontjából azonban mindenképpen érdemes tárgyalni, hogy az átlagostól eltérő hőmérséklet hatásának modellbe építése hogyan jelenik meg.

2.2.4.1. Felhasznált adatok

A tanulmány empirikus részében felhasznált földgázfogyasztás-adatok az Eurostat adatbázisából származnak,⁶⁵ a hőmérsékletadatokat pedig az OMSZ Pestszentlőrinci mérőállomásán mérték.⁶⁶

21. ábra. A magyarországi földgázfogyasztás a hőmérséklet függvényében, 2006–2013



Forrás: Eurostat adatai alapján saját készítésű ábra (Excel).

⁶⁵ <http://ec.europa.eu/eurostat/>, ezer terajoule-ban (TJ)

⁶⁶ http://www.varaljamet.eoldal.hu/cikkek/climate_budapest.html, Celsius-fokban (°C)

A 21. ábra alapján látható a hőmérséklet és a földgázfogyasztás közötti kapcsolat jellege 2006 és 2013 között, havi átlaghőmérséklet és fogyasztásadatok tekintetében. Jól látható, hogy 15-16 °C feletti átlaghőmérséklet esetén a havi fogyasztás viszonylag stabil szint körül ingadozik, és ahogy csökken a havi átlaghőmérséklet, úgy emelkedik a fogyasztás értéke. Az említett küszöbérték alatt a kapcsolat jó közelítéssel még akár lineárisnak is mondható, azonban ha eltekintünk a legmagasabb fogyasztási értéktől (ami egyébként 2006 januárja, a vizsgált időszak egyik leghidegebb téli hónapja⁶⁷), akkor a lineáris kapcsolat már korántsem olyan egyértelmű.

Az 15-16 °C-os küszöbérték figyelembe vétele alapvetően kétféleképpen történhet. Az egyik megoldás a HDD (*heating degree-day*) számítására, amikor a hagyományosan 16 °C küszöbértéktől lefelé vett eltéréseket számítjuk (és összegezzük)⁶⁸, a másik lehetőség, amikor 15 °C alatti hőmérsékletértékeknek (de csak ezeknek) a 18 °C küszöbértéktől vett eltéréseit számítjuk (és összegezzük). Képlettel a két megoldás a következő:

1. módszer:

$$HDD = \max(0, 16\text{ °C} - \text{hőmérséklet}),$$

2. módszer:

$$HDD = 18\text{ °C} - \text{hőmérséklet}, \text{ ha a hőmérséklet} \leq 15\text{ °C},$$

$$HDD = 0\text{ °C}, \text{ egyébként}.$$

A második módszer esetén a számítás ennél jóval szofisztikáltabban is körüljárható (lásd például *Howden–Crimp* [2001]); az empirikus eredmények mindenesetre azt igazolják, hogy az utóbbi megoldás általában jobb illeszkedést biztosít, illetve az Eurostat is e képlet szerint publikál.

Esetünkben a 2. módszer szerint számoltunk.⁶⁹ A 22. ábra a földgázfogyasztás és a hőmérséklet alapján számított HDD időbeli alakulását mutatja a 2006 és 2013 közötti időszakra. Jól látható a fogyasztásban meglevő csökkenő tendencia – ami nemcsak a 2008-as válság hatásával magyarázható –, valamint, hogy a hidegebb teleken magasabb

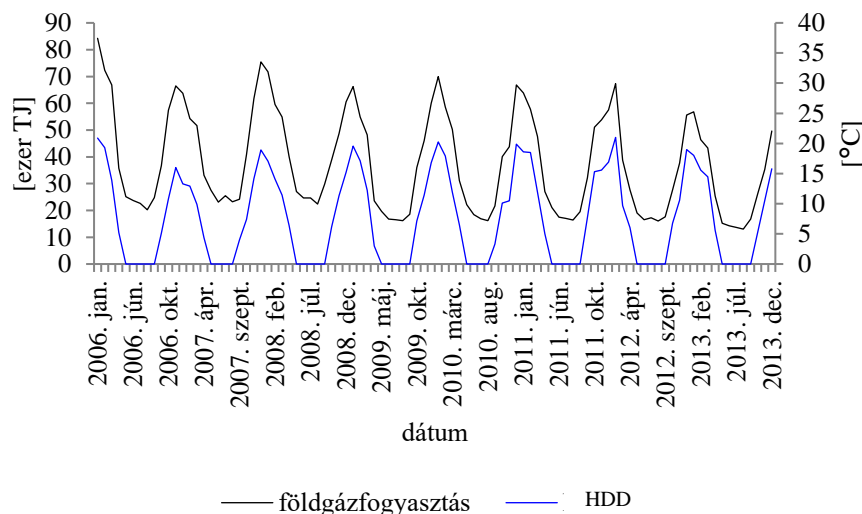
⁶⁷ Ennél hidegebb hónap csak 2012 februárja volt, azonban ekkor a csökkenő trend miatt jóval alacsonyabb volt a fogyasztás, a rendkívüli hideg ellenére is.

⁶⁸ Hasonlóan létezik hűtési napfok is (ún. *cooling degree-day*), amely a légkondicionálók nyári meleg miatti magasabb fogyasztásának leírásában tud segíteni. Itt egy adott küszöbértéktől felfelé vett eltéréseket szokás számítani, a küszöbérték tekintetében azonban kevésbé vannak „szabványok” úgy, mint a fűtési hatás tekintetében.

⁶⁹ Amennyiben rendelkezésre állnak napi adatok, az ezekből számított értékek aggregálásával kaphatjuk meg a havi és éves szinten az éves értékeket stb. Mivel 2006-ig visszamenőleg hiteles hőmérsékletadatok napi bontásban nem álltak rendelkezésünkre, és az Eurostat is csak 2009-ig közöl ilyen tartalmú adatokat, ezért a tanulmányban – napi adatok hiányában – a havi átlaghőmérsékletből számolt eredményekkel dolgoztunk.

a fogyasztás, míg melegebb teleken alacsonyabb, illetve felismerhető a korábban említett kiugró 2012. februári hideg hatása is.

22. ábra. A magyarországi földgázfogyasztás és a HDD alakulása, 2006-2013



Forrás: Eurostat adatai alapján saját számítás és saját készítésű ábra (Excel).

Összefoglalva az empirikusan tett megállapításainkat, önmagában a HDD-módszer tehát csak arra alkalmas, hogy a fűtési hatást modellezze, de a küszöbérték alatti megfigyelések esetében a kapcsolatot alapvetően lineárisan írja le. Azonban az ábrák alapján sejthető, hogy a küszöbérték alatt sem feltétlenül lineáris a kapcsolat.

A fejezetben többek között az utóbbi empirikus tény modellezésére és számszerűsítésére mutatunk be egy lehetséges megoldást. A bemutatott módszertant felhasználva azt is megvizsgáljuk, hogy a modell hogyan alkalmazható az extrém hőmérsékleti hatások kezelésére. Kézenfekvőnek tűnik a számított HDD-k átlagos értékektől való eltéréseit definiálni, és változóként az átlagos HDD-től vett eltéréseket használni. Így az eltérés lehet pozitív vagy negatív, attól függően, hogy az átlagos tendenciához képest viszonylag hideg vagy meleg hónap volt. Mivel a december, január (esetleg február) hónapok jellemzően hidegebbek a többinél, ezért célszerű az átlagtól vett eltéréseket havonta számolni. Azaz csak a januári hónapokat figyelembe véve kiszámoljuk a januári átlagos⁷⁰ HDD-t, és megnézzük minden egyes januári hónap HDD-értékének ezen átlagtól vett eltérését, majd ugyanígy járunk el a többi 11 hónap esetében. Így minden hónapban igaz az, hogy az átlagos HDD-től, mint referenciaszinttől számított eltérések átlaga pontosan zero (hiszen az átlagtól vett eltérések átlaga zero).

⁷⁰ Átlag alatt a mintabeli időszak átlagát értjük.

2.2.4.2. Becslési eredmények

A fejezetben bemutatott eredmények az X-13ARIMA-módszerrel, az X-13ARIMA-SEATS, illetve az Excel program felhasználásával készültek. A részletes lépéseket nem mutatjuk be, a fontosabb eredményeket, a modellezés közben meghozandó döntéseket és nehézségeket ismertetjük. Megvizsgáljuk azt is, hogy a HDD-eltérések, mint exogén változók (a szezonális kiigazítást végző programok elnevezésével élve az ún. felhasználó által definiált változók (*user-defined variables*)) a szezonális kiigazítás minőségét hogyan befolyásolják. Erre azért érdemes külön hangsúlyt fektetnünk, mert a földgázfogyasztás esetében a szezonalitást alapvetően a hőmérséklet befolyásolja, ahol a véletlen szerepe rendkívül nagy, és így a szezonális kiigazítás stabilitására is jelentős mértékben hatással lehet.⁷¹

Látszólag a szezonális kiigazítás a dolgozat témájától némileg távolabb eső terület. Itteni szerepeltetését azonban elsősorban az indokolja, hogy a profilozás során is lehet cél egy átlagos tendenciának a számítása, amikor mindenféle véletlen hatástól megtisztított idősort szeretnénk kapni. Az itt bemutatott megoldáshoz képest azonban profilozásnál ez a feladat valamivel összetettebb. A szezonális kiigazítás módszertanáról magyar nyelven ld. Sugár [1999a] és Sugár [1999b] munkáit.

6. táblázat: A regressziós modell főbb eredményei a hazai földgázfogyasztás példáján

| Változó | Koefficiens | Standard hiba | t-érték |
|--------------------|-------------|---------------|---------|
| konstans | -0,0010 | 0,0016 | -0,68 |
| LS_2008_okt | -0,2203 | 0,0494 | -4,46 |
| HDD_eltérés_jan | 0,0363 | 0,0075 | 4,85 |
| HDD_eltérés_feb | 0,0478 | 0,0066 | 7,30 |
| HDD_eltérés_márc | 0,0556 | 0,0101 | 5,53 |
| HDD_eltérés_ápr | 0,0749 | 0,0145 | 5,17 |
| HDD_eltérés_okt | 0,0439 | 0,0111 | 3,96 |
| HDD_eltérés_nov | 0,0427 | 0,0091 | 4,70 |
| HDD_eltérés_dec | 0,0320 | 0,0093 | 3,46 |
| csúcs_eltérés_nyár | 0,0001 | 0,0001 | 1,78 |

Forrás: saját számítási eredmények (X-13ARIMA-SEATS) és saját készítésű táblázat.

Ebben a fejezetben elsősorban a végeredményekre koncentrálva bemutatjuk a HDD-eltérések nélkül illetve a HDD-eltérésekkel számolt eredményeket is. Tanulmányként érdemes megjegyezni, hogy az extrém hőmérsékleti hatások explicit szerepeltetése egy jobban alátámasztható, jobban indokolható modellt eredményez például a különbö-

⁷¹ Ennek vizsgálatára felhasználjuk majd az X-13ARIMA-módszer egyik közismert diagnosztikáját (a csúszó tartományokat (*sliding spans*)). Az ezzel kapcsolatos eredményeket itt nem közöljük, azok a hivatkozott munkában megtalálhatóak (Mák [2015]).

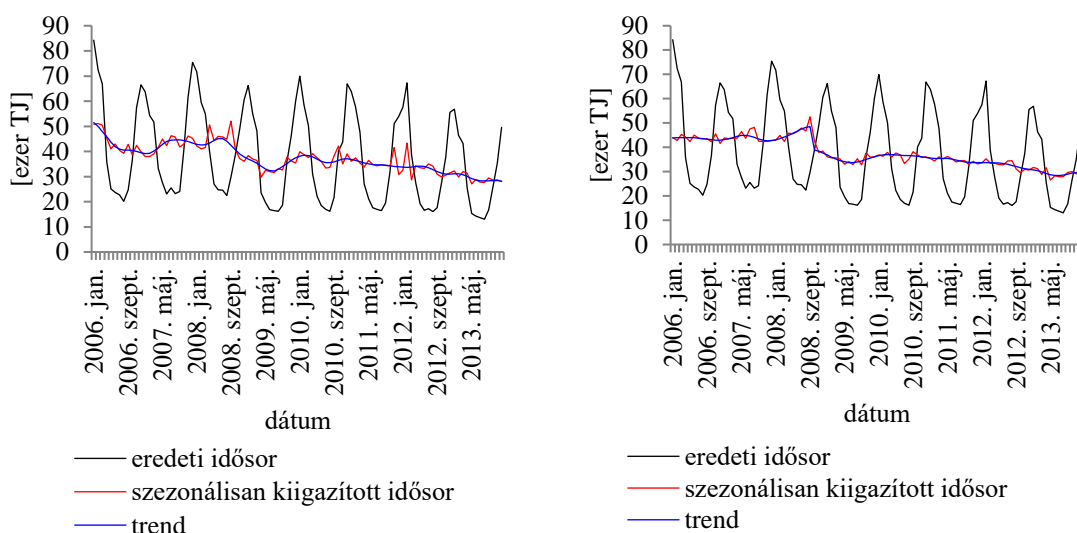
ző strukturális törések automatikus szelekciója kapcsán is. A fontosabb teszteket, diagnosztika eredményeket a hivatkozott munkában találja meg az Olvasó.

A HDD-eltérésekkel kapcsolatos eredményekre rátérve (lásd a 6. táblázatot), a június, július és augusztus hónapok HDD-eltéréseit nem is teszteltük, hiszen ezekben a nyári hónapokban a HDD és a HDD-eltérés értéke a vizsgált időszakban végig zero volt. A május és szeptember hónapok 1-1 évben mutattak ugyan pozitív HDD értékeket, azonban a HDD-eltérések itt nem bizonyultak szignifikánsnak.

Mivel a modell az eredeti idősor logaritmizálását követően készült el, ezért az idősor a különböző magyarázóváltozók exponenciális függvényeként írható fel.

Ebben az esetben – mint ismert – a paramétereket (jelöljük β -val) nem közvetlenül, hanem az exponenciális függvényformából adódóan az $\exp(\beta)$ értékekkel értelmezzük. Így például, ha a január hónap paramétere 0,0363, akkor $\exp(0,0363) = 1,0370$, tehát amennyiben a januári hónap 1 °C-kal hidegebb az átlagnál (azaz a HDD-eltérés értéke 1 °C-kal magasabb), akkor a földgázfogyasztás értéke *ceteris paribus* átlagosan 3,70 százalékkal lesz magasabb. Hasonlóan kalkulálhatók és értelmezhetők a többi hónap paraméterei is.

23. ábra. A szezonális kiigazítás végeredménye HDD-eltérések használata nélkül és HDD-eltérések használatával



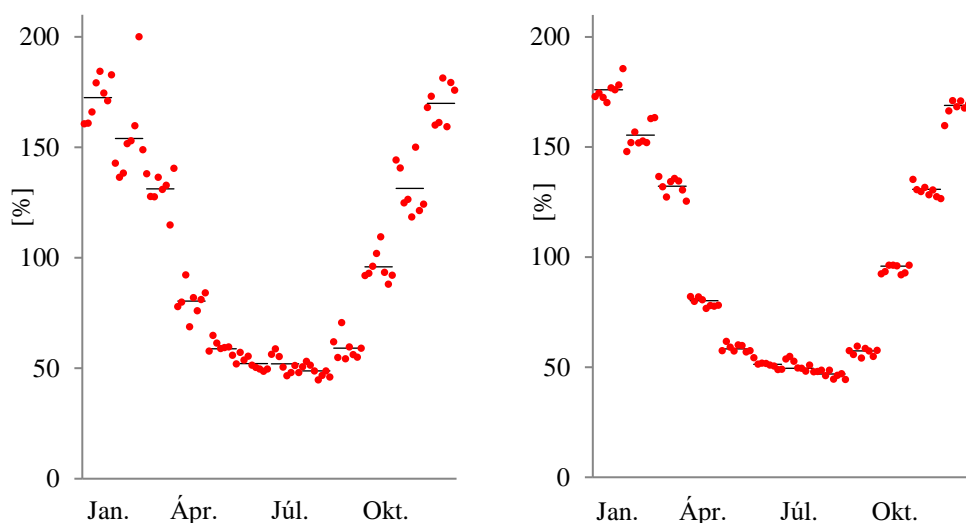
Forrás: saját számítási eredmények (X-13ARIMA-SEATS) és saját készítésű ábra (Excel).

Végezetül érdemes néhány ábrát megvizsgálni annak bemutatására, hogy a hőmérséklet beépítése a modellbe a szezonális kiigazítás szempontjából milyen hozadékkal jár.

A HDD-eltérések szerepeltetése nélkül a szezonális komponens hatásától megtisztított, tehát a szezonálisan kiigazított idősor meglehetősen cikkcakkos, hiszen a hőmérséklet okozta véletlen hatás még benne van (ld. a 23. ábrát). Ez különösen látványos a 2011/2012 telén, amikor februárban nagyon hideg volt. Ugyanígy cikkcakkos a válság környéki időszak, hiszen mint azt az eredeti tanulmányban említettük, ez a modell nem ismerte fel az válság során bekövetkező strukturális törést sem.

Hasonló megállapítás érhető tetten az SI-arányok⁷² vizsgálata esetén is, amikor hónapon belül ezek az értékek nagy változékonyságot mutatnak (ld. a 24. ábrát).

24. ábra. Az SI-arányok HDD-eltérések használata nélkül és HDD-eltérések használatával



Forrás: saját számítási eredmények (X-13ARIMA-SEATS) és saját készítésű ábra (Excel).

A HDD-eltérések modellbe építését követően egy jóval kevésbé csipkézett szezonálisan kiigazított idősort kapunk, hiszen a HDD-eltéréseket a szezonális komponens részeként definiáltuk. Jól felismerhető mind a trendben, mind a szezonálisan kiigazított

⁷² Az SI-arány a különböző (felhasználó által definiált vagy beépített) regressziós hatások nélküli szezonális és irreguláris komponens együttese (ld. alább) a szezonális kiigazítás keretrendszerében. Itt csak röviden megjegyezzük, hogy a szezonális kiigazítás kapcsán kiindulópontot jelentő komponensek illetve ezek kapcsolódásának a definiálása mit jelent. Additív modell esetén az idősor a következő módon írható fel:

$$Y = T + C + S + TD + H + O + I,$$

ahol:

- T – a hosszú távú trendet,
- C – a középtávú ciklus hatását,
- S – az éven belüli szabályos ingadozást leíró szezonaritást,
- TD – a munkanapok eltérő számának hatását,
- H – az ünnepnapok hatását,
- O – az outlier megfigyelések hatását,
- I – pedig a klasszikus véletlent (irreguláris komponens) jelenti.

Létezik még többek között multiplikatív és logadditív felírás is, amelyek ugyanezen komponensek szorzatszerű, illetve azok logaritmusának összezszerű felírását jelentik.

idősorban a szignifikánsnak bizonyuló válság alatti törés hatása (ld. a 23. ábrát). Folytatva az előző gondolatmenetet, ebben az esetben az SI-arányok is jóval kisebb ingadozást mutatnak egy-egy hónapon belül (ld. a 24. ábrát).

2.2.4.3. Az extrém (véletlen) hőmérsékleti hatás kiszűrésére

A 7. táblázatban a HDD-eltérés paraméterek becsült értékei láthatók, illetve a belőlük származtatható eredmények. Ezek szerint tehát, ha januárban az átlagosnál (a januári átlagos értéknél) 1 °C-kal van hidegebb, akkor a fogyasztás *ceteris paribus* átlagosan 3,70 (februárban ez az érték 4,90, márciusban 5,72) százalékkal magasabb.

Az eredmények nem meglepők, hiszen decemberben-januárban van jellemzően a leghidegebb, így ha ezekben a hónapokban az átlagnál még 1 °C-kal alacsonyabb a hőmérséklet, akkor nem emelkedik annyira a fogyasztás, mint a többi hónapban. Részben azért, mert a fűtési rendszereknek is van egy limitje, részben pedig azért, mert ha a hőmérséklet –1 °C-ról –2 °C-ra csökken, annak valószínűleg nagyobb a hatása, mint ha –5 °C-ról –6 °C-ra (például a spórolás miatt is). Ezzel magyarázható az is minden bizonnyal, hogy az áprilisi érték a legmagasabb

7. táblázat: A HDD-eltérés paramétereinek értéke és származtatott értelmezése a hazai földgázfogyasztás példáján

| Hónap | β | $\exp(\beta) - 1$ [százalék] |
|------------|---------|---------------------------------|
| Január | 0,0363 | 3,70 |
| Február | 0,0478 | 4,90 |
| Március | 0,0556 | 5,72 |
| Április | 0,0749 | 7,78 |
| Május | * | * |
| Június | ** | ** |
| Július | ** | ** |
| Augusztus | ** | ** |
| Szeptember | * | * |
| Október | 0,0439 | 4,49 |
| November | 0,0427 | 4,36 |
| December | 0,0320 | 3,25 |

* A paraméter nem szignifikáns.

** Paraméterbecslés nem történt (HDD-eltérés zéró).

Forrás: saját számítási eredmények (X-13ARIMA-SEATS) és saját készítésű táblázat (Excel).

Természetesen előfordulhat extrém hideg májusban és szeptemberben is (bár ez ritka, a fűtési szezon jellemzően október közepén indul), de emiatt nem lép fel többlet fűtési hatás, így a HDD-eltérés ezekben a hónapokban nem szignifikáns. Következésképpen, ha májusban és szeptemberben van HDD-eltérés, az ezzel történő kiigazítás sem indokolt. A januári hónapokban 1 °C-nyi HDD-eltérés *ceteris paribus* átlagosan

3,70 százalékkal magasabb fogyasztást jelent, így amennyiben egy január 1 °C-kal hidegebb volt az átlagosnál, akkor ennek a hónapnak a hőmérsékletkiigazított fogyasztását 3,70 százalékkal kell csökkenteni a valós fogyasztáshoz képest, hiszen az átlagosnál hidegebb idő okozta a többletfogyasztást; ha 2 °C-kal volt hidegebb, akkor kétszer ennyivel, és így tovább. A kiigazítás természetesen fordított irányban is érvényes, tehát ha egy január 1 °C-kal melegebb volt az átlagosnál, akkor ennek a hónapnak a hőmérsékletkiigazított fogyasztását 3,70 százalékkal kell növelni, hiszen az átlagosnál enyhébb idő okozta az alacsonyabb fogyasztást. Általánosságban a bemutatott modell esetében a hőmérsékletkiigazítás a következők szerint végezhető el:

$$Y \exp(-\beta * HDD\text{-eltérés}),$$

ahol

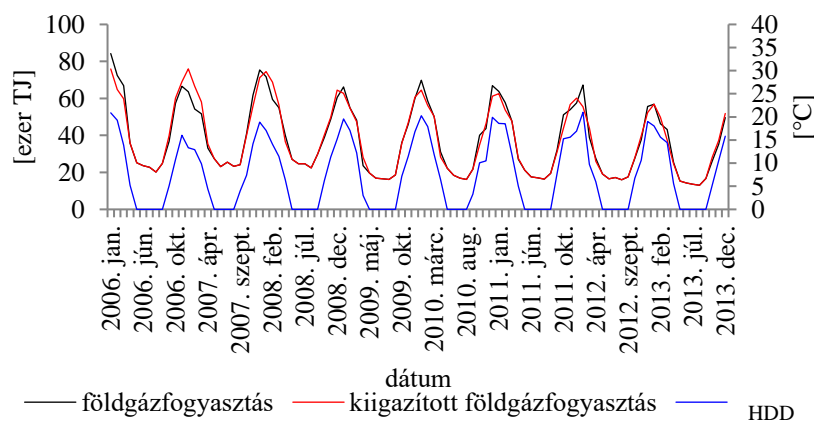
Y – az eredeti idősor adatait tartalmazza,

β – a becsült HDD-eltérés paramétereket jelöli,

HDD-eltérés – a HDD-eltérés havi időszora.

Az előbbi képlet alapján számolva, a hazai földgázfogyasztás hőmérsékletkiigazított időszora látható az 25. ábrán.

25. ábra. A hőmérséklettel kiigazított földgázfogyasztás alakulása, 2006–2013



Forrás: saját számítási eredmények (X-13ARIMA-SEATS) és saját készítésű ábra (Excel).

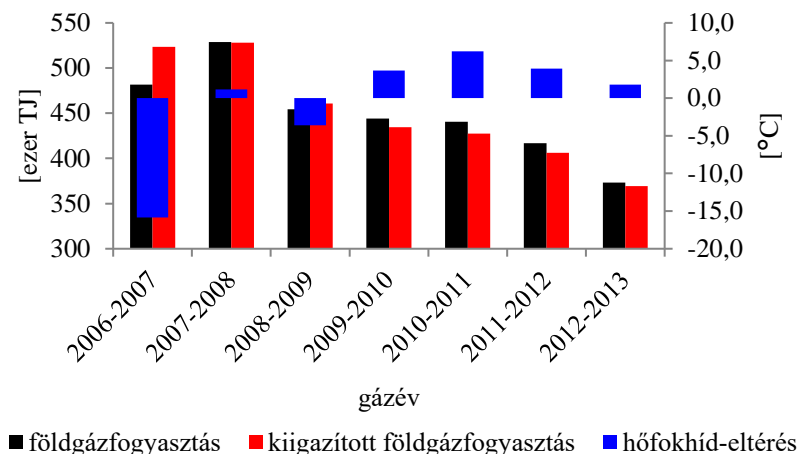
Jól látható, hogy például 2006/2007-ben a tél relatív enyhe volt, így ott a hőmérsékletkiigazított mennyiségek magasabbak a tény mennyiségeknél, még a kiemelten hideg 2012. februárban a kiigazítás lefelé történt.

A 26. ábrán az előbbi eredmények gázév szerint aggregáltan láthatók. Egy gázév július 1-jétől a következő naptári év június 30-ig tart⁷³, így a gázév a telet nem vágja

⁷³ 2015-től azonban változott a szabályozás, és egy három hónapos átmeneti időszakot követően a 2015-

ketté, és könnyebben értékelhető a relatív enyhe és hideg telek hatása (ld. a 2006–2007-es, illetve 2010–2011-es gázéveket).

26. ábra. A hőmérséklettel kiigazított földgázfogyasztás alakulása gázév szerint, 2006–2013



Forrás: saját számítási eredmények (X-13ARIMA-SEATS) és saját készítésű ábra (Excel).

2.2.4.4. A profilozás szempontjából megfogalmazható fontosabb konklúziók

A tanulmány annak ellenére, hogy havi bontású idősorral dolgozik, első ránézésre könnyen adaptálható megoldást közöl a napi, órás vagy akár negyedórás idősorok extrém időjárási (hőmérsékleti) hatásoktól történő megtisztítása tekintetében.

Mivel a profilt definiálhatjuk úgyis, hogy annak az időjárás (hőmérséklet) tipikus (átlagos) tendenciától való eltéréseinek a hatását nem kell tartalmaznia, ezért az itt bemutatott eredmények értékelhetőek egyfajta profilozásként is, hiszen itt is a tipikus (átlagos) hőmérsékleti viszonyok melletti fogyasztás kiszámítása történik meg. Ezt a módszert alkalmazza egyébként *Mutanen et al.* [2011] is. Nagy megfigyelési gyakoriságú idősorok esetén azonban ez a megoldás nem olyan előnyös, amit többek között a hőmérsékletben jelenlévő többszintű (napi és éves) szezonális megletével, fáziseltolódásokkal (előbb vagy később jön a nyár vagy a tél) is magyarázható. Nehézkesebbé válik például annak definiálása, hogy adott időpontban mi számít átlagos, várható hőmérsékletnek⁷⁴; illetve ez a fajta diszkretizálás okozhat túl nagy szórást is, ami miatt egy-egy paraméter nem tűnik szignifikánsnak. Egy ilyen megoldásnak a linearitás hiányának a kezelésére tehát elsősorban aggregált (például havi vagy negyedéves) idősorok esetén van igazán létjogosultsága.

2016-os gázév kezdete már 2015. október 1-je.

⁷⁴ *Mutanen et al.* [2011] olyan megoldást alkalmaz például, ahol hónapok illetve hét napjai alapján készít részátlagokat, és a véletlen hőmérsékleti hatás alatt ezekhez a részátlagokhoz képesti eltérés értendő.

Elsősorban az elemzés fókuszából adódóan az időben változó szóródással, a heteroszkedaszticitással itt kevésbé foglalkoztunk, azonban az eddigieket kiegészítendően fontos megállapítások tehetők ezzel kapcsolatban is.

A 21. ábra alapján a földgázfogyasztás a hőmérsékletfüggő és a nem hőmérsékletfüggő időszakokban gyakorlatilag konstans szórásúnak tűnik. A fejezetben lévő eredmények között ugyan nem szerepelt, de az értekezés további részéhez mindenképpen hasznos alapot jelent, ha a szezonális kiigazítást elvégezzük úgy, hogy **HDD-eltéréseket** helyett csak HDD-vel dolgozunk, illetve a **logaritmizálást** nem végezzük el. A D) melléklet alapján nagyon jól látszik, hogy amennyiben az itteni fejezetben alkalmazott modellt nem logaritmizált idősoron végezzük el, az SI-arányok hónapokon belüli stabilitásából veszítünk valamennyit (D) melléklet d) ábra), amennyiben pedig a HDD-eltérések helyett csak HDD-vel dolgozunk (D) melléklet a) és c) ábra), akkor még inkább.^{75,76}

Az is fontos megfigyelés egyébként, hogy az SI-arányok szabályos „íve” csak a HDD-eltéréseket tartalmazó modellek esetén jelentkezik. Ennek az lehet az oka, hogy ilyenkor a HDD illetve a regressziós hatások nélküli szezonális komponens összefüggnek, és ebben az esetben a HDD hatásának a kiszűrése önmagában – az értelmezés szempontjából is – kevésbé helytálló. A multikollinearitás (magyarázóváltozók – szoros – összefüggésének) problémájához hasonló probléma ez, olyan, mint amikor egy többváltozós regresszióban a HDD mellett szezonális (havi) *dummy* változókat is szerepeltetünk. Ilyenkor azonban nemcsak a becsült paraméterek értelmezése nehézkes, de a regressziós dekompozíciós elv megvalósítása is.

Érdemes megjegyezni azt is, hogy kismértékű heteroszkedaszticitás azért az itteni példában is van (azaz a hőmérsékletfüggő időszakok szóródása valamivel magasabb), azonban az időszak elején megjelenő nyári magasabb fogyasztási szintek miatt ez kevésbé látszik a 21. ábrán. Mivel aggregált és nem olyan hosszú (több 100 vagy 1000 elemű) idősorról van szó, ezekből az adottságokból fakadóan a heteroszkedaszticitás itt sokkal kevésbé kimutatható számszakilag (például statisztikai tesztekkel), grafikusan (származtatott eredményeken viszont) annál inkább (ld. a D) melléklet ábráit).

⁷⁵ A mellékletben szereplő eredmények esetén – a könnyebb összehasonlítás kedvéért – nem becsülünk havonta különböző HDD-eltérés paramétert.

⁷⁶ Hasonló konklúziók leszűrhetők egyébként egy SARIMA-modell reziduuma alapján is, a szemléletes értelmezési lehetőséget a szezonális kiigazítás keretrendszere és az SI-arányokon keresztül némileg elegánsabb vizsgálata adja.

Nagy megfigyelési gyakoriságú (például órás, negyedórás) idősorok jellemzően az itteninél sokkal inkább heteroszkedasztikusak (ahogyan azt a stilizált tények vizsgálatánál is láttuk), ráadásul a linearitás hiányának (esetleg az interakciós kapcsolatoknak) a kezelése még összetettebben merül fel, ezért ott ennek a két jelenségnek az együttes, helyes kezelése sokkal inkább kritikus, és egy alapvetően lineáris modell változóinak az ehhez hasonló praktikus transzformációja nem feltétlenül megvalósítható. Ráadásul az itteni megoldás következményei sem olyan kedvezőek, hiszen mind a logaritmizálás, mind a HDD-eltérések használta lényegében kezelte (tulajdonképpen „megszüntette”) a heteroszkedaszticitást.⁷⁷

A fentieket összegezve tehát, amennyiben a fogyasztási idősor véletlen komponensének a viselkedése fontos az elemzés szempontjából, akkor a hőmérséklet (HDD) teljes hatásának a kiszűrése technikai és értelmezésszerű szempontból sem igazán kedvező, a hőmérséklet véletlen hatásának a kiszűrése pedig a fogyasztás heteroszkedasztikus viselkedését mérsékelheti. Hasonló tendenciák nagy megfigyelési gyakoriságú idősorok esetén is érvényesek, az ennek a pontnak az elején említett nehézségek mellett.

A felsorolt eredmények, tapasztalatok tehát mind a **megfelelő módszertan** kiválasztása, mind a további **kutatási kérdések** megfogalmazása szempontjából fontosnak tekinthetők. Mindezekhez hozzájárul még az az elvi megfontolás is, hogy mivel az egyébként is jellemzően árrugalmatlan villamosenergia-fogyasztás alakulásának és bizonytalanságának egyik fontos forrása az időjárás (elsősorban a hőmérséklet), ezért a hatásának kiszűrése és egy attól megtisztított idősornak a vizsgálata mind a lehetséges elemzések, módszerek, mind a levonható következtetések szempontjából korlátozó tényező lehet.

A hőmérséklet (időjárás) hatásának a kezelését lezárandó és a következő fejezetet megelőlegezendő, érdemes az itt alkalmazott módszertan kapcsán kihangsúlyozni, hogy maga a profilozás sok tekintetben hasonlíthat a szezonális kiigazításhoz (például előigazítások, dekompozíciós logika, több lehetséges megoldás, stb.), és utóbbi módszertannak a komplexitása a profilozásra nézve is árulkodó lehet. Az értekezésben mi nem egy ilyen többlépcsős, komplex megoldás alkalmazásának a lehetőségét vizsgáljuk meg.

⁷⁷ A logaritmus transzformációt – ami más függvényformát és paraméterértelmezést eredményez – egyébként sokszor használják a gyakorlatban is, például úgy, hogy a fogyasztás logaritmusát ábrázolják a hőmérséklet függvényében (vö. az 1.4. fejezetbeli *scatter plot*-ok), mondván, hogy 1°C hatása a fogyasztás relatív (%-os) változásával sokkal jobban összefügg, mint annak abszolút változásával. Ezt modellszelekciós kritériumok ebben a tanulmányban is igazolták.

3. AZ ÉRTEKEZÉSBEN HASZNÁLT MÓDSZEREK ÉS PROFILOZÁSI CÉLÚ ALKALMAZÁSAIK BEMUTATÁSA

Az 1. fejezetben a fogyasztási görbéket jellemző ún. stilizált tények vizsgálata során alkalmaztunk olyan módszereket, amelyeket egyszerűbb logikájuk, de sok esetben kevésbé ismert vagy elterjedt használatuk miatt célszerűnek tartottunk röviden ott ismertetni.

Az ennek a fejezetnek az elején szereplő klasszikus (vagy annak tekinthető) sztochasztikus idősoros regressziós modellek a saját empirikus kutatási eredményeket tartalmazó (4.) fejezetében részben *benchmark*-ként, viszonyítási alapként szolgálnak, ezért csak röviden tekintjük át őket. A SARIMA-modellre a sztochasztikus idősor-elemzés alapmodelljeként szokás hivatkozni. Az ennek kiterjesztésével kapott periodikus autoregresszív modell stacioner idősorokon történő, az értekezés témájához kapcsolódó praktikus alkalmazásának a módszertani háttérével is – korábban ebben a témában írt, saját publikációt alapul véve (Mák [2014a], Mák [2014b]) – itt foglalkozunk.

A módszertani ismertetés jelentős hányadát a keverék-modellek, azon belül is a Gauss-féle keverék-modell módszertani áttekintése teszi ki; beleértve többek között a modellre épülő regressziós megoldás (Gauss-féle keverék-regresszió) részletesebb ismertetését. Mindezt az indokolja, hogy ez a terület kevésbé ismert, különösen a magyar nyelvű szakirodalomban és a gyakorlatban. A fejezetet a keverék-modellek energetikai (részben profilozási) területéről származó, korábbi eredményeinek a bemutatása zárja.

3.1. Klasszikus sztochasztikus idősoros regressziós modellek

A sztochasztikus idősor-elemzés egyik központi fogalma a **stacionaritás**. A stacionaritás jelentősége többek között abban áll, hogy alapvetően stacioner idősor modellezhető. Amennyiben ez nem teljesül, akkor valamilyen módon az idősort stacionerre kell tenni. Az értekezés szempontjából érdemes még megemlíteni, hogy a **véletlen** szerepe stacioner és nem-stacioner (integrált) idősor esetében más jellegű, mind a módszertani kezelés, mind a (gyakorlati) következmények tekintetében. A **klasszikus idősor-modellek felírása** előtt kevésbé formálisan ezeket a kérdéseket is tárgyaljuk. Részletesebb ismeretek a vonatkozó szakirodalmi munkákban találhatunk (ld. többek között Hamilton [1994], Maddala [2004], Ramanathan [2003] munkáit).

3.1.1. A stacionaritás fogalma és tesztelése

Gyengén vagy kovariancia stacionáriusnak nevezünk egy idősort, amennyiben várható értéke és varianciája (szórása) időben állandó, konstans. **Szigorúan stacionárius** egy idősor akkor, ha bármely $t_1, t_2 \dots t_m$ időhalmaz változóinak együttes valószínűség-eloszlása megegyezik bármely $t_{k+1}, t_{k+2} \dots t_{k+m}$ időhalmaz változóinak együttes valószínűség-eloszlásával, ahol k tetszőleges egész szám, azaz a változók együttes eloszlása csak a $t_1, t_2 \dots t_m$ időpontok egymástól való távolságától függ, t -től azonban nem. A szigorú stacionaritásból következik a gyenge stacionaritás, fordítva azonban nem igaz az állítás, ugyanis a magasabb rendű momentumok nem biztos, hogy időfüggetlenek.

A **stacionaritás megléte vagy hiánya** megállapításának alapvető eszközét jelentik az egységgyök tesztek⁷⁸. Ennek legelterjedtebb, és máig használt változata, az ún. ADF-teszt (*Augmented Dickey-Fuller-teszt* (Dickey-Fuller [1979])), KPSS-teszt (*Kwiatkowski-Phillips-Schmidt-Shin-teszt*, Kwiatkowski et al. [1986]), szezonalitást is tartalmazó idősorok esetén pedig például a HEGY-teszt szolgál a szezonális egységgyök meglétének tesztelésére (Hylleberg et al. [1990]).

3.1.2. A véletlen szerepe integrált idősorok esetében

Különösen energetikai idősorok esetén igaz az, hogy a stacionaritás hiányának alapvetően két forrása van: a **trend** illetve a **szezonális** jelenléte. Trend tekintetében alapvetően két eset lehetséges: az idősor determinisztikus vagy sztochasztikus trendet tartalmaz (azaz egységgyök van benne) – ritkább esetben mindkettőt. Párhuzamot vonva a szezonalitást illetően, az is modellezhető determinisztikus vagy sztochasztikus módon. A stacionaritás hiányának determinisztikus eredetű okai esetén a véletlen szerepe gyakorlatilag olyan, mint stacioner idősoroknál, erről a következő szakaszban írunk.

Amennyiben a stacionaritás hiányának sztochasztikus jellegű a forrása, azt egyszerű differenciaképzéssel ($y_t - y_{t-1}$), illetve szezonális differenciaképzéssel (általában $y_t - y_{t-s}$ ⁷⁹) szokás szűni, utóbbi alkalmazásának azonban két rendkívül markáns és komoly feltételezése van: egyrészt valamennyi (egy darab nem-szezonális és $(s - 1)$ darab szezonális) egységgyök megléte (ami miatt gyakran felléphet a túldifferenciázás problémája), másrészt nem-szezonális és a megfelelő szezonális komponensek

⁷⁸ Az egységgyök elnevezés a SARMA-modell 3.1.4. alfejezetben történő formális felírásához kötődik. Egy folyamat stacioner, amennyiben a származtatható differenciálegyenlet gyökei az egységkörön kívül helyezkednek el, és egységgyökfolyamat, amennyiben a gyökök az egységkörön helyezkednek el.

⁷⁹ Az s a szezon periodicitására utal.

függetlensége. A már korábban is hivatkozott saját tanulmányokban (Mák [2014a] illetve Mák [2014b]) részletesen tárgyaljuk ezen függetlenségi követelmény feloldásának többek között modellfeltételeket illetve a stacionaritás tesztelését érintő következményeit.⁸⁰

Az értekezés szempontjából azt érdemes inkább kiemelni, hogy a tanulmányokban bemutatott keretrendszer alkalmas arra, hogy megvizsgáljuk a **sokkok (véletlenek) időbeli lefutásának alakulását**, azaz azt, hogy melyek azok az időszakok (szezonomok), amelyek hatása hosszabb távon sokkal inkább **beépül** az idősor értékébe, illetve melyek azok, amikor a sokkok sokkal inkább **begyűrűznek**. Mivel a sokkok hosszú távú beépülése, megjelenése az idősorba szezonálisan különböző, ezért a modell az összefüggő (egymástól nem független) sztochasztikus trend és szezonális egyfajta modellezéseként is felfogható. Más oldalról megfogalmazva ez azt jelenti, hogy a szezonok kockázata nem azonos, és integrált idősor esetében ez a megoldás ennek a **szezonfüggő kockázatnak** a viselkedéséről is tud nyilatkozni.

3.1.3. A véletlen szerepe stacioner idősorok esetében

Stacioner idősorok esetében – szemben az integrált idősorokkal – a véletlen hosszú távon nem épül be az idősor várható értékébe.⁸¹ A klasszikus regressziós feltételrendszer alapján a véletlen tényező fehér zaj (*White Noise*, röviden *WN*), így a vele szemben a feltevéseink az alábbiak:

- várható értéke zéró,
- szórása állandó (azaz homoszkedasztikus),
- értékei egymástól (időben) függetlenek (azaz autokorrelálatlanok).

A gyakorlatban azonban itt sem teljesül a legtöbb esetben az, hogy a véletlen „igazi” véletlen, azaz független, azonos eloszlású (**FAE**⁸²) valószínűségi változó.

⁸⁰ A tanulmányokban a periodikus autoregresszív modellstruktúra felhasználásával mutatjuk be, hogy hogyan tesztelhető a stacionaritás megléte vagy hiánya akkor, amikor a sztochasztikus trend és szezonális egymástól nem függetlenek. A függetlenség feltételezésének feloldását, azaz azt, hogy a szezonális komponens értéke függ a trendtől (vagy fordítva), a gyakorlatban a komponensek multiplikatív összekapcsolódásaként szokás azonosítani. Így az ott ismertetett technika lényegében multiplikatív modellkörnyezetben végrehajtott egységgyök tesztelésként is felfogható. Vannak egyébként olyan technikák, amelyek az additív és multiplikatív modellkapcsolódás közötti választást segítik, ennek tárgyalása azonban meghaladja tanulmányunk kereteit (bővebben lásd Sugár [1999a], [1999b] munkáit).

⁸¹ Stacioner idősorok tekintjük tehát az olyan idősorokat is, amelyek különböző determinisztikus változók (például analitikus trend, *dummy* változók) beépítése mellett stacionerek.

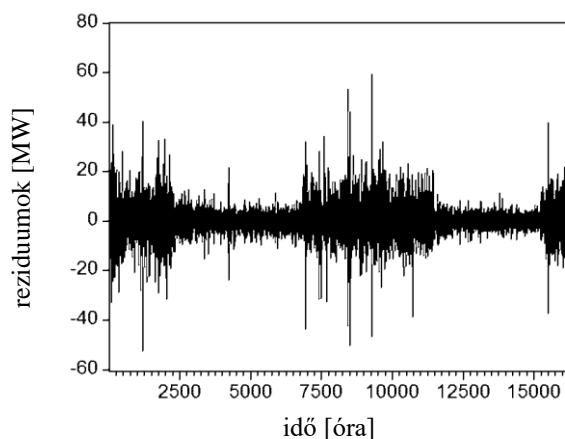
⁸² Vagy angolul: **IID**, *independent, identically distributed*.

Pénzügyi idősorok esetében például az egyik szembevetendő tulajdonság az időben változó szóródás, a heteroszkedaszticitás jelensége. A pénzügyi területen érvényesülő empirikus tapasztalat (az ún. *volatility clustering*, a volatilitás klasztereződése, ld. *Cont* [2005]) szerint bizonyos időszakokban hektikusabbak az árak, jobban szóródnak, bizonyos időszakokban kevésbé, és stacioner idősoroknál a kockázatot többek között ez jelenti.

Ez a heteroszkedaszticitás az energiapiacra azonban sokkal inkább érződik **fundamentális** okokkal magyarázhatónak (pl. napon belüli szezonális, évszakok, hőmérséklet), mint pénzügyi piacok esetében, ugyanakkor annál sokkal kevésbé feltérképezett illetve modellezett területről van szó.

Az 27. ábrán egy erősen téli hőmérsékletfüggő (fűtési/melegítési célú energiafelhasználás, hőfogyasztás) idősorra illesztett SARMA-modell reziduumaik alakulása látható. Egy valamivel több, mint két év hosszú, órás bontású idősorról van szó, ahol a téli időszakokban nagyobb, a nyári időszakokban kisebb a véletlen szórása.⁸³ A szóródást jellemző tendencia a(z alacsony) hőmérséklettel, illetve az emberi tevékenységgel, az abból származó „megnövekedett” véletlen hatással magyarázható.

27. ábra: Reziduumok alakulása egy SARIMA-modellben



Forrás: saját számítások (EViews) alapján saját készítésű ábra (EViews).

A pénzügyi piacok esetében a fundamentális alapú árfolyammodellek kevésbé rövidtávra tekintenek, rövidtávon pedig inkább a (G)ARCH-típusú modellek elterjedtek ((G)ARCH, (*generalized*) *autoregressive conditional heteroskedasticity*, ld. *Bollerslev* [1986], *Engle* [1982]), az itt felírásra kerülő ún. varianciaegyenletben (ami a szóródás

⁸³ A vízszintes tengelyen „csak” időpontok szerepelnek, nem dátumok.

modellezésének explicit megvalósulása) azonban a fundamentális változók alkalmazása mellett ezek használata már nem olyan kényelmes.⁸⁴

Érdemes még megjegyezni, hogy a heteroszkedaszticitásnak alapvetően több oka lehet. Energetikai idősorok esetén kiemelten vetődik fel **a linearitás hiányának** a megfelelő kezelése. Ennek a modellezésére is sokféle lehetőség van (a napfok-módszer is ilyen, vagy exploratív szemléletben a MARS-modellek⁸⁵ alkalmazása), így mindenképpen érdemes megvizsgálni azt, hogy a linearitás nem megfelelő kezelése vagy más változó kihagyása állhat-e például a heteroszkedaszticitás mögött.

Az időben változó szóródás, a heteroszkedaszticitás egyébként a becsült paraméterekre nincs káros hatással, azok továbbra is torzítatlanok maradnak, így például előrejelzésre ilyen modellek is nagyon jól használhatóak. A paraméterek standard hibái azonban torzítottak és inkonzisztensek lesznek.

3.1.4. SARMA-modell

A stacioner idősorok elemzésének alapvető eszközei az ún. ARMA-modellek. Az ARMA-(*autoregressive moving average*, azaz autoregresszív, mozgó átlagolású) modellek általános alakja a következő:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

ahol az ε_t hibatag fehér zaj ($\varepsilon_t \sim WN(0, \sigma^2)$) és $t = 1, 2, \dots, T$, T a megfigyelt idősor hossza, p az AR (autoregresszív), q pedig a MA (mozgóátlagolású) késleltetések számát jelöli. Az autoregresszív tagok értelmében a folyamat adott időszaki értéke függ az idősor korábbi értékeitől, **korábbi realizációitól**, a mozgóátlagolású tagok pedig azt jelentik, hogy az idősor adott időszaki értéke függ **a korábbi hibáktól, véletlenektől**. Az ARMA-modellek szezonális késleltetésekkel történő kibővítése megtalálható például

⁸⁴ (G)ARCH-modellek esetében a SARIMA-specifikációhoz képest (ld. 3.1.4. szakasz) nem feltételezzük az ε_t reziduális változóról, hogy fehér zaj, a **szórása az idő függvénye** lesz, mégpedig a következőképpen:

$$\varepsilon_t = \sigma_t \cdot u_t,$$

ahol az u_t ún. mögöttes reziduumok fehér zaj folyamatot követnek, azaz $u_t \sim WN(0, \sigma^2)$. A várható érték-egyenleten túl így becsülendő lesz még az ún. **varianciaegyenlet**, amely a reziduumok időben változó szórását modellezi. Egy általános GARCH(q,p) modell a következőképpen írható fel:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2.$$

ahol q az ARCH tagok késleltetési rendje, p pedig a GARCH tagok késleltetési rendje. Ennek értelmében egy adott időszak ún. feltételes varianciája (*conditional variance*) függ az előző időszaki reziduumok-négyzetektől, illetve korábbi napok feltételes varianciáitól, illetve az ún. feltétel nélküli variancia (*unconditional variance*, vagy hosszú távú variancia) is könnyen származtatható. A feltételes variancia más modellkeretben (keverék-regresszió) történő felírásával még később találkozunk a fejezetben.

⁸⁵ MARS, *multiple adaptive regression splines* (ld. bővebben Friedman [1991]).

Box és Jenkins [1976] munkájában. Az ún. **szezonális késleltetések** a SARMA (*seasonal autoregressive moving average*) modellben hasonló módon kerülnek beépítésre, jelölésben P -vel a szezonális AR (autoregresszív), Q -val pedig a szezonális MA (mozgóátlagolású) késleltetések számát szokás jelölni.⁸⁶

A SARMA-modellek paramétereinek becslése az ún. *Maximum Likelihood* (ML) eljárással történik.⁸⁷ Érdemes megjegyezni, hogy a különböző autoregresszív és mozgóátlagolású tagok mellett a modellekbe további ún. exogén változók is beépíthetők (így kapjuk az ún. SARMAX-modelleket), amelyek energetikai idősorok esetén tipikusan különböző, a szezonaritást rögzítő *dummy* változókat, vagy időjárást leíró változókat jelentenek. A regressziószámítással kapcsolatban magyar nyelven is bőséges szakirodalom áll rendelkezésre (ld. például Hunyadi-Vita [2003], Kerékgyártó et al. [2008]), a korábban már említett, a témát elsősorban az ökonometriai alkalmazások szemszögéből tárgyaló munkák mellett.

3.1.5. PAR-modell (Periodikus autoregresszív modell)

A következő szakasz két részletesebb tanulmány kiemelt részeit tartalmazza (ld. Mák [2014a] illetve Mák [2014b]), amelyek elsősorban integrált idősorokra koncentrálnak az alkalmazási lehetőségek tekintetében is.⁸⁸ Mivel a dolgozat szempontjából a stacioner folyamatok az érdekesek, ezért itt a stacioner esetre helyezzük a hangsúlyt az elemzési lehetőségek tekintetében.

3.1.5.1. A modell felírása, jelölések

Kiindulva a hagyományos p -ed rendű autoregresszív modell felírásából:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t,$$

⁸⁶ Itt csak terminológiai okokból említjük meg, hogy általánosan $SARIMA(p, d, q)(P, D, Q)_S$ módon jelöljük azokat a modelleket, ahol az eredeti idősor d egyszerű, és D szezonális differenciázás után tettük stacionerré (p és q az AR és MA, P és Q pedig a szezonális AR és szezonális MA késleltetések számának jelölésére már ismert).

⁸⁷ A megfelelő késleltetésszámok megválasztása általában korrelogram és/vagy modellszelekciós kritériumok felhasználásával történik. A gyakorlatban a legtöbbet alkalmazott, *Maximum Likelihood* alapú **modellszelekciós kritériumok** az Akaike információs kritérium ($AIC = -2(\ln L) + 2m$), Schwarz információs kritérium ($BIC = -2(\ln L) + m(\ln n)$). A fentiekben m jelöli a magyarázóváltozók számát, L pedig az optimalizált likelihood függvény értékét. A kritériumok minimuma a választandó modellt jelzi. Helyesen felépített SARMA-modell hibatagja fehér zaj, a **fehér zaj tesztelésének** a korrelogram mellett természetesen még több eszköze is van, többek között a Ljung-Box féle Q -statisztika (ld. Ljung-Box [1978]), vagy a Breusch-Godfrey-féle LM-teszt (ld. Breusch [1978], Godfrey [1978]).

⁸⁸ Ez az ún. periodikusan integrált autoregresszív modell, azaz PIAR (*Periodically Integrated Autoregressive*) modell.

ahol az ε_t hibatag fehér zaj ($\varepsilon_t \sim WN(0, \sigma^2)$); ennek kiterjesztése p -ed rendű periodikus autoregresszív modellre a következő (az egyszerűség kedvéért negyedéves megfigyelési gyakoriságú idősort feltételezünk):

$$y_{t,s} = \phi_{1,s}y_{t-1} + \phi_{2,s}y_{t-2} + \dots + \phi_{p,s}y_{t-p} + \varepsilon_t,$$

ahol a periodicitásnak megfelelően negyedéves idősorok esetén $s = 1, 2, 3, 4$. A felírából látható, hogy a $\phi_{p,s}$ p -ed rendű késleltetéshez tartozó paraméterek szezononként, periódusonként különböznek. A periódusról periódusra változó autoregresszív együtthatók empirikus alátámasztása az, hogy különböző idősorok esetén például az első negyedéves érték nem úgy függ az előző év negyedik negyedévi értékétől, mint a második negyedévié az elsőtől stb. Mivel periódusonként különböző autoregresszív együttható becslése történik meg, ez azt jelenti, hogy a modell alkalmas a periódusonként különböző autokovariancia becslésére is.

Érdemes még említést tenni a modell egyenletrendszerként történő felírásáról (ugyanis az értekezés empirikus részében ez a megközelítés a kézenfekvőbb), ahol az egyenletek száma – értelemszerűen – azonos a szezonok számával, azaz:

$$\Phi_0 Y_{T,s} = \Phi_1 Y_{T-1,s} + \Phi_2 Y_{T-2,s} + \dots + \Phi_p Y_{T-p,s} + E_T,$$

ahol $E_T = [\varepsilon_{T,1} \ \varepsilon_{T,2} \ \varepsilon_{T,3} \ \varepsilon_{T,4}]^T$ vektor változói fehér zajok ($\varepsilon_{T,s} \sim WN(0, \sigma^2)$) és $s = 1, 2, 3, 4$.

Az egyenletrendszerben szereplő változóink $Y_{T,s} = [y_{T,1} \ y_{T,2} \ y_{T,3} \ y_{T,4}]$ illetve $Y_{T-1,s} = [y_{T-1,1} \ y_{T-1,2} \ y_{T-1,3} \ y_{T-1,4}]$; a T -edik és a $(T-1)$ -edik évek negyedévei szerepelnek tehát az „éves” vektorokban. Láthatóan a negyedévente megfigyelt y_t változók indexe megváltozott. A t időváltozót, amelyet negyedéves gyakorisággal rögzítettünk, lecseréltük a T, s időváltozóra, amely szintén negyedévenként rögzít, de megmutatja azt is, hogy mikor melyik év, mely negyedévéről van szó.

A paramétereket tartalmazó mátrixok esetében az első index a késleltetés rendjére utal, a második pedig arra, hogy az adott késleltetési rend melyik periódus egyenleténél érvényes, így a négy késleltetést tartalmazó modellre egyszerűsítve az említetteket, a paramétermátrixok a következők:

$$\Phi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\phi_{1,2} & 1 & 0 & 0 \\ -\phi_{2,3} & -\phi_{1,3} & 1 & 0 \\ -\phi_{3,4} & -\phi_{2,4} & -\phi_{1,4} & 1 \end{bmatrix}, \text{ illetve } \Phi_1 = \begin{bmatrix} \phi_{4,1} & \phi_{3,1} & \phi_{2,1} & \phi_{1,1} \\ 0 & \phi_{4,2} & \phi_{3,2} & \phi_{2,2} \\ 0 & 0 & \phi_{4,3} & \phi_{3,3} \\ 0 & 0 & 0 & \phi_{4,4} \end{bmatrix}$$

Azaz például a Φ_1 mátrix első sora azt mutatja, hogy a T -edik év első negyedéve hogyan függ a $(T - 1)$ -edik év első, második, harmadik és negyedik negyedéveitől. A sorrend a mátrixban természetesen pont fordított, mint az időbeli késleltetés rendje indokolná, hiszen a T -edik év első negyedévé a $(T - 1)$ -edik év negyedik negyedéve előzi meg és együttthatója éppen ezért $\phi_{1,1}$.

A fentiek alapján már rekonstruálható is az egyenletrendszer, ami tehát az alábbi (átrendezés után, ld. az előző bekezdésben foglaltakat):

$$y_{T,1} = \phi_{1,1}y_{T-1,4} + \phi_{2,1}y_{T-1,3} + \phi_{3,1}y_{T-1,2} + \phi_{4,1}y_{T-1,1} + \varepsilon_{T,1}$$

$$y_{T,2} = \phi_{1,2}y_{T,1} + \phi_{2,2}y_{T-1,4} + \phi_{3,2}y_{T-1,3} + \phi_{4,2}y_{T-1,2} + \varepsilon_{T,2}$$

$$y_{T,3} = \phi_{1,3}y_{T,2} + \phi_{2,3}y_{T,1} + \phi_{3,3}y_{T-1,4} + \phi_{4,3}y_{T-1,3} + \varepsilon_{T,3}$$

$$y_{T,4} = \phi_{1,4}y_{T,3} + \phi_{2,4}y_{T,2} + \phi_{3,4}y_{T,1} + \phi_{4,4}y_{T-1,4} + \varepsilon_{T,4}$$

Érdeemes megjegyezni, hogy a mátrixoknak praktikus tartalma is van, hiszen Φ_0 az azonos, Φ_1 pedig a megelőző évbe eső negyedévekhez tartozó késleltetések paramétereit tartalmazza.

Összefoglalva tehát a bemutatott PAR-modellnek többféle felírása lehetséges. A legelső felírással a legkönnyebben interpretálható a modell alapötlete. Az egyenletrendszeres felírás ugyan idegennek tűnik az egyváltozós autoregresszív modellek hagyományos, egyegyenletes felírásához képest, a periodikusságból adódóan azonban egyrészt praktikus reprezentáció, másrészt bizonyos eredmények levezetésére is alkalmas. A modellszelekciós lehetőségek a klasszikus regressziós módszertannal azonosak.

3.1.5.2. Hosszú távú egyensúlyi szint (átlag) számolása⁸⁹

A korábbi pontban feltételeztük, hogy centrált az idősorunk, azaz a várható értéke zéró, így egyik felírás esetén sem szerepelt konstans tag a modellben. Vannak olyan gyakorlati alkalmazások, amikor ez a megoldás nem megfelelő. Ebben az esetben a modell mátrix formában történő felírása az alábbiak szerint módosul:

$$\Phi_0 Y_{T,s} = C + \Phi_1 Y_{T-1,s} + \Phi_2 Y_{T-2,s} + \dots + \Phi_p Y_{T-p,s} + E_T,$$

ahol C a peridusonkénti konstans paramétereket tartalmazza (amelyek lehetnek periódusonként azonosak, de különbözőek is). Amennyiben az idősor stacioner, definiálható az idősor hosszú távú átlaga, azaz várható értéke:

$$E(Y_{T,s}) = (\Phi_0 - \Phi_1 - \Phi_2 \dots - \Phi_p)^{-1} C,$$

⁸⁹ A cím nagyon általánosnak tűnik, az eredmény profilként történő felhasználása teszi az értekezés szempontjából kiemelt jelentőségűvé.

ahol $E(\cdot)$ a várható érték jelölésére szolgál. A levezetés nagyon egyszerű, hiszen amennyiben ez a hosszú távú átlag valóban létezik, ekkor $E(Y_{T,s}) = E(Y_{T-1,s}) = E(Y_{T-2,s}) = \dots = E(Y_{T-p,s})$, emellett $E(E_T) = 0$, így a fenti képlet könnyen adódik.⁹⁰

Az értekezésben olyan alkalmazásokkal találkozunk majd, ahol a konstans és az autoregresszív tagok mellett egyéb exogén változók is szerepelnek majd a magyarázó-változók között (pl. a fűtési illetve hűtési napfok értékek). *Espinoza et al.* [2005] nyomán a profil értelmezhető úgy, mint egy olyan napi óras vagy negyedórás bontású görbe, amelyből minden más változónak a hatása kiszűrésre került. Ez ekvivalens azzal a felfogással, hogy minden exogén változó értékét (pl. napfok értékek, különböző *dummy* változók) egyenlővé tesszük zéróval, és alkalmazzuk a fenti összefüggést. Így az előzőekben bemutatott hosszú távú egyensúlyi szint (átlag) definiálható egyfajta profilként.⁹¹ Erre az empirikus eredményeket tartalmazó fejezetben térünk még vissza.

3.2. Keverék-modellek

Az ún. véges számú komponenszt tartalmazó keverék-modellek (*Finite Mixture Models*) alkalmazása egyre több területen fordul elő, és sok jövőbeni gyakorlati alkalmazásnak is kiindulópontját, alapját jelenthetik. A keverék-modellek (*Mixture Models*, röviden MM) megjelentek már többek között biológiai, mezőgazdasági, orvosi, közgazdasági vagy éppen jelfelismerés, jelfeldolgozás területeken (ld. például *McLachlan-Basford* [1988], *McLachlan-Peel* [2000]). A keverék-modellek családja sok alkalmazást foglal magába. Így beleértendő különböző jelenségek eloszlásának ismert eloszlások (például normális eloszlás) keverékéből történő előállítás, illetve klaszterezés, diszkriminancia-elemzés vagy regressziós becslés keverék-modelleken alapuló elkészítése.

Mivel a dolgozatban alkalmazott módszertan a *Gauss*-féle keverék-modellen (*Gaussian Mixture Model*, röviden GMM), illetve annak regressziós alkalmazásán alapul, ezért részletesen ezeket tekintjük át. A módszer előnye és a dolgozat szempontjából történő alkalmazását alapvetően az indokolta, hogy a pontbecslések standard hibái a magyarázóváltozók aktuális értékének függvényében, azok feltételeként írhatóak fel, így

⁹⁰ Az eredmény egyébként nem más, mint a közismert AR(p) modell esetén a hosszú távú átlag összefüggésnek általánosítása, ami az alábbi:

$$E(y_t) = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p},$$

ahol a $\phi_1, \phi_2, \dots, \phi_p$ paraméterek a megfelelő autoregresszív tagok együtthatóit jelölik.

⁹¹ A cikk megemlíti egyébként, hogy a heteroszkedaszticitás kezelhető a PAR-modellben úgy (az itt is bemutatott alapmodell kiterjesztéseként), hogy periódusonként különböző reziduális szórást becsülünk (negyedórás esetben ez további 96 becsült paramétert jelentene). Ez azonban nem veszi figyelembe például heteroszkedaszticitásnak a hétköznapiok-hétvégek, vagy az évszakok közötti különbözőségét.

a módszer alkalmas lehet a heteroszkedaszticitás (fundamentálisabb alapú) kezelésére. **Ebben a fejezetekben több olyan levezetés található, amelyek más munkákban nem, illetve amelyek az általános többváltozós regresszióval való – első körben talán kevésbé nyilvánvaló – rokonságot mutatják.** A dolgozat emellett nagy megfigyelési gyakoriságú (negyedórás) idősorokkal dolgozik, ami az általános becslési eljárástól való eltérést is megköveteli, ami más empirikus munkában (ld. például a 3.3.3. szakasz példáját) szintén nem jelenik meg. Ennek módszertani hátterét a keverék-modellek véletlen mintavételt használó, klaszterezésen és diszkriminancia-analízisen alapuló felhasználásai jelentik (ld. többek között *Fräley-Raftery* [2000]).

3.2.1. A keverék-modell (MM) és a Gauss-féle keverék-modell (GMM) felírása

Legyenek adottak az egymástól független, $y = (y_1, y_2 \dots y_n)$ megfigyelések, ahol:

- y_i az i -edik megfigyelés attribútumait tartalmazó $(m \times 1)$ méretű vektor ($i = 1, 2 \dots n$),
- n a megfigyelések száma,
- m az attribútumok száma.⁹²

Feltételezzük, hogy a megfigyelések egy K komponenst tartalmazó keverék-eloszlásból állnak elő, amelynek sűrűségfüggvénye az alábbi módon írható fel:

$$f(y) = \prod_{i=1}^n \sum_{k=1}^K \tau_k \cdot f_k(y_i | \theta_k),$$

ahol:

- $f(\cdot)$ a keverék-eloszlás sűrűségfüggvénye,
- $f_k(\cdot)$ a k -adik komponens sűrűségfüggvénye,
- θ_k a k -adik komponens sűrűségfüggvényét leíró paramétereket jelöli,
- τ_k ún. *prior* valószínűség annak a valószínűsége, hogy az i -edik megfigyelés a k -adik komponensbe tartozik,
- k a komponensek jelölésére szolgál ($k = 1, 2, \dots K$),
- K a komponensek száma.

A legtöbb esetben – így a dolgozatban is – feltételezzük, hogy a k -adik komponens eloszlása normális, azaz $f_k(\cdot)$ a többváltozós normális eloszlás sűrűségfüggvényét

⁹² A nemzetközi, főként az adatbányász szakirodalomban használatos kifejezés, az attribútum azonos a regressziós terminológiában használt változó kifejezéssel.

jelöli μ_k átlagvektor és Σ_k kovariancia mátrix paraméterekkel, így a k -adik komponens sűrűségfüggvénye az alábbi módon írható fel:

$$f_k(y_i|\theta_k) = \varphi(y_i|\mu_k, \Sigma_k) = \frac{1}{|2\pi\Sigma_k|^{-1/2}} \exp\left[-\frac{1}{2}(y_i - \mu_k)^T \Sigma_k^{-1}(y_i - \mu_k)\right],$$

ahol:

- $\varphi(\cdot)$ a többváltozós normális eloszlás sűrűségfüggvényét,
- $|\cdot|$ a determinánst jelöli.

A komponenseket keverék-modellek csoportképzési, klaszterezési célú felhasználása esetében klaszterként szokás azonosítani, így a továbbiakban előfordulhat, hogy a klaszter és komponens kifejezéseket mi is felváltva használjuk, az első megoldás inkább az értelmezés, interpretáció, a második inkább a módszertani megközelítést hangsúlyozza.

3.2.2. Expectation-Maximization (EM) becslési eljárás

A keverék-modell paramétereinek becslése *Maximum Likelihood* (ML) módszerrel, az *Expectation-Maximization* (EM) algoritmussal történik (ld. többek között *Dempster et al.* [1977], *McLachlan-Krisnan* [1997]).⁹³ Az EM algoritmus ún. E-lépések (*Estimation step* vagy *E-step*) és M-lépések (*Maximization step* vagy *M-step*) iteratív elvégzéséből áll.

Az eljárás során a megfigyeléseinkre úgy tekintünk, mint egy nem teljes (hiányzó, nem megfigyelhető változóértéket is tartalmazó) adathalmazra. Ez azt jelenti, hogy úgy gondolunk rájuk, mint (y_i, z_i) változópárookra. Itt a z_i változó a nem megfigyelhető, ún. indikátorváltozót jelöli, amely azt mutatja meg, hogy egy-egy megfigyelés melyik komponensből származik. Azaz z_{ik} egyenlő 1-gyel, ha az i -edik megfigyelés a k -adik kom-

⁹³ Érdemes megjegyezni, hogy a keverék-modellben szereplő Σ_k kovariancia mátrix struktúrája nagyon sokféleképpen definiálható. Az általános (ún. *unconstrained*) megoldás szerint (ld. *Banfield-Raftery* [1993]):

$$\Sigma_k = \lambda_k D_k A_k D_k^T,$$

ahol:

- D_k a kovariancia mátrix ortogonális sajátvektorait,
- A_k a diagonális mátrixot (amelynek a diagonális elemei arányosak a mátrix sajátértékeivel),
- λ_k pedig ez az ún. arányosító tényezőt jelöli.

Ezek a paraméterek lehetnek klaszterenként különbözőek, illetve paraméterenként tehetőek megkötések is. Fraley és Raftery több tanulmányában is találhatók egyébként példák arra vonatkozóan, hogy az egyes esetek milyen más klaszterezési technikával ekvivalensek (például *Fraley-Raftery* [2000], *Fraley-Raftery* [2007]), hiszen a Σ_k kovariancia mátrix felbontása csak egy keretrendszer. A dolgozatban mi mindenhol az általános, megkötések nélküli (*unconstrained*) változatot fogjuk használni, ahogy azt több megvizsgált tanulmányban is tapasztaltuk (pl. *Srivastava et al.* [2013], *Eirola-Lendasse* [2013]).

ponensbe esik, és 0 egyébként.⁹⁴ Mivel azonban ezek a z_{ik} komponens tagságok hiányzó, nem megfigyelhető értékek, ezért az EM algoritmus során a komponens tagságot becsülni kell, ami a *posterior* p_{ik} valószínűségekből realizálódik.

Jelölje $\psi = (\tau_1, \tau_2 \dots \tau_K, \theta_1, \theta_2 \dots, \theta_K)$ a megbecsülni kívánt paramétereket, azaz a komponensek *prior* valószínűségeit és a normális eloszlás paramétereit. A *likelihood*-függvény az alábbi:

$$L(y) = \prod_{i=1}^n \prod_{k=1}^K f_k(y_i | \theta_k)^{z_{ik}},$$

a *loglikelihood*-függvény pedig:

$$\log(L(y)) = l(y) = \sum_{i=1}^n \sum_{k=1}^K z_{ik} \cdot \log(f_k(y_i | \theta_k)).$$

Tegyük fel, hogy van n darab megfigyelésünk $y = (y_1, y_2 \dots y_n)$, ekkor az $(r + 1)$ -edik iteráció a következő lépések elvégzését jelenti.

Az E-lépésben az r -edik iterációban megbecsült $\psi^{(r)}$ paraméterszett alapján kiszámoljuk minden i megfigyelésre a k -adik komponensbe esés p_{ik} *posterior* valószínűségét:

$$p_{ik}^{(r+1)} = \frac{\tau_k^{(r)} \cdot f(y_i | \theta_k^{(r)})}{\sum_{k=1}^K \tau_k^{(r)} \cdot f(y_i | \theta_k^{(r)})},$$

illetve ennek felhasználásával az ún. Q -függvény értékét, amely a teljes adathalmazra vonatkozó *loglikelihood* várható értéke az aktuális becsült paraméterek és a mintában megfigyelt változók értékei mellett, azaz:

$$Q(\psi | \psi^{(r)}) = \sum_{i=1}^n \sum_{k=1}^K p_{ik}^{(r+1)} \log(f_k(y_i | \theta_k)).$$

Az M-lépésben a kiszámított $p_{ik}^{(r+1)}$ *posterior* valószínűségek, mint súlyok felhasználásával kiszámoljuk a $\psi^{(r+1)}$ paraméterszett értékeit a Q -függvény maximalizálásával, azaz elvégezzük az alábbi optimalizációs feladatot:

$$\psi^{(r+1)} = \arg \max_{\psi} Q(\psi | \psi^{(r)}),$$

vagy másképpen felírva:

$$Q(\psi | \psi^{(r)}) \rightarrow \max_{\psi},^{95}$$

amelynek eredménye az optimális megoldást, azaz a becsülni kívánt paraméterek $(r + 1)$ -edik iterációban becsült értékeit adja eredményül:

⁹⁴ Természetesen így igaz az is, hogy $\sum_{k=1}^K z_{ik} = 1$ minden i megfigyelésr esetén.

⁹⁵ Az optimalizációs feladat kétféle felírása ugyanazt jelenti, a második talán egyszerűbb, az első kicsit formálisabb. Az *argmax* jelölés az *arguments of maxima* kifejezés rövidítése, egy függvény értelmezési tartományának azokat a pontjait jelöli, amelyeken a függvény helyettesítési értéke maximális, tehát a függvény maximumhelyeit.

$$\psi^{(r+1)} = (\tau_1^{(r+1)}, \tau_2^{(r+1)} \dots \tau_K^{(r+1)}, \theta_1^{(r+1)}, \theta_2^{(r+1)} \dots \theta_K^{(r+1)}).$$

Az M-lépés eredményeként a komponensek *prior* valószínűségei egyébként az alábbi képlettel írhatóak fel:

$$\tau_k^{(r+1)} = \frac{\sum_{i=1}^n p_{ik}^{(r+1)}}{n}.$$

A fenti E- és M-lépések addig kerülnek iteratív módon ismételésre, amíg a paraméterbecslések nem kezdenek el konvergálni, vagy el nem érünk egy maximális iterációszámot. A modellek közötti választás modellszelekciós kritériumok alapján történik (pl. *AIC*, *BIC*). A modellszelekciós kritériumok nemcsak a változók optimális számának, hanem a komponensek optimális számának meghatározásában is segítenek, ami – többek között – a modell alapú klaszterezés előnye az egyéb klaszterezési megoldásokkal szemben.⁹⁶

3.2.3. Empirikus példa Budapest napi földgázfogyasztási adatai példáján

A könnyebb érthetőség kedvéért egy egyszerűbb példán keresztül is bemutatjuk a keverék-modellek, a modell alapú klaszterezés előnyét egy hagyományos klaszterezési technika, a k-középpontú (K-Means, egyébként ugyanúgy iteratív) klaszterezés eredményeivel összehasonlítva. Számítási részeredmények a B) mellékletben találhatóak. A 28. ábrán a k-középpontú klaszterezés logikáját összefoglaló ún. pszeudo-kód látható.

A klaszterezési megoldások közötti különbséget Budapest napi földgázfogyasztásának alakulásán mutatjuk be.⁹⁷ A földgázfogyasztási illetve a hőmérsékleti idősor alakulása a 29. ábrán látható. A klaszterezést ennek a két változónak vagy attribútumnak (hőmérséklet illetve gázfogyasztás) a terében fogjuk végezni, tehát a csoportosítani kívánt megfigyelések az órák lesznek. A jobb oldali ábra alapján feltételezhető, hogy legalább két klaszter képzése szükséges, egyrészt a hőmérsékletfüggő időszakok, másrészt a nem hőmérsékletfüggő időszakok számára.

28. ábra: K-középpontú klaszterezés pszeudo-kódja

| | |
|---------------|---|
| I. | <i>k</i> darab kezdő klasztercentroid (véletlenszerű ⁹⁸) kiválasztása |
| REPEAT | |

⁹⁶ A *BIC*-kritérium képlete – az ismert módon – a következő: $BIC = -2 \cdot l(y) + \log(n) \cdot bp$, az *AIC*-kritérium képlete pedig: $AIC = -2 \cdot l(y) + 2 \cdot bp$, ahol *bp* a becsült paraméterek számát jelöli (ami a kovariancia-mátrix struktúrájától függően különböző lehet).

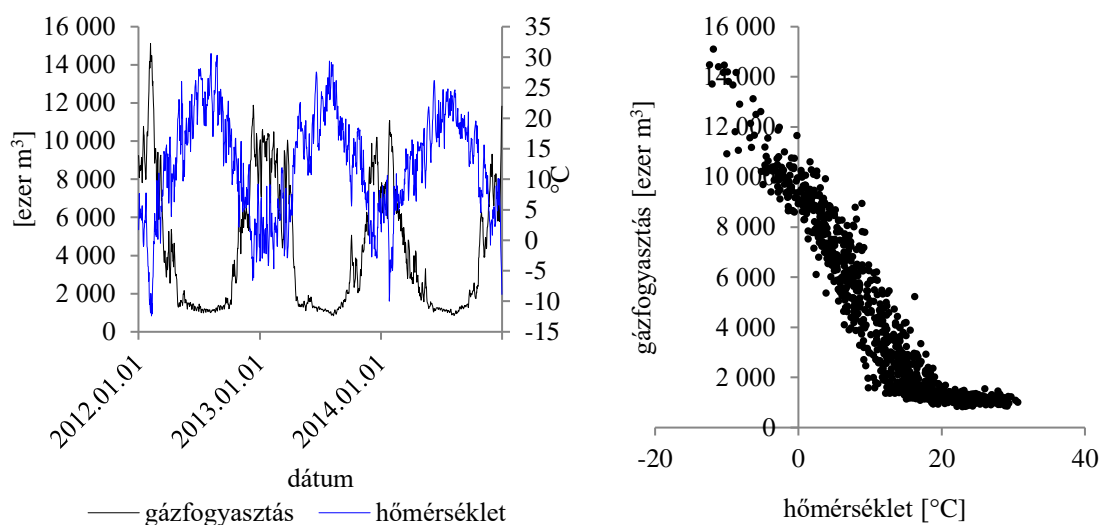
⁹⁷ Adatok forrása: Magyar Földgázszállító Zrt. (FGSz), www.fgsz.hu

⁹⁸ A kezdő klasztercentroidok véletlenszerű kiválasztása többféleképpen is biztosítható. A hagyományos k-középpontú klaszterezés illetve a K-Means++ alapvetően abban különbözik, hogy a K-Means++ esetében ezeknek a kiválasztása hatékonyabb, így futási idő alacsonyabb, mint a hagyományos megoldásnál, amelyik megoldás ráadásul sokszor lokális optimumba szorul.

| | |
|-------|--|
| 2. | <i>megfigyelések klaszterekhez történő hozzárendelése</i> |
| a. | megfigyelések és klasztercentroidok közötti távolságok számítása |
| b. | megfigyeléseknek ahhoz a klaszterhez történő hozzárendelése, aminek klasztercentroidjától számított távolság a legkisebb |
| 3. | <i>klasztercentroidok újraszámolása</i> |
| a. | klaszterenként átlagos megfigyelések számolása és |
| b. | klasztercentroidként történő azonosítása |
| UNTIL | <i>adott konvergenciakritérium nem teljesül</i> |

Forrás: saját készítésű ábra.

29. ábra: Budapesti napi gázfogyasztás és napi középhőmérséklet alakulása illetve középhőmérséklet-gázfogyasztás scatter plot



Forrás: saját készítésű ábra (Excel).

Csak a 2014-es évet kiválasztva a k-középpontú klaszterezés eredményei láthatóak a 30. ábrán. Mivel a k-középpontú klaszterezés esetén szükséges a klaszterek számának előzetes ismerete, ezért az eredményeket bemutatjuk 2, 3 illetve 4 klaszter képzése mellett is.

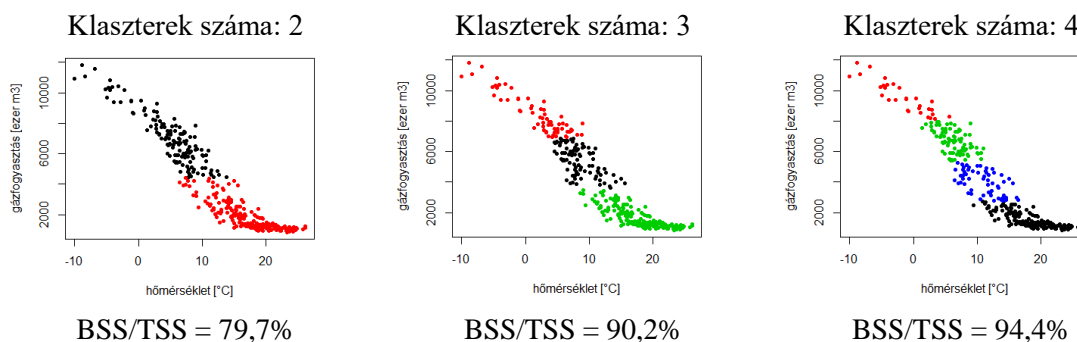
Az ábrán az alsó sorban az ún. külső és a teljes eltérés-négyzetösszegek (röviden BSS és TSS⁹⁹) hányadosai láthatóak. K-középpontú klaszterezés esetén erre az arányszámra (is) lehet támaszkodni az optimális klaszterszám megválasztása tekintetében¹⁰⁰. A klaszterek számát 2-től 10-ig megválasztva, a mellékletben szereplő 19. táblázat mutatja ezeket a hányadosokat. Amellett azonban, hogy a klaszterek megfelelő számának megválasztása ilyen módon elég kényelmetlen (4-5 klaszter, ahol a BSS/TSS értékek alakulása megtörik), az eredmények interpretálása sem túl szép, ugyanis a 30. ábra alap-

⁹⁹ *Between Sum of Squares és Total Sum of Squares.*

¹⁰⁰ Természetesen ennél jóval összetettebb, szofisztikált mutatószámok is léteznek a megfelelő klaszterszám megválasztására.

ján látszik, hogy a klaszterek mintha a hőmérséklet tengellyel párhuzamosan meghúzott sávok mentén szerveződnének (különösen jól látszik ez kevesebb klaszterszám választása esetén).

30. ábra: K-középpontú klaszterezés eredményei a napi középhőmérséklet-gázfogyasztás példán



Forrás: saját számítások (R) és saját készítésű ábra (R).

A k-középpontú klaszterezéssel szemben a keverék-klaszterezésnek megvan az az előnye, hogy ott a klaszterek számának megválasztása automatikusan, modellszelekciós kritérium (BIC) alapján történik (történhet), így a klaszterek számának *a priori* ismeretével, feltételezésével nem kell törődnünk. A mellékletben szereplő 20. táblázat alapján a BIC minimuma jelöli a választandó klasztert, ezért a választandó keverék-modell négy klasztert (komponenst) fog tartalmazni.

Ahogy az az előző fejezetben említettük, minden megfigyelés esetén kiszámításra kerülnek ún. klaszterbe tartozási valószínűségek, amelyeknek megfigyelésenkénti összege 1 (vagy 100%). Amennyiben a megfigyeléseket egyértelműen klaszterhez szeretnénk rendelni, a klasztertagságot a feltételes valószínűségek maximuma adja, vagyis minden megfigyelést oda sorolunk be, ahová az odatartozás valószínűsége a legnagyobb. Ennek alapján mutatja az 31. ábra bal oldala a létrejött 4 klasztert. A zöld színnel jelölt napok a nem hőmérsékletfüggő napok, a hőmérsékletfüggő napok pedig 3 részre oszthatóak. A pirossal jelöltek azok a napok, amikor az ún. fűtési hatás érvényesülni kezd, az ún. átmeneti időszakok, a feketével jelöltek az egyértelműen hőmérsékletfüggő napok, a kékkel jelöltek pedig a leghidegebb napokat jelölik. A különböző klaszterek esetén a fogyasztás hőmérsékletfüggése is más. Akár képzeletben a 4 pontfelhőre regressziós egyeneseket is illeszthetünk¹⁰¹, akár az ábrán látható ellipszisek¹⁰² tengelyeinek irányát is megvizsgálhatjuk ennek alátámasztására.

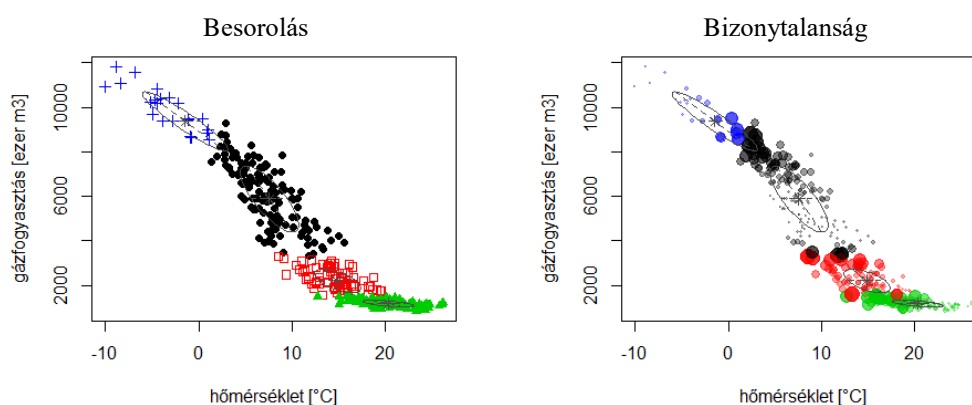
¹⁰¹ Később a keverék-modell regressziós alkalmazása ilyen logikát feltételez, csak nem pontfelhőnként külön-külön, hanem a *posterior* valószínűségek felhasználásával együttesen, súlyozott módon.

A 31. ábra bal oldalán a besorolás bizonytalanságát (angolul *uncertainty*) ábrázoljuk. A megfigyelés besorolásának a bizonytalanságát (u_i) az alábbi módon definiálhatjuk:

$$u_i = 1 - \max_k p_{ik},$$

azaz a maximális besorolási valószínűséget kivonjuk 1-ből, vagyis 100%-ból. Minél kisebb tehát egy megfigyelés esetén maximális *posterior* valószínűség, annál inkább bizonytalan besorolásúnak tekintjük azt.

31. ábra: Keverék-klaszterezés eredményei a napi középhőmérséklet-gázfogyasztás példán



Forrás: saját számítások (R) és saját készítésű ábra (R).

A felhasznált R programcsomag grafikáját használva a klaszterek színei a besorolást ábrázoló grafikonról öröklődnek, és annál nagyobb és sötétebb szimbólummal jelölünk egy megfigyelést, minél nagyobb a bizonytalansága. A konkrét értékek ismerete nélkül is látható, hogy „klaszterhatárokon” nagyobb a bizonytalanság, illetve hogy a leghidegebb (kék szín) illetve a hidegebb (kb. 10 °C alatti) napok „határán”, valamint az átmeneti (piros) napok esetében a legnagyobb a bizonytalanság.

Természetesen itt is számolhatóak klasztercentroidok (ld. 21. táblázat): alacsonyabb átlaghőmérsékletű klaszter esetén a napi átlagfogyasztás is magasabb. Ezek a normális eloszlásból adódóan tipikus, jellemző fogyasztási mintázatként is kezelhetők. **Ezt a megállapítást továbbvisszük a profilozással, a tipikus fogyasztási mintázatok készítésével foglalkozó 4. fejezetbe.**

Érdeemes megvizsgálni a napoknak a hónapok illetve a hét napjai (hétköznapi vagy hétvége) közötti megoszlását klaszterenként (ld. 22. és 23. táblázatok). Gyönyörűen kirajzolódik az időjárás (rendkívül hideg, hideg, átmenet és fűtést nem igénylő) és a

¹⁰² Az ellipszisek a kétdimenziós normális eloszlás $\text{átlag} \pm \text{szórás}$ értéknek megfelelő szintvonalát mutatják.

kapcsolódó gázfogyasztás hónapok közötti megoszlása, valamint az is, a hétköznapi és hétvégi megoszlása a klasztereken belül közel azonos (a hétvégi aránya mindenhol $2/7 \approx 28.6\%$ körüli), azaz nem különül el egy ún. hétvégi klaszter, a fűtési hatás sokkal inkább domináns, mint a heti naptári hatás.¹⁰³

Mivel ez a szakasz elsősorban didaktikai célokat szolgált, természetesen sok kérdés illetve elemzési lehetőség merül fel, amelyeknek nem tudunk helyet biztosítani kellő részletességgel. Ezek közül csak egyet említünk meg, ami az értekezés további részei szempontjából érdekes. Ez a következő: ha klaszterenként a megfigyelések eltérő mértékben szóródnak (itt például a nem hőmérsékletfüggő napok szóródása adott hőmérséklet mellett jóval kisebbnek tűnik, de ez mintha az extrém hideg napokon is igaz lenne), akkor ez milyen mértékben transzformálható a fogyasztás bizonytalanságának mérésére.¹⁰⁴ A módszertani fejezet további részében ennek a kérdésnek a módszertani aspektusával foglalkozunk, a későbbi empirikus fejezetben pedig negyedórás villamosenergia-fogyasztás idősoron mutatjuk be az eredményeket.

3.2.4. A Gauss-féle keverék-moddal kapcsolatos további módszertani kérdések

A fejezetben tárgyalunk néhány olyan módszertani jellegű kérdést, amelyet az EM algoritmus általános bemutatása során nem tettünk meg.

3.2.4.1. Kezdő klasztertagságok illetve posterior valószínűségek megadása

A hagyományos EM algoritmus veszélye, hogy könnyen lokális optimumban ragad vagy „önmagától” nem talál olyan megoldást, amely jól értelmezhető, jól interpretálható. Mivel a végeredmény függ a kezdő klasztertagságok vagy kezdő paraméterek megadásától, ezért lehetőség van ezek manuális megadására, de gyakran javasolt módszer az eljárás többszöri lefuttatása valamilyen véletlenszerű elven alapuló induló klasztertagságokkal vagy paraméterekkel (ilyen témában ld. például *Biernacki et al.* [2003] munkáját).

Lehetséges megoldás az is, hogy első lépésben más, módosított EM algoritmusokkal (pl. CEM, SEM, hierarchikus EM) számolunk, majd az általuk adott eredmények (kezdő klasztertagságok vagy paraméterek) alapján használjuk a hagyományos EM algoritmust.

¹⁰³ Ez azért is fontos megállapítás, mert a villamosenergia fogyasztási idősorokon teljesen más a szezonális különböző szintjeinek a súlya.

¹⁰⁴ A korábbi, hőmérsékleti véletlen hatás modellezésével foglalkozó 2.2. alfejezetben már foglalkoztunk hasonló kérdéssel.

3.2.4.2. Hierarchikus modell alapú klaszterezés

Az értekezésben használt 'mclust' R csomagban szereplő EM algoritmus esetén az alapértelmezett kezdő klaszterbesorolás az ún. hierarchikus EM algoritmus eredménye alapján történik. A hierarchikus modell alapú klaszterezés agglomeratív jellegű, tehát a kiindulásként minden megfigyelés egyedi klasztert képez. Az eljárás az ún. klasszifikációs *loglikelihood* (*classification loglikelihood*, röviden *cl*) maximalizálásán alapul, amely az alábbi módon írható fel:

$$cl(\psi|z_1, z_2 \dots z_n, y_1, y_2 \dots y_n) = \sum_{k=1}^K \sum_{\{i|z_i=k\}} \log(\tau_k f_k(y_i | \theta_k)),$$

ahol $\{i|z_i = k\}$ a k -adik komponensbe kerülő megfigyelések halmaza, együttese. Ennek az $\{i|z_i = k\}$ feltételnek a jelenléte miatt a bemutatott hagyományos EM algoritmus itt nem alkalmazható. A klaszterek lépésenkénti összevonása úgy történik, hogy a klasszifikációs *likelihood* értéke a lehető legnagyobb mértékben növekedjen minden egyes lépésben. Természetesen az utolsó lépésben minden megfigyelés egy klaszterbe kerül. A megoldás előnye, hogy a hierarchikus, agglomeratív megoldásból adódóan az összevonási lépések nem függenek a klaszterek számától.

3.2.4.3. A Gauss-féle keverék-modell diszkriminancia-analízisben történő alkalmazása

A keverék-modellek logikáját kihasználva készíthető diszkriminancia-analízis is, amely a klaszterezéssel ellentétben ún. *supervised* eljárás. Ebben az esetben ugyanis ismert csoportbesorolást modellezünk különböző (magyarázó)változók felhasználásával, míg a klaszterezés esetében csoportbesorolás nincs, hanem a valamilyen szempontból hasonlóknak ítélt megfigyeléseket soroljuk egy klaszterbe vagy csoportba (ún. *unsupervised* eljárás).

Keverék-modellek esetén több lehetőség is létezik erre (ld. *Fraley-Raftery [2000]*):

- *Eigenvalue Decomposition Discriminant Analysis* (EDDA), illetve
- *Mixture Discriminant Analysis*.

Előbbi csoportonkénti normális eloszlást feltételez, az utóbbi csoportonként megengedi a keverék-eloszlás illesztését.

Az értekezésben az EDDA diszkriminancia-analízis módszert használjuk. Ez lényegében nem más, mint egy (ún. diszkrét) M-lépés illetve E-lépés egymás utáni, egy-egy elvégzése. A diszkrét M-lépés elvégzése során a keverék-klaszterezés eredményé-

ből származó *posterior* valószínűségek közül megfigyelésenként a legnagyobbat 1-gyel tesszük egyenlővé, a többi 0-val. Az így kapott adathalmazra lefuttatunk egy M-lépést, majd a klaszterezésben részt nem vevő megfigyeléseket az E-lépésben hozzárendelhetjük ahhoz a klaszterhez, ahol a *posterior* valószínűség a legnagyobb.

3.2.5. A Gauss-féle keverék-modell regressziós alkalmazása (GMR)

A Gauss-féle keverék-regresszió Gauss-féle keverék-modellből történő származtatása úgy történik, hogy a változók közül kiválasztunk egyet, a későbbi eredményváltozót, és erre írunk fel regressziót a többi változó, mint magyarázóváltozók felhasználásával.

Az előző szakaszban használt jelölésrendszert a fejezetben némileg módosítjuk a regressziós alkalmazás végett:

- y_i az eredményváltozó értéke az i -edik megfigyelés esetén,
- x_i a magyarázóváltozók értékeit tartalmazó $(p \times 1)$ hosszúságú vektor az i -edik megfigyelés esetén ($i = 1, 2 \dots n$),
- n továbbra is a megfigyelések száma,
- p pedig a magyarázóváltozók száma.

Korábbiakban a változókat egységesen y_i -vel, a változók számát m -mel jelölük. Emellett, ahol indokolt, áttérünk a mátrixos jelölésre, a Gauss-féle keverék-regressziós szakirodalomban használt összefüggések bemutatásakor.

3.2.5.1. A Gauss-féle keverék-regresszió származtatása

A könnyebb érthetőség és a szakirodalmi eredményekkel való párhuzam megteremtése céljából érdemes az átlag és a kovariancia mátrix partícionálását a következők szerint elvégezni:

$$\mu^k = \begin{bmatrix} \mu_Y^k \\ \mu_X^k \end{bmatrix}, \text{ az alábbi dimenziókkal: } \begin{bmatrix} 1 \times 1 \\ p \times 1 \end{bmatrix},$$

illetve

$$\Sigma^k = \begin{bmatrix} \Sigma_{YY}^k & \Sigma_{YX}^k \\ \Sigma_{XY}^k & \Sigma_{XX}^k \end{bmatrix}, \text{ az alábbi dimenziókkal: } \begin{bmatrix} 1 \times 1 & 1 \times p \\ p \times 1 & p \times p \end{bmatrix},$$

ahol az átlagok illetve a kovariancia mátrixok egyaránt a p_{ik} *posterior* valószínűségekkel súlyozott átlagok illetve súlyozott kovariancia mátrixok, azaz:

$$\mu_Y^k = \frac{Y^T \text{diag}(W_k)}{\sum_{i=1}^n \text{diag}(W_k)}, \quad \mu_X^k = \frac{X^T \text{diag}(W_k)}{\sum_{i=1}^n \text{diag}(W_k)},$$

$$\begin{aligned}\Sigma_{XX}^k &= \frac{X^T W_k X}{\sum_{i=1}^n \text{diag}(W_k)}, & \Sigma_{XY}^k &= \frac{X^T W_k Y}{\sum_{i=1}^n \text{diag}(W_k)}, \\ \Sigma_{YX}^k &= \frac{Y^T W_k X}{\sum_{i=1}^n \text{diag}(W_k)}, & \Sigma_{YY}^k &= \frac{Y^T W_k Y}{\sum_{i=1}^n \text{diag}(W_k)}.\end{aligned}$$

ahol:

- X az x_i magyarázóváltozók $(n \times p)$ méretű mátrixa,
- W_k a p_{ik} súlyokat tartalmazó $(n \times n)$ méretű komponensenkénti diagonális mátrix,
- Y az y_i eredményváltozó értékeket tartalmazó $(n \times 1)$ méretű vektor,
- $k = 1 \dots K$ továbbra is a komponenseket jelöli.

Az $y_1, y_2 \dots y_n$ és $x_1, x_2 \dots x_n$ változókon a p_{ik} posterior valószínűségekkel, mint súlyokkal elvégzett súlyozott legkisebb négyzetek módszerével számolva, a komponensenként regressziós együtthatók az alábbiak:

$$\widehat{\beta}_k = (X^T W_k X)^{-1} X^T W_k Y,$$

vagy a particionált kovariancia mátrix felhasználásával más módon felírva is megkaphatóak:

$$\widehat{\beta}_k = \Sigma_{YX}^k (\Sigma_{XX}^k)^{-1}.$$

A komponensenkénti hibatag varianciája pedig az alábbi:

$$\widehat{\sigma}_k^2 = \frac{\sum_{i=1}^n p_{ik} (y_i - x_i^T \widehat{\beta}_k)^2}{\sum_{i=1}^n p_{ik}},$$

vagy mátrixos felírással:

$$\begin{aligned}\widehat{\sigma}_k^2 &= \frac{(Y - X \widehat{\beta}_k)^T W_k (Y - X \widehat{\beta}_k)}{\sum_{i=1}^n \text{diag}(W_k)} = \\ &= \frac{Y^T W_k Y - (\widehat{\beta}_k)^T X^T W_k Y - Y^T W_k X \widehat{\beta}_k + (\widehat{\beta}_k)^T X^T W_k X \widehat{\beta}_k}{\sum_{i=1}^n \text{diag}(W_k)} = \\ &= \frac{Y^T W_k Y - (\widehat{\beta}_k)^T X^T W_k Y - Y^T W_k X \widehat{\beta}_k + (\widehat{\beta}_k)^T X^T W_k X (X^T W_k X)^{-1} X^T W_k Y}{\sum_{i=1}^n \text{diag}(W_k)} = \\ &= \frac{Y^T W_k Y - (\widehat{\beta}_k)^T X^T W_k Y - Y^T W_k X \widehat{\beta}_k + (\widehat{\beta}_k)^T X^T W_k Y}{\sum_{i=1}^n \text{diag}(W_k)} = \\ &= \frac{Y^T W_k Y - Y^T W_k X \widehat{\beta}_k}{\sum_{i=1}^n \text{diag}(W_k)} = \frac{Y^T W_k Y - Y^T W_k X (X^T W_k X)^{-1} X^T W_k Y}{\sum_{i=1}^n \text{diag}(W_k)}\end{aligned}$$

amely a particionált kovariancia mátrixok felhasználásával felírva a következővel azonos:

$$\widehat{\sigma}_k^2 = \Sigma_{YY}^k - \Sigma_{YX}^k (\Sigma_{XX}^k)^{-1} \Sigma_{XY}^k.$$

Az eredmények $\widehat{\beta}_k$ és $\widehat{\sigma}_k^2$ tekintetében megegyeznek természetesen más munkák eredményeivel (ld. például *Srivastav et al.* [2013]), viszont itt a származtatásaik, levezetések is részletesen szerepelnek.¹⁰⁵

3.2.5.2. Az eredményváltozó feltételes várható értékének és szórásának számítása

A komponensek alapján az eredményváltozó feltételes várható értékének és feltételes szórásának (standard hibájának) számítása a komponensenkénti feltételes várható értékek és szórások alapján, a *posterior* valószínűségek, mint súlyok felhasználásával történik. Feltételezve tehát komponensenként a normális eloszlást, az y_i eredményváltozó feltételes sűrűségfüggvénye az alábbi módon írható fel *Srivastav et al.* [2013] alapján:

$$\Phi(y_i, \lambda(x_i)) = \sum_{k=1}^K p_{ik} \cdot \frac{1}{\sqrt{2\pi}s_{ik}} \cdot \exp\left(-\frac{1}{2}\left(\frac{y_i - m_{ik}}{s_{ik}}\right)^2\right),$$

ahol $\lambda(x_i) = \{p_{ik}, m_{ik}, s_{ik}\}$. Azaz, amint az a jelölésből is látható, a konkrét $\{p_{ik}, m_{ik}, s_{ik}\}$ paraméterek értéke függ az x_i magyarázóváltozók konkrét értékeitől.

A p_{ik} valószínűségek számítása már ismert:

$$p_{ik} = \frac{\tau_k \cdot f(y_i | \theta_k)}{\sum_{k=1}^K \tau_k \cdot f(y_i | \theta_k)}.$$

A többi paraméter értéke pedig az alábbi:

$$m_{ik} = \mu_Y^k + \Sigma_{YX}^k (\Sigma_{XX}^k)^{-1} (x_i - \mu_X^k),$$

illetve

$$s_{ik}^2 = \Sigma_{YY}^k - \Sigma_{YX}^k (\Sigma_{XX}^k)^{-1} \Sigma_{XY}^k.$$

¹⁰⁵ A kapott eredmények elsőre idegennek tűnhetnek, valójában a klasszikus többváltozós regresszióhoz képest mindössze a *posterior* valószínűségekkel történő súlyozás végett vannak különbségek. A klasszikus többváltozós regressziószámítás esetén ugyanis a becsült paraméterek képlete az itt is használt mátrixos felírással az alábbi:

$$\hat{\beta} = (X^T X)^{-1} X^T Y,$$

A becsült regressziós egyenlet:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip},$$

vagy másképpen:

$$\hat{y}_i = x_i^T \hat{\beta},$$

a hibatagok varianciája pedig:

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2}{n-p-1}.$$

A hasonlóság mostmár a klasszikus többváltozós regressziós illetve a keverék-regresszió eredményei között így nagyon könnyen felismerhető.

Könnyen látható, hogy az m_{ik} képlete nem más, mint a regressziós egyenletbe történő behelyettesítés, feltételezve, hogy nincsen konstans. Így a képletet a változók átlagtól vett eltéréseire írjuk fel, azaz:

$$m_{ik} - \mu_Y^k = \Sigma_{YX}^k (\Sigma_{XX}^k)^{-1} (x_i - \mu_X^k) = \widehat{\beta}_k (x_i - \mu_X^k),$$

Az általános regressziós terminológiából ismert, hogy amennyiben a regressziót konstans mellett becsüljük az eredeti változókkal, illetve konstans nélkül az eredeti változók átlagtól vett eltéréseivel, a $\widehat{\beta}_k$ becsült paraméterek ugyanazok lesznek.

Az s_{ik}^2 képlete pedig nem más, mint a komponensenkénti reziduális variancia.

A fentiek alapján könnyen továbbvezethető, hogy az eredményváltozó várható értéke és varianciája (azaz a standard hiba négyzete) az alábbiak szerint írható fel:

$$\hat{y}_i = \sum_{k=1}^K p_{ik} \cdot m_{ik},$$

és

$$\text{var}(\hat{y}_i) = \sum_{k=1}^K p_{ik} \cdot (s_{ik}^2 + m_{ik}^2) - (\sum_{k=1}^K p_{ik} \cdot m_{ik})^2.$$

Az eredményváltozó varianciájának származtatása során kétszer is kihasználtuk azt az ismert összefüggést, hogy a variancia a négyzetes átlag négyzetének és a számtani átlag négyzetének a különbsége.

3.2.5.3. Konfidencia-intervallum készítése az eredményváltozó értékére

A bemutatott keverék-regresszió esetében nincsen lehetőség a konfidencia-intervallum hagyományos módon történő számítására, ugyanis a mögöttes eloszlás a becsült keverék-eloszlás.¹⁰⁶ Az előző fejezetben bemutatott feltételes sűrűségfüggvény integrálja (Srivastav *et al.* [2013] alapján) az alábbi formában is felírható:

$$\Phi(y_i, \lambda(x_i)) = \sum_{k=1}^K \frac{p_{ik}}{2} \left(1 + \text{erf} \left(\frac{y_i - m_{ik}}{\sqrt{2} s_{ik}} \right) \right),$$

ahol $\text{erf}(\cdot)$ az ún. Gauss-féle hibafüggvényt (*error function*) jelenti. Ezt valószínűség-számítási és statisztikai területen használják, kapcsolódik a standard normális eloszlás eloszlásfüggvényéhez az alábbi módon:

$$\Phi(x) = \frac{1}{2} + \frac{1}{2} \text{erf} \left(\frac{x}{\sqrt{2}} \right) = \frac{1}{2} \left(1 + \text{erf} \left(\frac{x}{\sqrt{2}} \right) \right),$$

amely felírásból a fenti képlettel való kapcsolat már nagyon könnyen látható.

¹⁰⁶ Klasszikus többváltozós idősoros regressziós esetén a konfidencia-intervallum képlete – az előző lábjegyzetek jelöléseivel élve – az alábbi szimmetrikus intervallum: $\hat{y}_t \pm t_{1-\frac{\alpha}{2}}(n-p-1) \cdot \hat{\sigma}$, ahol az i index helyett t indexet használunk az idősoros tulajdonság miatt.

Az α megbízhatósági szintű konfidencia-intervallum alsó határának számolásához meg kell határoznunk, hogy $-\infty$ -tól integrálva mekkora y_L érték mellett lesz az integrál értéke $\frac{\alpha}{2}$, azaz azt az y_L értéket keressük, amely mellett teljesül az alábbi feltétel:

$$\frac{\alpha}{2} = \sum_{k=1}^K \frac{p_{ik}}{2} \left(1 + \operatorname{erf} \left(\frac{y_L - m_{ik}}{\sqrt{2}s_{ik}} \right) \right).$$

Az α megbízhatósági szintű konfidencia-intervallum felső határának számolásához meg kell határoznunk, hogy az y_U értéktől integrálva $+\infty$ -ig mekkora y_U érték mellett lesz az integrál értéke $\frac{\alpha}{2}$, azaz azt az y_U értéket keressük, amely mellett teljesül az alábbi feltétel:

$$1 - \frac{\alpha}{2} = \sum_{k=1}^K \frac{p_{ik}}{2} \left(1 + \operatorname{erf} \left(\frac{y_U - \hat{m}_{ik}}{\sqrt{2}\hat{s}_{ik}} \right) \right).$$

A fenti egyenletek megoldása történhet pl. a Newton-Raphson-módszerrel (ld. például *Srivastav et al.* [2013] munkájában is).

Az értekezésben ettől eltérő iteratív megoldást alkalmazva a konfidencia-intervallum felső határát úgy határoztuk meg, hogy a pontbecslésre, a pontbecsléshez képest egy annál jóval magasabb értékre, és a kettő átlagára is kiszámoltuk a keverék-eloszlás eloszlásfüggvényének az értékét. Amennyiben az átlaghoz tartozó érték nagyobb, mint $1 - \frac{\alpha}{2}$, a konfidencia-intervallum felső határa ezen átlag és a pontbecslés között van, így a leírt logikát erre az intervallumra ismételtük mindaddig, amíg az iterációnkénti átlag tekintetében az $1 - \frac{\alpha}{2}$ valószínűségtől csak minimális mértékben tértünk el. A megoldás a konfidencia-intervallum alsó határára hasonló módon elvégezhető. A leírt iteratív eljárás biztosan jó eredményhez konvergál, hiszen a keverék-eloszlás eloszlásfüggvénye – az eloszlásfüggvény tulajdonságából adódóan – monoton nem csökkenő.

3.2.6. A Gauss-féle keverék-regresszió alkalmazása idősorok esetén

A korábbi fejezetekben már említettük, hogy a Gauss-féle keverék-modellt nagyon sok esetben alkalmazzák idősorok esetén is, pl. beszéd- vagy hangfelismerésben. Ebben az esetben a feladat olyan módon válik többváltozósá, hogy a vizsgált idősor mellett a többi változót az idősor saját késleltetettjei jelentik.¹⁰⁷ *Eirola és Lendasse*

¹⁰⁷ A Gauss-féle keverék-modell regressziós alkalmazása esetén pl. beszéd- vagy hangfelismerés területén gyakorlatilag csak autoregresszív változókat használatára hagyatkoznak, tipikusan olyan konstrukcióban, hogy egy adott, t -edik időszaki jel mellé például a 2, 4, 6... időszakkal, vagy a 6, 12, 18... időszakkal megelőző késleltetésekkel képeznek több változót. Autoregresszív szó helyett azonban sokkal inkább az

[2013] áttekintik a *Gauss*-féle keverék-modell regressziós alkalmazását idősorok esetében, beleértve annak előrejelzésre illetve hiányzó adatok interpolálására történő alkalmazását.

Amennyiben adott tehát a $z = (z_1, z_2 \dots z_T)$, azaz T hosszúságú, stationer idősor, akkor képezhető az alábbi adatmátrix az idősor késleltetettjeinek felhasználásával:

$$y = \begin{bmatrix} z_1 & z_2 & \dots & z_d \\ z_2 & z_3 & \dots & z_{d+1} \\ \vdots & \vdots & & \vdots \\ z_{n-d+1} & z_{n-d+2} & \dots & z_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-d+1} \end{bmatrix},$$

ahol az y mátrix egy sora egy d elemű vektor. Azaz összességében d darab változónk van, ami a késleltetettekben áll össze, és a késleltetés végett a teljes rendelkezésre álló mintánk rövidül, a hossza $(n - d + 1)$ lesz. Ennek ismeretében a *Gauss*-féle keverék-modell a normális eloszlású sűrűségfüggvényekkel a következőképpen írható fel:

$$f(y) = \prod_{i=1}^{n-d+1} \prod_{k=1}^K \tau_k \cdot \varphi(y_i | \mu_k, \Sigma_k),$$

ahol $\varphi(y_i | \mu_k, \Sigma_k)$ a többváltozós normális eloszlás sűrűségfüggvénye, komponensenként μ_k átlaggal és Σ_k kovariancia mátrixszal, valamint τ_k keverék-súlyokkal. Innentől fogva a becslési eljárás azonos az előző fejezetekben bemutatottakkal, valamint a regressziós alkalmazás is ugyanúgy érvényes.

A késleltetés miatt elméletileg szükséges lehet az eredmények korrekciója, ugyanis a változók egymástól nem függetlenek. *Eirola és Lendasse* [2013] egy olyan megoldást javasol, ami az EM algoritmussal becsült paraméterek utólagos korrekcióját végzi el. Eredményeik alapján a korrekcióknak igazán nagy jelentősége azonban csak akkor van, ha sok komponenst becsülünk.¹⁰⁸

ún. beágyazott késleltetés (*delay embedding*) kifejezés elterjedt. Ezekről ld. bővebben *Shekofteh-Almasganj* [2013] illetve *Povinelli et al.* [2004].

¹⁰⁸ A korrekció szükségességének oka, hogy a késleltetett változók a késleltetéstől eltekintve megegyeznek az eredeti idősorral. Így követelmény, hogy valamennyi változónak a komponensekhez tartozó keverék-súlyokkal súlyozott átlaga (globális átlag) azonos legyen, a komponensekhez tartozó keverék-súlyokkal súlyozott kovariancia mátrixnak (globális kovariancia mátrix) pedig szimmetrikusnak illetve Toeplitz-típusúnak kell lennie. Ennek minden csökkenő diagonálisa azonos, azaz:

$$\Sigma = \begin{bmatrix} r(0) & r(1) & \dots & r(d) \\ r(1) & r(0) & \dots & r(d-1) \\ \vdots & \vdots & & \vdots \\ r(d) & r(d-1) & \dots & r(0) \end{bmatrix},$$

ahol $r(k)$ a z idősor k -adik késleltetés melletti autokovariancia együtthatóját jelöli. Ennek magyarázata praktikusán az, hogy például az idősor és első késleltetettje, valamint az idősor első és második késleltetettje stb. közötti autokovariancia értékének azonosnak kell lennie.

3.3. Keverék-modellek és energetikai alkalmazásaik

Még a saját empirikus kutatási eredményeket megelőlegezendő, érdemes a keverék-modellek néhány kedvező tulajdonságát kiemelni, amit más kutatások és az értekezés is hasznosítani kíván. Egyik kézenfekvő alkalmazási terület a terhelési görbéket jellemző tulajdonságok modell alapú leképezése, ahol az a mögöttes feltételezés, hogy a terhelési értékek eloszlása **többmódusú**, és a terhelési értékek alapvetően ezen módusok, mint tipikus értékek köré szerveződnek, sűrűsödnek. Az ún. *Gauss*-féle keverék-modell esetén ennek értelmezésben is praktikus hasznát lehet venni. Másik hasznos alkalmazási terület, amikor olyan jelenséget kívánunk vizsgálni, aminek az eloszlása **nem írható le valamilyen klasszikus eloszlás** típussal. Ez jelentheti akár azt is, hogy több módus van, vagy heteroszkedaszticitás (időben változó szóródás) van jelen.

Ebben a fejezetben azt tekintjük tehát át, hogy a keverék-modelleket eddig milyen területeken használták az energetikai modellezések során, hiszen az értekezés is az ilyen gyakorlati alkalmazásokhoz kíván hozzájárulni.

3.3.1. Napi fogyasztási görbék tipizálása

Többek között annak is köszönhetően, hogy az utóbbi időkben egyre inkább rendelkezésre állnak egyedi fogyasztókról is nagy megfigyelési gyakorisággal mérési adatok, a tipikus fogyasztási mintázatok feltárása folyamatosan jelenlévő és vizsgált téma, és a *Gauss*-féle keverék-modellek alkalmazása több kutatásban is megjelent már.

Az alkalmazások egy jelentősen hányada a korábban áttekintett módszerekhez hasonlóan alapvetően napi bontású fogyasztási görbék klaszterezésére épít. Mivel az értekezés saját empirikus eredményeinek megértése szempontjából releváns, ezért ezek közül néhányat részletesebben is bemutatunk; a saját empirikus eredmények és újdonság-tartalom kihangsúlyozása céljából ezeknek a munkáknak bizonyos eredményeit saját adatokon is reprodukáljuk.

McKenna et al. [2014] munkájában tipikus fogyasztási mintázatok feltárása történik meg okos mérőkkel mért vízfogyasztási adatokon. A keverék-modell itt úgy jelenik meg, hogy egy (átlagos) napi fogyasztási görbére úgy tekintenek, mint egy sűrűségfüggvényre, ami normális eloszlású sűrűségfüggvények keverékeként áll elő. Természetesen a sűrűségfüggvénnyel szembeni követelmény, hogy integrálja – azaz a sűrűségfüggvény alatti terület nagysága – 1 (100%) legyen. Így a keverék-sűrűségfüggvény becslése az alak rögzítésére jó. A végső eredményhez szükséges a sűrűségfüggvény helyettesítési értékeinek átskálázása úgy, hogy az átskálázást követően a görbe alatti

terület a napi össz fogyasztással legyen egyenlő. *McKenna et al.* [2014] munkájában a keverék-eloszlás paramétereinek (átlag, szórás illetve keverék-komponensek aránya) felhasználásával maguk a fogyasztók is klaszterezhetők, csoportosíthatók, ami itt a hagyományos K-középpontú (K-Means) klaszterezéssel történik meg, abból adódóan egyébként elég nehézkesen, hogy a paraméterek speciális (keverék-eloszlás) tulajdonságát nem használják ki.

Fontos felismerése még a tanulmánynak, hogy az órás bontású idősor helyett sokkal jobb, stabilabb eredmények kaphatóak félórás interpolált adatokkal (ún. *Hermit*-féle interpolációs polinom¹⁰⁹ alkalmazása), ami mindenképpen ígéretes és biztató a nagy frekvenciájú mérési adatok elemzése szempontjából; habár az interpoláció egyben az idősor simítását is jelenti, ami eredeti félórás mérésekkel szemben bizonyos tendenciákat elmoshat, kisíthat. Összességében tekintve azonban a megoldás egyik hátránya, hogy a keverék-sűrűségfüggvény illesztése napi átlagokon történik meg, nem pedig napi adatokra külön-külön. Utóbbi esetben a fogyasztónkénti tipikus napi fogyasztási görbe származtatását is külön meg kellene oldani (például úgy, hogy fogyasztónként klaszterezéssel napi profilokat állítunk elő, majd ezeknek a napi profiloknak a felhasználásával csoportosítjuk a fogyasztókat – itt azonban kezelni kellene például azt, ha a napi profilok száma fogyasztónként eltér, azaz lényegében a szerzők által alkalmazott napi átlagokon történő modellezés ennek a problémának egy elegáns kikerülése).

Hino et al. [2013] hasonlóan napi fogyasztási görbék klaszterezését végzi el, mind idősoros (egy fogyasztó napi fogyasztási görbéinek csoportosítása), mind keresztmetszeti (több fogyasztó ugyanazon napi fogyasztási görbéinek csoportosítása) dimenzióban. Munkájukban egy napi fogyasztási görbére ők is úgy tekintenek tehát, mint egy sűrűségfüggvényre, ami normális eloszlású sűrűségfüggvények keverékeként áll elő.

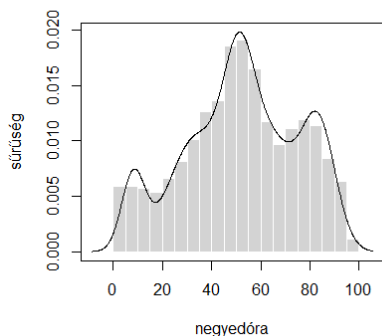
A napok klaszterezése azonban már jóval inkább szofisztikált módszertannal történik meg, hierarchikus klaszterezés alkalmazása mellett a szerzők által továbbvezetett ún. szimmetrikus általánosított Kullback-Leibler-(divergencia-) távolságmérték felhasználásával, amelyet alapvetően eloszlások különbözőségének mérésére használnak.¹¹⁰ A tanulmány másik nagy előnye az, hogy a mérési adatok másodpercesek, ami a sűrűségfüggvények illesztését különösen praktikussá teszi (ugyanis egy nap $24 \cdot 60 \cdot 60 =$

¹⁰⁹A *Hermit*-féle interpolációról ld. bővebben: https://en.wikipedia.org/wiki/Hermite_interpolation.

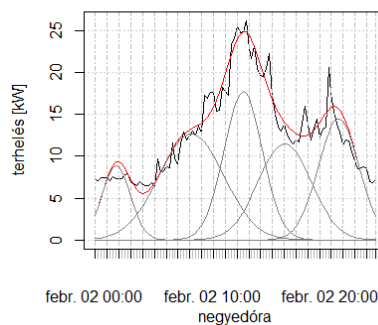
¹¹⁰ Normális eloszlás esetén viszonylag egyszerű, zárt képlete van, normális eloszlások keverékéből előállított eloszlások esetén pedig közelíthető az értéke (ld. *Hershey-Olsen* [2007]), ld. még a saját empirikus kutatási eredményeket tartalmazó fejezet is.

86400 másodpercből áll). A következő ábrák a dolgozatban is használt néhány görbe esetében mutatják be a napi görbékre illesztett keverék-sűrűségfüggvényeket.

32. ábra: Keverék-sűrűségfüggvény illesztése napi terhelési görbére (V109 görbe)



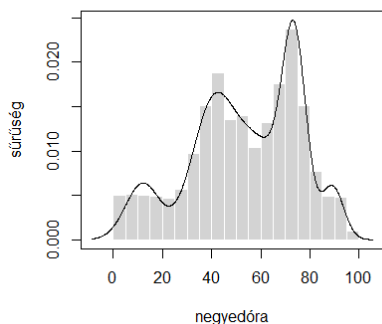
a) Becsült keverék-sűrűségfüggvény és negyedórás terhelési értékek alapján készített hisztogram.



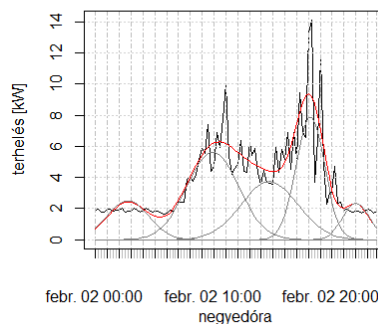
b) Átskálázott keverék-sűrűségfüggvény és a normális eloszlású komponensek, illetve a negyedórás terhelési görbe.

Forrás: saját számítási eredmények (R) és saját ábra (R).

33. ábra: Keverék-sűrűségfüggvény illesztése napi terhelési görbére (V148 görbe)



a) Becsült keverék-sűrűségfüggvény és negyedórás terhelési értékek alapján készített hisztogram.



b) Átskálázott keverék-sűrűségfüggvény és a normális eloszlású komponensek, illetve a negyedórás terhelési görbe.

Forrás: saját számítási eredmények (R) és saját ábra (R).

A megoldás tehát nagyon jó eszköz többek között annak vizsgálatára, hogy, mikor vannak a **napon belüli csúcsok**, illetve hosszabb távon vizsgálódva azok hogyan tolódnak el, hogyan változnak. Ez nemcsak egyedi fogyasztók szintjén, de aggregált, akár országos fogyasztás szintjén is érdekes. A csúcsok eltolódásának vizsgálata egyébként a hosszabb távú fogyasztás-megtakarítást, fogyasztás-átcsoportosítást célzó kutatások egyik fontos kérdése is, így az ittenieken túlmutató, inkább **stratégiai** szempontú kutatási terület.

3.3.2. Fogyasztás eloszlásának modellezése keverék-sűrűségfüggvénnyel

McKenna et al. [2014] illetve *Hino et al.* [2013] munkáiban az alapötlet az volt, hogy a keverék-sűrűségfüggvény a napi fogyasztási karakterisztikát írja le, mégpedig úgy, hogy amikor több energia fogy el ugyanannyi (egységnyi) idő alatt, akkor nagyobb a sűrűségfüggvény helyettesítési értéke, míg máskor alacsonyabb (azaz egységnyi energia elfogyasztásának az esélye (*likelihood*-ja) bizonyos időszakokban nagyobb, míg más időszakokban alacsonyabb). *Singh et al.* [2010] munkájában félórás fogyasztási adatok eloszlásának modellezése történik meg keverék-sűrűségfüggvények becslése által, azaz nem annak a modellezése, hogy egységnyi fogyasztás megtörténte **mikor** a legesélyesebb, hanem azé, hogy melyek a legjellemzőbb (legesélyesebb) **fogyasztási szintek**, ami körül a félórás fogyasztások tömörülnek, klasztereződnek.¹¹¹

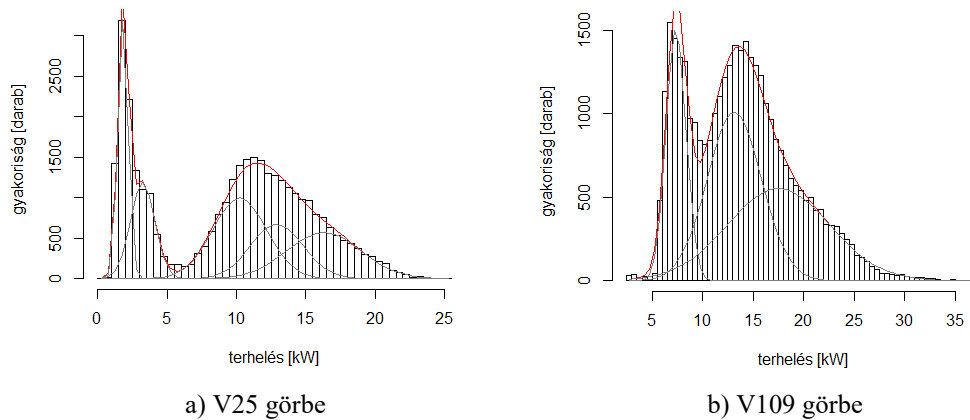
A következő ábrákon néhány fogyasztói görbe negyedórás terhelési adataiból összeállított hisztogram, illetve az illesztett keverék-sűrűségfüggvények láthatóak, beleértve a normális eloszlású komponenseket is. A V25 vállalat esetében alapvetően a hétvégi alacsonyabb illetve a hétközi magasabb fogyasztási szintek a tipikusak, összességében azonban öt komponenst becsültünk. Hasonló a helyzet a V109 esetében, ott feltehetően a két magasabb fogyasztásokkal jellemezhető komponens a nyári (hőmérsékletfüggő) illetve nem nyári (nem hőmérsékletfüggő) fogyasztási szinteket jelöli.

Sokkal előremutatóbbnak és transzparensnek tűnik azonban az elemzések többváltozós kiterjesztése (pl. hőmérséklet figyelembe vétele). Ilyen megoldásra mutatunk példát még ebben a fejezetben, illetve maga a dolgozat is hasonló utat követ. A többváltozós kiterjesztés iránti igény megjelenik egyébként *Hino et al.* [2013] esetében is (ld. az előző szakaszt), ott azonban ennek a plusz információnak a beépítése valamilyen többlépcsős megoldás keretében képzelhető el.

Láttuk tehát, hogy egyik fenti munkában sem jelent meg még a fogyasztás és az arra ható különböző tényezők, az időjárás vagy például a napszakok hatásának együttes vizsgálata, ezeknek a felhasználása a keverék-modellezés során. A következőkben erre mutatunk egy példát.

¹¹¹ *Singh et al.* [2010] munkájában jelenik meg emellett két fogyasztási idősor együttes eloszlásának a modellezése is keverék-modell segítségével, ami a portfólióhatás vizsgálatához jelenthet majd a jövőben fontos alapokat.

34. ábra: Keverék-sűrűségfüggvény illesztése terhelési értékek empirikus eloszlására

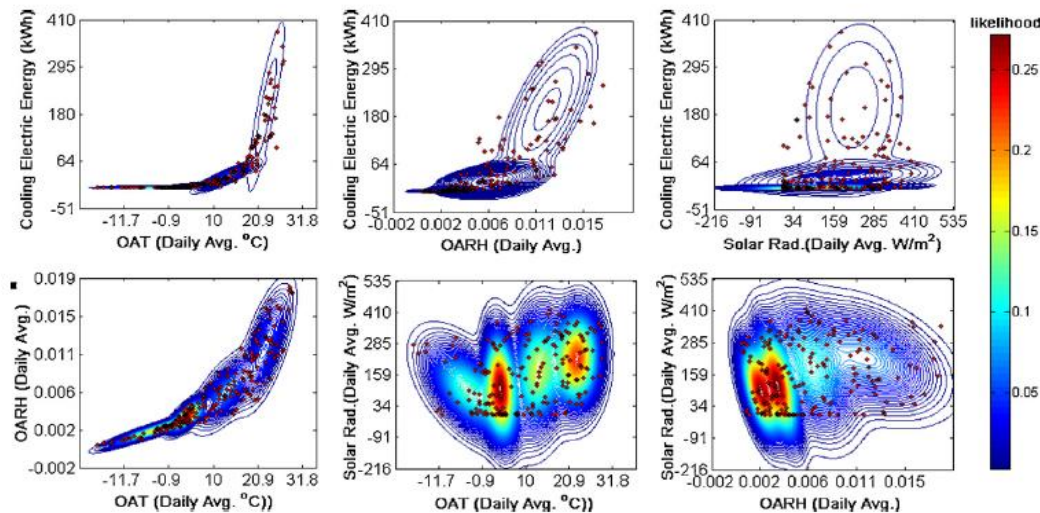


Forrás: saját számítási eredmények (R) és saját ábra (R).

3.3.3. Fogyasztás eloszlásának modellezése keverék-sűrűségfüggvénnyel és regressziós alkalmazással

A keverék-regresszió alkalmazására viszonylag kevés példa van, ilyen *Srivastav et al.* [2013] munkája. A tanulmányban bemutatják, hogy épületi hűtési (*cooling*) energia hogyan modellezhető keverék-modell keretében a hőmérséklet, páratartalom és nap-sugárzás változók felhasználásával, mind napi, mind órás bontású idősorok esetében.

35. ábra: Keverék-modell becslése és változópáronkénti illeszkedés ellenőrzése



Forrás: Srivastav et al. [2013]

Cikkükben azonban a keverék-modell alkalmazásának előnye nem elsősorban abban jelenik meg, hogy a kialakult komponensek móduszai felfoghatók úgy, mint tipikus fogyasztások, hanem a keverék-regresszió alkalmazása révén olyan, tulajdonképpen lokalizált konfidencia-intervallumok számíthatóak, amelyek az aktuális magyarázóvál-

tozók függvényében írhatóak fel. A szerzők alkalmazási területként a villamosenergia fogyasztás-megtakarítások árazásával kapcsolatos lehetőségeket jelölik meg, hiszen amennyiben a fogyasztás kockázata szezonálisan változik, úgy az abból a fogyasztásból realizálható megtakarítás mennyiségének a kockázata is változik.

Visszatérve a komponensek móduszaira, a megoldás hátránya (vagy inkább nem feltétlenül előnye), hogy egy-egy komponens (vagy klaszter) tipikus értéke – a többdimenziós normális eloszlás átlagvektora – nem egy tipikus napi lefutatót ad eredményül, mint ami a profilozással, a tipikus fogyasztási mintázatokkal foglalkozó szakirodalomban és gyakorlati alkalmazásokban megszokott. **Előnye** ugyanakkor, hogy a megbízhatósági intervallumokat lehet közvetlenül származtatni (közvetett megoldásra is van példa, a tapasztalatok alapján azonban ez nem olyan jó megoldás, ld. bővebben *Subbarao et al.* [2011] ún. kNN-módszerét röviden a 4.2. fejezetben), illetve viszonylag gyors (szemben az ún. *Gauss*-folyamat (*Gaussian Process*) alapú megoldásokkal, ld. például *Heo-Zavala* [2012], *Manfren et al.* [2013]).

Technikailag a változók komponensenkénti többdimenziós normalitásának feltételezése végett a regresszióban az eredményváltozónak választott változó konfidencia-intervalluma is normális eloszlások keverékéből származtatható, zárt képlet nincsen rá. A munkájuk így kapcsolódik a keverék-modellek olyan alkalmazási területeihez is, amikor a modellezésnek az a célja, hogy **feloldja** azt a feltételezést, miszerint a modell (regresszió) **hibatagjai független, azonos (normális)** eloszlású véletlen változók. A gyakorlatban energetikai fogyasztási idősorok esetében az időben változó szóródás, a heteroszkedaszticitás a legszembetűnőbb példája annak – a normalitás hiánya mellett (utóbbi árfolyam idősorok esetében még kevésbé teljesül) – hogy az említett FAE tulajdonság nem teljesül.

A tanulmány eredményeinek a tárgyalásakor a szerzők művében némileg elveszik, de a keverék-modell többváltozós alkalmazásának a mögöttes feltételezése, lényege éppen az, hogy az adott változók dimenzióiban mért megfigyelések klaszterezhetők, csoportosíthatók, és **klaszterenként**, csoportonként a megfigyeléseket **más-más kovariancia-struktúra**, szóródás jellemzi (amelynek következményei a regressziós illesztés nyomán természetesen majd a **hibatagokban** érhetőek tetten).

4. A FOGYASZTÁS BIZONYTALANSÁGÁNAK FIGYELEMBE VÉTELE A PROFILOZÁSBAN – EMPIRIKUS EREDMÉNYEK

A gyakorlatban a profilozással szemben egyre inkább követelmény, hogy a tipikus, várható fogyasztási karakterisztika mellett annak bizonytalanságát is képes legyen jól leírni, ezért a módszertan megválasztásánál ezt is szem előtt tartottuk.

A fogyasztási idősorok stilizált tényeinek vizsgálatánál sejthető volt, hogy a bizonytalanság valamilyen mértékben összefügg olyan változókkal, amelyekkel maguknak az idősoroknak a karakterisztikája is összefügg (például hőmérséklet, különböző szintű szezonális, stb.).

A fejezetben bemutatjuk a profilozásnak egy fenti követelményeknek is megfelelő, várható értéket és szóródást együttesen figyelembe vevő megvalósítási módját, beleértve azt is, hogy így milyen többletinformációk kinyerésére nyílik lehetőség. Az eredményeket klasszikusnak tekinthető modellekkel összehasonlítva tárgyaljuk, ahol a bizonytalanság figyelembevétele alapvetően nem cél, vagy háttérbe szorul és ennél fogva nehezen származtatható.

4.1. Tipikus fogyasztási mintázatok készítése

Mint azt láttuk, a profilozási szakirodalom – függetlenül attól, hogy csak egy-egy fogyasztói görbére vagy hasonló görbék csoportjára fókuszál – törekszik végeredményként a tipikus napi (fogyasztási) profilok származtatására (ez történhet valamilyen, főként napi diszkretizáláson alapuló klaszterezéssel, vagy különböző változók hatásának kiszűrésén alapuló regressziós módszertannal is). Ennek vitathatatlan előnye az egyszerűség és a könnyű interpretáció. Az értekezés önálló eredményeit hangsúlyozandó, az összehasonlítási alap (*benchmark*) egy regressziós logikát használó megoldás lesz.

Mivel a keverék-modell outputja közvetlenül nem összevethető a klasszikus profilozási eljárások outputjával, ezért az összehasonlítás sokkal inkább kvalitatív jellegű lehet, kiemelve a különböző módszerek előnyeit illetve hátrányait. Egyben a fejezet címét („tipikus fogyasztási mintázatok” kifejezés) is meg kívánjuk ezzel magyarázni, ugyanis módszertantól függően a **tipikus fogyasztási mintázat** lehet valamilyen napi profil, de lehet például az itt bemutatásra kerülő keverék-modell paraméterszettje is (*prior* vagy keverék-valószínűségek illetve átlagok és kovariancia mátrixok együtte-

se).¹¹² Az 2.1.3. fejezetben is használt **görbetulajdonság** (mint a görbére vonatkozó tömörített információ) lényegében a tipikus fogyasztási mintázatnak a statisztikai, módszertani megfelelője, amelyet sokszor – kicsit félrevezetően – profilnak is szoktunk nevezni. Ennek alapvetően az lehet az oka, hogy a görbetulajdonság nagyon sok esetben egy (jellemzően különféle hatásoktól szűrt) tipikus napi profilgörbének nevezett görbe formájában jelenik meg.

4.1.1. Keverék-modell alkalmazása tipikus fogyasztási mintázat készítésére

A keverék-modell profilozási célú alkalmazását bemutatjuk részletesen a vállalati portfólión, majd a fontosabb eredmények kihangsúlyozásával egyedi görbék esetén is. Mindezt a számítások során használt módszertan lépéseinek a bemutatása előzi meg.

4.1.1.1. Az alkalmazott módszertan leírása

A keverék-modell esetén a negyedórák klaszterezése történik meg, amely minden egyes, egy éves görbe esetén 35040 darab negyedóra ($365 \text{ nap} * 96 \text{ negyedóra/nap}$), azaz megfigyelés kezelését jelenti. Mivel a keverék-modellek egyik hátránya, hogy sok megfigyelés esetén lassabban működnek (a célfüggvény optimalizálásának magas futási ideje miatt), illetve a kezdő klasztertagságok kiszámításához használt hierarchikus modell alapú klaszterezést sem olyan praktikus sok megfigyelés mellett használni (mint ahogy a hierarchikus klaszterező eljárásokat általában sem), ezért a keverék-klaszterezést a dolgozatban keverék-diszkriminancia-elemzéssel kiegészítve használjuk (EDDA-módszer, ld. például *Banfield-Raftery* [1993]). A fontosabb lépések az alábbiak szerint összegezhetőek:

1. A teljes 2011-es évből vettünk egy véletlen mintát, ez $35040 / 8 = 4380$ negyedórát jelent.
2. Ennek a 20%-ára, azaz $4380 / 5 = 876$ negyedórára végeztünk el a klaszterezést, korlátozatlan kovariancia mátrixot¹¹³ feltételezve.¹¹⁴
3. A klaszterezésre épülően az EDDA-módszert az alábbiak szerint végeztük el:

¹¹² A fejezet címében a profilozás helyett azért szerepel inkább a tipikus fogyasztási mintázatok kifejezés, mert a profilozással foglalkozó munkák jelentős részénél a profil valamilyen tipikus napi lefutás formájában jelenik meg. A dolgozatban alkalmazott megoldás esetén a tipikus, jellemző értékek a komponensek paramétereiként közvetlenül adódnak, viszont – mint az alfejezet végén látni fogjuk – ugyanúgy jól interpretálhatóak.

¹¹³ A megvizsgált szakirodalomban is jellemzően ezt használják, ld. a 3. fejezetet.

¹¹⁴ Mind a teljes év 8-ad részének, mind a 20%-nak a választása szubjektív, de tapasztalati úton történt meg, több variáció kipróbálását követően.

- a. A 2. lépésben kiszámolt *posterior* valószínűségek alapján képzett valószínűségi besorolást diszkretizáltuk, azaz egy-egy negyedórát abba a klaszterbe soroltunk, amelyik komponens esetén a *posterior* valószínűség maximális, és ezt a maximális *posterior* valószínűséget 1-gyel (100%-kal), a többit pedig 0-val tettük egyenlővé.

Ezen diszkretizált besorolás, illetve a magyarázóváltozók felhasználásával elvégeztük a diszkriminancia-elemzést, ami egy egyszeri M-lépés és egy egyszeri E-lépés elvégzéséből áll.

- b. Az M-lépés eredményeként megkaptuk a becsült paramétereket (*prior*- vagy keverék-valószínűségek illetve átlagok és kovariancia mátrixok),
- c. az E-lépésben pedig a becsült paramétereket felhasználva a klasztertagok *posterior* valószínűségeit is.

Az M-lépés csak a klaszterezés során használt negyedórákat veszi figyelembe, az E-lépés viszont már a teljes véletlen mintát (4380 darab negyedóra). Természetesen az E-lépésben akár az összes (azaz 35040 darab) negyedóra is besorolható.¹¹⁵

A megoldás előnye alapvetően a gyorsaság és átláthatóság, de ezzel párhuzamosan kihasználja a modell alapú klaszterezésnek azt a kedvező tulajdonságát is, hogy az alapvetően a struktúra felismerését célozza, így a kisebb mintaelemszám nem jelent problémát. Ez különösen hasznos többek között abból a szempontból, hogy görbénként csak 1 évnyi adat áll rendelkezésre, így nagyon kicsi például az extrém hideg vagy meleg napok száma, amikor a hőmérséklet-terhelés kapcsolat más lehet, mint tipikus, kevésbé extrém téli vagy nyári napokon. A véletlen mintavétel előnye még, hogy így eljárva az elemzésbe bevont negyedórák tekinthetők egymástól függetlennek. A keverékmodell becslése során ugyanis feltételezzük a függetlenséget, ami szinte soha, egyetlen egy gyakorlati alkalmazás esetén sem állja meg a helyét, de a véletlen mintavétel által a feladat így jobban megfelel a modellfeltevéseknek.

A negyedórás terhelési idősorok mellett az alábbi változókat használtuk fel a negyedórák klaszterezése során:

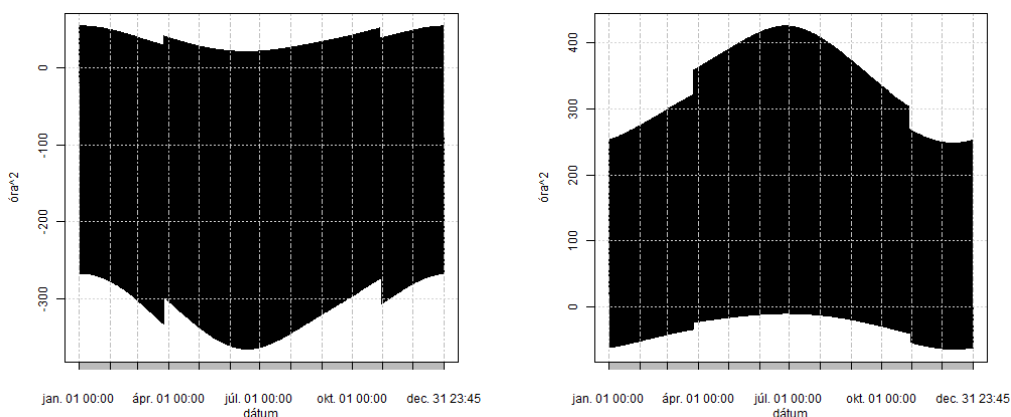
1. napi középhőmérséklet adatok,

¹¹⁵ Ha nem használjuk fel az összes negyedórát, akkor a fel nem használt negyedórák például mintán kívüli értékelésekhez használhatóak (ld. később a regressziós alkalmazással foglalkozó fejezetet).

2. az adott negyedóra esetében napon belül vizsgálva a napfelkeltétől, illetve a napnyugtától számított előjeles, percben mért négyzetes eltérések,
3. egy időszakos (azaz egy negyedórás), illetve egy napos (azaz 96 negyedórás) késleltetések.

A napi középhőmérséklet a téli illetve nyári fűtési hatás kezelése miatt fontos, egyben a(z éves) szezonális kezelésére használt *proxy* változóként is tekinthetünk rá. Az egy napos késleltetés alkalmas a heti szezonális kezelésére (gondoljunk a péntek-szombati vagy vasárnap-hétfői napok egymástól egy napnyi (azaz 96 negyedórnyi) távolságra lévő negyedóráira). A napfelkeltétől illetve napnyugtától vett négyzetes távolság lényege, hogy a napon belüli szezonális (időbeliséget) helyettesítik és a csoportképzést ebben a dimenzióban támogatják majd. A négyzetre emelést követően az előjel megőrzése a napon belüli elhelyezkedés miatt számít. Az utóbbi változók éves lefutásában nagyon szépen visszaköszön a napok hosszának a tendenciája; nem elsősorban a változó konkrét értékei számítanak, hanem sokkal inkább az „alak”, ami nagyban emlékeztethet az 1.4. alfejezet szintvonalábráin látott ún. naplemente-hatás éves lefutására.¹¹⁶ A négyzetre emelésnek köszönhetően a napfelkelte változó a nap végét, a naplemente változó a nap elejét „húzza szét”, aminek a klaszterezésnél lesz majd haszna (ld. később).

36. ábra: Napfelkelte illetve naplemente időpontjától számított előjeles, négyzetes eltérések alakulása



Forrás: saját készítésű ábra (R).

A keverék-modell alkalmazása során létrejött komponensek nem feltétlenül kell, hogy önálló csoportot alkossanak. Gyakran előfordul, hogy az egymáshoz hasonló,

¹¹⁶ Hasonló konstrukcióval a 4.1.2. szakaszban is fogunk találkozni, a klasszikus megoldások ismertetésénél.

egymáshoz közel eső komponenseket egy csoportnak tekintjük. Mivel egy komponenst egy normális eloszlás sűrűségfüggvénye reprezentál, a komponensek távolságának a mérésére praktikusán használható az ún. Kullback-Leibler-divergenciamérték (vagy -távolságmérték), amely valószínűség-eloszlások közötti távolság mérésére szolgál. Amennyiben $p_i(x)$ és $p_j(x)$ két eloszlásfüggvény, ennek definíciója az alábbi (ld. például Cover-Thomas [1991]):

$$D(p_i, p_j) = \int_{-\infty}^{\infty} p_i(x) \log \frac{p_i(x)}{p_j(x)} dx.$$

Amennyiben a két eloszlásfüggvény normális eloszlás, μ_i és μ_j várható értékekkel, valamint σ_i és σ_j szórásokkal, a képlet az alábbi zárt formára egyszerűsödik:

$$D(p_i, p_j) = \frac{1}{2} \left[\log \left(\frac{\sigma_j^2}{\sigma_i^2} \right) + \frac{\sigma_i^2}{\sigma_j^2} - 1 + \frac{(\mu_i - \mu_j)^2}{\sigma_j^2} \right].$$

A Kullback-Leibler-távolságmérték – mivel entrópia mutató¹¹⁷ – nem szimmetrikus¹¹⁸, ezért mind a $D(p_i, p_j)$ és a $D(p_j, p_i)$ értékeket ki szokás számolni, és távolságmértékként ezek átlagát használni, azaz:

$$D_{KL}(p_i, p_j) = \frac{D(p_i, p_j) + D(p_j, p_i)}{2},$$

ahol $D_{KL}(p_i, p_j)$ jelöli a fejezet hátralévő részében is használatos értéket.

4.1.1.2. *Eredmények a vállalati portfólió terhelési görbéje esetén*

A korábban ismertetett változók felhasználásával készült eredményeket *scatter plot*-ok tekintetében látjuk a 37. ábrán a vállalati portfólió esetében. A különböző színű pontfelhők különböző komponenseket jelölnek. Az ellipszisek középpontjában a komponensek átlaga szerepel, az ellipszisek pedig az átlag \pm szórás értéknek megfelelő vonalat rajzolják ki a változópaárak tekintetében – ugyanúgy, ahogy azt a 3.2.3. fejezetben láttuk Budapest napi földgázfogyasztása példáján.

A 37. b) ábrán látható, hogy a módszerrel a terhelés alakulásában általában eltérő hétfői illetve pénteki napok hogyan különülnek el (rendre az y és x tengelyekkel majdnem párhuzamos pontfelhőkről van szó), a c) és d) ábrákon pedig a napszakok klasztereződése látható. Ezeknél a jobb alsó sarokban szereplő piros illetve kék pontfelhők a

¹¹⁷ Szokás relatív entrópia néven is emlegetni, ld. például <http://mathworld.wolfram.com/RelativeEntropy.html>.

¹¹⁸ A hasonlósági mértékekkel vagy távolságmértékekkel szemben szokás megfogalmazni olyan követelményt, hogy azok szimmetrikusak legyenek, tehát a hasonlóság vagy távolság nagysága ne függjön a sorrendtől (azaz A távolsága B-től ugyanakkora legyen, mint B távolsága A-tól). Bonyolultabb, komplexebb mértékek esetén ez sokszor nem teljesül, ilyenkor szokás a hasonlóságot vagy távolságot mindkét irányban kiszámolni és azok átlagával dolgozni.

kora reggeli órákat jelölik, és míg a c) ábrán nem különülnek el abban a két dimenzióban olyan jól, addig a d) ábrán már igen, ezért hasznos szerepeltetni mind a napfelkeltétől, mind a naplementétől vett eltérés változókat a klaszterezés során. Az e) ábrával összevetve nemcsak a még jobb elkülönülésük látszik, de az is, hogy az egyik a nyári, a másik a téli reggeli negyedórát jelöli, és nagyon jól körvonalazódik a terhelés hőmérsékletfüggése is, különösen a nappali órákban. Az ellipszisek elhelyezkedését már az 1.4. fejezetben vizsgált *scatter plot*-ok alapján is prognosztizálhattuk, ott azonban még csak a hőmérséklet és a terhelés változókat vettük figyelembe.

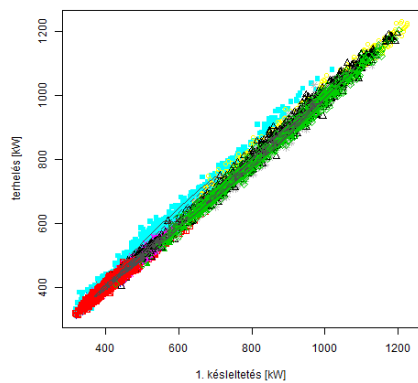
Bár a *scatter plot*-okból sejthető, meg lehet megvizsgálni, hogy egy-egy komponensbe melyik negyedórát kerültek. A 2-3-8. komponensekben főleg reggeli, a 4-9-10-11. komponensekben elsősorban esti, a többi komponensben inkább csúcsidőszaki órák vannak. Hasonlóan megvizsgálható, hogy a hétköznapiak illetve a hétvégék leírásában melyik komponensnek van nagyobb illetve kisebb szerepe. A 38. ábra az órák, a hónapok illetve a hétköznapiak és hétvégék komponensek közötti megoszlását mutatja.¹¹⁹

Csak néhány komponenst kiemelve még, a 6. komponens például a nyári hétköznapi csúcsidőszaki negyedórát jelöli, amikor a hűtési hatás miatt nagy a légkondicionáló készülékek használata. A 9. illetve 10. komponensek pedig a téli és a nyári hónapok esti (részben lecsengési) időszakainak negyedórát jelölik. Érdekes – de abszolút logikus – eredmény például az is, hogy a reggeli felfutási időszakokat tömörítő 3. komponensnek a hétfői negyedórákban jóval kisebb aránya van, mint hétköznapiakon. Természetesen a komponensek megoszlása nem ilyen „tisztá”, hiszen egy negyedórát abba a komponensbe sorolunk, ahol a *posterior* valószínűség a legnagyobb, márpedig a legtöbb negyedóra nem egyetlen egy komponens realizációjaként áll elő.

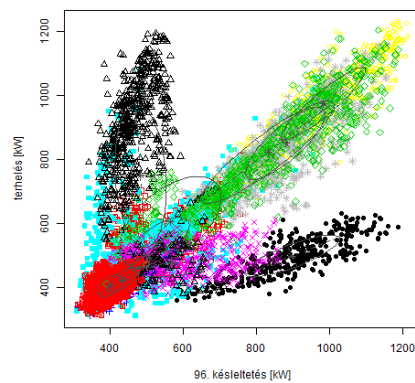
Mindezekhez szorosan kapcsolódva, a 38. ábrán a komponensek közötti Kullback-Leibler-távolságmérték alapján készített dendrogram látható. Az összevonás Ward-féle összevonási logika alapján készült (ld. még később).

¹¹⁹ Természetesen az itt szereplők mellett még a többi változópárra is el lehet készíteni a *scatter plot*-okat, ami kevésbé látványos és hasznos. Az itteni ábrákon a változópárok egyike mindig terhelés, a regressziós alkalmazásra gondolva a majdani eredményváltozó.

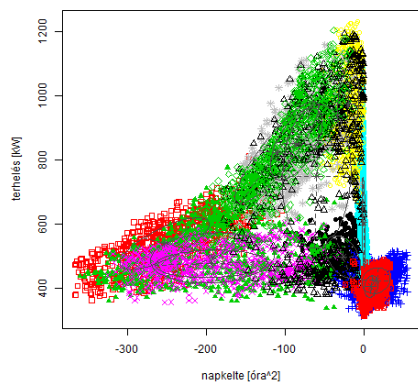
37. ábra: Keverék-modell becslése és változópáronkénti illeszkedés ellenőrzése a vállalati portfólió esetén



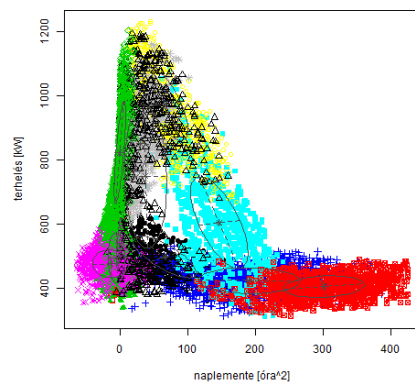
a) Terhelés és terhelés 1. késleltetettje



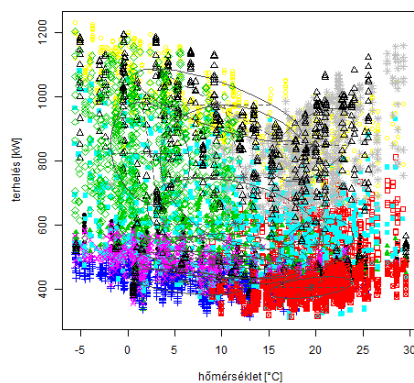
b) Terhelés és terhelés 96. késleltetettje



c) Terhelés és napfelkelte időpontjától számított négyzetes eltérés



d) Terhelés és napnyugta időpontjától számított négyzetes eltérés

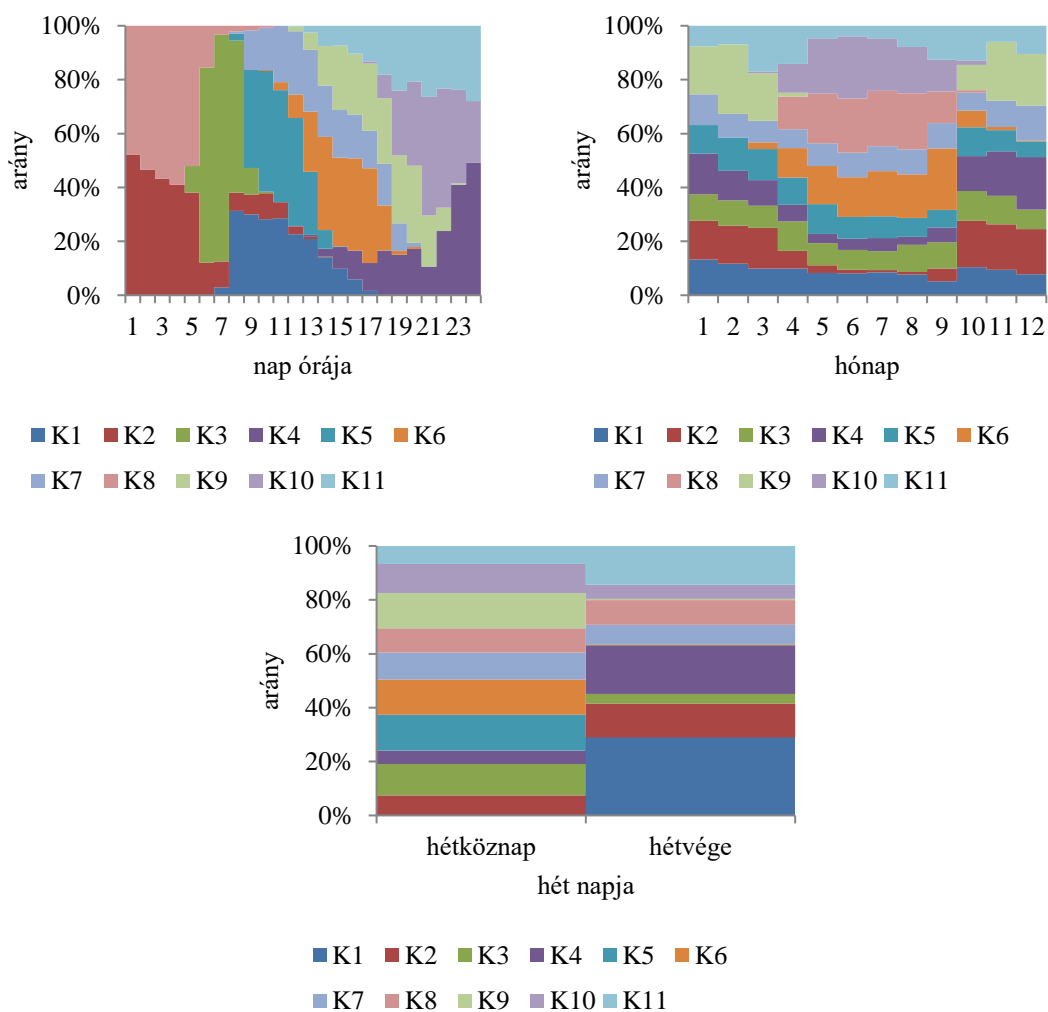


e) Terhelés és hőmérséklet

Forrás: saját számítások (R) és saját készítésű ábra (R).

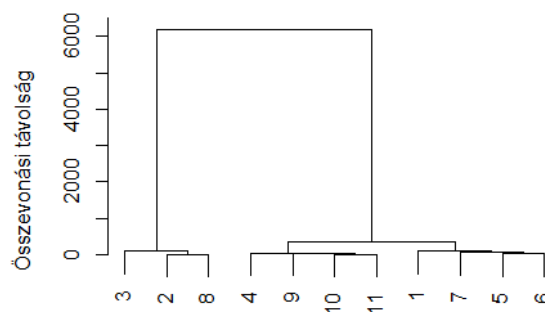
A kétdimenziós ábrák és a komponenseken belül megoszlások alapján gyakorlatilag mindegyik komponenst tekinthetjük önállóknak. A dendrogram azonban jó lehet arra, hogy megállapítsuk, hogy melyik komponensek hasonlóak inkább, vagy akár – nagyon

38. ábra: Komponensek összetétele a vállalati portfólió esetén¹²⁰



Forrás: saját számítások (Excel) és saját készítésű ábra (Excel).

39. ábra: Komponensek közötti távolság alapján készült dendrogram a vállalati portfólió esetén



Forrás: saját számítások (R) és saját készítésű ábra (R).

¹²⁰ A K1, K2 ... K11 az 1., 2. ... 11. komponenseket jelölik. A jelölés hasonló lesz a következő pontban is.

átfedő komponensek esetén – az összevonás is lehetséges a könnyebb értelmezhetőség kedvéért.¹²¹ Ha csoportosítani szeretnénk a komponenseket, akkor három nagyobb csoport különíthető el (ezeket már a 38. ábra alapján is meg lehetett sejteni):

- 1. csoport: 2., 3. és 8. komponens (reggeli negyedórák),
- 2. csoport: 4., 9., 10. és 11. komponens (esti negyedórák),
- 3. csoport: 1., 5., 6. és 7. komponens (csúcsidőszaki negyedórák).

Az összevonási lépéseket látva úgy tűnik, hogy a napon belüliség az elsődleges klaszterképzési tényező, azaz az, hogy a nap melyik időszakában vagyunk, sokkal jobban determinálja a klaszterképződést, mint például a hőmérséklet, ami csak kisebb mértékben módosít egy-egy nap terhelési értékeinek az eloszlásán (ld. például az 1.4. fejezetben a dobozábrákat is).

4.1.1.3. *Eredmények egyedi fogyasztók terhelési görbéi esetén*

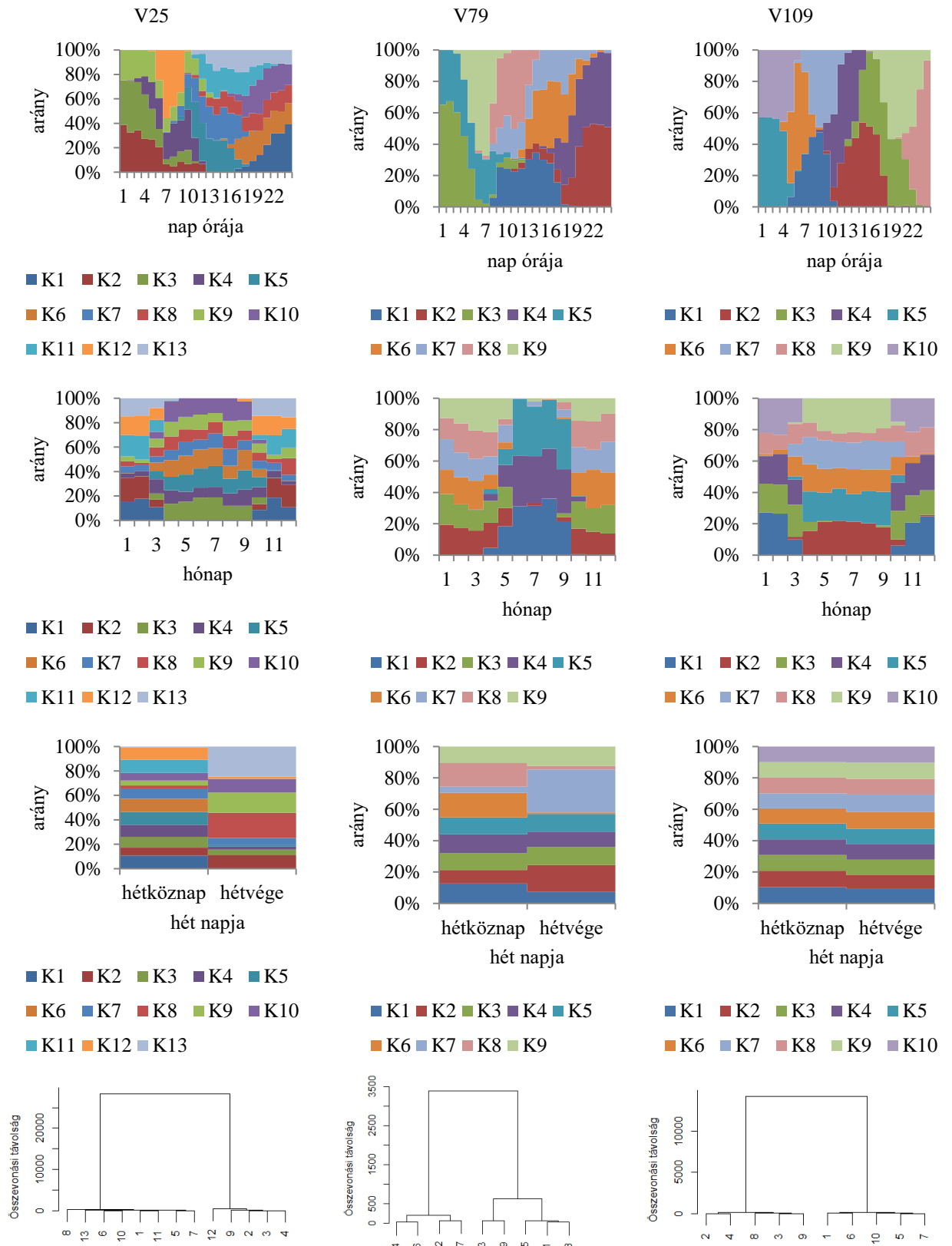
Az egyedi görbék kapcsán nagyon hasonló következtetések vonhatóak le, mint a vállalati portfólió esetében. A következő ábrán az előző pontból már ismert, a komponensek Kullback-Leibler-távolságmértéke alapján előállított dendrogram, illetve a komponensek összetétele látható néhány egyedi görbe esetében.

Az órák komponenseken belüli arányait összevetve a dendrogramokkal, mindhárom görbe esetében a napon belüli elhelyezkedés determinálja leginkább a csoportképződést itt is. Egy-két említésre méltó eredményt külön kiemelünk.

A V25 esetében például beazonosíthatóak azok a komponensek, amik a péntek estét követő hosszabb szombati lecsengést, illetve a hétfő reggeli gyors felfutást rögzítik, ezek rendre a 9. illetve a 4. komponensek. Mivel ezeknek a variabilitása nem igazán különbözik az év során, jól láthatóan ezek a komponensek minden hónapban körülbelül ugyanakkora arányt tesznek ki. Nagyon látványos, hogy a V79 esetében három komponens (az 1., a 4. és az 5.) írja le a nyári terhelés alakulását, ami elsősorban azzal magyarázható, hogy ebben az időszakban a görbe szóródása jóval meghaladja a többi időszakét (ld. még a 4.2. alfejezetet is később). Mivel a V109 esetében a hétköznapi és a hétvégi napok között nagyon kevés különbség van, ezért nem meglepő, hogy a hétköznapi és a hétvégék a komponensek között egyenletesen oszlanak el (azaz egy-egy komponensnek a „súlya” a hétköznapi és hétvégék leírásában közel ugyanakkora).

¹²¹ Hasonló, szintén entrópia alapú komponens-összevonási logika megtalálható egyébként Baudry et al. [2010] munkájában is.

40. ábra: Komponensek összetétele és a komponensek közötti távolság alapján készült dendrogramok egyedi görbék esetén



Forrás: saját számítások (R és Excel) és saját készítésű táblázat (R és Excel).

4.1.2. Klasszikus idősoros regresszió alkalmazása tipikus fogyasztási mintázat készítésére

A klasszikus idősoros regressziós alkalmazás a módszertani fejezetben is ismertett periodikus autoregresszív (PAR-) modellen alapul.

4.1.2.1. Az alkalmazott módszertan leírása

A regressziós megoldás a profilkészítést célozva alapvetően abból indul ki, hogy a regressziós logikának köszönhetően bizonyos változóknak a hatása leválasztható az idősről. Röviden megfogalmazva a felírt regresszió minden exogén változójának értékét zéróval egyenlővé téve, az ún. tipikus napi profil (*typical daily profile*, TDP) a periódusonkénti (negyedóránkénti) hosszú távú átlaggal azonos (ld. 3.1.5. szakasz), amelytől időközben a véletlen és az exogén változók hatására az aktuális fogyasztás eltér.

A regresszióban felhasznált exogén változókat tartalmazza az alábbi táblázat.

8. táblázat: A regresszió során használt magyarázóváltozók és rövid leírásuk (PAR-modell)

| | |
|--|---|
| negyedórák megkülönböztetésére szolgáló <i>dummy</i> változók | a <i>dummy</i> változó értéke 1, ha az adott nap adott negyedórájáról van szó, 0 egyébként |
| hétvégek megkülönböztetésére szolgáló <i>dummy</i> változó | a <i>dummy</i> változó értéke 1, ha hétvégi nap negyed-órájáról van szó, 0 egyébként |
| negyedórák megkülönböztetésére szolgáló interakciós változók a hétvégi napokon | a megfelelő negyedórák <i>dummy</i> változóinak illetve a hétvégek megkülönböztetésére szolgáló <i>dummy</i> változó interakcióiként (azaz szorzataként) előálló változók |
| ünnepnapokat és egyéb speciális napokat (pihenő illetve áthelyezett napokat) jelölő <i>dummy</i> változók ¹²² | a megfelelő <i>dummy</i> változó értéke 1, ha ünnepnap, pihenő nap illetve áthelyezett munkanap negyed-órájáról van szó, 0 egyébként |
| az ún. naplemente-hatást leíró változó | a napnyugta időpontnak a 18:00-tól vett előjeles eltérése, ld. később a 41. ábrát |
| fűtési napfok változó (<i>heating degree-day</i> , HDD) | a hőmérsékletnek a 12°C-tól lefelé vett eltérése a fűtési hatás becslésére |
| hűtési napfok változó (<i>cooling degree-day</i> , CDD) | a hőmérsékletnek a 21°C-tól felfelé vett eltérése a hűtési hatás becslésére |

Forrás: saját készítésű táblázat.

A 41. ábra alsó sorában a regressziós modellhez is használt hőmérséklet illetve a naplemente-hatás változók láthatóak. A változók modellben történő szerepeltetését

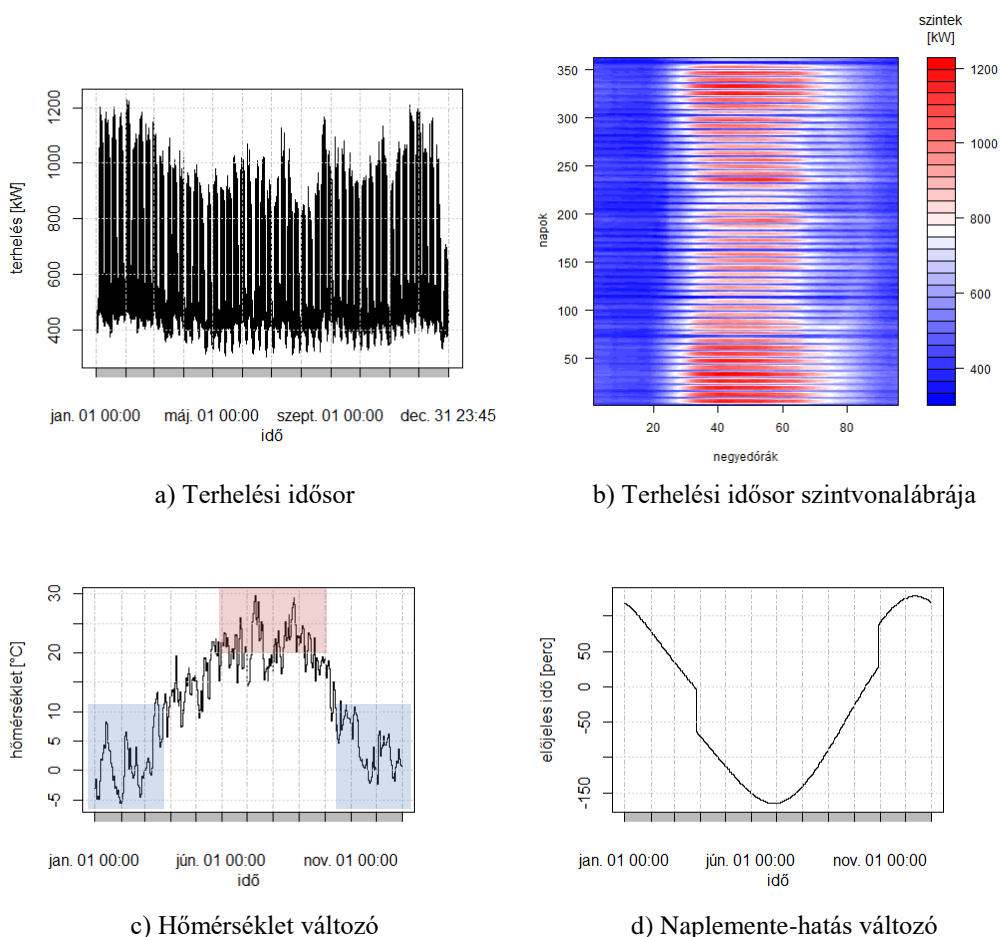
¹²² Ezek az alábbiak (a 2011-es teljes évre vonatkozóan):

- ünnepnapok: január 1., március 15., április 24-25., május 1., június 12-13., augusztus 20., október 23., november 1., december 24-26.
- pihenőnapok: március 14., október 31.
- áthelyezett munkanapok: március 19., november 5. illetve az áthelyezett munkanapok közé soroltuk még az ún. két ünnep közé eső napokat (december 27-től 31-ig), amelyek hétköznapiak ugyan, de a legtöbb helyen (ha csak lehet) szabadságolások vannak, ugyanúgy, mint az áthelyezett munkanapokon.

megindokolandó, az ábra felső soraiban maga a vállalati portfólió terhelési idősora, illetve annak a korábbi fejezetben már bemutatott ún. szintvonalábrája látható.

A naplemente-hatásnak a táblázatban feltüntetett módon történő szerepeltetésére az angol rendszerirányító (*Elexon* [2013]) által is használt megoldás alkalmazását jelenti. Nagyon hasonló egyébként ez ahhoz, amikor a nap hossza (napkeltétől napnyugtáig eltelt időszak hossza órában mérve, ld. *Sugár* [2011]) változót használjuk. Az itt alkalmazott megoldás ettől annyiban több, hogy az óráátállítás miatti ugrások is megjelennek a változóban, ahogy azt a szintvonalábrákon is észre lehetett venni a terhelési idősorok alakulásában. Az itt használt naplemente-változó lefutása szinte egy az egyben visszaadja a szintvonalábrán is tükröződő alakot, 90 fokos elforgatás után.¹²³

41. ábra: A hőmérséklet és naplemente hatása a terhelésre a vállalati portfólió esetén



Forrás: saját számítások (R) és saját készítésű ábra (R).

A fűtési illetve a hűtési napfok változók rögzített 12 illetve 21 °C-os küszöbértékek melletti használata nem biztos, hogy minden idősor esetében a legoptimálisabb, a

¹²³ Ez az utóbbi hatás az országos rendszerterhelés idősoron sokkal jobban látszódik.

legjobb illeszkedést biztosító megoldást nyújtja. Ezek alkalmazása ugyanakkor a mi esetünkben kielégítő lesz, a legjobb illeszkedést (például maximális determinációs együtthatót, R^2 -et produkáló) megoldások jelentősen nem változtatnának az eredményeken. Sugár [2011] alapján egyébként ez a két küszöbérték az, amely a magyar országos rendszerterhelés esetében a legjobb illeszkedést biztosítja.¹²⁴ Az ábrán a kék illetve piros téglalapokkal elfedett területek a hőmérsékletnek ezt a szezonálisan eltérő hatását célozzák érzékeltetni.

Néhány megjegyzést érdemes fűzni még elsősorban a *dummy* változók konstrukciójához. Általában regressziós modellt úgy szoktunk felírni, hogy egy konstans mellett a különböző *dummy* változók becsült paraméterei ezt a konstanszt tolják el (amennyiben a változók értéke 1-et vesz fel). Például egyetlen konstans tag mellett definiálhatunk 95 *dummy* változót a negyedórák hatásainak becslésére valamelyik kiválasztott negyedóra, mint referenciakategória mellett. Ezzel ekvivalens megoldás konstans nélkül 96 *dummy* változó szerepeltetése, de a periodikus autoregresszív modell 96 egyenletből álló modellként történő felírására gondolva 96 konstanszt becsülve is ugyanazt a megoldást kapjuk.

Röviden összefoglalva a lényegét, a tipikus napi profilok becslése során minden, a táblázatban szereplő – exogén – változó hatását szűrjük, és a 96 konstans alapján kalkuláljuk a 96 negyedóra hosszú távú, minden exogén változó hatásától megtisztított hosszú távú egyensúlyi szintjét. Ezek a tipikus napi profilok szerepelnek majd a görbéket jellemző tömörített információként, azaz görbetulajdonságként a továbbiakban.

4.1.2.2. Eredmények a vállalati portfólió és egyedi fogyasztók terhelési görbéi esetén

Ebben a pontban már csak a PAR-modellből származtatott végeredmények, azaz a tipikus napi profilok tárgyalásával foglalkozunk.

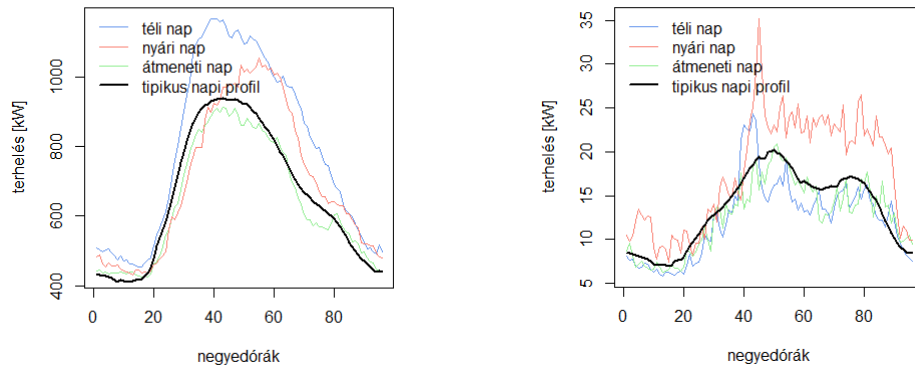
A 42. ábrán a vállalati portfólió illetve a V109 görbe egy-egy téli, nyári és átmeneti időszaki, illetve a számított tipikus napi profil lefutása látható (a többi görbére vonatkozóan ld. az E) melléklet első ábráját).

Nagyon jól látszódik, hogy az exogén változók hatása kiszűrésre került, így a görbék alakja leginkább az átmeneti időszak – nem hőmérsékletfüggő – napi alakjához hasonlít. Emellett természetesen a tipikus napi profil alapvetően egy sima, a véletlen hatástól is megtisztított, zajmentes görbét jelent. Érdemes megfigyelni, hogy a téli illetve

¹²⁴ Ugyan földgáz esetében a fűtési napfoknál a 16°C-os küszöbérték használata megszokott, villamosenergia esetében az alacsonyabb érték azzal magyarázható, hogy a villamosenergiával működtetett fűtési rendszerek fogyasztása akkor válik érezhetővé, amikor már sokkal hidegebb van.

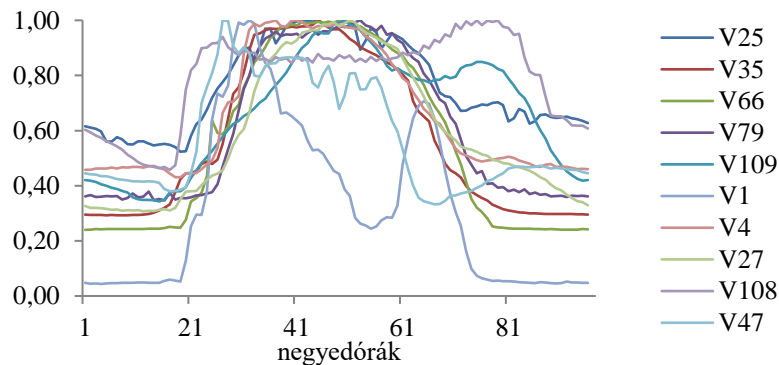
a nyári hőmérséklet hatás nem ugyanazokban a negyedórákban módosít a görbe alakján: a téli inkább reggel illetve a nap első felében, a nyári inkább a nap közepén, délután. (A V109 görbe esetében csak nyári hőmérsékletfüggés van, ott ez kevésbé látványos.)

42. ábra: Vállalati portfólió és V109 görbe napi lefutásai illetve a tipikus napi profilok (TDP-k) alakulása



Forrás: saját számítások (R) és saját készítésű ábra (R).

43. ábra: Normalizált tipikus napi profilok (TDP-k) alakulása egyedi görbék esetén



Forrás: saját számítások (R) és saját készítésű ábra (R).

A 43. ábrán több görbéhez kalkulált – normalizált¹²⁵ – tipikus napi profilok szerepelnek. Jól látható, hogy ez a klasszikusnak tekinthető módszer is sok tulajdonságot megragad (pl. *peak-off-peak* arányok, napon belüli csúcsidőszak(ok) elhelyezkedése, stb.). Az is látszik ugyanakkor, hogy nemcsak a zaj, de minden, ami a profillal kapcsolatos kockázatnak a forrása lehet, kiszűrésre kerül. Mindez tartalmát tekintve nem olyan

¹²⁵ A normalizálás itt a megjelenítést támogatja, a profilsorozatok képzésénél pedig azért használjuk, hogy a görbe nagyságrendje ne befolyásolja az eredményeket. A normalizálás az alábbi művelet elvégzését jelenti: TDP_t / TDP_{max} , ahol TDP_t a TDP t -edik negyedórabeli értéke, TDP_{max} pedig a TDP maximuma.

előnyös, de a görbék simasága, zaj(- és lényegében kockázat)mentessége a klaszterezés során hasznos tulajdonság.

4.1.3. Profilszoportok képzése

Az értekezés fókuszja inkább az egyedi fogyasztók viselkedésének feltérképezése, ezért ezt a fejezetet részben a korábbi kutatási eredmények kiegészítésének, részben pedig további kutatási irányok megalapozásának is tartjuk. Ennek ellenére a kapcsolat az előző fejezethez azért szoros: az egyedi görbénként megvalósított információátvitel minősége így könnyebben ellenőrizhető. Módszertanilag pedig – a keverék-modellek oldalán – a görbék összehasonlítása esetén alkalmazott Kullback-Leibler-távolságmérték használatában manifesztálódik.

A különböző fogyasztási mintázattal rendelkező görbék viszonyát alapvetően két-féle dimenzióban vizsgálhatjuk:

1. egyrészt a tipikus fogyasztási mintázatok alapján a hasonló terhelési karakterisztikával jellemezhető fogyasztókat homogén csoportokba (profilszoportokba) sorolhatjuk,
2. másrészt minden egyes görbe esetén egy (absztrakt) mérőszám (távolságmérték konkrét értéke) alapján tudunk rangsorolni, hogy a görbe profilja mennyire tér el a teljes portfólió profiljától.

A profilszoportok képzését természetesen a klasszikusnak nevezett regressziós megoldás, illetve a keverék-modellek esetén is elvégezzük.

4.1.3.1. Az alkalmazott módszertan leírása

A keverék-modellek esetében a távolság mérésére a Gauss-féle keverék-eloszlások esetén érvényes Kullback-Leibler-távolságmértéket használjuk. Mivel a keverék-modellek paramétereit a görbét jól leíró görbetulajdonságoknak tekintjük, ezért ezek távolságának, különbözőségének mérésére ez a távolságmérték különösen kézenfekvő. Tegyük fel, hogy $\tilde{p}_i(x, \theta_i)$ és $\tilde{p}_j(x, \theta_j)$ (vagy röviden $\tilde{p}_i(x)$ és $\tilde{p}_j(x)$) két Gauss-féle keverék-eloszlásfüggvény. Ekkor a távolság mérték közelítő képlete (Hershey-Olsen [2007] alapján) az alábbi módon írható fel:

$$D(\tilde{p}_i, \tilde{p}_j) = \int_{-\infty}^{\infty} \tilde{p}_i(x) \log \frac{\tilde{p}_i(x)}{\tilde{p}_j(x)} dx =$$

$$= \sum_{m=1}^M \pi_m \cdot \log \frac{\sum_{m*=1}^M \pi_{m*} \cdot \exp(-D(\Phi_m, \Phi_{m*}))}{\sum_{m**=1}^{M**} \omega_{m**} \cdot \exp(-D(\Phi_m, \Phi_{m**}))},$$

ahol:

- $\{\pi_{m*}\}_{m*=1}^{M*}$ illetve $\{\omega_{m**}\}_{m**=1}^{M**}$ a Gauss-féle keverék-eloszlások keverék-valószínűségeit jelölik, ahol M^* illetve M^{**} a komponensek száma,
- Φ a megfelelő normális eloszlású sűrűségfüggvényeket,
- $\theta_i = \{\pi_{m*}, \mu_{m*}, \sigma_{m*}\}_{m*=1}^{M*}$ illetve $\theta_j = \{\pi_{m**}, \mu_{m**}, \sigma_{m**}\}_{m**=1}^{M**}$ pedig a normális eloszlású komponensek paramétereit jelölik.

Természetesen a távolságmérték itt sem szimmetrikus, ezért a fejezetben használt távolságmérték a $D(\tilde{p}_i, \tilde{p}_j)$ és a $D(\tilde{p}_j, \tilde{p}_i)$ értékek átlaga:

$$D_{KL}(\tilde{p}_i, \tilde{p}_j) = \frac{D(\tilde{p}_i, \tilde{p}_j) + D(\tilde{p}_j, \tilde{p}_i)}{2},$$

ahol $D_{KL}(\tilde{p}_i, \tilde{p}_j)$ jelöli a szakasz hátralévő részében a görbék összehasonlítása esetén használatos távolságmértéket.

Előnye a távolságmértéknek, hogy kihasználja azt a feltételezést, hogy **egy-egy komponens a többdimenziós normális eloszlás ír le. Nem szükséges emellett az sem, hogy a két keverék-eloszlás komponenseinek a száma azonos legyen.**

A klasszikus megoldás esetében a görbék közötti távolságot euklideszi távolságmérték alapján definiáljuk, az alábbi ismert módon:

$$D_{eucl}(TDP_j, TDP_k) = \sqrt{\sum_{i=1}^{96} (TDP_{ij} - TDP_{ik})^2},$$

ahol TDP_j illetve TDP_k a j -edik illetve k -edik görbék tipikus napi profiljait, TDP_{ij} és TDP_{ik} pedig ezeknek az i -edik negyedórán felvett értékeit jelölik.

A távolságokat minden esetben normalizált görbékre számoltuk.

Mind a klasszikus, mind a keverék-modellek esetében hierarchikus klaszterezéssel, Ward-féle összevonási logikával dolgozunk. A Ward-módszer vagy Ward-féle minimum variancia módszer az agglomeratív hierarchikus klaszterezésen belül sokszor használt algoritmus. Lényege a klasztereken belüli variancia minimalizálása. A hierarchikus klaszterezés logikájának megfelelően a klaszterek lépésenkénti összevonása úgy történik meg, hogy a klasztereken belüli variancia az összevonás eredményeként a lehető legkisebb mértékben növekedjen.

4.1.3.2. Hasonló fogyasztási mintázattal rendelkező görbék csoportosítása¹²⁶

A következő ábrákon a kétféle megoldás páronkénti távolságokat tartalmazó távolságmátrixai, illetve az ezek alapján lefuttatott, hierarchikus klaszterezéssel előállított dendrogramok láthatóak, egyelőre csak néhány kiválasztott görbe esetében. A távolságmátrixokban vastaggal kiemeltük a legnagyobb illetve a legkisebb távolság értékeit.

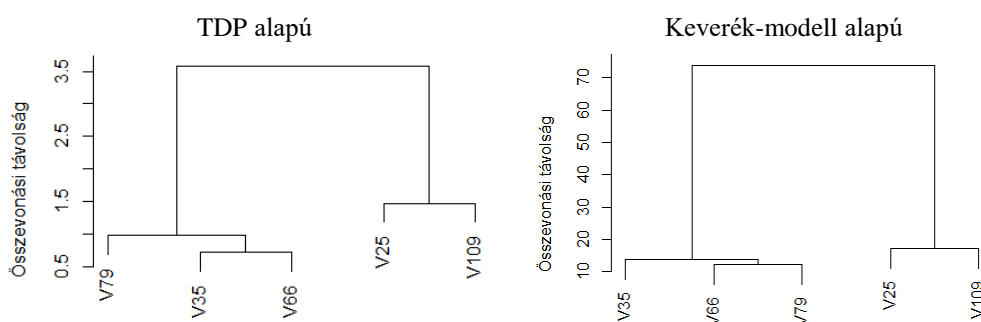
A dendrogramokon tükröződő összevonási távolságok alapján megállapítható, hogy alapvetően mindkét megoldás ugyanazt a csoportbesorolást adja eredményül. A konklúzió sokkal inkább abból a szempontból érdekes, hogy a kétféle megoldás külön-külön miért vezetett mégis azonos eredményre.

44. ábra: Távolságmátrixok a TDP alapú klaszterezés illetve a keverék-klasztározás esetén

| Euklideszi távolságmátrix (TDP alapú) | | | | | | Kullback-Leibler távolságmátrix (keverék-modell alapú) | | | | | |
|--|-------------|-------------|------|------|------|---|-------|-------|--------------|-------|------|
| | V25 | V35 | V66 | V79 | V109 | | V25 | V35 | V66 | V79 | V109 |
| $d =$ | 0 | | | | | $d =$ | 0 | | | | |
| | 2,26 | 0 | | | | | 32,08 | 0 | | | |
| | 2,55 | 0,71 | 0 | | | | 70,74 | 14,57 | 0 | | |
| | 1,87 | 0,88 | 0,93 | 0 | | | 25,88 | 12,19 | 12,18 | 0 | |
| | 1,47 | 2,15 | 2,31 | 1,68 | 0 | | 17,10 | 16,86 | 73,84 | 17,14 | 0 |

Forrás: saját számítások (R) és saját készítésű ábra.

45. ábra: Görbék közötti távolság alapján készült dendrogram



Forrás: saját számítások (R) és saját készítésű ábra (R).

Az TDP alapú megoldás esetében a V35 és V66 kerül először összevonásra, hiszen ezeknél mind a csúcsidőszaki, mind a csúcsidőszakon kívüli lefutás nagyon hasonló. A V79 tőlük alapvetően abban különbözik, hogy a csúcsidőszakon kívüli terhelés szintje arányaiban valamivel magasabb a másik kettőhöz képest. Ehhez képest a másik két görbe esetében az esti terhelés szintje jóval magasabb, és teljesen más a görbék alak-

¹²⁶ Az eredmények könnyebb nyomon követéséhez ld. az E) melléklet ábráit.

ja is, a V109 esetében például két csúcsterhelés is van napon belül (a második valamivel enyhébb).

A keverék-modellen alapuló megoldás szerint a két leginkább hasonló görbe a V66 és V79, ami nem meglepő, hiszen hétvégén mindkettőnél gyakorlatilag zsinór a terhelés – és ami különösen fontos: **a hétköznapi csúcsidőszakon kívüli terhelés szintje gyakorlatilag azonos a hétvégi terhelés flat szintjével**. Egyébként mindkettőnél csak nyári hőmérsékletváltozás van, a V35 tőlük annyiban különbözik, hogy ott a téli hőmérsékletváltozás is jelen van – minden bizonnyal ez az oka, hogy harmadikként a V35 csatlakozik ehhez a klaszterhez, hiszen **a hétköznapi off-peak és a hétvégi terhelés szintje** ugyanúgy viszonyul egymáshoz, mint a V66 és a V79 esetében. A V25 és V109 esetében a terhelés hétvégén nem konstans, a V25-nél van egy hosszú szombati lecsengés, a V109 esetében pedig olyanok a hétvégi lefutások is, mint hétköznapokon. Azonban nem a „hétvégén terhelés megléte” az igazán meghatározó tényező a V25-nél (hiszen az csak a szombati napok első felét teszi ki), hanem az, hogy a V35-V66-V79 hármasához képest a hétvégi terhelés szinte jóval alacsonyabb, mint a hétköznapi off-peak terhelés szintje. A fentiek kellően alátámasztják azt is, hogy miért a V25 a legkilógóbb a görbék között (ennél a legnagyobb a többi görbétől vett távolságok átlaga (távolságmátrix első oszlopa), valamivel magasabb, mint a V109 esetében (távolságmátrix utolsó sora)).

Technikailag újdonságot nem tartalmaz, de az előbb leírtakat támasztja alá az alábbi táblázat, ahol a görbék vállalati portfóliótól vett távolsága Kullback-Leibler-távolságban illetve euklideszi távolságban mérve látható:

9. táblázat: Fogyasztói görbék portfóliótól vett távolsága

| Távolság / Görbe | V25 | V35 | V66 | V79 | V109 |
|---------------------------|-------|-------|-------|-------|-------|
| Euklideszi távolság | 0,79 | 1,63 | 1,91 | 1,29 | 1,13 |
| Kullback-Leibler-távolság | 23,68 | 33,36 | 79,19 | 63,01 | 20,41 |

Forrás: saját számítások (R) és saját készítésű táblázat.

A Kullback-Leibler-távolság alapján a portfólióhoz legközelebb lévő görbék a V25 és V109, hiszen azoknál nem zsinór a terhelés hétvégén, ahogy a portfóliónál sem.

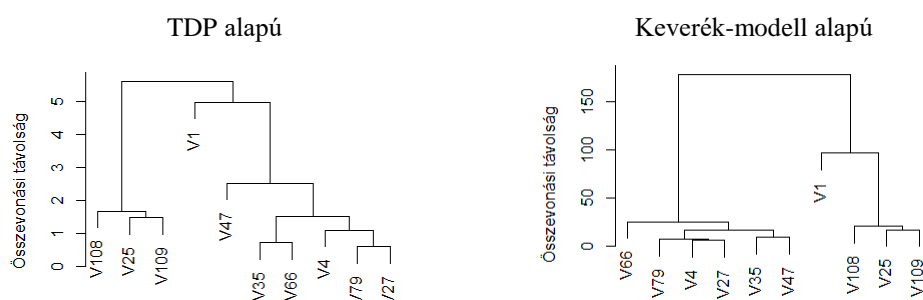
Annak oka, hogy az euklideszi távolság esetén a távolságok különbözősége arányaiban jóval kisebb, minden bizonnyal az, hogy ez a hétvégéket „nem látja”, és csak a hétköznapi napi alakból tud következtetni. Valószínűleg azért tűnik legközelebbinek a

V25, mert a reggeli felfutás illetve az esti lecsengés tekintetében ez hasonlít – hétköznapokon – a legjobban a vállalati portfólióhoz.

4.1.3.3. Hasonló fogyasztási mintázattal rendelkező görbék csoportosítása (kibővített példa)

A kibővített példa esetén a különbség nemcsak az „ajánlott” klaszterek számában különbözik (hiszen a TDP alapú modell esetében négy (esetleg három), a keverék-modell esetében inkább három klaszter választása lehet a megfelelő,¹²⁷ ld. a 46. ábrát), hanem a besorolás alakulásában. Az egyértelműen látszik, hogy mindkét megoldás a V1-et *outlier* megfigyelésnek tekinti, hiszen mindkét besorolás esetén külön egyelemű klasztert alkot. Az *outlier* tulajdonságot az okozza, hogy ennél a fogyasztónál a **csúcsidőszakon kívüli és a hétvégi terhelés gyakorlatilag zéró**, míg a többi görbénél magasabb. A V47 a tipikus napi profil lefutása alapján is (ld. korábban a 43. ábrán) valahol a V1 és az ötelemű klaszter között helyezkedik el. Onnan egyébként a TDP alapú megoldás eredménye könnyen lekövethető. A 47. ábrán ugyanezek a napi profilok láthatóak, az ábra bal oldalán az azonos klaszterbe kerülők azonos színnel szerepelnek.

46. ábra: Görbék közötti távolság alapján készült dendrogram (kibővített példa)

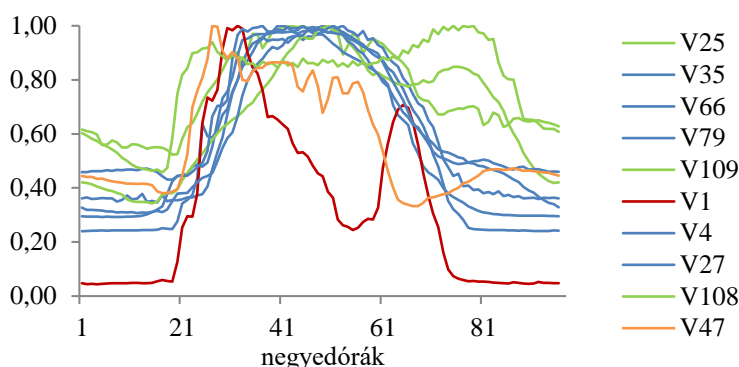


Forrás: saját számítások (R) és saját készítésű ábra (R).

A keverék-modell esetében valószínűleg azért hasonlít jobban a V25-V108-V109 hármas alkotta klaszterhez a V1, mert ezek azok, ahol van egy **második esti csúcs** (vagy egy viszonylag magas esti terhelési szint a V25 esetében). Ezt a tulajdonságot a TDP alapú megoldás nem ismeri fel, és a V1 ahhoz a klaszterhez kerül közelebb, ahol a csúcsidőszakon kívüli terhelések alacsonyabbak (ld. a hatelemű klasztert a 46. ábra bal oldalán).

¹²⁷ Amennyiben az *outlier* V1 görbét mindkét esetben külön klaszternek tekintjük, ld. még a főszövegben is.

47. ábra: Normalizált tipikus napi profilok (TDP-k) klaszterezésének eredménye



Forrás: saját számítások (R) és saját készítésű ábra (Excel).

A keverék-modell esetében a V47-tel együtt kialakuló hatelemű klasztert tekintve van igazán különbség a kétféle megoldás között. Valamennyi görbe esetén jóval kisebb a hétvégi terhelés volumene a hétköznapiakhoz képest (a V66 és a V79 esetén gyakorlatilag konstans a várható értéke). A klaszterképződés alakulásának valószínűleg az a vezérlőelve, hogy a V35 és a V47 esetén van egy nagyon erős téli hőmérséklet-hatás, míg a V66 és V79 esetében csak a nyári hőmérséklet-hatás erős. A hőmérséklet a *peak-off-peak arányokat* módosítja a hőmérséklet-hatás nélküli időszakhoz képest, márpedig a TDP alapú modell tulajdonképpen csak azt tudja figyelembe venni a tipikus napi profil konstrukciója miatt.

Bár a kétféle megoldás végigkövetése, visszafejtése nem volt olyan kényelmes, a vizsgált példák megmutatták a keverék-modell előnyeit, ami itt többek között a napon belüli profil fundamentálisabb jellegű figyelembevételén és a hőmérséklet-hatás klaszterképződésben betöltött szerepén keresztül jelent meg.

Mivel a bemutatott két megoldás (a TDP alapú illetve a keverék-modell alapú) filozófiájában jelentősen különbözik, ezért a fejezetben nem tértünk külön ki rá, de mindenképpen érdemes megemlíteni, hogy a különböző megoldások mekkora **becsült paraméterszámmal** dolgoznak. A klasszikus megoldás esetében a becsült paraméterek száma **294**¹²⁸. A keverék-modell esetében a becsült paraméterek számát az alábbi képlet adja meg:

$$(K - 1) + K \left[d + \frac{d(d+1)}{2} \right],$$

¹²⁸ Amennyiben a magyarázóváltozókat tartalmazó 8. táblázat alapján ellenőrizzük a 294 az alábbiak összegeként adódik: 96 negyedórás *dummy*, 1 hétvége *dummy*, 95 interakciós *dummy*, 3 ún. különleges nap *dummy*, 1 naplemente-hatás, 1 fűtési hatás, 1 hűtési hatás változó, illetve 96 autoregresszív együttható.

ahol a K a becsült komponensek számát, d a változók számát jelenti. A képlet első tagja a becsült *prior* valószínűségek száma, a szögletes zárójelben pedig a komponensenként becsült átlagok illetve a becsült kovariancia mátrix paramétereinek a száma jelenik meg. Mivel a mi esetünkben a változók száma mindig 6, ezért a komponensek számának a függvényében a becsült paraméterek száma a 10. táblázatban szerepel. 10-11 komponensszám tehát az, ami a klasszikus megoldásokkal nagyságrendileg azonos becsült paraméterszámot jelent. Az esetek többségében ennél sokkal több komponenst nem becsültünk. Az eltérések pedig nagyságrendileg sem olyan nagyok; különösen, ha figyelembe vesszük azt, a keverék-modell esetén nemcsak a várható értéket, de a varianciát is becsüljük, ezáltal jóval több érdemi, használható információ rögzítése történik meg. A görbénkénti becsült paramétereket illetve modellszelekciós kritériumok értékeit egyébként a 11. táblázat tartalmazza.

10. táblázat: Becsült paraméterek száma a komponensek függvényében a keverék-modell esetén

| Komp. száma | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------------|------------|-----|-----|-----|-----|-----|-----|-----|-----|------------|
| Becs. par. száma | 27 | 55 | 83 | 111 | 139 | 167 | 195 | 223 | 251 | <u>279</u> |
| Komp. száma | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Becs. par. száma | <u>307</u> | 335 | 363 | 391 | 419 | 447 | 475 | 503 | 531 | 559 |

Forrás: saját számítások és saját készítésű táblázat.

11. táblázat: Modellszelekciós kritériumok értéke és becsült paraméterek száma a keverék-modellek esetén

| Görbe | V25 | V35 | V66 | V79 | V109 |
|------------------|---------|---------|---------|---------|---------|
| BIC | 140 376 | 122 375 | 102 456 | 109 914 | 136 212 |
| Komp. száma | 13 | 11 | 13 | 9 | 10 |
| Becs. par. száma | 363 | 307 | 363 | 251 | 279 |
| Görbe | V1 | V4 | V27 | V108 | V47 |
| BIC | 60 362 | 93 320 | 101 882 | 138 566 | 93 149 |
| Komp. száma | 13 | 11 | 7 | 11 | 11 |
| Becs. par. száma | 363 | 307 | 195 | 307 | 307 |

Forrás: saját számítások (R) és saját készítésű táblázat.

4.1.4. Eredmények, következtetések összegzése

A fejezet egyetlen terhelési görbére vonatkozó eredményeinek a bemutatása részben a keverék-modell logikájának a megértését, részben a volumenkockázattal kapcsolatos fejezetek megalapozását szolgálta. Az, hogy a keverék-modellek által előállított komponensek paraméterei, mint tipikus fogyasztási mintázatok statisztikai megfelelői (avagy a görbetulajdonságok) mennyivel nyújtanak realisabb képet egy-egy görbéről, az a görbék klaszterezése, több görbe összehasonlítása nyomán derült ki igazán. Azt ta-

pasztaltuk, hogy a keverék-modell a teljes görbe struktúráját sokkal jobban képes rögzíteni (mind a napon belüli alak, mind a *peak-off-peak*, hétköznap / hétvége tekintetében), így a **H2 hipotézist nem tudjuk elvetni, a keverék modell, azáltal, hogy görbénként az adott görbét egyedileg jellemző komponenseket becsüli, sokkal realisabb profilok kialakítását támogatja, mint a klasszikus megoldások.**

A keverék-modellen alapuló megoldás egyértelmű előnye abban áll, hogy a csoportképzés során a terhelés teljes eloszlását figyelembe veszi, tulajdonképpen **tömörített információval** dolgozik, míg a klasszikus megoldásoknál sokszor inkább **információvesztésről** beszélünk. Az itteni példákban ez úgy jelent meg, hogy a tipikus napi profiloknál (TDP-knél) például a hétvége, a hőmérséklet, stb. hatása is kiszűrésre került.

A fejezetben használt keverék-modell **kovariancia-struktúra** alapján csoportosít, így a klasszikus értelemben vett tipikus napi profil közvetlenül nem adódik. Az előállított tipikus fogyasztási mintázatok, az ún. komponensek nem a napi diszkretizálás logikáját követik, nem a napok mentén szerveződnek, hanem a felhasznált változók együttes alakulása, eloszlása mentén. Tipikus napi lefutások természetesen itt is előállíthatók (a keverék-modellre épülő regressziós módszernek köszönhetően, ld. a következő fejezetet), azonban egy tipikus (vagy várható) napi lefutás több komponens keverékéből áll majd elő.

Ez természetesen nem jelenti azt, hogy a klasszikus (napi profilgörbén alapuló) megoldások teljes mértékben haszontalanok volnának, ugyanis a görbék varianciájának nagy részét a terhelés napon belüli variabilitása adja és ezt ezek is képesek rögzíteni. Vállalati görbéknel pedig általában a hétvégi terhelés olyan kicsi, hogy az így elhanyagolt információ kevésbé jelent nagy veszteséget. Ráadásul az éves, heti szezonális alapján lehet előzetesen a fogyasztókat makrokategóriákba sorolni, és azokon belül klaszterezni.

A második példa alapján már sokkal jobban fény derült a keverék-modell előnyére, hogy az a napon belüli alakot is sokkal inkább fundamentálisan kezeli (ld. például napon belüli csúcsidőszakok, hőmérsékletváltozás). Ennek oka feltehetőleg abban van, hogy a Kullback-Leibler-távolságmérték révén a szórást (kovarianciát) is figyelembe veszi, és **az alak mellett a kockázat alapján is csoportosít, így nemcsak a hasonló profillal, de a hasonló kockázati karakterisztikával rendelkező görbék kerülnek ennél a megoldásnál egy csoportba.**

4.2. A fogyasztás bizonytalanságának modellezése

Mivel a fogyasztás bizonytalanságának, a volumenkockázatnak a leírása a véletlen komponens viselkedésének a jellemzését jelenti, a feladat feltételezi valamilyen modell meglétét, amelynek alapján a véletlen komponens realizációi keletkeznek. Természetesen a gyakorlatban ez a feltétel nem mindig teljesül, így léteznek egyszerűbb, heurisztikus eszközök is a volumenkockázatnak a mérésére, amelyek konkrét modell hiánya esetén is iránymutatóak.

Klasszikus idősoros regressziós modellek (SARMA illetve PAR¹²⁹) felhasználásával ellenőrizzük, hogy a véletlen komponens állandó szóródását feltételező megoldások a fogyasztási bizonytalanságot (milyen mértékben) nem kezelik, és **időszaktól függően alul- illetve felülbecsülik azt**. Az eredmények arra engednek következtetni, hogy a fogyasztási idősorok **heteroszkedasztikus** viselkedése alapvetően szezonális és naptári hatás jellegű, azaz:

1. ahogy a többszintű szezonális megjelenik a fogyasztási idősorok várható értékében, ugyanúgy a szóródásban is jellemzően megjelenik ez a többszintűség,
2. így kézenfekvőnek tűnik olyan megoldás választása, ami a fogyasztási idősor varianciáját ugyanúgy a magyarázóváltozók függvényében írja fel, mint magának az idősornak a várható értékét.

A Gauss-féle keverék-modellre épített regresszió éppen ilyen, a fenti követelményeknek eleget tevő lehetőséget kínál számunkra. Ennek felhasználásával ún. feltételes standard hibákat illetve konfidencia-intervallumokat számolunk, és bemutatjuk, hogy azok mennyiben konzisztensek a véletlen komponens realizációival. Az eredményeket több egyedi görbén is bemutatjuk, kihangsúlyozva az elért eredmények általánosabb felhasználhatóságát, és a klasszikusnak nevezett megoldásokhoz képest (egyébként pénzben is mérhető) gyakorlati jelentőségét.

4.2.1. Volumenkockázat a klasszikus idősoros regressziós modellekben

Ebben a szakaszban a klasszikus idősoros regresszió alkalmazásával kapott következtetések mellett egy, a gyakorlatban is könnyen implementálható, ún. heurisztikus módszert is bemutatunk a volumenkockázat mérésére.

¹²⁹ A módszertani fejezetben részleteztük, hogy technikailag a különbség mindössze abban van, hogy a PAR-modell periódusonként (azaz a mi esetünkben negyedóránként) különböző autoregresszív együttthattót becsül (ami így – *ceteris paribus* – 95 darabbal több becsült paramétert jelent). A klasszikus SARMA-logika értelemszerű kiterjesztése miatt a PAR-modellt is klasszikus modellnek tekintjük.

4.2.1.1. *Volumenkockázat heurisztikus mérése*

Lo-Wu [2003] az előrejelzési kockázat modell hiányában történő mérésére egy ún. **kockázati indexet** (*risk index*) definiálnak. Mivel az előrejelzési hibák az előrejelzett (várható) értéktől vett eltérések realizációinak tekinthetők, ezért ez a volumenkockázat mérésének egy lehetséges mérőszáma. Magyarországon a negyedóra az a legkisebb intervallum, amelyre vonatkozóan előre lehet kereskedni, így a kockázati indexet is ennek megfelelő bontásban érdemes számolni, így a továbbiakban $k = 1, 2 \dots 96$ a napon belüli negyedórákat jelöli. Lényegét tekintve a kockázati index egy szórás tartalmú mutató, számítása a következő:

$$R_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (L_{k_i} - \overline{L_{k_i}})^2},$$

ahol:

- R_k a k -adik negyedóra ún. kockázati indexe,
- n a visszatekintési periódus hossza (időszakok száma),
- L_{k_i} a k -adik negyedórában történő változás értéke a visszatekintési periódus i -edik időszakában,
- $\overline{L_{k_i}}$ a k -adik negyedórában történő változások átlaga a visszatekintési periódus alatt.

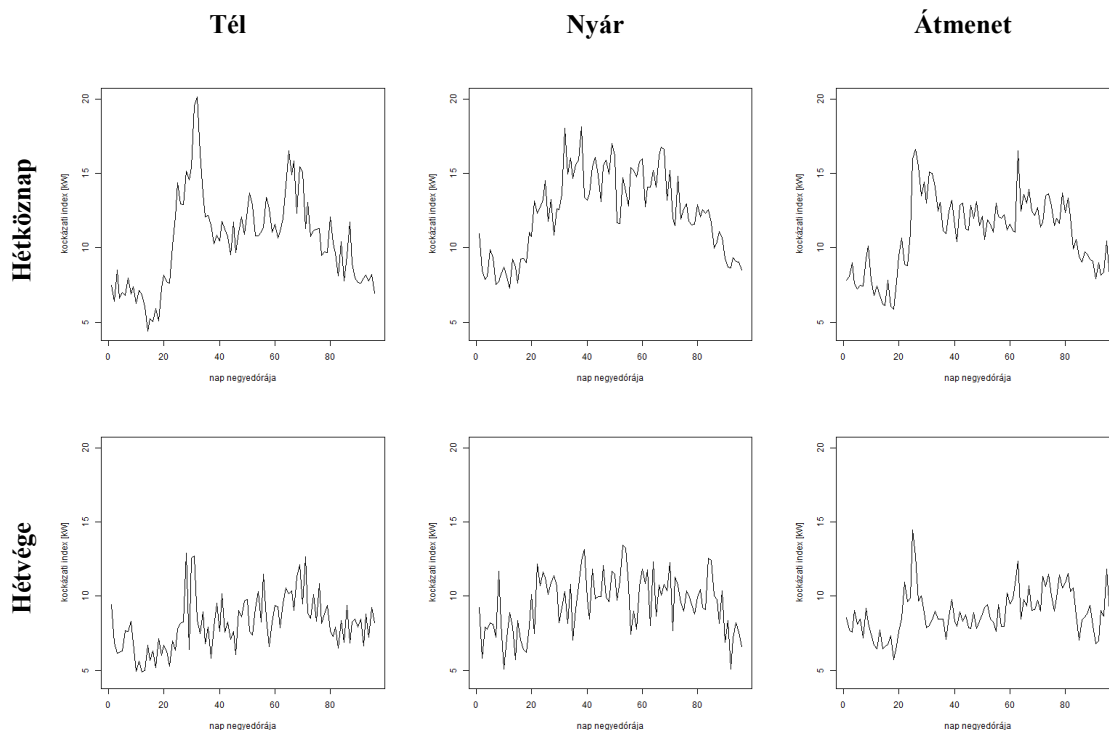
A korábbi eredmények alapján feltételezett többszintű szezonalitásból adódóan érdemes a kockázati index alakulását évszakonként, illetve hétköznapra és hétvégékre külön-külön számolni. Az eredményeket csak a vállalati portfólió vonatkozásában vizsgáljuk meg, az erre vonatkozó kockázati indexek láthatóak a következő ábrákon,¹³⁰ ahol a zajosság ellenére az alapvető tendenciák azért kirajzolódnak abban a tekintetben, hogy mikor lehet magasabb vagy alacsonyabb a volumenkockázat. Ezek összegzése előtt azonban érdemes az eredményekkel szembeni fenntartásoknak is helyet adni.

A bemutatott megoldás előnye ugyan, hogy a megfelelő minta kiválasztásával a bizonytalanság számítása a **manuálisan, szubjektíve** legjellemzőbbnek ítélt időszak (például az elmúlt két hét) alapján történhet meg. Módszertanilag kedvező tulajdonság még, hogy a mutatószám az azonos órákra eső változások szórásán alapul, és a **szórás** egyébként is **a kockázat, a bizonytalanság mérőszámának tekinthető**. Hátrányként tudható be ugyanakkor az önkényesség, azaz tulajdonképpen az, hogy a visszatekintési periódus hossza csak tapasztalati úton határozható meg, és ha túl rövid időszakot válasz-

¹³⁰ Az ábrák a negyedórás bontású, teljes évre vonatkozó idősorra készültek.

tunk, akkor a **kis mintaelemszám** okoz gondot (túl zajos lesz a kockázati index alakulása), ha pedig túl hosszú időszakot, akkor a „**lokalizáltság**” nem lesz érvényes.

48. ábra: Kockázati indexek alakulása a vállalati portfólió esetén



Forrás: saját számítások (R) és saját készítésű ábra (R).

A számítások mellé nem társul mindezek mellett **fundamentális indokolás** sem, hiszen valamely nap volumen-kockázatát a megelőző napok alapján kalkuláljuk, amely egy nagyon egyszerű, „naiv” értékelés. Másrészt – még nagy mintaelemszám esetében is – ezek a kockázati indexek meglehetősen **hektikusak** maradnak, nem olyan simák, legfeljebb bizonytalan tendenciák olvashatók ki, ami a praktikus használatot megnehezíti. Ilyen egyszerű számítási logika mellett egyébként az sem dönthető el egyértelműen, hogy a nagyobb kiugrások **valósak-e**, tényleg az adott időszak nagyobb kockázatát tükrözik-e; ami tehát mindenképpen valamilyen modell alapú megközelítés irányába mutat.¹³¹

¹³¹ Alternatívaként sokszor alkalmazott megoldás még az előző hét azonos órájához képesti eltérések szóródásának a vizsgálata. Ez sokkal inkább logikus párja annak, amikor a gyakorlatban az adott nap terhelését az egy héttel korábbi nap terhelésének felhasználásával prognosztizáljuk – így az üzleti felhasználásban ez az ún. heti kockázati index sokkal inkább elterjedt. Bár kvalitatív jelleggel hasonlóak a következtetések ennél a mutatónál is, azonban az itt vizsgált kockázati index az, ami praktikus összehasonlítható a modell alapú eredményekkel. Egyébként az Enoro Smart Energy Management egyik alkalmazása (szoftvere) ilyen – heti késleltetésen alapuló – módon (is) méri egy-egy fogyasztó görbéjének változékonyságát, variabilitását (ld. <http://www.enoro.com/>). Természetesen ilyenkor az ünnepnapok, munkanap áthelyezések, óraátállítás miatti 23 illetve 25 órás napok figyelembe vétele nélkül számolunk.

4.2.1.2. A volumenkockázat karakterisztikájának modell alapú feltárása

A görbék klasszikus idősoros technikával történő modellezését SARMA- illetve PAR-regresszióval végezzük tehát el. Mivel az egyedi görbék esetében csak egy évnyi adat áll rendelkezésre, ezért a becslést a teljes időszakra készítettük, ennél fogva mintán kívüli teljesítmény nem értékelhető, azonban maguknak a mintán kívüli számításoknak – azaz az előrejelzéseknek – a vizsgálatát érdemes külön témaként tekinteni.

A célunk itt elsősorban a probléma feltáró (esetleg ellenőrző) jellegű bemutatása arra vonatkozóan, hogy a klasszikus megoldások melyik időszakokban és milyen mértékben becsülik felül illetve alul a volumenkockázatot. Az egyes lépéseket részletesebben csak a vállalati portfólió esetén mutatjuk be, a több görbére történő összehasonlító elemzést külön pont tartalmazza.

12. táblázat: A klasszikus idősoros regresszió során használt magyarázóváltozók és rövid leírásuk

| | |
|---|---|
| órák ¹³² megkülönböztetésére szolgáló <i>dummy</i> változók | a megfelelő <i>dummy</i> változó értéke 1, ha a nap adott órájának a negyedórájáról van szó, 0 egyébként |
| hétvégek ¹³³ megkülönböztetésére szolgáló <i>dummy</i> változó | a <i>dummy</i> változó értéke 1, ha hétvégi nap negyedórájáról van szó, 0 egyébként |
| órák megkülönböztetésére szolgáló interakciós változók a hétvégi napokon | a megfelelő órák <i>dummy</i> változóinak illetve a hétvégek megkülönböztetésére szolgáló <i>dummy</i> változó interakcióiként (azaz szorzataként) előálló változók |
| ünnepnapok és egyéb speciális napokat (pihenő illetve áthelyezett napokat) jelölő <i>dummy</i> változók | a megfelelő <i>dummy</i> változó értéke 1, ha ünnepnap, pihenő nap illetve áthelyezett munkanap negyedórájáról van szó, 0 egyébként |
| az ún. naplemente-hatást leíró változó | a napnyugta időpontnak a 18:00-tól vett előjeles eltérése, ld. még a 41. ábrát |
| fűtési napfok változó (<i>heating degree-day, HDD</i>) | a hőmérsékletnek a 12°C-tól lefelé vett eltérése a fűtési hatás becslésére |
| hűtési napfok változó (<i>cooling degree-day, CDD</i>) | a hőmérsékletnek a 21°C-tól felfelé vett eltérése a hűtési hatás becslésére |

Forrás: saját készítésű táblázat.

A regressziós becslés során használt magyarázóváltozókat tartalmazza a 12. táblázat. A változók a negyedórák hatásának eltérő (órán belül különböző vagy azonos) kezelésétől eltekintve ugyanazok, amelyekkel a tipikus fogyasztási mintázatok készítése során a 4.1. fejezetben is dolgoztunk. A fentiekhez annyit érdemes mindenképpen már itt hozzáfűzni, hogy a **fejezetben használt keverék-regressziónak az egyik nagy elő-**

¹³² Lehetne a negyedórák megkülönböztetésére szolgáló *dummy* változókkal is dolgozni (ugyanúgy, ahogy a korábbi profilozással foglalkozó alfejezetben tettük), így azonban 23 helyett 95, majd az interakciós változók szerepeltetésével még több becsült paraméterrel kellene dolgoznunk. A két megoldás közötti választás természetesen modellszelekciós kritériumokkal ellenőrizhető, ettől azonban itt eltekintünk.

¹³³ Természetesen a hét napjainál is lehetne annyi *dummy* változót használni, hogy a hét minden napját megkülönböztessük, azaz 1 helyett 6 *dummy* változót is használhatunk, azonban itt is a takarékoság elve vezérel minket.

nye az, hogy olyan **változótranszformációkat**, mint milyeneket például a hőmérsékletből számított napfok változók esetén készítünk, **nem kell elvégeznünk, hiszen a modell a nemlinearitást automatikusan kezeli**. Itt nemcsak a nemlinearitásnak, hanem az automatizmusnak is hangsúlyos a szerepe, hiszen egyedi görbék esetében – a napfoknál maradva – a küszöbértékek megválasztása manuálisan nagyon nehézkes és körülményes lehet.

A klasszikus regressziós modellekhez visszatérve, a közülük történő választás a Schwarz-Bayes információs (BIC) kritérium alapján történik (ld. a 13. táblázatot).

13. táblázat: A klasszikus idősoros regressziós modellek illeszkedését jellemző mutatószámok alakulása¹³⁴ a vállalati portfólió esetén

| Modell | BIC | RMSE [kW] | MAPE [%] | MAE [kW] |
|---------------------------------|-------------------|-----------|----------|----------|
| ARMA(1, 0) | 272 989,81 | 11,81 | 1,56 | 9,05 |
| ARMA(2, 0) | 272 980,76 | 11,81 | 1,56 | 9,05 |
| ARMA(0, 1) | 358 614,89 | 40,04 | 4,90 | 29,59 |
| ARMA(0, 2) | 332 575,50 | 27,61 | 3,50 | 20,77 |
| ARMA(1, 1) | 273 009,57 | 11,80 | 1,56 | 9,05 |
| SARMA(1, 0)(1, 0) ₉₆ | 270 413,01 | 11,51 | 1,52 | 8,83 |
| SARMA(1, 0)(0, 1) ₉₆ | 271 077,58 | 11,48 | 1,52 | 8,84 |
| SARMA(0, 1)(0, 1) ₉₆ | 348 205,91 | 34,50 | 4,32 | 25,79 |
| PAR(1) | 270 483,40 | 11,28 | 1,49 | 8,64 |
| PAR(2) | 270 212,60 | 11,03 | 1,46 | 8,47 |
| PAR(3) | 270 506,20 | 10,92 | 1,44 | 8,36 |

Forrás: saját számítások (R) és saját készítésű táblázat.

¹³⁴ Az alábbi táblázatban szereplő mutatószámok a modellilleszkedés illetve az előrejelzések jóságának értékelésekor használatosak (y_i az idősor realizálódott értékeit, \hat{y}_i a modell alapján becsült értékeket jelenti, h pedig mintán belüli illeszkedés vizsgálatokor az idősor hossza, előrejelzés esetén az előrejelzési horizont hossza):

| Mutatószám neve | Mutatószám számítási módja |
|--|--|
| Átlagos négyzetes hiba (Mean Squared Error) | $MSE = \frac{\sum_{i=1}^h (y_i - \hat{y}_i)^2}{h}$ |
| Átlagos négyzetes hiba gyöke (Root Mean Squared Error) | $RMSE = \sqrt{\frac{\sum_{i=1}^h (y_i - \hat{y}_i)^2}{h}}$ |
| Abszolút százalékos hiba (Absolute Percentage Error) | $APE_i = \frac{ y_i - \hat{y}_i }{y_i}$ |
| Átlagos abszolút százalékos hiba (Mean Absolute Percentage Error) | $MAPE = \frac{\sum_{i=1}^h APE_i}{h}$ |
| Átlagos abszolút hiba (Mean Absolute Error) | $MAE = \frac{\sum_{i=1}^h y_i - \hat{y}_i }{h}$ |

Az *RMSE* nagyobb súllyal veszi figyelembe a nagyobb hibákat (ezért is jobb inkább a mintán belüli illeszkedés értékelésére), míg a *MAPE* az átlagos hiba %-os kifejezésére alkalmas, így mértékegységfüggetlen mutató, ami összehasonlítások esetében hasznos tulajdonság.

Kizárólag a **BIC** kritérium minimumát tekintve a PAR(2)-t kellene választanunk, de a 96-tal több becsült paraméterrel szemben a modellszelekciós kritériumon elért csökkenés nem olyan jelentős, így a választandó modelleknek a SARMA(1, 0)(1, 0)₉₆ illetve a PAR(1) tekintendők. Az eredményekből látszik egyébként az is, hogy jóval rosszabb az illeszkedése azoknak a modelleknek, amelyekben az időbeliséget csak mozgóátlagolású taggal (azaz a hibában) kezeljük.^{135,136}

Ugyan a táblázat mutatószámai (a pontos definíciókat ld. lábjegyzetben) közül mintán belüli értékelés során az **RMSE**-nek a számítása a szokványos, nagyságrendi tájékozódás végett érdemes megvizsgálni a többit is, amelyek inkább az előrejelzések, a mintán kívüli értékelések használata során lennének fontosak. A legtöbb, illeszkedéssel és előrejelzéssel foglalkozó munka a relatív (%-os) hibákat részesíti előnyben, így első sorban a **MAPE**-t. Ennek hátránya azonban, hogy a hibák jellemzően a csúcsidőszakban nagyobbak, amikor – általában – a terhelési értékek is magasabbak, így egy abszolút értékben nagyobb hiba relatíve alacsonyabbnak tűnhet. Hasonlóan egy alacsonyabb terhelés mellett elkövetett hiba, a terhelés arányában sokkal nagyobbbnak tűnhet, indokolatlanul megnövelve ezzel a MAPE értékét. A MAPE értékét tehát éppen a drágább (csúcs)időszakokban elkövetett hibák csökkentik, és az olcsóbb (völgy)időszakokban elkövetettek növelik. Mindezek figyelembe vétele mellett a **MAE** mutató tűnik a legszerencsésebb választásnak energetikai idősorok esetén¹³⁷, hiszen az elkövetett hibákból fakadó eltéréseket majd úgyis naturáliában, [kWh]-ban kell figyelembe venni.

¹³⁵ Figyelembe véve azonban azt, hogy a vizsgált idősor esetében a terhelések csúcserké 1000-1200 [kW] körül mozog, bármelyik modell illeszkedése jónak mondható, de amennyiben a jobb illeszkedésen (és pontosabb előrejelzéseken) van a hangsúly, érdemes valamilyen autoregresszív modellt választani. Amennyiben viszont a paraméterek (béták) értelmezésén van a hangsúly, azok *ceteris paribus* típusú értelmezése miatt csak olyan modellek esetében lehetséges, ahol nincsen autoregresszív tag, így ebben az esetben a csak mozgóátlagolású tagokat tartalmazó modelleket választhatjuk.

¹³⁶ A paraméterek értelmezése önmagában természetesen hasznos, hiszen támogatja az idősor dekompozícióját a magyarázóváltozók mentén, azonban gyakorlati jelentősége valamivel talán kisebb. Mivel a célunk nem elsősorban a pontosságnak, hanem a volumen-kockázatnak az értékelése, ez a fókusz is azt támogatja, hogy a legjobban illeszkedő (így autoregresszív tagot is tartalmazó) modellekkel dolgozzunk tovább, és azok kapcsán vizsgáljuk meg a reziduumok viselkedését részletesebben is.

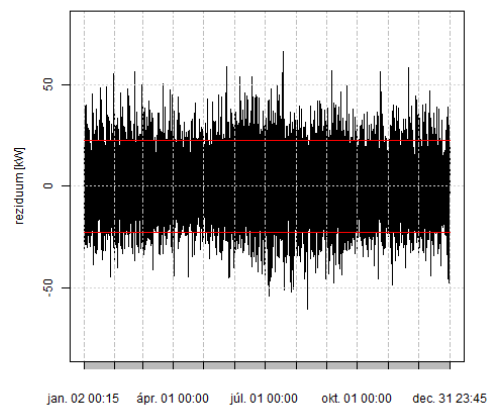
Későbbi fejezetekben látni fogjuk, hogy egyedi fogyasztói görbék esetében is jellemző az, hogy az időbeliséget nem mozgóátlagolású, hanem autoregresszív tag ragadja meg pontosabban. Ez feltehetően arra utal, hogy valamilyen nagyon enyhe trend van az idősorokban. Ez a jelenség a gyakorlatban a mintán kívüli előrejelzések során érhető tetten. Még nagyon bonyolult és komplex modellek esetében is előfordul az, hogy több periódusos előrejelzések esetén több időszakot alul- vagy felülbecsülünk. Ennek valószínűleg vagy ez az enyhe trend a magyarázata, vagy olyan – gyakran előforduló, de nehezen felismerhető – szinteltolások, amelyeket esetleg csak bonyolultabb technikákkal lehetne megragadni. A mintán kívüli teljesítmény vizsgálata többek között ezért is kezelendő külön területként.

¹³⁷ Nem is beszélve az energiapiaci árfolyam idősorokról, ahol előfordulhatnak negatív árak is; az APE illetve MAPE ezekben az órákban negatív értéket venne fel, így az alkalmazása semmiképpen nem előnyös.

Rátérve a hibák vizsgálatára, a 49-50. ábrán a SARMA(1, 0)(1, 0)₉₆-modell reziduuma, illetve a körjük szerkesztett 95%-os megbízhatósági szintű konfidencia-sáv látható a teljes évre illetve az 1.4. fejezetben is vizsgált három kiválasztott hétre.

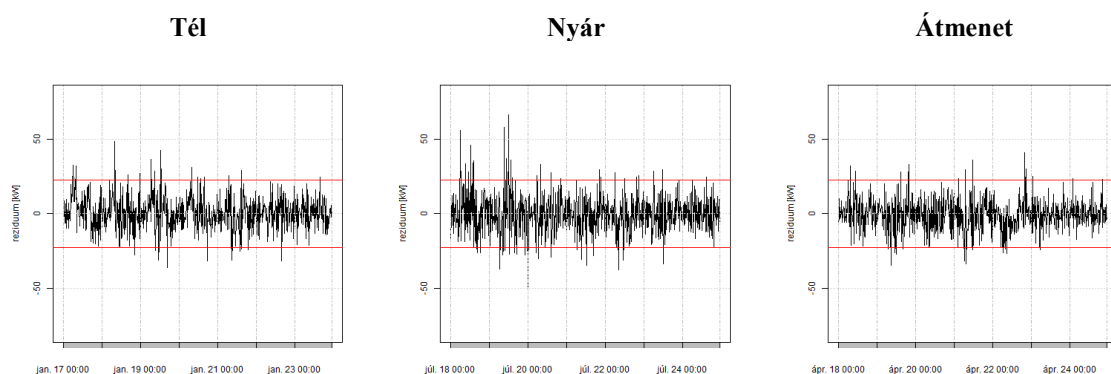
Természetesen a konfidencia-sáv végig konstans, és összességét tekintve a reziduumok zöme a konfidencia-sávon belül van. Az illesztett modell abban az esetben tekinthető megfelelőnek, amennyiben például a 95%-os megbízhatósági szintű konfidencia-intervallumon kívül a reziduumok 5%-a esik (vagy általánosságban $(1-\alpha)$ %-os megbízhatósági szint esetén α %-a). Ennek az értéke ebben az esetben 5.61 % (ld. 14. táblázat), tehát a modell összességében, **globálisan** nézve tekinthető helyesnek, hiszen az eltérés az 5%-tól elhanyagolható.

49. ábra: SARMA(1, 0)(1, 0)₉₆-modell reziduumainak alakulása a vállalati portfólió esetén



Forrás: saját számítások (R) és saját készítésű ábra (R).

50. ábra: SARMA(1, 0)(1, 0)₉₆-modell reziduumainak alakulása a vállalati portfólió esetén néhány kiválasztott hétre



Forrás: saját számítások (R) és saját készítésű ábra (R).

Annak az oka, hogy a konfidencia-sávon kívül eső megfigyelések aránya közel esik az elvárthoz, feltehetően a regressziós modell számítási logikájával magyarázható. A becslési eljárásnak ugyanis a lényege az, hogy az eredményváltozónak a tény illetve a modell alapján becsült értékei közötti távolság négyzetes értelemben minimális legyen. A **nagyobb eltérések** végett a modell **standard hibája** – mint szórás tartalmú mutató – nagyon **nagy** lesz, így ha a reziduumok összességét tekintjük, a konfidencia-sávon kívül eső megfigyelések aránya – pontosan a széles konfidencia-sáv végett – közelítőleg teljesül (annak ellenére, hogy az ábrák alapján – a másik oldalról tekintve – bizonyos időszakokban feleslegesen szélesnek tűnik a konfidencia-sáv).

Az előző ábrákat elkészíthetjük a többi, hasonlóan jó illeszkedésű modell esetében is, de az eltérések elhanyagolhatóak (ezt a 14. táblázat és az 51. ábra PAR(1) és ARMA(1, 0)-modellekre vonatkozó, közel hasonló eredményei is engedik sejtetni).¹³⁸

14. táblázat: Konfidencia-intervallumon (CI95)¹³⁹ kívül eső megfigyelések aránya és átlagos konfidencia-intervallum nagysága a vállalati portfólió esetén (klasszikus regressziós modellek)

| Modell | Arány [%] | Átlagos CI95 [kW] |
|---------------------------------|-----------|-------------------|
| SARMA(1, 0)(1, 0) ₉₆ | 5,61 | 45,12 |
| PAR(1) | 5,75 | 44,01 |
| ARMA(1, 0) | 5,66 | 46,30 |

Forrás: saját számítások (R) és saját készítésű táblázat.

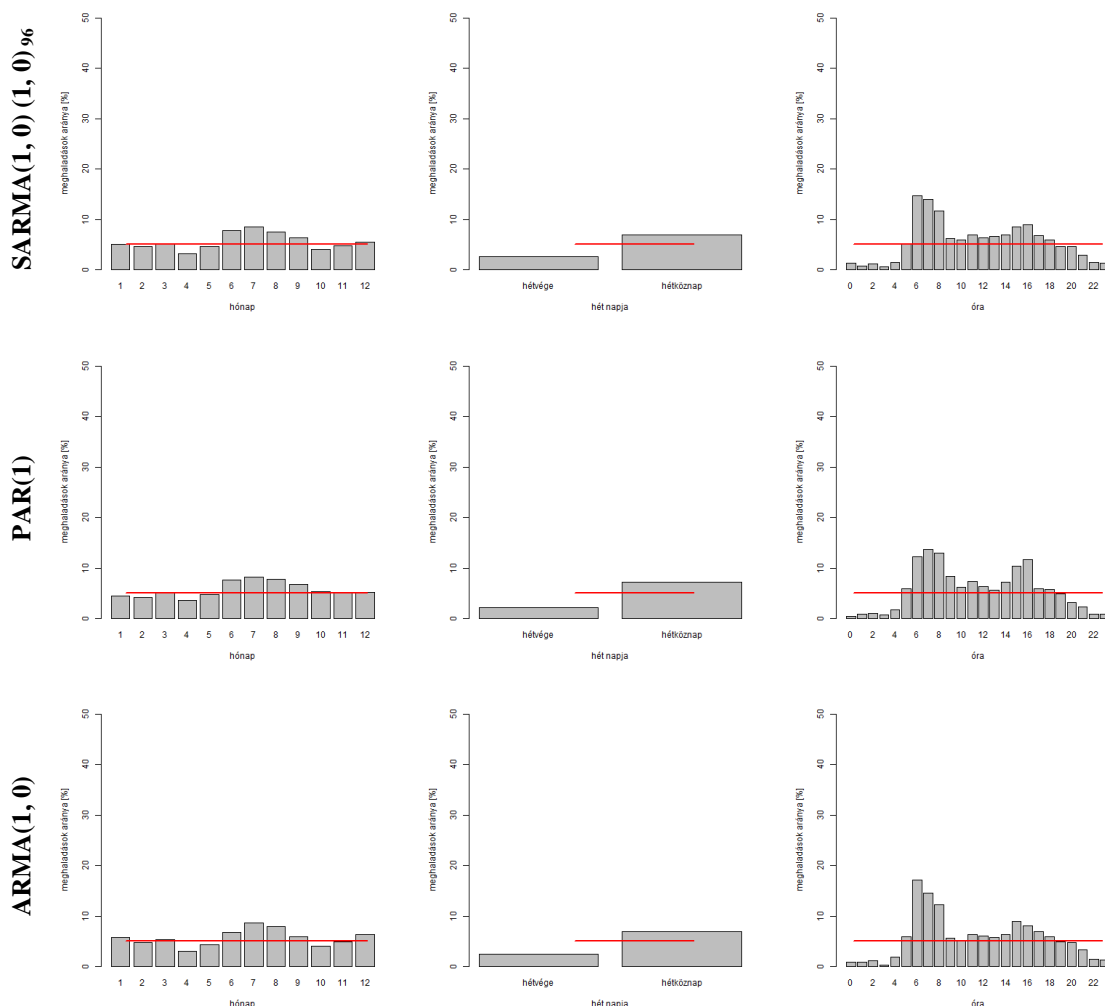
A hibák feltételezett szezonális jellegű heteroszkedasztikus viselkedésének részletebb ellenőrzésére érdemes azt is megvizsgálni, hogy a konfidencia-intervallumon kívül eső megfigyelések aránya különböző időszakokban (például évszakok, hét napjai, napon belüli órák) hogyan alakul (ld. az 51. ábrát). Amennyiben az állandó szóródás feltételezése helyes, a különböző időszakokban a konfidencia-sávon kívül eső megfigyelések arányában az – elvárt, ebben az esetben – 5%-os szinttől (ezt az ábrán piros vonal jelzi) csak véletlenszerű eltérések megengedettek. A téli illetve nyári hónapokban, amikor a fogyasztás hőmérsékletfüggő valamivel nagyobb a konfidencia-intervallumon kívül eső megfigyelések aránya. Hétvégéken a kívül eső megfigyelések aránya jóval 5%

¹³⁸ A konfidencia-intervallumok átlagos méretében a több becsült paraméternek köszönhető kisebb standard hiba hatása tükröződik.

¹³⁹ Itt és a következő táblázatokban a CI95 a 95%-os megbízhatósági szintű konfidencia-intervallum jelölésére szolgál.

alatt van¹⁴⁰, hétköznapokon viszont jelentősen meghaladja azt, hasonlóan a reggeli illetve a csúcsidőszaki órákhoz.

51. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya a vállalati portfólió esetén (klasszikus regressziós modellek)



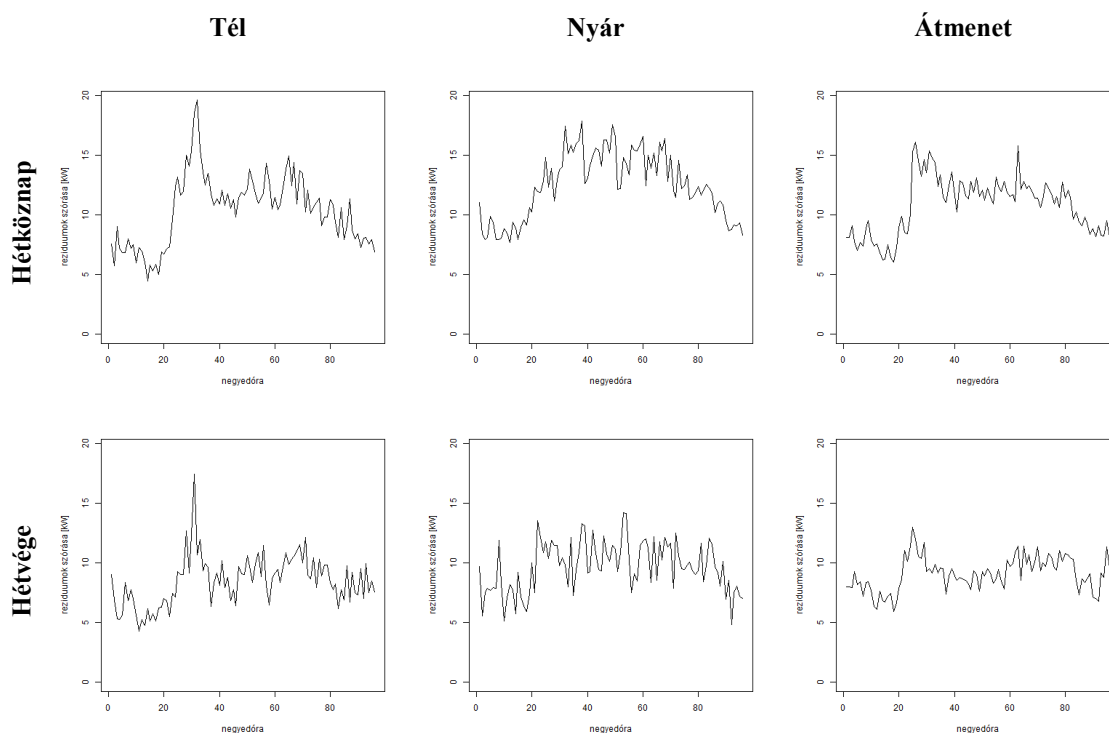
Forrás: saját számítások (R) és saját készítésű ábra (R).

Ez a heteroszkedasztikus viselkedés nyilvánvalóan bármilyen heteroszkedaszticitás-teszt alapján könnyen ellenőrizhető. Az itteni eredmények hozzáadott értéke elsősorban abban jelenik meg, hogy feltártuk azt is, hogy a **heteroszkedaszticitás karakterisztikája** alapvetően milyen. Ez a logika inkább **pénzügyekben** használt **személetet** követ, ahol szintén fontos az, hogy a volatilisabb és kevésbé volatilisabb **időszakok kockázatát helyesen és konzisztensen mérjük**, azonban ott a heteroszkedaszticitás teljesen más jellegű és ennél fogva másfajta módszerek alkalmazását indokolja.

¹⁴⁰ Az 49-50. ábrákon éppen ez volt a leglátványosabb: a végig konstans konfidencia-sáv a hétfégi időszakban illetve a hétköznapi völgyórákban jóval szélesebb, mint amit a reziduumok alakulása (szóródása) szemmel láthatóan indokol.

A klasszikus modellek tekintetében a konklúzió tehát elsősorban arra a hiányosságra hívja fel a figyelmet, hogy a konfidencia-sávon kívül eső hibák eloszlása nem egyenletes, azaz időszaktól függően a volumenkockázatot **nagyságrendekkel alul- vagy felülbecsüljük**, így az eredmények arra engednek következtetni, hogy valamilyen **időfüggő, feltételes standard hiba** (bizonyos időszakokban nagyobb, más időszakokban kisebb) számítása szükséges. Az időtől függően helyes nagyságrendekről a reziduumok szórásának az 51. ábra bontásához illeszkedő kiszámolása mindenképpen jó alapot jelenthet. Az 52. ábrán a reziduumok negyedóránkénti szórásai láthatóak a kockázati index számolásánál is alkalmazott bontás mellett.

52. ábra: Reziduumok szórása a SARMA(1, 0)(1, 0)_%-modellben a vállalati portfólió esetén



Forrás: saját számítások (R) és saját készítésű ábra (R).

Az ábrák gyakorlatilag teljesen azonosak a kockázati index megfelelő ábráival, a **tapasztalatok összességében az alábbiak** tehát:

- a hétfégek volumenkockázata alacsonyabb, a hétköznapoké magasabb,
- a reggeli, délelőtti felfutások, és az esti, 18-20 óra körüli napi lecsengések volumenkockázata jellemzően a legmagasabb,
- a napon belüli, kb. 10-18 óra közötti „igazi” csúcsidőszak volumenkockázata alacsonyabb, mint a felfutási illetve lecsengési időszak kockázata, de magasabb, mint az éjszakai illetve hajnali óráké.

A napon belüli felfutási illetve lecsengési időszakok viselkedése alapvetően az **emberi tevékenységgel** magyarázható, ami csúcsidőszakon kívül elhanyagolható, és éppen a napi felfutási és lecsengési (legnagyobb gradienssel rendelkező) időszakokban a leginkább bizonytalan. Ez alól a nyár kivétel, amikor a csúcsidőszak kockázata a legmagasabb, ami főként a villamosenergia hűtési célú felhasználása miatt lehetséges. A fontosabb következtetéseket, konklúziókat a következő pontban összegezzük.

4.2.1.3. *Eredmények, következtetések összegzése*

Mivel mind a kockázati index, mind a megfelelő bontásban számolt szórása a reziduumoknak, tartalmát tekintve szórásstartalmú mutatók, így egyformán alkalmasak a kockázat mérésére. Azonban a modellből számolt reziduumoknak önmagában – a modellalapúságon túl – látszólag kevés a hozzáadott értéke a kockázati indexekhez képest, hiszen a reziduumok alakulása ellentmond a modellből származó – konstans – konfidencia-intervallumnak.

A kockázati index illetve a reziduumok szórásának nagyságrendi egybeesése egyébként a **linearitás** feltétele miatt áll fenn. A kockázati indexet a terhelés változásainak szórásából számoltuk, és ezeknek a változásoknak az átlaga az, amelyet egy lineáris modell rögzíteni képes. A szórás pedig a lineáris modell által hátrahagyott hibák szórásában jelenik meg, ezért hoz a két megoldás hasonló eredményt.¹⁴¹

Első ránézésre egyébként elképzelhető lenne, hogy a reggeli felfutási illetve esti lecsengési időszakok magasabb szóródása mögött például a nem jól kezelt napfelkelte illetve naplemente (ami az év során változik) hatása, vagy más szezonális hatás áll. Mivel a SARMA(1, 0)(1, 0)₉₆ esetében éppen a 96-os késleltetés miatt ez a hatás nagyon jól kezelve van (az illeszkedést mérő mutatószámok javulása is az ARMA(1, 0)-modellhez képest nagyrészt ennek köszönhető), ezért a jelenség mögött feltehetően **nem** valamiféle **kihagyott, nem megfelelően kezelt szezonális hatás** áll. A kockázati indexhez képest a probléma **modell alapú** megközelítése tehát azért is elegánsabb megoldás, mert a kihagyott releváns változó lehet oka a heteroszkedasztikus viselkedésnek. Mindezek azonban modell hiányában nem ellenőrizhetőek.

A fejezetben a jobb illeszkedésre koncentrálna némileg elhanyagoltuk azt a tulajdonságot, hogy a periodikus autoregresszív modell azáltal, hogy periódusonként (tehát

¹⁴¹ A kockázati indexhez hasonlóan, abszolút (nem %-os relatív) változásokat abban az esetben szokás számolni, ha a jelenség alakulásáról feltételezhető, hogy az időszakonkénti abszolút változás tekinthető állandóak, azaz a jelenség alakulását lineáris tendencia jellemzi. A SARMA-modell mögött is pontosan ez a lineáris feltételrendszer húzódik meg.

negyedóránként) más és más autoregresszív együttthatót becsül, ezáltal feltételezi, hogy az idősort jellemző **autokovariancia függvény** – természetesen a definiált exogén változók kontrollja mellett – nem állandó, hanem csak **periodikusan állandó**.¹⁴² A lehetőség elméletileg adott arra, hogy a periódusonként különböző autoregresszív együttthatók becslése változtasson valamit a reziduumok viselkedésén, hiszen a negyedóránként különböző autoregresszív együttthatók tulajdonképpen valamilyen negyedóránként különböző szezonális hatást kezelnek. A 14. táblázat és a 51. ábra eredményei alapján azonban a konfidencia-sávon kívül eső reziduumok arányának mintázata nem változott jelentősen a periódusonként változó autoregresszív együtttható bevezetésének köszönhetően.

Itt mindenképpen érdemes megemlíteni, hogy *Subbarao et al.* [2011] ún. kNN-módszere olyan logikát követ, hogy a fogyasztás alakulását leíró modell reziduumaiból számol – lényegében historikus módszerrel – konfidencia-intervallumot, azzal a kiegészítéssel, hogy valamely időpont esetében egy jól definiált távolságmérték alapján hasonlónak tekinthető múltbéli időpontok hibáit használja fel mintaként,¹⁴³ amely mintából az intervallumhatárok percentilisekként adódnak. Az intervallumhatárok időfüggőségét a távolságmérték ugyan biztosítja, de mivel maguk a hibák szórása is elég nagy (ld. például a 52. ábrát, de gondoljunk például a kiugró értékekre vagy a multikollinearitás káros hatására), ez a zajosság a konfidencia-intervallumokban is megjelenik, ami nem olyan kedvező tulajdonság.

4.2.2. Volumenkökázat modellezése keverék-regresszióval

A volumenkökázat számszerűsítése során használt modell azonos a korábbi, tipikus fogyasztási mintázatok készítésével foglalkozó fejezetben (4.1. alfejezet) is használttal. Mivel ott a teljes rendelkezésre álló évnek csak egy részét használtuk fel a becsléshez, ez lehetőséget biztosít arra, hogy az értékelést a maradék, mintán kívüli időszakra is elvégezzük.

¹⁴² Természetesen lehetőség van annak tesztelésére, hogy szükséges-e periódusonként különböző autoregresszív együtttható – azaz periódusonként különböző autokovariancia – becslése, ez megoldható F-próbával, a 13. táblázat szerint a modellszelekciós kritériumok felhasználásával, vagy arra épülő *likelihood* alapú teszttel is.

¹⁴³ A távolságmérték logikája az euklideszi távolság logikáján nyugszik, és az adott időpont és többi időpont magyarázóváltozóinak a távolságát azoknak az eredményváltozóra kifejtett parciális hatásával (többváltozós lineáris regresszió esetén a béta becsült paraméterekkel) súlyozza. A konfidencia-intervallum számításánál a k darab legközelebbi szomszéd időpontok hibái kerülnek figyelembe vételre. Azt, hogy a konfidencia-intervallum mennyiben lesz a magyarázóváltozók konkrét értékének a függvénye, a k értékének a megválasztása befolyásolja, ami mindenképpen jelent egyfajta önkényességet, ugyanakkor a túl alacsony és a túl magas érték sem megfelelő igazán (kis minta miatti hektikus viselkedés illetve magyarázóváltozó konkrét értékétől való függőség elkenése).

Ennek a szakasznak a célja megvizsgálni azt, hogy – az előző szakaszban már ismertetett logikát folytatva – a keverék-modellen alapuló regresszióval tudunk-e olyan, feltételes standard hibát számolni, illetve konfidencia-intervallumot származtatni az eredményváltozó (a görbe terhelése) tekintetében, amelyek az adott időpont volumen-kockázatát konzisztensen tükrözi.

A klasszikus idősoros illetve keverék-regressziós eredmények ebben a szakaszban történő **összehasonlítása** – a tipikus fogyasztási mintázatokkal foglalkozó alfejezethez hasonlóan – szintén csak **kvalitatív jellegű** lehet, hiszen a különböző modellkonstrukcióból adódóan nem is ugyanazokkal a magyarázóváltozókkal dolgozunk. Az eredményeket valamivel részletesen ismét a vállalati portfólió esetén követjük végig, az egyedi görbék esetében már elsősorban a fontosabb következtetésekre térünk ki – a klasszikus megoldásokkal történő összehasonlítással párhuzamosan.

4.2.2.1. A volumen-kockázat karakterisztikájának rögzítése

A 15. táblázat a modell mintán belüli illetve mintán kívüli illeszkedését jellemző mutatószámokat tartalmazza. A célunk nem elsősorban az illeszkedésnek vagy az előrejelzésnek a javítása, a hangsúly sokkal inkább azon van, hogy a keverék-regresszió körülbelül ugyanúgy teljesít, mint a klasszikus idősoros regressziós modellek.

15. táblázat: A modellek illeszkedését jellemző mutatószámok alakulása a vállalati portfólió esetén

| Modell | Predikció | RMSE [kW] | MAPE [%] | MAE [kW] |
|---------------------------------|--------------|-----------|----------|----------|
| SARMA(1, 0)(1, 0) ₉₆ | Mintán belül | 11,51 | 1,52 | 8,83 |
| GMR | Mintán belül | 12,23 | 1,60 | 9,27 |
| | Mintán kívül | 12,28 | 1,62 | 9,38 |

Forrás: saját számítások (R) és saját készítésű táblázat.

A 95%-os konfidencia-intervallumon kívül eső megfigyelések aránya (ld. 16. táblázat) valamivel meghaladja az 5%-ot (mint ahogyan a klasszikus modellek esetében is),¹⁴⁴ a releváns megállapítás azonban sokkal inkább a rákövetkező 53. ábra alapján keletkezik.

16. táblázat: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya és átlagos konfidencia-intervallum nagysága a vállalati portfólió esetén

| Modell | Predikció | Arány [%] | Átlagos CI95 [kW] |
|--------|-----------|-----------|-------------------|
|--------|-----------|-----------|-------------------|

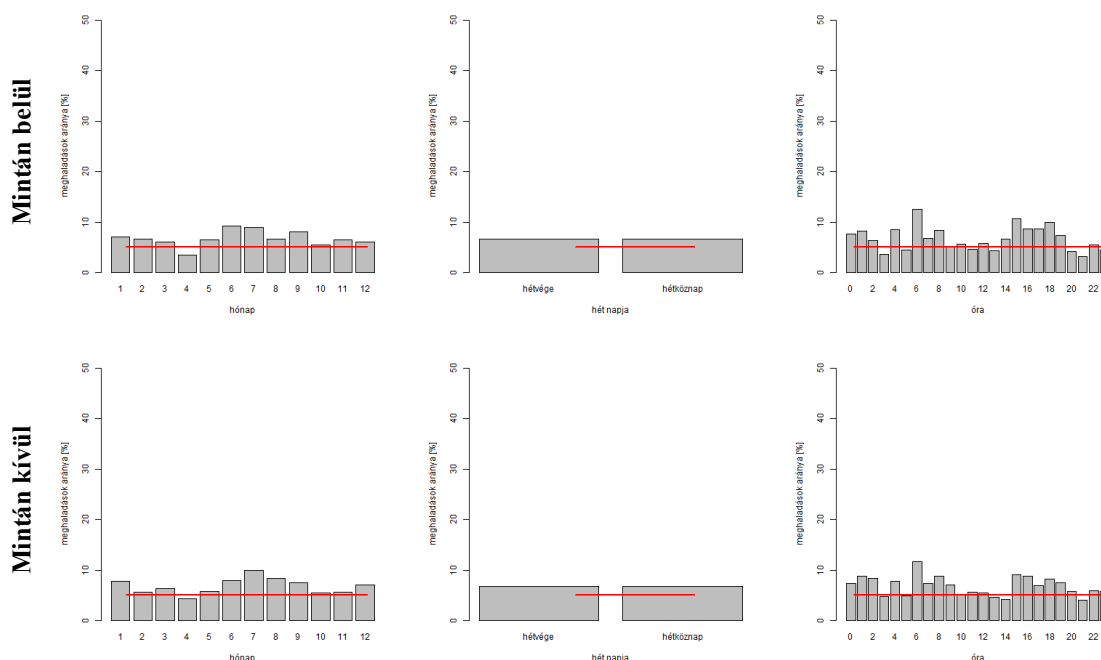
¹⁴⁴ Annak, hogy a meghaladások aránya az 5%-tól eltér, több oka is lehet, például releváns magyarázóváltozó hiánya, nem megfelelő bontású magyarázóváltozó, esetleg a hibatagra vonatkozóan nem megfelelő eloszlás feltételezése stb.

| | | | |
|---------------------------------|--------------|------|-------|
| SARMA(1, 0)(1, 0) ₉₆ | Mintán belül | 5,61 | 45,12 |
| GMR | Mintán belül | 6,69 | 43,89 |
| | Mintán kívül | 6,82 | 43,94 |

Forrás: saját számítások (R) és saját készítésű táblázat.

A konfidencia-intervallumon kívül eső megfigyelések eloszlása sokkal egyenletesebb, mint a klasszikus modellek esetében volt, azaz minden hónapban, hétköznap és hétvégén, valamint napon belül is körülbelül ugyanakkora a konfidencia-intervallumon kívül eső megfigyelések aránya. A lényegi megállapítás a napon belüli **egyenletesebb** eloszláson, az időszakosan (szabályosan) jelentkező egyenlőtlenségek kisimításán van.

53. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya a mintában és mintán kívül a vállalati portfólió esetén (keverék-regresszió)



Forrás: saját számítások (R) és saját készítésű ábra (R).

A következőkben ellenőrizhetjük számszakilag is, hogy a különböző negyedórákra becsült **standard hibák** egybecsengenek-e a különböző negyedórák **reziduumainak szóródásaival**. A standard hibák (vagy feltételes szórások) számolása természetesen a 3.2.5. módszertani szakaszban bemutatottak szerint történt. Az eredményváltozó feltételes szórása, az eredményváltozó feltételes várható értékéhez hasonlóan (a hivatkozott fejezet jelölésével élve az \hat{y}_i és a $var(\hat{y}_i)$ gyökéről – mint szórásról – van szó) a konkrét magyarázóváltozó értékek feltétele mellett értendő, és az itteni eredmények tükrében ez jelenti a standard hibák időfüggő számolásának módszertani megvalósulását.

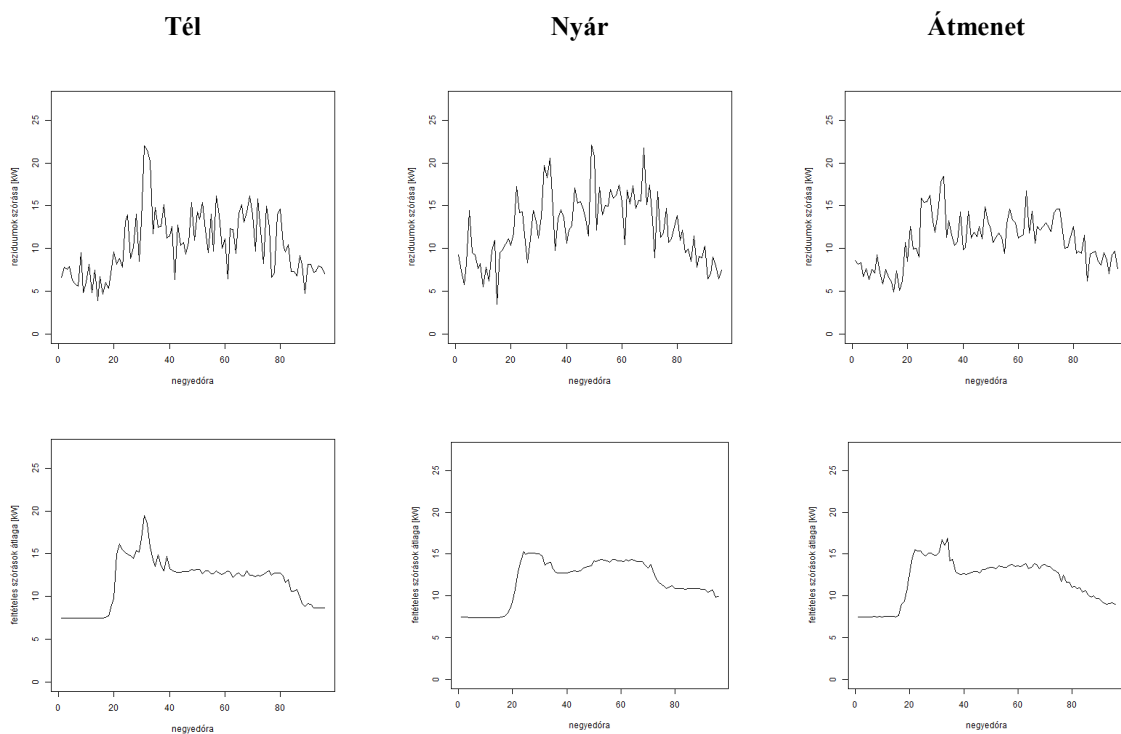
Mivel a feltételes szórások és természetesen a reziduumok is negyedóráról negyedóra-ra különbözhetnek, ezért hogy kompakt képet kapjunk ezek alakulásáról, érdemes évszakok (tél, nyár, átmenet) illetve hét napjai (hétköznapi, hétvége) szerinti bontásban megvizsgálnunk minden negyedóra-ra a reziduumok szórását, illetve a keverékmodell alapján becsült feltételes szórások átlagát (ld. 54-55. ábrák). Ez a fajta diszkretizálás már megjelent a heurisztikus mutatóknál illetve a klasszikus idősoros modelleknél is. Ez itt elsősorban az eredmények nagyságrendi validálását szolgálja. Külön érdemes kihangsúlyozni, hogy a feltételes szórások tekintetében átlagokat közlünk, azok a különböző negyedórákban – a magyarázóváltozók függvényében – különbözőek lehetnek.

Elmondhatjuk tehát ismét immár modell-alapon számolt bizonytalansági mérőszámok (azaz standard hibák) alapján, hogy a reggeli, hajnali órák kockázata a legalacsonyabb, a reggeli felfutások és esti lecsengések kockázata pedig relatíve magas, illetve napközben (a csúcsidőszaknak megfelelő órákban) a hétvégék kockázata alacsonyabb, mint a hétköznapiokon. Az itteni ábrák a reziduumok szórása tekintetében egyébként a korábbi szakaszban bemutatottakhoz nagyon hasonló alakot tükröznek. Némi eltérés abból adódhat, hogy – szemben az előző fejezetekben használt megoldásokkal – a keverékmodell képes **nemlineáris** kapcsolatok automatikus, exploratív felismerésére (szemben a korábbi, linearitást feltételező módszerekkel).

A konklúzióhoz még egy fontos megjegyzést érdemes fűzni. Igaz ugyan, hogy az ebben a fejezetben kapott eredmények nem mondanak ellent sem a feltételezéseinknek, sem a fejezet korábbi részeiben kapott eredményeknek abban a tekintetben, hogy a különböző időszakok volumen-kockázata hogyan alakul. Ami ennél sokkal fontosabb az az, hogy a modell, amelyet becsültünk olyan standard hibákat számol, amelyek a reziduumok szórásával konzisztensek. A reziduumok szórását illetve a standard hibák átlagát az ábrákon úgy kell tehát vizsgálni, hogy a görbék alakja hasonló kell, hogy legyen (az y tengelyek terjedelme azonos, ami mindenképpen segíti az összehasonlítást). A keverék-regresszió haszna többek között ebben a **konzisztenciában** és a vele járó modell alapú megközelítésben érhető tetten.¹⁴⁵

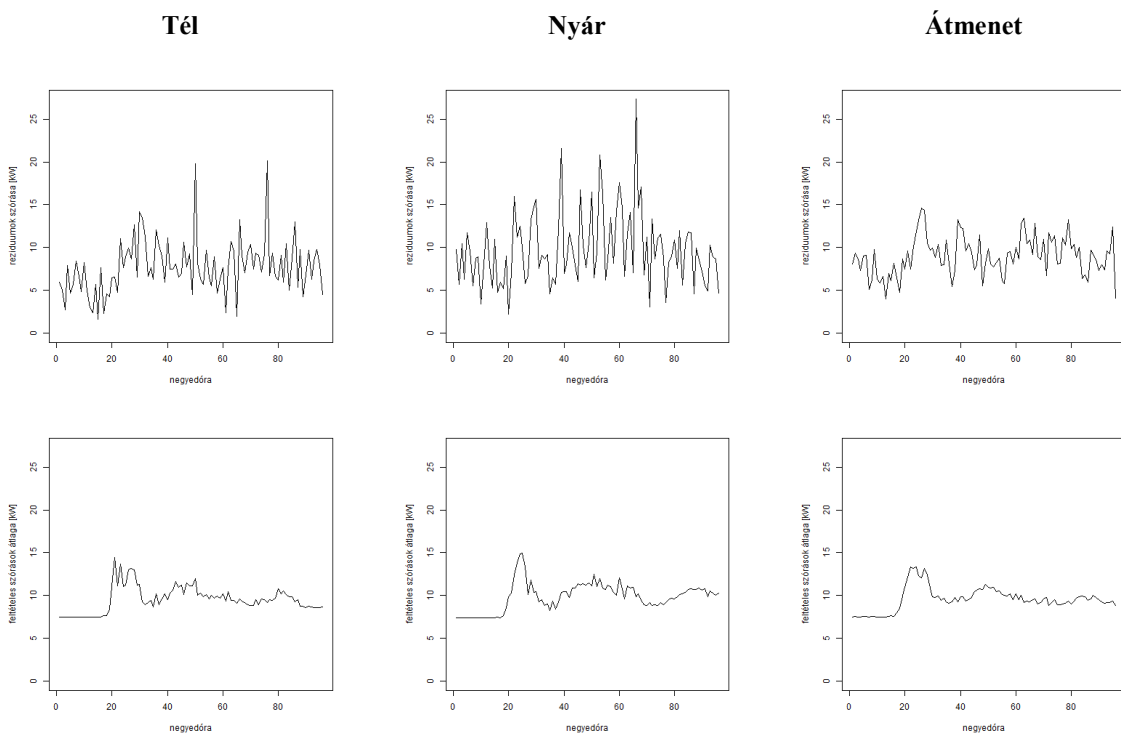
¹⁴⁵ Az 54-55. ábrákon egyébként csak a mintán belüli eredmények szerepelnek, a mintán belüli és mintán kívüli eredmények között gyakorlatilag nincsen különbség.

54. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a vállalati portfólió esetén (hétköznapok)



Forrás: saját számítások (R) és saját készítésű ábra (R).

55. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a fogyasztói portfólió esetén (hétfői napok)



Forrás: saját számítások (R) és saját készítésű ábra (R).

4.2.2.2. Egyedi görbék volumenkockázata és összehasonlító elemzése

A 17. táblázatban néhány egyedi fogyasztói görbére vonatkozóan szerepelnek az illeszkedést jellemző mutatók. Nincsen tehát jelentős különbség a klasszikus regressziós illetve a keverék-modellek becslései között illeszkedés tekintetében, sőt, sok esetben a keverék-regresszió még valamivel jobban is teljesít.

17. táblázat: A modellek illeszkedését jellemző mutatószámok alakulása az egyedi görbék esetén

| Görbe | Modell | Predikció | RMSE [kW] | MAPE [%] | MAE [kW] |
|-------|------------|--------------|-----------|----------|----------|
| V25 | ARMA(1, 0) | Mintán belül | 0,93 | 9,11 | 0,65 |
| | GMR | Mintán belül | 0,93 | 8,40 | 0,65 |
| | | Mintán kívül | 1,01 | 8,51 | 0,69 |
| V35 | ARMA(1, 0) | Mintán belül | 0,66 | 4,62 | 0,36 |
| | GMR | Mintán belül | 0,62 | 4,40 | 0,36 |
| | | Mintán kívül | 0,59 | 4,35 | 0,34 |
| V66 | ARMA(1, 0) | Mintán belül | 0,40 | 5,06 | 0,24 |
| | GMR | Mintán belül | 0,36 | 4,43 | 0,22 |
| | | Mintán kívül | 0,38 | 4,48 | 0,22 |
| V79 | ARMA(1, 0) | Mintán belül | 0,39 | 19,90 | 0,48 |
| | GMR | Mintán belül | 0,64 | 16,58 | 0,41 |
| | | Mintán kívül | 0,64 | 16,63 | 0,41 |
| V109 | ARMA(1, 0) | Mintán belül | 1,64 | 9,22 | 1,23 |
| | GMR | Mintán belül | 1,59 | 8,60 | 1,17 |
| | | Mintán kívül | 1,64 | 8,95 | 1,21 |
| V1 | ARMA(1, 0) | Mintán belül | 0,10 | 36,55 | 0,05 |
| | GMR | Mintán belül | 0,10 | 34,77 | 0,05 |
| | | Mintán kívül | 0,10 | 33,78 | 0,05 |
| V4 | ARMA(1, 0) | Mintán belül | 0,17 | 11,15 | 0,11 |
| | GMR | Mintán belül | 0,18 | 11,55 | 0,11 |
| | | Mintán kívül | 0,17 | 11,48 | 0,11 |
| V27 | ARMA(1, 0) | Mintán belül | 0,25 | 11,36 | 0,18 |
| | GMR | Mintán belül | 0,25 | 11,18 | 0,18 |
| | | Mintán kívül | 0,25 | 11,32 | 0,18 |
| V108 | ARMA(1, 0) | Mintán belül | 1,62 | 11,09 | 1,23 |
| | GMR | Mintán belül | 1,54 | 10,25 | 1,17 |
| | | Mintán kívül | 1,59 | 10,32 | 1,20 |
| V47 | ARMA(1, 0) | Mintán belül | 0,18 | 10,32 | 0,11 |
| | GMR | Mintán belül | 0,19 | 10,13 | 0,11 |
| | | Mintán kívül | 0,19 | 10,84 | 0,11 |

Forrás: saját számítások (R) és saját készítésű táblázat.

Sokkal fontosabb azonban a 18. táblázat alapján leszűrhető következtetés. A konfidencia-intervallumon kívül eső megfigyelések aránya itt is eltér némileg az 5%-tól, nagyságrendileg a SARMA-modelleknél számoltakhoz hasonlóak. Azonban a későbbi

ábrákon látni fogjuk, hogy a konfidencia-intervallumon kívül esés sokkal egyenletesebben oszlik el – hasonlóan ahhoz, amit a vállalati portfólió esetében tapasztaltunk.

18. táblázat: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya és átlagos konfidencia-intervallum nagysága az egyedi görbék esetén

| Görbe | Modell | Predikció | Arány [%] | Átlagos CI95 [kW] |
|-------|------------|--------------|-----------|-------------------|
| V25 | ARMA(1, 0) | Mintán belül | 6,29 | 3,64 |
| | GMR | Mintán belül | 7,96 | 3,13 |
| | | Mintán kívül | 7,87 | 3,32 |
| V35 | ARMA(1, 0) | Mintán belül | 4,41 | 2,59 |
| | GMR | Mintán belül | 7,12 | 1,78 |
| | | Mintán kívül | 7,09 | 1,67 |
| V66 | ARMA(1, 0) | Mintán belül | 7,89 | 1,55 |
| | GMR | Mintán belül | 7,23 | 0,99 |
| | | Mintán kívül | 7,72 | 1,00 |
| V79 | ARMA(1, 0) | Mintán belül | 7,90 | 2,62 |
| | GMR | Mintán belül | 6,35 | 1,92 |
| | | Mintán kívül | 7,07 | 1,87 |
| V109 | ARMA(1, 0) | Mintán belül | 5,72 | 6,45 |
| | GMR | Mintán belül | 6,33 | 5,61 |
| | | Mintán kívül | 7,04 | 5,64 |
| V1 | ARMA(1, 0) | Mintán belül | 6,99 | 0,38 |
| | GMR | Mintán belül | 6,70 | 0,23 |
| | | Mintán kívül | 6,69 | 0,24 |
| V4 | ARMA(1, 0) | Mintán belül | 5,79 | 0,68 |
| | GMR | Mintán belül | 7,91 | 0,56 |
| | | Mintán kívül | 7,66 | 0,56 |
| V27 | ARMA(1, 0) | Mintán belül | 5,95 | 0,97 |
| | GMR | Mintán belül | 6,77 | 0,87 |
| | | Mintán kívül | 7,04 | 0,85 |
| V108 | ARMA(1, 0) | Mintán belül | 5,28 | 6,35 |
| | GMR | Mintán belül | 7,50 | 5,46 |
| | | Mintán kívül | 7,90 | 5,48 |
| V47 | ARMA(1, 0) | Mintán belül | 6,11 | 0,72 |
| | GMR | Mintán belül | 7,20 | 0,55 |
| | | Mintán kívül | 7,44 | 0,55 |

Forrás: saját számítások (R) és saját készítésű táblázat.

Gyakorlati szempontból sokkal fontosabb eredmény az, hogy a keverék-regresszióknak köszönhetően sokkal kisebb az konfidencia-intervallum átlagos mérete. A 18. táblázat utolsó oszlopában szerepelnek ezek az értékek. Már korábban, a vállalati portfóliónál is feltüntettük ezeket, azonban itt sokkal szemléletesebb tárgyalni. SAR-

MA-modellek esetén a 95%-os megbízhatósági intervallum az ismert módon az alábbi szerint adódik:

$$\hat{y} \pm 1.96 \cdot s_{\hat{y}},$$

ahol \hat{y} a pontbecslés értéke, $s_{\hat{y}}$ a standard hiba, $t_{1-\frac{\alpha}{2}}(\infty) = 1.96$ pedig a megbízhatósági szorzó értéke nagy (∞ -nel jelölt) mintaelemszám esetén. A konfidencia-intervallum szélessége ilyenkor a szimmetriából adódóan $2 \cdot 1.96 \cdot s_{\hat{y}}$, és nyilvánvalóan az állandó szóródás feltételezése végett a konfidencia-intervallum átlagos szélessége is ennyi lesz mindenhol.

Keverék-modell esetén még csak nem is közvetlenül a standard hibával dolgozunk (hiszen a hiba nem is feltétlenül normális eloszlású és a konfidencia-intervallum nem is feltétlenül szimmetrikus minden esetben, azaz a keverék tulajdonság a hibára is öröklődik). A módszertani fejezetben (3.2.5. alfejezet) leírt módon számolt intervallum határok különbségeként kaphatjuk meg a konfidencia-intervallum szélességét, azaz az alábbi módon:

$$\hat{y}_u - \hat{y}_L,$$

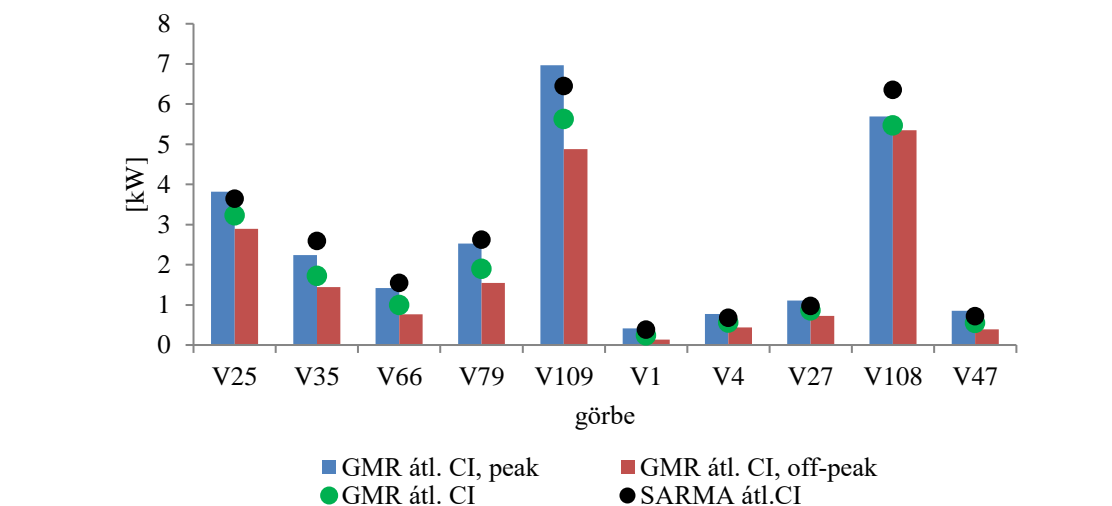
ahol \hat{y}_u a konfidencia-intervallum felső (*upper*), \hat{y}_L pedig a konfidencia-intervallum alsó (*lower*) határát jelöli. Ezek a különbségek időpontról időpontra eltérőek, ezek minden időpontra történő átlagolásából adódnak a táblázat utolsó oszlopában lévő eredmények a keverék-modell esetében.

Vannak olyan esetek (V35, V66, V79, V1 illetve V47), ahol a konfidencia-intervallum átlagos szélessége több mint 25%-kal csökken, de a többi esetben is tapasztalható több mint 10%-os csökkenés. A **konfidencia-intervallum szélességét átlagában** praktikus bemutatni, gyakorlati haszna egy görbe vizsgált intervallum alatti lefutása szempontjából **összességében** is jelentős. Ez azt jelenti ugyanis, hogy nagyságrendekkel kisebb intervallumot tudunk adni egy görbe napi, heti, havi stb. alakulását tekintve, amely intervallumba a tény értékek adott megbízhatósággal esnek.

Az 56. ábra grafikusan is szemlélteti az átlagos konfidencia-intervallumok alakulását. Mivel a kereskedési gyakorlatban a hétköznapi-hétféle megkülönböztetéshez képest a csúcs-völgy (*peak-off-peak*) megkülönböztetés valamivel kézenfekvőbb, ezért az oszlopok a *peak* illetve *off-peak* időszakok átlagos intervallum méreteit mutatják a keverék-regresszió esetében. (SARMA-modell esetében a konfidencia-intervallum szélessége nyilván időszaktól függetlenül mindig ugyanakkora.) Mivel ezek az eredmények (pl. az *off-peak* kockázata akár fele is lehet a *peak* kockázatának) a későbbi 62-71. ábrákról

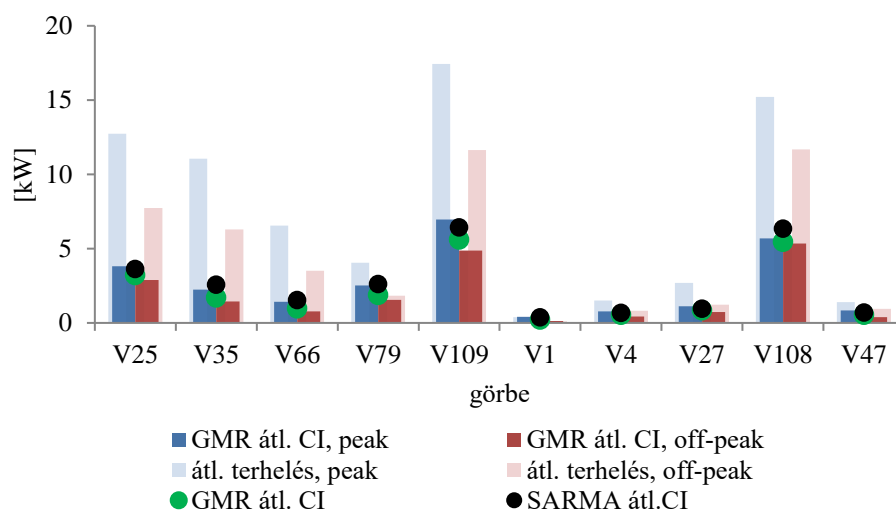
negyedórás szinten is jól visszaellenőrizhetőek, azok mellett igazán hasznos információval az 57. ábra szolgál még, ahol ezek az intervallum-méretetek a görbék ugyanazon időszaki átlagos terheléseivel vannak párba állítva.

56. ábra: Átlagos konfidencia-intervallum SARMA- és keverék-modellek esetén az egyedi görbék esetén¹⁴⁶



Forrás: saját számítások (R) és saját készítésű ábra (Excel).

57. ábra: Átlagos konfidencia-intervallum SARMA- és keverék-modellek esetén az egyedi görbék esetén (kibővített ábra)



Forrás: saját számítások (R) és saját készítésű ábra (Excel).

¹⁴⁶ A CI itt és az 57. ábra esetében a 95%-os megbízhatósági szintű intervallum rövid jelölésére szolgál.

A korábbi fejezet nyomán azért érdemes megvizsgálni, hogy egy-egy görbe esetén a SARMA-modell hibái hogyan alakultak, és mennyiben tudott ehhez alkalmazkodni a keverék-modell a heteroszkedaszticitás konzisztensebb modellezése által. Az 58-61. ábrákon megtekinthetők a konfidencia-intervallumon kívül eső megfigyelések arányai mindkét esetben. Néhány kiemelt görbénél a reziduumok szórásai illetve a feltételes szórások átlagai az 62-71. ábrákon hasonló, azonos tendenciákat írnak le minden esetben.¹⁴⁷ Az ábrákon a feltételes szórások alakulása jóval – még a viszonylag kis mintaelemszám ellenére is – simább a reziduumok szórásához képest minden esetben, hiszen a feltételes szórások számított értékek. A modellalapúság előnye tehát mindenképpen jól látszik, akár a heurisztikus módszerhez, akár valamilyen, a realizálódott reziduumokon alapuló megoldáshoz (pl. a korábban már említett kNN-módszert) képest.

Néhány fontosabb eredményt érdemes kiemelni – a teljesség igénye nélkül – az egyedi görbék tekintetében.

A V25 és a V35 esetében jól látszik például, hogy a hétköznapiok-hétvégék, és a kora hajnali órák esetén értünk el látványos javulást abban a tekintetben, hogy a reziduumok konfidencia-intervallumon kívül esése egyenletesebb. A V25 standard hibáiban jól látszik emellett az is, hogy a téli illetve a nyári hőmérsékletváltozás a bizonytalanság napon belüli alakulására is eltérően hat (a téli inkább reggel, a nyári inkább a délutáni órákban).

A V66 valószínűleg egy „szigorú” munkarend szerint működő üzem, ahol csak hétköznapiokon, csúcsidőben fordult elő a konfidencia-intervallumon kívülre esés a SARMA-modell esetében, amit a keverék-modell nagyon szépen kezelt. Emellett a nyári (egyébként inkább extrémnek mondható) hőmérsékletfüggés valamivel magasabb standard hibát eredményez a nyári csúcsidőszakban.

Különleges a V79, hiszen ennek a görbének a nyári viselkedése nagyon eltér a többi időszaktól, elsősorban éppen a variancia tekintetében, és a konfidencia-intervallumon kívül kerülés alapvető forrása ez volt a SARMA-modell esetében, a „különleges” módon különböző kovariancia-struktúra felismerésének az eredménye itt tehát nagyon jól látszik.

A V109 szintén nyári hőmérsékletváltozással és emellett napközbeni két csúcsfogyasztással is jellemezhető görbe. Ez a két tulajdonság tükröződik a SARMA-modell

¹⁴⁷ Csak a mintán belüli eredmények szerepelnek itt is, mert a mintán belüli és a mintán kívüli eredmények között elhanyagolható a különbség.

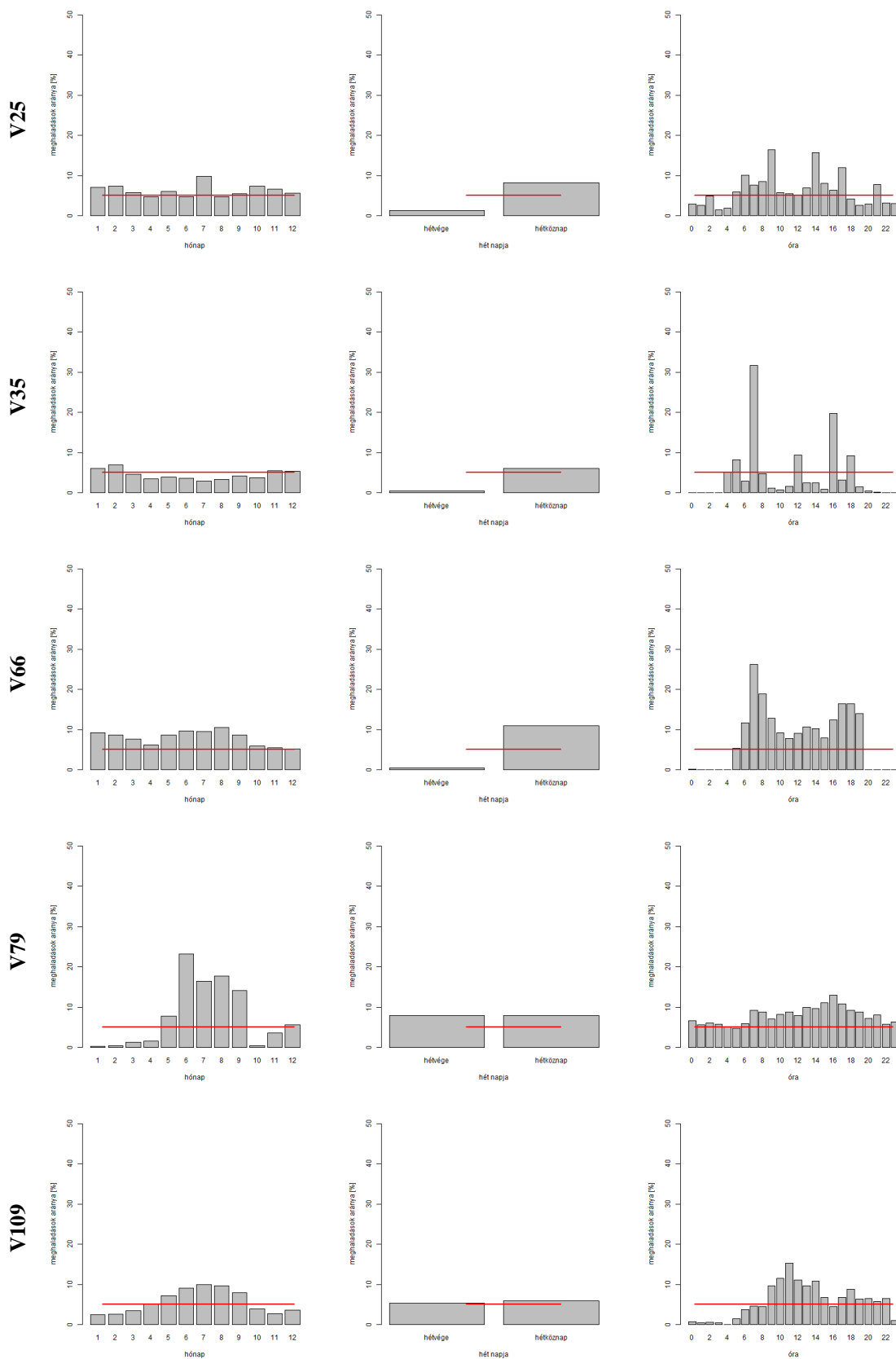
reziduumainak a viselkedésében, illetve a feltételes szórások átlagában is (a második csúcs egyébként jóval enyhébb).

A V1 esetében is nagyon jól észrevehető a téli hőmérsékletfüggés okozta magasabb bizonytalanság a SARMA-modell reziduumainak alakulásában, itt minden bizonnyal a keverék-modell magasabb standard hibákat becsül.

A V27, mint a hőmérséklethatással gyakorlatilag egyáltalán nem terhelt görbe esetében csak a csúcs- illetve völgyidőszakok bizonytalanságának eltérő nagyságrendjei láthatóak, a V108 esetében pedig a viszonylag korai reggeli illetve késő esti csúcs nagyobb bizonytalansága említésre méltó.

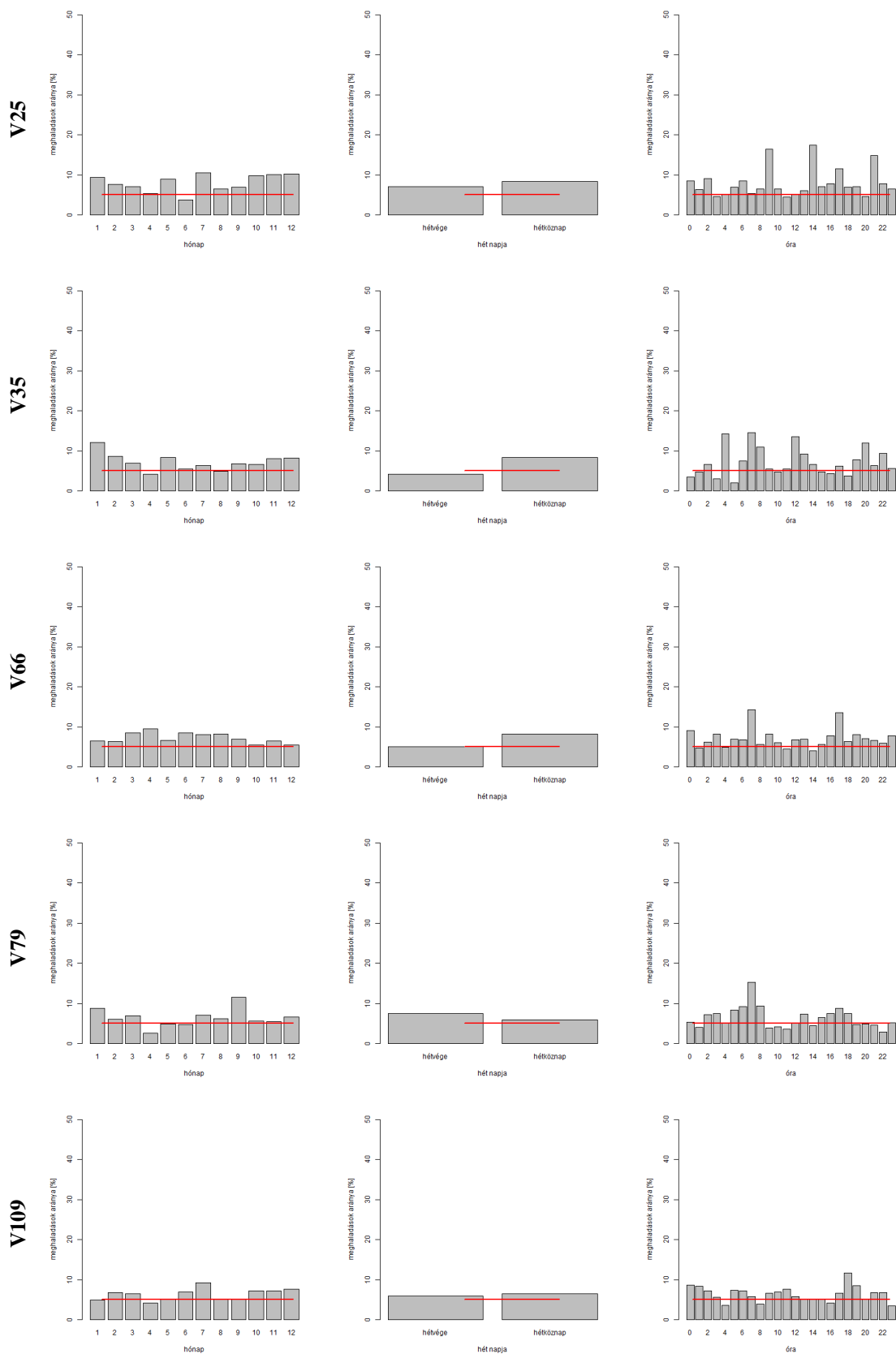
A V47 esetében mind a hőmérsékletfüggés, mind a napon belüli alak bizonytalansága jól tükröződik a standard hibák alakulásában.

58. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya az egyedi görbék esetében (SARMA-modell)



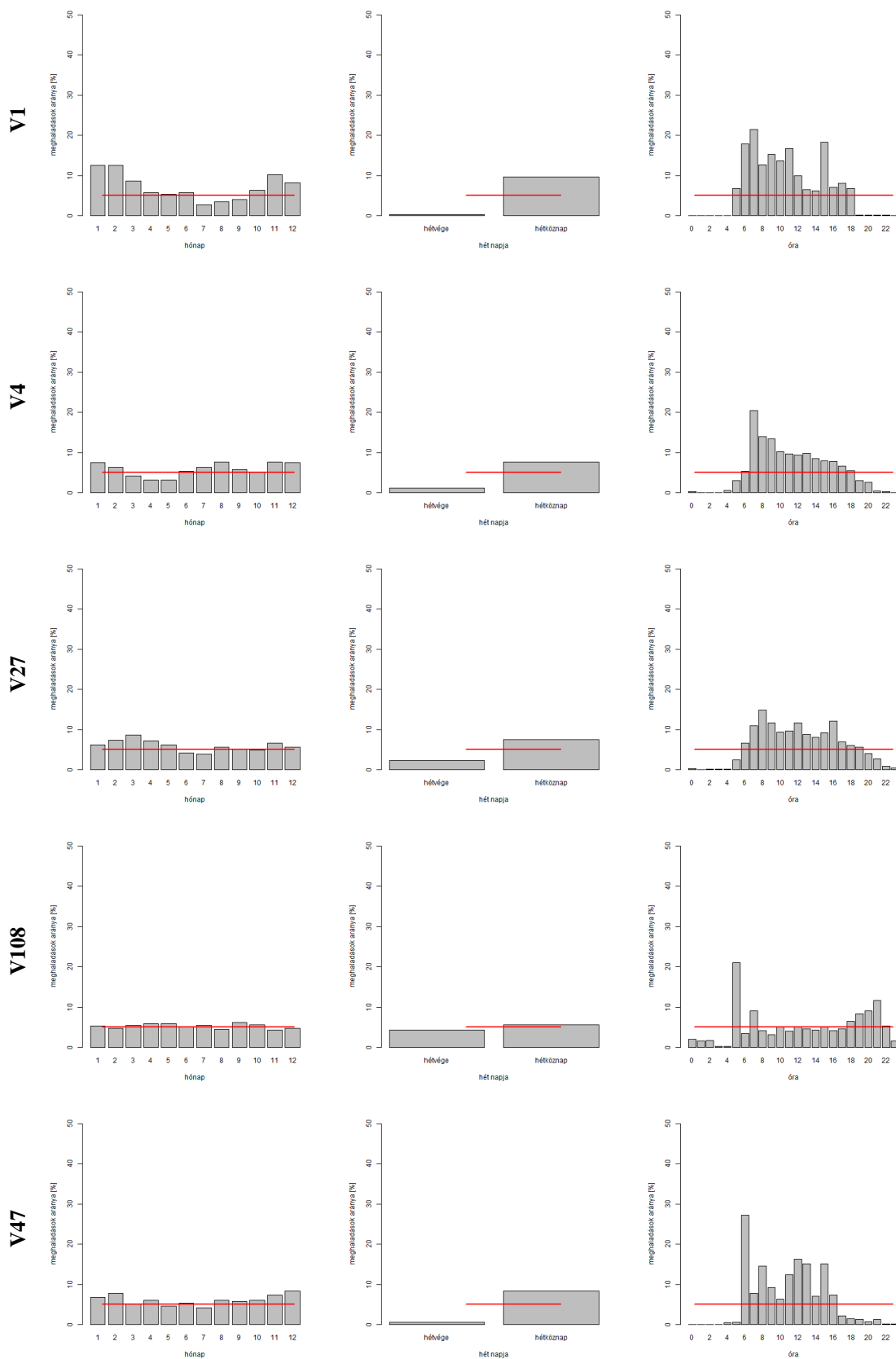
Forrás: saját számítások (R) és saját készítésű ábra (R).

59. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya az egyedi görbék esetében (keverék-regresszió)



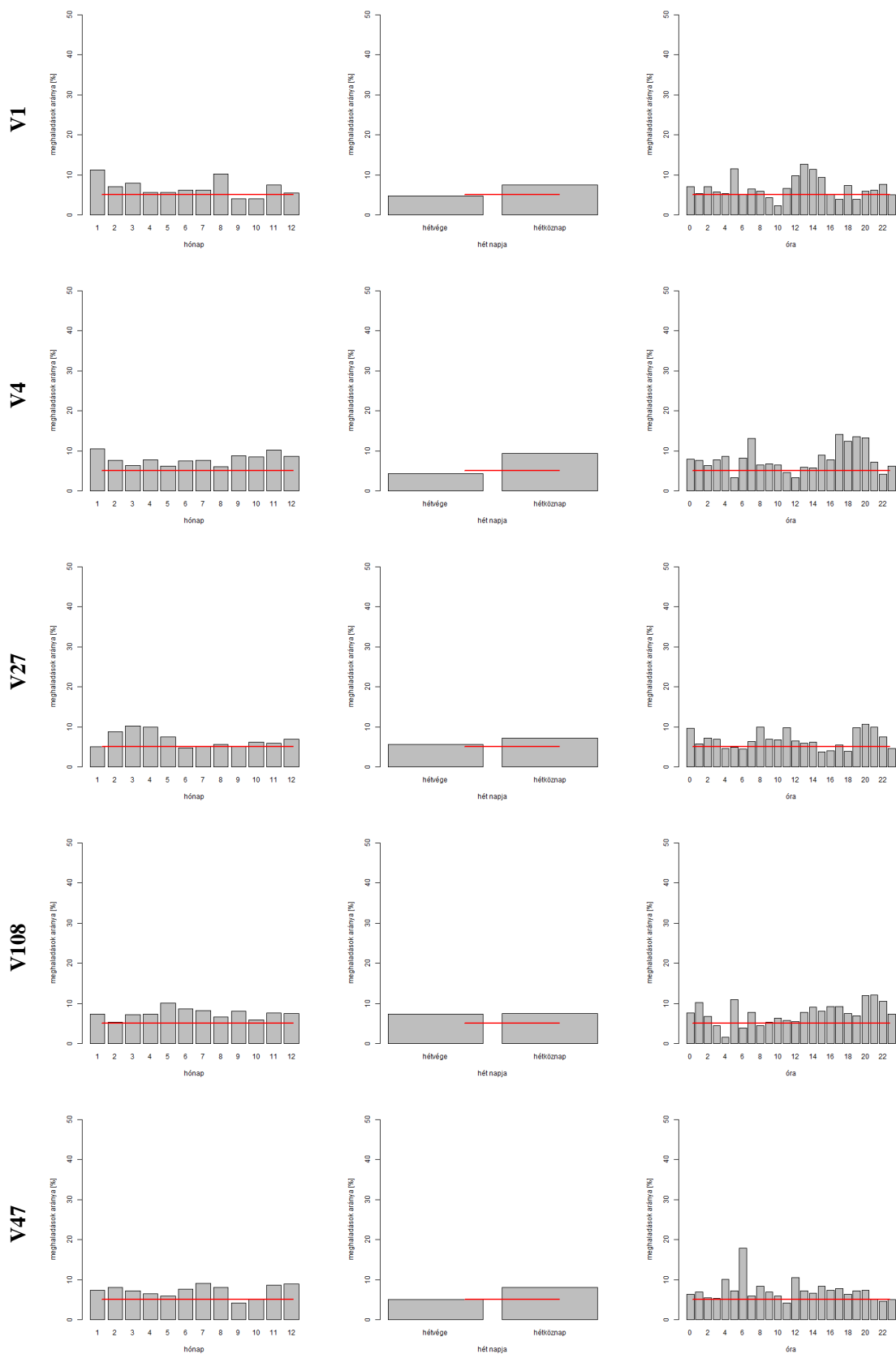
Forrás: saját számítások (R) és saját készítésű ábra (R).

60. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya az egyedi görbék esetében (SARMA-modell)



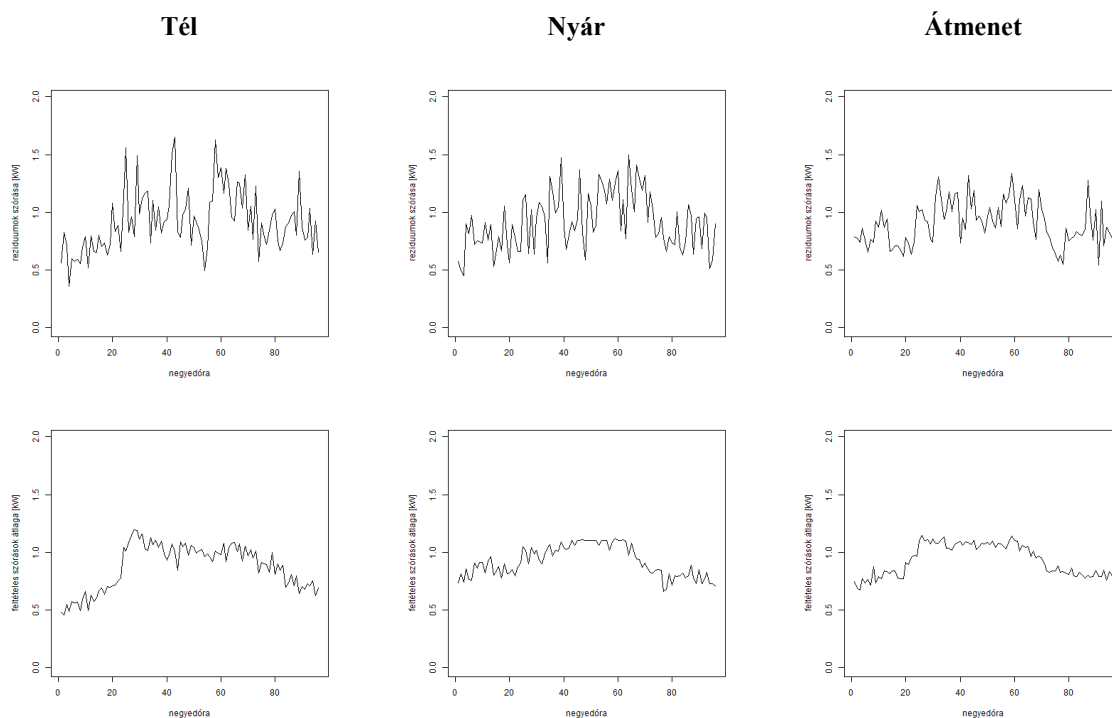
Forrás: saját számítások (R) és saját készítésű ábra (R).

61. ábra: Konfidencia-intervallumon (CI95) kívül eső megfigyelések aránya az egyedi görbék esetében (keverék-regresszió)



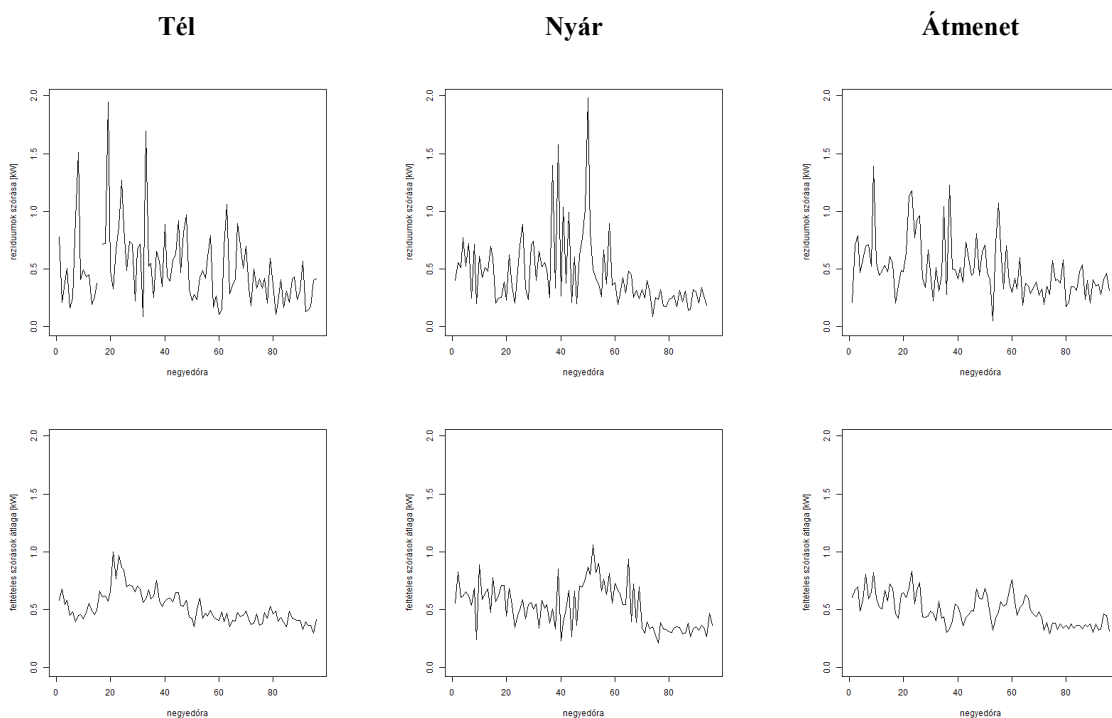
Forrás: saját számítások (R) és saját készítésű ábra (R).

62. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V25 görbe esetében (hétköznapi)



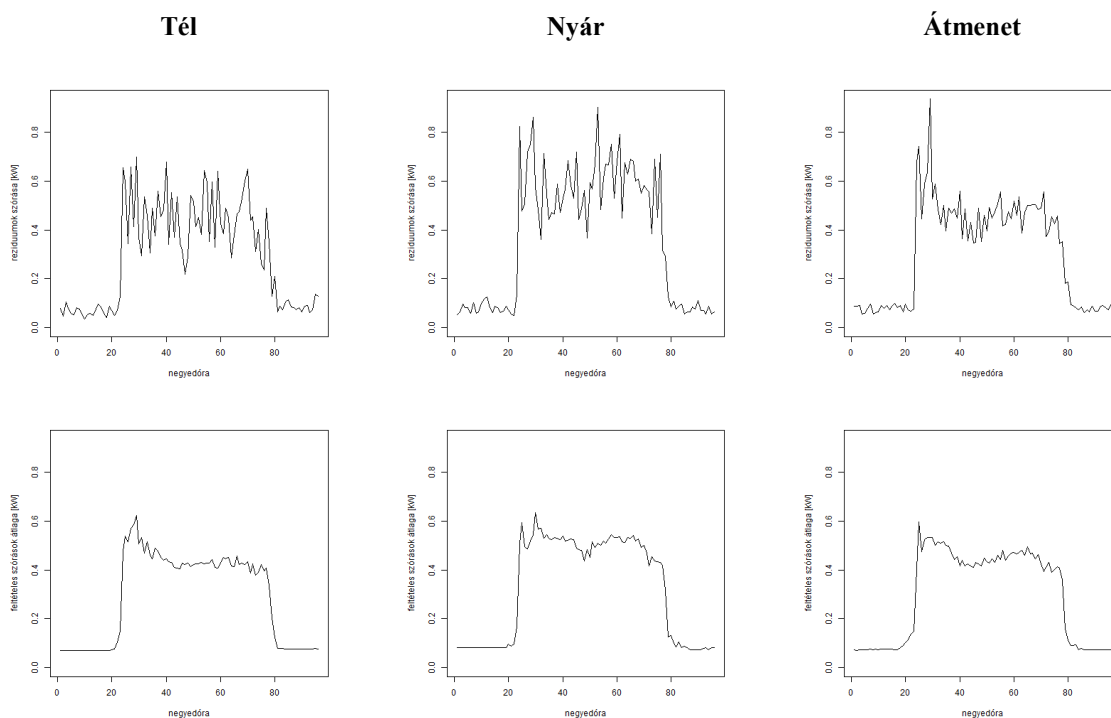
Forrás: saját számítások (R) és saját készítésű ábra (R).

63. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V25 görbe esetében (hétvégi napok)



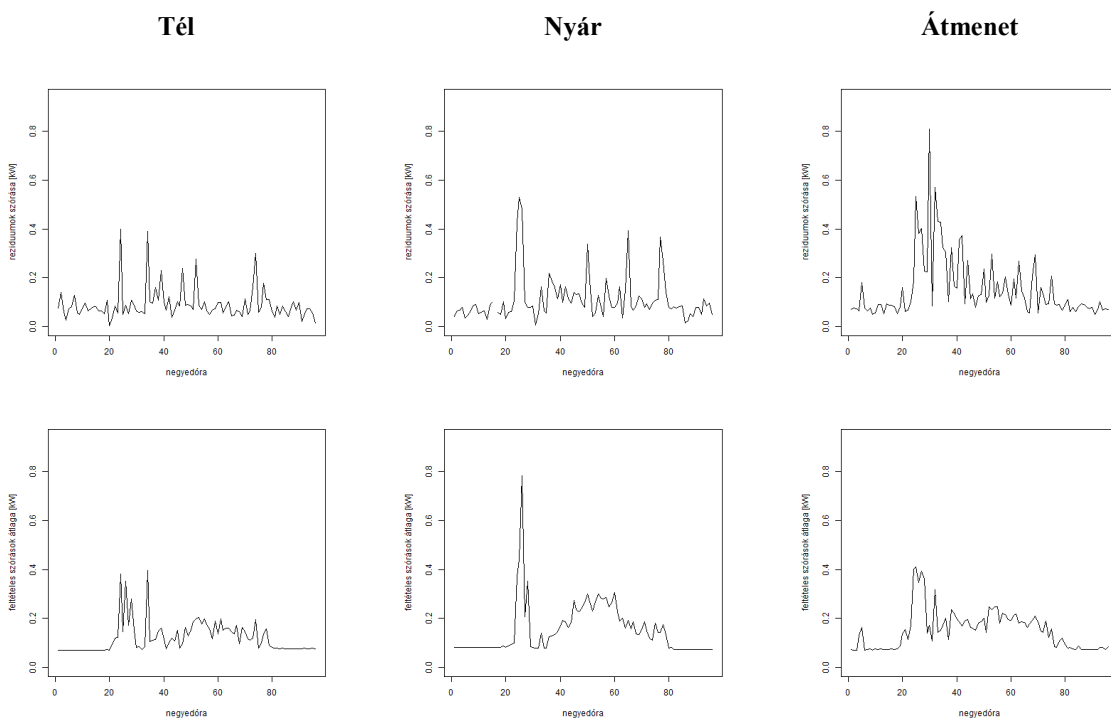
Forrás: saját számítások (R) és saját készítésű ábra (R).

64. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V66 görbe esetében (hétköznapi napok)



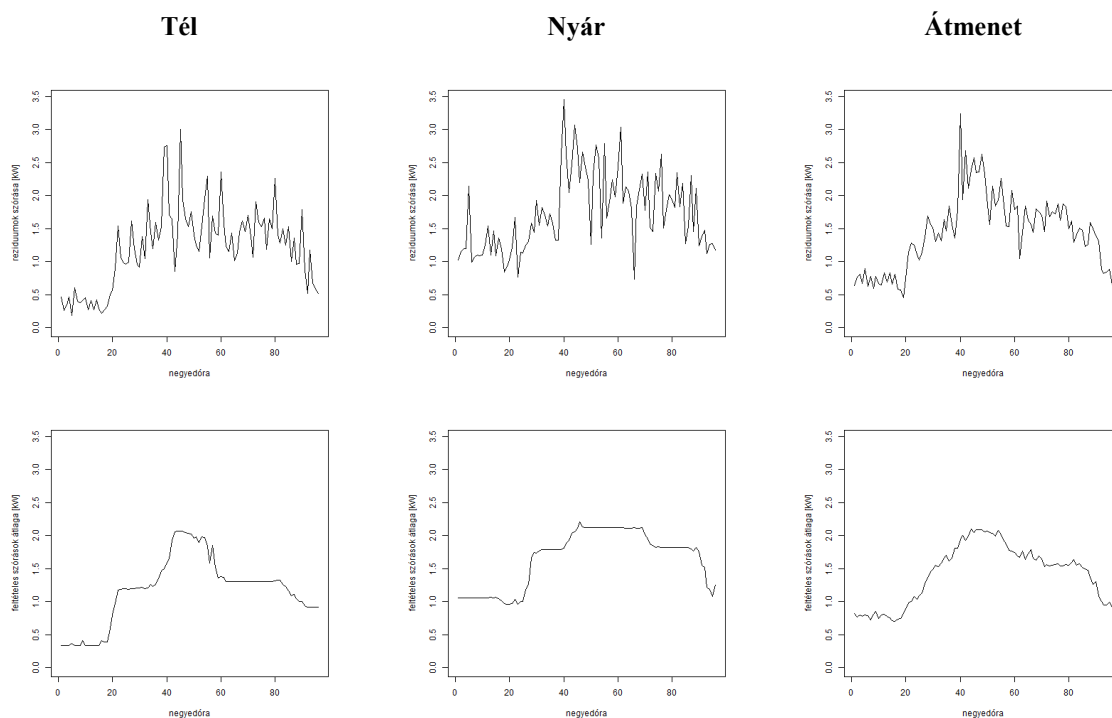
Forrás: saját számítások (R) és saját készítésű ábra (R).

65. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V66 görbe esetében (hétvégi napok)



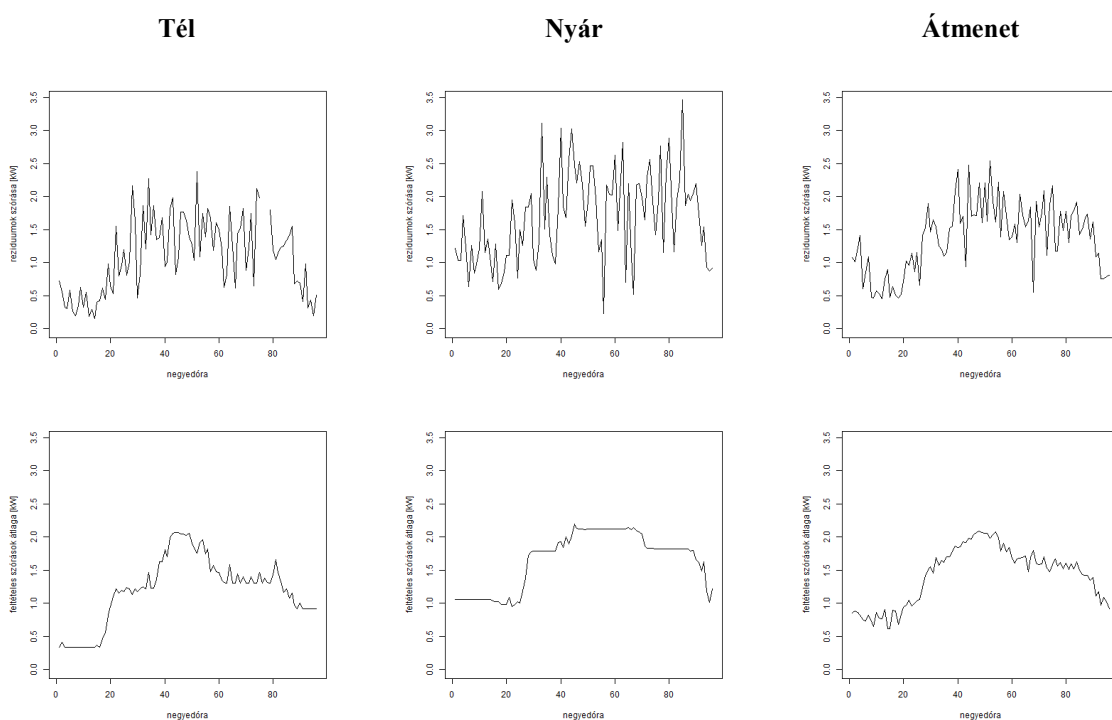
Forrás: saját számítások (R) és saját készítésű ábra (R).

66. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V109 görbe esetében (hétköznapi napok)



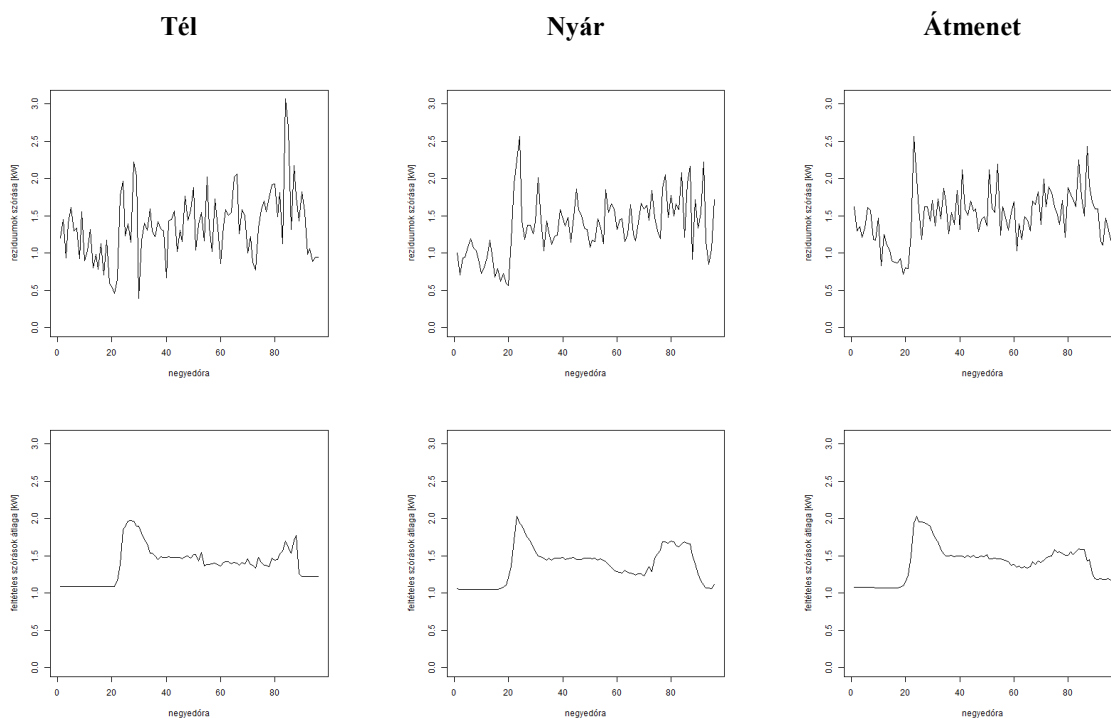
Forrás: saját számítások (R) és saját készítésű ábra (R).

67. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V109 görbe esetében (hétvégi napok)



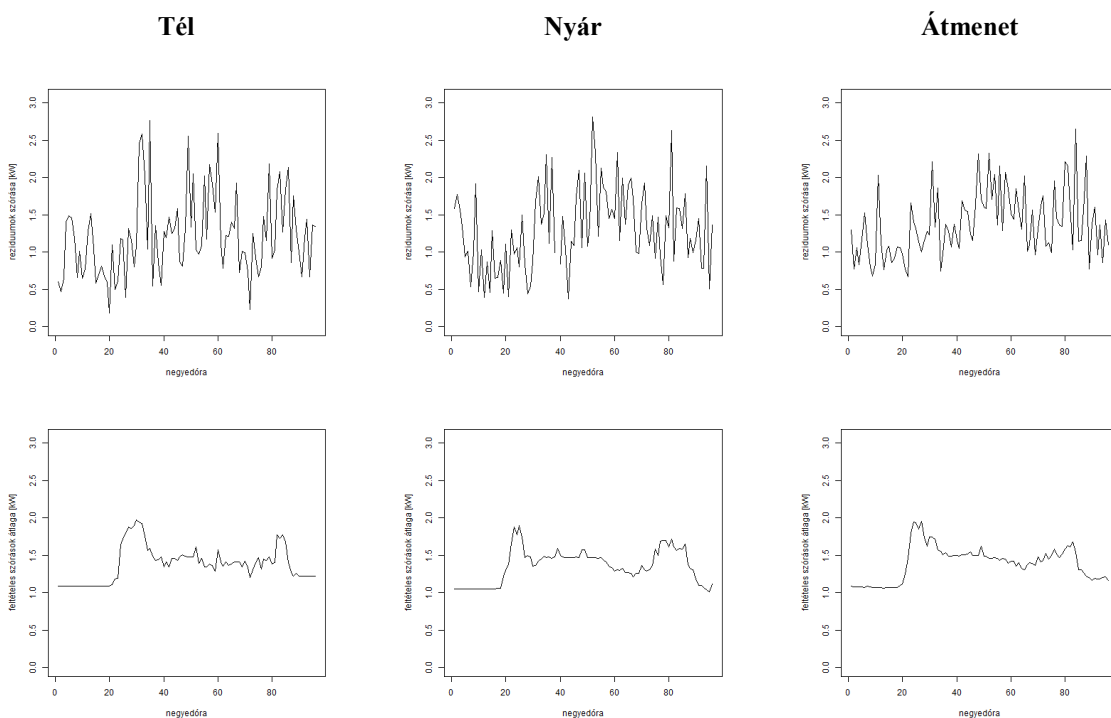
Forrás: saját számítások (R) és saját készítésű ábra (R).

68. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V108 görbe esetében (hétköznapok)



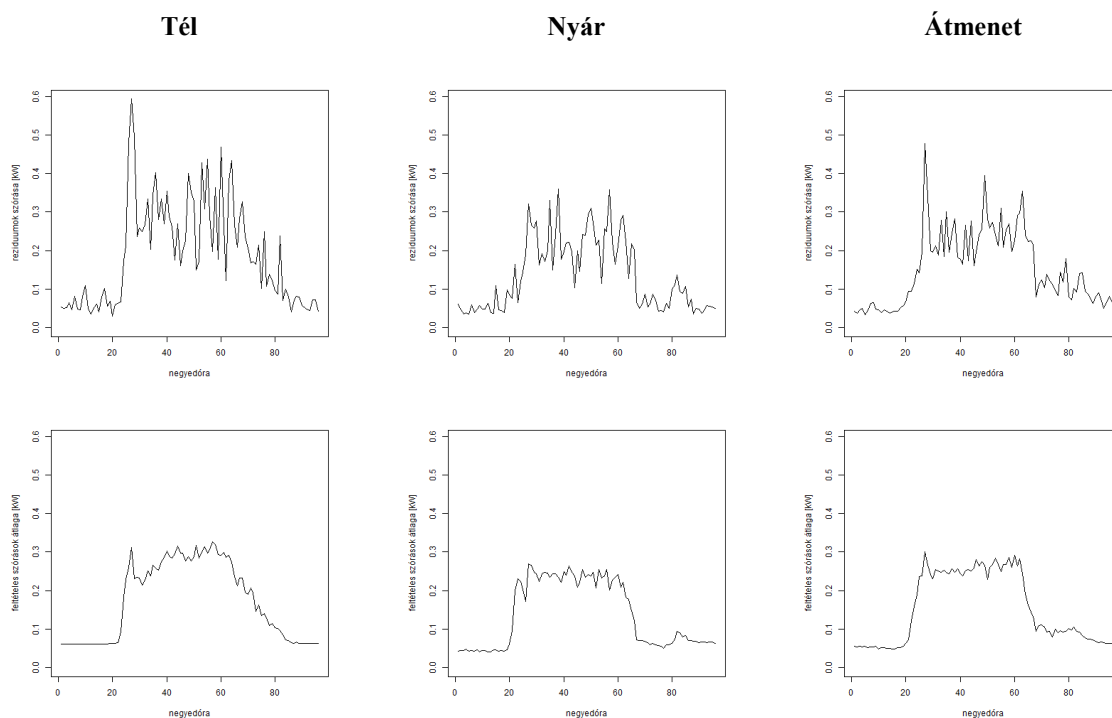
Forrás: saját számítások (R) és saját készítésű ábra (R).

69. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V108 görbe esetében (hétvégi napok)



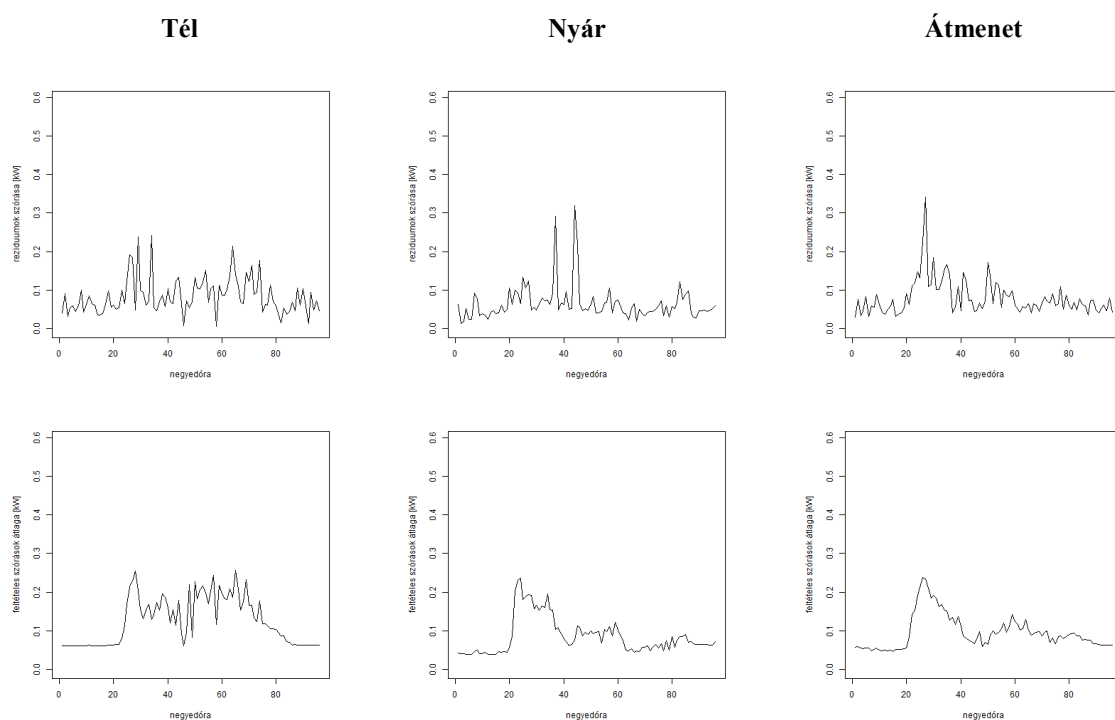
Forrás: saját számítások (R) és saját készítésű ábra (R).

70. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V47 görbe esetében (hétköznapok)



Forrás: saját számítások (R) és saját készítésű ábra (R).

71. ábra: Keverék-regresszió reziduumaik szórása és a feltételes szórások átlaga a mintában a V47 görbe esetében (hétvégi napok)



Forrás: saját számítások (R) és saját készítésű ábra (R).

4.2.2.3. *Eredmények, következtetések összegzése*

Ebben a szakaszban az egyedi görbék esetében is megvizsgáltuk, hogy a klasszikus idősoros regressziós megoldások által adott konstans standard hiba milyen mértékben inkonzisztens az empirikus tapasztalatokkal. Az előző szakaszban is bemutatott logika mellett ellenőriztük, hogy különböző időszakokban mekkora a hibák szórása, illetve mekkora a konfidencia-intervallumon kívül eső megfigyelések aránya. Amennyiben a homoszkedaszticitás feltételezése helyes, 95%-os megbízhatósági szinten minden esetben a megfigyelések 5%-ának kellene a konfidencia-intervallumon kívül szerepelni (kismértékű eltérések természetesen megengedettek). Az eredmények egybehangzóan azt mutatják, hogy bizonyos időszakokban a konfidencia-intervallumon kívül eső megfigyelések aránya tendenciaszerűen, jelentősen meghaladja az 5%-ot, míg más időszakokban jóval elmarad attól. A viselkedésben azonban fellelhető egyértelműen egy szabályos, szezonális tendencia, ami görbénként különböző ugyan, de elmondhatjuk, hogy nemcsak a fogyasztás karakterisztikáját, hanem annak bizonytalanságát, szóródását is a görbét egyedileg jellemző többszintű szezonális viselkedés írja le.

Ellenőriztük, hogy a keverék-regresszió által becsült standard hibák és konfidencia-intervallumok a hibákkal konzisztensen alakulnak-e. Tapasztalataink alapján a bemutatott módszer alkalmas az idősorokban lévő heteroszkedasztikus viselkedés megragadására, és a konfidencia-intervallumon kívül eső megfigyelések aránya sokkal egyenletesebben szóródik a megbízhatósági szint által indikált arány körül. A keverék-regresszió által becsült hibák némileg eltérnek a klasszikus megoldások által produkált hibáktól (az illeszkedésben általában egyébként nincsenek jelentős eltérések), azonban ez valószínűleg abból adódik, hogy a keverék-regresszió képes a nemlineáris kapcsolatok megragadására.

Mindezek alapján sem a H3, sem a H4 hipotézist nem tudjuk elvetni. Azaz a H4 hipotézis alapján a volumenkockázat időben nem állandó, a kockázat különböző exogén változók, szezonális és naptári hatások függvényében időben változó, azaz jellemzően nagyobb:

- **hétköznapokon, mint hétvégéken és ünnepnapokon,**
- **a hétköznapi csúcsidőszakban, mint azon kívül, és**
- **azokban az időszakokban, amikor a fogyasztás időjárásfüggő.**

Bár utólag nyilvánvalónak tűnik, de nem várt eredmény volt az, hogy nagyon sok görbe esetében a reggeli felfutás vagy az esti lecsengés kockázata a legnagyobb.

A H3 hipotézis alapján pedig igaz az, hogy a hibák állandó szóródásának feltevése mellett bizonyos időszakokban a volumenkockázatot alul-, bizonyos időszakokban felülbecsüljük. Ehhez kapcsolódóan kedvező eredmény volt az, hogy helyes (heteroszkedasztikus) modell becslése mellett összességben vagy átlagában nagyságrendekkel szűkebb konfidencia-intervallumot tudunk mondani egy-egy görbe alakulására vonatkozóan. Az empirikus tapasztalat is az egyébként, hogy a klasszikus idősoros regresszió standard hibái olyan nagyok és olyan széles konfidencia-intervallumot eredményeznek, amik a gyakorlatban az idősorok egészét tekintve is szinte semmitmondóak és használhatatlanok.

A standard hibák alakulásában rendre felismerhető volt a napon belüli csúcsidőszakok elhelyezkedése, illetve jellemzően az is, hogy a téli illetve a nyári hőmérséklet-hatás akár a reggeli felfutások, akár a napközbeni csúcsidőszakok bizonytalanságára hogyan hat. A standard hibák nagyságrendjét tekintve (akár összevontan, ld. 56-57. ábrák; akár szezonálisan különválasztva vizsgáljuk, ld. 62-71. ábrák) elmondható, hogy görbétől függően a völgyidőszakok bizonytalansága sokszor fele vagy még kisebb a csúcsidőszakok bizonytalanságához képest. Ezen völgyidőszakok bizonytalanságát összevetve a klasszikus idősoros regresszió által becsült eredményekkel a nagyságrendek közötti különbség még inkább szemléletes.

ÖSSZEFOGLALÁS – AZ ÉRTEKEZÉS FŐBB MEGÁLLAPÍTÁSAI, EREDMÉNYEI

Az értekezésben a fogyasztói profilok, illetve a fogyasztás alakulásában a véletlen hatásból származó volumenkockázat modellezésével foglalkoztunk. A fejezetben összefoglaljuk, hogy az empirikus eredmények kapcsán milyen következtetések fogalmazhatóak meg.

A) Fogyasztási idősorok stilizált tényeinek vizsgálata

Több egyedi görbét is megvizsgáltunk, hogy melyek azok a tulajdonságok, amelyek egy-egy görbe karakterisztikáját meghatározhatják, amelyeket a tipikus fogyasztási mintázatnak a meghatározásakor is figyelembe kell venni. Ezek a vizsgálatok jellemzően nem klasszikus statisztikai próbák voltak, hanem olyan egyszerűbb számítások vagy ábrák, amelyeket viszonylag ritkán alkalmaznak egy-egy görbe tömör leírásakor, jellemzésekor.

A vizsgálatok és eredményeik röviden az alábbiakban összegezhetők:

- Szintvonalábra (*contour plot*) segítségével megvizsgáltuk, hogy a terhelési értékek **eloszlása** milyen a teljes **év során**, így feltárhatóak többek között az alábbiak:
 - a csúcsidőszaki, csúcsidőszakon kívüli, hétköznapi, hétvégi terhelések szintje, a csúcsidőszakok napon belüli elhelyezkedése hogyan változik az év során,
 - melyik görbék esetében van jelentős hatása az ünnepnapoknak,
 - a hőmérséklet hatásáról is vonhatóak le következtetések,
 - feltárható az is egyértelműen, hogy melyik görbék esetében jelenik meg erősen a naplemente miatti ún. világítási hatás (vagy naplemente-hatás), amelynek ilyen transzparens kimutatására empirikus példát sehol nem találtunk, inkább csak heurisztikákra szoktak hagyatkozni.
- *Scatter plot* (magyarul szórásdiagram) segítségével ellenőriztük a **hőmérséklet-függőséget**, elsősorban azt, hogy az évszakonként vagy a hét egyes napjain ez görbénként mennyiben különböző, illetve a terhelési értékek hogyan **csoportosulnak, tömörülnek** a hőmérséklet függvényében.

- Dobozábrák (*box plot*) segítségével megvizsgáltuk a terhelési értékek **napon belüli eloszlását**, hétköznapokon és hétvégéken, illetve téli, nyári és átmeneti időszakokban; ezzel ellenőriztük, hogy
 - a napon belüli **eloszlás** görbénként mennyire stabil vagy instabil,
 - a napon belüli eloszlást a különböző szezonális tényezők vagy az időjárás hogyan **módosítják**, illetve
 - különböző görbék esetében mikor nagyobb a véletlen hatás szerepe.

A vizsgálatok eredményeit figyelembe véve elmondhatjuk, hogy **a villamosenergia-fogyasztási görbék varianciájából a legnagyobb hányad jellemzően a napon belüli szezonalitással magyarázható, azaz a H1 hipotézist nem tudtuk elvetni**. Ennek alapján arra a megállapításra jutottunk, hogy a tipikus napi profilok képzése – amely a gyakorlatban bevett megoldás – lényegében megfelelő, azonban **a tipikus fogyasztási mintázatok nem feltétlenül a napi bontások mentén képződnek**, és nem feltétlenül a napi bontások mentén a **leghatékonyabb a modellezésük**.

B) Keverék-modell alkalmazása tipikus fogyasztási mintázatok készítésére

A hagyományos megoldásoktól eltérően az értekezésben nem napi görbék klaszterezése történik meg, hanem az idősor bontásának megfelelően negyedórás időpontoké, és a tipikusnak nevezett eredmények is ennek alapján keletkeznek. Az alkalmazott, ún. *Expectation Maximization* eljárással (EM) becsült *Gauss*-féle keverék-modell (*Gaussian Mixture Model*, GMM) esetében azok az időpontok kerülnek egy klaszterbe, amely időpontokban azon klaszter esetén fordulnak elő a változó értékei együttesen a legnagyobb eséllyel. Az értekezés célja kettős: nemcsak a tipikus, jellemző fogyasztási mintázatot kívánjuk megkeresni, hanem a fogyasztással kapcsolatos bizonytalanság számszerűsítése is érdekes. Ezért nagyobb hangsúlyt kap a fogyasztás mellett a fogyasztás késleltetettjeinek illetve a potenciálisan elérhető exogén változóknak az együttlmozgása is.

Az említett modell alapú klaszterezési módszertan több, a profilozás kapcsán felmerülő problémát az ismert megoldásoknál jobban képes kezelni. Így:

- Az egyedi fogyasztói szokások tipizálását úgy végeztük el, hogy **nem** volt szükség a fogyasztási idősorok **előigazítására**, többek között például:
 - a kiugró értékek szűrésére, vagy
 - a(z extrém) hőmérséklet vagy egyéb időjárási hatásnak a kiszűrésére.

- A megoldás lényegében egy **többváltozós** megoldást jelent, ami – a gyakorlatban és tudományos munkákban sokszor alkalmazott megoldásoktól eltérően – a fogyasztási idősor értékeit nem önmagában csoportosítja, hanem egyéb időbeliséget vagy szezonalitást leíró változókkal (pl. időjárás) együtt. Így természetesen további exogén változókkal való bővítés lehetősége is fennáll. Ez a lehetőség a legtöbb, profilozással foglalkozó módszer esetén fel sem merül igazán, többek között – például – az előigazításokkal kapcsolatos nehézségek miatt.

A keverék-modell konstrukciójából adódóan a módszer előnye, hogy mind a változók közötti interakciós kapcsolatok, mind a linearitás hiánya figyelembe vételre kerül, anélkül, hogy explicit módon definiálnánk ezeket a hatásokat. Ez a klaszterenként eltérő kovariancia mátrix becslésében jelent meg, hiszen ezeknek a hatások szerepeltetése mögött általában az a feltételezés húzódik meg, hogy **a változók kovariancia-struktúrája nem azonos a teljes mintában**. Utóbbi tulajdonság különösen a modell regressziós alkalmazásánál fontos, ugyanis a változók közötti (**nemlineáris** vagy **interakciós**) kapcsolatok és a **heteroszkedaszticitás** együttes kezelése több odafigyelést igényel. Az eredményeink mind a profilozással, mind a bizonytalansággal kapcsolatban meggyőzőek abban a tekintetben, hogy a fogyasztást jellemző stilizált tények fundamentálisan megragadhatóak és számszerűsíthetőek.

Erre egy, az országos **földgázfogyasztás** alakulásával kapcsolatos **korábbi saját eredmény** kapcsán is felhívtuk a figyelmet. Ugyanitt megállapítottuk azt is, hogy a **hőmérséklet teljes hatásának**, illetve a **hőmérsékleti véletlen hatásnak** a kiszűrése is sok előnytelen következménnyel járhat, különösen akkor, ha a fogyasztási idősorok bizonytalanságát is modellezni szeretnénk (előbbi esetben a hőmérsékleti hatás gyakran nem különíthető el, nem választható le megfelelő módon a regressziós dekompozíciós logika alapján, utóbbi esetben pedig a heteroszkedasztikus tulajdonság szűrődik ki). Ráadásul elvi megfontolások is amellet szólnak, hogy a fogyasztás időjárásfüggő részét ne válasszuk le (ugyanis az időjárás a fogyasztás várható értékét nagyban befolyásolja, illetve a bizonytalansággal is összefüggésben lehet), hanem valamilyen többváltozós megoldásban gondolkozzunk.

Bemutattuk, hogy a keverék-modell komponensek paraméterei (a többdimenziós normális eloszlás átlagvektorai, a kovariancia mátrixok) tömörített információként is felfoghatóak, amelyek segítettek több egyedi fogyasztói görbe klaszterezésében, csoportosításában is. Az eredményeket **összehasonlítottuk** egy klasszikusnak tekinthető (reg-

ressziós) megoldással. A távolság méréséhez az ún. **Kullback-Leibler**-távolságmértéket használtuk, ami egyben alkalmas arra is, hogy egy-egy görbe komponenseinek egymástól való távolságát meghatározzuk.

A profilcsoportok képzését a dolgozatban arra használtuk elsősorban, hogy a jobb információtömörítést igazoljuk. Az eredmények azt mutatták, hogy a csoportok képződése sokkal inkább a fogyasztást jellemző fundamentális jellegű tulajdonságok mentén történik, mint többek között a hétköznapi *peak-off-peak* fogyasztás aránya, a hétvégi fogyasztás szintje a hétköznapihoz képest, a hőmérséklet-függőség jellege (utóbbi a szezonális *peak-off-peak* arányt is befolyásolhatja), napon belüli csúcsidezőszakok elhelyezkedése, stb.

A módszer több, **módszertani szempontból is kedvező tulajdonsággal** bír. A klasztereket reprezentáló **tipikus** fogyasztás például természetes módon megkapható a becsült többdimenziós normális eloszlású komponensek móduszaiként (átlagaiként). Ez feloldja azt a sokszor előforduló problémát, hogy a klasztert reprezentáló tipikus érték mi legyen (általában az átlagot szokták használni), ugyanis tartalmilag a módusz a tipikus, jellemző középérték. Ugyanígy inkább módszertani jellegű előny, hogy a keverék-modell nem érzékeny a kis **mintaelemszámra**, ugyanis – modell alapú megoldásról lévén szó – struktúrát ismer fel, amely tulajdonságot egyébként a számítások során ki is használtunk. Előnynek tekinthető még az is, hogy az optimális klaszterszám kiválasztása modellszelekciós kritérium útján, objektív módon történhet.

Ehhez kapcsolódóan kevésbé módszertani jellegű, inkább szemléletbeli különbség, hogy a klasszikus regressziós esetben a különböző kategória-kimeneteket jellemzően 1-0 kimenetű *dummy* változókkal rögzítjük. A keverék-modell esetében ezeknek a szerepét a komponensek (a módszertani fejezetben z_i -vel jelölt, a becslés során p_{ik} *posterior*, komponensbe tartozási valószínűségekkel helyettesített komponenstagságok) veszik át, ennél fogva nemcsak a várható érték, de a szóródás leírásában is kiaknázható a kategória-jellegű információ.

Mindezen eredmények alapján megállapítottuk, hogy a keverék-modellen alapuló profilcsoportképzés a klasszikus megoldásokhoz képest sokkal realisabb eredményt ad, és a keverék-modellek számtalan előnye mellett a csoportképzést nemcsak a várható értéket, de a szóródást, azaz lényegében a bizonytalanságot, kockázatot is figyelembe véve végzi el. Az ezzel kapcsolatos H2 hipotézis így szintén nem került elvetésre.

C) Heurisztikus és klasszikus sztochasztikus idősoros módszerek alkalmazása a fogyasztás bizonytalanságának mérésére

A véletlen görbénkénti viselkedésének ellenőrzésére a következő vizsgálatokat végeztük el. A klasszikus idősoros megoldások, a SARMA- illetve PAR-regresszió alapján számolható standard hibák, illetve a gyártott konfidencia-intervallum alapján megállapítottuk, hogy az állandó szórás feltételezése mellett összességében ugyan közel jól becsüljük egy-egy görbe bizonytalanságát, de bizonyos időszakokban felül-, bizonyos időszakokban alulbecsüljük a kockázatot, így a konstans konfidencia-intervallum feltételezése nem felel meg az empirikus tapasztalatoknak. Ezt úgy vizsgáltuk, hogy ellenőriztük **a 95%-os konfidencia-intervallumon kívül eső megfigyelések arányát**. Amennyiben az intervallum „helyes”, minden hónapban, a hétvégén és hétköznapokon illetve minden (negyed)órán körülbelül a megfigyelések 5%-ának kell az intervallumon kívül esnie (véletlenszerű eltérések természetesen megengedettek). A tapasztalatok alapján azonban görbétől függően ugyan, de jellemzően a csúcsidőszaki, a napon belüli felfutási-lecsengési időszakokban, a hétköznapokon, illetve nyáron és télen ez jóval 5% felett van, egyébként pedig jóval alatta.¹⁴⁸

A klasszikus **idősoros regressziós megoldások hibái (reziduuma)** szórásának vizsgálata alapján az alábbi következtetésre jutottunk a fogyasztás időben változó kockázata tekintetében:

- a csúcsidőszaki fogyasztás bizonytalansága magasabb,
- a csúcsidőszakon kívüli fogyasztás bizonytalansága alacsonyabb,
- nagyon sok görbe esetében a napon belüli reggeli felfutások és nap végi lecsengések bizonytalansága a legnagyobb,
- azokban az időszakokban, amikor a fogyasztás időjárás- (hőmérséklet-) függő, jellemzően a fogyasztás kockázata is nagyobb *ceteris paribus*.¹⁴⁹

A tapasztalatok görbétől függően különbözőek lehetnek, mindenesetre nagyságrendileg teljesen egybecsengenek egyébként a gyakorlatban is alkalmazott heurisztikus mutatószámok (kockázati index) számolásával kapott eredményekkel, a modell alapú

¹⁴⁸ A PAR-regresszió teljesítménye egyébként abban a tekintetben lehet elsőre meglepő, hogy ez a módszer a periódusonként (negyedóránként) változó autoregresszív együtthatók révén periodikusan változó autokovarianciát becsül, ami legalább részben kezelhetné a heteroszkedaszticitást, de a dolgozat eredményei alapján ez az időfüggő autokovariancia nem bizonyult elégséges megoldásnak.

¹⁴⁹ Az időjárás (hőmérséklet) – mint tudjuk – maga is sztochasztikus változó, és annak haszna, hogy a profilozás során az időjárás hatását nem szűrtük, ebben a fontos megállapításban köszön vissza.

megközelítésnek elsősorban a megalapozottságában van az előnye (ld. például kihagyott változó kérdése, időbeli függőség kezelése, stb.).

A felsoroltaknál sokkal komolyabb, megalapozottabb állítások kevésbé fogalmazhatóak meg, aminek részben az az oka, hogy a reziduumok (évszakok, hét napjai alapján történő) „csoportosításának” finomítása semmiképpen nem előreutató, ráadásul a számított eredmények zajos, hektikus jellegéből sokszor csak bizonytalan megállapítások tehetők még az idősoros modellek esetén is.

A H3 hipotézist nem tudtuk tehát elvetni, miszerint időszaktól függően a kockázatot hol alul-, hol felülbecsüljük egy-egy görbe esetén a klasszikusnak mondható, állandó szórást feltételező regressziós megoldások esetén.

D) Keverék-modell alkalmazása a fogyasztás bizonytalanságának mérésére

Az összegzett tapasztalatok alapján a profilkészítés kapcsán is alkalmazott keverék-modell regressziós alkalmazásával olyan, ún. **feltételes, időfüggő standard hibákat**, illetve konfidencia-intervallumokat készítettünk, amelyek a fogyasztás kockázatával összhangban vannak.

Ellenőriztük, hogy a keverék-regresszió alapján gyártott konfidencia-intervallumok mennyiben felelnek meg az elvárásoknak (a 95%-os konfidencia-intervallumok a keverék-regresszió esetén is elkészültek), vagy máshogyan fogalmazva a hibák szórása (pl. órákra, hétköznapiakra/hétvégékre, hónapokra számolva) mennyiben cseng egybe a keverék-regresszió alapján kalkulált standard hibákkal.

Az eredmények alapján a keverék-regresszió jól képes reprezentálni a fogyasztás időben változó bizonytalanságát (a konfidencia-intervallumon kívül eső megfigyelések aránya sokkal egyenletesebb, mint a klasszikus modellek esetében), és jellemzően nagyságrendileg egybe esnek azokkal a várakozásokkal, amiket a heurisztikus mérőszámok és a SARMA-modell hibatagjai alapján meg tudunk fogalmazni. Az eltérések forrása ebben a tekintetben alapvetően az lehet, hogy a SARMA-modell lineáris, míg a keverék-regresszió nem, így a nemlineáris kapcsolatok jobb leképezése révén a hibák kapcsán kaphatunk némileg eltérő eredményeket.

A keverék-regresszió alkalmazásának előnye, hogy a standard hibák a **magyarázóváltozók függvényeként, a magyarázóváltozók feltétele mellett** írhatóak fel, így a hibákban is meglévő szezonális viselkedést ugyanazokkal a változókkal írtuk fel, mint amellyel magának a fogyasztási görbének is a szezonális viselkedését. A standard hibák nemcsak azt tükrözték vissza, hogy napon belül, héten belül vagy az év során mely idő-

szakok bizonytalansága nagyobb, de azt is, hogy görbénként eltérő mértékben ugyan, de a téli hőmérséklet hatás inkább a reggeli, míg a nyári hőmérséklet hatás inkább a délutáni időszakok bizonytalanságát növeli.

A klasszikus- illetve keverék-regressziós számítások alapján tehát **a H4 hipotézist el tudtuk fogadni, azaz valóban igaz, hogy a fogyasztás kockázata időben jellemzően nem állandó, hétköznapokon, csúcsidőszakban és időjárásfüggő időszakokban jellemzően magasabb; azaz alapvetően hasonlóan többszintű szezonális jellemzi, mint magát a fogyasztást is.**

Az **eredmények fontossága** abban rejlik többek között, hogy

- profil és a fogyasztás bizonytalansága (volumenkockázat modellezése) egy modellkereten belül készülhet el,
- a keverék-regresszió alkalmazása jó, ígéretes eredményekkel szolgál, és energiapiaci alkalmazása viszonylag újnak tekinthető,
- a keverék-modell ilyen regressziós célú alkalmazásai esetén a hibák *backtest*-je egy tanulmányban sem jelent meg,
- maga a regressziós alkalmazás (és a klaszterezés bizonyos lépései) az alkalmazott *R Project* program(csomag)ban nem szerepel(nek) közvetlenül, így annak elkészítése is a kutatás részét képezte.

A megfogalmazott hipotézisekben tárgyaltakon túl további fontos eredmény, hogy **a keverék-regresszió által készített átlagos konfidencia-intervallum szélessége görbéről görbére változó mértékben ugyan, de jóval kisebb, mint amit a klasszikus megoldások esetén kaptunk.** Utóbbi eredmény azért is hasznos, mert a klasszikus idősoros modellek esetében készülő konfidencia-intervallum a legtöbb esetben nagyon széles, gyakorlati felhasználásra nem igazán megfelelő. A fejezetben megvizsgáltuk azt is, hogy a konfidencia-intervallum átlagos szélessége hogyan változik *peak* illetve *off-peak* időszakokban, összevetve ugyanezen időszakok átlagos terheléseivel.

Az átlagolás miatt csak nagyságrendi megállapítások tehetők, de minden bizonnyal az áralakulást is hasonló bizonytalansági karakterisztika jellemzi. Ez mindenképpen a fogyasztói befolyásolás kiemelt céljaira, azok szükségességére hívja fel a figyelmet, gondoljunk például a fogyasztási görbe kisimításából, a fogyasztás-átcsoportosításból származó csúcs-völgy arány vagy a tény-terv eltérésekből adódó kiegyenlítő energia költségének csökkentésére.

Az eredmények értékeléséhez mindenképpen érdemes hozzáfűzni azt, hogy a használt módszerek nemcsak annak vizsgálatát és mérését tették lehetővé, hogy mikor nagyobb a fogyasztás bizonytalansága, hanem azt is, hogy pontosan mennyivel. A potenciális gyakorlati (akár klasszikus, akár például fogyasztói befolyásolással kapcsolatos) felhasználások tekintetében utóbbi sem nélkülözhető.

TOVÁBBFEJLESZTÉSI LEHETŐSÉGEK, AZ EREDMÉNYEK HASZNOSÍTHATÓSÁGA

A dolgozat empirikus részében több helyen is utaltunk az eredmények gyakorlati hasznosíthatóságára, illetve ezekhez kapcsolódóan a további kutatási irányokra. Ezeket összegezzük ebben a fejezetben.

Lehetséges további kutatási lehetőség például a keverék-modellen alapuló profilsoportok létrehozásának vizsgálata több (száz vagy ezer) görbével, és összehasonlító elemzések végzése a klasszikus megoldásokkal. A dolgozatban a hangsúly sokkal inkább a bizonytalansággal kapcsolatos értékelésen, mint egy ilyen nagyobb volumenű elemzésen van, ezért mindenképpen érdemes ellenőrizni azt, hogy azáltal, hogy a keverék-modell lényegében a görbe alakulásában megjelenő teljes információt tömöríti, mennyiben kapunk eltérő profilsoport besorolást. **A hangsúly módszertani – és gyakorlati szempontból is – a jobb információ-tömörítés és a bizonytalanság figyelembevételének kiaknázásán van.** Egy ilyen kiterjesztett vizsgálat során természetesen felmerülhetnek olyan, a csoportképzés során általában felmerülő kérdések, mint az optimális klaszterszám megválasztása, a klaszterezés eredményének megfelelőségét vizsgáló mutatószámok vizsgálata, stb. Ezek a kérdések ilyen nagy mennyiségű adathalmaz kezelése esetén szükségszerűen vizsgálandók.

A gyakorlatban sokszor probléma az is, hogy egyes fogyasztói görbék esetén nem áll rendelkezésre egy teljes éves idősor sem. Mivel a keverék-modell a **kis mintaelemszámra** kevésbé érzékeny, ezért – reális határokon belül – érdemes megvizsgálni, hogy jobb teljesítményt nyújt-e, használhatóbb megoldást produkál-e a mintaelemszámigényesebb megoldásokhoz képest olyan esetekben, amikor csak tört évre van információnk.

Igaz ugyan, hogy a keverék-regresszió a reziduumokkal konzisztens standard hibákat adott, azonban globálisan nézve így is több a konfidencia-intervallumon kívül eső megfigyelések aránya, mint amekkorát a megbízhatósági szint alapján várunk (de mint láttuk, nagyságrendileg a SARMA-modellek teljesítménye is hasonló). Érdemes megvizsgálni, hogy amennyiben a normális eloszlás helyett **más eloszlások keverékével** dolgozunk, az javít-e az eredményeken. Az, hogy a konfidencia-intervallumon kívül eső megfigyelések aránya nagyobb, mint amit a megbízhatósági szint indokol, valamilyen vastag szélű eloszlás szükségességére utal. Ennek ellenőrzése és ilyen mennyiségű és

heterogén görbehalmaz esetén egy általános, jól alkalmazható megoldás keresése mindenképpen izgalmas kutatási feladatot jelent.

Egy további lehetséges kutatási irány – akár az előbb említett jelenségnek a kezelésére is – a hőmérsékleten kívüli **további időjárási (esetleg más) változók** bevonása az elemzésbe. Időjárási változók tekintetében természetesen nélkülözhetetlen azok megfelelő minősége, hiszen a szakirodalom ezek tekintetében még annyira sem egységes, mint a hőmérséklet esetében (sokszor a kapcsolat meglétét tekintve sem). Más kérdés, hogy mivel a hőmérséklet hatása messze a legerősebb, ezért például hosszabb távú tervezésnél vagy egy fogyasztó teljes éves fogyasztásának vizsgálata során a hőmérséklet elegendő lehet, és a további változók bevonásának akkor lehet igazi hozadéka, ha például elkülönítetten mért energiafelhasználást szeretnénk modellezni. Mindez természetesen megfelelő műszaki infrastruktúrát is igényel – például az adatrögzítés gyakoriságát illetően a fogyasztási és az exogén változók esetében is).

Az értekezés során sokat hangsúlyoztuk, hogy a „klasszikus” profilozási megoldásoknak a hőmérsékleti hatásoktól bármilyen formában megtisztított görbék esetén kicsit kevesebb a mozgástere, különösen, ha figyelembe vesszük azt, hogy a **fogyasztás időjárásfüggés** esetén például **kevésbé befolyásolható**, kevésbé árrugalmas. Ezért az eredmények az időjárásfüggő bizonytalanság mérése tekintetében mindenképpen hasznos kiindulópontot jelenthetnek a kapcsolódó kutatások számára.

Ami a jövőben mindenképpen ígéretes továbbfejlesztési lehetőség lehet, az a **portfólióhatás** vizsgálata, különös tekintettel a véletlenek összefüggésének modellezése tekintetében. Ennek egy közelítő becslése lehet, ha különböző időszakokra lineáris korrelációs együtthatókat számolunk. Az F) mellékletben is szereplő rövid példa alapján valószínűsíthető, hogy a volumenkockázat diverzifikálhatóságának mértéke időben változó, hiszen a reziduumok korrelációja¹⁵⁰ is időfüggő. A számolás mindenesetre elképzelhető a keverék-modell keretrendszerébe illesztve is. Mivel a keverék-modell a változók különböző kovariancia-struktúrával rendelkező komponenseit becsüli, és ez a hibákra is transzformálódik, ez nagyban megkönnyítheti a hibák együttmozgásnak, kovarianciájának, lényegében a portfólióhatásnak a modellezését egyetlen lépésben.

Minden statisztikai modell esetén fontos a mintán kívüli teljesítmény értékelésére. Ebben a tekintetben a dolgozatban korlátozott volt a lehetőségünk, hiszen csak egy éves görbék álltak rendelkezésre. Az egy időszakra előre történő, statikus előrejelzések érté-

¹⁵⁰ Sőt, annak szignifikanciája vagy annak hiánya.

kelése lényegében ebben a dolgozatban is megtörtént, így különösen érdekes terület a több időszakra előre történő, dinamikus **előrejelzések** készítése és értékelése.

Vannak a fentiek mellett olyan potenciális kutatási területek is, amelyek a dolgozat fókuszától némileg eltérnek, azonban mindenképpen érdemes róluk említést tenni. A korábbi kutatási eredményekkel foglalkozó fejezetben szerepelt olyan (egyébként empirikusan is reprodukált megoldás), ahol egy-egy napi görbét normális eloszlású sűrűségfüggvények keverékeként modelleztünk. A módszer potenciálisan alkalmas lehet a napon belüli **csúcs időpontjának a becslésére** is. Ennek napjainkban fontos szerepe van, hiszen nagyon sok olyan körülmény, tendencia van (például elektromos autók elterjedése), amelyek ha nagyobb volument öltenek, teljes mértékben átalakíthatják a rendszer szintű napi profilt is a napon belüli csúcsok eltolódásával – mind időben, mind nagyságrendben.

Érdemes megjegyezni, hogy a dolgozat első hipotézisének megfogalmazását éppen az a megfigyelés indukálta, hogy a profilozás során jellemzően napi profil görbék alapján dolgozunk. Mivel a görbék varianciájának nagy hányadát a napon belüli szezonális teszi ki, ezért ezek a megoldások nem adnak annyira félrevezető eredményeket villamosenergia-görbék esetében. Érdemes éppen ezért megvizsgálni **más energianemeket** (pl. földgáz, ahol sok fogyasztónál a fűtési hatás dominál) nemcsak a profilozás, de a volumenkövetés szempontjából is, ahol a variancia nagy hányadát nem a napon belüli szezonális dominálja, hanem sokkal inkább az időjárás. Itt a keverék-modellen alapuló profilozásnak akár még az itt bemutatottaknál is nagyobb hozzáadott értéke lehet a klasszikus megoldásokhoz képest.

Hasonlóan érdekes terület a keresleti oldal mellett a **kínálati** (villamosenergia-piac esetében lényegében az erőművi) **oldal** bizonytalanságának a vizsgálata. Itt a nehézséget sokszor az jelenti, hogy az időjárásfüggő termelők esetében a termelés helyén mért, lokális időjárási (szélsebesség, napsugárzás, felhőtakaró nagysága, pára stb.) adatokra van szükség, viszont a klasszikus meteorológiai adatszolgáltatások információi sokszor még közelítőleg sem bizonyulnak megfelelőnek. Ugyanakkor a linearitás hiánya vagy a változók közötti interakciós kapcsolatok fundamentális feltárása és ezzel párhuzamos a bizonytalanság számszerűsítése (ld. például a termelői menetredek megbízhatóságának az értékelését) itt is igény, és a kettő együttes modellezésére a kínálati oldalon is nagyon kevés példa van.

Ehhez kapcsolódóan szintén fontos kérdés az, hogy a **kereslet- illetve a kínálat** oldal mind profilban, mind a profil bizonytalansága tekintetében hogyan **hangolhatóak**

össze. Különösen itt érdemes kiemelni a háztartási okos mérés jövőben egyre inkább várható terjedését, ahol a majdan rendelkezésre álló adatmennyiség feldolgozása – a kisfogyasztói viselkedés pontosabb ismeretével – a háztartási méretű energiatermelési projektek értékelésében is hasznos adalékkal szolgál majd.

MELLÉKLETEK

A) A számításokhoz használt programcsomagok illetve fontosabb funkciók

A dolgozatban használt számítások nagy része az **R Project** statisztikai programcsomag felhasználásával készült. Ebben a fejezetben áttekintő jelleggel összefoglaljuk, hogy milyen *package*-eket, funkciókat használtunk. Ez fontos abból a szempontból, hogy a dolgozat módszertani íve jobban látszódik, illetve vannak olyan elemei is a számításoknak, amelyek elvégzéséhez szükséges számításokat, utasításokat külön függvényben el kellett készíteni (ilyen például a keverék-modellre épülő regresszió). Vannak olyan számítási eredmények is a dolgozatban, amelyek nem az R felhasználásával készültek, ezeket itt is külön kiemeljük.

A dolgozat jelentős hányadát kitevő empirikus kutatási eredmények között használt legfontosabb módszertanok és az azok megvalósításához használt *package*-k az alábbiak (4. fejezet):

- a SARMA-regressziók (*seasonal autoregressive moving average regression*) a 'stats' *package*¹⁵¹ felhasználásával készültek,¹⁵²
- a PAR-regressziók (*periodic autoregressive regression*) a 'partsm' *package*¹⁵³ felhasználásával készültek,
- az EM-eljárással (*Expectation-Maximization*) becsült Gauss-féle keverék modellek az 'mclust' *package*¹⁵⁴ felhasználásával készültek,
- a Gauss-féle keverék-modellen alapuló EDDA nevezetű diszkriminanciaanalízis utasításai az 'mclust' által adott eredményekre épülően saját készítésű függvények, ugyanis az R *package*-ben ez nem szerepel beépített függvényként,
- a Gauss-féle keverék-modellen alapuló regressziók utasításai az 'mclust' által adott eredményekre épülően saját készítésű függvények, ugyanis az R *package*-ben regressziós alkalmazási célú függvények nem szerepelnek.

¹⁵¹ <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>

¹⁵² Néhány esetben – a kutatás korábbi fázisaiban – a SARMA-regressziók készítése során a **Gretl** (<http://gretl.sourceforge.net/>), szintén nyílt forráskódú, elsősorban ökonometriai, regressziós alkalmazásokhoz kiváló szoftvert is használtuk.

¹⁵³ <https://cran.r-project.org/web/packages/partsm/index.html>

¹⁵⁴ <https://cran.r-project.org/web/packages/mclust/index.html>

További, az empirikus kutatási eredmények között kisebb súllyal szereplő vagy a korábbi kutatási, szakirodalmi eredmények egyszerűbb reprezentálása, bemutatása során használt funkciók:

- a keverék-sűrűségfüggvények illesztése (3.3. alfejezet) az *'mclust'* package felhasználásával történt,
- k-középpontú klaszterezés (3.2.3. szakasz) és hierarchikus klaszterezés (4.1. fejezet) a *'stats'* package felhasználásával történt,
- a szintvonalábrák (1.4. alfejezet) a *'graphics'* package¹⁵⁵ felhasználásával készültek,
- a dobozábrák (1.4. alfejezet) a *'ggplot2'* package¹⁵⁶ felhasználásával készültek,
- hierarchikus klaszterezés során a Kullback-Leibler-távolságmérték számítása a *'monomvn'* package¹⁵⁷ felhasználásával történt, a normális eloszlások keverékére történő közelítés saját készítésű függvény *Hershey-Olsen* [2007] alapján.

A hőmérsékleti véletlen hatás szerepéről (és szezonális kiigazításra való hatásáról) szóló – szintén saját korábbi kutatásból származó – eredmények X13-ARIMA szezonális kiigazítási eljárással készültek, **X13-ARIMA-SEATS**¹⁵⁸ programcsomag felhasználásával (az X13-ARIMA-SEATS programcsomaghoz egyébként létezik a *'seasonal'* nevezetű *R interface package*¹⁵⁹, amelyen keresztül az X13-ARIMA-SEATS funkciói meghívhatók); a heteroszkedaszticitás szemléltetésére használt 27. ábra (3.1.3. szakasz) pedig **EViews**¹⁶⁰ programcsomagban készült SARIMA-regresszió eredménye.

¹⁵⁵ <https://stat.ethz.ch/R-manual/R-devel/library/graphics/html/00Index.html>

¹⁵⁶ <https://cran.r-project.org/web/packages/ggplot2/index.html>

¹⁵⁷ <https://cran.r-project.org/web/packages/monomvn/index.html>

¹⁵⁸ <https://www.census.gov/srd/www/x13as/>

¹⁵⁹ <https://cran.r-project.org/web/packages/seasonal/index.html>

¹⁶⁰ <http://www.eviews.com/home.html>

B) Empirikus példa Budapest napi földgázfogyasztási adatai példáján – számítási részeredmények

A melléklet a Budapest napi földgázfogyasztási adatai példán a 3. fejezetben bemutatott k-középpontú klaszterezés és keverék-klaszterezés néhány fontos részeredményét tartalmazza.

19. táblázat: BSS/TSS hányadosok alakulása k-középpontú klaszterezés esetén a napi középhőmérséklet-gázfogyasztás példán

| klaszterek száma | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------------|------|------|------|------|------|------|------|------|------|
| BSS/TSS (%) | 79,7 | 90,2 | 94,4 | 96,4 | 97,8 | 98,2 | 98,3 | 98,7 | 99,0 |

Forrás: saját számítások (R) és saját készítésű táblázat.

20. táblázat: BIC mutató alakulása keverék-klaszterezés esetén a napi középhőmérséklet-gázfogyasztás példán

| klaszterek száma | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------|----------|----------|----------|-----------------|----------|----------|
| BIC | 17 127,6 | 16 328,3 | 16 221,0 | 16 213,2 | 16 254,4 | 16 304,1 |

Forrás: saját számítások (R) és saját készítésű táblázat.

21. táblázat: Klasztercentroidok keverék-klaszterezés esetén a napi középhőmérséklet-gázfogyasztás példán

| változó / klaszter | 1. (kék) | 2. (fekete) | 3. (piros) | 4. (zöld) |
|--------------------------------------|----------|-------------|------------|-----------|
| hőmérséklet [°C] | -1,45 | 7,57 | 14,89 | 20,31 |
| gázfogyasztás [ezer m ³] | 9 376,74 | 5 889,05 | 2 211,55 | 1 184,04 |

Forrás: saját számítások (R) és saját készítésű táblázat.

22. táblázat: Napok klasztereken belüli megoszlása a különböző hónapok esetén a napi középhőmérséklet-gázfogyasztás példán (keverék-klaszterezés)

| klaszter/hónap | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | Össz. |
|----------------|------|------|------|------|------|------|------|------|------|------|------|------|-------|
| 1. (kék) | 36,4 | 31,8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4,6 | 27,3 | 100,0 |
| 2. (fekete) | 16,4 | 15,0 | 17,9 | 5,0 | 0 | 0 | 0 | 0 | 0 | 7,1 | 20,7 | 17,9 | 100,0 |
| 3. (piros) | 0 | 0 | 7,6 | 29,1 | 26,6 | 0 | 0 | 0 | 10,1 | 26,6 | 0 | 0 | 100,0 |
| 4. (zöld) | 0 | 0 | 0 | 0 | 8,1 | 24,2 | 25,0 | 25,0 | 17,7 | 0 | 0 | 0 | 100,0 |

Forrás: saját számítások (R) és saját készítésű táblázat.

23. táblázat: Napok klasztereken belüli megoszlása a hétköznapok / hétvégék esetén a napi középhőmérséklet-gázfogyasztás példán (keverék-klaszterezés)

| klaszter/hét napja | hétköznap | hétvége | Össz. |
|--------------------|-----------|---------|-------|
| 1. (kék) | 72,7 | 27,3 | 100,0 |
| 2. (fekete) | 74,3 | 25,7 | 100,0 |
| 3. (piros) | 67,1 | 32,9 | 100,0 |
| 4. (zöld) | 71,0 | 29,0 | 100,0 |

Forrás: saját számítások (R) és saját készítésű táblázat.

C) Terhelési idősorok stilizált tényeinek vizsgálata

A melléklet a terhelési görbék stilizált tényeinek vizsgálatáról szóló 1.4. alfejezethez kapcsolódóan tartalmaz néhány kiegészítő számítási eredményt.

A 24. táblázat a magyar országos rendszerterhelés heti idősorainak leíró statisztikáit mutatja napi bontásban, amit a dobozábrákkal párhuzamosan érdemes vizsgálni.

24. táblázat: A magyar rendszerterhelés heti idősorainak leíró statisztikái napi bontásban¹⁶¹

| Mutató | Szezon | H | K | Sze | Cs | P | Szo | V |
|----------------|---------|----------|----------|----------|----------|----------|----------|----------|
| \bar{Y} [MW] | tél | 5 396,45 | 5 501,18 | 5 539,88 | 5 408,05 | 5 340,32 | 5 124,29 | 4 887,36 |
| | nyár | 5 151,63 | 5 314,92 | 5 142,26 | 4 837,76 | 4 929,93 | 4 476,53 | 4 230,46 |
| | átmenet | 4 899,53 | 4 883,30 | 4 881,34 | 4 869,16 | 4 795,88 | 4 384,30 | 3 827,43 |
| Q_1 [MW] | tél | 4 821,07 | 5 043,25 | 5 015,85 | 4 891,91 | 4 914,03 | 4 796,97 | 4 488,83 |
| | nyár | 4 443,42 | 4 731,33 | 4 759,73 | 4 519,87 | 4 586,45 | 4 237,23 | 3 861,42 |
| | átmenet | 4 480,57 | 4 707,46 | 4 663,91 | 4 601,99 | 4 561,74 | 4 209,40 | 3 694,68 |
| Me [MW] | tél | 5 798,47 | 5 835,29 | 5 883,33 | 5 772,15 | 5 638,47 | 5 283,88 | 4 997,30 |
| | nyár | 5 492,63 | 5 568,60 | 5 352,26 | 4 992,08 | 5 038,09 | 4 573,68 | 4 455,15 |
| | átmenet | 5 202,25 | 5 098,33 | 5 117,09 | 5 072,33 | 4 978,13 | 4 481,00 | 3 862,72 |
| Q_3 [MW] | tél | 5 900,50 | 5 911,02 | 6 064,95 | 5 916,05 | 5 698,83 | 5 521,76 | 5 236,65 |
| | nyár | 5 766,76 | 5 901,37 | 5 512,43 | 5 232,89 | 5 362,40 | 4 795,27 | 4 555,81 |
| | átmenet | 5 311,14 | 5 238,27 | 5 241,25 | 5 277,33 | 5 182,99 | 4 634,17 | 4 097,67 |
| σ [MW] | tél | 753,89 | 621,33 | 672,01 | 674,40 | 594,00 | 488,13 | 524,51 |
| | nyár | 754,41 | 694,19 | 519,59 | 490,81 | 541,02 | 416,50 | 452,93 |
| | átmenet | 596,95 | 477,43 | 506,92 | 520,90 | 474,31 | 377,57 | 368,16 |
| V [%] | tél | 13,97 | 11,29 | 12,13 | 12,47 | 11,12 | 9,53 | 10,73 |
| | nyár | 14,64 | 13,06 | 10,10 | 10,15 | 10,97 | 9,30 | 10,71 |
| | átmenet | 12,18 | 9,78 | 10,38 | 10,70 | 9,89 | 8,61 | 9,62 |
| α_3 | tél | -0,93 | -0,87 | -0,80 | -0,86 | -0,93 | -0,51 | -0,29 |
| | nyár | -0,78 | -0,62 | -0,85 | -0,84 | -0,76 | -0,69 | -0,67 |
| | átmenet | -1,03 | -1,05 | -1,07 | -0,97 | -0,99 | -0,78 | -0,58 |

Forrás: saját számítások (R) és saját készítésű táblázat (R).

Az 25. táblázat néhány görbe esetén a hónapok, hét napjai illetve az órák/negyedórak által megmagyarázott varianciarányadok (H^2 -mutatók) értékeit mutatja.¹⁶²

¹⁶¹ A 24. táblázatban szereplő leíró statisztikák rendre az alábbiak: átlag, alsó kvartilis, medián, felső kvartilis, szórás, relatív szórás, illetve α_3 ferdeségi mutató. Mivel ezek ismert mutatók, ezért csak az utóbbi kettőhöz fűzünk megjegyzést. A relatív szórás a szórás illetve az átlag hányadosa, mivel a szóródást %-osan méri, ezért a különböző görbék összehasonlíthatóak abban a tekintetben, hogy hol nagyobb a szóródás mértéke. Az α_3 pedig ún. momentum alapú ferdeségi mutató, pozitív értékei jobbra hosszán elnyúló (*positively, right skewed*), negatív értékei balra hosszán elnyúló (*negatively, left skewed*), zero körüli értékei pedig közel szimmetrikus eloszlásra utalnak.

¹⁶² A feladatra úgy kell gondolnunk, mint egy vegyes kapcsolatra, ahol a mennyiségi ismerv a görbe idősor, a csoportképző ismérvek pedig a hónapok, a hét napjai illetve az órák, negyedórak. Ennek alapján

25. táblázat: Szezonális ismérvek által megmagyarázott varianciarányad görbénként

| Görbe | Csoportképző ismerv | | | |
|----------------------------------|---------------------|-----------|-------|------------|
| | Hónap | Hét napja | Órák | Negyedórak |
| V25 | 3,45 | 49,76 | 11,48 | 11,87 |
| V66 | 0,46 | 18,63 | 52,52 | 52,84 |
| V96 | 4,94 | 0,51 | 69,90 | 70,19 |
| V109 | 12,52 | 1,38 | 62,14 | 62,84 |
| Vállalati portfólió | 3,66 | 23,72 | 48,07 | 48,37 |
| Országos rendszerterhelés | 9,68 | 11,59 | 60,84 | 61,50 |

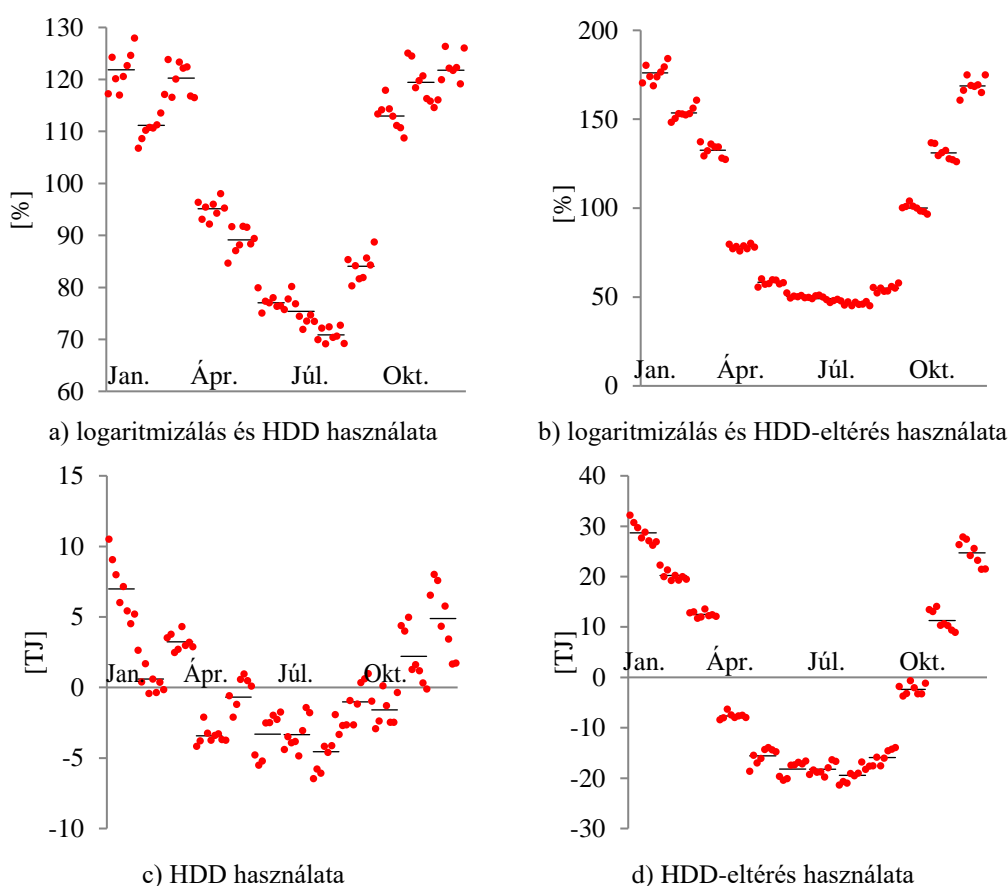
Forrás: saját számítások (R) és saját készítésű táblázat.

számolhatunk teljes, belső és külső eltérés-négyzetösszeget, melyek alapján meghatározható, hogy a csoportképző ismerv mekkora hányadot magyaráz a görbe idősor varianciájából.

D) SI-arányok a hazai földgázfogyasztás szezonális kiigazítása során

A melléklet az 2. fejezetben bemutatott szezonális kiigazítás során kapott SI-arányokat mutatja be az eredményváltozó szerepeltetésének mikéntje (logaritmikus transzformáció használata vagy sem), illetve a napfok változó eltérő módon történő kezelése (HDD vagy HDD-eltérés használata) mellett. Az eredmények könnyebb összehasonlíthatósága végett a napfok változóhoz minden modellváltozatban egy paramétert (bétát) becsülünk, így a több becsült paraméteren elért növekmény az illeszkedés jóságában nem torzít az itteni eredményeken.¹⁶³ A négy különböző modellváltozat során számolt SI-arányokat¹⁶⁴ mutatja az alábbi ábra.

72. ábra: SI-arányok különböző szezonális kiigazítási modellváltozatok esetén



Forrás: saját számítások (X13-ARIMA-SEATS, Excel) és saját készítésű ábrák (Excel).

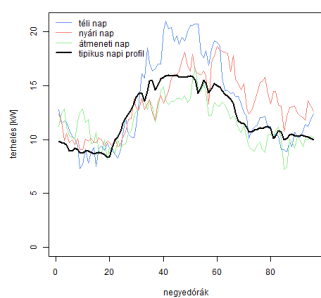
¹⁶³ Az, hogy szükséges-e az eredményváltozót logaritmizálnunk, illetve hogy a napfokot hogyan szerepeltessük, modellszelekciós kritériumokkal ellenőrizhető a programcsomag keretein belül. Ennek alapján a javasolt modell az, ami a főszövegben is szerepel (logaritmikus transzformáció, és havonta különböző HDD-eltérés paraméter becslése), de az SI-arányok alakulásában nincs olyan nagy különbség, ha a hónapokra azonos HDD-eltérés paramétert becsülünk (ld. 24. ábra a főszövegben és 72. b) ábra itt).

¹⁶⁴ Additív modell esetén (a fenti példában akkor, amikor az eredményváltozó értékét nem logaritmizáljuk) ezek a trend leválasztásával nem hányadosként, hanem különbségként adódnak, de ettől az elnevezésbeli problémától eltekintünk.

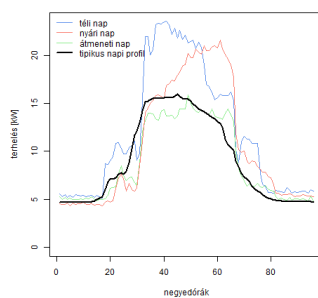
E) Görbék tipikus napi profiljai és heti idősoros ábrái

A melléklet a 4. fejezetben vizsgált egyedi görbék néhány idősoros ábráját tartalmazza.

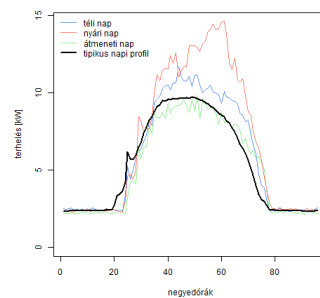
73. ábra: Egyedi fogyasztói görbék és a vállalati portfólió tipikus napi profiljai



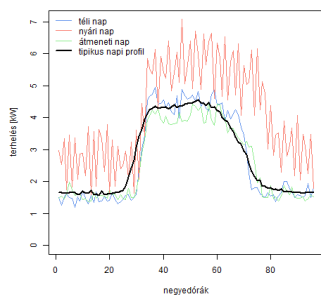
a) V25 görbe



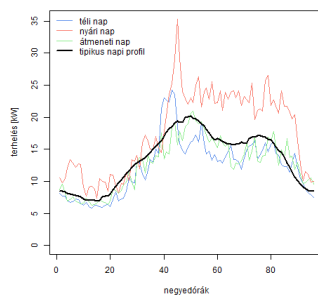
b) V35 görbe



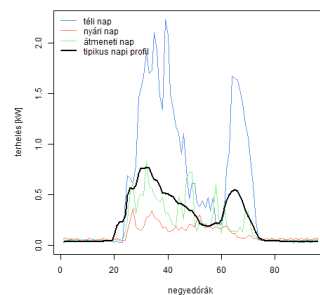
c) V66 görbe



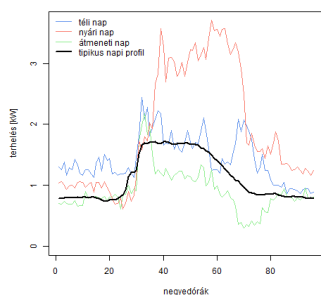
d) V79 görbe



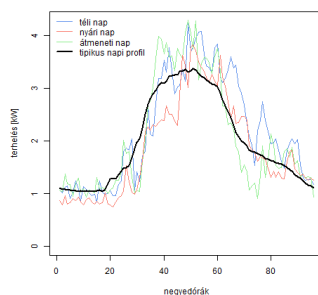
e) V109 görbe



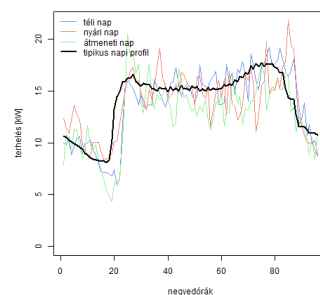
f) V1 görbe



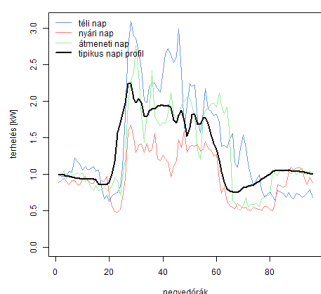
g) V4 görbe



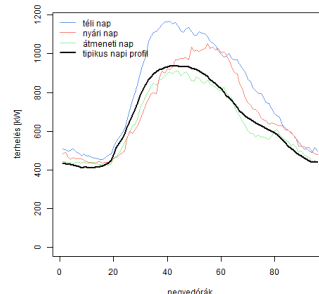
h) V27 görbe



i) V108 görbe



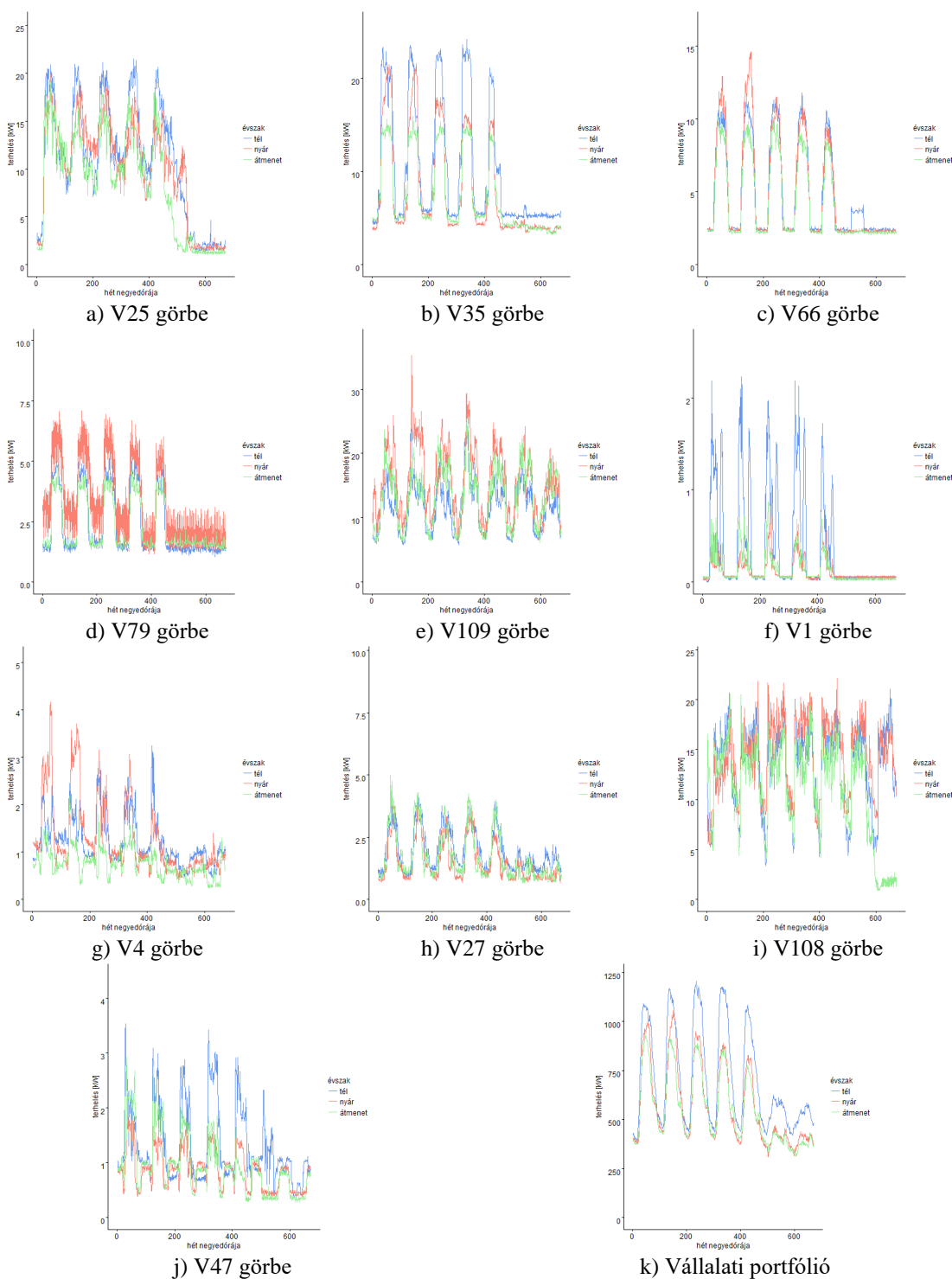
j) V47 görbe



k) Vállalati portfólió

Forrás: saját számítások (R) és saját készítésű ábrák (R).

74. ábra: Egyedi fogyasztói görbék és a vállalati portfólió heti idősorai szezononként

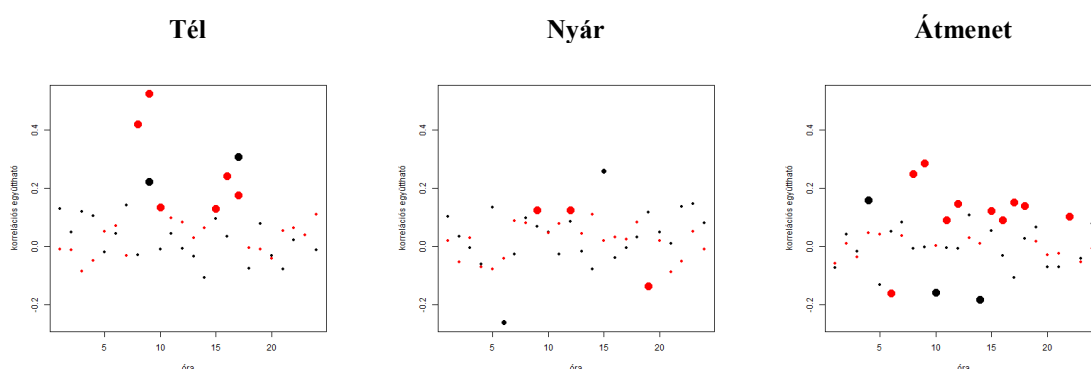


Forrás: saját számítások (R) és saját készítésű ábrák (R).

F) A volumenkockázat diverzifikálhatósága

A 75. ábrán a V35 illetve az értekezésben is többször szereplő vállalati portfólió standardizált (feltételes szórással normált¹⁶⁵) reziduumaik közötti lineáris korrelációs együtthatók értékei láthatóak. Az 5%-os szignifikancia szinten szignifikáns együtthatókat a nagyobb méretű pontok jelzik (a hétköznapokat piros, a hétvégéket fekete színnel jelöltük). Megállapítható, hogy az együtthatók értéke főként a hétköznapokon, a csúcsideőszakban különbözik szignifikánsan nullától, és jellemzően pozitív irányú kapcsolatot jelez.

75. ábra: Lineáris korrelációs együtthatók értéke a V35 egyedi görbe és a vállalati portfólió standardizált reziduumaik között¹⁶⁶



Forrás: saját számítások (R) és saját készítésű ábrák (R).

A teljes mintára a lineáris korrelációs együttható értéke egyébként 0,0671 (p-érték = 0,0000), ami gyenge, pozitív irányú (szignifikáns) kapcsolatot mutat az egyedi görbe és a portfólió reziduumaik között, ami tehát elmosza a szezonálisan jelentkező tendenciákat.

Természetesen a lineáris korrelációs együtthatók akkor mérik helyesen a kapcsolat erősségét, ha a változók közötti kapcsolat lineáris. Annyi azonban így is megállapítható, hogy a volumenkockázat nemdiverzifikálható része is időszaktól függően lehet kisebb vagy nagyobb, és állandó kapcsolatszorosságot feltételezve az értéke felül- vagy alulbecsült lesz.

A fentiek teljes körű vizsgálata, modellezése meghaladja az értekezés kereteit, ez a rövid (inkább illusztratív jellegű) példa azonban nagyon jól felhívja a figyelmet arra, hogy éppen a nagyobb szórású időszakok kockázata csökkenthető kisebb mértékben

¹⁶⁵ Így a heteroszkedaszticitás nem befolyásolja az eredményeket.

¹⁶⁶ A hétköznapok piros, a hétvégék fekete színnel vannak jelölve.

diverzifikációval, amit mindenképpen érdemes figyelembe venni a modellezések és a gyakorlati felhasználások során.

A jelenségre – mint egyfajta stilizált tényre – sem tudományos, sem gyakorlati munkákban nem találtunk példát, ennél fogva a tulajdonságot megragadó, publikusan elérhető modell sem létezik.

Az itt bemutatott példában a kapcsolatszorosságot viszonylag *ad-hoc* módon tettük idő- (magyarázóváltozó-) függővé. Ugyanígy lehet időfüggő kapcsolatszorosságot számolni például *wavelet*-transzformáció segítségével, ahol ráadásként még arra is választ kaphatunk, hogy a kapcsolat, az együttmozgás milyen frekvencián áll fenn. Hasonlóan jó megoldás lehet az ún. lokális korrelációs együttthatók (ld. *Tjostheim–Hufthammer* [2013]) alkalmazása, amire pénzügyi területen van sok példa, ennek a problémája azonban az, hogy nem időben, hanem a változók értékkészletét tekintve értendő a „lokalizált” előtag. Mindezek mellett a megoldás akár a keverék-modell keretrendszerében is elképzelhető.

IRODALOMJEGYZÉK

2007. évi LXXXVI. törvény a villamos energiáról

- Banfield, J. D. – A. E. Raftery, A. E. [1993]: Model-based Gaussian and non-Gaussian clustering. *Biometrics*. 49 pp. 803–821.
- Baudry, J.-P. – Raftery, A. E. – Celeux, G. , Lo, K. – Gottardo, R. [2010]: Combining mixture components for clustering. *Journal of Computational and Graphical Statistics*. 19 (2) pp. 332-353.
- Biernacki, C. – Celeux, G. – Gérard, G. [2003]: Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*. 41 pp. 561-575.
- Bollerslev, T. [1986]: Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. 31 (3) pp. 307-327.
- Box, G. E. P. – Jenkins, G. M. [1970]: Time Series Analysis: Forecasting and Control. Holden Day. San Francisco.
- Brealey, A. R. – Myres, C. S. [2005]: Modern vállalati pénzügyek. Panem Kft. Budapest.
- Breusch, T. S. [1978]: Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers*. 17 pp. 334–355.
- Carpaneto, E. – Chicco, G. – Napoli, R. – Scutariu, M. [2003]: Customer Classification by Means of Harmonic Representation of Distinguishing Features. Paper for *IEEE Bologna Power Tech Conference*, June 23th-26th, Bologna, Italy.
- Carpaneto, E. – Chicco, G. – Napoli, R. – Scutariu, M. [2006]: Electricity customer classification using frequency-domain load pattern data. *Electrical Power and Energy Systems*. 28 pp. 13-20.
- Chicco, G. [2012]: Overview and performance assesment of the clustering methods for electrical load pattern grouping. *Energy*. 42 pp.68-80.
- Chicco, G. – Napoli, R. – Piglion, F. – Postolache, P. – Scutariu, M. – Toader, C. [2005]: Emergent electricity customer classification. *IEE Proceedings – Generation, Transmission and Distribution*. 152 (2) pp. 164-172.
- Cont, R. [2001]: Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*. Vol. 1, pp. 223-236.
- Cover, T. M. – Thomas, J. A. [1991]: Elements of Information Theory. New York: Wiley.
- Dempster, A. P. – Laird, N. M. – Rubin, D. B. [1977]: Maximum Likelihood from Incomplete Data Via the EM-Algorithm. *Journal of Royal Statistical Society B*. Vol. 39. pp. 1-38.
- Dickey, D. A. – Fuller, W. A. [1979]: Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*. 74 (366) pp. 427–431.
- Eirola, E. – Lendasse, A. [2013]: Gaussian Mixture Models for Time Series Modelling, Forecasting, and Interpolation. *Advances in Intelligent Data Analysis XII. Lecture Notes in Computer Science*. (8207) pp. 162-173
- Elxon [2013]: Load Profiles and their use in Electricity Settlement. Guidance. Version 2.0. Elxon Ltd. UK.
- Engle, R. F. [1982]: Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*. 50 (4) pp. 987-1007.

- Espinoza, M. – Joye, C. – Belmans, R. – De Moor, B. [2005]: Short-Term Load Forecasting, Profile Identification and Customer Segmentation: A Methodology based on Periodic Time Series. *IEEE Transactions on Power Systems*. 20 (30) pp. 1622-1630.
- Fraley, C. – Raftery, A. E. [2000]: Model-Based Clustering, Discriminant Analysis, and Density Estimation. Technical Report no. 380. Department of Statistics, University of Washington.
- Fraley, C. – Raftery, A. E. [2007]: Model-based Methods of Classification: Using the mclust Software in Chemometrics. *Journal of Statistical Software*. 18 (6) pp. 1-13.
- Friedman, J. H. [1991]: Multivariate Adaptive Regression Splines. *The Annals of Statistics*. 19 (1) pp. 1-67.
- Godfrey, L. G. [1978]: Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*. 46 pp. 1293-1301.
- Hamilton, J. D. [1994]: Time Series Analysis. Princeton University Press. Princeton, New Jersey.
- Heo, Y. – Zavala, V. M. [2012]: Gaussian process modeling for measurement and verification of building energy savings. *Energy and Buildings*. 53 pp. 7-18.
- Hershey, J. R. – Olsen, P. A. [2007]: Approximating the Kullback-Leibler divergence between Gaussian mixture models. *IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07* (4) pp. IV-317-IV-320.
- Hino, H. – Shen, H. – Murata, N. – Wakao, S. [2013]: A Versatile Clustering Method for Electricity Consumption Pattern Analysis in Households. *IEEE Transactions on Smart Grid*. 4 (2) pp. 1048-1057.
- Howden, S. M. – Crimp, S. [2001]: Effect of Climate and Climate Change on Electricity Demand in Australia. *CSIRO Sustainable Ecosystems*. Canberra.
- Hunyadi, L. – Vita, L. [2003]: Statisztika közgazdászoknak. Aula Kiadó Kft., Budapest.
- Hylleberg, S. – Engle, R. F. – Granger, C. W. J. – YOO, B. S. [1990]: Seasonal Integration and Cointegration. *Journal of Econometrics*. 44. pp. 215-238.
- Jaques, J. – Preda, C. [2013]: Functional data clustering: a survey. Research Report. *Research Centre Lille – Nord Europe*.
- Junghans, G. [2015]: Portfolio risk management in a highly complex multi-regional market: Case study of Baltic market. *2nd Annual Intelligent Risk and Portfolio Optimisation for the Energy Markets*. 22nd-23rd September 2015, Berlin, Germany.
- Kerékgyártó, Gy. – L. Balogh, I. – Sugár, A. – Szarvas, B. [2008]: Statisztikai módszerek és alkalmazásuk a gazdasági és társadalmi elemzésekben. Aula Kiadó Kft., Budapest.
- Kwiatkowski, D. – Phillips, P. C. B. – Schmidt, P. – Shin, Y. [1992]: Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*. 54 (1-3) pp. 159-178.
- Leng, T. K. – Cheong, C. W. – Hooi, T. S. [2014]: Impact of global financial crisis on stylized facts between energy markets and stock markets. *Proceedings of the 3rd International Conference on Mathematical Sciences*. AIP Conf. Proc. Vol. 1602, pp. 994-1001.
- Levy, G. [2013]: Electricity contract risk with portfolio effects. *EnergyRisk risk-net/energy-risk*. Technical Paper. pp. 40-46.
- Li, X. – Bowers, C. P. – Schnier, T. [2010]: Classification of Energy Consumption in Buildings With Outlier Detection. *IEEE Transactions on Industrial Electronics*. 57 (11) pp. 3639-3644.
- Liao, T. W. [2005]: Clustering of time series data – a survey. *Pattern Recognition*. 28 pp. 1857-1874.
- Ljung, G. M. – Box, G. E. P. [1978]: On a Measure of a Lack of Fit in Time Series Models. *Biometrika*. 65 (2) pp. 297-303.

- Lo, K. L. – Wu, Y. K. [2003]: Risk assesment due to local demand forecast uncertainty int he compoeti-
tive supply industry. *IEE Proceedings – Generation, Transmission and Distribution*. 150 (5) pp.
573-581.
- Macedo, M. N. Q. – Galo, J. J. M. – de Almeida, L. A. L. – de C. Lima, A. C. [2015]: Demand side ma-
nagement using artificial neural networks in a smart grid environment. *Renewable and Sustainable
Energy Reviews*. 41 pp. 128-133.
- Maddala, G. S. [2004]: Bevetés az ökonometriába. Nemzeti Tankönyvkiadó, Budapest.
- Mák, F. [2014a]: Egységgyöktesztek alkalmazása szezonalitást is tartalmazó idősorok esetében energia-
tőzsde-adatok példáján. *Statistikai Szemle*. 92. évf. 7. sz. 647–679. old.
- Mák, F. [2014b]: Analyzing Interrelated Stochastic Trend and Seasonality on the Example of Energy
Trading Data. *Society and Economy*. 36 (2) pp. 233-261.
- Mák, F. [2015]: Az időjárás véletlen hatásának szerepe a szezonális kiigazítás során, a hazai földgázfo-
gyasztás példáján. *Statistikai Szemle*. 93. évf. 5. sz. 417–441. old.
- Manfren, M. – Aste, N. – Moshksar, R. [2013]: Calibration and uncertainty analysis for computer models
– A meta-model based approach for integrated building energy simulation. *Applied Energy*. 103
pp. 627-641.
- Marossy, Z. [2010]: A spot villamosenergia-árak elemzése statisztikai és ökonofizikai eszközökkel. PhD
értekezés. *Budapesti Corvinus Egyetem*. Budapest.
- Mathieu, J. L. – Price, P. N. – Kilicote, S. – Piette, M. A. [2011]: Quantifying Changes in Building Elect-
ricity Use, With Application to Demand Response. *IEEE Transactions on Smart Grid*. 2 (3) pp.
507-518.
- McKenna, S. A. – Fusco, F. – Eck, B. J. [2014]: Water demand pattern classification from smart meter
data. *Procedia Engineering*. 70 pp. 1121-1130.
- McLachlan, G. J. – Basford, K. E. [1988]: Mixture Models: Inference and Applications to Clustering.
Marcel Dekker.
- McLachlan, G. J. – Krishnan, T. [1997]: The EM Algorithm and Extensions. Wiley.
- McLachlan, G. J. – Peel, D. [2000]: Finite Mixture Models. Wiley.
- Mutanen, A. – Ruska, M. – Repo, S. – Järventausta, P. [2011]: Customer Classification and Load Profil-
ing Method for Distribution Systems. *IEEE Transactions on Power Delivery*. 26 (3) pp. 1755-
1763.
- Panapakidis, I. P. – Alexiadis, M. C. – Papagiannis, G. K. [2012]: Load Profiling in the Deregulated
Electricity Markets: A Review of the Applications. *2012 9th International Conference ont he Eu-
ropean Energy Market*. pp. 1-8.
- Panapakidis, I. P. – Papadopoulos, T. A. – Christoforidis, G. C. – Papagiannis, G. K. [2014]: Pattern
recognition algorithms for electricity load curve analysis of buildings. *Energy and Buildings*. 73
pp. 137-145.
- Pitt, B. [2000]: Applications of Data Mining Techniques to Electric Load Profiling. PhD Thesis. *Univer-
sity of Manchester Institute of Science and Technology*.
- Povinelli, R. J. – Johnson, M. T. – Lindgren, A. C. – Ye, J. [2014]: Time Series Classification Using
Gaussian Mixture Models of Reconstructed Phase Spaces. *IEEE Transactions on Knowledge and
Data Engineering*. 16 (6) pp. 779-783.
- Ramanathan, R. [2003]: Bevezetés az ökonometriába alkalmazásokkal. Panem Kft., Budapest.

- Räsänen, T. – Voukantsis, D. – Niska, H. – Karatzas, K. – Kolehmainen, M. [2010]: Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*. 87 pp. 3538-3545.
- Shekofteh, Y. – Almasganj, F. [2013]: Feature Extraction Based on Speech Attractors in the Reconstructed Phase Space for Automatic Speech Recognition Systems. *ETRI Journal*. 35 (1) pp. 100-108.
- Singh, R. – Pal, B. C. – Jabr, R. A. [2010]: Statistical Representation of Distribution System Loads Using Gaussian Mixture Model. *IEEE Transactions on Power Systems*. 25 (1) pp. 29-37.
- Srivastav, A. – Tewari, A. – Dong, B. [2013]: Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models. *Energy and Buildings*. 65 pp. 438-447.
- Subbarao, K. – Lei, Y. – Reddy, T. A. [2011]: The Nearest Neighborhood Method to Improve Uncertainty Estimates in Statistical Building Energy Models. *ASHRAE Transactions*. 117 (2) pp. 459-471.
- Sugár, A. [1999a]: Szezonális kiigazítási eljárások (I.). *Statisztikai Szemle*. 77 (9) 705–721. old.
- Sugár, A. [1999b]: Szezonális kiigazítási eljárások (II.). *Statisztikai Szemle*. 77 (10-11) 816–832. old.
- Sugár, A. [2011]: A hőmérséklet hatásáról a villamosenergia- és gázfogyasztás magyarországi példáján. *Statisztikai Szemle*. 89 (4) 379–398. old.
- Tjostheim, D. – Hufthammer, K. O. [2013]: Local Gaussian correlation: A new measure of dependence. *Journal of Econometrics*. 172 pp. 33-48.
- Tsekouras, G. J. – Hatziargyriou, N. D. – Dialynas, E. N. [2007]: Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers. *IEEE Transactions on Power Systems*. 22 (3) pp. 1120-1128.
- Tsekouras, G. J. – Kotoulas, P. B. – Tsikeris, C. D. – Dialynas, E. N. – Hatziargyriou, N. D. [2008]: A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. *Electric Power Systems Research*. 78 pp. 1494-1510.
- Váradi, K. [2012]: Likviditási kockázat a részvénytőzsián. PhD értekezés. *Budapesti Corvinus Egyetem*. Budapest.
- Varian, H. R. [2004]: Mikroökonómia középfokon. *Akadémiai Kiadó Zrt.* Budapest.
- Verdú, S. V. – García, M. O. – Senabre, C. – Marín, A. G. Franco, F. J. G. [2006]: Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps. *IEEE Transaction on Power Systems*. 21 (4) pp. 1672-1682.

Internetes források:

Speciális időjárási hatások kezelésével kapcsolatos források:

https://en.wikipedia.org/wiki/Wind_chill

<https://en.wikipedia.org/wiki/Humidex>

Módszertani fogalmakkal kapcsolatos források:

mathworld.wolfram.com/RelativeEntropy.html

https://en.wikipedia.org/wiki/Hermite_interpolation

Az elemzésekhez használt adatok forrásai (a megjelölést ld. mindig az aktuális fejezetben):

ec.europa.eu/eurostat/

<https://www.fgsz.hu>

<https://www.hupx.hu>

<https://www.mavir.hu>

www.varaljamet.eoldal.hu/cikkek/climate_budapest.html

Volumenkockázat heurisztikus mérésével kapcsolatos források:

www.enoro.com/

Egyéb internetes források:

<https://wikiszotar.hu/>

SAJÁT PUBLIKÁCIÓK JEGYZÉKE

A) A dolgozat témakörével kapcsolatos magyar nyelvű publikációk

Referált folyóirat:

- Mák, F. [2011]: Egységgyöktesztek alkalmazása strukturális törések mellett a hazai benzinár példáján. *Statisztikai Szemle*. 89. évf. 5. sz. 545–573. old.
- Mák, F. [2014]: Egységgyöktesztek alkalmazása szezonalitást is tartalmazó idősorok esetében energiatőzsde-adatok példáján. *Statisztikai Szemle*. 92. évf. 7. sz. 647–679. old.
- Mák, F. [2015]: Az időjárás véletlen hatásának szerepe a szezonális kiigazítás során, a hazai földgázfogyasztás példáján. *Statisztikai Szemle*. 93. évf. 5. sz. 417–441. old.

B) A dolgozat témakörével kapcsolatos idegen nyelvű publikációk

Referált folyóirat:

- Mák, F. [2014]: Analyzing Interrelated Stochastic Trend and Seasonality on the Example of Energy Trading Data. *Society and Economy*. 36 (2) pp. 233-261.

C) A dolgozat témakörével kapcsolatos, nem publikált munkák

- Mák, F. – Sugár, A. – Tóth, B. [2009]: A magyarországi gázpiaci és gázszolgáltatásokhoz köthető egyéb termékek keresleti elemzése. *Tanulmány a Fővárosi Gázművek Zrt. részére*.
- Mák, F. – Sugár, A. [2011]: Lokális hőigényekre vonatkozó előrejelzési modellek. *Tanulmány az E.ON Energiaszolgáltató Kft. részére*.
- Mák, F. [2015]: Hőmérsékletkorrekciós modell. *Tanulmány és modell dokumentáció a Magyar Energetikai és Közmű-szabályozási Hivatal részére*.