



Gazdálkodástani Doktori Iskola

TÉZISFÜZET

Mák Fruzsina

Fogyasztási kockázat a villamosenergia-piacon

Profilozás a fogyasztási bizonytalanság figyelembevétele mellett

című doktori értekezéséhez

Témavezetők:

Dr. Oravecz Beatrix, Ph.D.

egyetemi adjunktus

Dr. Sugár András, Ph.D.

egyetemi docens, tanszékvezető

Budapest, 2017

Statisztika Tanszék

TÉZISFÜZET

Mák Fruzsina

Fogyasztási kockázat a villamosenergia-piacon

Profilozás a fogyasztási bizonytalanság figyelembevétele mellett

című doktori értekezéséhez

Témavezetők:

Dr. Oravecz Beatrix, Ph.D.

egyetemi adjunktus

Dr. Sugár András, Ph.D.

egyetemi docens, tanszékvezető

© Mák Fruzsina

TARTALOMJEGYZÉK

1. KUTATÁSI ELŐZMÉNYEK ÉS A TÉMA INDOKOLÁSA	1
2. A FELHASZNÁLT MÓDSZEREK ÉS AZOK INDOKOLÁSA	6
3. AZ ÉRTEKEZÉS FŐBB TUDOMÁNYOS EREDMÉNYEI ÉS AZOK HASZNOSÍTHATÓSÁGA.....	13
4. FŐBB HIVATKOZÁSOK	22
5. A TÉMAKÖRREL KAPCSOLATOS SAJÁT PUBLIKÁCIÓK JEGYZÉKE.....	25

1. KUTATÁSI ELŐZMÉNYEK ÉS A TÉMA INDOKOLÁSA

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 **fogyasztáscsökkentés** 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 rendszerszintű szabályozási nehézségek kezelésében, vagy éppen a rendszerszintű költségek, veszteségek csökkentésére való törekvésekben manifesztálódnak.

1.1. Az értekezés célja

Annak ellenére, hogy az energiapiaci kockázatok egyik legjelentősebb forrása alapvetően az ár, a fentiek nyomán nemcsak a portfólió-szintű, 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 összefüggenek, egyrészt keresztmetszetileg (az egyedi fogyasztói görbék összege a portfóliószintű görbe), másrészt idő-

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

sorosan (a menetrendezés során elkövetett előrejelzési hibák eredményezte kiegyenlítő energia költsége rakódik rá a portfólióra az üzleti év során).

Mivel a villamosenergia-rendszer minden időpillanatban fennálló **kereslet-kínálati egyensúlya** 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. Ezekben a piacokon rövidtávon például az egyensúlyt tarifális ösztönzéssel elérni kívánó 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. Ezekben 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 fogyasztói profilok – amelyek egy alapvetően kvantitatív módszertani folyamat eredményeként állnak elő, és leírják, hogy a fogyasztás hogyan függ különböző szezonális, naptári vagy egyéb hatásoktól² – vizsgálatával, modellezésével kapcsolatban rendkívül széleskörű, de természetesen korántsem egységes szakirodalom és gyakorlati alkalmazások állnak rendelkezésre.

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 szűkszerűen 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.

² A szakirodalom alapján egyébként nincs egységes, általános érvényű definíció.

1.2. Szakirodalmi háttér és a megfogalmazott hipotézisek

A profilozással foglalkozó szakirodalom alapján a fogyasztói profilok illetve profilsoportok képzése az ún. kétlépeses 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 előállítása (1. lépés) tekinthető inkább hangsúlyosnak és iparág-specifikusnak.

A kétlépeses vagy kétlépcsős klaszterezés lényege tehát, hogy első lépésben megtörténik az alapadatokból valamilyen releváns tulajdonság (*feature*) kinyerése, majd a klaszterezés ebben a térben zajlik.

Az egy-egy fogyasztói görbét jellemző, ún. görbetulajdonságok sokféleképpen képezhetőek. Ezeknek a megoldásoknak egy lehetséges csoportosítását tartalmazza az 1. táblázat. Jól látható, hogy a legtöbb esetben a fogyasztói görbét jellemző tulajdonságok előállítása valamilyen **napi reprezentatív terhelési görbe** (*representative load curve*, röviden RLC) alapján történik meg, és ez az alapja a későbbiekben a görbék egyedi jellemzésének illetve a hasonló karakterisztikával jellemezhető görbéket tartalmazó profilsoportok képzésnek is.

1. táblázat: Profilozás során előállított görbetulajdonságok lehetséges csoportosítása

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

Forrás: saját gyűjtés és saját szerkesztésű táblázat.

Ugyan ezek az ún. napi reprezentatív terhelési görbék 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

egyébként természetesen megoldható az, hogy **különböző feltételek melletti** (például nyári, téli, átmeneti időszakai vagy a hét különböző napjaira vonatkozó stb.) **napi profilokat** is előállítsunk.

Fontos konklúzió még, hogy az értekezésben vizsgált kutatási eredmények alapján az időjárás (legtöbb esetben a hőmérséklet) teljes (ritkábban a véletlen) hatása kiszűrésre kerül profilozás során (ld. a táblázatban szereplő forrásokat is).

Saját korábbi eredmények (ld. *Mák* [2015]), illetve annak kiegészítése alapján azonban megállapítottuk, hogy amennyiben a fogyasztási idősor véletlen komponensének a viselkedése fontos az elemzés szempontjából, akkor a hőmérséklet teljes hatásának a kiszűrése technikai és értelmezésbeli szempontból sem igazán kedvező, a hőmérséklet véletlen hatásának a kiszűrése pedig a fogyasztás heteroszkedasztikus viselkedését mérsékelheti.

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 időornak a vizsgálata mind a lehetséges elemzések, módszerek, mind a levonható következtetések szempontjából korlátozó tényező lehet.

A fenti lényegi eredmények alapján az értekezésben elsősorban az alábbi **területekre** koncentráltunk:

- 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 – részben a feldolgozott szakirodalom, részben korábbi saját eredmények alapján – az alábbi **hipotéziseket** fogalmaztuk meg:

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

- H2: Az ún. klasszikus (tipikus napi profilgörbén alapuló) megoldásokhoz képest reálisabb 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.

2. A FELHASZNÁLT MÓDSZEREK ÉS AZOK INDOKOLÁSA

Mivel az értekezésben felhasznált módszertant a korábbi kutatási eredmények tapasztalatai mellett alapvetően a fogyasztási idősorok stilizált tényei kapcsán tett megfigyelések alapján választottuk, ezért a fejezetben röviden összegezzük a stilizált tények kapcsán tett fontosabb megállapításokat, illetve azt, hogy a keverék-modell miért kézenfekvő ezek jelentős részének a kezelésére.

2.1. Fogyasztási idősorok stilizált tényeinek vizsgálata

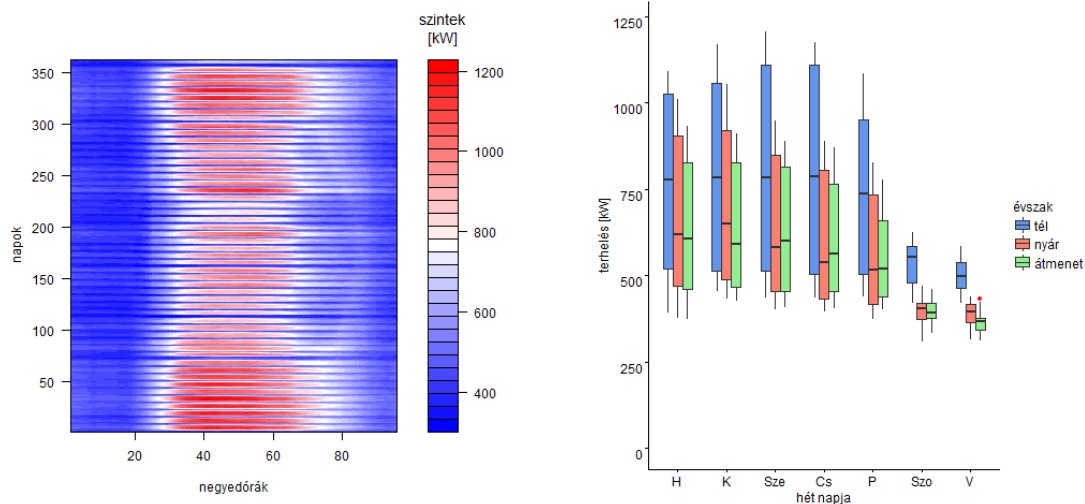
Több egyedi görbét is megvizsgáltunk, hogy melyek azok a tulajdonságok, amelyek egy-egy görbe karakterisztikáját meghatározhatják, amelyeket a tipikus fogyasztási mintázatnak a meghatározásakor is figyelembe kell venni. Ezek a vizsgálatok jellemzően nem klasszikus statisztikai próbák voltak, hanem olyan 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:
 - o a csúcsidezőszaki, csúcsidezőszakon kívüli, hétköznapi, hétvégi terhelések szintje, a csúcsidezőszakok napon belüli elhelyezkedése hogyan változik az év során,
 - o melyik görbék esetében van jelentős hatása az ünnepnapoknak,
 - o a hőmérséklet hatásáról is vonhatóak le következtetések,
 - o 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 **hőmérsékletfüggőséget**, 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**, hétköznapiokon és hétvégéken, illetve téli, nyári és átmeneti időszakokban; ezzel ellenőriztük, hogy
 - o 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.

1. ábra: Egy vállalati portfólió szintvonalábrája és dobozábrája



Forrás: saját készítésű ábra (R).

A vizsgálatok eredményeit figyelembe véve elmondhatjuk, hogy a **villamosenergia-fogyasztási görbék varianciájából a legnagyobb hányad jellemzően a napon belüli szezonális 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.**

2.2. Keverék-modell alkalmazása az időben változó kovariacia-struktúra modellezésére

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. A módszertanról ld. többek között *Eirola és Lendasse* [2013], *Frøley-Raftery* [2000], *Frøley-Raftery* [2007].

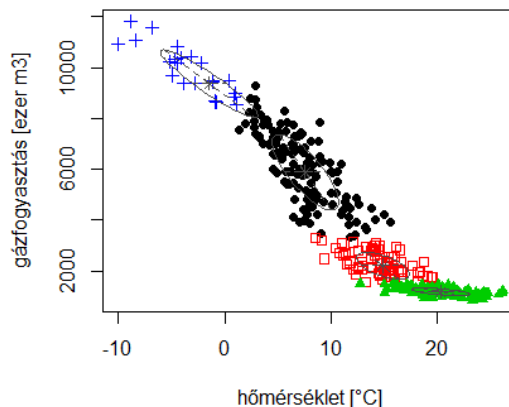
Az értekezés célja kettős: nemcsak a tipikus, jellemző fogyasztási mintázatot kívánjuk megkeresni, hanem a fogyasztással kapcsolatos bizonytalanság számszerűsítése is érdekes.

Ezért nagyobb hangsúlyt kap a fogyasztás mellett a fogyasztás késleltetettjeinek illetve a potenciálisan elérhető exogén változóknak az együttmozgása 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:
 - o a kiugró értékek szűrésére, vagy
 - o 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 szezonaritást 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.

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



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

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 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 manifesztálódni.

Az említett módszertani nehézségre egy, az országos **földgázfogyasztás** alakulásával kapcsolatos **korábbi saját eredmény** kapcsán is felhívtuk a figyelmet³. 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 szűródik 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 gondolkozunk.

2.3. A keverék-modell és a keverék-regresszió rövid formális felírása

Formálisan felírva a keverék-modellek lényege röviden az alábbiakban összegezhető.

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:

- 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,⁴
- $f(\cdot)$ a keverék-eloszlás sűrűségfüggvénye, $f_k(\cdot)$ a k -edik komponens sűrűségfüggvénye,
- θ_k a k -edik komponens sűrűségfüggvényét leíró paramétereket jelöli, τ_k pedig a k -edik komponensbe tartozás ún. *prior* valószínűsége,
- k a komponensek jelölésére szolgál ($k = 1, 2, \dots K$), K a komponensek száma.

³ Az eredmények az X13-ARIMA-SEATS szezonális kiigazító program felhasználásával készültek.

⁴ A nemzetközi, főként az adatbányász szakirodalomban használatos kifejezés, az attribútum azonos a regressziós terminológiában használt változó kifejezéssel.

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]). 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, ami azt mutatja meg, hogy egy-egy megfigyelés melyik komponensből származik. 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} valószínűségeiben 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 komponensek sűrűségfüggvényének a paramétereit. A *likelihood*-függvény az alábbi:

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

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:

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

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

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

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

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

azaz a $\psi = (\tau_1, \tau_2 \dots \tau_K, \theta_1, \theta_2 \dots, \theta_K)$ paraméterszett becslendő paraméterei a *prior* valószínűségek mellett a komponensenkénti átlagvektorok illetve kovariancia-mátrixok lesznek.

A Gauss-féle keverék-regresszió (*Gaussian Mixture Regression*, röviden GMR) 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. Technikailag a keverék-regresszió a súlyozott legkisebb négyzetek módszerének a keretrendszerében írható fel.⁵

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.

Az értekezésben bemutatjuk, hogy az m_{ik} képlete nem más, mint a komponensenkénti regressziós egyenletbe történő behelyettesítés eredménye, az s_{ik}^2 képlete pedig nem más, mint a komponensenkénti reziduális variancia az x_i magyarázóváltozók konkrét értékei mellett.

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

⁵A keverék-regresszió értelmezés és implementálás szempontjából is praktikus formális felírása eltér a klasszikus regressziós módszertan által használt jelölésrendszertől, azonban a dolgozat eredményei között szerepel a kettő közötti kapcsolat bemutatása is. Ezt felhasználva a komponensenkénti regressziós koefficiensek illetve a komponensenkénti reziduális variancia az alábbi módon írhatóak fel (W_k a p_{ik} súlyokat tartalmazó $n \times n$ méretű komponensenkénti diagonális mátrix):

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

illetve

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

ahol a súlyozástól eltekintve könnyen felismerhetőek a klasszikus regressziós módszertan legkisebb négyzetek módszerének alkalmazásával kapott megfelelő összefüggései.

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

Az értekezés megfelelő fejezete több olyan levezetést, értelmezést is tartalmaz különösen a keverék-regresszióhoz kapcsolódóan, 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 fentiek az eredmények nagyban támogatták a keverék regresszió *R Project* programcsomagban történő implementációját is. A *Gauss*-féle keverék-modellen alapuló regressziók utasításai az *'mclust'* *package* által adott eredményekre épülően saját készítésű függvények, ugyanis az *R package*-ben regressziós alkalmazási célú függvények nem szerepelnek.

3. AZ ÉRTEKEZÉS FŐBB TUDOMÁNYOS EREDMÉNYEI ÉS AZOK HASZNOSÍTHATÓSÁGA

Az értekezésnek mind a profilozással, mind a bizonytalansággal kapcsolatos alfejezeteiben szerepelnek ún. klasszikusnak tekinthető megoldások is, részben *benchmark*-ként, az új eredményekkel történő szembeállítás, összehasonlítás céljával, részben pedig feltáró jelleggel. Mivel az elért eredmények a klasszikus megoldásokkal történő szembeállítás nyomán értékelhetőek, ezért a fontosabb következtetéseket ennek tükrében összegezzük. A fejezet végén az értekezés eredményeinek, fontosabb megállapításainak a hasznosíthatóságát tárgyaljuk a továbbfejlesztési lehetőségek ismertetése mellett.

3.1. Keverék-modell alkalmazása tipikus fogyasztási mintázatok készítésére

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 (regressziós) 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 profilszempontok 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 hétköznapokhoz képest, a hőmérsékletfüggőség 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 **min-taelemszámra**, 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 komponensstagságok) 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 profilsoportképzés 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.

3.2. 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 csúcsidezőszi, a napon belüli felfutási-lecsengési időszakokban, a hétköznapokon, illetve nyáron és télen ez jóval 5% felett van, egyébként pedig jóval alatta.⁶

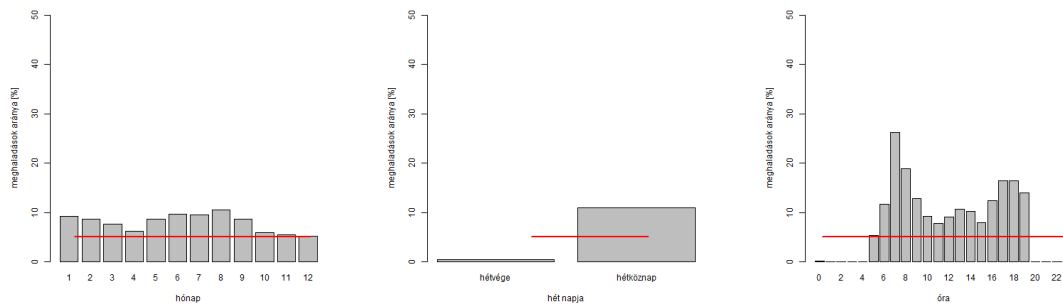
A klasszikus **idősoros regressziós megoldások hibái (reziduuma)** szórásának vizsgálata alapján az alábbi következtetésre jutottunk a fogyasztás időben változó kockázata tekintetében:

- a csúcsidezőszi fogyasztás bizonytalansága magasabb,

⁶ 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.

- a csúcsidezőszakon 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*.⁷

3. ábra: Konfidencia-intervallumon kívül eső megfigyelések aránya egy egyedi görbe esetén



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

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 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 tehetőek 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.

3.3. Keverék-modell alkalmazása a fogyasztás bizonytalanságának mérésére

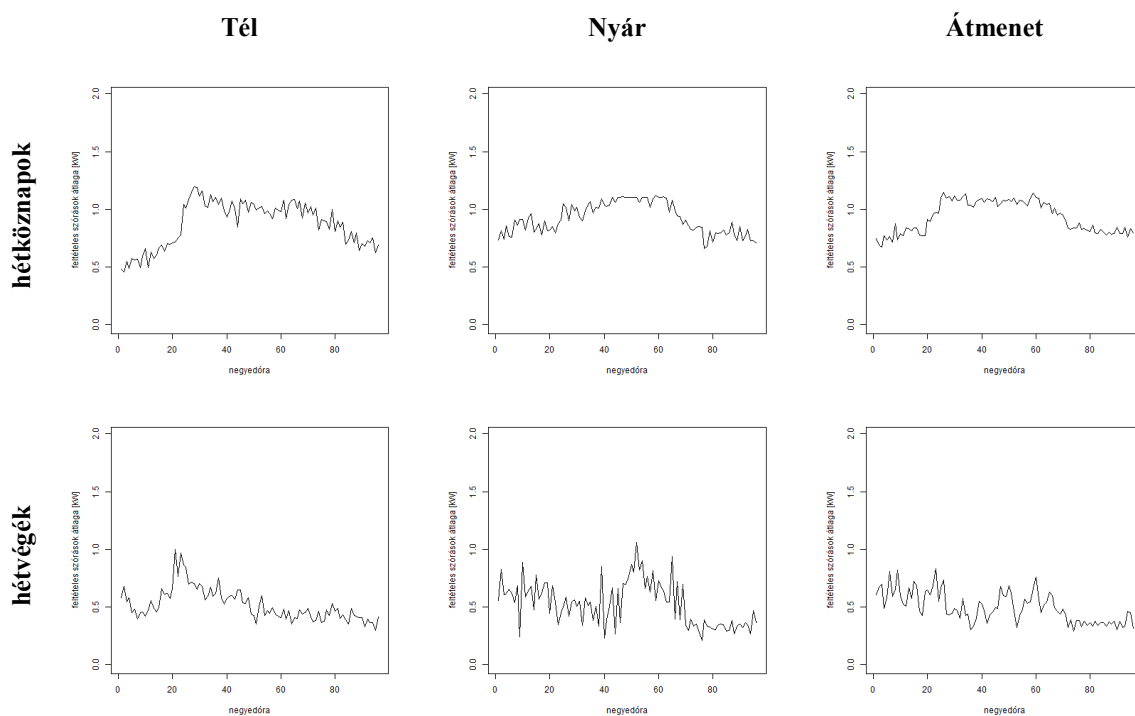
Az összegzett tapasztalatok alapján a profilkészítés kapcsán is alkalmazott keverék-modell regressziós alkalmazásával olyan, ún. **feltételes, időfüggő standard hibákat**, illetve

⁷ 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.

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. órára, hétköznapokra/hétfévkre, hónapokra számolva) mennyiben cseng egybe a keverék-regresszió alapján kalkulált standard hibákkal.

4. ábra: Keverék-regresszió által számolt standard hibák átlaga egy egyedi görbe esetén



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

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 időszakok 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.**

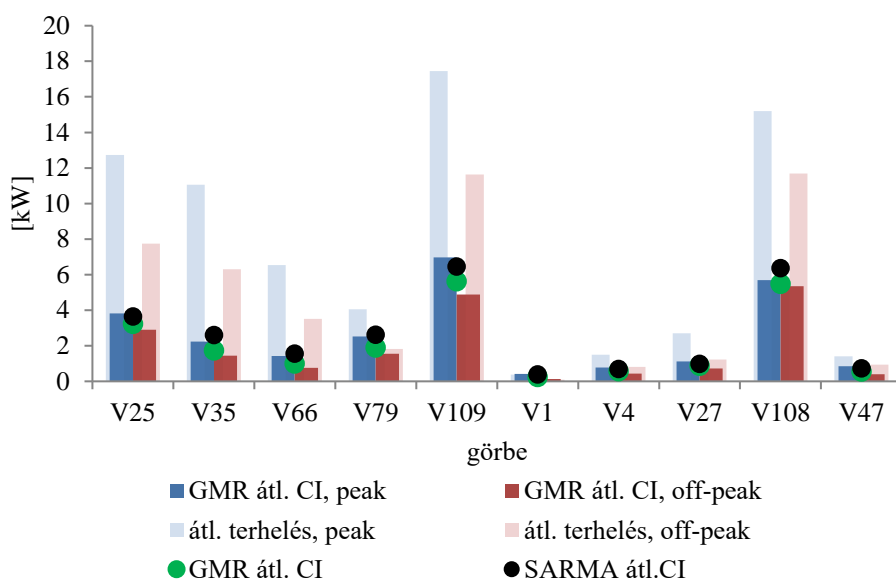
Az **eredmények fontossága** abban rejlik többek között, hogy

- profil és a fogyasztás bizonytalansága (volumenkockázat modellezése) egy modellkeleten 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.

5. ábra: Átlagos konfidencia-intervallum SARMA- és keverék-modellekben az egyedi görbék esetén⁸



Görbe	V25	V35	V66	V79	V109	V1	V4	V27	V108	V47
Csökkenés mértéke [%]	11	33	36	28	13	38	18	11	14	24

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

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.

3.4. Az értekezés eredményeinek hasznosíthatósága, továbbfejlesztési lehetőségek

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 profilsoprotok 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 profilsoprot

⁸ Az ábrán a CI a 95%-os megbízhatósági szintű konfidencia-intervallum rövid jelölésére szolgál.

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 mintaelemszám-igényesebb megoldásokhoz képest olyan esetekben, amikor csak tört évre van információnk.

Igaz ugyan, hogy a keverék-regresszió a reziduummokkal 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 időjárás-

függő 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 korrelációja⁹ 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 értékelése 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 rendszerszintű 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 teszti 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 volumenkockázat szempontjára.

⁹ Sőt, annak szignifikanciája vagy annak hiánya.

ból 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 (szélsébség, 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.

4. FŐBB HIVATKOZÁSOK

- 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.
- Box, G. E. P. – Jenkins, G. M. [1970]: Time Series Analysis: Forecasting and Control. Holden Day. San Francisco.
- 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. – Piglione, 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.
- 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
- 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.
- Hamilton, J. D. [1994]: Time Series Analysis. Princeton University Press. Princeton, New Jersey.
- 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.
- 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.
- 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.
- Lo, K. L. – Wu, Y. K. [2003]: Risk assesment due to local demand forecast uncertainty int he compoetitive 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 management 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 energiatőzsde- adatok példáján. *Statisztikai 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 földgázfogyasztás példáján. *Statisztikai Szemle*. 93. évf. 5. sz. 417–441. old.
- Mathieu, J. L. – Price, P. N. – Kilicote, S. – Piette, M. A. [2011]: Quantifying Changes in Building Electricity 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 Profiling 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 European 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. *University of Manchester Institute of Science and Technology*.
- 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.
- 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.
- Tsekouras, G. J. – Hatzargyriou, 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. – Hatzargyriou, 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.
- 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.

A hivatkozások teljes listája az értekezésben található.

5. A TÉMAKÖRREL KAPCSOLATOS SAJÁT PUBLIKÁCIÓK JEGYZÉKE

5.1. 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 szezonaritást 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.

5.2. 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.