



BUDAPESTI CORVINUS EGYETEM
EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Szakdolgozat

A hőhullámok által okozott mezőgazdasági aszálykár-kifizetések becslése Magyarországon

BIZTOSÍTÁSI ÉS PÉNZÜGYI MATEMATIKA MSc
AKTUÁRIUS SPECIALIZÁCIÓ

Készítette:

NGUYEN HOANG AN

Konzulens:

KOVÁCS LÁSZLÓ

Köszönetnyilvánítás

Ezúton szeretném kifejezni hálámat témavezetőmnek, Kovács Lászlónak, aki szakmai támogatásával, útmutatásával és a belém fektetett munkájával jelentősen hozzájárult a kutatásom sikeréhez. Külön köszönöm, hogy a felmerülő kérdésekre, problémákra mindig magas szintű szakértelemmel adott választ, ezzel is emelve a szakdolgozat minőségét. Szintén hálás vagyok a munkahelyi vezetőmnek, Doró Attilának a szakdolgozatírással kapcsolatos tanácsaiért és bátorításáért, amely nélkül ez a kutatás nem valósulhatott volna meg.

Tartalomjegyzék

1. BEVEZETÉS	4
2. SZAKIRODALMI ÁTTEKINTÉS	7
3. MATEMATIKAI MÓDSZERTAN	10
3.1. BIZTOSÍTÁSI SZERZŐDÉS FELÉPÍTÉSE	10
3.1.1. HŐHULLÁM DEFINIÁLÁSA	10
3.1.2. SZERZŐDÉS DEFINIÁLÁSA	11
3.1.3. SZERZŐDÉS KIÉRTÉKELÉSE	12
3.2. HŐMÉRSÉKLET ELŐREJELZÉSE SZTOCHASZTIKUSAN . .	14
3.2.1. AZ ARCH FOLYAMATOK	16
3.2.2. A LIKELIHOOD FÜGGVÉNY	19
3.3. EXTRÉMÉRTÉK-ELMÉLET	22
3.3.1. A MAXIMUM HATÁRELOSZLÁSA	22
3.3.2. KÜSZÖB FELETTI MEGHALADÁSOK MÓDSZERE . .	25
3.3.2.1. KÜSZÖB MEGVÁLASZTÁSA	26
3.3.3. PARAMÉTEREK BECSLÉSE	28
3.4. KÁRKIFIZETÉS BECSLÉSE	30

4. MODELLEZÉSI EREDMÉNYEK	32
4.1. ADATOK BEMUTATÁSA	32
4.2. HŐMÉRSÉKLET MODELLEZÉSE ÉS ELŐREJELZÉSE	33
4.2.1.DETERMINISZTIKUS MODELLEZÉS	33
4.2.2.AR(P) MODELL MEGVÁLASZTÁSA	36
4.2.3.ARCH MODELL	40
4.2.4.ELŐREJELZÉS	42
4.2.5.HÓHULLÁMOK BEKÖVETKEZÉSI VALÓSZÍNŰSÉGE .	45
4.3. EXTRÉMÉRTÉK-MODELLEK	46
4.3.1.KÜSZÖB FELETTI MEGHALADÁSOK	46
4.3.2.BLOKKMAXIMUMOK	50
4.3.3.KÁRKIFIZETÉS BECSLÉSE A HŐMÉRSÉKLET FI- GYELEMBEVÉTELÉVEL	51
5. KONKLÚZIÓ	55

1. fejezet

Bevezetés

A globális éghajlatváltozás egyik legjelentősebb hatása a szélsőséges időjárási események gyakoriságának és intenzitásának a növekedése. Ezek közül is kiemelt figyelmet érdemelnek a hőhullámok, amelyek nem csak az emberi egészséget és környezetet veszélyeztetik, hanem jelentős hatással vannak a mezőgazdasági termelésre is. Hőhullámnak nevezünk egy olyan természeti katasztrófát, ami meleg napok sorozatából áll (meleg napnak minősül egy nap, ha a maximum hőmérséklet meghalad egy küszöbértéket), több tanulmányban is általában 3 nap a minimum (*Fenner et. al, 2018, Campbell és Diebold, 2005*). 2003 és 2018 között a hőhullámok hosszának mediánja 5-6 nap volt, az átlag az ennél több, kb. 8 nap (*Garcia Leon et. al, 2021*). Ezek az események jelentős gazdasági és társadalmi károkat okozhatnak, különösen az időjárástól erősen függő mezőgazdaságban. A mezőgazdasági termelés szoros kapcsolatban áll az időjárási körülményekkel és a magas hőmérséklet a terméshozam csökkenésével jár. A szélsőséges hőmérséklet negatív hatást gyakorol a növények fotoszintézisére, csökkenti a termékek minőségét és mennyiségét és ezek a hatások az élelmiszer árak alakulásában is megjelenhetnek. Szakdolgozatom célja, összefüggést találgatni a hőhullámok és mezőgazdasági károk egy részhalmaza között (nem terjeszt-

hető ki ez a teljes országra, mivel egy nagyobb magyar biztosító adataival fogok dolgozni, ami persze nem fedi le az egész piacot, illetve olyan károk is lehetnek, amit az állam fedez, vagy olyan is előfordulhat, hogy a gazdáknak kell maguknak). Ehhez első sorban hőmérsékleti idősoros adatokat használok, amit két részben fogok megvizsgálni: egy determinisztikus, illetve egy sztochasztikus rész. A modell determinisztikus részében az idősorban rejlő trendet, illetve szezonális hatásokat szeretném megragadni, a sztochasztikus részében, pedig az idősorban fennmaradó autokorrelációt és szórást szeretném modellezni, majd ezen információk segítségével előrejelezni a következő 1 évre. A hőmérséklet változáson túl mezőgazdasági aszálykár-adatokat is belevonok a kutatásomba, hogy összefüggéseket állítsak fel a hőmérséklet-változások és a gazdasági veszteségek között. A mezőgazdasági károk esetében extrémérték-elméleti modell segítségével szeretném megbecsülni a nagyobb mezőgazdasági károk bekövetkezési valószínűségét. Az extrémérték-elméleti modellek alkalmazása indokolt, mivel az aszálykárok viszonylag ritkán bekövetkező események, emiatt a hagyományos statisztikai módszerek, amelyek a múltbeli megfigyelések gyakoriságán alapulnak, nem nyújtanak kellően megbízható alapot a jövőbeli kockázatok becslésére. Ez a módszertan kifejezetten a ritka, de potenciálisan jelentős hatású események modellezésére szolgál, így lehetőséget ad arra, hogy a szélsőséges aszályesemények valószínűségéről és várható hatásairól megalapozottabb következtetéseket vonjunk le. Az elemzés részeként részletesen bemutatom a későbbi fejezetekben a használt modellek matematikai hátterét, amelyek közül kiemelt szerepet kapnak az idősoros modellek (ezen belül is főként az ARMA és ARCH/GARCH modellek, amiket a későbbiekben részletesen kifejtek) és az extrémérték-eloszlások (blokk maximum, illetve küszöbérték módszerek) vizsgálatai. Ezek a technikák segítenek pontosabban felmérni a hóhullámok időbeli hatásait, valamint lehetőséget nyújtanak arra, hogy előrejelzéseket készítssek a jövőbeli esetleges károkról. Az eredményeim hozzájárulhatnak a kárcsökkentési stratégiák és biztosítási modellek kidolgozásához, amelyek a fenntartható mezőgazda-

sági termelést támogatják, illetve reményeim szerint hozzájárulhat az éghajlatváltozás következményeinek jobb megértéséhez és a mezőgazdasági alkalmazkodási stratégiák fejlesztéséhez.

2. fejezet

Szakirodalmi áttekintés

A hőmérséklet alakulásának vizsgálata különösen fontos a biztosítási piacon, nagy relevanciája van a mezőgazdasági károk kialakulásában. *Karl Larsson* (2023) szót ejt a hőhullámok vizsgálatának fontosságáról: halált (2010-ben 55 ezer halottat eredményezett Európában), kórházi ápolást okoz, ezen kívül a vasutak ideiglenesen használhatatlanná válhatnak, valamint az áramfelhasználás is megnő, ami áramszünethez vezethet (*Karl Larsson, 2023*). A hőhullámokhoz köthető károk jelentősek, (a 2003, 2010, 2015 és 2018-as éveket vizsgálva) az európai GDP 0,3-0,5%-át teszi ki (*Garcia Leon et. al., 2021*). Továbbá a munkaerőre is hatással van, sokkal kevésbé lesz hatékony a gazdaság és ez akadályozza a növekedést. Szintén *Garcia Leon* (2021) szemlélteti, hogy 1960 és 2019 között a meleg napok száma duplájára nőtt (meleg napok esetén a küszöbértéket az 1970 és 2000 közti időszak hőmérsékletének a 90. percentilisét vette). A WMO (*World Meteorological Organization*) (2021) jelentése szerint a szélsőséges időjárás több száz milliárd dolláros veszteséget okozott, ezeknek a 4%-a a nagy melegnek tudható be. A cikk arra is kitért, hogy az elmúlt hét év volt a valaha legmelegebb hét év (2015-2021). A 2021-es globális átlaghőmérséklet 1,11 °C-kal volt magasabb az iparosodás előtti időszakhoz képest.

A hóhullám definiálása problémát vethet fel, mivel kérdés lehet az hogy hány egymást követő meleg nap esetén, illetve milyen küszöbérték esetén mondhatjuk azt, hogy hóhullám következett be. Egy korábbi tanulmány (*Fenner et. al, 2019*) több definícióját is vizsgálja a hóhullámnak. A legtöbb esetben 3 egymást követő nap esetén van szó hóhullámról, illetve az egyik esetben 6 nap volt ez az idő. Azonban ha bekövetkeznek ilyen események, legtöbb esetben csak rövid ideig lesz magas a hőmérséklet (egymást követő napokat tekintve) (*Perkins és Alexander, 2013*), ezért nem érdemes 6 napos időintervallumot használni. *Fenner et. al* (2019) ezt megerősítik, cikkük szerint a rövidebb de melegebb hóhullámok sokkal gyakoribbak, mint a hosszabb de kevésbé meleg hóhullám. Ezen kívül szintén kérdés lehet, hogy napi maximum hőmérsékletet használjunk, vagy napi átlag hőmérsékletet. A kutatásom célját tekintve érdemesebb lenne napi maximumot használni, mivel ha az átlaghőmérsékletet vesszük az nem szemlélteti annyira jól, hogy milyen meleg elérése esetén fog bekövetkezni valamilyen kár (mezőgazdasági károkat tekintve jelentős különbség lehet, ha a hőmérséklet szórása nagyobb egy adott napon, így talán többet mondó az, ha a maximumot használom az elemzésben és nem az átlagot). Emellett érvelnek *Fenner et. al* (2019), illetve *Smoyer, Tomic* (2003). Vannak korábbi cikkek (*Cao és Wei 2001, Torro Meneu és Valor 2001*), amikben az átlaghőmérsékletet használják, viszont ezek a modellek sokkal korlátozottabbak voltak (pl. az átlaghőmérséklet kisimítja az extrém értékeket, egy 40 fokos délután már nagy kárt okozhat), illetve több feltevessel éltek a szerzők.

Az így definiált biztosítási szerződéseket (ahol egy előre meghatározott kifizetés történik egy adott szélsőséges időjárási körülmény bekövetkezése esetén) a pénzügyi világban "időjárási derivatíváknak" nevezzük. Ebben az esetben az alaptermék (aminek az alakulása határozza meg a kifizetést) lehet bármilyen időjáráshoz köthető mérhető dolog, ilyen lehet például a hőmérséklet, csapadék vagy a havazás is (*Campbell és Diebold, 2005*).

A kutatásban fontos lehet az is hogy a fentebb meghatározott paramétereket (hőhullám hossza, hőmérsékletre vonatkozó küszöbérték, napi maximum hőmérséklet) több esetben is megvizsgáljuk. *Karl Larsson* (2023) esettanulmányában megvizsgálja (Berlinben) a 3-tól 11 nap hosszú tartó hőhullámokat, illetve a számuknak az alakulását különféle küszöbértékek esetén (ez 30 °C és 36 °C között 1 fokos növekményekkel minden lehetséges kombináció). Ezen tanulmány is alátámasztja, hogy jellemzőbb a rövidebb de melegebb időszak.

Kérdést vethet fel az is, hogy az országon belül a hőmérséklet nem mindenhol ugyanakkora, illetve hogy ez esetben melyik területen mért hőmérséklettel dolgozunk. A városokban jellemzően magasabb a hőmérséklet (*Fenner et. al, 2014*), viszont a mezőgazdasági területek pedig máshol helyezkednek el jellemzően. A térbeliség hatását is figyelembe kell venni (*spatial variation*), így a kutatásomban Magyarország 3 különböző pontjában mért maximum hőmérsékleteknek a minimumát fogom venni ennek a problémának a kiküszöbölésére. Ez a három város Budapest, Debrecen, illetve Nagykanizsa. Ezt a térbeli hatást *Karl Larsson* (2023) szintén így vonja be az elemzésébe (ő is 3 külön városban mért maximum hőmérsékletet használ), így azt gondolom, hogy indokolt a szakdolgozatomban is ezt a módszert alkalmazni.

3. fejezet

Matematikai módszertan

3.1. Biztosítási szerződés felépítése

Az alábbi matematikai modellt *Karl Larsson* (2023), *Parametric Heatwave Insurance* című tanulmánya alapján állítottam fel. A korábbi fejezetben meghatározott hőhullám definíciója szerint fogom felépíteni a modellt.

3.1.1. Hőhullám definiálása

Definíció. Legyen $A(n, a, \tau, Z) = \{Z_t \geq a \text{ legalább } n \text{ egymást követő napon keresztül, } t \in \tau\}$, ahol:

n = napok száma

a = hőmérséklet küszöbértéke

τ = mérési időszak (pl. nyári hónapok)

Z_t = napi maximum hőmérséklet t időben ($^{\circ}\text{C}$)

Tehát azokat az eseményeket nézzük ahol a napi hőmérséklet meghalad egy előre rögzített küszöbértéket. Az előző fejezetben említett térbeliség bevonása miatt Z_t helyett $\tilde{Z}_t = \min(Z_t^1, Z_t^2, Z_t^3)$, ahol a felső index a 3 külön helyen mért napi maximum

hőmérsékletet indikálja. Az a küszöbértéket pedig kétféleképpen lehet meghatározni, statikusan és dinamikusan. A statikus módszernél csak előre meghatározzuk az egyes küszöbértékeket (pl. 28, 30, 32, 34, 36), a dinamikus módszernél pedig az elmúlt évek tapasztalataira támaszkodunk: $P(Z_t \leq a) = 0,9$, azaz az idősor 90. percentiliséét határozzuk meg, különböző időszakokra visszanyúlva (pl. 2000-2024-ig vett idősor 90. percentilise, 2010-2024-ig vett idősor 90. percentilise és így tovább). Továbbá a napok száma meghatározásához is több értéket fogok használni. A korábban említett tanulmányokra hivatkozva 3 napra veszem a legrövidebb hőhullámot, majd megvizsgálom itt is különféle értékeket. A τ esetén pedig a május-szeptember időszakot fogom megvizsgálni. Fontos megemlíteni, hogy diszjunkt és különálló hőhullámokat vizsgálom (ez annyit jelent hogy például egy 6 napos hőhullámot nem szedhetek szét két 3 naposra, ez a biztosítási szerződés kifizetésénél játszik fontos szerepet). A szerződés kifizetéséhez meg kell határozni a bekövetkezett biztosítási események számát, ennek a definícióját a következő fejezetben határozom meg.

3.1.2. Szerződés definiálása

Definíció. Legyen $\Gamma(n, a, \tau, Z)$ az $A(n, a, \tau, Z)$ események száma a τ mérési időszakban.

Ezek után tudjuk egy adott biztosítási szerződés kifizetés függvényét felírni:

Definíció. Legyen $\Phi(T, K, n, a, \tau, Z) = K \cdot \mathbf{1}_{\Gamma(n, a, \tau, Z) \geq 1}$, ahol

$\mathbf{1}$ = az esemény indikátorfüggvénye

K = kifizetés mértéke kár bekövetkezése esetén

T = szerződés lejárat dátuma

A definícióban fontos megemlíteni, hogy itt több káresemény bekövetkezése esetén

is csak egy kifizetést feltételez a modell. A lejárat dátumot azért használom a kifizetésfüggvényben, hogy kiküszöböljem azokat az eseteket, amikor a káresemény nem a szerződéses időszakba esik bele, ebben az esetben ez a mérési időszak vége lesz, tehát $T = \tau_b$, ha $\tau = [\tau_a, \tau_b]$. Ez a biztosítási szerződés azonban egy leegyszerűsített feltételezéssel él, miszerint több káresemény esetén is csak egy kifizetés van, így definiálok egy másik kifizetésfüggvényt, ami már más kifizetést eredményez, ha több káresemény következik be.

Definíció. Legyen $\Phi_m(T, K_m, n, a, \tau, Z) = K_m \cdot \mathbf{1}_{\Gamma(n, a, \tau, Z)=m}$, $m = 0, 1, \dots, N(n, \tau)$, ahol

$m = \text{káresemény száma}$

$K_m = \text{kifizetés mértéke } m \text{ káresemény esetén}$

$N(n, \tau) = \text{a mérési időszakban a maximális lehetséges hőhullámok száma (ez szerződésben előre meghatározott paraméter)}$

Triviális hogy a mérési időszakban lehetséges maximális hőhullámok száma függ az n és τ paraméterektől.

3.1.3. Szerződés kiértékelése

Ahogy a korábbi fejezetben is említésre került (*Campbell és Diebold (2005)*), egy ilyen biztosítási szerződés tekinthető egy időjárási derivatívának, ahol a hőmérséklet az alaptermék. Ahhoz hogy ki tudjuk értékelni a fent definiált szerződéseket, abból kell kiindulnunk, hogy Z_t egy valószínűségi változó az $(\Omega, \mathcal{F}_t, \mathbb{P})$ filtrált valószínűségi mezőn, ahol $\mathcal{F}_t = \sigma(Z_s, s \leq t)$ a Z természetes filtrációja, és $\mathbb{P}$ egy valószínűségi mérték. Ahhoz hogy meghatározzuk a szerződés árát a $\Phi(T, K, n, a, \tau, Z)$ várható jelenértékét egy másik ekvivalens mérték alatt kell meghatározni, azaz $\mathbb{E}^{\mathbb{Q}}(\Phi(T, K, n, a, \tau, Z))$, ahol $\mathbb{Q} \sim \mathbb{P}$. Ezt nevezzük kockázatmentes mértéknek. Ez a modell akkor

használandó, ha azt feltételezzük, hogy a hőmérséklet ingadozásából fakadó kockázatnak van ára a piacon. Mivel a hőmérséklet nem egy általános pénzügyi termék, amivel kereskednek a piacon, így ezt az ekvivalens mértéket nagyon nehéz lenne meghatározni. Az egyszerűség kedvéért feltesszük, hogy $\mathbb{P} = \mathbb{Q}$. Ezt több másik kutatásban is feltételként szerepel (*Campbell és Diebold 2005, Karl Larsson 2023*), így én is ezzel fogok tovább dolgozni.

Definíció. Legyen $V(t, T) = e^{-r \cdot (T-t)} \cdot \mathbb{E}(K \cdot \mathbf{1}_{\Gamma(n, a, \tau, Z) \geq 1} \mid \mathcal{F}_t)$, ahol folytonos kamatozást feltételezünk és r ennek megfelelően valamilyen log hozam és T a lejárat.

Ekkor $V(t, T)$ a várható kifizetés diszkontált jelenértéke lesz. Fontos megemlíteni, hogy itt azzal a feltételezéssel él a modell, hogy a szerződés lejáratakor történik a kifizetés. Kihasználva azt, hogy $K \sim \mathcal{F}_t$ (mérhető), és hogy indikátor várható értéke megegyezik az esemény valószínűségével, ezt tovább tudjuk írni:

$$V(t, T) = e^{-r \cdot (T-t)} \cdot K \cdot \mathbb{P}(\Gamma(n, a, \tau, Z) \geq 1 \mid \mathcal{F}_t) \quad (3.1)$$

A másik szerződés típusnál analóg módon:

$$V_m(t, T) = e^{-r \cdot (T-t)} \cdot K_m \cdot \mathbb{P}(\Gamma(n, a, \tau, Z) = m \mid \mathcal{F}_t), \quad m = 0, 1, \dots, N(n, \tau) \quad (3.2)$$

A kutatásom második felében a kifizetés becslésére fogok fókuszálni és egy változóként fogok rá tekinteni. Ez azonban annyi módosítást igényel a modellben, hogy K helyett $\mathbb{E}(K)$ fog szerepelni az egyenletben, ez persze konstans lesz, így szintén kiemelhető a mérhetőség miatt:

$$V(t, T) = e^{-r \cdot (T-t)} \cdot \mathbb{E}(K) \cdot \mathbb{P}(\Gamma(n, a, \tau, Z) \geq 1 \mid \mathcal{F}_t) \quad (3.3)$$

A szakdolgozatom célja elsősorban az, hogy megbecsüljem a jövőbeli károkat, ezzel a modellel viszont a jövőbeli károk ismeretében a szerződés árát lehet meghatározni. Kérdés, hogy K értéke mekkora lehet, azaz ha bekövetkezik egy káresemény, akkor mekkora veszteséget okoz a mezőgazdaságban. Tehát a K -ra adott becslés eredménye fogja kiegészíteni ezt a modellt, és akkor kapjuk meg, hogy mekkora lesz a várható kár. Ennek meghatározása előtt a hőmérsékletet szeretném előrejelezni, hogy egy pontos becslést kapjak a következő kárvalószínűségre vonatkozóan.

3.2. Hőmérséklet előrejelzése sztochasztikusan

A hőmérséklet függ a véletlentől, így ahhoz hogy előre tudjuk jelezni a hőmérsékletet, fel kell írjuk a folyamatot egy determinisztikus, illetve egy sztochasztikus rész összegeként:

$$Z_t = S_t + X_t \quad (3.4)$$

ahol S_t az idősor determinisztikus része, és X_t az idősor sztochasztikus része. *Karl Larsson* (2023)-ban additív modellt feltételez a berlini hőmérsékletre, így az idősorban én is ilyen modellel fogok dolgozni, ami annyit jelent, hogy az idősor determinisztikus része a trend és szezonális hatások összegeként áll elő:

$$S_t = T_t + G_t \quad (3.5)$$

ahol T a trendhatás, G pedig a szezonális hatás. Az utóbbi 50 évben nem mutatható ki szignifikáns gyorsulás a globális felmelegedésben (*Communications Earth and Environment*, 2024), így a T_t folyamatra egy lineáris modellt fogok illeszteni a trend megragadásához. A szezonalitást Fourier sorokkal fogom közelíteni, mivel minden

periodikus függvény felírható szinusz/koszinusz függvények összegeként:

$$T_t = \beta_0 + \beta_1 t \quad (3.6)$$

$$G_t = \sum_{n=1}^N (\alpha_n \cdot \cos\left(\frac{2\pi n t}{P}\right) + \gamma_n \cdot \sin\left(\frac{2\pi n t}{P}\right)) \quad (3.7)$$

ahol P az idősor periodicitása, az én esetemben 365, α_n és γ_n pedig a szinusz és koszinusz tagok súlyai. Az idősor sztochasztikus része pedig $X_t = Z_t - S_t$ lesz és az X folyamatról azt feltételezem hogy egy $AR(p)$ folyamat, *Karl Larsson, (2023)*-ra hivatkozva (azaz függ a saját múltjától p késleltetésig), így felírható a következőképpen:

$$X_t = \sum_{i=1}^p \lambda_i X_{t-i} + u_t \quad (3.8)$$

ahol $u_t = \sigma_t \epsilon_t$ és ϵ_t 0 várható értékű, független azonos eloszlású változó. Ezt az X folyamatot nevezzük p -rendű autoregresszív folyamatnak. Fel tudjuk írni Z_t folyamatot a következőképpen:

$$Z_t = \eta_t + \sigma_t \epsilon_t \quad (3.9)$$

ahol η_t és σ_t $\mathcal{F}_{t-1}$ mérhetőek, $\epsilon_t \sim N(0, 1)$ és függetlenek egymástól és a múlttól. Ekkor η_t lesz Z_t feltételes várható értéke, σ_t pedig a feltételes szórásnégyzete $\mathcal{F}_{t-1}$ -re nézve. A fentiekből következik az alábbi:

$$\eta_t = S_t + \sum_{i=1}^p \lambda_i (Z_{t-i} - S_{t-i}) \quad (3.10)$$

Az u_t hibatag szórása időben nem állandó. Ezeknek a hatásoknak a kezelésére szolgálnak az ARCH folyamatok. A modellben az ARCH hatás feltételezése nem triviális, viszont korábbi tanulmányra támaszkodva (*Ahcan, 2012*) a hőmérsékletben

van ARCH hatás, így azt gondolom, hogy jogos lehet ennek a modellnek az illeszkedését is megvizsgálni.

3.2.1. Az ARCH folyamatok

A fentebb említett időben nem állandó szórás kezelésére alkalmasak az ARCH folyamatok. Az alábbi modellt *Robert F. Engle*, (1982) alapján állítottam össze. Tekintsük a (3.9)-es egyenletet, ahol az η_t folyamattal megragadtuk az idősorban a trendet, szezonális hatást, illetve maradéktagban lévő AR hatást. Ennek a modellnek a hibatagjairól feltételezzük, hogy időben nem állandó a szórása (más szóval heteroszkedasztikus), azaz σ_t időfüggő változó. Induljunk ki egy egyszerű AR(1) folyamatból:

$$u_t = \lambda u_{t-1} + \epsilon_t \quad \epsilon_t \sim WN(0, \sigma^2) \quad (3.11)$$

Az idősoros modellekben az előrejelzések a feltételes várható értékek illetve feltételes varianciák:

$$\mathbb{E}(u_t \mid u_{t-1}) = \mathbb{E}(\lambda u_{t-1} + \epsilon_t \mid u_{t-1}) = \mathbb{E}(\lambda u_{t-1} \mid u_{t-1}) + \mathbb{E}(\epsilon_t \mid u_{t-1}) = \lambda u_{t-1} \quad (3.12)$$

A második lépésnél azt használtuk fel, hogy a λu_{t-1} mérhető a feltételre nézve, illetve, hogy az ϵ_t független a feltételtől. A feltételes szórásnégyzetre is hasonlóan a mérhetőséget és függetlenséget felhasználva:

$$D^2(u_t \mid u_{t-1}) = D^2(\lambda u_{t-1} + \epsilon_t \mid u_{t-1}) = D^2(\lambda u_{t-1} \mid u_{t-1}) + \sigma^2 = \sigma^2 \quad (3.13)$$

Engle (1982) a heteroszkedaszticitás megragadásához bevezet egy exogén változót (ϕ_t), ami megbecsüli a varianciát azzal a feltétellel, hogy $\mathbb{E}(\phi_t) = 0$. Ebben az

esetben a modell így fog kinézni:

$$u_t = \epsilon_t \phi_{t-1} \quad (3.14)$$

Ebből következik, hogy az u_t szórásnégyzete $D^2(u_t) = \phi_{t-1}^2 \sigma_t^2$ lesz. Az előrejelzés hibája egy exogén változótól fog függeni. Ezzel a megközelítéssel az a probléma, hogy megköveteli a változékonny variancia okainak a meghatározását, ahelyett hogy úgy kezelné a feltételes várható értéket és feltételes szórásnégyzetet, mint egy időben változó értéket. Ennek a problémának a kiküszöbölésére alkalmas az alábbi modell:

$$u_t = \sigma_t \epsilon_t \quad \epsilon_t \sim N(0, 1) \quad (3.15)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 \quad (3.16)$$

ahol, ϵ_t -k függetlenek egymástól. A (3.15)-ös egyenlettel leírjuk az idősort, gyakorlatilag egy standard normális eloszlású változót szorzunk meg minden időpontban a σ_t változóval (ami megadja az idősor volatilitását), és a σ_t^2 -re írunk fel egy egyenletet, amivel megragadjuk az időben nem állandó szórást. Ez az ARCH(1) modell, általánosságban az ARCH(p) modell pedig így néz ki:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2 \quad \alpha_i > 0, \quad i = 0, 1, \dots, p \quad (3.17)$$

Ennek a továbbfejlesztett verziója a GARCH(p, q), amit *Bollerslev* (1986) írt le. Ez a modell annyiban különbözik az ARCH-tól, hogy a varianciát önmagának a múltjával is magyarázzuk:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i u_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad \alpha_i, \beta_j > 0, \quad i = 0, \dots, p \quad j = 0, \dots, q \quad (3.18)$$

A (3.17) és (3.18) egyenletek az idősor szórását kezelik (és az idősor trendjének időbeli változását, illetve az autokorrelációt nem kezeli), így ezt olyan változókra szeretnénk ráilleszteni, amiben már nincsen semmilyen trend, illetve szezonális hatás.

Definíció. Az $X_1, X_2, \dots$ idősor gyengén stacionárius, ha $\mathbb{E}(X_t) = \text{konstans}$ és $\text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_0, X_h)$, tehát csak a két időpont különbségétől függ, azaz eltolás invariánsak.

A GARCH(p, q) egyenlet megoldására *Bollerssev* feltételt is adott, hogy mikor lesz gyengén stacionárius:

Tétel. A GARCH(p, q) egyenlet megoldása gyengén stacionárius, ha:

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$$

$$\text{Ekkor } u_t \text{ fehérzaj és } D^2(u_t) = \frac{\omega}{1 - \left(\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \right)} < \infty$$

Ha nem stacionárius idősorra illesztünk ARCH/GARCH modellt, akkor a modellfeltételek sérülnek és nem lesz pontos az előrejelzés. A továbbiakban az egyszerűség kedvéért az ARCH modellel fogok dolgozni. A korábban bevezetett $\mathcal{F}_t$ filtrációra tekinthetünk úgy is, mint a t időpontig rendelkezésre álló információ. Itt σ_t értéke kiszámítható a múltbeli értékekből és ϵ_t pedig standard normális eloszlású, valamint független a múlttól, azaz $\mathcal{F}_{t-1}$ -től. A modellben feltesszük, hogy az $u_t \mid \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$, azaz a múltba vett feltételes eloszlása normális. Ekkor a σ_t^2 felírható az alábbi változók függvényeként:

$$\sigma_t^2 = \sigma(u_{t-1}, u_{t-2}, \dots, u_{t-p}, \alpha) \quad (3.19)$$

ahol, p az ARCH modell rendje, illetve α az ismeretlen paraméter vektor. Általánosságban azt feltételezzük, hogy $\mathbb{E}(u_t) = \phi_t \beta_t$, azaz késleltetett endogén és exogén

változók lineáris kombinációjaként áll elő, amik benne vannak az $\mathcal{F}_{t-1}$ σ -algebrában. Ebben az esetben persze $u_t \mid \mathcal{F}_{t-1} \sim N(\phi_t \beta_t, \sigma_t)$ lenne, azonban a mi esetünkben az AR(p) modellből kapott hibatagnak az időfüggő varianciáját szeretnénk vizsgálni, a hibatagnak a várható értéke pedig 0, ezért elegendő a (3.15)-es egyenletben felírt esetet vizsgálni, ahol a várható érték 0.

3.2.2. A Likelihood függvény

Tekintsük a korábban már definiált ARCH(1) folyamatot:

$$u_t = \sigma_t^{\frac{1}{2}} \epsilon_t \quad (3.20)$$

Az egyszerűség kedvéért használtam itt $\sigma_t^{\frac{1}{2}}$, hogy később a számolásnál σ_t -vel tudjak számolni a variancia esetében. Itt azt feltételeztük, hogy az u_t $\mathcal{F}_{t-1}$ -re vett feltételes eloszlása normális, azonban ebből nem következik az, hogy az u_t vektor együttesen normális eloszlású. Az ARCH(p) modellben a paramétereket maximum likelihood módszerrel becsüljük, így ahhoz, hogy a likelihood függvényt meg tudjuk határozni szükségünk van az együttes sűrűségfüggvényre. Mivel az u_t hibatagok egymástól függetlenek, így az együttes sűrűségfüggvényt a feltételes sűrűségfüggvények szorzataként kapjuk meg. A log-likelihood függvényt pedig a feltételes log-likelihoodok összegeként kapjuk meg. Legyen l az átlagos log-likelihood, l_t pedig a t. elem log-likelihoodja feltéve, hogy T elem van. Az u_t feltételes sűrűségfüggvényét ismerjük, ebből persze az együttes sűrűségfüggvényt ezeknek a produktumából kaphatjuk meg:

$$f_{u_1, \dots, u_t} = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_t}} e^{\frac{-u_t^2}{2\sigma_t}} \quad (3.21)$$

Itt egyébként lehetne feltételezni t eloszlást is az u_t hibatagokra, természetesen akkor annak megfelelően kellene változtatni a feltételes sűrűségfüggvényeket. A továbbiakban én normális eloszlást fogok feltételezni a hibatagokra. A log-likelihood függvény pedig a következőképpen fog kinézni:

$$l(u_1, \dots, u_T) = \sum_{i=1}^T \left(-\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma_t - \frac{u_t^2}{2\sigma_t} \right) \quad (3.22)$$

Ebből:

$$l_t = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma_t - \frac{u_t^2}{2\sigma_t} \quad (3.23)$$

$$l = \frac{1}{T} \sum_{t=1}^T l_t \quad (3.24)$$

Az ismeretlen α paramétervektor megbecsléséhez maximalizálnunk kell a likelihoodokat (a konstansoknak nincs relevanciájuk ebben az esetben, így azt már el is hagyhatjuk). Nézzük az elsőrendű deriváltat:

$$\frac{\partial l_t}{\partial \alpha} = -\frac{1}{2\sigma_t} \frac{\partial \sigma_t}{\partial \alpha} + \frac{u_t^2}{2\sigma_t^2} \frac{\partial \sigma_t}{\partial \alpha} = \frac{1}{2\sigma_t} \frac{\partial \sigma_t}{\partial \alpha} \left(\frac{u_t^2}{\sigma_t} - 1 \right) \quad (3.25)$$

A másodrendű derivált pedig pont a Hesse mátrixot fogja megadni nekünk:

$$\frac{\partial^2 l_t}{\partial \alpha \partial \alpha^\top} = -\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t}{\partial \alpha} \frac{\partial \sigma_t}{\partial \alpha^\top} \frac{u_t^2}{\sigma_t} + \left(\frac{u_t^2}{\sigma_t} - 1 \right) \frac{\partial}{\partial \alpha^\top} \left(\frac{1}{2\sigma_t} \frac{\partial \sigma_t}{\partial \alpha} \right) \quad (3.26)$$

Ha a második tagnak a feltételes várható értékét vesszük $\mathcal{F}_{t-1}$ -re, akkor az 0 lesz:

$$\mathbb{E} \left(\frac{u_t^2}{\sigma_t} - 1 \mid \mathcal{F}_{t-1} \right) = \mathbb{E} \left(\frac{u_t^2}{\sigma_t} \mid \mathcal{F}_{t-1} \right) - 1 = \frac{1}{\sigma_t} \mathbb{E}(u_t^2 \mid \mathcal{F}_{t-1}) - 1 \quad (3.27)$$

Itt felhasználtuk, hogy $\sigma_t \sim \mathcal{F}_{t-1}$. Ha u_t^2 előállítását behelyettesítjük:

$$\frac{1}{\sigma_t} \mathbb{E}(\sigma_t \epsilon_t^2 \mid \mathcal{F}_{t-1}) - 1 = \frac{1}{\sigma_t} \sigma_t \mathbb{E}(\epsilon_t^2) - 1 = D^2(\epsilon_t) + \mathbb{E}^2(\epsilon_t) - 1 = 0 \quad (3.28)$$

Itt szintén felhasználtuk σ_t mérhetőségét, illetve azt is hogy ϵ_t független az $\mathcal{F}_{t-1}$ σ -algebrától. Ugyan ezzel a gondolatmenettel az $\frac{u_t^2}{\sigma_t}$ feltételes várható értéke pedig 1 lesz. Mivel a Fischer-féle információs mátrix a Hesse mátrix negatív várható értéke az összes megfigyelésre vonatkozóan, így a következőképpen fog kinézni ebben az esetben:

$$\mathbb{I}(\alpha) = -\mathbb{E} \left(\frac{\partial^2 l(\alpha)}{\partial \alpha \partial \alpha^\top} \right) = \mathbb{E} \left(\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t}{\partial \alpha} \frac{\partial \sigma_t}{\partial \alpha^\top} \right) \quad (3.29)$$

A várható érték ML becslése az átlag lesz, így az információs mátrix konzisztens becslése pedig:

$$\hat{\mathbb{I}}(\alpha) = \frac{1}{2T} \sum_t \left(\frac{1}{\sigma_t^2} \frac{\partial \sigma_t}{\partial \alpha} \frac{\partial \sigma_t}{\partial \alpha^\top} \right) \quad (3.30)$$

Ha egy p-rendű ARCH folyamatról van szó, akkor a varianciát fel tudjuk írni vektorokkal is:

$$\sigma_t = z_t \alpha \quad (3.31)$$

ahol $z_t = (1, u_{t-1}^2, \dots, u_{t-p}^2)$ és $\alpha = (\alpha_0, \dots, \alpha_p)^\top$. Itt egyből látszik, hogy $\frac{\partial \sigma_t}{\partial \alpha} = z_t$, amiből a (3.25)-es egyenletbe behelyettesítve következik, hogy $\frac{\partial l_t}{\partial \alpha} = \frac{1}{\sigma_t} z_t \left(\frac{u_t^2}{\sigma_t} - 1 \right)$. Ekkor a Fischer-féle információs mátrix így fog kinézni:

$$\hat{\mathbb{I}}(\alpha) = \frac{1}{2T} \sum_t \left(\frac{1}{\sigma_t^2} z_t^\top z_t \right) \quad (3.32)$$

Az információs mátrix a hibatagok eloszlásában játszik fontos szerepet. Az $\hat{\alpha}$ becült paramétervektor aszimptotikusan normális (ha $T \rightarrow \infty$):

$$\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \mathbb{I}^{-1}(\alpha)) \quad T \rightarrow \infty \quad (3.33)$$

3.3. Extrémérték-elmélet

A következő módszer abban fogja kiegészíteni a modellemet, hogy becslést ad az esetleges extrém károk nagyságára. Az extrémértékek statisztikai modellezésének elméleti hátterét *Embrechts, Klüppelberg és Mikosch* 1997-ben írt könyve alapján mutatom be.

3.3.1. A maximum határeloszlása

Kiindulásként legyenek $X_1, X_2, \dots$ független azonos eloszlású valószínűségi változók, és legyen F a közös eloszlásfüggvény. Továbbá legyen $M_1 = X_1$ és $M_n = \max(X_1, \dots, X_n)$, $n \geq 2$. Fontos megemlíteni, hogy ez az elmélet működhet extrém kicsi értékek modellezésére is, abban az esetben a változók minimumát át tudjuk írni:

$$\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n) \quad (3.34)$$

Az extrém értékek az eloszlás felső végpontjához "közel" vannak, így M_n aszimptotikus viselkedése is az eloszlás széléhez köthető. M_n eloszlása pedig a következőképpen írható fel:

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) = F(x) \cdot \dots \cdot F(x) = F^n(x) \quad (3.35)$$

Definíció. Legyen $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$. Ezt nevezzük az eloszlás felső végpontjának.

Könnyen látható, hogy minden $x < x_F$ -re $\mathbb{P}(M_n < x) = F^n(x) \rightarrow 0$ és minden $x > x_F$ -re $\mathbb{P}(M_n < x) = F^n(x) \rightarrow 1$. Ezen felül M_n határeloszlására vagyunk kíváncsiak. A határeloszlás megállapításához *Fischer* és *Tippet* adott szükséges feltételt. Ehhez viszont előbb be kell vezetnünk egy fogalmat:

Definíció. F eloszlás *max-stabilis*, ha minden n -hez létezik c_n, d_n , hogy $F^n(x) = F(c_n x + d_n)$.

Tétel. Legyenek $X_1, X_2, \dots$ független azonos eloszlású változók, $M_n = \max(X_1, \dots, X_n)$. Léteznek $c_n > 0$ és $d_n \in \mathbb{R}$ sorozatok, hogy: $\mathbb{P}\left(\frac{M_n - d_n}{c_n} < x\right) \rightarrow H(x) \iff H$ max-stabilis eloszlású

Tétel. Legyenek $X_1, X_2, \dots$ független azonos eloszlású változók, $M_n = \max(X_1, \dots, X_n)$. Ha léteznek $c_n > 0$ és $d_n \in \mathbb{R}$ sorozatok, hogy: $\frac{M_n - d_n}{c_n} \xrightarrow{d} H$ (eloszlásbeli konvergencia), ahol H nem elfajuló eloszlás, akkor H biztosan általánosított extrémérték eloszlás (továbbiakban *GEV*).

Az általánosított extrém érték eloszlások eloszlásfüggvénye a következőképpen néz ki:

$$H_{\xi, \mu, \sigma}(x) = \begin{cases} e^{-(1+\xi \frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}}, & \xi \neq 0 \\ e^{-e^{-\frac{x-\mu}{\sigma}}}, & \xi = 0 \end{cases} \quad \mu \in \mathbb{R}, \sigma > 0, \xi \in \mathbb{R}, \quad 1 + \xi \frac{x-\mu}{\sigma} > 0 \quad (3.36)$$

Az eloszlás ξ paraméter függvényében 3 különböző esetre bontható szét:

1. Fréchet, ha $\xi = \frac{1}{\alpha} > 0$:

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ e^{-x^{-\alpha}}, & x > 0 \end{cases} \quad \alpha > 0 \quad (3.37)$$

2. Weibull, ha $\xi = -\frac{1}{\alpha} < 0$:

$$\Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0 \quad (3.38)$$

3. Gumbel, ha $\xi = 0$:

$$\Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R} \quad (3.39)$$

Az utóbbi felsorolásban elhagytam a normáló tényezőket, mert az eloszlások bemutatása szempontjából nem volt releváns (ezek a paraméterek csak az eloszlás várható értékét és szórását állítják a kívánt értékekre, viszont az eloszlás sűrűségfüggvényének alakját nem változtatják), a gyakorlatban persze ugyan úgy ki kell vonni a μ -t és osztani kell a σ -val, hogy a megfelelő eloszlásfüggvénnyel tudjunk számolni. Fontos megemlíteni, hogy a 3 eloszlás között természetesen van kapcsolat: $X \sim \Phi_\alpha \iff \ln X^\alpha \sim \Lambda \iff -X^{-1} \sim \Psi_\alpha$. Tekintsük azt az esetet, amikor X Fréchet eloszlású:

$$\mathbb{P}(X < x) = e^{-x^{-\alpha}}, \quad x > 0 \quad (3.40)$$

$$\mathbb{P}(\ln X^\alpha < x) = \mathbb{P}(X^\alpha < e^x) = \mathbb{P}(X < e^{\frac{x}{\alpha}}) = e^{-e^{-x}} \quad (3.41)$$

A másik ekvivalencia ugyan ilyen gondolatmenettel kiszámítható. Az eloszlás várható értéke $\xi < 1$ -re lesz csak véges:

$$\mathbb{E}(X) = \begin{cases} \mu + \frac{\sigma(\Gamma(1-\xi)-1)}{\xi}, & \xi \neq 0, \quad \xi < 1 \\ \mu + \sigma\gamma, & \xi = 0 \\ \infty, & \xi \geq 1 \end{cases} \quad (3.42)$$

Itt γ az Euler-Macheroni konstansot, a Γ pedig a gamma függvényt jelöli. A variancia pedig $\xi < \frac{1}{2}$ esetén lesz véges:

$$\mathbb{D}^2(X) = \begin{cases} \sigma^2 \frac{\Gamma(1-2\xi)-\Gamma^2(1-\xi)}{\xi^2}, & \xi \neq 0, \quad \xi < \frac{1}{2} \\ \sigma^2 \frac{\pi^2}{6}, & \xi = 0 \\ \infty & \xi \geq \frac{1}{2} \end{cases} \quad (3.43)$$

3.3.2. Küszöb feletti meghaladások módszere

Az extrém károk modellezésére létezik egy másik megközelítés is, ahol nem a maximum eloszlását keressük meg, hanem definiálunk egy megfelelő küszöbértéket, és az afeletti értékeket tekintjük extrém értékeknek. Tekintsük egy X változót F eloszlásfüggvénnyel, x_F felső végponttal, valamint legyen u a küszöbérték.

Definíció. Valamilyen fix $u < x_F$ küszöbre legyen

$F_u(x) = \mathbb{P}(X - u < x \mid X > u)$, $0 \leq x < \infty$ a küszöb feletti rész feltételes eloszlása
 $e(u) = \mathbb{E}(X - u \mid X > u)$ a várható küszöb túllépés

A feltételes valószínűség tovább bontható:

$$\mathbb{P}(X - u < x \mid X > u) = \frac{\mathbb{P}(u < X < u + x)}{\mathbb{P}(X > u)} = \frac{F(x + u) - F(u)}{1 - F(u)} \quad (3.44)$$

Hasonlóan a 3.3.1. fejezethez F_u határeloszlását szeretnénk megkeresni $u \rightarrow \infty$ esetén.

Tétel. A fenti feltételek mellett, ha léteznek $a(u) > 0$ és $b(u)$ sorozatok, hogy
 $F_u(a(u)x + b(u)) \rightarrow G(x)$ $u \rightarrow \infty$, ahol G nem elfajuló eloszlás, akkor G általánosított
Pareto eloszláscsaládba tartozik (továbbiakban GPD).

G eloszlásfüggvénye a következőképpen néz ki:

$$G_{\xi, \sigma, u}(x) = \begin{cases} 1 - (1 + \xi \frac{x-u}{\sigma})^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - e^{-\frac{x-u}{\sigma}} & \xi = 0 \end{cases} \quad x \geq 0 \text{ ha } \xi \geq 0, \quad 0 \leq x \leq -\frac{1}{\xi} \text{ ha } \xi < 0 \quad (3.45)$$

A várható érték ebben az esetben is csak $\xi < 1$ esetén, a variancia pedig $\xi < \frac{1}{2}$ esetén lesz véges.

$$\mathbb{E}(X) = u + \frac{\sigma}{1 - \xi}, \quad \xi < 1 \quad (3.46)$$

$$\mathbb{D}^2(X) = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}, \quad \xi < \frac{1}{2} \quad (3.47)$$

Hasonlóan a GEV eloszlásoknál itt is a ξ paraméter alapján tudjuk meghatározni a konkrét eloszlást:

1. $\xi > 0 \Rightarrow X \sim \text{Pareto}$, ahol $\xi = \frac{1}{\alpha}$ és α az eloszlás skálaparamétere
2. $\xi = 0 \Rightarrow X \sim \text{Exp}(\frac{1}{\sigma})$
3. $\xi < 0 \Rightarrow X \sim \text{Béta}$

Pareto eloszlás esetén egy vastag farkú eloszlásról beszélünk (farok eloszlásfüggvénye polinomiális sebességgel tart a nullába), nagyobb valószínűséggel következnek be értékek az eloszlás jobb széléről (nem úgy mint pl. egy lognormális eloszlású változó esetén), míg a Béta eloszlás a jobb szélén gyorsabban lecseng, ami azt jelenti, hogy kisebb eséllyel következnek be a nagyobb értékek. Az exponenciális eloszlás pedig örökifjú tulajdonsága szerint:

$$\mathbb{P}(X > u + x \mid X > u) = \frac{\mathbb{P}(X > u + x)}{\mathbb{P}(X > u)} = \frac{e^{-\lambda(u+x)}}{e^{-\lambda u}} = e^{-\lambda x} = \mathbb{P}(X > x) \quad (3.48)$$

3.3.2.1. Küszöb megválasztása

A küszöb megválasztásához több javasolt módszert is ki fogok próbálni, a következőket *Max Rydan* (2018) alapján dolgoztam fel. Több megközelítést is említ a cikk, ebből az egyik lehet az, ha a megfigyelések 90. percentilisét választjuk meg küszöbnek. Ezen kívül azt is jónak tartja a cikk, ha k extrém megfigyelés és n mintaelemszám esetén a $k = \sqrt{n}$, vagy a $k = \frac{n^{\frac{2}{3}}}{\log(\log n)}$ teljesül. Egy másik módszer esetében az átlagos küszöbmeghaladás alapján döntünk. Integrálással belátható, hogy alkalmas

u küszöbértékre az átlagos küszöbmeghaladási függvény lineáris (u-ban):

$$\mathbb{E}(X - u \mid X > u) = \frac{\sigma + \xi u}{1 - \xi} = \frac{\sigma}{1 - \xi} + \frac{\xi}{1 - \xi}(u - \mu) \quad (3.49)$$

Ehhez azonban szükségünk van a GPD eloszlás sűrűségfüggvényére, ami a következő:

$$f(x) = \begin{cases} \frac{1}{\sigma} \left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)^{-1-\frac{1}{\xi}}, & \xi \neq 0 \\ \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}}, & \xi = 0 \end{cases} \quad (3.50)$$

Látható, hogy $\xi \geq 0$ esetén $\mu \leq x < \infty$. $0 < \xi < 1$ -re és $u \geq \mu$ -re

$$e(u) = \mathbb{E}(X - u \mid X > u) = \frac{\mathbb{E}(X - u)}{\mathbb{P}(X > u)} = \frac{\int_u^\infty (x - u) \frac{1}{\sigma} \left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)^{-1-\frac{1}{\xi}} dx}{\int_u^\infty \frac{1}{\sigma} \left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)^{-1-\frac{1}{\xi}} dx} \quad (3.51)$$

Legyen $y = 1 + \xi \left(\frac{x-\mu}{\sigma}\right)$, ekkor $dx = \frac{\sigma}{\xi} dy$ és $x = \mu - \frac{\sigma}{\xi} + \frac{\sigma y}{\xi}$, továbbá legyen $w = 1 + \xi \frac{u-\mu}{\sigma}$.

A helyettesítést elvégezve

$$e(u) = \frac{\int_w^\infty \frac{1}{\xi} \left(\frac{\sigma y}{\xi} + \mu - \frac{\sigma}{\xi} - u\right) y^{-1-\frac{1}{\xi}} dy}{\int_w^\infty \frac{1}{\xi} y^{-1-\frac{1}{\xi}} dy} = \frac{\sigma}{\xi} \frac{\int_w^\infty y^{-\frac{1}{\xi}} dy}{\int_w^\infty y^{-1-\frac{1}{\xi}} dy} + \mu - u - \frac{\sigma}{\xi} \quad (3.52)$$

Innen az integrálok könnyen kiszámolhatóak, így ebből azt kapjuk, hogy:

$$e(u) = \frac{\sigma}{1 - \xi} + \frac{\xi}{1 - \xi}(u - \mu) \quad (3.53)$$

Ebből pedig látszik, hogy küszöb feletti értékek lineárisak (a kapott egyenlet u-ban lineáris), így ennek megfelelően is tudunk választani megfelelő küszöböt (úgy kell megválasztani, hogy az $e(u)$ függvény lineáris legyen u-ban). Ezt azonban inkább csak vizuálisan lehet a gyakorlatban.

3.3.3. Paraméterek becslése

A paraméterek becslése alapvetően nem extrémérték elméleti téma, de a modellemben fontosnak tartom ennek a meghatározását. Mivel csak korlátozott számban vannak adataim a kárkifizésekről, így nem maximum-likelihood módszerrel becsültem meg a modell paramétereit, hanem L-momentumok módszerével. Ez a módszer nem valószínűségi alapon számolja ki az optimális paramétereket, így konfidenciaintervallumot sem tudunk ebből számolni, viszont kevés megfigyelés esetén sokkal robusztusabb ez a módszer, mint a maximum-likelihood (*Urych (2000)*). Legyen $X_{r:n}$ az n elemű mintából a r . legkisebb elem (az én esetemben az $n - r + 1$. legnagyobb kárkifizetés). Ekkor az r . L-momentum a következő:

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathbb{E}(X_{r-k:r}) \quad (3.54)$$

A várható érték pedig a következő képlet alapján számolható ki:

$$\mathbb{E}(X_{r:n}) = r \binom{n}{r} \int_0^1 x(u) u^{r-1} (1-u)^{n-r} du \quad (3.55)$$

$u = F(x)$ helyettesítéssel, ahol $x(u)$ az X kvantilisfüggvénye:

$$x(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}, \quad 0 < u < 1 \quad (3.56)$$

Legyenek továbbá $\tau = \frac{\lambda_2}{\lambda_1}$, és $\tau_r = \frac{\lambda_r}{\lambda_2}$, $r \geq 3$. Ebben az esetben τ az eloszlás volatilitását, τ_3 a ferdeségét és τ_4 a csúcsosságát írja le. A λ_r értékek segítségével fogjuk megbecsülni az eloszlásoknak a paramétereit:

$$1. \quad \xi = \frac{3\tau_3 - 1}{\tau_3 + 1}$$

$$2. \quad \sigma = (1 - \xi)(2 - \xi)\lambda_2$$

$$3. \mu = \lambda_1 - (2 - \xi)\lambda_2$$

A küszöbmeghaladások módszerénél μ ismert, megegyezik az alkalmasan megválasztott küszöb értékével (u). Ebben az esetben:

$$\xi = 2 - \frac{\lambda_1 - \mu}{\lambda_2} \quad (3.57)$$

$$\sigma = (1 - \xi)(\lambda_1 - \mu) \quad (3.58)$$

Az én esetemben a λ_1 , λ_2 és λ_3 értékekre lesz csak szükség. Feltéve, hogy van egy m elemű minta a 3.54 egyenletben leírt paraméterek a következő formulával becsülhetők:

$$\hat{\lambda}_1 = b_0 \quad (3.59)$$

$$\hat{\lambda}_2 = 2b_1 - b_0 \quad (3.60)$$

$$\hat{\lambda}_3 = 6b_2 - 6b_1 - b_0 \quad (3.61)$$

ahol,

$$b_r = \frac{1}{m} \sum_{j=r+1}^m \frac{(j-1) \cdot (j-2) \cdot \dots \cdot (j-r)}{(m-1) \cdot (m-2) \cdot \dots \cdot (m-r)} X_{j:m}, \quad r \geq 1 \quad (3.62)$$

$$b_0 = \frac{1}{m} \sum_{j=1}^m X_{j:m}, \quad r = 0 \quad (3.63)$$

A kutatásom során heti, illetve havi bontású adattal is dolgoztam, előbbinél volt sok megfigyelésem, így ott alkalmazható volt a maximum-likelihood becslés, a havi adatoknál viszont jobbnak gondoltam, ha az L-momentumok módszerével becsülöm meg a paramétereket. *Delicado és Goría* (2008) részletesen kifejtette, hogy kisebb minták esetén ($n \leq 100$) miért érdemesebb az L-momentumok módszerét alkalmazni a maximum-likelihood helyett.

3.4. Kárkifizetés becslése

A 3.2 és 3.3 fejezetekben részletesen tárgyalt módszereket fogom most egyben felhasználni, hogy egy becslést kapjak a hőhullámok által okozott károkra. Ahogy korábban is felírtam, a hőmérséklet idősorát fel lehet bontani egy determinisztikus, illetve egy sztochasztikus rész összegeként: $Z_t = S_t + X_t$, ahol S_t tartalmazza a idősor determinisztikus trend és szezonális hatását, míg X_t pedig az idősorban rejlő AR és ARCH hatást, valamint a már nem modellezhető fehérzaj hibatagot. A modell előrejelzésében szimulációt használok a sztochasztikus részhez és megvizsgálom, hogy az összes esetben hányszor fordult elő egy adott hőhullám (pl. előrejelzek a modellel 10000-szer, és megnézem, hogy hány előrejelzésben fordult elő, hogy legalább n egymást követő napon a hőmérséklet elért egy küszöbértéket). A vizsgált időszakot leszűkítem a május-szeptember időintervallumra. Az így előállított szimulációval kapok egy becslést egy adott hőhullám bekövetkezésének valószínűségére. A kárkifizetés becslésére, pedig az extrém-érték elmélettel felállított modellből becsülöm meg:

$$\mathbb{P}(X > z \mid X > u) = \frac{\mathbb{P}(X > z)}{\mathbb{P}(X > u)} \quad (3.64)$$

Ebből átrendezéssel megkapjuk, hogy:

$$\mathbb{P}(X > z) = \mathbb{P}(X > z \mid X > u)\mathbb{P}(X > u) \quad (3.65)$$

ahol a feltételes valószínűség az pont $1 - F_u(x)$, ahol $x = z - u$, mert eredetileg az F_u függvény az $X - u$ feltételes eloszlását adja meg. A $\mathbb{P}(X > u)$ általában historikus adatok alapján becsülik meg, de az én esetemben ez a megközelítés nem alkalmazható, mivel nem veszi figyelembe az időt, télen ugyan akkora valószínűséget fog becsülni egy extrém kifizetésre, mint nyáron, ami persze nem igaz. Ehelyett

a küszöbtúllépési valószínűséget azzal fogom becsülni, hogy a vizsgált időszakban mekkora eséllyel fog bekövetkezni egy hőhullám. Ez azt jelenti, hogy múltbeli adatok alapján megvizsgálom, hogy az egyes extrém kifizetéseknél milyen hőhullám volt az azt megelőző időszakban (pl. egy hónapra visszamenőleg), és az alapján meghatározom, hogy milyen hőhullám paraméterek esetén (azaz milyen hosszú és milyen meleg) haladja meg a kifizetés a küszöbértéket. Ennek segítségével szolgál a fentebb említett szimulációs előrejelzés.

4. fejezet

Modellezési eredmények

4.1. Adatok bemutatása

A kutatásomban kétféle adatot használtam, hőmérsékleti, illetve mezőgazdasági aszály-kár-kifizetési adatokat. Ezek mind idősoros adatok voltak, a hőmérsékletről 2002 januártól 2024 augusztusig, a kárkifizetésekről pedig 2013 júliusától 2020 novemberéig. A kárkifizetési adatok egy nagyobb biztosítótól származnak, így adatvédelmi okokból nem forintban fogom értelmezni az eredményeket, hanem egy általam definiált fiktív pénznemben (peták). Fontos megemlíteni, hogy az adatok nem tartalmazzák az aszálykárokat országos szinten, ezért nem is lehet ezt az elemzést kiterjeszteni egész Magyarországra, viszont az adatot szolgáltató intézmény nagy piaci részesedéssel rendelkezik, így reprezentatív lehet a kutatásban. A bevezetésben említett térbeliség bevonása miatt 3 helyről gyűjtöttem ki a hőmérsékleti adatokat:

Z_1 : Budapest napi maximum hőmérséklete

Z_2 : Nagykanizsa napi maximum hőmérséklete

Z_3 : Debrecen napi maximum hőmérséklete

Az elemzéshez a 3.1.1 fejezetben tárgyalt $\tilde{Z} = \min(Z_1, Z_2, Z_3)$ fogom használni. Továbbá az adatokat a European Climate Assessment & Dataset (ECAD) oldaláról töltöttem le. A kárkifizetési adatok pedig havi, illetve heti bontásban vannak, mindkét esetben meg fogom vizsgálni, hogy milyen eloszlás illeszkedik az adatokra.

4.2. Hőmérséklet modellezése és előrejelzése

4.2.1. Determinisztikus modellezés

A modellem célja az lenne, hogy előre tudjam jelezni a 2025 május - szeptemberi időszakra a hőhullámokat, vagy legalább is valószínűséget adni az egyes hőhullámok bekövetkezésére. Ehhez elsősorban vizsgáljuk meg, hogy a múltban milyen hosszú, illetve meleg események következtek be.

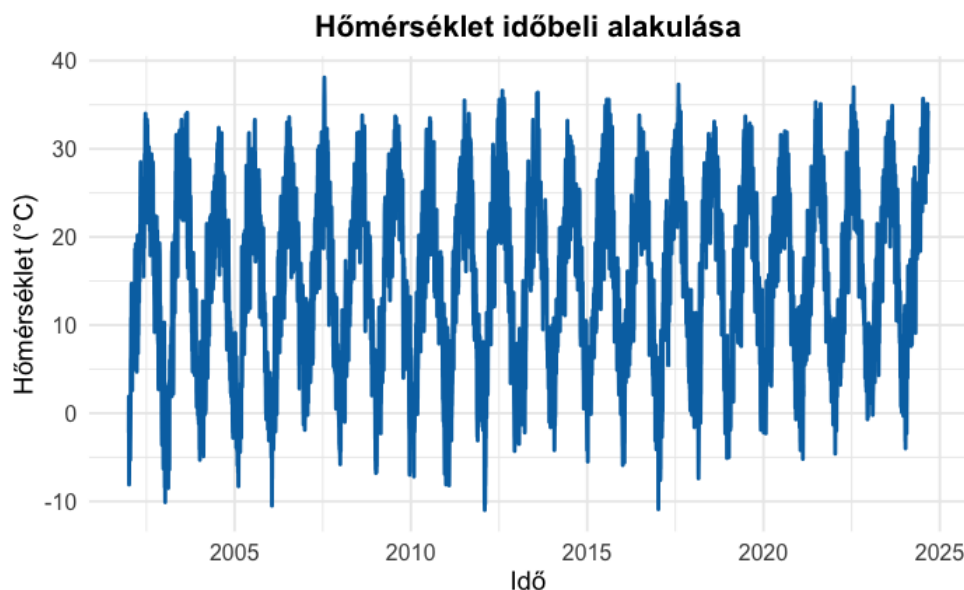
4.1. táblázat. Hőhullámok száma az egész időszakra tekintve

Hossz \ Küszöbérték	28 °C	30 °C	32 °C	34 °C	36 °C
3	95	47	18	4	1
4	66	31	11	1	1
5	52	23	7	1	0
6	41	16	5	0	0
7	32	10	5	0	0
8	22	7	3	0	0
9	17	6	3	0	0
10	13	4	2	0	0
11	9	3	2	0	0

A 4.1 táblázat különböző hossz és küszöbértékekre mutatja meg, hogy a múltban

adott paraméterek mellett hányszor fordult elő az az esemény. Pl.: Összesen ötször fordult elő olyan, hogy legalább 6 egymást követő napon volt 32 fok vagy annál több. A küszöbértékeket a teljes idősor 90. percentilisétől vettem 2 fokonként növelve.

A modellezéshez először tekintsük meg a hőmérséklet időbeli alakulását:



4.1. Ábra. A hőmérséklet időbeli alakulása

Tisztán látszódnak a szezonális hatások a modellben, a nyári hónapokban magasabb téli hónapokban pedig alacsonyabb, illetve egy nagyon enyhe pozitív trend is megfigyelhető az adatokban. Utóbbit egy egyszerű lineáris regresszióval fogom kiszűrni, a szezonalitást pedig Fourier együtthatós módszerrel szűröm ki, a (3.7) egyenlet alapján. A lineáris regresszió esetében az együttható szignifikánsan eltért nullától, így valóban megfigyelhetünk egy növekvő trendet az idősorban, évente átlagosan 0.125 fokot emelkedik a hőmérséklet. A szezonális szűrése esetén pedig $N = 3$ -at választottam (Campbell és Diebold 2005). α_n jelöli a koszinusz tagok súlyait, illetve γ_n jelöli a szinusz tagok súlyait:

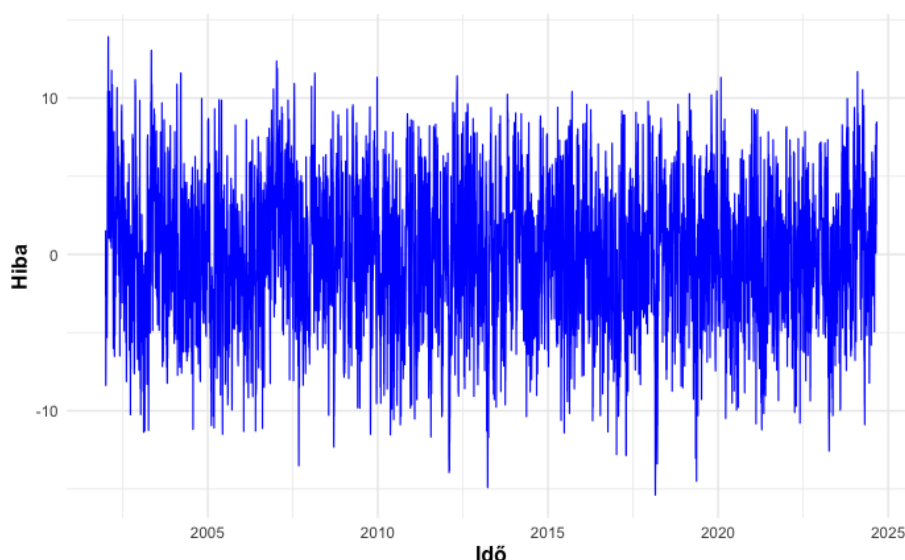
4.2. táblázat. Fourier együtthatók

α_1	-12.29^{***}
α_2	-1.13^{***}
α_3	-0.08
γ_1	-3.47^{***}
γ_2	0.24^{***}
γ_3	-0.48^{***}
t	0.00034^{***}

Megjegyzés: A * a 10%-on , a ** az 5%-on, míg a *** az 1%-on szignifikáns együtthatókat mutatja. A t együtthatója az átlagos napi hőmérséklet változást mutatja cp.

A determinisztikus részt $S_t = T_t + G_t + u_t$ egyenlettel írtuk le, ahol T_t az idősor lineáris trendje és G_t a Fouries sor, és u_t egy véletlentől függő hibatag. Mivel additív modellt feltételezek, így az eredeti idősorból kivonva a trend és szezonális hatást, megkapjuk ezt a 4.2 ábrán látható hibatagot. Az így adódó idősorra elvégeztem egy stacionaritás vizsgálatot. ADF (Augmented Dickey-Fuller) tesztet és KPSS (Kwiatkowski-Phillips-Schmidt-Schin) tesztet használtam ehhez. Az ADF teszt lényegében az idősor megváltozására felírt egyenlet egyik együtthatójának nullától való eltérését teszteli: $\Delta Y_t = \alpha + \delta Y_{t-1} + \sum_{i=1}^p \gamma_i \Delta Y_{t-i} + u_t$ egyenletben a $\delta = 0$ voltát teszteli. Ha az együttható 0, abban az esetben van egységgyök az idősorban, így nem lesz stacionárius (ez az állítás a teszt nullhipotézise). A KPSS teszt esetében az idősorra az $Y_t = \mu + \delta t + \sum_{i=1}^t \eta_i + u_t$, ahol μ konstans, δt írja le az idősor trendjét, u_t a stacionárius komponens és η_i -k írják le a sztochasztikus trendet (η_i -k független azonos eloszlású változók 0 várható értékkel és σ_η^2 varianciával). A teszt H_0 -ja azt mondja ki, hogy $\sigma_\eta^2 = 0$, azaz trend mentén stacionárius. Mindkét esetben azt mutatta a

hipotézisvizsgálat, hogy stacionáriusnak tekinthető az idősor (az ADF teszt p-értéke 10^{-6} nagyságrendű) még 1%-os szignifikancia szinten is. Ezzel ellentétben a KPSS teszt H_0 -ja az, hogy stacionárius és ez esetben is stacionáriusnak tekinthető az idősor 5%-os szignifikancia szinten (a KPSS teszt p-értéke 31.28%).

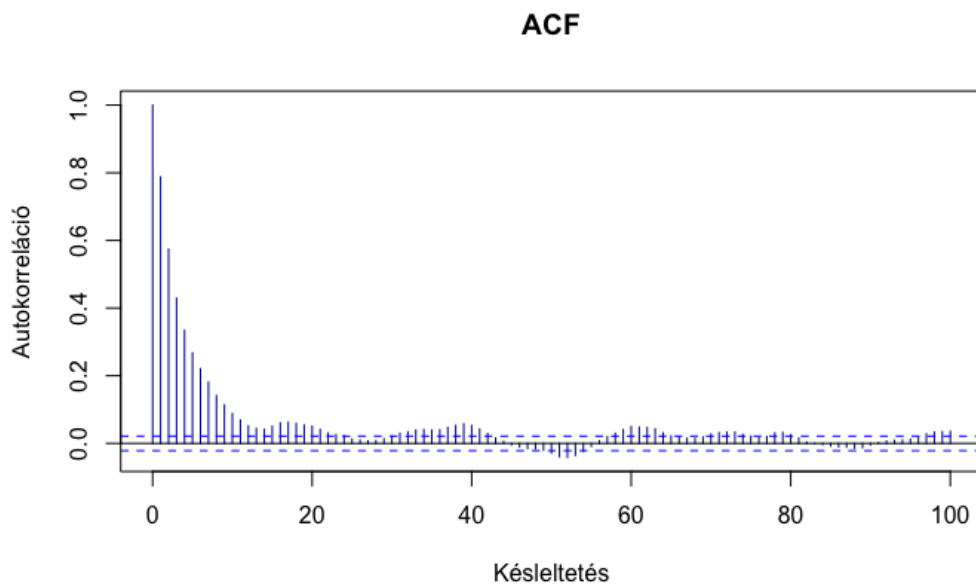


4.2. Ábra. Determinisztikus modell hibatagja

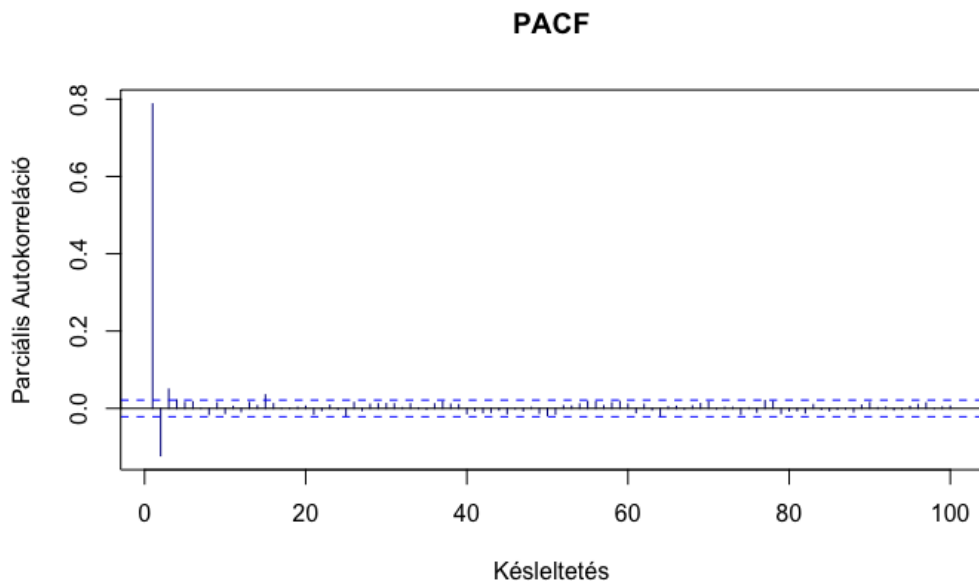
4.2.2. AR(p) modell megválasztása

A stacionaritás vizsgálata után egy fehérzaj tesztet is elvégeztem, ehhez BG (Breusch-Godfrey) tesztet használtam, aminek a nullhipotézise az, hogy fehérzaj az idősor, tehát nincsen benne autokorreláció. A tesztet 200 késleltetésig néztem, és a p-értéke 10^{-16} nagyságrendű lett, így az állítható, hogy van szignifikáns autokorreláció a folyamatban 5%-os szignifikancia szinten. A módszertani részben részletesebben tárgyalt ARMA folyamattal fogom modellezni a hibatagban megmaradó autokorrelációt, ehhez azonban nézzük meg először egy korrelogramon, hogy hogyan viszonyul az idősor a késleltetett értékekhez. A 4.3 és 4.4 ábrákból tisztán látszik, hogy az ACF szépen

lassan lecseng a 0-ba, a PACF esetében pedig az első és második késleltetés ki-metsz a 95%-os konfidenciaintervallumból (tehát szignifikánsan eltér 0-tól 95%-os megbízhatósággal) és utána a többi érték már mind 0-nak tekinthető. Ez pont egy AR hatásra utal az idősorban, ahogyan *Karl Larsson* (2023)-ban is láthatjuk. Az ábrák alapján elsősorban egy AR(3) modellt illesztenék, de annak érdekében, hogy biztos ne maradjon benne semmi autokorreláció, megvizsgáltam az AR(1), AR(2), AR(3), illetve AR(4) modelleket, majd az együtthatók empirikus szignifikancia-szintje és az információs kritériumok alapján döntöttem. Magasabb rendű modellt már nem tartottam célszerűnek megvizsgálni, mivel úgy komplexebb lesz a modell és könnyen beleeshetünk a túlillesztés hibájába.



4.3. Ábra. Napi hőmérséklet autokorrelációja (ACF)



4.4. Ábra. Napi hőmérséklet parciális autokorrelációja (PACF)

Az Akaike és Bayes-i információs kritériumok egy olyan mérőszámot adnak, ami figyelembe veszi a modell illeszkedését és a komplexitását. Ezen mutatók segítségével könnyen tudunk modellszelekciót elvégezni.

$$AIC = 2k - 2 \ln(\hat{L}) \quad (4.1)$$

$$BIC = k \ln(l) - 2 \ln(\hat{L}) \quad (4.2)$$

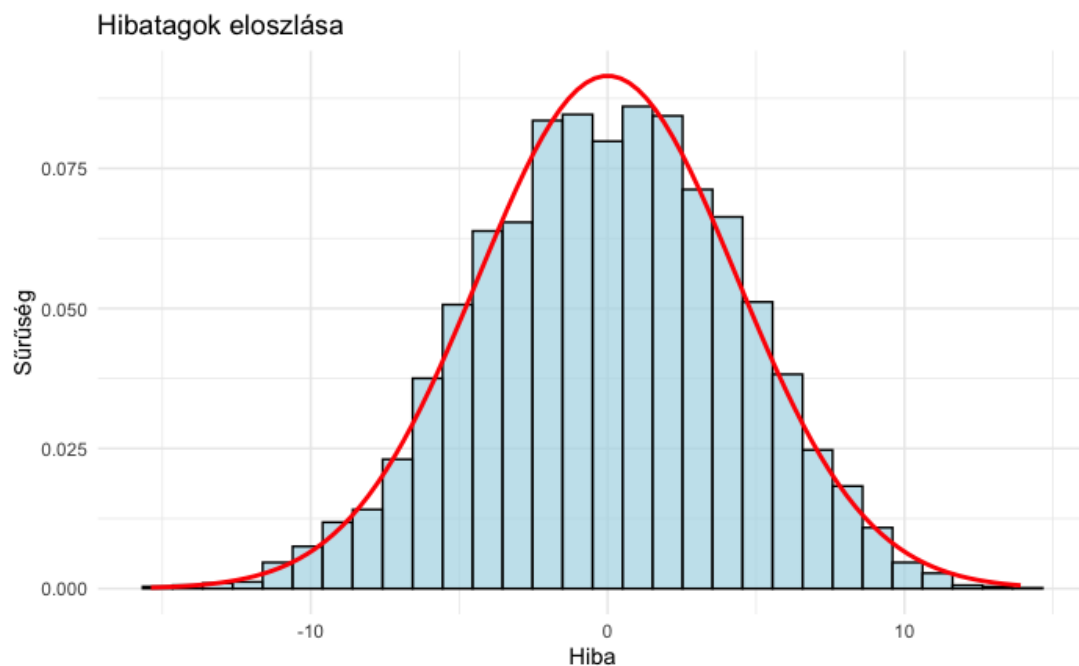
ahol k a becsült paraméterek száma és $\hat{L}$ a likelihood függvény maximuma. Akkor lesz preferáltabb egy modell ha ennek a mutatónak az értéke kisebb. Az egyenletekből látszik, hogy a BIC jobban bünteti a felesleges magyarázóváltozókat.

4.3. táblázat. AR(p) modellek együtthatói és AIC/BIC értékei

Együtthatók	AR(1)	AR(2)	AR(3)	AR(4)
λ_1	0.788***	0.884***	0.891***	0.889***
λ_2	-	-0.122***	-0.167***	-0.163***
λ_3	-	-	0.05***	0.03**
λ_4	-	-	-	0.0002
AIC	39849	39726	39707	39705
BIC	39870	39754	39742	39747

Megjegyzés: Az * a 10%-on, a ** az 5%-on, míg a *** az 1%-on szignifikáns együtthatókat mutatja, illetve a tengelymetszetek egyik esetben sem voltak szignifikánsak.

A BIC az AR(3) modellt preferálja, míg az AIC az AR(4) nagyon kicsivel jobban mint AR(3)-at, viszont feleslegesnek tartom belevonni a 4. késleltetett, mivel amúgy sem tér el szignifikánsan 0-tól a 4.4 ábra alapján. A továbbiakban az AR(3) modellel fogok dolgozni, és az így keletkező hibatagban vélhetően már nem lesz autokorreláció. Utóbbi állítást az újra elvégzett BG-teszt is alátámasztja, a H_0 -t el tudtuk fogadni 95%-os p-érték mellett. A modellben erre a hibatagra normális eloszlást feltételeztem, így annak a tesztelése sem elhanyagolható, erre a Kolmogorov-Szmirnov próbát használtam, aminek a nullhipotézise az, hogy normálisnak tekinthető a hibatag. A p-érték alapján (0.055) épp hogy el tudjuk fogadni a H_0 -t 5%-os szignifikancia szinten.



4.5. Ábra. Hibatagok eloszlása

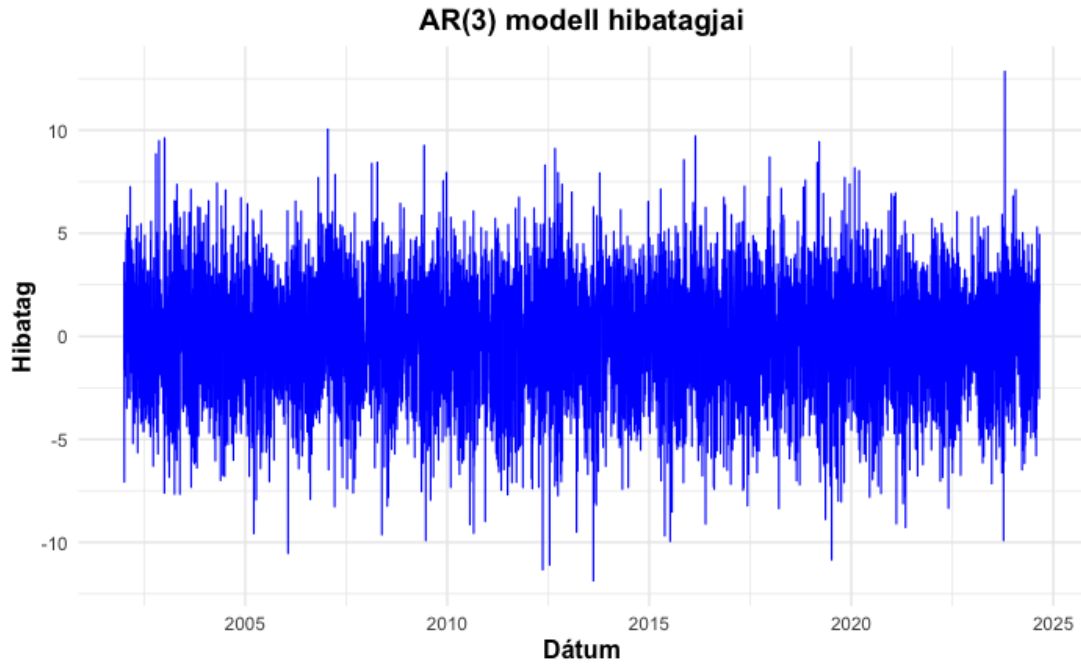
A piros vonal mutatja a normális eloszlás sűrűségfüggvényét, látszik, hogy 0-nál talán nem követi le annyira az elméleti eloszlást, de mindent összevetve a normális eloszlás nem egy rossz közelítés.

4.2.3. ARCH modell

Az eredeti idősorban kezeltük a trendet, szezonális hatást, illetve autokorrelációt, viszont a szórás esetleges időfüggését még nem. Emlékeztetésképpen, $u_t = \sigma_t \epsilon_t$, $\epsilon_t \sim N(0, 1)$, és a σ_t^2 -re írtunk fel egy egyenletet:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2 \quad \alpha_i > 0 \quad (4.3)$$

Tehát az idősorban nem állandó kilengéseket szeretnénk kezelni ezzel az egyenlettel.



4.6. Ábra. AR(3)modell hibatagja

Nem egyértelmű az ARCH hatás (azaz időfüggő szórás) jelenléte, viszont ha a hibatagok négyzetére vizsgáljuk meg a PACF értékeket, akkor vannak késleltetések, amik szignifikánsan eltérnek 0-tól. Ez alátámasztja azt, hogy van ARCH hatás az idősorban. Konkrét hipotézisvizsgálattal alátámasztva, Engle-féle ARCH tesztet használtam, ami felépít egy segédregressziót a hibatagok négyzetére:

$$u_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2 \quad (4.4)$$

A teszt nullhipotézise az, hogy α_i -k együttesen 0-nak tekinthetők, azaz nincs benne ARCH hatás. A tesztstatisztika értéke $N \cdot R^2 \sim \chi_p^2$ mellett, ahol N az idősor hossza és R^2 a segédregresszió magyarázóereje. A teszt p-értéke 0.28% lett, ami azt jelenti, hogy el kell vetnünk a H_0 -t 5%-os szignifikancia szinten, van jelen ARCH hatás. Ez a teszt a GARCH hatás voltáról nem állít semmit, így az előző részhez hasonlóan

fogom kiválasztani a legjobban illeszkedő modellt: az együtthatók szignifikanciája és információs kritériumok alapján.

4.4. táblázat. ARCH(p)/GARCH(p,q) modellek együtthatói és AIC/BIC értékei

Együtthatók	α_1	α_2	α_3	β_1	β_2	β_3	AIC	BIC
ARCH(1)	0.061***	-	-	-	-	-	4.792	4.795
ARCH(2)	0.061***	0.017	-	-	-	-	4.792	4.796
ARCH(3)	0.061***	0.014	0.042***	-	-	-	4.791	4.795
GARCH(1,1)	0.031***	-	-	0.907***	-	-	4.788	4.791
GARCH(1,2)	0.048***	-	-	0.262***	0.603***	-	4.787	4.791
GARCH(2,2)	0.048***	10^{-8}	-	0.262	0.603***	-	4.788	4.793
GARCH(3,1)	0.031***	10^{-10}	10^{-8}	0.907***	-	-	4.788	4.794
GARCH(1,3)	0.057***	-	-	0.082	0.428*	0.336	4.787	4.792

Megjegyzés: Az * a 10%-on, a ** az 5%-on, míg a *** az 1%-on szignifikáns együtthatókat mutatja, illetve a tengelymetszetek egyik esetben sem voltak szignifikánsak.

A modellek közül számomra a GARCH(1, 2) bizonyult a legjobbnak, így egy kombinált modellel (AR(3) + GARCH(1, 2)) fogok tudni előrejelezni.

4.2.4. Előrejelzés

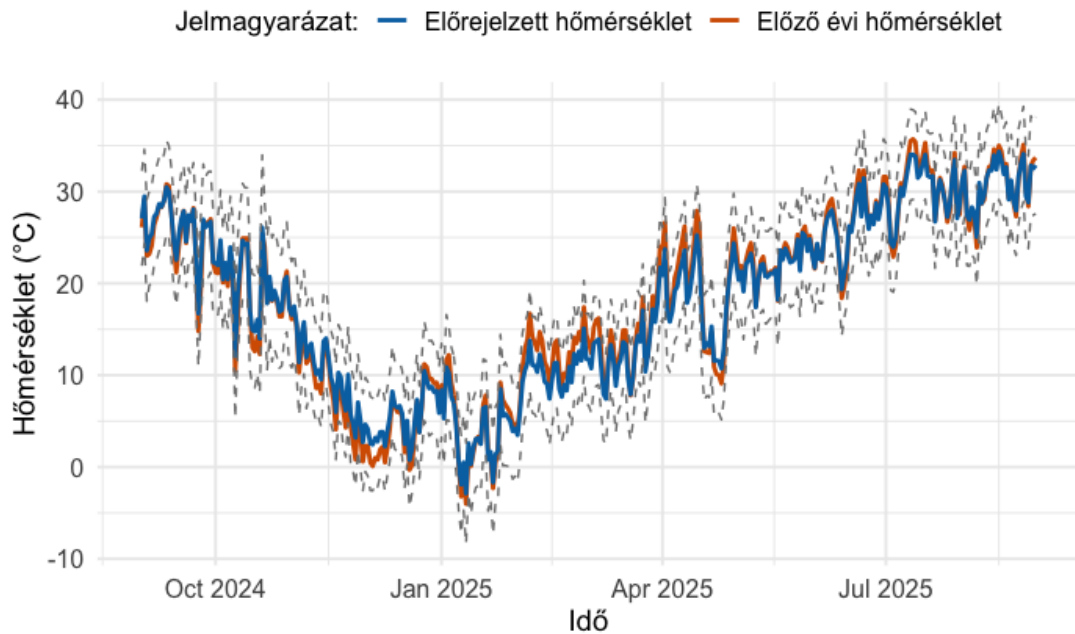
365 napos dinamikus előrejelzést végeztem (2024 szeptembertől 2025 szeptemberig lévő időszakra), ami azt jelenti, hogy mindig csak 1 napra jelzek előre a modellel, majd a becsült értéket figyelembe véve újraillesztem a modellt és úgy számolom ki a következő napi hőmérsékletet. A sztochasztikus rész előrejelzéséhez a várható értéket fogom venni, majd ehhez hozzáadva a trendet és a szezonális hatást kapom meg a becsült hőmérsékletet. A szórás becsléséhez az GARCH modellt használtam, az

alábbi ábra mutatja, hogy várhatóan melyik időszakokban lesz nagyobb a kilengés a következő 1 évben.



4.7. Ábra. Szórás előrejelzése GARCH modellből

Reálisnak tartom az eredményt, a legnagyobb szórást az október-novemberi időszakra mutatja, amikor kezd nagyon lehűlni az idő, illetve tavasszal április környékére is becsül a modell egy nagyobb szórást. Az elmúlt évek tapasztalatai alapján volt, hogy hó is esett áprilisban, így abszolút indokoltnak tartom a nagyobb szórást abban az időszakban. A napi maximumok előrejelzése igen csak jól illeszkedett az előző évihez:



4.8. Ábra. Hőmérséklet előrejelzése GARCH modellből

A 4.8 ábra alapján eléggé egybeesik az előző éves hőmérséklettel. Az AR modell hosszútávon bekonvergál a várható értékhez, viszont a dinamikus előrejelzési technikával ez kiküszöbölhető. Attól függetlenül, hogy ilyen jól illeszkedett az előrejelzés az előző évihez, fontosnak tartom hogy validáljam a modell jóságát, így számoltam egy 99%-os VaR (Value at Risk) értéket (az az érték ami fölé az adatok 99%-a esik) a modellből és megnéztem, hogy a valós adatok hány százaléka esik ezen érték fölé. A várható érték előrejelzésénél eddig csak az AR(3) modellt használtam, viszont számolt VaR értékhez már a GARCH modellel megbecsült szórást is belevontam. A modell konfidenciaintervallumát a valós adatok 1.6%-a hagyta el (az 1% helyett). Emellé még elvégeztem egy Kupiec tesztet, ami azt teszteli, hogy a modell által rosszul becsült megfigyelések (amik a konfidenciaintervallumon kívül estek) elfogadhatónak tekinthetőek-e. A teszt p-értéke 0.22 lett, ami azt jelenti, hogy elfogadjuk a H_0 -t,

hibahatáron belülnek tekinthető ez az eltérés.

4.2.5. Hőhullámok bekövetkezési valószínűsége

Az eredeti célom az volt, hogy megadjam egy valamilyen paraméterekkel rendelkező hőhullám bekövetkezési valószínűségét. Egy ilyen sztochasztikus modell esetében elegendő szimulációval közelíthető az elméleti valószínűség. A modellben a 3.15 egyenletben az ϵ_t tag az egyetlen olyan ami a véletlentől függ. Erre standard normális eloszlást feltételeztem, így az előrejelzésben csak ezeket a tagokat szimulálom le, majd rászorozva a GARCH modellből becsült szórással, illetve hozzáadva az AR(3), trend és szezonális komponenseket megkapjuk az előrejelzést. 10000 szimulációt végeztem el 365 napra, így ezekből már könnyen ki lehet számolni, hogy hány esetben következett be az adott hőhullám. Mérési időszaknak a május-szeptemberi intervallumot vettem. A modellt természetesen tovább lehet bonyolítani, úgy is hogy külön vesszük a hónapokat és mindegyik lesz egy mérési időszak, így külön valószínűségeket kapva az egyes hónapokra, de engem most kifejezetten a melegebb hónapok érdekelnek, így nem szedtem azokat külön. Az egyes valószínűségek azt mutatják, hogy mekkora eséllyel fog bekövetkezni egy olyan esemény május és szeptember között, ahol legalább n egymást követő napon a küszöbértéket meghaladja a hőmérséklet, azaz a 3. fejezetben használt jelölésekkel: $\mathbb{P}(\Gamma(n, a, \tau, Z) \geq 1 \mid \mathcal{F}_t)$. A 4.1 táblázat alapján 34°C-os küszöbértéknél nem volt megfigyelhető hőhullám, ami legalább 6 napig tartott volna, viszont a szimuláció alapján $\mathbb{P}(\Gamma(7, 34, \tau, Z) \geq 1 \mid \mathcal{F}_t) = 10.57\%$, ami egy viszonylag magas érték, ahhoz képest, hogy az utóbbi 22 évben nem volt ilyen megfigyelhető esemény. Összességében ezzel a modellel tudunk egy becslést adni, hogy mekkora valószínűséggel következik be egy hőhullám, ami gyakorlatilag azt jelenti, hogy mekkora eséllyel fog kár bekövetkezni. Ezen kívül már csak a kár nagyságára kell adnunk egy becslést.

4.5. táblázat. Hőhullámok bekövetkezési valószínűsége a 2025 május-szeptemberi időszakra nézve

n \ a	28°C	30°C	32°C	34°C	36°C
3	99.95%	98.17%	84.56 %	54.99%	26.95%
5	99.12%	88.02%	56.09%	25.41%	8.12%
7	94.21%	69.03%	32.2%	10.57%	2.58%
9	83.5%	48.52%	16.73 %	4.19%	0.87%
11	69.98%	32.37%	9.25%	1.81%	0.25%

Megjegyzés: Az n az egymást követő napok számát, míg a a küszöbértéket jelzi.

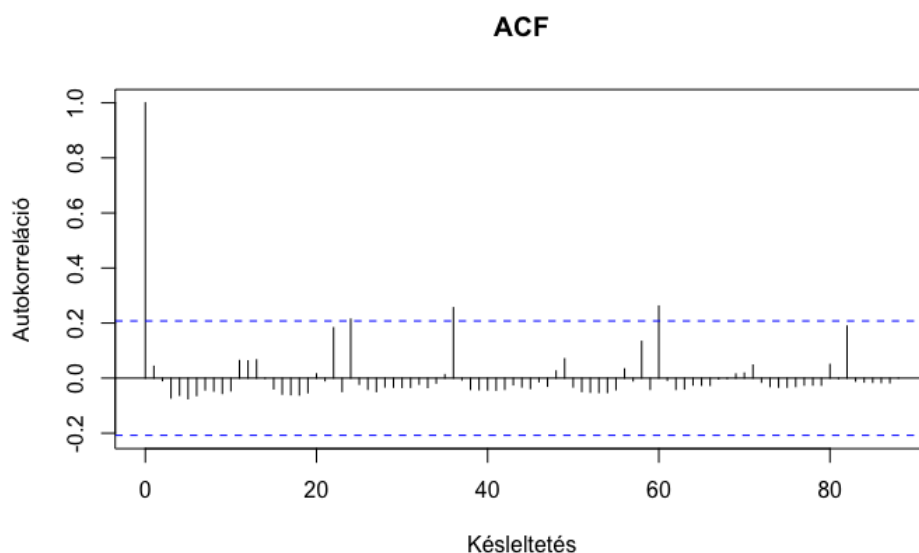
4.3. Extrémérték-modellek

Az elemzésem célja, hogy extrémérték-modellek segítségével írjam le a hőmérsékleti adatokhoz kapcsolódó szélsőséges eseményeket. Két, a szakirodalomban széles körben alkalmazott megközelítést vizsgálok meg: a küszöb feletti meghaladások módszerét, valamint a blokkmaximumok módszerét. Mindkét eljárás célja a ritka, de jelentős mértékű kilengések statisztikai modellezése, amelyek különösen fontosak a kockázatkezelés és előrejelzés szempontjából.

4.3.1. Küszöb feletti meghaladások

Az aszálykár-kifizetési adatokat havi bontásban használtam fel a modellemhez, elsősorban megvizsgáltam, hogy stacionáriusnak tekinthető-e az idősor, és mind az ADF teszt és a KPSS teszt is a stacionaritás mellett volt (5%-os szignifikancia szinten).

Ezen kívül az idősor autokorrelációját is megvizsgáltam, és két helyen talán épp hogy kimetsz a nullára épített 95%-os konfidenciaintervallumból, de a legtöbb esetben 0-nak mondható. A BG tesztet itt is elvégeztem, a p-érték 0.96 lett, így 5%-os szignifikancia szinten el tudom fogadni a H_0 -t, autokorrelálatlannak tekinthető az idősor, ami alátámasztja az előző állításomat. Számomra ez azért fontos, mert függetlenséget feltételezek a modellben, persze a korrelálatlanságból nem következik a függetlenség, de ha már a gyengébbik feltétel sem teljesülne, úgy nem lenne alkalmas az idősor az extrémérték modellhez.



4.9. Ábra. Aszálykár-kifizetések autokorrelációja

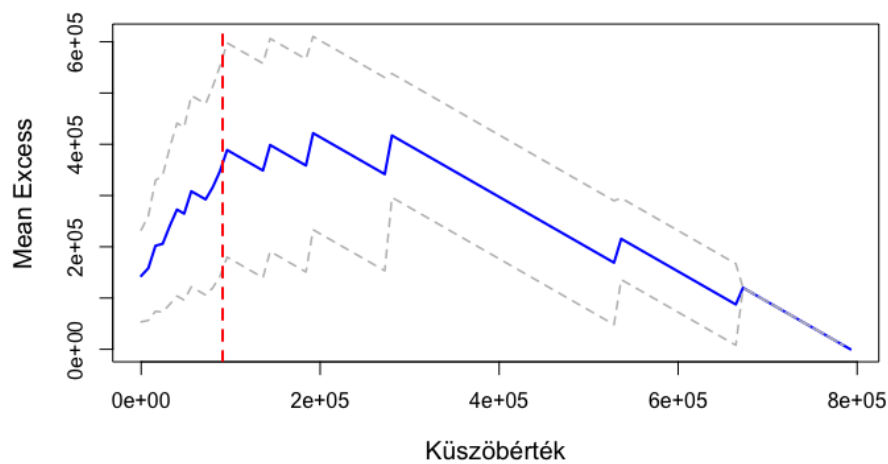
Különböző küszöbértékekre fogok illeszteni eloszlást, majd a becsült paraméterek alapján ki fog derülni, hogy valójában melyik extrém eloszlás illeszkedik az idősorra. Az első esetben vizuálisan határoztam meg a küszöbértéket, azt felhasználva, hogy a meghaladások várható értéke lineárisak u-ban. 91000-nél húztam meg a határt, bár nem látszik egyértelműen a 4.10-es ábrán a kevés megfigyelés miatt. Ebben az esetben 9 extrém megfigyelés van. A többi esetben a 3. fejezetben tárgyalt képletekkel

dolgoztam és az alábbi ábra mutatja az alkalmas küszöbértékeket:

4.6. táblázat. Küszöbértékek különböző módszerek esetén

	Feltétel	Küszöbérték	Extrém megfigyelések száma
1.	Ábra alapján	91000	8
2.	90. percentilis	77477	9
3.	$k = \sqrt{n}$	76000	10
4.	$k = \frac{n^{\frac{2}{3}}}{\log(\log(n))}$	33670	14

Itt k az extrém megfigyelések számát, n pedig az összes megfigyelt értékek száma.



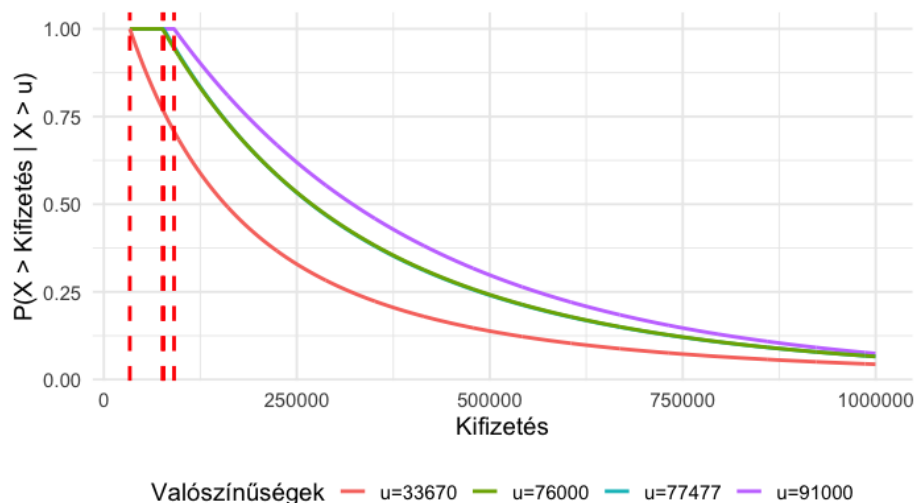
4.10. Ábra. Átlagos küszöb meghaladások

A kevés megfigyelés miatt a már korábban tárgyalt L-momentumok módszerével becsültem meg a paramétereket:

4.7. táblázat. Becsült paraméterek

	u	σ	ξ	Határeloszlás
1.	91000	328065.6	0.04813	Pareto
2.	77477	257715.7	0.1922	Pareto
3.	76000	261593.83	0.18383	Pareto
4.	33670	154528.3	0.400332	Pareto

A küszöb feletti meghaladások módszerével több különböző küszöbértéken is illesztettem extrémérték-eloszlásokat az adatokra. Az eredmények alapján minden vizsgált küszöb esetén Pareto eloszlás volt a határeloszlás. Ez azt mutatja, hogy a modell robusztus a küszöbérték megválasztásával szemben, vagyis nem érzékeny a pontos küszöbszintre. A Pareto eloszlás illeszkedése arra utal, hogy az extrém kárkifizetések eloszlása vastag farkú, vagyis a kiugróan nagy kifizetések nem gyorsan csengenek le, hanem a nagy károk bekövetkezése is reálisan számottevő valószínűséggel bír, ami különösen fontos szempont a biztosítói kockázatkezelés és tőkekövetelmény számítás szempontjából. Egy hátránya az L-momentum módszernek, hogy nem valószínűségi alapú, így konfidenciaintervallumot nem lehet meghatározni belőle úgy, mint a maximum-likelihood módszernél. A modellek összehasonlítása során a farokeloszlás függvények viselkedését vizsgáltam az extrém értékek tartományában. A négy modell közül három nagyon hasonló eredményt adott, míg a negyedik esetében, ahol jelentősen alacsonyabb küszöbértéket használtam, a valószínűségek végig kisebbek voltak a többinél. Fontos megjegyezni, hogy mind a négy esetben véges várható értéke és szórása lesz az eloszlásnak, mivel a ξ paraméterek minden esetben kisebbek $\frac{1}{2}$ -nél.



4.11. Ábra. Különböző modellek farokeloszlása

Megjegyzés: A piros szaggatott vonalak az egyes modellek küszöbértékeit mutatják.

4.3.2. Blokkmaximumok

Az elemzés során két blokkmaximumos modellt is készítettem, az egyik esetben havi, a másikban éves blokkmaximumokat alkalmazva. Havi blokkmaximumok alatt az egyes hónapokban a legnagyobb kifizetést értem (heti szinten aggregált adatok esetén), ugyan ilyen logika alapján értendő az éves blokkmaximum (itt havi szintű aggregáció esetén nézzük a legnagyobb kifizetést az egyes években). Ez a módszer a változók maximumára ad egy határeloszlást, ebben különbözik a küszöb meghaladások módszerétől. A becsült paraméterek a következők lettek:

4.8. táblázat. Becsült paraméterek

	Blokkmaximum	μ	σ	ξ	Határeloszlás
1.	Éves	199006	211429	-0.4933	Weibull
2.	Havi	1367.2	5977.55	0.8144	Fréchet

Mindkét modellt az L-momentumok módszerével becsültem, mivel a rendelkezésre álló havi adatok száma is mindössze 89 megfigyelésből állt, így a maximum likelihood becslés helyett ezt a megközelítést indokoltabbnak láttam (*Ulrych, 2000*). A két különböző időbeli aggregáció eredményeként azonban eltérő határeloszlások illeszkedtek a maximumra: a havi blokkok esetében a Fréchet eloszláshoz, míg az éves blokkok esetében a Weibull eloszláshoz tart. Ez az eredmény azt mutatja, hogy a választott blokkstruktúra jelentős hatással van a határeloszlás típusára, vagyis a modellérzékenység szempontjából nem elhanyagolható, hogy milyen időbeli aggregációval dolgozunk. Az eloszlások azt mutatják, hogy havi szinten aggregálva az adatokat egy vastag farkú eloszláshoz tart, tehát havi szinten jellemzőbbek a kiugróan nagy értékek, míg éves szinten inkább egy gyorsan lecsengő eloszlásról van szó, azaz az éves károkra kevésbé jellemzőek az extrém nagy aszálykárok. Mivel az alkalmazott modell végeredménye ilyen mértékben függ a blokkmaximumok megválasztásától, ezt módszert ebben az esetben nem tartom megbízhatónak. Az ilyen érzékenység különösen problémás lehet a gyakorlatban, hiszen eltérő időablakok eltérő kockázati következtetésekhez vezethetnek, ami ronthatja az eredmények robusztusságát és gyakorlati hasznosíthatóságát.

4.3.3. Kárkifizetés becslése a hőmérséklet figyelembevételével

A végső számolás során a 77477-es küszöb mellett döntöttem, mivel az csak minimálisan tér el a harmadik modell által javasolt értéktől. Az első modell által

használt küszöböt ugyan szemrevételezéssel választottam ki, azonban úgy gondolom, hogy nagyon nehéz meghatározni csupán egy ábra alapján a pontos küszöböt, így szerintem nem tekinthető teljesen megbízhatónak az első modell. A negyedik modell pedig lényegesen eltér az összes többitől, főként azért, mert jóval alacsonyabb küszöbértéket alkalmaz. Mindezek alapján a második ($u=77477$) modell eredményeit tartom a legmegbízhatóbbnak, mivel ez kellően közel áll a szintén statisztikai módszeren alapuló harmadik modellhez, ugyanakkor a kiválasztott küszöbérték mögött gyakorlati érvek is vannak (*Max Rydan*, 2018). Az így adódó végleges eloszlás ($u=77477$ mellett illesztett extrémérték eloszlás) illeszkedését Kolmogorov-Szmirnov próbával ellenőriztem. Létrehoztam a megfelelő paraméterekkel az eloszlásfüggvényt, majd a teszt segítségével megvizsgáltam, hogy az adatok illeszkednek-e rá. A teszt p-értéke 0.6656 lett, így 5%-os szignifikancia szinten elfogadjuk a nullhipotézist, tehát az adatok illeszkednek a becsült eloszlásra. A kárkifizetések becsléséhez azt a megközelítést alkalmaztam, hogy kiválasztottam azokat a hónapokat, amelyekben extrém kárkifizetés történt (azaz ahol a kifizetés mértéke meghaladta a küszöbértéket), majd megnéztem, hogy ezekben a hónapokban fordult-e elő valamilyen hőhullám (esetleg az előző hónapot is), így létrehozva a kárkifizetések és a hőmérséklet közti kapcsolatot a modellben. A 4.9 táblázatban összeszedtem azokat a hónapokat amikor extrém kifizetés volt (azaz a kifizetés meghaladta a 77477 petákat), illetve a hozzájuk tartozó kifizetést, valamint azt, hogy volt-e valamilyen hőhullám abban az időszakban, és ha volt akkor milyen paraméterekkel rendelkezett. 4 olyan hónap volt, amihez nem tartozott semmilyen hőhullám, de még is magas volt a kifizetés, ezek viszont magyarázhatóak azzal, hogy az előző hónapban lévő meleg volt hatással rá és elképzelhető, hogy a kárbejelentés és a kifizetés között eltelt egy kis idő. Egyedül a 2020 árilisában lévő extrém kárra nem mondható ez, arra sajnos nem találtam konkrét magyarázatot, hogy az mitől is következhetett be.

4.9. táblázat. Extrém kifizetésű hónapok

	Hónap	Kifizetés	Hőhullám
1.	2013-08	792420.92	n=5, a=32
2.	2014-07	280000	n=7, a=28
3.	2014-08	137801.8	-
4.	2015-08	670744.9	n=11, a=32
5.	2015-09	83389.84	-
6.	2018-08	792546.2	n=4, a=32 és n=7, a=30
7.	2018-09	91140	-
8.	2020-04	186865.54	-
9.	2020-06	533715.76	n=3, a=30

Sutanto és tsai. (2024)-ben cikkében kitértek arra, hogy az aszálykárok esetében a hosszabb ideig tartó, de kevésbé meleg hőhullámok nagyobb kárt okozhatnak, mint a rövidebb, de extrém forró hőhullámok. Erre hivatkozva szeretném megválasztani azt a hőhullámot, ami esetén extrém kárt várunk (tehát a $P(X > 77477)$). 2014 júliusában volt 280 ezer petáknyi kifizetés, ez volt a legalacsonyabb kifizetésű hónap, ahol detektáltam hőhullámot, így feltehető hogy már egy 7 napig tartó, legalább 28°C-os hőhullám esetén már bekövetkezik egy extrém kár. Ennek a valószínűsége a 4.5 táblázatból kiolvasható: 94.21% eséllyel következik be egy ilyen hőhullám a 2025 május-szeptemberi időszakban. A korábbi jelöléseket használva:

$$\mathbb{P}(\Gamma(n, a, \tau, Z) \geq 1 \mid \mathcal{F}_t) = 94.21\% \quad (4.5)$$

Mind ezeket használva egy példán szeretném megmutatni, hogyan számíthatjuk annak a valószínűségét, hogy pl. lesz egy 200000-nél nagyobb kifizetés a jelenlegi mérési

időszakban. A 3.45 eloszlásfüggvénybe behelyettesítve a megfelelő paramétereket, majd ezt kivonva 1-ből kapjuk meg a $\mathbb{P}(X > 200000 \mid X > 77477)$ valószínűséget, majd ezt megszorozva a 4.5 egyenletben leírt valószínűséggel kapjuk meg annak az esélyét, hogy 200000 petáknál magasabb lesz a kárkifizetés:

$$\mathbb{P}(X > 200000) = \mathbb{P}(X > 200000 \mid X > 77477)\mathbb{P}(X > 77477) \approx \quad (4.6)$$

$$\approx (1 - (1 - (1 + \xi \frac{x-u}{\sigma})^{-\frac{1}{\xi}}))\mathbb{P}(\Gamma(n, a, \tau, Z) \geq 1 \mid \mathcal{F}_t) = \quad (4.7)$$

$$= (1 + 0.1922 \cdot \frac{200000 - 77477}{257715.7})^{-\frac{1}{0.1922}} \cdot 0.9421 = 0.5978 \quad (4.8)$$

Tehát 59.78% valószínűséggel fog bekövetkezni egy 200 ezernél nagyobb kár a május szeptemberi időszakban. A kárkifizetésre adott becslés (aminek segítségével be tudjuk árazni a szerződést) pedig X várható értéke lesz:

$$\mathbb{E}(X) = u + \frac{\sigma}{1 - \xi} = 77477 + \frac{257715.7}{1 - 0.1922} = 396511.04 \quad (4.9)$$

Ezt megszorozva a $\mathbb{P}(\Gamma(n, a, \tau, Z) \geq 1 \mid \mathcal{F}_t)$ mennyiséggel, majd diszkontálva megkaphatjuk a 3.3 egyenletben definiált szerződés árát:

$$V(t, T) = e^{-r(T-t)} \cdot \mathbb{E}(K) \cdot \mathbb{P}(\Gamma(n, a, \tau, Z) \geq 1 \mid \mathcal{F}_t) = 373553.05 \cdot e^{-r(T-t)} \quad (4.10)$$

5. fejezet

Konklúzió

A dolgozat célja az volt, hogy a hőhullámok által okozott extrém mezőgazdasági károkat kvantitatívan megbecsüljem hőmérsékleti és kárkifizetési adatok elemzése révén. A kutatás során két fő módszertani irányt követtem: egyrészt a hőhullámok bekövetkezési valószínűségének becslését végeztem el hőmérsékleti idősorok alapján, másrészt pedig az extrém káresemények valószínűségi eloszlásának modellezését valós kárkifizetési adatok felhasználásával. A kutatásom során egy olyan modellt próbáltam felépíteni, ami figyelembe veszi az adatok idősoros szerkezetét, így képes megbízható valószínűséget adni a vizsgált időszakra. Az eredmények alapján bizonyos esetekben viszonylag magas bekövetkezési valószínűséget kaptam még olyan hőhullámra is, amelyek az elmúlt húsz év során egyáltalán nem fordultak elő. Ez azt jelenti, hogy a múltbeli megfigyelések hiánya nem jelenti azt, hogy a jövőben a hasonló események kizárhatóak lennének.

A káresemények súlyosságának modellezésére extrémérték-elméleti módszertant alkalmaztam, ezen belül is a küszöb feletti meghaladások módszerét. Az alkalmazáshoz megfelelő küszöbérték megválasztása kulcsfontosságú, mivel ez jelentősen befolyásolja az eredmények robusztusságát. A dolgozatomban bemutatott elemzések

alapján azonban megállapítható, hogy a modell a küszöbérték megválasztására nem érzékeny, az eredmények konzisztensnek és stabilnak bizonyultak különböző, észszerűen megválasztott küszöbök esetén is. Ez különösen fontos a gyakorlati alkalmazás szempontjából, hiszen növeli a modell prediktív képességébe vetett bizalmat.

Ez a megközelítés jól alkalmazható lehet biztosítói kockázatkezelésben, mezőgazdasági biztosítási termékek tervezésében, illetve kockázatalapú tőkeallokáció során is. Fontos kiemelni, hogy az alkalmazott hőmérsékleti modell az idősoros adatok autokorrelációját, szórását, trendjét és szezonális mintázatát is figyelembe vette, így az előrejelzések a lehető legnagyobb mértékben tükrözték a valós folyamatok időbeli dinamikáját.

Természetesen ennek a modellnek is vannak korlátai, mivel a kiinduló adatbázis nem tartalmazza az egész országra vonatkozó aszálykárokat. Egy nagyobb magyar biztosítótól származnak az adatok, ahogyan azt a 4.1. fejezetben is említettem, viszont a szóban forgó cégnek nincs 100%-os piaci részesedése. Ezen kívül van szó olyan károkról is, amit az állam térít, illetve olyanok is, ahol a gazdának kell viselnie a kárt, így ezek az esetek sem kerülnek bevonásra az elemzésben.

Irodalomjegyzék

- [1] Ahčan, A. *Statistical analysis of model risk concerning temperature residuals and its impact on pricing weather derivatives*. Insurance: Mathematics and Economics, 50(1), 131–138, 2012.
- [2] Asquith, W. H. *L-moments and TL-moments of the generalized lambda distribution*. Computational Statistics & Data Analysis, 51(9), 4484–4496, 2007.
- [3] Beaulieu, C., Gallagher, C., Killick, R., Lund, R., & Shi, X. *A recent surge in global warming is not detectable yet*. Communications Earth & Environment, 5(1), 576, 2024.
- [4] Cao, M., & Wei, J. *Pricing weather derivative: an equilibrium approach*. Rotman School of Management, University of Toronto, Working Paper No. 99-002, 1999.
- [5] Campbell, S. D., & Diebold, F. X. *Weather forecasting for weather derivatives*. Journal of the American Statistical Association, 100(469), 6–16, 2005.
- [6] Chen, Y., & Li, Y. *An inter-comparison of three heat wave types in China during 1961–2010: Observed basic features and linear trends*. Scientific Reports, 7(1), 45619, 2017.

- [7] Delicado, P., & Goría, M. N. *A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution*. Computational Statistics & Data Analysis, 52(3), 1661-1673. 2008.
- [8] Embrechts, P., Klüppelberg, C., & Mikosch, T. *Modelling extremal events: for insurance and finance* (Vol. 33). Springer Science & Business Media, 2013.
- [9] Engle, R. F. *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. Econometrica, 50(4), 987–1007, 1982.
- [10] Fenner, D., Holtmann, A., Krug, A., & Scherer, D. *Heat waves in Berlin and Potsdam, Germany—Long-term trends and comparison of heat wave definitions from 1893 to 2017*. International Journal of Climatology, 39(4), 2422–2437, 2019.
- [11] Forster, P. M., Smith, C. J., Walsh, T., Lamb, W. F., Lamboll, R., Hauser, M. & Zhai, P. *Indicators of Global Climate Change 2022: annual update of large-scale indicators of the state of the climate system and human influence*. Earth System Science Data, 15(6), 2295–2327, 2023.
- [12] García-León, D., Casanueva, A., Standardi, G., Burgstall, A., Flouris, A. D., & Nybo, L. *Current and projected regional economic impacts of heatwaves in Europe*. Nature communications, 12(1), 5807. 2021.
- [13] Larsson, K. *Parametric heat wave insurance*. Journal of Commodity Markets, 31, 100345, 2023.
- [14] MacKenzie, C. A., & Winterstein, S. R. *Comparing L-moments and conventional moments to model current speeds in the North Sea*. Industrial Engineering Research Conference, Reno, NV, USA (pp. 21-25). 2011.

- [15] Pandey, M. D., Van Gelder, P. H. A. J. M., & Vrijling, J. K. *The use of L-moments in the peak over threshold approach for estimating extreme quantiles of wind velocity*. Probabilistic Engineering Mechanics, 16(1), 1–12. 2001.
- [16] Perkins, S. E., & Alexander, L. V. *On the measurement of heat waves*. Journal of Climate, 26(13), 4500–4517, 2013.
- [17] Rydman, M. *Application of the peaks-over-threshold method on insurance data*. Master’s Thesis, Aalto University, School of Science, Department of Mathematics and Systems Analysis, 2018.
- [18] Smoyer-Tomic, K. E., Kuhn, R., & Hudson, A. *Heat wave hazards: an overview of heat wave impacts in Canada*. Natural Hazards, 28, 465–486, 2003.
- [19] Sutanto, S. J., Zarzoza Mora, S. B., Supit, I., & Wang, M. *Compound and cascading droughts and heatwaves decrease maize yields by nearly half in Sinaloa, Mexico*. npj Natural Hazards, 1(1), 26, 2024.
- [20] Torró, H., Meneu, V., & Valor, E. *Single factor stochastic models with seasonality applied to underlying weather derivatives variables*. The Journal of Risk Finance, 2(4), 17-30. 2001.
- [21] Weitzman, M. L. *GHG targets as insurance against catastrophic climate damages*. Journal of Public Economic Theory, 14(2), 221–244, 2012.
- [22] Adatok: *European Climate Assessment & Dataset*.
<https://www.ecad.eu/dailydata/predefinedseries.php>

Ábrajegyzék

4.1. A hőmérséklet időbeli alakulása	34
4.2. Determinisztikus modell hibatagja	36
4.3. Napi hőmérséklet autokorrelációja (ACF)	37
4.4. Napi hőmérséklet parciális autokorrelációja (PACF)	38
4.5. Hibatagok eloszlása	40
4.6. AR(3)modell hibatagja	41
4.7. Szórás előrejelzése GARCH modellből	43
4.8. Hőmérséklet előrejelzése GARCH modellből	44
4.9. Aszálykár-kifizetések autokorrelációja	47
4.10. Átlagos küszöb meghaladások	48
4.11. Különböző modellek farokeloszlása	50

Táblázatjegyzék

4.1. Hőhullámok száma az egész időszakra tekintve	33
4.2. Fourier együtthatók	35
4.3. AR(p) modellek együtthatói és AIC/BIC értékei	39
4.4. ARCH(p)/GARCH(p,q) modellek együtthatói és AIC/BIC értékei . .	42

4.5. Hőhullámok bekövetkezési valószínűsége a 2025 május-szeptemberi időszakra nézve	46
4.6. Küszöbértékek különböző módszerek esetén	48
4.7. Becsült paraméterek	49
4.8. Becsült paraméterek	51
4.9. Extrém kifizetésű hónapok	53

MI-eszközhasználati nyilatkozat

Alulírott, Nguyen Hoang An, nyilatkozom, hogy szakdolgozatom elkészítése során az alább felsorolt feladatok elvégzésére a megadott MI-alapú eszközöket alkalmaztam:

Feladat	Felhasznált eszköz	Felhasználás helye	Megjegyzés
Nyelvhelyesség ellenőrzése, illetve LATEX kód írása	ChatGPT	Teljes dolgozat	Helyesírás ellenőrzés, illetve az egyes táblázatok és képletek kódja LATEX-ban

A felsoroltakon túl más MI-alapú eszközt nem használtam.