

OTDK-dolgozat

2025

Neumann János Egyetem

**A KÖRNYEZETI KÉRDÉSEKRE IRÁNYULÓ FIGYELEM ÉS A
ZÖLD BEFEKTETÉSEK KAPCSOLATA RENDELLEN IDŐJÁRÁSI
KÖRÜLMÉNYEK ESETÉN**

**THE RELATIONSHIP BETWEEN ENVIRONMENTAL
AWARENESS AND GREEN INVESTMENTS UNDER ABNORMAL
WEATHER CONDITIONS**

KANYÁRI BENCE PÉTER
NEMZETKÖZI GAZDASÁG ÉS GAZDÁLKODÁS MSC

Konzulens: Granát Marcell

Kézirat lezárásának dátuma: 2024.05.17.

Tartalomjegyzék

1.	BEVEZETÉS	1
2.	ADATOK	5
2.1.	HÍRADATOK	5
2.2.	GOOGLE KERESÉSEK ADATAI	6
2.3.	IDŐJÁRÁSI ADATOK	7
2.4.	RÉSZVÉNY- ÉS VÁLLALATIADATOK	9
2.5.	IPARÁGI ESG ADATOK	10
3.	MÓDSZERTAN	11
3.1.	LDA TÉMAMODELLEZÉS	11
3.2.	ZÖLD ÉS BARNA PORTFÓLIÓK KIALAKÍTÁSA	12
3.3.	VEKTOR-AUTOREGRESSZÍV MODELL	14
4.	EREDMÉNYEK	17
4.1.	Az LDA TÉMAMODELL EREDMÉNYEI	17
4.2.	A VAR-MODELL EREDMÉNYEI	18
5.	KONKLÚZIÓ	24
6.	IRODALOMJEGYZÉK	26
	MELLÉKLETEK	I

1. BEVEZETÉS

Az elmúlt években, a fenntartható növekedést célzó befektetések egyre nagyobb népszerűségnek örvendenek. A fenntartható növekedést célzó befektetések nem csak pénzügyi, hanem környezeti, társadalmi és irányítási (Environmental, Social, Governance, ESG) szempontokat is figyelembe vesznek, melyek közül a környezeti szempont a legjelentősebb. A környezetbarátnak tartott befektetéseket gyakran nevezik „zöld” befektetésekknek, míg a „barna” befektetéseket ezen ellenkezőjeként tartják számon (Pástor és mtsai., 2022).

Korábbi kutatások eredményei azt mutatják, hogy a zöld részvények jellemzően jobban teljesítenek, mint a barnák, amikor az éghajlatváltozás aggodalmi és a környezeti témák népszerűsége növekszik. Pástor és mtsai. (2021) a fenntarthatóságnak az eszközárakra gyakorolt hatását modellezték és arra következtetésre jutottak, hogy a zöld eszközök – így a részvények is – jobban teljesítenek akkor, ha az éghajlatváltozással kapcsolatos aggodalmak növekednek. Ardia és mtsai. (2020) az S&P500 indexben szereplő vállalatok napi részvényárfolyamainak az elemzésével empirikusan is igazolják a Pástor és mtsai. (2021) által felvázolt elméleti eredményeket a 2010 januárja és 2018 júniusa közötti időszakban. Csillag J. és mtsai. (2022) tanulmányukban a NASDAQ tőzsdén 2010 és 2020 között kereskedett részvényeket vizsgálták és arra következtetésre jutottak, hogy az ESG pontszámok jelentősen befolyásolják a jövőbeli hozamokat akkor, amikor a médiában gyakrabban jelennek meg a környezetvédelmi kérdésekkel kapcsolatos hírek. Choi és mtsai. (2020) megmutatják, hogy a lakossági befektetők a környezetükben tapasztalt rendellenes időjárás esetén megválnak a magasabb szén-dioxid kibocsátással rendelkező vállalatok részvényeitől. Továbbá megállapítják, hogy az emberek módosítják az éghajlatváltozással kapcsolatos megítélésüket akkor, ha a szokásosnál melegebb hőmérsékletet tapasztalnak. Mindebből arra következtethetünk, hogy a befektetők éghajlatváltozással kapcsolatos aggodalmi rendellenes időjárási körülmények esetén is nőhetnek, amik szintén befolyásolhatják a fenntartható fejlődést biztosító befektetésekkel kapcsolatos nézeteiket.

Ha a befektetők több hírt olvasnak a fenntarthatósági és környezeti témával kapcsolatban, vagy szélsőségesebb időjárást tapasztalnak akkor az megváltoztathatja az ezzel kapcsolatos nézeteiket, és ennek függvényében dönthetnek befektetéseikről. Az emberek – és így a befektetők – korlátozott figyelmük miatt nagyobb valószínűséggel figyelnek oda a környezetvédelmi és az éghajlatváltozással kapcsolatos problémákra akkor, ha gyakrabban szembesülnek velük (Hirshleifer & Teoh, 2003). A személyes tapasztalatok, mint például a rendellenes időjárás, is sokat számítanak és növelik a globális felmelegedéssel kapcsolatos

aggodalmakat (Li és mtsai., 2011). Ráadásul a szélsőségesen meleg években, a médiában megnő az éghajlatváltozással kapcsolatos problémák iránti figyelem (Schmidt, 2015), ami azt jelenti, hogy a szélsőséges időjárás hatása a médián keresztül fel is erősödhet. Emiatt, a környezeti témájú hírek médiamegjelenése lehet, hogy csak közvetve befolyásolja a befektetők fenntarthatósággal kapcsolatos felfogását és ezáltal a zöld befektetések hozamát.

Jelen tanulmány, a fentebb említett szakirodalmi tudást, azzal egészíti ki, hogy a „zöld” és a „barna” részvények teljesítményének a vizsgálatok egyaránt figyelembe veszem a környezeti problémákra irányuló figyelem és a szokásostól eltérő időjárási körülmények hatását. Korábbi kutatások eddig, vagy csak az egyik tényező hatását vizsgálták a fenntartható befektetésekre, vagy csak a két említett tényező egymásra gyakorolt hatását elemezték. A kétféle hatás együttes vizsgálata pedig lehetőséget ad arra, hogy elkülönítsem ezen tényezők közvetlen hatását. Ráadásul a környezeti problémákra irányuló figyelem meghatározásánál a szakirodalomban alkalmazott mindkét módszert, a hírekben előforduló környezettel és az éghajlatváltozással kapcsolatos tematikájú cikkek, valamint az ilyen tematikájú Google keresések volumenét is figyelembe veszem.

Tanulmányomban tehát azt elemzem, hogyan alakul a „zöld” és a „barna” vállalatok részvényeiből álló portfóliók hozamkülönbözete akkor, amikor a környezeti téma és az ehhez kapcsolódó tematikájú hírek médiamegjelenése növekszik, valamint a befektetők a szokásostól eltérő időjárási körülményeket tapasztalnak.

A fentiek alapján, azzal a hipotézissel élek, hogy azokon a napokon, amikor (1) nagyobb a környezeti témájú hírekre irányuló médiafigyelem, valamint (2) a szokásostól sokkal inkább eltérőek az időjárási körülményeket, megnő a „zöld” és a „barna” részvényekből álló portfóliók hozamkülönbözete.

Az LDA témamodellezés segítségével a Wall Street Journal¹ (WSJ) hírportálon közzétett híreken a környezeti témájú hírek megjelenésének a gyakoriságát vizsgálom. A modell segítségével azonosítok egy környezeti témát, majd meghatározom ezen témájú híreknek az adott napon közzétett összes hírhez viszonyított relatív gyakoriságát. Ez a relatív gyakoriság pedig a vizsgált időszak minden napjára megadja a környezeti téma előfordulásának az arányát a médiában. A WSJ választását alapvetően az indokolja, hogy a pénzügyi piacokon a legfontosabb médiumok közé tartozik, valamint a hírportál minőségi és releváns tartalmakkal szolgál környezeti problémák és az éghajlatváltozás miatt aggódó befektetők számára (Engle és mtsai., 2020). Da és mtsai. (2011) azonban megemlíti, hogy a WSJ-ban megjelent

¹ <https://www.wsj.com>

környezeti témájú újságcikkek, csak akkor garantálják a befektetők nagyobb figyelmét az adott téma – esetünkben a környezeti téma – iránt, ha ténylegesen el is olvassák az adott cikket. Emiatt, az említett szerzők által is javasolt, a fenntartható befektetésekre, valamint az éghajlatváltozásra irányuló Google keresések volumenét is felhasználom az elemzésen során, mint a környezettel kapcsolatos témákra irányuló figyelem proxyja.

Tanulmányomban az S&P500 indexben szereplő jelenlegi vállalatokat elemzem a 2010. január 1. és 2022 december 31. közötti időszakban. A zöld és a barna portfólió kialakítása során a Refinitiv (Eikon, 2022) iparági ESG pontszámait használom. Az iparági környezeti pillér pontszámok, alapján sorbarendezelem a vállalatokat, majd az alacsony (magas) pontszámmal rendelkező vállalatokból létrehozok egy zöld (barna) portfóliót.

Elemzésem során egy több változós vektor-autoregresszív modellt alkalmazok. Ennek segítségével megvizsgálom, hogy a médiában megjelenő, környezeti témával foglalkozó cikkek aránya és a környezeti kifejezésekre történő Google keresési volumenadatok miképpen hatnak a zöld és a barna portfóliók hozamkülönbségére. A modelleket lefuttatom a szokásostól eltérő időjárási körülményeket tartalmazó változókkal is – melyeket exogén változóként szerepeltetek a modellben – és megvizsgálom, hogy ezen változók szerepeltetése mennyire befolyásolja modell eredményeit és a következtetéseket. Az időjárást leíró változókat, azért exogén változókként szerepeltetem, mivel feltételezem, hogy az időjárásra a modell többi változója nincs hatással és ezen megkötést alkalmazva pontosabb eredményeket kaphatok. Az egyes modelleket lefuttatom különböző időszakokra, annak érdekében, hogy megvizsgáljam az eredmények robusztusságát. Ezen túlmenően egy periódusos előrejelzéseket is készítek az egyes modellek esetében és megvizsgálom, hogy a rendellenes időjárási körülményeket leíró változók szerepeltetése a modellben szignifikánsan javítja-e ezen előrejelzéseket.

Az eredmények azt mutatják, hogy a médiában (WSJ hírportálon) megjelenő környezeti tematikájú cikkek arányának növekedése, növeli a zöld és a barna portfóliók hozamkülönbségét, valamint egyes időszakokban Granger-okozza is azt. A környezettel kapcsolatos kifejezések Google keresési volumenének növekedése ugyanakkor csökkenti a zöld és a barna portfóliók hozamkülönbségét és Granger-okság ezen változók között nem áll fenn. Az eredmények alapján megállapítható az is, hogy a szokásostól eltérő időjárási körülményeket leíró változók szerepeltetése a modellben érdemben nem változtatja meg a kapott eredményeket. Az előrejelzések elemzéséből azt a következtetést vonhatjuk le, hogy ugyan a rendellenes időjárást leíró változók szerepeltetése esetén a vizsgált mintán jobb előrejelzéseket kapunk, viszont a két modell előrejelzései közötti különbség nem szignifikáns.

Jelen tanulmány a következőképpen épül fel. A 2. fejezetben ismertetem az elemzésem során felhasznált adatokat, az egyes adattisztítási lépéseket, valamint az adatokból létrehozott különböző változókat. A 3. fejezetben bemutatom az LDA témamodellezésnek és a vektor-autoregresszív modelleknek a módszertanát, valamint a zöld és a barna portfólióknak a kialakítását. A 4. fejezetben tárgyalom az eredményeket, majd végezetül az 5. fejezetben összefoglalom a következtetéseket és azok értelmezésének a korlátait.

2. ADATOK

Ebben a fejezetben ismertetem az elemzésem során használt adatbázisokat. Először a Wall Street Journal hírportálon megjelenő hírek, majd a Google keresések adatait mutatom be. Ezt követően rátérek az időjárási adatokra, végezetül pedig a részvényadatok ismertetésére.

2.1. HÍRADATOK

A médiában megjelenő környezeti (zöld) témájú hírek gyakoriságának a vizsgálatához, szükségünk van megfelelő szöveges adatokra. Ennek érdekében, elemzésem során a Wall Street Journal hírportálon megjelenő, nyilvánosan elérhető hírek törzsszövegeinek a szöveges adatait használom.

A WSJ hírportál online hírarchívuma 1998-tól kezdődően, az év minden napjára vonatkozóan tartalmazza az adott napon megjelenő híreket. A WSJ kifejezetten az üzleti és pénzügyi világot, valamint a gazdaságot érintő problémákkal, kérdésekkel, újdonságokkal foglalkozik és napi szinten beszámol a tőkepiacokat érintő fejleményekről, illetve globális tényezőkről. Annak ellenére, hogy a hírarchívum 1998-tól kezdődően elérhető, csupán 2010. január 1-től kezdődően vizsgálom a benne megjelenő szöveges adatokat. Összesen 695.406 cikk szöveges adatait elemzem. A havonta elérhető hírcikkek száma 2010 és 2017 között átlagosan meghaladta az 5000-et, ezt követően viszont jelentősen csökkent (lásd az 1. ábrát). Ennek ellenére, még így is bőven elegendő számú dokumentum áll rendelkezésre az elemzéshez, hiszen nincs olyan hónap, ahol a hírcikk száma ne haladná meg a 2500-at.



1. ábra: A WSJ hírportálról letöltött cikkek havi számának alakulása

Forrás: saját szerkesztés

2.2. GOOGLE KERESÉSEK ADATAI

A környezeti témákra irányuló médiafigyelem meghatározására, a WSJ hírportálon megjelenő zöld témájú hírek gyakorisága mellett, a Google Trends² adatait is felhasználom.

A Google Trends lényegében egy index, amely a Google felhasználóknak egy adott témára vagy kifejezésre történő kereséseinek a relatív mennyiségét adja meg egy adott időszakban. Az indexet, a Google kereséseknek egy véletlen és torzítatlan mintájából számítják, majd egy 0-tól 100-ig terjedő skálán jelenítik meg, ahol a 100-as érték az adott kifejezésre történő keresések volumenének a maximumát jelenti adott időszak és lokáció esetén. A keresési volumen számítása során nem a keresések mennyiségét veszik figyelembe, hanem az adott témára, kifejezésre történő keresések relatív arányát az összes keresésen belül.

Elemzésem során a környezeti témával kapcsolatos kifejezésekre történő Google keresések volumenét használok fel a befektetők fenntarthatóság és környezettudatosság iránti érdeklődésének a meghatározására. A környezettel kapcsolatos kifejezéseket Kvam és mtsai. (2022) tanulmánya alapján választottam ki. Ezen kifejezések között szerepel például a klímaváltozás (Climate change) a globális felmelegedés (Global warming), vagy a megújuló energia (Renewable energy). A kifejezések teljes listája a mellékben (M1) találhatóak.

A Google keresési volumen adatokat 2010. január 1-től kezdődően töltöttem le, egészen az általam vizsgált időszak végéig, azaz 2022. december 31-ig. Fontos megjegyezni, hogy a Google által szolgáltatott adatok gyakorisága függ a vizsgált időszak hosszától. Az elemzésem szempontjából lényeges, napi gyakoriságú adatok, nagyjából csak egy 8 hónapos hosszúságú időszak esetén állnak rendelkezésre. Ez abból a szempontból okoz problémát, hogy a fentebb említett számítási módszer miatt, két eltérő időszakból származó nap adatainak az összehasonlítása közvetlenül nem lehetséges.

A Google keresési adatok aggregálása során elsőnek létrehozok egy, a napi ingadozásokat és a hosszútávú trendeket egyaránt megragadó változót. Wanidwaranan és Padungsaksawasdi, (2022) által használt módszert követve először 2 hónapos blokkokban letöltöttem a napi, valamint a vizsgált időszak teljes egészére a havi keresési volumen adatokat minden környezettel kapcsolatos kifejezés esetén. Ezt követően a napi keresési volumen adatokat kiigazítottam a havi keresési volumen adatokkal. Ezen eljárás a napi keresési gyakoriságok relatív értékét adja meg a teljes vizsgált időszakra. A havi adatokkal kiigazított gyakorisági index (SVI) értékét az alábbi képlet alapján számoltam ki minden keresési kifejezésre.

² <https://trends.google.com>

$$SVI_i = \frac{\text{napi keresési volumen adatok} \times \text{havi keresési volumen adatok}}{100}. (1)$$

Az egyes környezettel kapcsolatos kifejezések SVI értékeinek az álagaként meghatároztam az átlagos kiigazított gyakorisági indexet (MSVI):

$$MSVI_t = \sum SVI_{it}, (2)$$

ahol SVI_t az i-edik kifejezés SVI értéke a t-edik napon.

Az átlagos kiigazított gyakorisági index értékének a meghatározását követően Da és mtsai. (2011) alapján, létrehoztam egy abnormális gyakorisági indexet (ASVI). Ezen index érték következő módon határoztam meg:

$$ASVI_t = \log(MSVI_t) - \log[Med(MSVI_{t-1}, MSVI_{t-8}, \dots, MSVI_{t-70})], (3)$$

ahol $\log(MSVI_t)$ az MSVI értékének a logaritmus a t-edik napon, $\log[Med(MSVI_{t-1}, MSVI_{t-8}, \dots, MSVI_{t-70})]$, pedig az adott napot megelőző 7 hét azonos napjain megfigyelhető MSVI értékek mediánjának a logaritmus. Ha az $MSVI_t$ egy adott t időpontban nulla, akkor a $\log(MSVI_t)$ értéket szintén nullának veszem. Az abnormális gyakorisági index számítása mögött az a logika van, hogy egy hosszabb időablak mediánja a figyelem normális szintjét ragadja meg oly módon, hogy robosztus marad a közelmúltbeli kiugrásokra (Da és mtsai., 2011). Az AVSI számításának további előnye, hogy kiszűri a trendeket és a szezonalitást az idősorból. Az elemzésem további részében ezen abnormális gyakorisági indexet használom.

2.3. IDŐJÁRÁSI ADATOK

Ahhoz, hogy megállapítsuk, hogy a befektetők mikor tapasztalnak szélsőséges időjárási körülményeket, szükségünk van olyan változókra, melyekkel mérni tudjuk ezeket. Ennek érdekében, elemzésem során a napi csapadékmennyiséget és a napi maximumhőmérsékletet használom a szélsőséges időjárás meghatározására. Továbbá, számításaim során, az időjárási változók gazdaságra gyakorolt hatásainak a vizsgálatánál alkalmazott módszerekhez (Dell és mtsai., 2012; Mendelsohn és mtsai., 2006) hasonlóan, figyelembe veszem, hogy az adott területen megfigyelt időjárási körülmények a népesség hány százalékát érintették.

A napi maximumhőmérséklet, valamint csapadékmennyiség meghatározására a Parameter-elevation Relationships on Independent Slopes Model (PRISM)³ adatbázis adatait használom. Az adatbázist közzé tevő PRISM Climate Group, meteorológiai állomások széles köréből gyűjt adatokat, ellenőrzi ezek minőségét és különböző modellek segítségével (Daly és

³ <https://prism.nacse.org>

mtsai., 2008), az USA egész területére vonatkozóan közöl különböző felbontású és gyakoriságú adatokat (Daly, 2013). Az általam használt konkrét adatbázis 1981-től kezdődően tartalmaz napi gyakoriságú időjárási változókat – mint a napi csapadékmennyiség, vagy maximum hőmérséklet – kb. 4 kilométeres (2,5 ívperces) rácsfelbontásban. Az adatokat az elemzésemben szereplő többi változóhoz hasonlóan a 2010. január 1. és 2022 december 31. közötti időszakra töltöttem le. Ugyanezen adatbázis tartalmaz továbbá, havi normál értékeket is az egyes változóra vonatkozóan. A normál értékek az 1990 és 2020 közötti 30 éves időszak adatai alapján lettek meghatározva (Daly, 2013). A napi csapadékmennyiség és a napi maximumhőmérséklet esetén ezen havi normál értékeket is letöltöttem, melyeket az elemzésem további részében felhasználok.

A népesség adatokhoz a 2010-es amerikai népszámlálás eredményei alapján a Columbia University Center for International Earth Science Information Network (CIESIN)⁴ által készített kb. 1 kilométer (0,5 ív másodperces) felbontású adatokat használtam. Az adatbázis a népességre, háztartásokra és lakhatásra vonatkozó változókat is tartalmaz (CIESIN, 2017), én viszont csak az elemzésemhez szükséges, a teljes népességre vonatkozó adatokat töltöttem le.

A népesség adatokat, valamint az időjárást leíró változókat tartalmazó adatbázisok felbontása eltérő. Annak érdekében, hogy ez ne okozzon problémát a kisebb felbontású – népesség adatokat tartalmazó – adatbázis adatait aggregáltam. Ennek következtében már a népesség adatokat adatbázis felbontása is kb. 4 kilométeres (2,5 ívperces) rácsfelbontású. Problémát jelent ugyanakkor, hogy a két adatbázis koordináta rendszere eltérő. Az időjárásra vonatkozó adatokat tartalmazó adatbázis a North American Datum of 1983 (NAD 83), míg a népesség adatokat tartalmazó adatbázis a The World Geodetic System 1984 (WGS 84) koordináta rendszert használja. A két koordináta rendszer azonban szinte azonosnak tekinthető, hiszen a kettő közötti eltérés nagyjából 2 méter (White, 2013). Mivel ez a 2 méter nem jelentős eltérés egy kb. 4 kilométeres felbontás esetén, így a két koordinátarendszert az elemzésem további részében azonosnak tekintem.

Mind a két időjárási változó – a napi csapadékmennyiség és a napi maximumhőmérséklet – esetében, minden egyes napra kiszámoltam az adott változó népességszámmal súlyozott átlagát. Ennek érdekében az időjárási változókat tartalmazó adatbázis minden egyes rácsponthoz hozzárendelem a népesség adatokat tartalmazó adatbázis legközelebb eső

⁴ Center for International Earth Science Information Network (CIESIN), Columbia University. 2017. U.S. Census Grids (Summary File 1), 2010. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://doi.org/10.7927/H40Z716C>

rácspontját, majd kiszámolom ezen változók súlyozott átlagát. Eredményként megkapom a napi csapadékmennyiség és a napi maximumhőmérséklet népességszámmal súlyozott átlagát az USA területére vonatkozóan. Ugyanezen számításokat elvégeztem a két vizsgált változó havi normál értékeire is.

A szélsőségek megragadására, az egyes változóknak az adott havi normál értékétől vett eltérését használok. Ennek következtében két olyan változót kapok, amely azt adja meg, hogy egy adott napi maximumhőmérséklet, illetve csapadékmennyiség, mennyivel tért el a normális értékétől. Az alkalmazott módszer további előnye, hogy eltávolítja a szezonális ingadozásokat az eredeti idősorból.

2.4. RÉSZVÉNY- ÉS VÁLLALATIADATOK

Az S&P500 részvényindexben szereplő vállalatok részvényeinek a napi adatait használok az elemzésem során. Az adatokat a Refinitiv Datastream adatbázisából töltöttem le a 2010. január 1. és 2022. december 31. közötti időszakra. A vizsgált időszak minden egyes kereskedési napjára letöltöttem az elemzésemben felhasznált változók értékeit. Ezen változók a Teljes hozam index (Total Return Index, TOTRET) és a piaci érték (Market Value, MV). A Teljes hozam index egy olyan megtérülési mutatószám, amely minden olyan tényező hatását tartalmazza, amely befolyásolja a részvény hozamát, beleértve az árfolyamváltozást, a részvényfelaprózódást és az osztalékfizetést. A piaci érték pedig az adott részvény piaci kapitalizációját adja meg egy adott időpontban. A Teljes hozam index alapján meghatároztam az adott részvény napi hozamát is. Az i -edik részvény hozama a t -edik időpontban:

$$\text{RETURN}_{it} = \frac{\text{TOTRET}_{it}}{\text{TOTRET}_{it-1}} - 1, (4)$$

ahol TOTRET_{it} i -edik részvény Teljes hozam indexe a t -edik időpontban.

Az adattisztítás során töröltem azon nap megfigyeléseit, amelyik napon az összes részvény hozama nulla volt, hiszen ez azt jelenti, hogy ezeken a napokon nem kereskedtek az amerikai tőzsdéken. Ilyen eset például akkor fordulhat elő, ha az adott nap valamilyen ünnepnapra, vagyis munkaszüneti napra esik és az adatbázis mégis tartalmaz ezen napra vonatkozó adatokat. Összesen 179 ilyen nap volt az adatbázisban és ezen lépés következtében 57.065 megfigyelést töröltem.

Az adattisztítást követően az adatbázis összesen 503 vállalatra vonatkozóan 1.541.622 megfigyelést tartalmazott, amely a 2010. január 1. és 2022. december 31. közötti 3212 kereskedési nap adatait foglalja magában. Az 503 vállalatból 439 esetben, az adatbázis a teljes vizsgált időszak minden kereskedési napjára vonatkozóan tartalmaz megfigyeléseket. A

maradék 64 vállalatra vonatkozó adatok azonban nem fedik le a teljes vizsgált időszakot. Ennek az az oka, hogy ezen vállalatok 2010. január 1-ét követően léptek a tőzsdére. Ezen vállalatok esetében, a tőzsdére lépés időpontjától kezdődően, az adatbázis összes kereskedési napjára vonatkozóan vannak megfigyelések, egészen a vizsgált időszak végéig, megszakítás nélkül.

2.5. IPARÁGI ESG ADATOK

A zöld és a barna portfóliók kialakítására a Refinitiv által alkalmazott, az ESG pontszámok számítása során használt, úgynevezett iparági lényegességi mátrix (materiality matrix) adatait használom. Ezen mátrix egy 1-től 10-ig terjedő skálán határozza meg, hogy az adott iparág esetében az egyes ESG-hez kapcsolódó kategóriák mennyire lényegesek. Az összesen 10 kategória közül én csak 3-at, a környezeti pillérbe (Environmental pillar) tartozó kategóriákat használom. Ezen 3 kategória a kibocsátás (Emission), az innováció (Innovation) és az erőforrás-felhasználás (Resource use). Az egyes kategóriákat több tényezőkre osztják, majd ezen tényezők proxyjaként használt változók iparágakkénti mediánértékét használják a lényegességi mátrix meghatározása során (Eikon, 2022).

Az egyes iparágakhoz tartozó környezeti pillér pontszámot, az ezen pillér kategória pontszámainak az átlagaként határoztam meg. A környezeti pillér pontszám ennek következtében egy 1-től 10-ig terjedő skálán mutatja, hogy a környezeti szempont mennyire lényeges az egyes iparágakban. Minél nagyobb ezen pontszám értéke, az adott iparág annál jobban veszélyezteteti a környezetet. A magas környezeti pillér pontszámú iparágakba ezáltal a „barna” részvények, az alacsony pontszámú iparágakba pedig a „zöld” részvények tartoznak.

3. MÓDSZERTAN

3.1. LDA TÉMAMODELLEZÉS

A témamodellzés egy nem felügyelt gépi tanulási módszer, amely során az algoritmus a szavak előfordulási mintázata alapján hoz létre különböző témákat, melyekhez hozzárendeli a dokumentumok szövegeit (Grimmer & Stewart, 2013). Elemzésem során az egyik legelterjedtebb algoritmust a látens Dirichlet-allokációt (Latent Dirichlet Allocation, LDA) használom. Ezen algoritmus koncepciója azon a feltételezésen alapszik, hogy minden téma szavak valószínűségi keveréke és minden dokumentum pedig témák valószínűségi keveréke. Minden téma tehát, valamilyen arányban előfordul minden dokumentumban és a témák a szavak valamilyen keverékei (Blei és mtsai., 2003).

Az LDA egy olyan statisztikai módszer, amely automatikusan megtalálja az egymáshoz kapcsolódó szavakat, melyeket témákba sorol és megbecsüli az egyes témák arányát a dokumentumokban, esetünkben a hírcikkekben. A modell eredményül megadja, hogy az egyes szavak milyen valószínűséggel származnak egy adott témából (β), valamint, hogy az egyes témák milyen arányban fordulnak elő a hírcikkekben (γ).

A modell egyik hangolható hiperparamétere a témák száma. A témák számának pontos meghatározása nem egyértelmű. Túl kevés téma esetén az eredmények értelmezése nehezzé válik, hiszen túl átfogó eredményeket ad. A túlságosan sok téma sem előnyös, mivel a kelleténél jobban klasztereződnek a témák, így nagymértékben hasonlítani fognak egymásra (Greene és mtsai., 2014). Az optimális témaszám meghatározása az LDA modellek esetében általában a perplexitás vagy a kizárólagosság és a szemantikai kohézió alapján történik (Bischof & Airolidi, 2012; Jacobi és mtsai., 2018). Elemzésemben viszont, mivel kiemelten a környezeti témára fókuszálok, a Csillag J és mtsai. (2022) által használt keretrendszert alkalmazom.

A modellt megbecsülöm több, eltérő témaszámmal, és azt a legkisebb témaszámmal rendelkező modellt választom, ahol elsőként, egyértelműen megjelenik a környezeti téma. Csillag J és mtsai. (2022) módszeréhez hasonlóan a modellt 2-től 30-ig minden páros témaszám mellett megbecsülöm. A témaszám megválasztására azért ezt a módszert alkalmazom, mert ha nem elégséges témaszámú modellt használok, akkor nem jelenik meg a meg környezeti téma. A túl sok témát tartalmazó modell alkalmazása pedig felesleges, hiszen előfordulhat, hogy több környezeti téma is megjelenik, melyek megfelelő aggregálása további nehézségeket jelentene.

Az LDA témamodell nem közvetlenül a dokumentumokat elemzi, hanem egy ezekből létrehozott, úgynevezett dokumentum-kifejezés mátrixot. Ez egy olyan mátrix, amely megadja, hogy az egyes dokumentumokban egy adott kifejezés (szó), milyen gyakran jelenik meg. A

dokumentum-kifejezés mátrix létrehozásához alkotnunk kell először egy korpuszt a vizsgált dokumentumokból. Ezen lépésnél Roberts és mtsai. (2019) keretrendszerét alkalmaztam. Eltávolítottam a korpuszból a számokat, az írásjeleket, a leggyakrabban használt kifejezéseket és szótövesítést végeztem. A kapott korpuszból létrehoztam a dokumentum-kifejezés mátrixot. Mivel az LDA modell alkalmazása az így kapott korpusz – és az abból létrehozott dokumentum-kifejezés mátrix – összes szavára számításigényes és okszerűtlen lenne (Jacobi és mtsai., 2018), tovább szűkítettem a mátrixot. Griffiths & Steyvers (2004), Grün & Hornik (2011), és Jacobi és mtsai. (2018) javaslatára elhagytam a nagyon ritka, maximum 100-szor előforduló kifejezéseket, valamint a nagyon gyakori, a dokumentumoknak legalább a 25 százalékában szereplő kifejezéseket.

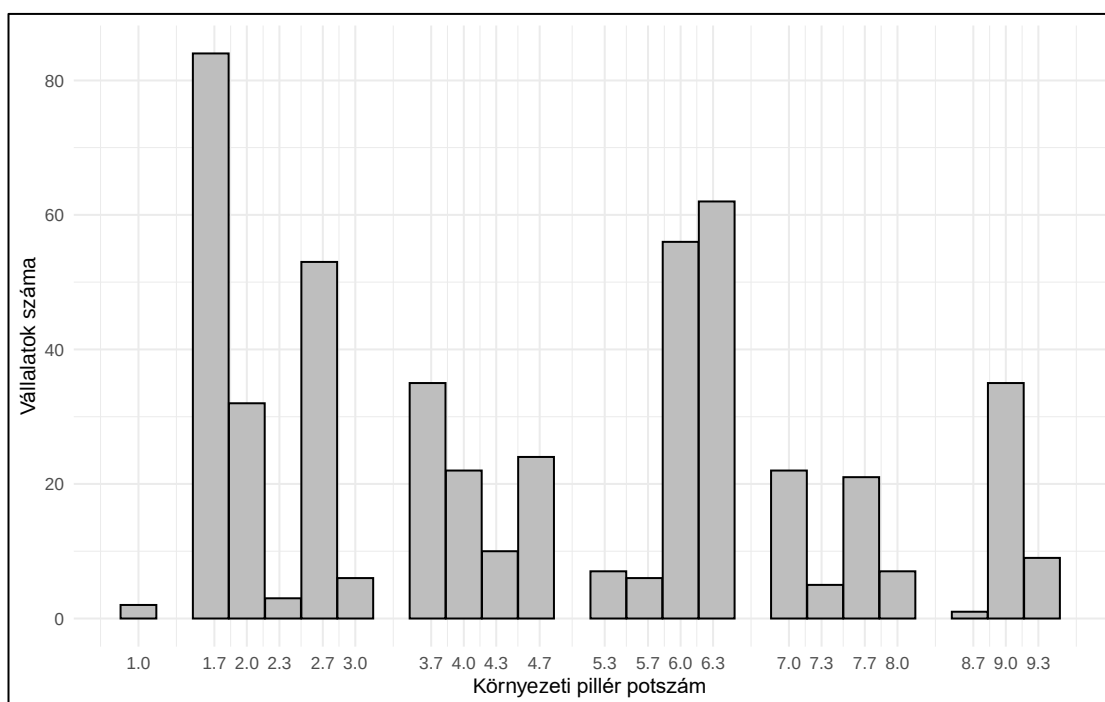
A témák azonosításához azokat a szavakat használom, amelyek a legnagyobb valószínűséggel származnak az adott témából. Ez gyakorlatilag azt jelenti, hogy témaként, megnézzük a legnagyobb β -val rendelkező szavakat és ez alapján határozzuk meg az adott témát. A témák azonosítása után megvizsgálom, hogy a modell által becsült γ értékek alapján a környezeti téma milyen arányban fordul elő az egyes dokumentumokban. Ezt követően, a környezeti téma γ értékeinek veszem a naponkénti súlyozott átlagát, ahol a súlyok az adott napon megjelenő hírcikkekben szereplő szavaknak a száma. Az így kapott idősort használom a továbbiakban a hírekben megjelenő környezeti téma gyakoriságának a proxyjaként.

Az elemzésem során csupán a 2010. január 1. és 2012. december 31. közötti hírekre futtatom le a témamodellt. Ennek az az oka, hogy a modell futtatása ilyen sok dokumentum esetén rendkívül időigényes. A 2012 utáni hírek esetében az azt megelőző időszakra futtatott modell eredményeit felhasználva becsülöm meg a környezeti téma előfordulásának az arányát az egyes dokumentumokban.

3.2. ZÖLD ÉS BARNA PORTFÓLIÓK KIALAKÍTÁSA

A zöld és a barna részvényportfólió kialakítása során az iparági környezeti pillér pontszámokat használom. A zöld portfólióba az iparágak vállalati kerültek, melyek környezeti pillér kisebb, mint 2, míg a barna portfólióba a 7-esnél nagyobb pontszámmal rendelkező vállalatok kerültek. Általában az egy változó szerint kialakított portfólióknál azt elvet alkalmazzák, hogy az egyes vállalatokat az adott változó szerint sorba rendezik, majd egyenlő nagyságú csoportokra osztják és az első és az utolsó csoportot vizsgálják. Elemzésemben azért nem ezt a módszert alkalmazom, mert az általam vizsgált vállalatok környezeti pillér pontszáma nem kellően heterogén ahhoz, hogy a sorba rendezés után egyenlő nagyságú csoportokat lehessen kialakítani. Ennek az az oka, hogy az adott iparágba tartozó vállalatok pontszáma megegyezik.

A 2. ábra az egyes környezeti pillér pontszámokhoz tartozó vállalatok számát mutatja. A portfólió kialakításához használt küszöbértékeket úgy határoztam meg, hogy a vállalatok számának a különbsége a két kialakított portfólióban minél kisebb legyen. Továbbá figyelembe vettem azt is, hogy a csoportokkénti környezeti pillér pontszámok átlagának a különbsége minél nagyobb legyen és mindkét portfólióba kellő számú vállalat kerüljön. Ennek következtében a zöld portfólió 86, míg a barna 78 vállalat részvényeit tartalmazza.



2. ábra: Az egyes környezeti pillér pontszámokhoz tartozó vállalatok száma

Forrás: saját szerkesztés

A zöld és a barna portfólió kialakítása során az egyes részvényeket a vállalatok piaci értékei szerint súlyoztam. Mindkét portfólió esetében kiszámoltam az adott portfólió napi hozamát az adott részvényhozamok súlyozott átlagaként, ahol a súlyokat a portfólióban szereplő vállalatok piaci értéke alapján határoztam meg. Ennek megfelelően a k portfólió hozama a t -edik időpontban:

$$R_{kt} = \frac{MV_{it-1}}{\sum_{i=1}^n MV_{it-1}} RETURN_{it}, \quad (5)$$

ahol MV_{it} , $RETURN_{it}$ a portfólióban szereplő i -edik vállalat piaci értéke, valamint a részvényének a hozama a t -edik időpontban, n pedig az adott portfólióban szereplő vállalatoknak a száma.

A zöld és a barna portfólió hozamkülönbségének (spread) a meghatározásához az (6) képlet alapján kiszámoltam az egyes portfóliók napi hozamát, majd a zöld portfóliónak a hozamából kivontam a barna portfóliónak a hozamát:

$$spread_t = R_{zöld,t} - R_{barna,t}, (6)$$

ahol $R_{zöld,t}$ és $R_{barna,t}$ rendre a zöld, valamint a barna portfólió hozama t-edik időpontban.

3.3. VEKTOR-AUTOREGRESSZÍV MODELL

A környezeti témára irányuló figyelemnek és a szélsősége időjárásnak a zöld és a barna portfólió hozamkülönbségére való hatását, vektor-autoregresszív modell (VAR) segítségével vizsgálom. A VAR modell az egyik legelterjedtebb és legkönnyebben használható modell a többváltozós idősorok elemzésére (Zivot & Wang, 2006). A modell lényegében az egyváltozós autoregresszív modell kiterjesztése többváltozós idősorokra. A VAR modellekben tehát egyszerre több idősor szerepel, melyek általában endogén változói a modellnek, vagyis a modellen belül határozódnak meg. Ha exogén változókat is szerepeltetünk a modellben, akkor VARX modellről (vektor-autoregresszív modell exogén változóval) beszélünk.

A modell elkészítéséhez szükségünk vagy egy, az endogén változókhoz kapcsolódó késleltetési paraméter (p) meghatározására. A szakirodalomban ezt általában valamilyen információs kritérium alapján határozzák meg. Tanulmányom, én az Akaike-féle információs kritériumot használom. Ha exogén változókat is szerepeltetünk a modellben, akkor szükségünk van egy, az ezen változókhoz kapcsolódó késleltetési paraméter (q) meghatározására is, melynek értékét szintén az Akaike-féle információs kritérium alapján határozom meg. Egy VARX(p, q) modell lényegében egy olyan rendszer, ahol az endogén változók utolsó p darab értékei hatással vannak saját magukra, valamint a többi endogén változóra. Az exogén változók utolsó q darab értékei pedig szintén hatással vannak az exogén változókra.

A modell feltétele, hogy a benne szereplő idősorok stacionerek legyenek. A stacionaritás tesztelése több módon is lehetséges. Tanulmányom a kiterjesztett Dicky-Fuller-tesztet (ADF-teszt) (Said & Dickey, 1984) alkalmazom a stacionaritás ellenőrzésére. A teszt nullhipotézise, hogy az adott idősor nem stacioner, így amennyiben el tudom vetni a nullhipotézist az adott idősort stacionernek tekintem. A dolgozatomban a stacionaritás tesztelésénél mind végig 5%-os szignifikanciaszintet alkalmazok. Amennyiben valamelyik idősor az ADF-teszt alapján nem tekinthető stacionernek, akkor megfelelő számú differenciálást alkalmazok.

A modell további feltétele, hogy az egyes endogén változók modell reziduumaik fehérzaj folyamatot követnek, valamint a karakterisztikus polinom gyökei az egységkörön belül helyezkednek el. Amennyiben ezen modellezési kritériumok nem teljesülnek a modellből

levont következtetések tévesek lehetnek. A reziduumok fehérzaj voltát Portmanteau-teszttel tesztelem. Ezen teszt nullhipotézise, hogy a reziduális autokavarianciák nullának tekinthetők, amely azt jelenti, hogy a reziduumok együttesen fehérzaj folyamatot követnek.

Elemzésem során elsőnek a vizsgált időszak teljes egészére futtatok VAR modelleket. Összesen két különböző modellt vizsgállok. Mindkét modell esetében megvizsgálom, hogy van-e Granger értelemben oksági kapcsolat az endogén változók között és elemzem az impulzus-válaszfüggvényeket.

A Granger-okság az endogén változók közötti ok-okozati kapcsolatok vizsgálatára alkalmas. A Granger-oksági teszt lényegében egy F-próba, melynek nullhipotézise, hogy az okozati változóra felírt, a modell által becsült egyenletben, az oksági változó minden késleltetettjének együtthatója együttesen nullának tekinthető. Ennek következtében Granger-okság akkor áll fenn, ha nullhipotézis elutasításra kerül.

VAR modellek esetében általában nem a modell által becsült együtthatókat szokás elemezni, hanem az úgynevezett impulzus-válaszfüggvényeket (IRF). Ennek oka, hogy a modellek legtöbbször nagyon sok paramétert tartalmaznak és változók közötti multikollinearitás miatt, a becsült paraméterek standard hibái torzítottak lesznek. Emiatt nehéz következtetéseket levonni, hiszen a t-tesztek is félrevezető eredményt adhatnak. Az IRF azt mutatja meg, hogy egy adott endogén változó modellből származó reziduumainak egységnyi szórású pozitív irányú sokkja, hogyan hat az egész rendszerre. Ennek következtében a IRF alkalmas arra, hogy a sokkok hatását és lecsengését vizsgáljuk. Elemzésemben elsősorban azt vizsgálom, hogy az egyes endogén változók maradéktagjainak a sokkjai, hogyan hatnak a zöld és barna portfóliók hozamkülönbségére.

Az első modellben csak a Google keresési volumen adatokat, a hírcikkekben előforduló környezeti téma arányát és a zöld és a barna portfólió hozamkülönbségét szerepeltetem. A második modell esetében az első modell változói mellett exogén változóként szerepeltetem a két időjárási változót is. Ezen megfontolás azon a feltételezésen alapszik, hogy a napi csapadékmennyiségre és a napi maximumhőmérsékletre a modell egyik változója nincs hatással, így ezen változók a modellen kívül határozódnak meg. Endogén változóként való szerepeltetésük a modellben ennek következtében nem indokolt. Az eredmények robusztusságának az érdekében a modelleket lefutatom az első, valamint az utolsó két év adatai nélkül is.

A kétfajta modellt megbecsülöm minden egyes 5 éves időszakra és megvizsgálom van-e különbség a két modell egy periódusra szóló előrejelzési hibái között. Azért 5 éves időszakokat nézek, mert ez az időszak már kellően sok megfigyelést tartalmaz ahhoz, hogy a modellek

paramétereit egyértelműen meg lehessen becsülni. Illetve az előrejelzési pontok is egy 8 éves időszakot fednek le, amely elég hosszú időszak ahhoz, hogy össze lehessen hasonlítani a kétfajta modell előrejelzéseit. A két előrejelzés esetén kiszámolom az átlagos négyzetes előrejelzési hibákat (MSFE) a (7) egyenlet szerint.

$$MSFE_i = 1000 \frac{1}{N} \sum_{t=1}^N e_{it}^2, (7)$$

ahol $MSFE_i$ az i-edik modell átlagos négyzetes előrejelzési hibája, e_{it} az i-edik modell előrejelzési hibája a t-edik időpontban, N pedig az időszakok száma. A két modell MSFE értékének az összehasonlításával azonban nem tudjuk elönten, hogy a különbség statisztikailag szignifikánsnak mondható-e. Ennek következtében Diebold-Mariano-teszttel (Diebold & Mariano, 2002) tesztelem, hogy a két előrejelzés pontossága között van-e szignifikáns különbség. A teszt nullhipotézise azt mondja ki, hogy a két előrejelzés ugyan olyan jó, tehát nincs közöttük szignifikáns különbség. Amennyiben tehát szignifikáns különbség mutatható ki az előrejelzések között a VARX modell javára, akkor abból arra következtethetünk, hogy a rendellenes időjárási körülményeket figyelembevételével jobban előre tudjuk jelezni a zöld és a barna portfóliók hozamkülönbségének az alakulását.

hub to Gulf Coast refineries. With the technical aspects now complete, the companies will start shipping oil this weekend to the Gulf Coast from Cushing, Okla., where the recent boom in domestic oil production has brought inventories to historic highs. The wave of new oil, together with a lack of ways to bring it out of Cushing, had brought benchmark U.S. oil prices as much as \$20 below European benchmark Brent. Enterprise and Enbridge said they will expand the 150,000-barrel-a-day pipeline's capacity to 400,000 barrels a day in the first quarter of 2013."

„The monsoon in India where crops are heavily dependent on the rains is likely to be normal this year, the India Meteorological Department said Friday. The June-September monsoon rainfall is likely to be 98% of the long-term average, the department said. In 2009, monsoon rainfall was 22% below the long-term average, making it the worst in 37 years. This despite an initial prediction by the Meteorological Department that the rains would be just 4% below average. Apart from the volume of rain, other factors such as the timing of the start of the season as well as the geographical spread of rainfall are important in determining the impact on crop output."

„Dwindling oil production along Alaska's northern edge means the famed Trans Alaska Pipeline carries less than one-third the volume it once did and the crude takes five times as long to get to its destination. Unless a technological solution can be found, the arcane physics of crude flow may put an end to the multibillion dollar, 48-inch-wide steel pipeline and determine the fate of the largest oil field ever found in the U.S."

Látható ezen példákból, hogy a tárgyalt első téma, elsősorban olyan cikkekben jelenik meg, melyek az energiaágazattal kapcsolatos fejleményeket, valamint a különböző természeti katasztrófáknak és a rendellenes időjárási körülményeknek a lehetséges hatásait tárgyalják. Ennek következtében az azonosított környezeti téma előfordulásának a gyakorisága valóban alkalmas arra, hogy a továbbiakban a környezeti témákra irányuló médiafigyelem megfelelő mérőszámaként használjam. A továbbiakban környezeti téma előfordulásának a gyakoriságát (γ) használom az elemzésem során.

4.2. A VAR-MODELL EREDMÉNYEI

Ebben a fejezetben azt vizsgálom, hogy eltérő-e a környezeti témákra irányuló figyelmet leíró változók (γ , AVSI) hatása a zöld és a barna portfóliók hozamkülönbségére (spread), akkor, ha figyelembe vesszük a napi maximumhőmérséklet (t_{max}) és a napi csapadékmennyiség (ppt) havi normál értékeitől való eltérését. Elsőként a vizsgált időszak

egészére futtattam egy VAR modellt a spread, gamma és AVSI változókkal, majd egy VARX modellt futtattam, ahol a t_{max} és p_{pt} változókat exogén változóként szerepeltettem a modellben.

Elsőként ADF-teszteltem, hogy stacionerek-e az egyes idősorok. Mivel a gamma idősor a teszt alapján nem volt stacioner, így vettem ezen idősor logdifferenciáját, amely már stacionernek bizonyult. Ezt követően lefuttattam a modelleket és megvizsgáltam a Granger-oksági kapcsolatokat. A modelleket lefuttattam az első és az utolsó két-két év elhagyásával is és elvégeztem ugyan ezen lépéseket. A p és q késleltetéseket minden esetben az Akaike-féle információs kritérium alapján határoztam meg. Ezen modellek eredményeit az 1. táblázat tartalmazza. A táblázatból látható, hogy a Portmanteau-tesztek p -értékei jóval nagyobbak, mint a legnagyobb szokásos, 10%-os szignifikanciaszint, így ez azt jelenti, hogy a modellek reziduumaik fehérzaj folyamatot követnek. A modellek karakterisztikus polinomjainak a gyökei is jóval kisebbek egynél, ennek következtében minden modellezési feltétel teljesül.

1. táblázat: VAR és VARX modellek eredményei

	2010 - 2022		2010 - 2020		2012 - 2022	
	VAR(13)	VARX(13,1)	VAR(13)	VARX(13,1)	VAR(13)	VARX(13,1)
Granger-okság	p-érték		p-érték		p-érték	
gamma $\mapsto$ spread	4,1%	3,8%	15,5%	15,4%	4,9%	4,7%
AVSI $\mapsto$ spread	45,0%	45,0%	49,9%	49,8%	55,6%	55,7%
spread $\mapsto$ gamma	23,2%	21,9%	56,4%	54,9%	42,6%	41,1%
AVSI $\mapsto$ gamma	83,4%	82,9%	77,1%	77,2%	57,7%	57,7%
spread $\mapsto$ AVSI	37,0%	37,4%	63,7%	62,3%	34,0%	32,8%
gamma $\mapsto$ AVSI	59,9%	60,0%	75,4%	76,2%	82,2%	82,1%
Portmanteau-teszt (p-érték)	23,1%	23,4%	52,0%	52,0%	26,6%	28,6%
Legnagyobb gyök	0,88	0,88	0,87	0,89	0,87	0,89

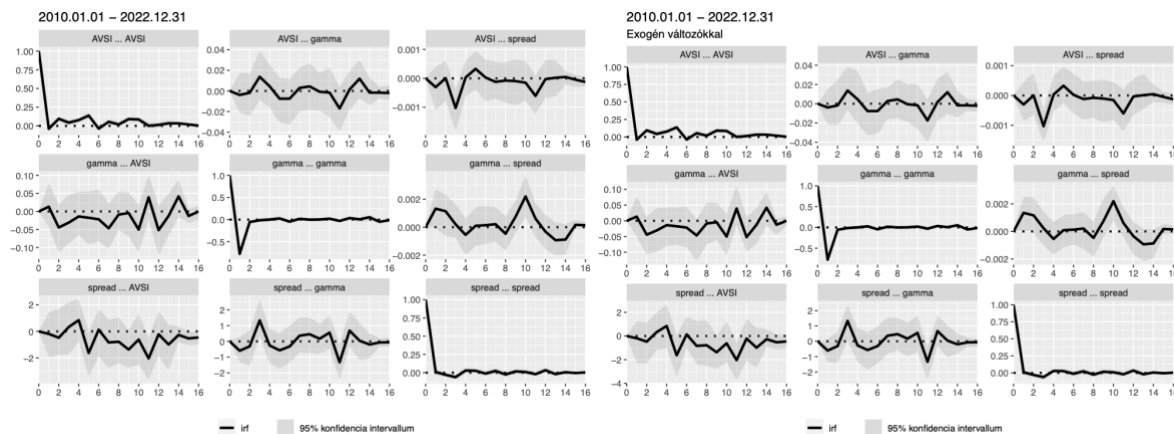
A táblázatban $VAR(p)$, azaz az időjárási változókat (t_{max} , p_{pt}) nem tartalmazó modell, valamint a $VARX(p,q)$, azaz az időjárási változókat exogén változóként tartalmazó modellek eredményeit mutatja három különböző időszak esetén. A táblázatban az egyes endogén változók Granger-oksági tesztjének a p -értékei láthatóak, ahol az 5%-os szignifikáns értékek ki vannak emelve. Az utolsó előtti sorban az egyes modellek Portmanteau-tesztjének p -értékei láthatóak, az utolsó sor pedig a legnagyobb gyökeket tartalmazza a modellek karakterisztikus polinomjainak.

Forrás: saját szerkesztés

Az 1. táblázatból kiolvasható az is, hogy a gamma Granger-okozza a spread változót a 2010 és a 2022, valamint a 2012 és a 2022 közötti időszakok esetében. Ez onnan látszik, hogy a Granger-okságot tesztelő F-próba által kapott p -értékek alapján elvethető az a nullhipotézis, hogy a gamma változó nem Granger-okozza a spread változót 5%-os szignifikanciaszint mellett. Ez az oksági kapcsolat azt jelenti, hogy a hírekben megjelenő környezeti téma arányának múltbeli értékei szignifikánsan magyarázzák a zöld és a barna portfóliók

hozamkülönbségének jelenlegi értékét. Az összes többi lehetséges esetben nem áll fenn Granger-okság az endogén változók között. Ez többek között azt is jelenti, hogy a Google keresések volumenének múltbeli értékei nincsenek szignifikáns hatással a zöld és a barna portfóliók hozamkülönbségének jelenlegi értékére. Megfigyelhetjük továbbá, hogy a gamma és a spread változó közötti Granger-okság függ a vizsgált időszak

A Granger-oksági kapcsolatok elemzése után a különböző időszakokra lefuttatott modellek esetében megvizsgáltam az impulzus-válaszfüggvényeket. A 5. ábra a teljes vizsgált időszakra lefuttatott modellek IRF-ét mutatja. Az ábrákon látható, hogy nincsen lényegi különbség az exogén változókat nem tartalmazó, valamint az azokat tartalmazó modellek között. A sokkok vizsgálatánál elsősorban az érdekel bennünket, hogy az AVSI, valamint a gamma becsült reziduumjainak egységi szórású pozitív sokkjai, hogy hatnak a zöld és a barna portfóliók hozamkülönbségére, vagyis a spread-re. Az AVSI-t érő sokk a spread értékét csökkenti és ez a negatív hatás a harmadik periódusban szignifikánsnak is mondható. A gamma differenciázott értékét érő sokk pedig növeli a spread értékét és ez a hatás az exogén változókat tartalmazó modell esetében már az első periódusban szignifikáns. Továbbá mindkét modell esetében a 10. periódusban figyelhető meg a legnagyobb hatás, ami mindkét esetben szignifikáns.



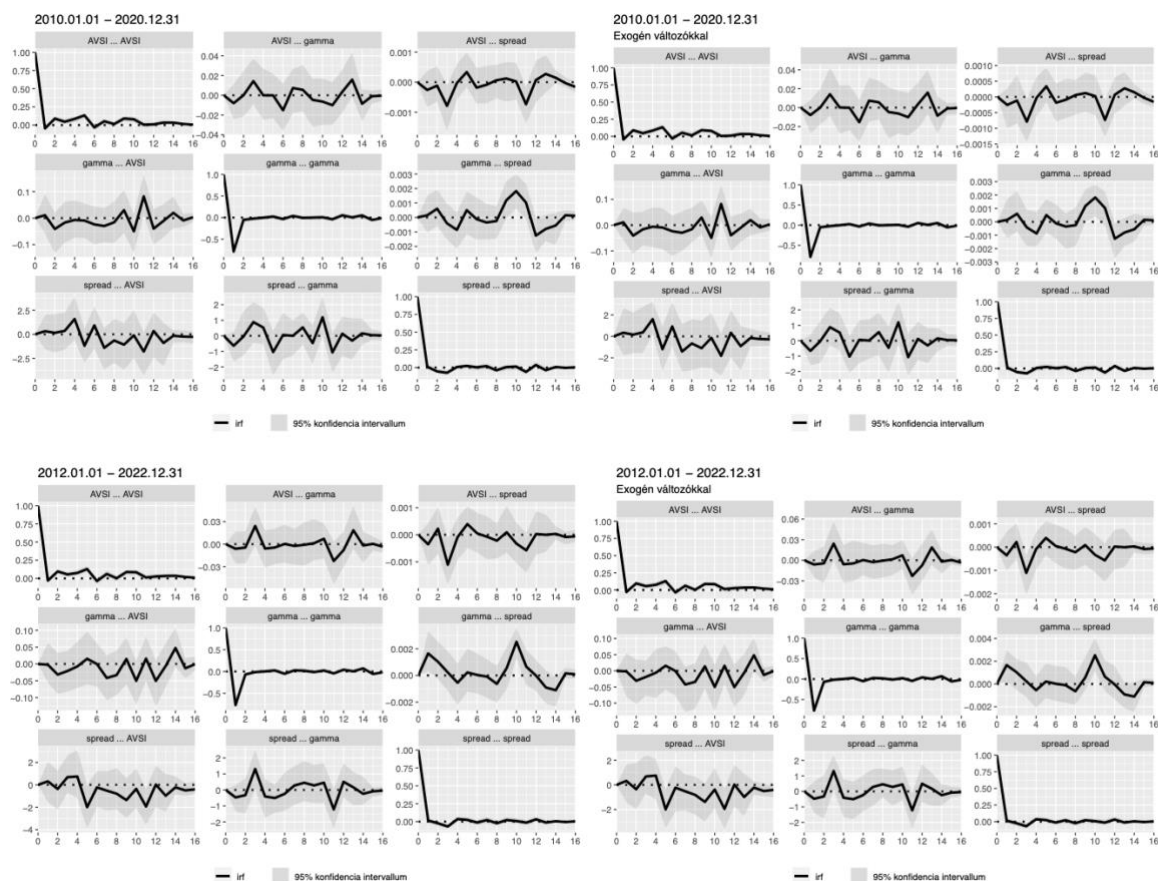
4. ábra: Impulzus-válaszfüggvények a 2010 és 2022 között futtatott modellekre

Az ábrán az exogén változókat nem tartalmazó, valamint az azokat is tartalmazó modellek impulzus-válaszfüggvényei (IRF) láthatóak a 2010.01.01 és a 2022.12.31. közötti időszakra. Az egyes IRF-eknél az első változó az a változót jelenti, amelyben a sokk keletkezett, a második (a ... utáni) változó pedig az amelyikre való hatását láthatjuk.

Forrás: saját szerkesztés

Az első, valamint az utolsó két év elhagyásával futtatott modellek impulzus-válaszfüggvényei is nagyon hasonló eredményt mutatnak (lásd a 6. ábrát). Az AVSI-t érő

sokkok csökkentik, a gamma logdifferentenciázott értékeit érő sokkok pedig alapvetően növelik a spread értékét. Az időjárás változókat nem tartalmazó, valamint az azokat exogén módon szerepeltető modellek között pedig jelentős különbségek nem figyelhetőek meg.



5. ábra: Impulzus-válaszfüggvények a 2010 és 2020 között, valamint a 2012 és 2022 között futtatott modellekre

Az ábrán az exogén változókat nem tartalmazó, valamint az azokat is tartalmazó modellek impulzus-válaszfüggvényei (IRF) láthatóak a 2010.01.01 és a 2020.12.31., valamint a 2012.01.01 és a 2022.12.31. közötti időszakokra. Az egyes IRF-eknél az első változó az a változót jelenti, amelyikben a sokk keletkezett, a második (a ... utáni) változó pedig az amelyikre való hatását láthatjuk.

Forrás: saját szerkesztés

Az eredmények alapján összességében azt figyelhetjük, meg a vizsgált időszak alatt a környezeti témájú hírek arányának jelenlegi és múltbeli értékei szignifikánsan és pozitívan befolyásolják a zöld és a barna portfóliók hozamkülönbségének jövőbeli értékeit. Viszont a kapcsolat szignifikanciája változik attól függően, hogy milyen időszakot vizsgálunk. Összességében tehát egyértelmű kapcsolat nem állapítható meg a két változó között, viszont az IRF-ek alapján látható, hogy még ha nem is minden esetben szignifikáns a kapcsolat, a környezeti témájú hírek aránya növeli a zöld és a barna portfóliók hozamkülönbségét. Ez az

megállapítás megfelel az előzetes hipotézisemnek is és konzisztens a korábbi szakirodalmi eredményekkel. Ezzel ellentétben, ugyanakkor a Google keresési volumenadatok növekedése alapvetően csökkenti a spread értékét az IRF-ek alapján. Ez már sokkal inkább meglepő eredmény, viszont érdemes kihangsúlyozni, hogy amíg a környezeti témájú hírek elemzésénél, kifejezetten a gazdasági élettel és befektetésekkel foglalkozó hírportál cikkeit vettük figyelembe, addig a Google keresési volumenadatok nem kifejezetten a befektetők figyelmét tükrözik, így ez valamelyest magyarázatot adhat a kapott eredményekre. Azt is láthattuk, hogy az időjárás változók kontrol változókként való szerepeltetése a modellben, alapvetően nem változtatta meg a kapott eredményeket. Ez tehát azt jelenti, hogy a befektetőknek a környezeti kérdésekkel kapcsolatos nézeteit inkább az ezen témára irányuló médiafigyelem, mintsem a rendellenes időjárási körülmények határozzák.

Elemzésem utolsó részében az exogén változókat nem tartalmazó (VAR), valamint az azokat is szerepeltető modellek (VARX) előrejelzéseit vizsgálom. A 2010.01.02 és 2015.01.01. közötti időszakkal kezdődően, minden egyes 5 éves időszakra megbecsültem a modelleket, egy-egy periódussal mindig előre haladva, és a következő periódusra előrejelzéseket végeztem. A p és a q paramétereket ezekben az esetekben is a Akaike-féle információs kritérium alapján határoztam meg. Ennek következtében összesen 1990 darab előrejelzést végeztem a 2015.01.02. és a 2023.12.31. közötti időszak minden egyes kereskedési napjára vonatkozóan. A modellezési feltételek minden esetben teljesült. A gamma idősor logdifferenciálását követően minden idősor stacioner volt, minden időszak esetében. Az egyes modellek reziduumaival továbbá együttesen fehérzajnak tekinthetők a Portmanteau-teszt szerint a legtöbb időszakban, viszont ahol a Portmanteau-teszt nullhipotézisét elvetettük, ott a Ljung-Box-teszt szerint az egyes endogén változók reziduumaival már fehérzaj folyamatot követtek. Végül a modellek karakterisztikus polinomjainak a gyökei is az egységkörön belül helyezkedtek el minden esetben. A (7) képlet szerint kiszámoltam VAR és VARX modellek átlagos négyzetes előrejelzési hibáit, valamint elvégeztem a Diebold-Mariano-tesztet két modell előrejelzéseire, amely során az $L(x) = x^2$ veszteségfüggvényt használtam. Ezen eredményeket a 2. táblázatban foglaltam össze.

2. táblázat: Átlagos négyzetes előrejelzési hibák és a Diebold-Mariano-teszt eredménye

	MSFE	DM-teszt	
		DM	p-érték
VAR	0,15239	0,12	90,5%
VARX	0,15235		

A táblázatban VAR, azaz az időjárási változókat ($tmax$, ppt) nem tartalmazó modell, valamint a VARX, azaz az időjárási változókat exogén változóként tartalmazó, minden 5 éves időszakra megbecsült modellek egy periódusos előrejelzéseinek az átlagos négyzetes előrejelzési hibáit mutatja (MSFE) a 2015.01.02 és a 2023.12.31. közötti időszakra vonatkoztatva. Továbbá a két modell előrejelzései Diebold-Mariano-tesztjének a teststatisztikája (DM), valamint a p-értéke láthatóak.

Forrás: saját szerkesztés

A 2. táblázatból látható, hogy a VARX modell átlagos négyzetes előrejelzési hibája kisebb, mint a VAR modellé. Ez azt jelenti, hogy a vizsgált mintán az időjárási változókat tartalmazó modell pontosabb előrejelzéseket adott átlagosan, mint a csak a környezeti témákra irányuló figyelmet leíró változókat tartalmazó modell. A Diebold-Mariano-teszt alapján azonban a két előrejelzés közötti eltérés nem szignifikáns, hiszen a teszt azon nullhipotézisét nem tudom elvetni semmilyen szokásos szignifikanciaszinten, hogy szignifikáns eltérés van a két előrejelzés között.

Ezen eredmények alapján megállapíthatjuk, hogy a szokásostól eltérő időjárás figyelembevételével szignifikánsan nem tudjuk javítani a zöld és a barna portfóliók hozamkülönbszetére adott előrejelzéseket. Ez az eredmény ráerősít arra, hogy a befektetőknek a környezeti kérdésekkel kapcsolatos nézeteit lényegesen nem befolyásolják a rendellenes időjárási körülmények.

5. KONKLÚZIÓ

Tanulmányomban a környezeti kérdésekre irányuló figyelemnek és a rendellenes időjárási körülményeknek elemeztem a „zöld” és a „barna” részvényekből álló portfóliók hozamkülönbszetére gyakorolt hatását. A kutatásomat a 2010. január 1. és 2022. december 31. közötti időszakra végeztem el és az S&P500 indexben szereplő vállalatok részvényeit vizsgáltam. A környezeti kérdésekre irányuló figyelmet egyrészt az LDA témamodellezés segítségével vizsgáltam, másrészt a környezettel kapcsolatos Google keresések volumenét is felhasználtam. A rendellenes időjárási körülményeket a napi maximumhőmérséklet és a napi csapadékmennyiség normál értékeitől való eltéréssel határoztam meg az USA területére vonatkozóan. A zöld és a barna portfóliók kialakítására az iparági környezeti pillér pontszámokat használtam. Ezt követően vektor-autoregresszív modellekkel vizsgáltam a környezeti témára irányuló figyelmet leíró változóknak a zöld és barna portfóliók hozamkülönbszetére gyakorolt hatását és megvizsgáltam, hogy az eredményeket miképpen befolyásolják a szokásostól eltérő időjárási körülményeket leíró változók exogén módon való szerepeltetése a modellben. Továbbá az egyes modellekkel előrejelzéseket is készítettem és megnéztem, hogy az időjárási változók szerepeltetésével jobb előrejelzést tudok-e adni a zöld és barna portfóliók hozamkülönbszetére.

Az eredmények alapján azt a következtetést vonhatjuk le, hogy a hírekben megjelenő környezeti kérdések arányának növekedése növeli, a környezettel kapcsolatos kifejezések Google keresési volumenének növekedése viszont csökkenti a zöld és a barna portfóliók hozamkülönbszetét. Ezeket a némileg egymással ellentmondásos eredményeket valamilyen mértékben az magyarázhatja, hogy az elemzésemben egy kifejezetten a gazdasági és pénzügyi világgal foglalkozó hírportál cikkeit vizsgáltam. Ennek következtében, az ezen a hírportálon megjelenő cikkek valószínűleg jobban befolyásolják a befektetők nézeteit, mivel hiteles és aktuális kérdések jelennek meg bennük. Ezzel ellentétben a Google keresések nem közvetlenül a befektetők érdeklődését tükrözik, sokkal inkább a témára irányuló átlagos figyelmet ragadják meg. Az időjárási változók szerepeltetése a modellben ugyanakkor nem változtatta meg az eredményeket. Ebből azt a következtetést vonhatjuk le, hogy a befektetőknek a környezeti témával kapcsolatos nézeteit alapvetően nem befolyásolják a rendellenes időjárási körülmények. Ezen megállapításra ráerősít az az eredmény is, hogy a szokásostól eltérő időjárást leíró változók szerepeltetésével a modellben nem tudunk szignifikánsan jobb előrejelzést adni a zöld és a barna portfóliók hozamkülönbszetére.

A kutatásomnak a legfontosabb korlátja, hogy csak azon vállalatokat elemeztem, amelyek a tanulmányom készítésének az időpontjában is a tőzsdén voltak és szerepeltek az S&P500 indexben. Ennek következtében az vizsgált időszakban az indexből kikerülő vállalatokat nem vizsgáztam, így ez jelentősen torzíthatja az eredményeimet és az azokból levont következtetéseket. Ez a probléma az előrejelzések készítésénél is jelentős torzítást okozhat, hiszen korábbi időpontokban nem voltak megfigyelhetőek, mely vállalatok lesznek tagjai az indexnek. Ugyan ez igaz az iparági ESG adatokra is melyek a vizsgált időszak elején nem voltak megfigyelhetőek. Mindezek ellenére, mivel alapvetően a modellek előrejelzései közötti különbségeket elemeztem és nem az előrejelzés jóságát, így mindez kevésbé okozott problémát. Összességében tehát azt lehet mondani, hogy az eredmények valamelyest torzítottak lehetnek, így mindenképpen érdemes lenne az általam elvégzett elemzéseket egy olyan, több vállaltot tartalmazó mintán is elvégezni, amelyek tartalmazzák az indexben már nem szereplő vállalatokat. Mindez tehát teret ad egy további kutatásnak is.

Fontos megjegyezni továbbá, hogy nem adtam választ arra, hogy mi okozza a környezeti kérdésekről szóló eltérő médiafigyelmet. Ahhoz azonban, hogy konkrétan fel tudjuk tárni a zöld befektetések és a környezeti témákra irányuló figyelem közötti kapcsolatot, ezen kérdést mindenképpen meg kell válaszolnunk. Ehhez azonban további elemzések szükségesek.

6. IRODALOMJEGYZÉK

- Ardia, D., Bluteau, K., Boudt, K., & Inghelbrecht, K. (2020). Climate change concerns and the performance of green versus brown stocks. *Available at SSRN 3717722*.
- Bischof, J., & Airoidi, E. M. (2012). Summarizing topical content with word frequency and exclusivity. *Proceedings of the 29th international conference on machine learning (icml-12)*, 201–208.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
- Center for International Earth Science Information Network (CIESIN), Columbia University. 2017. Documentation for the U.S. Census Grids (Summary File 1), 2010. *Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC)*.
- Choi, D., Gao, Z., & Jiang, W. (2020). Attention to global warming. *The Review of Financial Studies*, 33(3), 1112–1145.
- Csillag J, B., Granát, M., & Neszveda, G. (2022). Media Attention to Environmental Issues and ESG Investing. *Financial and Economic Review*, 21(4), 129–149.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The journal of finance*, 66(5), 1461–1499.
- Daly, C. (2013). Descriptions of PRISM spatial climate datasets for the conterminous United States. *PRISM doc*, 14.
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology: a Journal of the Royal Meteorological Society*, 28(15), 2031–2064.
- Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3), 66–95.
- Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1), 134–144.
- Eikon, R. (2022). *Environmental, social and governance scores from Refinitiv*. London: Refinitiv Eikon.
- Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebe, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3), 1184–1216.
- Greene, D., O’Callaghan, D., & Cunningham, P. (2014). How many topics? Stability analysis for topic models. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I 14*, 498–513.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl_1), 5228–5235.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267–297.
- Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of statistical software*, 40, 1–30.
- Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of accounting and economics*, 36(1–3), 337–386.

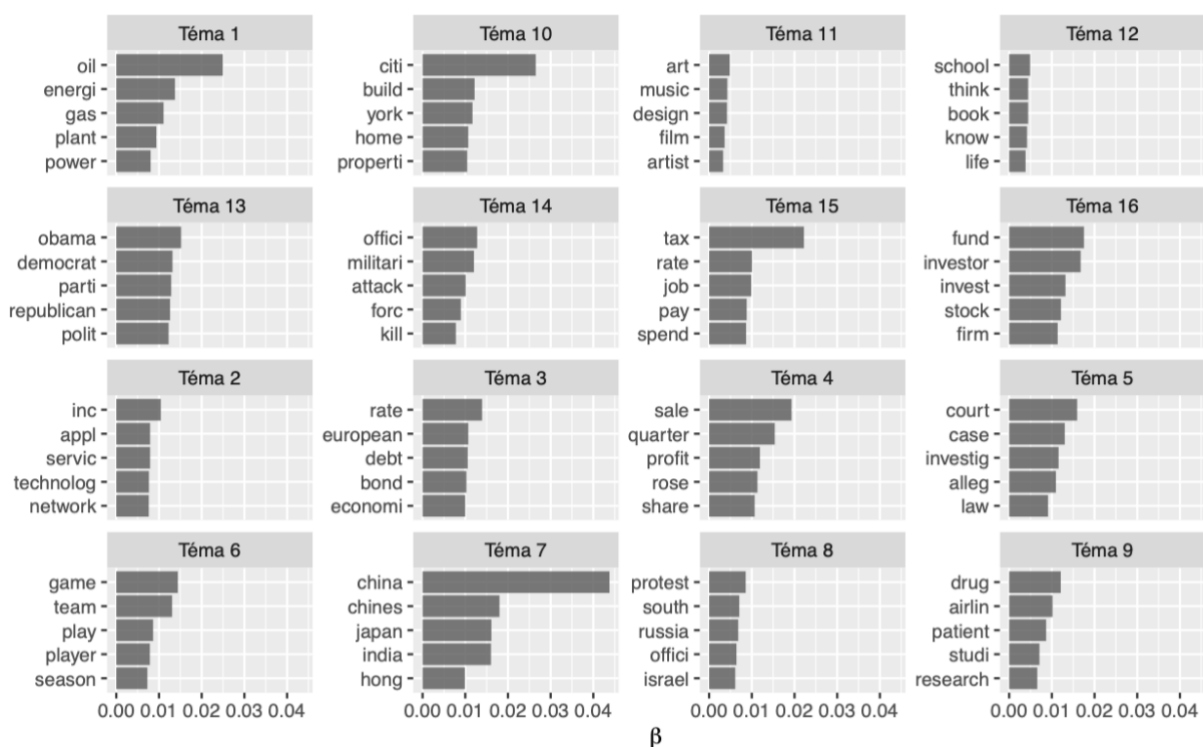
- Jacobi, C., Van Atteveldt, W., & Welbers, K. (2018). Quantitative analysis of large amounts of journalistic texts using topic modelling. In *Rethinking Research Methods in an Age of Digital Journalism* (o. 89–106). Routledge.
- Kvam, E., Molnár, P., Wankel, I., & Ødegaard, B. A. (2022). Do sustainable company stock prices increase with ESG scrutiny? Evidence using social media. (*March 15, 2022*). <http://dx.doi.org/10.2139/ssrn.4057988>
- Li, Y., Johnson, E. J., & Zaval, L. (2011). Local warming: Daily temperature change influences belief in global warming. *Psychological science*, 22(4), 454–459.
- Mendelsohn, R., Dinar, A., & Williams, L. (2006). The distributional impact of climate change on rich and poor countries. *Environment and development economics*, 11(2), 159–178.
- Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2), 550–571.
- Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2022). Dissecting green returns. *Journal of Financial Economics*, 146(2), 403–424.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, 91, 1–40.
- Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599–607.
- Schmidt, G. (2015). Thoughts on 2014 and ongoing temperature trends. *Blog post, Real Climate*, 22.
- Wanidwaranan, P., & Padungsaksawasdi, C. (2022). Unintentional herd behavior via the Google search volume index in international equity markets. *Journal of International Financial Markets, Institutions and Money*, 77, 101503.
- White, S. A. (2013). VDatum: Vertical Datum Transformation Tool. *Presented to the Hydrographic Services Review Panel*.
- Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®*, 385–429.

MELLÉKLETEK

M1: A Google Trends adatok során használt környezettel kapcsolatos kifejezések

Kifejezés
Acid rain
Biodiversity
Carbon footprint
Climate change
Conservation
Deforestation
Endangered species
Environment
Global warming
Globalization
Greenhouse gas
Ozone layer
Pollution
Renewable energy

M2: Az egyes témákban legnagyobb valószínűséggel előforduló kifejezések



Forrás: saját szerkesztés