

OTDK-dolgozat

2025

NEUMANN JÁNOS EGYETEM
MNB INTÉZET: FENNTARTHATÓ PÉNZÜGYEK KÖZPONT
NEMZETKÖZI GAZDASÁG ÉS GAZDÁLKODÁS MESTERSZAK



Neumann János Egyetem
Gazdaságtudományi
Kar



HÍREKBŐL FOGYASZTÓI VÁRAKOZÁSOK:
TOPIK ALAPÚ ELŐREJELZÉS

NEWS-DRIVEN CONSUMER EXPECTATIONS:
TOPIC-BASED FORECAST

Hollósi István

Konzulens: Granát Marcell Péter

Kézirat lezárásának dátuma: 2024. 04. 15.

TARTALOMJEGYZÉK

1.	BEVEZETÉS	1
2.	ADATOK ÉS MÓDSZERTAN	4
2.1.	AZ ADATOK: VÁRAKOZÁSOK, HÍREK ÉS MAKROGAZDASÁGI MUTATÓK	4
2.1.1.	<i>Fogyasztói inflációs várakozás és bizalmi index</i>	<i>4</i>
2.1.2.	<i>Online hírportálok cikkei</i>	<i>5</i>
2.1.3.	<i>Makrogazdasági mutatók</i>	<i>5</i>
2.2.	MÓDSZERTAN: LDA TOPIC MODELLEZÉS ÉS GÖRDÜLŐ ELŐREJELZÉS	6
3.	EREDMÉNYEK	10
3.1.	FOGYASZTÓI BIZALMI INDEX ELŐREJELZÉSE	10
3.2.	HÁZTARTÁSOK INFLÁCIÓS VÁRAKOZÁSÁNAK ELŐREJELZÉSE	12
4.	KONKLÚZIÓK	15
	IRODALOMJEGYZÉK	16
	MELLÉKLETEK.....	I
1.	MELLÉKLET: VÁRAKOZÁSOK HAVI VÁLTOZÁSÁNAK IDŐSORAI	II

TÁBLÁZAT- ÉS ÁBRAJEGYZÉK

1. táblázat: a fogyasztói inflációs várakozás és bizalmi index leíró statisztikái	5
2. táblázat: az elemzésbe bevont cikkek évek és hírportálok szerinti megoszlása	5
3. táblázat: makrogazdasági magyarázó változók a modellekben	6
4. táblázat: a fogyasztói bizalmi index előrejelző modelljeinek értékelése	11
5. táblázat: a háztartások inflációs várakozásának modelljeinek értékelése	14

1. ábra: az alkalmazott step-forward validation eljárás ábrázolása	7
2. ábra: a fogyasztói bizalmi index és annak előrejelzései a havi változás előrejelzésekből kumulálva	11
3. ábra: a háztartások inflációs várakozása és annak előrejelzései a havi változás előrejelzésekből kumulálva	13
4. ábra: a fogyasztói bizalmi index havi változása és annak előrejelzései	II
5. ábra: a háztartások inflációs várakozásának havi változása és annak előrejelzései	II

ABSZTRAKT

A gazdasági stabilitás és hatékony gazdaságpolitika szempontjából elengedhetetlen a háztartások gazdasági várakozásainak megértése. A tanulmány a hírek és a fogyasztói bizalmi index, valamint inflációs várakozások közötti összefüggéseket vizsgálja. Mintegy 2 millió online cikkből építettem LDA topik modelleket, amelyekkel előrejeleztem a fogyasztói várakozásokat. Míg a havi elmozdulásokat nem jelezték előre modelljeim, a változások összegzésével képzett idősorok jól visszaadták a tény idősorok számait. Kutatásom a várakozások hírekkel történő előrejelzésének abból adódó korlátjára is rávilágít, hogy számos esetben a témák legfeljebb egy időben jelennek meg a sokkokkal.

Understanding households' economic expectations is essential for economic stability and effective economic policy. The study examines the relationships between news, consumer confidence index, and inflation expectations. Using LDA topic models built from approximately 2 million online articles, I forecasted consumer expectations. While the models did not predict monthly fluctuations, they well represented aggregate changes over time. Additionally, my research also highlights the limitations of predicting expectations based on news, as in many cases, topics do not appear before the shocks themselves.

1. BEVEZETÉS

A fogyasztói várakozások mutatói a gazdaságok széles körben figyelemmel kísért barométerei, köszönhetően előrejelző képességüknek, amelyet számos tanulmány bizonyított (például Adrangi & Macri, 2011; Kwan & Cotsomitis, 2006; Li, 2011). Mivel a fogyasztói döntéseken keresztül hatással vannak a vállalatok beruházásaira, a pénzügyi piacokra és a gazdaság- és monetáris politika hatékonyságára is, a fogyasztói várakozások és azok mozgatóinak jobb megértése hozzájárulhat a hatékonyabb gazdaság- és monetáris politika kialakításához. A várakozások, különösképpen az inflációs várakozások, sokat kutatott terület, ugyanakkor az, hogy pontosan milyen tényezők is alakítják azokat, vitatott kérdés. Azt biztos, hogy a média jelentősen alakítja a fogyasztói várakozásokat, mint az emberek többsége számára (az egyik) elsődleges információforrás (Doms & Morin, 2004).

Tanulmányomban az online hírek fogyasztói inflációs várakozásra és bizalmi indexre gyakorolt hatását vizsgáltam, különös tekintettel arra, hogy *lehetséges-e a hírportálok hírei alapján előrejelezni a háztartások gazdasági várakozásait és azok változását?*

Az *information rigidity* koncepciója szerint a gazdasági szereplők nem mindig friss vagy teljes információkat használnak fel döntéseik meghozatalakor. Ennek oka az, hogy az információ megszerzése és feldolgozása költséges folyamat (Mankiw et al., 2003; Coibion & Gorodnichenko, 2015). Egyetlen egyén sem rendelkezik elegendő erőforrással, hogy minden olyan eseményt figyelemmel követhessen, ami potenciális hatással lehet döntéseire. Ezért átvadják ezt a feladatot egy szakosodott hírszolgáltatónak, aki a nevükben követi az eseményeket és csak egy szűrt összeállítást jelenít meg számukra (Nimark & Pitschner, 2019). Az egyén számára rendelkezésre álló információt azonban nem csak ez a jelenség torzítja és szűkíti, hanem egyrészt a médiának az a jellemzője, hogy a változásokat hajlamos túlhangsúlyozni, elsősorban azért, mert a status quo nem bír hírértékkel (Van Raaij, 1989; Van Dalen et al., 2017), másrészt a kormányzatok és jegybankok szintén használják a médiát a várakozások horgonyzásához.

A merevségnek azonban egy másik oka is lehet, ha a háztartások ritkán vizsgálják felül és változtatják inflációs várakozásaikat. Carroll (2003) egy olyan modellt épített, ahol a háztartások várakozásai professzionális előrejelzésekből táplálkoznak híradások által közvetítve. Azt találta, hogy egy átlagos háztartás évente egyszer változtat várakozásain. A

háztartások azonban még abban az esetben sem biztos, hogy megváltoztatják várakozásaikat, ha tényszerű információval rendelkeznek a fundamentumokról: Binder (2021) azt találta, hogy nagy mértékű pozitív inflációs sokk hírére sem változtak jelentősen a háztartások inflációs várakozásai rövid távon.

Azonban még ha azt is feltételeznénk, hogy azonos információ birtokában alakítják ki várakozásaikat, hozzák meg döntéseiket a háztartások, akkor sem várhatunk azonos várakozásokat és döntéseket. Az inflációs várakozásokat egy sor szociodemográfiai tényező befolyásolja, úgymint a nem, jövedelem, végzettség vagy a háztartás összetétele. Jellemzően azok a háztartások, ahol a mindennapi kiadások is gondot jelentenek, illetve az alacsonyabb pénzügyi tudatosságú egyedülállók jelentősen magasabb inflációs várakozással bírnak (Bruine de Bruin et al., 2010). Malmendier és Nagel (2016) generációs különbséget is talált az infláció várakozások kialakításában: megállapították, hogy az egyének hajlamosak az életük során tapasztalt inflációt felülsúlyozni várakozásaikban, valamint, hogy a fiatalabbak jellemzően nagyobb mértékben tudják változtatni várakozásaikat, amennyiben változnak a körülmények.

Emellett a hírfogyasztási szokások jelentősen eltérnek az egyes társadalmi csoportok összevetésében: ha csak a legolvasottabb hazai online sajtótermék weboldalait vizsgáljuk, azok elérése hasonlóan alakul nem, korcsoport vagy végzettség tekintetében, ugyanakkor az oldalon töltött idő jelentős szórást mutat mindhárom szociodemográfiai változó mentén (Médiapiaci jelentés 2023, 2023).

A háztartások inflációs várakozásának azonban egy kézenfekvő és jelentős negatív értelemben vett horgonyzója tud lenni az, ha gyakori kiskereskedelmi árváltozással szembesülnek. Ulrike és Juan (2019) tanulmánya szerint a kiskereskedelemben tapasztalt gyakori árváltozással való szembesülés erősebben befolyásolja a várakozásokat, mint a jövedelem-arányos kiadás.

Carroll (2003) felveti, hogy az elérhető gazdasági híradások mennyiségének növekedésével nőhet a várakozások változtatásának, felülvizsgálatának valószínűsége, amit Doms és Morin (2004) igazolt tanulmányában. Megállapították továbbá, hogy a média a fogyasztók gazdaságról alkotott képét alapvetően három szinten befolyásolja: egyrészt közli a gazdaság aktuális állapotát, arról az aktuális szakértői véleményeket, másrészt a közlés tonalitása, harmadrészt a gazdasággal kapcsolatos hírek mennyisége is fontos tényezője a várakozások alakulásának (Doms & Morin, 2004).

A fogyasztói bizalmi indexet (és az inflációs várakozást sem feltétlenül) kizárólag a szigorú értelemben vett gazdasági hírek befolyásolhatják. Larsen, Thorsrud és Zhulanova (2021) tanulmányukban LDA topik modellt használva azt találták, hogy az inflációs várakozást olyan témák is szignifikánsan befolyásolják az Egyesült Államokban, mint az egészségügy vagy az Internet. Larsen és Thorsrud (2022) egy másik tanulmányban szintén topik modellt alkalmazva eszközárak előrejelzésére építettek modellt Norvég üzleti napilapok cikkeiből, míg Yukhymenko (2022) szintén inflációs várakozásokat magyarázott hasonló módszertannal.

Tanulmányomban LDA topik modellt illesztettem 10 év online hírportálokon megjelent híreire 24 hónapos gördülő ablakkal, majd az így kapott topik modellekre LASSO regressziót alkalmazva előrejelzést készítettem a várakozások havi változásaira.

Azt találtam, a szakirodalommal egyetértésben, hogy a várakozások igen ragadósak, jellemzően csak jelentős sokkhatások mozdítják meg azokat. Random walk folyamattal leírhatóak, a jelentős szintváltozásokat csak egy periódus késéssel tudtam előrejelezni. Ennek oka, hogy a vizsgált időszakban jelentős elmozdulást előidéző események az eseményeket megelőző 23 hónap jelentős témáival nem leírhatóak (a Covid19-pandémia), vagy ha a téma jelenléte már észlelhető is volt, az korábban nem járt együtt a fogyasztói várakozások változásával (Ukrajna orosz inváziója). Az orosz-ukrán háború a tekintetben az is figyelemre-méltó, hogy azt a részvénytőke legalább 50 nappal korábban elkezdte árazni, ugyanakkor ugyanabban az időszakban a *global economic policy uncertainty index* még nem mutatott növekedő kockázatot (Granát et al., 2023).

További fontos eredményem, hogy a makrogazdasági mutatókkal kontrollált modellek nem magyaráznak jobban, az azokkal készült előrejelzés nem illeszkedik jobban a tényadatokra. A friss statisztikai adatok nem tudnak javítani a háztartások várakozásainak előrejelzésén a tisztán hírekből épített modellhez képest, mivel e mutatók kisebb-nagyobb elmozdulásait a háztartások legritkább esetben tapasztalják meg azonnal, azok a hírekben is jellemzően csak akkor jelennek meg, ha nagy kiugrás tapasztalható bennük.

Dolgozatom felépítése a következő: elsőként ismertetem a felhasznált adatokat és forrásait, majd az előrejelzés módszertanát, ezt követően a fogyasztói bizalmi indexet, majd a háztartások inflációs várakozását előrejelző modellek eredményeit mutatom be, végül tanulmányomat áttekintve összefoglalom azt.

2. ADATOK ÉS MÓDSZERTAN

2.1. AZ ADATOK: VÁRAKOZÁSOK, HÍREK ÉS MAKROGAZDASÁGI MUTATÓK

2.1.1. Fogasztói inflációs várakozás és bizalmi index

Kutatásom során az Európai Bizottság Gazdasági és Pénzügyi Főosztálya (DG ECFIN) által koordinált, havi rendszerességű kérdőíves fogyasztói felmérés Magyarországra vonatkozó inflációs várakozás és bizalmi index mutatóit jeleztem előre. A kutatást Magyarországon 1000 fős havi megkérdezésekkel minden hónap első napjaiban végzi a GKI Gazdaságkutató Zrt. Az adatfelvétel ilyen időzítése lehetővé teszi, hogy a tárgyhónapot az azt megelőző naptári hónappal magyarazzuk.

A fogyasztói inflációs várakozás mutatója egy balansz mutató, így nem a tényleges várakozásokat tükrözi, hanem az infláció gyorsulására és lassulására számító háztartások arányának súlyozott különbségét a mintában. A kérdés *„Összehasonlítva a jelenlegivel, Ön szerint hogyan alakulnak az árak a következő 12 hónapban?”*, amire egy ötfokú aszimmetrikus skálán válaszolhatnak a megkérdezettek, ahol a négyes érték jelenti, hogy nem fognak változni az árak. Minden esetben szezonális hatásoktól tisztított adatsorokat használtam.

A fogyasztói bizalmi index a háztartások gazdasággal kapcsolatos optimizmusának összetett mutatója, ami azonban az inflációs várakozásokat direkt módon nem tartalmazza. Négy balansz mutató számtani átlaga, amelyek a háztartás szubjektív pénzügyi helyzete az elmúlt 12 hónaphoz hasonlítva és a következő 12 hónapra vonatkozó várakozások, az ország gazdasági kilátásaira vonatkozó várakozások és nagyértékű fogyasztási cikkekre szánt diszkrécionális jövedelemmel kapcsolatos várakozások (The Joint Harmonised EU Programme of Business and Consumer Surveys, 2024).

1. táblázat: a fogyasztói inflációs várakozás és bizalmi index leíró statisztikái

A táblázat az előrejelzendő várakozások leíró statisztikáit mutatja a 2016. február – 2023. december időszakra vonatkozóan: átlagokat, szórásokat, minimumokat, 25. percentiliseket, mediánokat, 75. percentiliseket és maximumokat.

	Átlag	Szórás	Minimum	25. percentilis	Medián	75. percentilis	Maximum
Fogyasztói inflációs várakozás (index)	35.26	13.73	14	22	35.5	44.5	63.2
Fogyasztói bizalmi index	-15.88	14.00	-49.2	-24	-10	-5	1.9

Forrás: Eurostat, saját szerkesztés

2.1.2. Online hírportálok cikkei

Hét vezető magyar online hírportál (24.hu, blikk.hu, hvg.hu, index.hu, origo.hu, portfolio.hu, telex.hu) tíz évre visszanyúló cikkgyűjteménye állt rendelkezésemre a korpusz megépítéséhez a Neticle Plc. jóvoltából. Az egyes hírportálok lefedettsége nem egyenletes, időben folyamatosan javul, azonban még ha minden oldal minden cikke nem is áll rendelkezésemre minden időpontra, a topik modellel azonosított témák részesedését minden esetben jól leírják az elérhető híroldalak, tekintve azok általános hírportál jellegét, azokon legfeljebb tonalitásukban térhetnek el a téma említések. Az adattisztítás után mintegy 1,95 millió online újságcikkre illesztettem LDA topik modelleket gördülő ablakban, így azonosítva egyes témák prevalenciáját.

2. táblázat: az elemzésbe bevont cikkek évek és hírportálok szerinti megoszlása

A táblázatban a Neticle Media Intelligence által rendelkezésemre bocsátott cikkgyűjtemény cikkeinek darabszám szerinti megoszlása található éves bontásban híroldalanként. Az egyes domainek tartalmazzák az sub-domaineken megjelent cikkeket is.

	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
24.hu	16	30	73	3770	51863	27351	40659	39343	39771	40481
origo.hu	27989	39712	36977	30045	50719	59852	64734	64809	62584	57881
blikk.hu	38583	47048	17167	60248	70405	59426	50906	61118	51851	70740
hvg.hu	1	0	1	2	0	9996	34972	33870	36168	35370
index.hu	249	305	312	699	36821	35825	37561	35442	43295	44331
portfolio.hu	16153	16365	13248	15069	18636	17003	22767	26459	33500	33813
telex.hu	0	0	0	0	0	0	5791	22251	21655	21281

Forrás: Neticle Media Intelligence, saját szerkesztés

2.1.3. Makrogazdasági mutatók

A makrogazdasági mutatókkal kontrollált regressziókhoz az Eurostat Magyarországra havi szinten elérhető mutatóit használtam, melyek az alábbiak voltak:

3. táblázat: makrogazdasági magyarázó változók a modellekben

A táblázat a kontrollált regressziókhoz használt havi mutatók leíró statisztikáit mutatja a 2016 február – 2023 december időszakra vonatkozóan: átlagokat, szórásokat, minimumokat, 25. percentiliseket, mediánokat, 75. percentiliseket és maximumokat. A forrás minden esetben az Eurostat. A csillaggal jelölt idősorok első késleltetéssel vett változását használtam a modellezéshez.

	Átlag	Szórás	Minimum	25. percentilis	Medián	75. percentilis	Maximum
Ipari termelés (havi változás)	0.28	4.65	-29.00	-1.15	-0.10	1.65	17.30
Munkanélküliség*	3.91	0.50	3.20	3.50	3.80	4.20	5.50
EUR/HUF árfolyam*	343.70	31.29	304.37	313.57	334.38	365.06	418.31
EMU konvergencia kritérium*	4.01	2.22	1.83	2.53	3.10	4.56	10.25
HICP (év/év változás)*	6.36	6.91	-0.30	2.50	3.40	7.45	26.20
Építési engedélyek száma*	188.73	37.20	102.60	165.20	188.00	212.55	275.00
Építőipari termelés*	134.31	27.63	73.20	112.95	146.10	154.95	180.60
Termelői árak az iparban	0.43	1.08	-3.90	-0.10	0.20	0.80	4.20
Ipari termelés (2015 bázison)*	113.68	9.60	72.30	107.25	116.50	120.75	128.60
Tőkejavak forgalma (2015 bázison)	103.23	10.75	41.40	98.90	102.40	111.35	119.00
Fogyasztói javak forgalma (2015 bázison)*	114.14	9.77	94.90	105.95	114.90	121.95	133.00
Szolgáltatásiipari termelés (havi változás)	0.46	3.13	-18.60	-0.75	0.60	2.25	6.20
Szolgáltatásiipari forgalom*	157.80	36.92	103.30	129.50	150.25	183.48	226.10
Kiskereskedelem volumene*	122.31	10.05	103.40	114.15	124.90	128.50	144.90
Vendégéjszakák száma	30.02	118.62	-93.90	-3.29	4.71	11.86	607.55
Folyó fizetési mérleg*	-155.5	650.4	-2419.6	-402.2	-65.9	297.5	920.3
Pénzügyi számla	-226.3	864.8	-2674.5	-690.1	-88.1	176.5	1643.9
HICP (havi változás)*	0.51	0.70	-0.50	0.10	0.30	0.75	3.70
Energia felhasználás	31812	15522	13341	16956	27826	45438	69485

Forrás: Eurostat, saját szerkesztés

2.2. MÓDSZERTAN: LDA TOPIC MODELLEZÉS ÉS GÖRDÜLŐ ELŐREJELZÉS

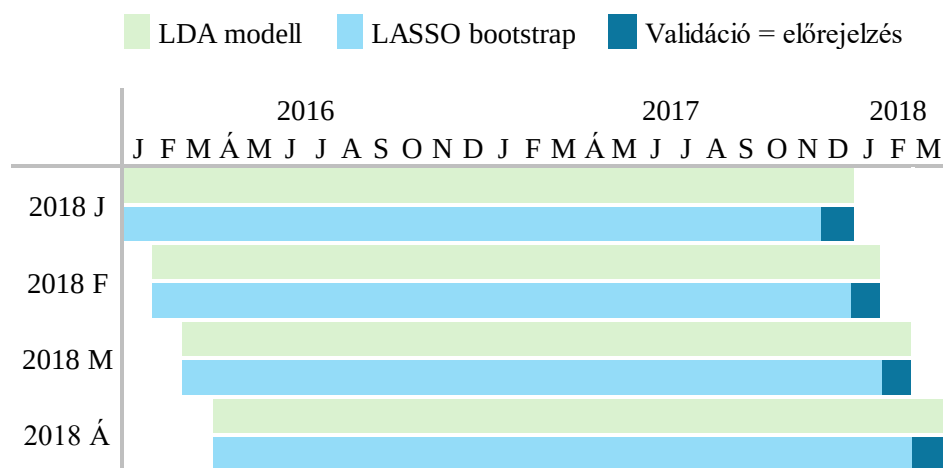
Kutatásom fő célja egy olyan előrejelző modell építése volt, amely mentes az előretekinthető torzítástól (*look ahead bias*). Gyakori torzítás a hasonló tanulmányokban, hogy a topik modellt a teljes vizsgált időszak adataira illesztik. Ez különösen problémás lehet a hírek esetében, ahol az állandó témakörök mellett (például oktatás, egészségügy, bulvár), törvényszerűen jelennek meg újabb és újabb újdonságok, amiket a korábban azonosított topikok kevésbé vagy nem is tudnak leírni. De még ha ez nem is állna fent, mivel a topik elemzés során korlátos számú témakörre egyszerűsítjük le a hírek tartalmát, elkerülhetetlen, hogy idővel a releváns témakörök változzanak. A topikok evolúcióját vizsgálta tanulmányában Rieger, Jentsch és Rahnenführer, akik egy gördülő LDA modellt javasoltak, ami lehetővé teszi

olyan LDA alapú topik idősorok előállítását, amelyek konzisztensek az LDA modell korábbi állapotaival (Rieger et al., 2021). Ennek vizsgálata kívül esik tanulmányom hatókörén.

Előrejelző modelletem egy egyszerű *step-forward validation* dizájnban építettem fel, ahol a korpusz 24 havi szeleteire illesztettem topik modellt havi lépésekkel görgetve, majd LASSO regressziót futtattam az első 23 hónap adataira, amivel előrejeleztem a 24-ik hónap fogyasztói inflációs várakozásait, majd egy második modellel a bizalmi index értékeit, mindvégig azok egy hónappal eltolt értékeit használva. Az alábbi ábrán szemléltetve az eljárást, például LDA modellt illesztettem a 2016. január - 2017. decemberi hírekre, majd ezt a havi idősort felhasználva LASSO regressziót illesztettem a várakozások 2016. február - 2017. decemberi adatsorára, egy hónappal elcsúsztatva azt, majd az így kapott modellel előrejeleztem a 2018 januári várakozásokat, pontosabban a havi változást. Az így kapott előrejelzéseket aztán egy idősorba összegeztem, tehát minden egyes hónap előrejelzéséhez egyedi LDA modellt használtam és várakozásonként külön LASSO modellt, ami összességében 2x95 külön modell a teljes időszakra.

1. ábra: az alkalmazott step-forward validation eljárás ábrázolása

Az alábbi ábra szemlélteti a modellezés dizájnját. Minden előrejelzett havi értékhez két sáv tartozik az ábrán, amelyek a következő folyamatot szemléltetik: (1) LDA topik modell illesztés 24 havi adatra, (2) az így kapott havi idősor első 23 hónapjával LASSO modell tréningje, ami minden esetben a következő havi várakozásokat magyarázza, (3) majd legvégül az így tanított modellel az utolsó havi LDA topik megoszlásokkal a következő havi fogyasztói inflációs várakozás és bizalmi index havi változásának előrejelzése. Az egymás után futtatott modellek eredményeit összegezve kapunk olyan teljes időszakra vonatkozó előrejelzést, ami minden esetben adott t időpontban elérhető t-23 időpontig visszatekintő adatokból jelzi előre a t+1 értéket, mentesen az előretekinő torzítástól a topikmodell tekintetében is.



Forrás: saját szerkesztés

A Blei, Ng és Jordan által bemutatott (Blei et al., 2003) Latent Diriclet Allocation (LDA) az egyik legnépszerűbb felügyelet nélküli algoritmus a természetes nyelvi feldolgozás (NLP) területén, köszönhetően egyszerűségének és rugalmasságának, valamint annak a tulajdonságának, hogy a szövegek hasonló osztályozására képes, mint az emberek (Chang et al., 2009). Az eljárás lényege két alapvetésben foglalható össze: (1) minden dokumentum topikok keveréke, vagyis minden dokumentum számos topik kifejezéseit tartalmazhatja egy bizonyos arányban, (2) míg minden topik kifejezések keveréke. Az LDA egy olyan módszer, amely egyszerre becsli mindkettőt: a dokumentumokhoz kapcsolódó témákat és a témákat jellemző kifejezéseket (Silge & Robinson, 2017).

A modellezést megelőzően az általános gyakorlatnak megfelelő tisztításokat végeztem el a szöveges adatbázison: tokenizálást követően *stemmeltem* (ez a szótőre alakításnál kevésbé pontos ám jelentősen gyorsabb eljárás), eltávolítottam az ún. *stopwordöket* (azokat a leggyakrabban használt kifejezéseket, amelyek önálló jelentéssel nem bírnak, mint például a névelők, igekötők), végül a túl kevés vagy túl sok szövegben előforduló kifejezéseket is kiszűrtem (tekintve, hogy azok osztályozásra ezért kevésbé alkalmasak). Az eljárás igényli, hogy előre definiáljuk a létrehozandó topikok számát, amit szűk szakirodalomra támaszkodva 80-ban határoztam meg (Larsen & Thorsrud, 2019; Thorsrud, 2020; Larsen et al., 2021).

A várakozások magyarázásához Least Absolute Shrinkage and Selection Operator (LASSO, Tibshirani, 1996) eljárást használtam, mivel a magyarázóerővel bíró topikok (vagyis az, hogy milyen típusú hírek befolyásolják ténylegesen a háztartások inflációs várakozását és a bizalmi indexet) előre nem ismertek. Ezen felül az egymást követő időszakokra egyesével illesztett LDA modellek egy nem konzisztens magas dimenziószámú topik teret eredményeznek, ahol hasonlósági alapon ugyan nyomon követhető lehet adott topik „hangerejének” alakulása és az azt leíró szófelhő változásai is, de 1:1 megfeleltetésük periódusról periódusra nem lehetséges. A LASSO eljárás a nem fontos változók paramétereinek becslését nulla felé zsugorítja, egyszerűbb és szűkebb modelleket eredményezve. A modell paraméterezéséhez bootstrapping mintavételi eljárást használtam a modellenkénti alacsony elemszám miatti túlillesztés kockázatát mérséklendő (17 hónapra illesztettem alkalmanként, 6 hónapon teszteltem és az utolsó hónap a bootstrapping eljárás után a validáció volt), minden esetben a legmagasabb r-négyzetet eredményező modellt választva.

Az előrejelzések formalizálva:

$$\Delta F(\pi_{t+12} | t + 1) = a + \left(\sum_{n=1}^M (b_{1,n} | t) * (T_n | t) \right) + (b_2 | t) * \Delta F(\pi_{t+12} | t) + \varepsilon_t \quad (1)$$

ahol $\Delta F(\pi_{t+12} | t + 1)$ a háztartások következő 12 hónapra vonatkozó inflációs várakozásának havi változása $t+1$ időpontban, M a LASSO-val kiválasztott nem nulla együtthatójú topikok száma $b | t$ a béta együtthatók a t időpontban illesztett regresszióból, $T | t$ a t időpontban illesztett LDA modell topikjai, míg $\Delta F(\pi_{t+12} | t)$ a háztartások következő 12 hónapra vonatkozó inflációs várakozásának havi változása t időpontban. Hasonlóan:

$$\Delta CCI_{t+1} = a + \left(\sum_{n=1}^N (b_{1,n} | t) * (T_n | t) \right) + (b_2 | t) * \Delta CCI_t + \varepsilon_t \quad (2)$$

ahol ΔCCI_{t+1} a háztartások bizalmi indexének havi változása $t+1$ időpontban, N a LASSO-val kiválasztott nem nulla együtthatójú topikok száma $b | t$ a béta együtthatók a t időpontban illesztett regresszióból, $T | t$ a t időpontban illesztett LDA modell topikjai, míg ΔCCI_t a háztartások bizalmi indexének havi változása t időpontban.

3. EREDMÉNYEK

Annak eldöntésére, hogy lehetséges-e a magyar fogyasztók inflációs várakozásának és bizalmi indexének előrejelzése a hírportálok hírkínálata alapján, 95 periódusra készítettem előrejelzést a havi változásokra vonatkozóan, majd ezekből a havi változásokat összegezve készítettem előrejelzett idősorokat. Kutatásom során azt találtam, hogy míg a havi elmozdulásokat nem jelezték előre modelljeim, a változások összegzésével képzett idősorok jól visszaadták a tény idősorok számait, noha számos sokk-esemény környezetében csak egy periódus késéssel. Ez az előrejelző képesség makrogazdasági változók bevonása mellett is megmaradt, mi több a havi gyakorisággal elérhető gazdasági mutatók érdemben nem tudták javítani a modellek illeszkedését és előrejelző képességét.

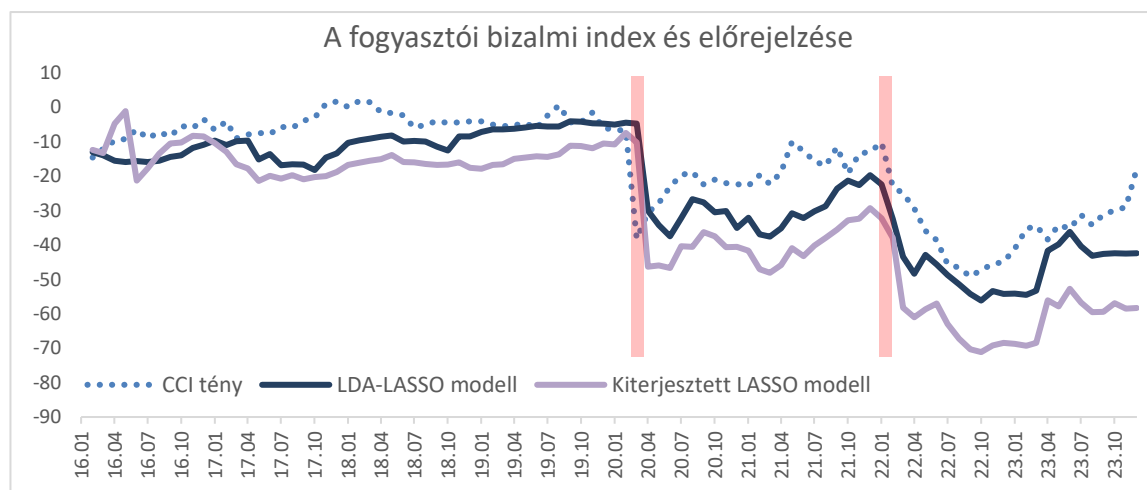
3.1. FOGYASZTÓI BIZALMI INDEX ELŐREJELZÉSE

A mutató havi szintű változását egyértelműen nem tudja előrejelezni sem a tisztán hírekből épített modell, sem a gazdasági mutatókkal kontrollált változat. Ha azonban figyelembe vesszük azt, hogy az 1000 fős mintán kérdőíves megkérdezéssel gyűjtött mutató szórása a teljes vizsgált időszakon 4,77, míg a jelentős szintváltozásokkal járó sokk-hónapok (2020 március-április: Covid19-pandémia és 2022 február: Ukrajna orosz inváziója) eltávolítása után mindösszesen 3,20, akkor már egyáltalán nem meglepő, hogy ezek az apró változások nem előrejelezhetőek hónapról hónapra.

A fogyasztói bizalmi index kapcsolata a reálgazdasággal és a híradásokkal egy még nem tisztázott terület, okkal feltételezhetjük azonban, hogy az idősorok hosszabb időszakon mozognak együtt (például Alsem et al., 2008; Malovaná et al., 2021). Az alábbi grafikonon és táblázatban a havi változások előrejelzésére épített modellek előrejelzéseiből összegzett előrejelzett idősort és a tény adatokat láthatjuk és a modellek értékeléséhez használt főbb mutatókat.

2. ábra: a fogyasztói bizalmi index és annak előrejelzései a havi változás előrejelzésekből kumulálva

Az ábrán a fogyasztói bizalmi index idősorát láthatjuk pöttyözött vonallal, sötétebb folytonos vonallal az LDA-LASSO modell havi változásra vonatkozó előrejelzéseinek kumulált idősorát, míg világosabb folytonos vonallal a makrogazdasági mutatókkal kiegészített regresszió előrejelzéseinek kumulált idősorát. Függőleges vonalakkal jelölve a vizsgált időszak két jelentősebb elmozdulása: a Covid19-pandémia és Ukrajna orosz inváziója.



Forrás: Eurostat és saját számítások

4. táblázat: a fogyasztói bizalmi index előrejelző modelljeinek értékelése

A táblázatban a kizárólag LDA topikokra és a bizalmi index egygel korábbi értékére futtatott (LDA-LASSO modell) és a makrogazdasági változókra kontrollált modell (AUX-LASSO) főbb mutatóit láthatjuk. Az első oszlopokban az eredeti előrejelzés adatai, ahol a havi változás előrejelzése történt (DIF), míg a harmadik és negyedik oszlopokban az első oszlopokból kumulált előrejelzés található.

Mivel minden periódusra saját paraméterekkel becsültem, a pontos magyarázóváltozók egyértelműen nem azonosíthatók, helyettük (#)-vel jelölve a modellek átlagában a LASSO által nem nulla becsült együtthatójú változók átlagos számát tüntetem fel.

	LDA-LASSO DIF	AUX-LASSO DIF	LDA-LASSO CUM	AUX-LASSO CUM
Átlagos R^2 ^a	0,18	0,18		
Tény-Előrejelzés korreláció	-0,03	-0,05	0,89	0,89
MAPE (%)			239	413
MdAPE (%)			63	116
RMSE			35,41	43,18
Topikok magyarázóerővel (#)	10,67	7,74		
Gazdasági mutatók magyarázóerővel (#)		3,48		

^a Az itt szerelt R^2 értékek a bootstrapping minták teszt adatain történő illeszkedések átlagát jelentik az összes időszakra futtatott LASSO regressziók átlagában (95 periódus átlaga).

Forrás: saját számítások

A modellek előrejelző képessége igen gyenge, azok a legtöbb változást csak utólag képesek követni. Figyelemreméltó ugyanakkor az a tény, hogy a havi makrogazdasági mutatók és azok változásai nem javították az előrejelzés pontosságát, hanem rontották, nagyobb átlagos előrejelzési hibákat eredményezve. A fogyasztói bizalmi index idősorának perzisztenciáját szemlélve az a benyomásunk lehet, hogy a naiv előrejelzés igen jó eredményt adhat: az index szintjében csak igen ritkán következik be jelentős változás, így minden periódust jól becsülhetünk az előző periódus szintjével.

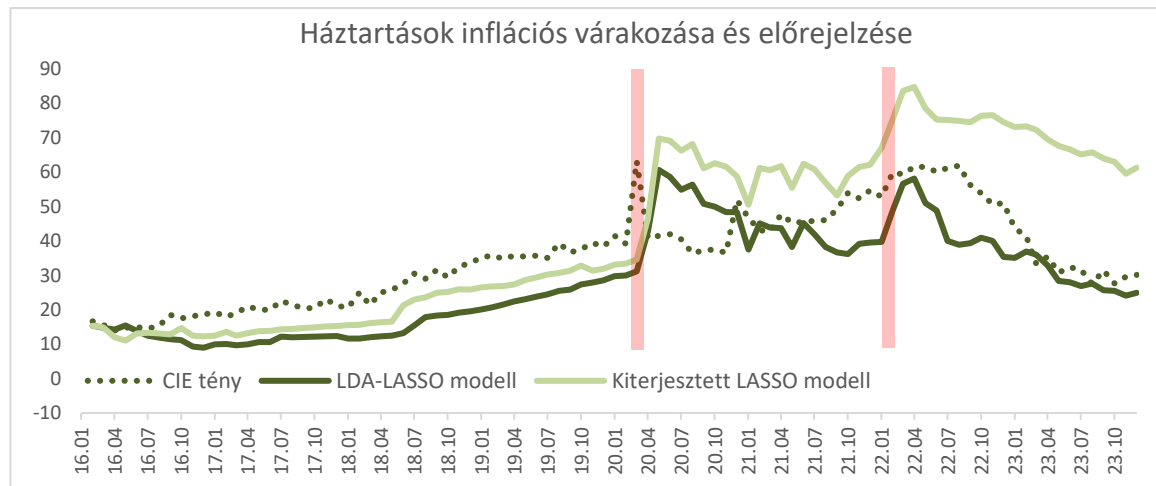
A bizalmi index előrejelző modellem teljesítménye valószínűleg javítható lehet szentiment elemzés hozzáadásával: még ha a sokk-hatású hirtelen felfutó topikokhoz esetleg nem is attributál megfelelő fontosságot az egyes periódusok modellje, ha az előrejelzendő hónap előtti 23 hónap topikjainak nem csak a mennyiségi, de szentiment adatait is bevonnám a LASSO regressziókba, minden bizonnyal a trendfordulókat jobban leíró modelleket kapnék. Nem gépi tanulás módszerével, de hasonló módon jártak el Alsem és szerzőtársai, akik erős korrelációt találtak (0,60) a fogyasztói bizalmi index és az általuk képzett gazdasági híreket és azok szentimentjét magában foglaló index között (Alsem et al., 2008).

3.2. HÁZTARTÁSOK INFLÁCIÓS VÁRAKOZÁSÁNAK ELŐREJELZÉSE

A háztartások inflációs várakozását képes volt előrejelezni mind az LDA topikokból épített, mind a makrogazdasági mutatókkal kiegészített modell, ugyanakkor az előrejelzés pontosnak nem mondható és a bizalmi indexnél megfigyelhető késés itt is megfigyelhető a nagy szintugrást kiváltó sokk-események környezetében. Elmondható továbbá, hogy a modellek hajlamosak túlreagálni bizonyos eseményeket és a tényértéknél jelentősen nagyobb változásokat előrejelezni.

3. ábra: a háztartások inflációs várakozása és annak előrejelzései a havi változás előrejelzésekből kumulálva

Az ábrán a háztartások inflációs várakozásának indexének idősorát láthatjuk pöttyözött vonallal, sötétebb folytonos vonallal az LDA-LASSO modell havi változásra vonatkozó előrejelzéseinek kumulált idősorát, míg világosabb folytonos vonallal a makrogazdasági mutatókkal kiegészített regresszió előrejelzéseinek kumulált idősorát. Függőleges vonalakkal jelölve az vizsgált időszak két jelentősebb elmozdulása: a Covid19-pandémia és Ukrajna orosz inváziója.



Forrás: Eurostat és saját számítások

A havi makrogazdasági mutatók az inflációs várakozás esetében sem tudták javítani az előrejelzések pontosságát, bár ebben az esetben nem is rontották azt. A széles körben alkalmazott értékelés szerint a 29%-os átlagos abszolút előrejelzési hiba (MAPE) érték az LDA-LASSO modell esetében az *ésszerű előrejelzés* minőségi kategóriába esik (Lewis, 1982). A hónapról hónapra történő elmozdulásokat azonban ebben az esetben sem sikerült előrejelezni. A szentiment-dimenzió bevonása ebben az esetben is tovább javíthatja a pontosságot, a bizalmi indexnél leírtaknak megfelelően.

5. táblázat: a háztartások inflációs várakozásának modelljeinek értékelése

A táblázatban a kizárólag LDA topikokra és a bizalmi index egygel korábbi értékére futtatott (LDA-LASSO modell) és a makrogazdasági változókra kontrollált modell (AUX-LASSO) főbb mutatóit láthatjuk. Az első oszlopokban az eredeti előrejelzés adatsorai, ahol a havi változás előrejelzése történt (DIF), míg a harmadik és negyedik oszlopokban az első oszlopokból kumulált előrejelzés található.

Mivel minden periódusra saját paraméterekkel becsültem, a pontos magyarázóváltozók egyértelműen nem azonosíthatóak, helyettük (#)-vel jelölve a modellek átlagában a LASSO által nem nulla becsült együttthatójú változók átlagos számát tüntetem fel.

	LDA-LASSO DIF	AUX-LASSO DIF	LDA-LASSO CUM	AUX-LASSO CUM
Átlagos R^2 ^a	0,21	0,22		
Tény-előrejelzés korreláció	-0,08	-0,10	0,82	0,80
MAPE (%)			29	38
MdAPE (%)			30	28
RMSE			11,15	17,20
Topikok magyarázóerővel (#)	8,76	5,60		
Gazdasági mutatók magyarázóerővel (#)		3,38		

^a Az itt szereplő R^2 értékek a bootstrapping minták teszt adatain történő illeszkedések átlagát jelentik az összes időszakra futtatott LASSO regressziók átlagában (95 periódus átlaga).

Forrás: saját számítások

Az inflációs várakozások előrejelzésének pontosításánál további lehetőség lehet az, ha nem a balansz mutatót próbáljuk előrejelezni, hanem a tényleges várt infláció értékét, amit a kérdőív tartalmaz is. Kutatási kérdésem hatókörén kívül esik, de fontos kérdés, hogy turbulens időszakokban a média hogyan tudja horgonyozni a háztartások várakozásait.

4. KONKLÚZIÓK

Összegezve eredményeim vegyes képet mutatnak: míg a fogyasztói bizalmi indexet az eljárásom nem volt képes előrejelezni elfogadható pontossággal, a háztartások inflációs várakozását igen, azonban csak magas átlagos hiba mellett. Az LDA topikok mindkét esetben előrejelző képességgel bírtak a havi makrogazdasági mutatókkal bővített modellekben is, sőt a fogyasztói bizalmi index esetében a havi gazdasági mutatók bevonása rontott az előrejelzések pontosságán. A várakozások mindkét esetben igen perzisztensnek bizonyultak, azokat hónapról hónapra történő változások csak igen kevés esetben mozdítják meg jelentősen. A várakozások hírekkel történő előrejelzésének nehézsége, hogy amennyiben azokat olyan sokkhatások alakítják, amelyek a korábbi adatokban nem szerepelnek, a várakozások elmozdulása és a sokkhatás topik lenyomata közötti összefüggés valós időben nem rendelkezésre álló információ, azt csak egy periódus késéssel használhatjuk modelljeinkben. További nehézség, hogy a vizsgált várakozás idősorokban a nagyobb elmozdulások jellemzően egy periódus ideig tartottak, miközben az azt mozgótopik tartós magas prevalenciája mellett a következő periódusban – a felfutásból tanulva – a LASSO a változás további felfutását jelzi előre. (Ne feledjük, hogy a várakozások időben egy periódussal eltolt értékeivel számoltam, így ténylegesen az előrejelzett várakozások a magyarázó változók adott értékeit követő hónapra vonatkoznak!)

Eredményeim összhangban állnak Yukhymenko (2021) eredményeivel a tekintetben, hogy első differenciára a háztartások inflációs várakozása esetében ő is igen alacsony R^2 értékeket talált.

A fogyasztói várakozások a reálgazdasággal és a híradásokkal hosszabb időszakon együtt mozognak (például Alsem et al., 2008; Malovaná et al., 2021), ám azokat jellemzően követő indikátornak tartják, elsősorban azok ragadósága miatt. Az azokban jelentkező hirtelen változások a vizsgált időszakban előre nem látható, a korábbi időszakokat leíró topikokkal teljességében nem megragadható események indukálták, amik ilyen módon valós időben nem előrejelezhetőek. Módszertanom lehetséges tovább fejlesztése a szentiment dimenziójának bevonása lehet: például még ha a Covid19-pandémia esetében hirtelen hatalmas hatással bíró téma jelenik is meg, aminek hatását a modell nem tudja a valós idő miatt megragadni, az ezt kísérő szentiment változás más topikok esetében jó előrejelzője lehet a várakozások változásának.

IRODALOMJEGYZÉK

Szakirodalom

- Adrangi, B., & Macri, J. (2011). Consumer confidence and aggregate consumption expenditures in the United States. *Review of Economics & Finance*, 1, 1-18.
- Alsem, K. J., Brakman, S., Hoogduin, L., & Kuper, G. (2008). The impact of newspapers on consumer confidence: does spin bias exist?. *Applied Economics*, 40(5), 531-539.
- Binder, C. C. (2021). Household expectations and the release of macroeconomic statistics. *Economics Letters*, 207, 110041.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Bruine de Bruin, W., Vanderklaauw, W., Downs, J. S., Fischhoff, B., Topa, G., & Arman-tier, O. (2010). Expectations of inflation: The role of demographic variables, expectation formation, and financial literacy. *Journal of Consumer Affairs*, 44(2), 381-402.
- Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of economics*, 118(1), 269-298.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 22.
- Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8), 2644-2678.
- Doms, M. E., & Morin, N. J. (2004). Consumer sentiment, the economy, and the news media. *FRB of San Francisco Working Paper*, (2004-09).
- Granát, M. P., Lehmann, K., Nagy, O., & Neszveda, G. (2023). Expect the unexpected: Did the equity markets anticipate the Russo-Ukrainian war?. *Finance Research Letters*, 58, 104301.
- Kwan, A. C., & Cotsomitis, J. A. (2006). The usefulness of consumer confidence in forecasting household spending in Canada: A national and regional analysis. *Economic Inquiry*, 44(1), 185-197.
- Larsen, V. H., Thorsrud, L. A., & Zhulanova, J. (2021). News-driven inflation expectations and information rigidities. *Journal of Monetary Economics*, 117, 507-520.
- Larsen, V. H., & Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, 210(1), 203-218.
- Larsen, V. H., & Thorsrud, L. A. (2022). Asset returns, news topics, and media effects. *The Scandinavian Journal of Economics*, 124(3), 838-868.
- Lewis, C. D. (1982). Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. Butterworth Scientific.
- Li, C. (2011). Consumer expectation and output growth: The case of China. *Economics Letters*, 113(3), 298-300.
- Malmendier, U., & Nagel, S. (2016). Learning from inflation experiences. *The Quarterly Journal of Economics*, 131(1), 53-87.
- Malovaná, S., Hodula, M., & Frait, J. (2021). What does really drive consumer confidence?. *Social Indicators Research*, 155(3), 885-913.
- Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about inflation expectations. *NBER macroeconomics annual*, 18, 209-248.
- Médiapiaci jelentés 2023. (2023). Nemzeti Média- és Hírközlési Hatóság.
- Nimark, K. P., & Pitschner, S. (2019). News media and delegated information choice. *Journal of Economic Theory*, 181, 160-196.

- Rieger, J., Jentsch, C., & Rahnenführer, J. (2021, November). RollingLDA: An update algorithm of Latent Dirichlet Allocation to construct consistent time series from textual data. In *Findings of the Association for Computational Linguistics: EMNLP 2021* (pp. 2337-2347).
- Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. " O'Reilly Media, Inc."
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288.
- The joint harmonised EU programme of business and consumer surveys. (2024). European Commission, Directorate-General for Economic and Financial Affairs.
- Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, 38(2), 393-409.
- Ulrike, M., & Juan, O. (2019). Exposure to Daily Price Changes and Inflation Expectations. *NBER working paper*, 26237.
- Van Dalen, A., de Vreese, C., & Albæk, E. (2017). Economic news through the magnifying glass: How the media cover economic boom and bust. *Journalism Studies*, 18(7), 890-909.
- Van Raaij, W. F. (1989). Economic news, expectations and macro-economic behaviour. *Journal of Economic Psychology*, 10(4), 473-493.
- Yukhymenko, T. (2021). The role of the media in the inflation expectation formation process. *Visnyk Nacional'noho Banku Ukraïny*, (253), 4-26.

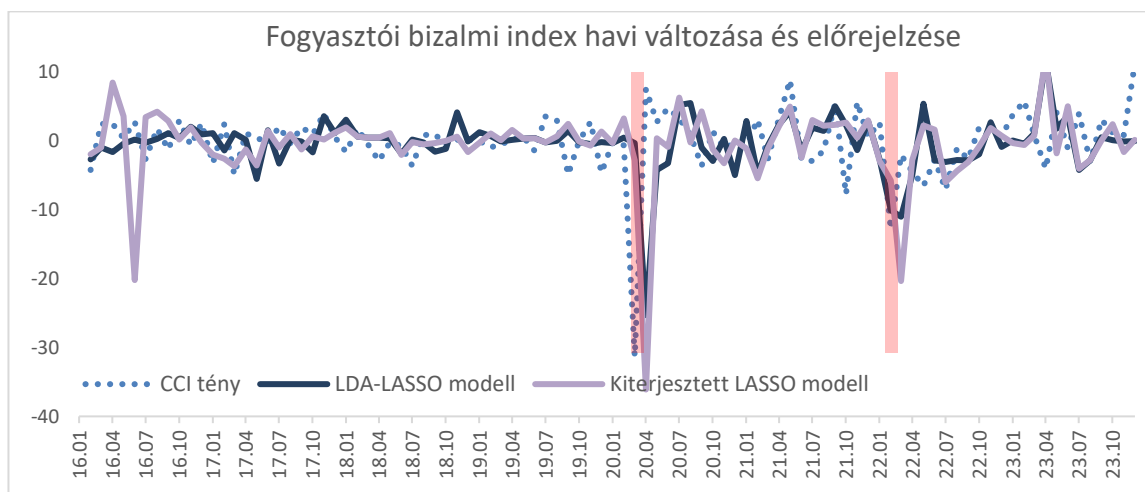
MELLÉKLETEK

1. MELLÉKLET: VÁRAKOZÁSOK HAVI VÁLTOZÁSÁNAK IDŐSORAI

Az alábbi ábrákon a gördülő előrejelzés havi változásokra vonatkozó értékeit és a tényszámok láthatóak. Az *Eredmények* fejezetben található ábrák ezek összegzésével álltak elő.

4. ábra: a fogyasztói bizalmi index havi változása és annak előrejelzései

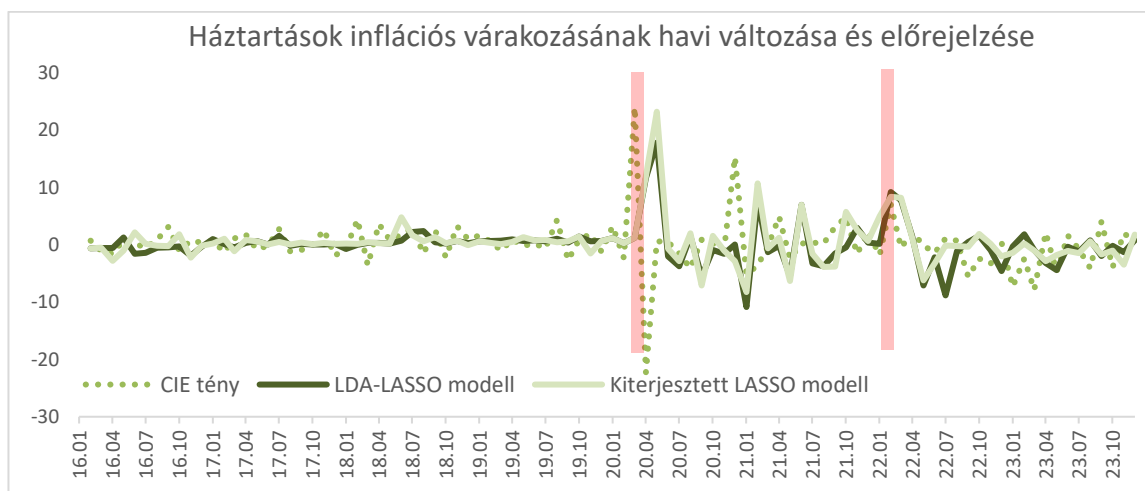
Az ábrán a fogyasztói bizalmi index havi eltéréseit (CCI) láthatjuk a megelőző hónaphoz hasonlítva pöttyözött vonallal, sötétebb folytonos vonallal az LDA-LASSO modell előrejelzéseit, míg világosabb folytonos vonallal a makrogazdasági mutatókkal kiegészített regresszió előrejelzéseit. Függőleges vonalakkal jelölve a vizsgált időszak két jelentősebb elmozdulása: a Covid19-pandémia és Ukrajna orosz inváziója.



Forrás: Eurostat és saját számítások

5. ábra: a háztartások inflációs várakozásának havi változása és annak előrejelzései

Az ábrán a háztartások inflációs várakozásának indexének eltéréseit láthatjuk a megelőző hónaphoz hasonlítva pöttyözött vonallal, sötétebb folytonos vonallal az LDA-LASSO modell előrejelzéseit, míg világosabb folytonos vonallal a makrogazdasági mutatókkal kiegészített regresszió előrejelzéseit. Függőleges vonalakkal jelölve a vizsgált időszak két jelentősebb elmozdulása: a Covid19-pandémia és Ukrajna orosz inváziója.



Forrás: Eurostat és saját számítások