

TDK dolgozat

2024

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



MNB INTÉZET
FENNTARTHATÓ PÉNZÜGYEK KÖZPONT



Neumann János Egyetem
Gazdaságtudományi
Kar

**A JEGYBANKI KOMMUNIKÁCIÓ TÉMÁI ÉS AZOK PIACRA
GYAKOROLT HATÁSAI**

TÁNCZOS GERGELY DÁNIEL

Konzulens: Granát Marcell

Kézirat lezárásának dátuma: 2024. április 15.

Absztrakt

A dolgozat a jegybanki kommunikációval és azok piacra gyakorolt hatásával foglalkozik, azon belül hogy meghatározhatóak e olyan jegybanki kommunikációban szereplő témák, melyek szignifikánsan képesek befolyásolni a piac volatilitását. A kutatás során 2018 és 2023 között az FOMC által kiadott nyilatkozatok és kamatdöntő ülések után tartott jegybankelnöki beszédek szövegei kerülnek felhasználásra. A dolgozat ezen a szövegeken topik modell segítségével azonosítja a szövegekben előforduló témákat és azok megoszlását az egyes szövegekben. A Standard & Poors index segítségével kerülnek meghatározásra a piacon tapasztalt volatilitások. Lineáris regresszió segítségével ad választ a dolgozat a kutatási kérdésre, felhasználva a volatilitások és a témák súlyát. A dolgozat a szakirodalomnak megfelelően kapcsolatot talál a kommunikációkban szereplő témák és a piaci volatilitás között.

Tartalomjegyzék

1. Bevezetés	4
2. Adatok	5
2.1. Szövegelemzés	5
2.1.1. Identifikáció	14
2.2. Tőzsdei árfolyamok	16
2.2.1. Leíró statisztika	17
3. Módszertan	21
4. Eredmények	22
5. Konklúzió	24
6. Appendix	25
Irodalomjegyzék	29

Ábrák jegyzéke

1.	Topikok megoszlása adott szövegek esetén - nyilatkozatok	12
2.	Topikok megoszlása adott szövegek esetén - sajtótájékoztatók teljes szövege	14
3.	Topikok megoszlása adott szövegek esetén - sajtótájékoztatók kérdésektől szűrve	15
4.	Topik eloszlások strukturált topik modell esetén - sajtótájékoztatók teljes . .	16
5.	Adatok volatilitása 30 perces ablakokban	18
6.	Adatok volatilitása 60 perces ablakokban	18
7.	Árfolyamok varianciája az egyes napokon	19
8.	Nyilatkozatok utáni fél óra varianciái	20
9.	Sajtótájékoztatók utáni egy óra varianciái	21
10.	Regressziók eredményei	24
11.	Topik eloszlások strukturált topik modell esetén - nyilatkozatok	29
12.	Topik eloszlások strukturált topik modell esetén - sajtótájékoztatók szűrt . .	30
13.	Egyes szövegek ideje körül márt varianciák eloszlása	30

Táblázatok jegyzéke

1.	Vizsgált szövegek időpontjai (a nyilatkozatok minden időpontban elérhetőek, a sajtótájékoztatók csak 5 esetben nem elérhetőek)	6
2.	Leggyakoribb kifejezések topikonként - nyilatkozatok, $k = 4$	8
3.	Leggyakoribb kifejezések topikonként - nyilatkozatok, $k = 5$	8
4.	Leggyakoribb kifejezések topikonként - nyilatkozatok, $k = 6$	8
5.	Topikok metszetében lévő szavak - nyilatkozatok	9
6.	Leggyakoribb kifejezések topikonként, tisztítás után - nyilatkozatok, $k = 4$.	9
7.	Leggyakoribb kifejezések topikonként, tisztítás után - nyilatkozatok, $k = 5$.	10
8.	Leggyakoribb kifejezések topikonként, tisztítás után - nyilatkozatok, $k = 6$.	10
9.	Perplexity értékek az egyes modellek esetén - nyilatkozatok	11
10.	Nyilatkozatokon végzett modellezés eredményeinek interpretálása	11
11.	Perplexity értékek az egyes modellek esetén - sajtótájékoztatók	13
12.	Teljes sajtótájékoztatókon végzett modellezés eredményeinek interpretálása	13
13.	Kérdésektől szűrt sajtótájékoztatókon végzett modellezés eredményeinek interpretálása	13
14.	A felhasznált adatok időbeli eloszlásának összegzése	17
15.	Eltávolított szavak a nyilatkozatok szövegeiből	25
16.	Eltávolított szavak a sajtótájékoztatók szövegeiből	26

17.	Leggyakoribb szavak topikonként és modellenként a sajtótájékoztatók szövegeinél	26
18.	Sajtótájékoztatók teljes szövegeinél a topikok metszetében szereplő szavak .	27
19.	Sajtótájékoztatók kérdésektől megtisztított szövegeinél a topikok metszetében szereplő szavak	27
20.	Nyilatkozatok és beszédek UTC-be átalakított időpontjai	28
21.	Struktúrált topik modellek összefoglaló táblázata	29

1. Bevezetés

A jegybanki kommunikációval és annak piacra gyakorolt hatásával több ismert tanulmány is foglalkozik Sack and Kohn (2003). A kamatdöntés hatását vizsgálták a kötvény-, deviza-, és tőkepiacokra is, többek között azok volatilitására is (Fišer and Horváth 2010). A kommunikációban szereplő témák is számtalan kutatás fókuszába kerültek (Cieslak and Schrimpf 2019). A nyilatkozatok és sajtótájékoztatók megjelenése szignifikánsan magasabb volatilitáshoz vezetnek a tőzsdei árfolyamok tekintetében (Lucca and Moench 2015).

A dolgozat kutatási kérdése ugyanakkor arra fókuszál, hogy megragadható-e egy olyan téma a jegybanki kommunikációban, ami a vizsgált időszakban magasabb volatilitást okozott a piacon mint a többi kommunikációban felhasznált téma. A topikok (témák) tartalma és a piaci volatilitás közötti szignifikáns kapcsolat igazolt (Jegadeesh and Wu 2015). Jelen dolgozat arra keresi a választ, hogy meghatározhatóak-e olyan jegybanki kommunikációban szereplő témák, melyek képesek befolyásolni a piaci volatilitást. A kommunikációban megjelenő témák közötti különbségtétel nem került olyan nagy reflektorfénybe, így kevesebb kutatás foglalkozott ezzel a témával.

A dolgozat a 2018 és 2023 közötti 6 év FOMC által kiadott nyilatkozatok és a kamatdöntés utáni jegybankelnöki beszédek szövegei segítségével végez elemzést. A tanulmány látens Dirichlet allokáció segítségével határozza meg az egyes szövegekben szereplő témákat és ezen témák adott szövegben betöltött súlyát. A piacon megfigyelt volatilitást a Standard & Poors indexből határozza meg a dolgozat 30 és 60 perces ablakokban. A nyilatkozatok esetén 30 perces ablakot használ a kutatás, tekintettel arra, hogy a nyilatkozat kiadása után 30 perccel kezdődik a jegybankelnöki beszéd, mely új információkat tud a piac számára biztosítani. Emiatt hogy csak a nyilatkozat hatását mérje a kutatás, 30 perc volatilitását vizsgálja. A sajtótájékoztatók esetében 60 perces ablakot vesz alapul a dolgozat, hiszen több kutatás is igazolta, hogy egy óra alatt a piac az újonnan kapott információk döntő részét beárazza. A kialakult volatilitások és a kommunikációban szereplő témák aránya segítségével ad választ a tanulmány a kutatási kérdésre.

A kutatás eredménye arra a válaszra jut, hogy léteznek olyan kommunikációban előforduló témák, melyek szignifikánsan képesek befolyásolni a piaci volatilitást. A sajtótájékoztató szövegei esetében az árstabilitással, makrógazdasági helyzettel és a gazdasági növekedéssel kapcsolatos témák kommunikációban betöltött súlyának emelkedése ceteris paribus magasabb volatilitást okoz a piacon. A nyilatkozatok esetében a munkaerőpiac és a monetáris politika okoz szignifikánsan magasabb volatilitást. A regresszióban szereplő eredmények egy része torzított, ami annak köszönhető hogy a volatilitások eloszlása nem normális.

2. Adatok

A dolgozat az FOMC által kiadott nyilatkozatok és a sajtótájékoztatók szövegein végzi az elemzést az R programcsomagot felhasználva (R Core Team 2022). A sajtótájékoztatókat a sajtó munkatársainak kérdéseitől és az erre adott válaszoktól megtisztítva is figyelembe veszi a dolgozat, így 3 eltérő szövegen végez elemzést. A szövegekben szereplő témák (topikok) meghatározására felügyelet nélküli gépi tanulási eszközt használ a kutatás, azon belül is látens Dirichlet-allokációt (Latent Dirichlet Allocation, LDA). Az LDA modellek két fontos feltevésre épülnek: egyrészt minden topik meghatározható bizonyos szavak keverékeként, másrészt minden szöveg meghatározható topikok keverékeként (Blei, Ng, and Jordan 2003). A jegybanki kommunikációra vonatkoztatva tehát ez azt jelenti, hogy minden jegybanki nyilatkozat és sajtótájékoztató több téma keveréke, továbbá minden téma szavak keverékéből áll.

2.1. Szövegelemzés

Minden adat a Federal Reserve (<https://www.federalreserve.gov>) oldaláról származik („Federal Reserve,” n.d.). A dolgozat a 2018 és 2023 közé eső Fed nyilatkozatokkal és sajtótájékoztatókkal foglalkozik. Az FOMC naptárban megtalálható egyéb adatokat – mint például az FOMC Minutes – nem használja fel a kutatás. Ennek oka, hogy ezen szövegek lényegesen később kerülnek kiadásra, rendszerint több héttel az ülés után. Ezen szövegek elenyésző új információt tartalmaznak a nyilatkozatok és a beszédekhez viszonyítva.

A vizsgált 6 év során több nyilatkozatot adott ki az FOMC, mint amit a dolgozat felhasznál a kutatás során. Ezt azért teszi a dolgozat, mert minden kiadott nyilatkozat felhasználása torzítást vitt volna a kutatásba. Összesen 3 alkalommal tartott az FOMC előre nem tervezett ülést a vizsgált időszakban: 2019. október 4.-én, 2020. március 3.-án továbbá 2020. március 15.-én. Az előre nem tervezett ülések körül nagy piaci mozgások jellemzőek, amik nem csak az ülésen elhangzottak, hanem magának a nem tervezett ülés tényének köszönhető. Az az információ, hogy a Fed nem tudott várni a következő ülésig és ezért szükségesnek látta egy váratlan ülés megtartását fontos információ a piac számára, emiatt jellemzőek a nagyobb volatilitások is. 2020 márciusában a covid válság kezdete adott okot a döntéshozóknak a váratlan ülés tartására. A kutatás a fenti 3 időpontban kiadott FOMC nyilatkozatot kiveszi az elemzésből, mert azok torzító hatással bírnának az eredmény tekintetében: nem csak az ott lehangzottak piacra gyakorolt hatása kerülne bele az elemzésbe, hanem a váratlan ülés tényéből adódó heves piaci átárazások is.

A továbbiakban a dolgozat 48 darab nyilatkozat szövegével dolgozik, ezeknek az időpontjait az 1. táblázat tartalmazza. Az adatok másik részét a sajtótájékoztatók képezik.

A weboldalra feltöltött pdf fájlok kerültek letöltésre és feldolgozásra (Press Conference Transcript PDF). A korábban ismertetett okok miatt a 2020. március 3.-ai és 2020. március 15.-ei két kiadott sajtótájékoztató kikerült az elemzés keretéből. Pár alkalommal előfordult, hogy az FOMC adott ki nyilatkozatot, de nem került sor sajtótájékoztatóra, ilyen alkalom volt például 2020 március 23. Emiatt összesen 43 darab sajtótájékoztató szöveg kerül felhasználásra, ezek pontos időpontjait az 1. táblázat mutatja. A sajtótájékoztatók szövegei igen hosszúak, többször 30 oldalt is meghaladják. Ennek egyik oka, hogy tartalmazzák a sajtó munkatársainak a kérdéseit és az erre adott jegybankelnöki válaszokat is. A sajtótájékoztatók 43 szövegét a kutatás kettébontja: kérdések-válaszok nélküli és azokat tartalmazó részre. A továbbiakban ismertetett elemzéseket a dolgozat összesen 3 eltérő szöveghalmazon végzi el, egyrészt a kiadott 48 darab nyilatkozat szövegein, másrészt a 43 sajtótájékoztató teljes szövegén, harmadrészt ezen 43 sajtótájékoztató kérdésektől és válaszoktól megtisztított szövegein. Ezek alapján meghatározható, hogy a sajtó kérdései és az erre adott válaszok mennyire gyakorolnak nagy hatást az elemzés eredményére.

1. táblázat. Vizsgált szövegek időpontjai (a nyilatkozatok minden időpontban elérhetőek, a sajtótájékoztatók csak 5 esetben nem elérhetőek)

Dátum	Sajtótájékoztató	Dátum	Sajtótájékoztató
2018-01-31	Nincs	2021-01-27	
2018-03-21		2021-03-17	
2018-05-02	Nincs	2021-04-28	
2018-06-13		2021-06-16	
2018-08-01	Nincs	2021-07-28	
2018-09-26		2021-09-22	
2018-11-08	Nincs	2021-11-03	
2018-12-19		2021-12-15	
2019-01-30		2022-01-26	
2019-03-20		2022-03-16	
2019-05-01		2022-05-04	
2019-06-19		2022-06-15	
2019-07-31		2022-07-27	
2019-09-18		2022-09-21	
2019-10-30		2022-11-02	
2019-12-11		2022-12-14	
2020-01-29		2023-02-01	
2020-03-23	Nincs	2023-03-22	
2020-04-29		2023-05-03	

Dátum	Sajtótájékoztató	Dátum	Sajtótájékoztató
2020-06-10		2023-06-14	
2020-07-29		2023-07-26	
2020-09-16		2023-09-20	
2020-11-05		2023-11-01	
2020-12-16		2023-12-13	

Először a központosítások és a számjegyek eltávolítására kerül sor, ezt követik a stop szavak (SW) kiszedése. Ezen szavak kiszedése a szövegből a zaj csökkentéséhez vezethet az elemzés során, ezzel javítva az elért modell eredményét (Bholat et al. 2015). Mindegyik szöveg esetén 5 darab stopszavakat tartalmazó szótár értékeit távolítja el az elemzés. Ezek a szótárak a „stopwords” csomagból származnak (Benoit, Muhr, and Watanabe 2021). Az alapvető angol nyelvi stopszavakon túl a „nlk”, a „marimo”, a „snowball” és a „smart” szótárak szavai kerülnek törlésre (Benoit, Muhr, and Watanabe 2021). Ezen szótárak értékkészlete között jelentős átfedés van, ugyanakkor bizonyos szavakat csak egy-egy szótár tartalmaz. Az alaposság okán alkalmazza mindegyiket a dolgozat mindegyik szöveg esetén. A nyilatkozatoknál egy egyedi lista szavai is eltávolításra kerülnek. Ezek olyan szavak, melyek a kutatás szempontjából nem bírnak jelentéssel és nem kerültek eltávolításra a korábbi szótárak használata során. Ezek a szavak a 15. táblázatban szerepelnek. Több kifejezés is része a szövegnek, melyek eltérnek egymástól írásmódban, de jelentésük tekintetében azonosak. Egy adott szó és annak bármilyen ragozott formája külön szavaknak minősülnek az algoritmus számára, ugyanakkor jelentésben megegyeznek egymással, emiatt nem kezelésük jelentős zajhoz vezethet (Bholat et al. 2015). Ezt a problémát áthidalandó, a dolgozat szöveg sztemmelést végez minden kifejezésen. Ennek az eljárásnak az eredményeképpen minden kifejezés leválasztásra kerül mindenféle ragjaitól, jelzőitől és képzőitől. Így a korábban ragozott alakok és az eredeti szótári formájuk azonos kifejezésként kerülnek kezelésre. A stop szavaktól megtisztított és sztemmelt szövegeken a dolgozat létrehozza a dokumentum kifejezés mátrixot (document feature matrix, dfm). Ebben a lépésben kerülnek a szöveges adatok átalakításra szám adtokká, melyekkel az algoritmusok képesek dolgozni (Sebők, Ring, and Máté 2021). Ez a dfm 48 dokumentumot és 407 kifejezést rendel össze egymással a nyilatkozatok szövegei esetén. A topikmodellek létrehozását ugyanakkor nehezítik azok a kifejezések, melyek kevés dokumentumban fordulnak csak elő, hiszen nem segítik a topikok közötti megkülönböztetést. Ennek okán a dokumentum kifejezés mátrixból eltávolításra kerülnek azok a kifejezések, melyek 5 alkalomnál kevesebbszer szerepelnek vagy csak 1 dokumentumban találhatóak meg. Ezután a szűrés után kialakult dfm már csak 262 kifejezést tartalmaz. A megtisztított dfm

mátrixot használja fel a tanulmány a későbbi topikmodellezés során. Több topikszámmal is lefutatásra kerül a modell. A 4, 5 és 6 topikot tartalmazó modellek topikonkénti leggyakoribb 10 kifejezését rendre a 2. 3. és 4. táblázat mutatja.

2. táblázat. Leggyakoribb kifejezések topikonként - nyilatkozatok, $k = 4$

Topik	Leggyakoribb kifejezések (szótövesített formában)
1	committe, inflat, polici, econom, monetari, remain, rate, assess, addit, continu
2	committe, secur, market, feder, econom, inflat, support, condit, reserv, term
3	committe, inflat, employ, continu, polici, econom, expect, condit, support, goal
4	inflat, committe, rate, market, econom, expect, feder, condit, labor, strong

3. táblázat. Leggyakoribb kifejezések topikonként - nyilatkozatok, $k = 5$

Topik	Leggyakoribb kifejezések (szótövesített formában)
1	committe, inflat, polici, monetari, econom, remain, rate, assess, activ, continu
2	committe, secur, market, feder, econom, support, reserv, term, agenc, price
3	committe, inflat, employ, continu, polici, expect, econom, condit, goal, support
4	inflat, committe, rate, market, econom, expect, feder, labor, condit, strong
5	committe, inflat, econom, rate, pressur, monetari, polici, target, relat, secur

4. táblázat. Leggyakoribb kifejezések topikonként - nyilatkozatok, $k = 6$

Topik	Leggyakoribb kifejezések (szótövesített formában)
1	committe, inflat, polici, monetari, econom, remain, rate, assess, activ, continu
2	secur, market, feder, committe, agenc, reserv, mortgageback, support, desk, oper
3	inflat, committe, econom, rate, market, object, feder, remain, measur, labor
4	inflat, committe, expect, rate, market, econom, condit, feder, labor, strong
5	committe, inflat, econom, rate, pressur, monetari, polici, target, relat, secur
6	committe, inflat, employ, econom, polici, continu, expect, condit, economi, support

Látható, hogy a modelleken belül a topikok között van átfedés a kifejezésekben. A „committe” kifejezés például a 4 topikot tartalmazó modell esetén része mind az első,

mind a második mind a harmadik topiknak is, ráadásul a leggyakoribb kifejezésként került azonosításra (lásd 2. táblázat). Továbbá más átfedéseket is láthatunk a fenti ábrákon. A topikmodellezésben megengedettek ezek az átfedések topikok között; több olyan egymástól elhatárolható téma meghatározható, melyek azonos szavakat is gyakran használnak. Ugyanakkor a nagymértékű átfedések nehezítik a topikok kialakítását. A fenti példában szereplő „committe” kifejezés 4 topikból 3 esetén a leggyakoribb kifejezés, így nem segíti a különbségtételt a témák között. Ennek okán a dolgozat a topikokon átívelő szavak esetén szűrést végez, ezzel is hozzájárulva az eredményekben kialakuló torzítás visszaszorításához. A négy topikot tartalmazó modell esetén a leggyakoribb 30 kifejezést vizsgálva meghatározásra kerülnek azok a szavak, melyek mindegyik topikban szerepelnek. Hasonlóan, az 5 topikot tartalmazó modell esetén a legalább 4 topikban szereplő szavak, a 6 topikot tartalmazó modell esetén ugyancsak a legalább 4 topikban előforduló szavak kerülnek kigyűjtésre. Ezeket a szavakat mutatja az 5. táblázat.

5. táblázat. Topikok metszetében lévő szavak - nyilatkozatok

Topik darabszám	Kifejezések (szótövesített formában)
4	committe, inflat, econom, continue, condit, feder, expect, market
5	committe, inflat, polici, monetari, econom, remain, rate, assess, continue, secur, condit, market, feder
6	committe, inflat, polici, econom, remain, rate, assess, activ, continu, secur, condit, strong, target, market, feder, labor, expect, monetari

A 6 elemű modell esetén azok a szavak, melyek csak 3 topikban szerepelnek, még hasznosak lehetnek a modellezés során, így azok nem kerülnek eltávolításra. Topikszámtól függetlenül a legalább négy topikban is előforduló kifejezések kerülnek kigyűjtésre (lásd 2. táblázat). A 2. táblázatban szereplő 3 kifejezéseket tartalmazó lista közös elemei azok, melyeket a kutatás eltávolít az eredeti adatokból, tekintettel arra, hogy ezen kifejezések nem segítik a topikmodellezést, gyakori előfordulásuk korlátját képezi az elemzésnek. Az újbóli adattisztítás után újra futtatásra kerülnek a modellek. Az így kialakult topikok leggyakoribb kifejezéseit a 6., 7. és 8. táblázatok mutatja.

6. táblázat. Leggyakoribb kifejezések topikonként, tisztítás után - nyilatkozatok, k = 4

Topik	Leggyakoribb kifejezések (szótövesített formában)
1	expect, labor, goal, financi, rate, polici, strong, assess, support, busi
2	remain, rate, target, monetari, outlook, incom, financi, measur, account, polici

Topik	Leggyakoribb kifejezések (szótövesített formában)
3	rate, secur, assess, polici, employ, strong, support, agenc, monetari, develop
4	activ, remain, expect, monetari, employ, polici, assess, pressur, fund, price

7. táblázat. Leggyakoribb kifejezések topikonként, tisztítás után - nyilatkozatok, k = 5

Topik	Leggyakoribb kifejezések (szótövesített formában)
1	expect, labor, rate, polici, strong, financi, goal, assess, support, busi
2	remain, rate, target, monetari, polici, incom, secur, mortgageback, outlook, develop
3	rate, polici, secur, support, strong, assess, agenc, develop, monetari, achiev
4	remain, activ, polici, expect, monetari, assess, pressur, price, fund, monitor
5	employ, expect, monetari, goal, adjust, financi, outlook, measur, target, fund

8. táblázat. Leggyakoribb kifejezések topikonként, tisztítás után - nyilatkozatok, k = 6

Topik	Leggyakoribb kifejezések (szótövesített formában)
1	polici, monetari, remain, asses, rate, activ, addit, financi, develop, secur
2	secur, support, agenc, reserv, mortgageback, desk, oper, open, treasuri, repurchas
3	employ, polici, support, expect, goal, accommod, remain, secur, assess, financi
4	monetari, polici, rate, pressur, target, remain, secur, agenc, goal, addit
5	employ, price, expect, financi, public, health, develop, term, activ, polici
6	rate, expect, labor, strong, fund, object, remain, target, measur, assess

Természetesen a fenti gyakori szavak eltávolítása után újra futatott topikmodell is tartalmazhat olyan szavakat, melyek több topikban (akár mindegyikben) is a leggyakoribb kifejezések közé tartozik. A fenti szűrért többször is el lehet végezni, ugyanakkor túl sok kifejezés eltávolítása megint nehezíti az elemzést, emiatt a kutatás csak egyszer hajtja végre a gyakori szavak fent leírt tisztítását. A kutatás – ahogy a fenti ábrákon is látható – több eltérő számú topikból álló modellt is futtat. Fontos feladat kiválasztani, hogy a dolgozat a további elemzések során melyik modellt vegye alapul. Természetesen nem csak a 4, 5 és 6 topikból álló modellt lehet érdemes megnézni, ugyanúgy elvégezhető a fent leírt elemzés tetszőleges számú topik esetén. Ugyanakkor a jegybanki kommunikációban megjelenő témák száma jellemzően nem túl magas, és nem is csak egy-két téma azonosítható. Ebből a megfontolásból került a 4, 5 és 6 elemű modell lefutására. Az optimális topikszám

kiválasztásához számtalan módszer és mérőszám áll rendelkezésre. A dolgozat a perplexity mutatószámot határozta meg az eltérő modellekre. A mutatószám azt vizsgálja, hogy a modell mennyire érzékeny az új adatokra, ezáltal a modell robusztusságát képes mérni, ahol a kisebb érték stabilabb modellt jelent (Kapadia 2022). A tisztított adatok véletlenszerűen kerültek 75-25% mentén egy tanuló és egy tesztelő adathalmazra szétbontva. A tanuló adatokon elkészített LDA modellek és a maradék 25%-nyi tesztelő adat segítségével kerültek meghatározásra a perplexity mutatók. A 9. táblázat szemlélteti a perplexity mutatók értékeit.

9. táblázat. Perplexity értékek az egyes modellek esetén - nyilatkozatok

	k = 4	k = 5	k = 6
Perplexity érték	244.4062	244.4352	153.9429

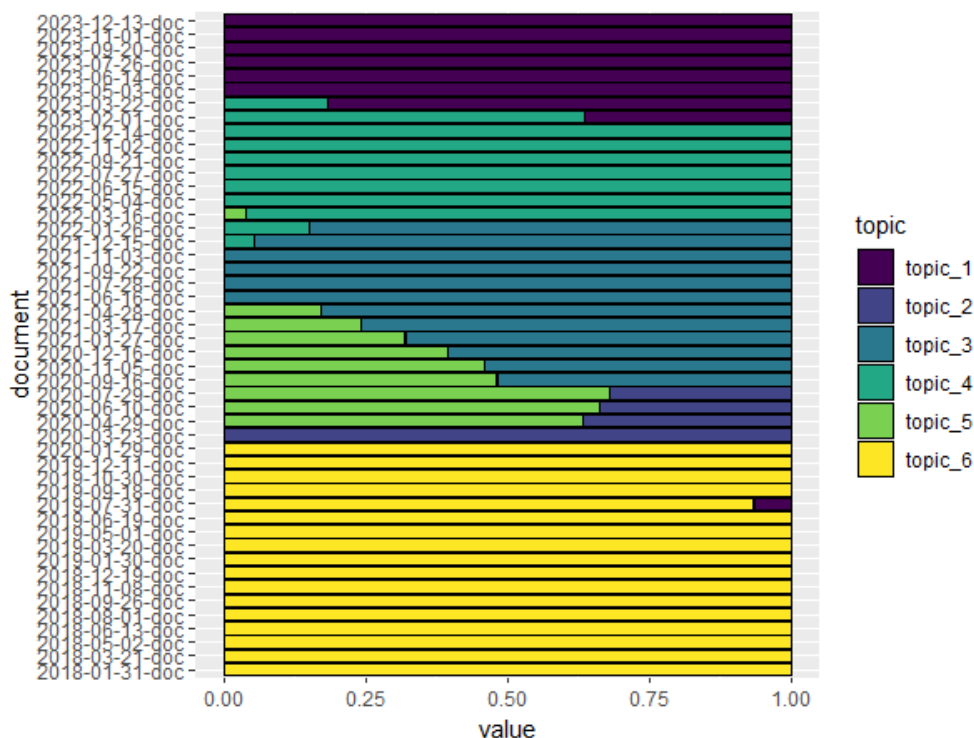
A tanuló és tesztelő adatokra történő szétbontás véletlenszerűen történik, emiatt minden futtatás más értékeket ad, ugyanakkor az értékek közötti sorrend nem változik új realizációk esetén se. Ahogy a 9. táblázatban is látható, a 6 topikból álló modell az, amelyik a legkevésbé érzékeny az új adatokra, ez alapján ezt a modellt lenne célszerű választani. Ugyanakkor a kialakult topikok felcímkézése, vagyis annak meghatározása, hogy adott topik milyen gazdasági témához tartozik, emberi feladat. Könnyen lehet, hogy egy adott mutató szerinti ideális topikszám nem összehangolható az emberi megítéléssel, annak felcímkézése problémákba is ütközhet. A topikokhoz értelmes témát rendelni rendkívül nehéz feladat, melyet csak emberi megítélés képes elvégezni. A korábbi táblázatok szereplő topikokba eső leggyakoribb kifejezések alapján a 10. táblázatban látható felcímkézést végzi el a dolgozat. Az indoklás oszlopban olyan gyakran előforduló kifejezések láthatóak, melyek az adott témakörhöz köthetőek. A 6. táblázat alapján egyéb felcímkézés is elvégezhető, ugyanakkor a kutatás a 10. táblázatban szereplő értékeket veszi alapul a későbbiek folyamán.

10. táblázat. Nyilatkozatokon végzett modellezés eredményeinek interpretálása

Topik	Megnevezés	Kapcsolódó kifejezések
1	Növekedés	Assessment, Remain, Active, Develop
2	Mennyiségi lazítás, Bankrendszer	Repurchase, Mortgagebacked
3	Munkaerőpiac	Employment, Goal, Policy
4	Monetáris politika	Monetary, Policy, Rate, Target

Topik	Megnevezés	Kapcsolódó kifejezések
5	Árstabilitás	Price, Financial
6	Előretekintés, várakozások	Expect, Object, Target, Remain

A kutatási kérdés tekintetében a lényeges információ, hogy milyen arányban oszlanak meg a topikok egyes szövegek esetén. Erre az LDA modell gamma paramétere ad választ. A topikok megoszlását az egyes szövegek esetén az 1. ábra mutatja.



1. ábra. Topikok megoszlása adott szövegek esetén - nyilatkozatok

Ahogy az 1. ábrán is látható több olyan dokumentum van, ahol a teljes szöveg egy adott topikhoz tartozik. Ez részben a rendelkezésre álló adatokkal magyarázható. A jegybanki kommunikációban előforduló témakörök között jelentős átfedés figyelhető meg a felhasznált szavak tekintetében. A példa kedvéért érdemes a ráta kifejezést elővenni, mely rengeteg kontextusban gyakran előfordulhat, mind inflációs ,- kamat,- de még munkanélküliségi rátaként is használható. A modellezésnek ez a fajta nehézsége visszaköszön az 1. ábrán látható megoszlásokon is.

A fent leírt módszereknek megfelelően dolgozza fel a kutatás a FED sajtótájékoztató szövegeit is. A stopszavak eltávolítása során az 5 korábban ismertetett szótár szavain túl egy egyedi lista értékei is eltávolításra kerülnek, ezek a szavak megtalálhatóak a 16. táblázatban. Eltávolításra kerülnek azok a kifejezések is, melyek a dfm mátrix szerint 5-nél

kevesebbszer vagy csak 1 dokumentumban fordulnak elő. A túl gyakran előforduló szavak kezelése miatt a topikok között átfedő kifejezések is eltávolításra kerülnek, hasonlóan, mint a nyilatkozatoknál. Ezeket a kifejezések mutatja az 18. és 19. táblázat. Jelen szövegek esetén is kiszámolásra került a perplexity mutatók értéke az eltérő topikszámokra, ezeket az alábbi táblázat mutatja (lásd 11. táblázat).

11. táblázat. Perplexity értékek az egyes modellek esetén - sajtótájékoztatók

Perplexity érték	k = 4	k = 5	k = 6
Teljes szöveg	917.1174	911.6247	905.186
Kérdésektől és válaszoktól mentes szöveg	352.628	337.2197	343.7763

A sajtótájékoztató szövegek kérdésektől megtisztított és azokat tartalmazó szövegeire elkészített LDA topikmodellek eredményeit a 17. táblázat ábrái mutatják. Minden topikba eső leggyakoribb kifejezések láthatóak az adott szövegre futatott modellek esetén. A perplexity értékek alapján a teljes szöveg esetén a 6, míg a kérdéseket nem tartalmazó szövegek esetén az 5 topikból álló modellek azok, melyek a legkevésbé érzékenyek az új adatokra (lásd 11. táblázat).

A 11. táblázat alapján rendel témákat a dolgozat az egyes szövegek 5 és 6 topikból álló modelljéhez (lásd 12. és 13. táblázat).

12. táblázat. Teljes sajtótájékoztatókon végzett modellezés eredményeinek interpretálása

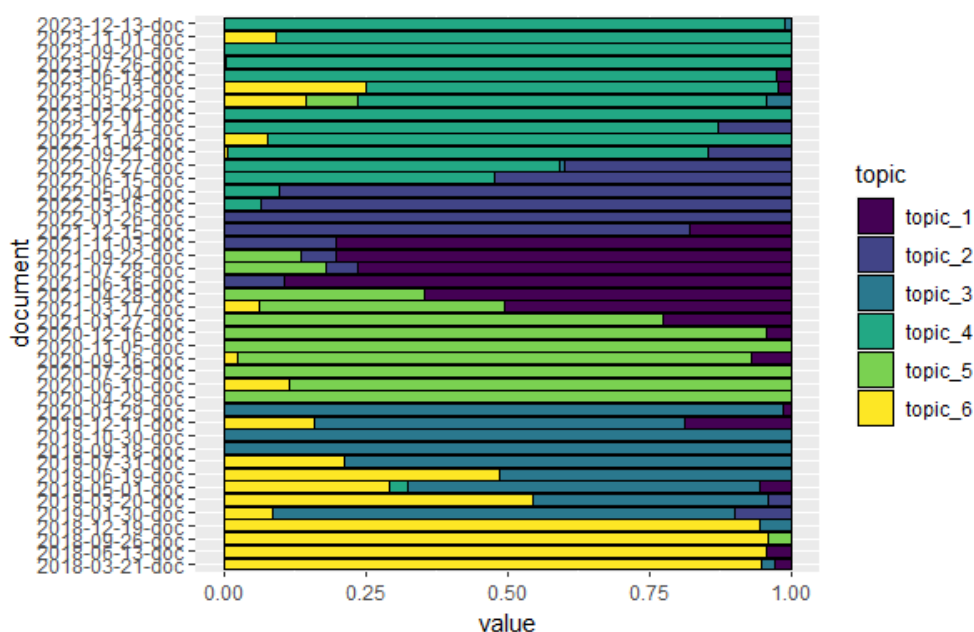
Topik	Megnevezés	Kapcsolódó kifejezések
1	Eszközvásárlás	taper, continu, price
2	Gazdasági makró helyzet	price, balance, sheet, demand, suppli
3	Monetáris politika	reserv, monetari, cut, fed
4	Növekedés	Growth, Forecast, Data, Condit
5	Munkaerőpiac	Work, Job, Unemployment
6	Bankrendszer	Bank, interest, project

13. táblázat. Kérdésektől szűrt sajtótájékoztatókon végzett modellezés eredményeinek interpretálása

Topik	Megnevezés	Kapcsolódó kifejezések
1	Árstabilitás	Price, high, increas, target

Topik	Megnevezés	Kapcsolódó kifejezések
2	Munkaerőpiac	employ, goal, progress, support, labor
3	Nem meghatározható, maradék kifejezések összessége	
4	Gazdaság támogatása	support, recoveri, employ, credit, provid, spend
5	Növekedés	growth, outlook, support

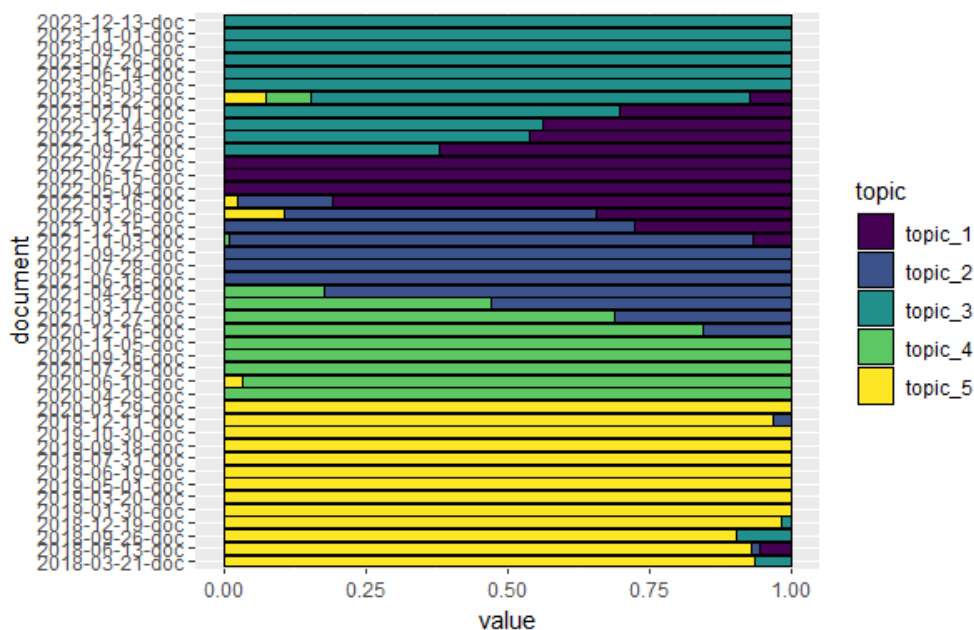
A 2. és a 3. ábra mutatja a sajtótájékoztatók szövegein a topikok megoszlását.



2. ábra. Topikok megoszlása adott szövegek esetén - sajtótájékoztatók teljes szövege

2.1.1. Identifikáció

Érdeemes kitérni a topikok eloszlását mutató 1., 2. és 3. ábrák sajátosságaira. Megfigyelhető az ábrák mindegyikén egyfajta idővariánság, vagyis a topikok időponttól függő jellege. Látható a korábbi ábrákon, hogy egyes topikok csak egy adott időpontban fordulnak elő, más beszédek alkalmával nem szerepelnek. Ez szembe megy a várakozásokkal, hiszen egy jegybanki kommunikációban előforduló témák rendszerint minden beszédben előfordulnak, hol kisebb, hol nagyobb súllyal. Ennek a problémának a kezelése érdekében a dolgozat a - fentebb leírt módokon tisztított - szövegeken struktúrált topikmodelleket futtat. A struktúrált topikmodellek (Structural Topic Model, STM) metaadatokkal képesek



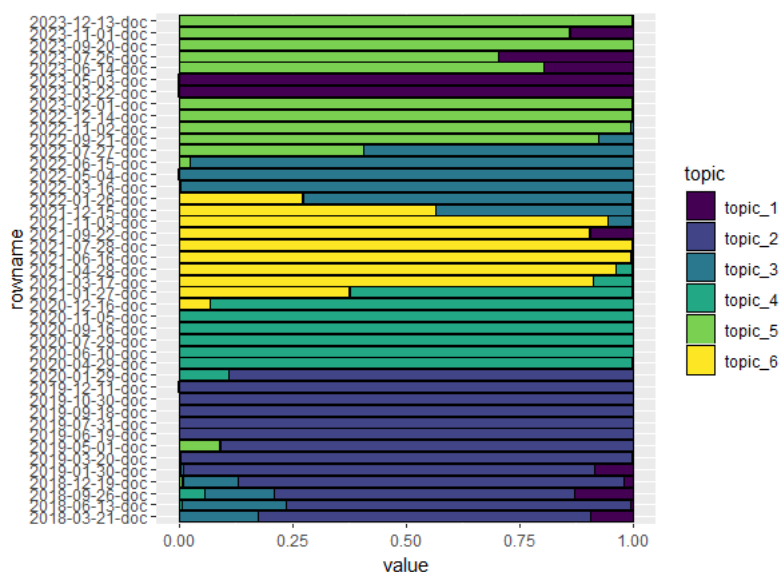
3. ábra. Topikok megoszlása adott szövegek esetén - sajtótájékoztatók kérdésektől szűrve

bővíteni az LDA modelleket, így megoldást tudnak adni a klaszterezettség problémájára (Roberts, Stewart, and Tingley 2019). A dolgozat által futatott STM modellek a dátum értékek mint metaadatok felhasználásával kerültek futtatásra. A 3 szövegen elkészített STM modellek összefoglaló táblázatát a 21. táblázat mutatja.

A sajtótájékoztatók teljes szövegein lefutott struktúrált topik modell esetén kialakult topik eloszlásokat a 4 ábra mutatja. A nyilatkozatok (lásd 11. ábra) és a szűrt sajtótájékoztatók szövegein kialakult eredmények (lásd 12. ábra) a mellékletben láthatóak. Az ábrák alapján az alábbi következtetést vonja le a tanulmány: a klaszterezettség nem szűnt meg, továbbá számottevő változás se figyelhető meg. Jelen modellek esetén ugyanaz a topik szám és stopszó eltávolítást végezte a dolgozat minden szöveg esetén mint az LDA modelleknél, ugyanakkor a topikok közötti összerendelés nem lehetséges a két modell esetén. Az STM és LDA modellek topikjai eltérőek, így az első topik mást jelent az egyes modellek esetén.

Tekintettel arra, hogy a klaszterezettség problémáját a stuktúrált topik modellek esetén se sikerült megoldani, a dolgozat a továbbiakban a korábban részletesebben ismertetett LDA modellek eredményeivel halad tovább. Fontos azonban kiemelni, hogy a dolgozat korlátjai közé tartozik azonban a klaszterezettség problémája. Ennek kezelése túlmutat a kutatás jelenlegi keretein, így a dolgozat ennek a problémának a kezelésével nem foglalkozik a továbbiakban, azt egy esetleges továbbfejlesztési irányként kezeli.

Összesen három eltérő szöveghalmazon történt szövegtisztítás; különös tekintettel a



4. ábra. Topik eloszlások strukturált topik modell esetén - sajtótájékoztatók teljes

stop szavakra és a legritkábban előforduló kifejezésekre helyezett hangsúlyt. A szövegek topikokra történő szétválasztását támogatandó, azok a kifejezések is eltávolításra kerültek a szövegekből, melyek több topik esetén is a leggyakoribb kifejezések voltak. Három eltérő topikszámú modell került futtatásra. Ezek közül a dolgozat a perplexity mutató alapján választott modellt. A választott modellek esetén minden topikhoz jegybanki kommunikációs témakört is rendelt a dolgozat. Meghatározásra került, hogy milyen eloszlása van az adott beszédeknek a topikok között. Az így kialakult klaszterezettséget strukturált topikmodellek alkalmazásával igyekezett kezelni a dolgozat, sikertelenül. Emiatt az LDA modellek eredményeit veszi alapul a dolgozat a későbbiek folyamán. A következő fejezetben a piaci mozgásokkal foglalkozik a dolgozat, majd az ebben a fejezetben látott eredmények és a piaci mozgások összekapcsolására kerül sor a későbbiekben.

2.2. Tőzsdei árfolyamok

Számtalan adat felhasználható a piaci mozgások elemzésére. Tekintettel arra, hogy dolgozat az amerikai jegybank kommunikációjára fókuszál, ezért a New York Stock Exchange-en vezetett Standard & Poors indexet (SPX) használja a kutatás. Azon túl, hogy nagy likviditással rendelkezik az index, rendelkezésre áll perces felbontású adat is az árfolyamok alakulásáról. Az árfolyamok közül a close bid árak kerülnek felhasználásra 2018. január 30. és 2023. március 31. közötti intervallumban. Az intervallum kezdeti értéke az első vizsgált nyilatkozat előtti nap, míg az intervallum vége az az időpont, ameddig rendelkezésre álltak a szükséges adatok. Az óráátállítás miatti egy órás eltolódás miatt a 6 óra 30 perces nyitvatartási ablak (Core Trading Session) helyett a kutatás 7 óra 30 perces

ablakot használ. Ebben az ablakba időponttól függetlenül belesik mind a nyitás mind a zárás is. Fontos megemlíteni, hogy a teljes dolgozat UTC-be átalakított adatokat használ, így az SPX adatok is UTC-ben szerepelnek. A 14. táblázat szemlélteti a felhasznált adatok kezdetét és végét.

14. táblázat. A felhasznált adatok időbeli eloszlásának összegzése

Esemény	Dátum	Idő
SPX adatok indulása	2018. 01. 30.	9:30 (EST), 14:30 (UTC)
1. Nyilatkozat megjelenése	2018. 01. 31.	14:00 (EST), 19:00 (UTC)
1. Jegybanki beszéd megtartása	2018. 03. 21.	14:30 (EDT), 18:30 (UTC)
SPX adatok vége	2023. 04. 01.	15:59 (EDT), 19:59 (UTC)
48. (utolsó) nyilatkozat megjelenése	2023. 12. 13.	14:00 (EST), 19:00 (UTC)
43. (utolsó) jegybanki beszéd megtartása	2023. 12. 13.	14:30 (EST), 19:30 (UTC)

A 14. táblázatban látható, hogy a nyilatkozatok 14:00-kor jelennek meg minden alkalommal, míg a beszédek fél órával később tartják. Fontos azonban kitérni az óraátállásra. A 48 nyilatkozatból 14 EST (Eastern Standard Time), míg 34 EDT (Eastern Daylight Time) időzónában szerepel. Mivel a dolgozat minden adatot UTC-re alakít át, így EDT esetén 4, míg EST esetén 5 órával szükséges korrigálni az értékeket. A nyilatkozatok és beszédek UTC-be átalakított értékeit a mellékletben található 20. táblázat mutatja. A 14. táblázatban látható, hogy a tőzsdei nyitvatartási időszakhoz történik az adatok igazítása. A 9:30-as nyitás 2018. január 30.-án EST zónába esett, így 5 órával szükséges korrigálni, azonban a 16 órás zárás 2023. március 31.-én csak 4 órás korrigálást vont maga után.

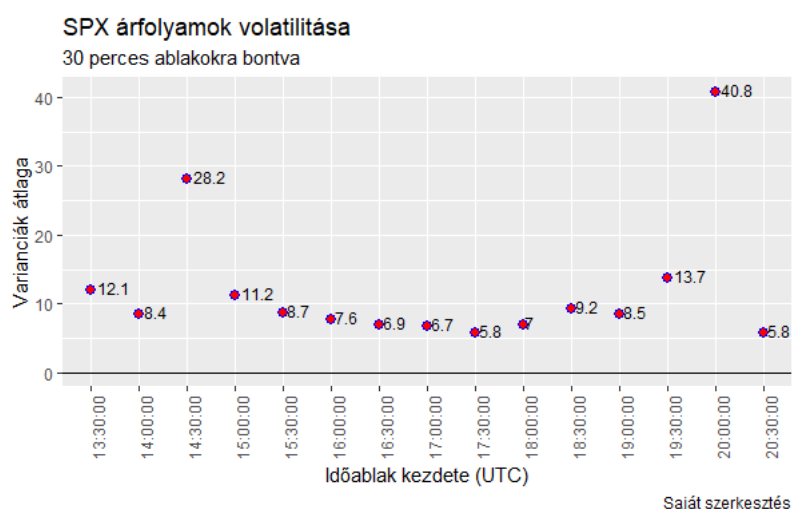
Az 1. és a 14. táblázatból következik, hogy van 6 olyan nyilatkozat és beszéd, melyek időpontjaira nincs SPX adat. Emiatt a kutatás csak 37 beszéd és 42 nyilatkozat hatását tudja vizsgálni a piacra vonatkoztatva. A topikmodellek robusztussága miatt azonban a korábbi fejezetbe szándékosan kerül felhasználásra mindegyik szöveg.

2.2.1. Leíró statisztika

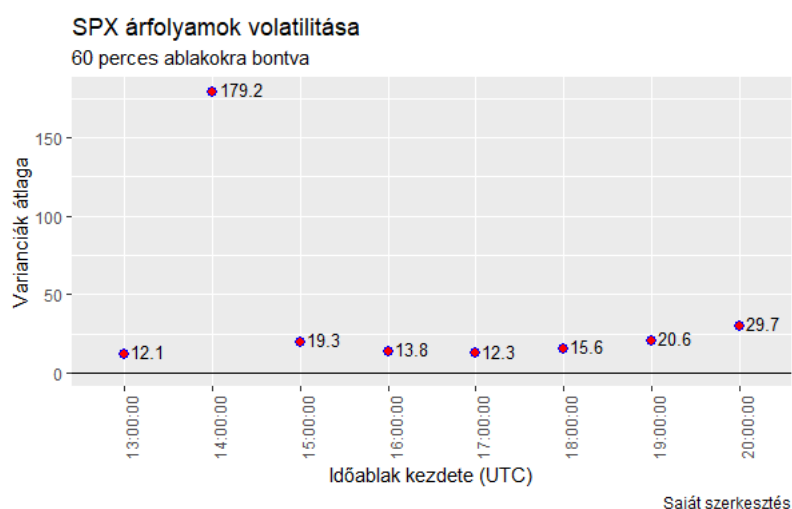
A kutatási kérdés arra fókuszál, hogy a beszédek alkalmával más-e a varianciája az SPX adatoknak. Elsőként fontos leíró statisztika segítségével megnézni, hogy mi jellemző

az adatokra, leginkább azok varianciájára. A 14. táblázatban szereplő intervallumon belül és a 13:30 és 21:00 intervallumon belül perces felbontásban összesen 849 150 darab SPX árfolyam megfigyelési adat áll rendelkezésre. Ezek az adatok UTC-ben szerepelnek, vagyis óraáttállítás nem érinti őket.

Egy kutatás szerint a nyilatkozatok kiadása után 40 percig szignifikánsan magasabbak a részvények hozamai mint más napokon, emiatt fontos lehet eltérő hosszúságú ablakokat is felhasználni az elemzés során (Rosa 2011). Mindegyik megfigyelés esetén meghatározásra kerül az a 30 és 60 perces időablak, melybe beletartozik. Minden napon belül így összesen 15 darab 30 perces és 8 darab 60 perces ablak található. Az SPX closebid árfolyamok varianciáját határozza meg a dolgozat minden ablak esetén, majd minden nap azonos ablakainak értékeit átlagolja (lásd 5., 6. ábra).



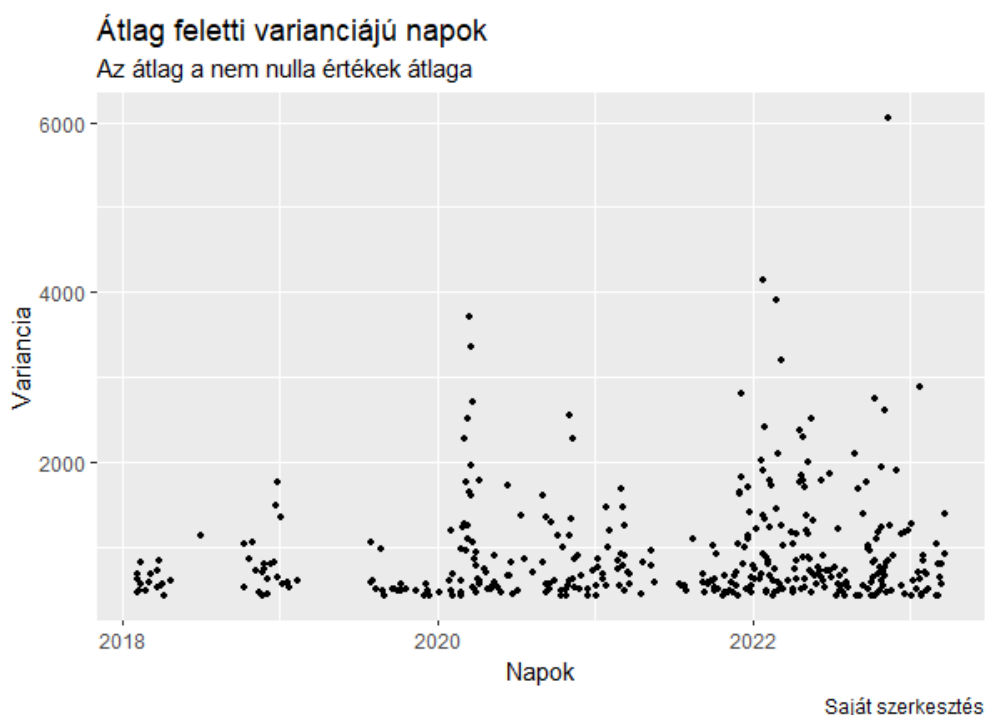
5. ábra. Adatok volatilitása 30 perces ablakokban



6. ábra. Adatok volatilitása 60 perces ablakokban

Első ránézésre látható pár kiugró érték, leginkább 14:00-14:30 és 20:00 magasságában. Ezek az adatok UTC-ben szerepelnek, továbbá egy adott esemény a dátumától függően EDT-ben és EST-ben is szerepelhet, vagyis mind a 4, mind az 5 órai korrekció elképzelhető. A 9:30-as New Yorki tőzsdei nyitás után kialakuló árfolyam volatilitás 13:30-nak vagy 14:30-nak is megfeleltethető UTC-ben kifejezve. Emiatt egy rendszeresen bekövetkező esemény (tőzsdei nyitás, zárás) a fenti ábrákon több megfigyelésben is szerepelhet. A kiugró értékek magyarázhatóak statisztikai adatok közzétételével, jegybanki kommunikációval vagy akár külső sokkokkal is (például Covid, háború). A 14 óra magasságában megfigyelt kilengés mögött feltehetően a tőzsdei nyitáson túl a munkanékülségi segélyre előző héten újonnan jelentkezők száma (initial claims) is meghúzódikék, mely jellemzően 8:30-kor kerül nyilvánosságra (U. S. Bureau of Labor Statistics, n.d.).

A 7. ábra szemlélteti, hogy a vizsgált időszakban több olyan nap is volt, ahol jelentős variancia figyelhető meg napon belül. Az ábrán 377 nap adata látható, tekintettel arra, hogy a nem nulla varianciájú napok közül csak az átlag (412,6439) feletti varianciájú napok kerültek ábrázolásra. Látható, hogy az adatok döntő többsége 1000 alatt marad, kevés nap figyelhető meg jelentősen magasabb varianciával (lásd 7. ábra).

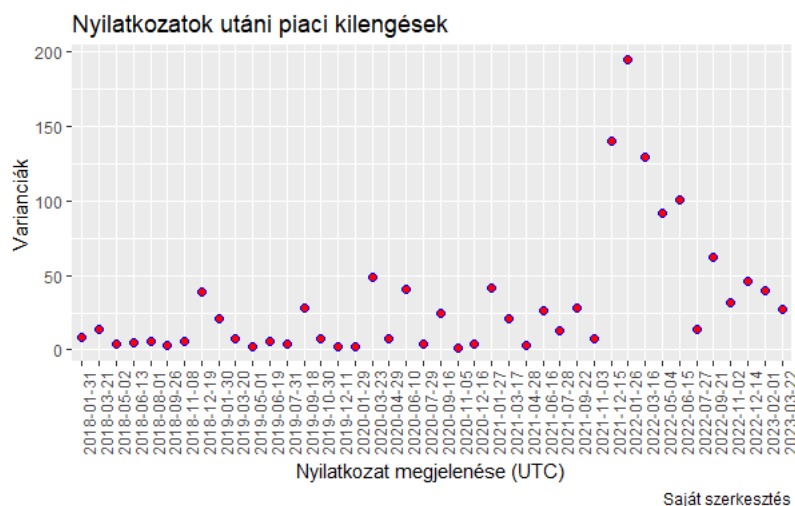


7. ábra. Árfolyamok varianciája az egyes napokon

A 7. ábrán látható legnagyobb varianciájú napok közé tartozik 2022. 11. 10.-e (amerikai félév választás), 2022. 01. 26. (kamatemelési ciklus elkezdése), 2022. 02. 25. (orosz-ukrán háború kitörése) illetve 2020. 03. 16. és 2020. 03. 20. (Covid lezárások). Látható,

hogy a nagy mozgások mögött jelentős események állnak, melyek indokolják a napon belüli nagy elmozdulásokat az árfolyamban.

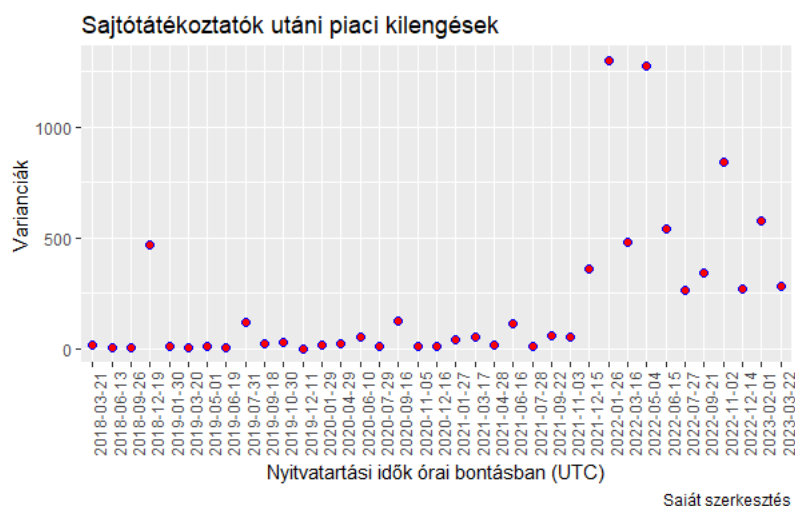
A kutatási kérdés tekintetében fontosabb azonban, hogy a beszédek után milyen variancia tapasztalható. Mivel a nyilatkozatok kiadása után fél óra eltéréssel kezdődik a sajtótájékoztató, így a nyilatkozatok esetében egy fél órás ablakot vizsgál a dolgozat. Ennek köszönhetően nem kerül figyelembevételre a sajtótájékoztató okozta piaci mozgás. A 30 perces ablakok varianciáit a 8. ábra szemlélteti mind a 42 esetben.



8. ábra. Nyilatkozatok utáni fél óra varianciái

A 9. ábra szemlélteti a helyi idő szerint 14:30-kor kezdődő sajtótájékoztatók (37 darab) utáni 60 percben az SPX varianciáit. Mivel a nyilatkozat és a sajtótájékoztató időben igen közel van egymáshoz, így átfedhetnek a hatások egymással, emiatt részben torzítottak lehetnek a sajtótájékoztató ideje alatt tapasztalt varianciák.

Összeségében elmondható, hogy mind a 30 mind a 60 perces ablakokat vizsgálva pár jelentősen kiugró érték megfigyelhető az SPX varianciájában. Ezek mögött alapvetően két tényező húzódik meg, melyek a fenti ábrákon is láthatóak. Elsőként azonosíthatóak olyan tényezők, melyek rendszeresen jelennek meg mindig azonos időben napon belül. Ilyen tényezők lehetnek a kiadott új statisztikai adatok (GDP, infláció, munkanélküliségi segélyre újonnan jelentkezők száma stb.). Ezek a tényezők - tekintve hogy mindig azonos időben jelennek meg napon belül - mindig ugyanabban a 30 illetve 60 perces ablakban kerülnek nyilvánosságra, így az általuk hordozott új információ mindig ugyanabban az ablakban okoz kiugró varianciát. A másik tényező ami a kiugró varianciák mögött meghúzódik egyedi események, melyek nem ismétlődő jellegűek. A fenti ábrákon látható, hogy a napok között is jelentős eltérés tapasztalható a varianciák tekintetében. Ennek okai olyan események, melyek csak az adott napra jellemzőek. Ilyen események az amerikai félidős választások, a



9. ábra. Sajtótájékoztatók utáni egy óra varianciái

kamatemelési ciklus elkezdése, a covid lezárások vagy az orosz-ukrán háború kitörése.

3. Módszertan

A jegybanki kommunikációkról és azok hatásairól számtalan cikk és tanulmány született az elmúlt években, évtizedekben. Ezen munkákat sokban megkönnyítette, hogy a jegybanki kommunikáció érthetőbb és transzparensőbb lett az évek során (Blinder et al. 2008). A jegybanki közlemények és beszédek a piaci szereplők nagy figyelmét élvezik minden alkalommal, hiszen gyakran ilyenkor szerez tudomást új információkról a piac. Ennek okán nem meglepő hogy a jegybanki kommunikációknak hatása van a piacra (Rosa 2011).

Egészen széles azon eszközök köre, melyre hatással van egy jegybanki kommunikáció. Fiser és Horváth (2010) azt találta, hogy a kis nyitott cseh gazdaságban a devizaárfolyamra jelentős hatásal tud lenni a jegybank mindössze kommunikáció révén. Amerikában a FED által kiadott nyilatkozatok szignifikánsan képesen befolyásolni az eszközök árának volatilitását (Sack and Kohn 2003). Ezen nyilatkozatok tőzsdei árakra gyakorolt hatása olykor jelentősebb, mint maga a monetáris politikai döntések (Rosa 2011). A kommunikáció során szerzett váratlan, meglepetés szerű információ is fontos hatásokkal bírhat. Egy 25 bázispontos váratlan kamatvágás 1%-pontos tőzsdei index emelkedést is okozhat (Bernanke and Kuttner 2005). A kapcsolat a jegybanki kommunikáció és a piac között nem csak cseh vagy amerikai jelenség. Egy kutatás szerint a Bank of England által kiadott közlemények szignifikánsan képesek befolyásolni a piaci kamatvárakozásokat (Reeves and Sawicki 2007).

A jegybanki kommunikáció és a piac közötti kapcsolat tehát igazolt. Ugyanakkor a jegybanki kommunikációs eszközök között jelentős eltérés is mutatkozhat. Különböző jegybankok kommunikációjának összehasonlítása az eltérő keretrendszer, mandátum és kommunikációs stratégia miatt nem egyszerű feladat (Reeves and Sawicki 2007). Jegybankon belül is sok eltérés mutatkozhat a kommunikációs eszközök területén. (Cieslak and Schrimpf 2019) tanulmánya szerint a monetáris és nem monetáris hírek más-más kommunikációs csatornán jellemzőek. A szerzők szerint a 2008-as válságban a nem monetáris hírek voltak piacmozgató hatásúak, és csupán 2013 után lett egyre nagyobb hatásúak a monetáris politikai hírek (Cieslak and Schrimpf 2019). A jegybanki kommunikáció - azon túl hogy új információt tud biztosítani a piac számára - nem mindig képes csökkenteni a piaci zajt (Blinder et al. 2008). Pedig több tanulmány azt találta, hogy a több, jobb és érthetőbb jegybanki kommunikáció képes csökkenteni a piacon tapasztalt volatilitást, zajt (Blinder et al. 2008), (Jansen 2011).

A kommunikáció és a piac közötti kapcsolat meglétén túl a monetáris politika előrejelzése is sokat nőtt (Blinder et al. 2008). Az előrejelzés gyakran fontosabbá válik mint a jelenlegi gazdasági helyzet értékelése. 18 év amerikai adatait felhasználva az derült ki, hogy a jegybanki előrejelzések a jövő kamatpályáját illetően fontosabbak mint a jelenlegi gazdasági állapotról szóló hírek a piaci mozgások tekintetében (Hansen and McMahon 2016).

Látható tehát, hogy számtalan tanulmány foglalkozott a jegybanki kommunikáció milyenségével, piacra gyakorolt hatásával. Globális jelenség, hogy a jegybanki kommunikáció más eszközökkel és csatornákkal ugyan, de számtalan piaci eszközre képes hatást gyakorolni, legyen szó devizaárfolyamról, tőzsdei árakról vagy épp azok volatilitásáról.

4. Eredmények

A tanulmány arra a kérdésre keresi a választ, hogy meghatározható e olyan téma, mely szignifikánsan képes befolyásolni a piac varianciáját. A korábbiakban a dolgozat 3 eltérő szövegen végzett LDA eljárás segítségével modelleket, melyek segítségével meghatározásra kerültek az egyes témák és azok megoszlása a szövegek között. Ezt követően az SPX adatokból kiszámolt varianciák is meghatározásra kerültek. Jelen fejezet ennek a két részeredménynek a kapcsolatát bontja ki.

Az 1., 2. és 3. ábrán látható topikarányokkal magyarázza a varianciák alakulását a dolgozat regresszió segítségével. A dolgozat 5%-os szignifikancia szinttel dolgozik. Fontos kiemelni, hogy a varianciák nem normál eloszlást követnek (lásd 13. ábra), így

a regresszió eredményében szereplő t próbafüggvény értékek torzítottak, ugyanakkor a koefficiens értékek torzítatlanok maradnak.

A 3 szövegen kialakult topikarányokkal futatott lineáris regressziók eredményeit a 10. ábra mutatja. Az első és a második modell esetében az eredményváltozó a 9. ábrán is szereplő varianciák, amik 60 perc varianciáját tartalmazzák. A nyilatkozatok esetében a modell a 8. ábra varianciáit magyarázza a topikarányokkal. Az 5 és 6 topikot tartalmazó modellek a regresszióban csak 4 és 5 magyarázó változót jelentenek a magyarázó változók között fennálló függvényszerű kapcsolat miatt.

Látható, hogy a nyilatkozatok esetében a legalacsonyabb a korrigált R^2 mutató értéke. Ennek oka lehet, hogy a nyilatkozatok jelentősen rövidebb szövegek, így a topikmodell eredménye nem olyan robosztusak mint a sajtótájékoztatók esetében. A sajtótájékoztatók esetén látható, hogy a kérdések és válaszok kiszűrése a szövegekből kisebb determinisztikus együtthatót eredményezett. Ez magyarázható azzal, hogy a hosszabb szöveg több információt tartalmazott, így tudott a modell magasabb magyarázóerőt elérni.

A sajtótájékoztatók teljes szövegén végzett elemzés azt mutatja, hogy a makrogazdasági helyzetről és a gazdasági növekedésről szóló témák (lásd 12. táblázat) súlyának emelkedése a beszédben növeli az SPX index varianciáját a beszéd kezdetekor induló 60 perces ablakban (lásd 10. ábra). Mindkét magyarázóváltozó béta értéke pozitív előjelű, továbbá lényegesen magasabb mint a többi béta együttható. Ezek az eredmények nem nevezhetők meglepőnek. Ezen témák beszédben betöltött arányának emelkedése azt mutatja, hogy a megszokottnál több szó esik a jegybanki kommunikáció során erről a kettő témáról. Ez maga után vonhat egy vártnál jobb vagy rosszabb jelenlegi vagy jövőbeli gazdasági helyzet kialakulását, mely természetes módon párosulhat magasabb piaci volatilitással.

A sajtótájékoztatók kérdéseket és válaszokat nem tartalmazó szövegein kialakult eredmények mást mutatnak. A második regressziós modell azt jelzi, hogy az árstabilitás beszédben betöltött súlya pozitív kapcsolatban van a piaci varianciával. Ennek megfelelően egy 1 %-pontos emelkedés az árstabilitás témájában 681-el növeli meg a piacon megfigyelt varianciát a vizsgált 60 perces időintervallumban.

A nyilatkozatok szövegeit vizsgálva szignifikáns kapcsolat a 3. és a 4. topik esetén mutatkozik meg, ezek a munkaerőpiac és a monetáris politikai témák (lásd 10. táblázat). Ezen változók együtthatói is pozitív előjelűek. Fontos kitérni rá, hogy a nyilatkozatok szövegei a legrövidebbek, emiatt ebben a modellben tapasztalható a legkisebb sztochasztikus hiba a maradékok esetén. A sajtótájékoztatók során nagyságrendileg magasabb hiba figyelhető meg a maradékok esetén.

A kutatási kérdése a dolgozatnak az, hogy meghatározhatóak-e olyan jegybanki

Regressziók eredményei			
	Szövegek		
	Sajtótájékoztatók teljes (1)	Sajtótájékoztatók szűrt (2)	Nyilatkozatok (3)
topik_1	-165.79 (150.88)	681.29*** (141.60)	2.40 (42.25)
topik_2	737.05*** (143.76)	96.17 (127.75)	48.97 (35.72)
topik_3	-119.69 (146.05)	274.46 (210.62)	44.30*** (16.14)
topik_4	304.01** (142.12)	-32.40 (122.39)	61.35*** (16.22)
topik_5	-75.91 (129.82)		-28.32 (30.58)
Konstans	119.34 (98.96)	57.26 (70.80)	9.70 (8.88)
N	37	37	42
R ²	0.68	0.50	0.34
Korrigált R ²	0.63	0.43	0.25
Maradékok elo. hiba	201.14 (df = 31)	249.71 (df = 32)	36.46 (df = 36)
F Statisztika	13.35*** (df = 5; 31)	7.86*** (df = 4; 32)	3.79*** (df = 5; 36)

10. ábra. Regressziók eredményei

kommunikációban szereplő témák, melyek magasabb volatilitáshoz vezetnek. A tanulmány 6 év adatain 3 különböző szövegen végzett többféle tisztítási eljárást, majd LDA modellek és perplexity mutatók segítségével határozta meg az FOMC kommunikációjában szereplő témákat és azok megoszlását az egyes beszédek alkalmával. Lineáris regresszió segítségével vizsgálta a dolgozat a topik arányok és a kommunikáció megjelenése utáni 30 illetve 60 perc SPX index varianciája közötti kapcsolatot. Mindegyik szöveg esetén 5%-os szignifikancia szint mellett meghatározható olyan téma, mely szignifikánsan képes befolyásolni (növelni) a piacon tapasztalt varianciát.

5. Konklúzió

Jelen tanulmány a jegybanki kommunikációkkal és azok piacra gyakorolt hatásaival foglalkozik. Számptalan tanulmány készült, mely az amerikai jegybank kommunikációját és az amerikai tőkepiac kapcsolatát vizsgálta. Jelen kutatás a jegybanki kommunikációban megjelenő témákkal, más néven topikokkal foglalkozik. Ezen topikok azonosítása révén a dolgozat arra keresi a választ, hogy meghatározhatóak e olyan témák, melyek szignifikánsan képesek befolyásolni a piacon tapasztalt volatilitást.

A dolgozat az FOMC által 2018 és 2023 között kiadott nyilatkozatok és a kamatdöntő ülések után tartott sajtótájékoztatók szövegeit használta fel a kutatás során. Ezen 6 év alatt

rendelkezésre álló szöveg előkészítése után látens Dirichlet allokáció (LDA) segítségével azonosítja a szövegekben fellelhető témákat és azok megoszlását az egyes szövegek esetén. Minden szöveg esetén a perplexity mutató segítségével választja ki a dolgozat a megfelelő topikszámot a 4, 5 és 6 értékek közül. Ennek a mutatónak köszönhetően azok az elemszámok kerülnek kiválasztásra, melyek esetén a modell a legkevésbé érzékeny az új adatokra.

Az Standard & Poors index (SPX) adatai segítségével határozza meg a dolgozat a piaci volatilitást. Ezen perces felbontású adatok mutatják az amerikai tőzsde árfolyamait adott időegységekre. Az SPX értékekből kerül meghatározásra 30 és 60 perces ablakok esetén a piaci volatilitás. Ezeket a volatilitásokat lineáris regresszió segítségével becsüli az elemzés a topik arányokat, mint magyarázó változókat felhasználva.

Az eredmények azt mutatják, hogy meghatározhatóak olyan témák, melyek szignifikánsan képesen a piaci volatilitást befolyásolni. Az egyes szövegek esetén futtatott modellek eltérő eredményeket adtak mind a topikok mind a hatás tekintetében. A makrogazdasági helyzetre, árstabilitásra, növekedésre, munkaerőpiacra és monetáris politikára vonatkozó jegybanki kommunikációk mindegyike adott típusú szövegre vonatkoztatva szignifikáns hatást mutatnak.

A dolgozat több korláttal és továbbfejlesztési iránnyal is rendelkezik. A látens Dirichlet allokáció segítségével elkészített topikmodellek klaszterezettek, a dátumtól erősen függenek a topikok megoszlásai. Ennek javítása érdekében a dolgozat struktúrált topikmodelleket alkalmazott, melyek azonban nem voltak képesek megszüntetni a problémát. A topikmodellek esetén jelentkező klaszterezettség további javítása fontos továbbfejlesztési irányát képezi a dolgozatnak. További korlátja a tanulmánynak a kialakult topikok és témák összerendelésében jelentkező subjektivitás. Az egyes topikokhoz tartozó leggyakoribb kifejezések és a jegybanki kommunikációban fellelhető témák összerendelése igen nehéz és szubjektív feladat, mely azonban az eredmények szempontjából kulcsfontosságú lépés.

6. Appendix

15. táblázat. Eltávolított szavak a nyilatkozatok szövegeiből

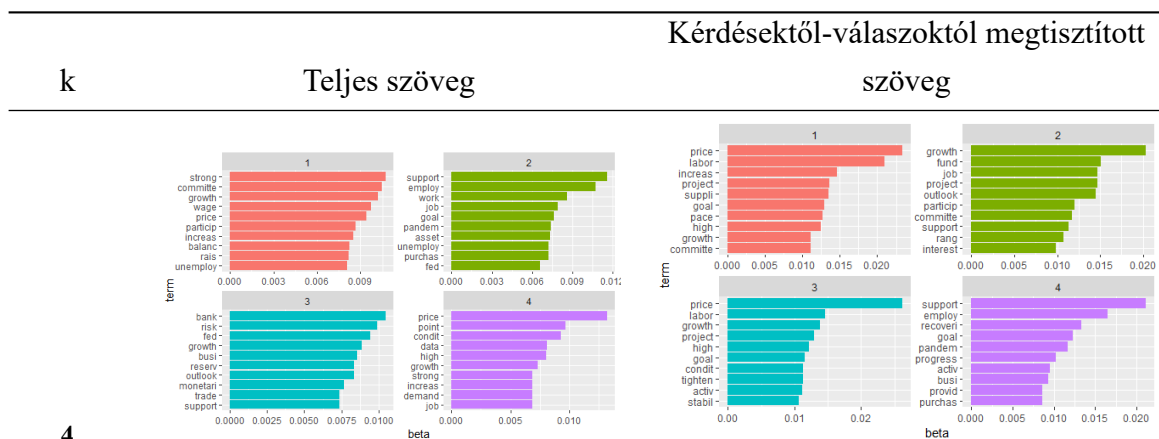
all	year	month
take	maximum	minimum
today	time	times
timing	stance	recent
include	including	range

ranges	percent	percents
run	running	reading
information	months	food
\$	powell	jerome
vice	williams	decide
decided	decides	john

16. táblázat. Eltávolított szavak a sajtótájékoztatók szövegeiből

percent	percents	make	made	powell	meet	meets
meeting	met	lot	want	wanted	wants	happen
happens	kind	long	good	move	moves	moved
level	talk	ill	time	times	question	questions
datum	datums	thing	things	mr	theyre	run
running	american	thought	weve	ah	end	havent
find	people	youre	youas	didnt	dont	doesnt
past	youve	im	\$	all	today	today's
year	years	month	months	take	maximum	minimum
back	®	¼	½	¾		

17. táblázat. Leggyakoribb szavak topikonként és modellenként a sajtótájékoztatók szövegeinél

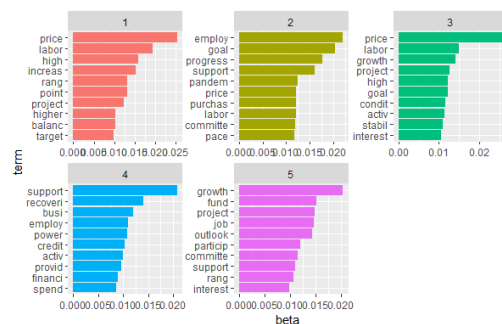
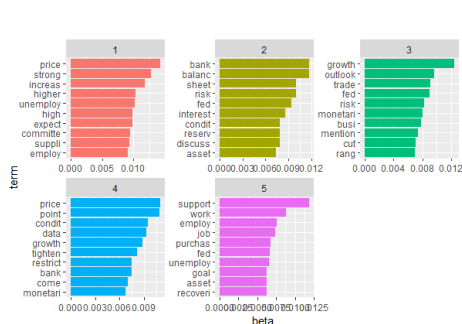


k

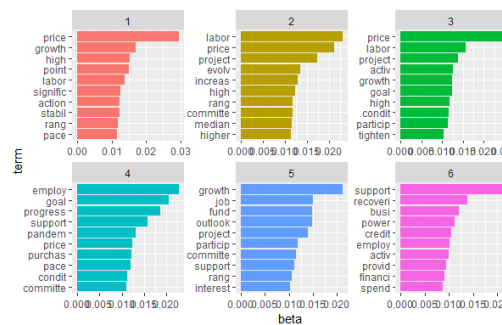
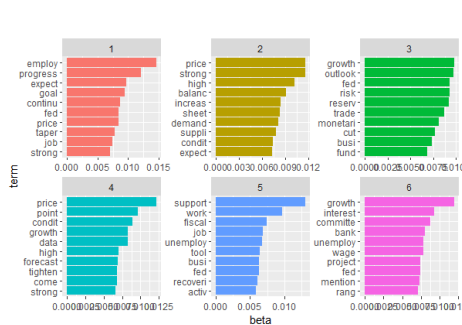
Teljes szöveg

szöveg

5



6



18. táblázat. Sajtótájékoztatók teljes szövegeinél a topikok metszetében szereplő szavak

economic	economy	economies	economical
economist	economists	expect	expecting
expects	financial	financials	inflation
inflationadjusted	inflationtargeting	inflationary	labor
labors	market	markets	policies
policy	rate	rates	rated

19. táblázat. Sajtótájékoztatók kérdésektől megtisztított szövegeinél a topikok metszetében szereplő szavak

continue	continues	continuing	economic	economy
economies	economical	economist	economists	expect
expecting	expects	federal	inflation	inflationadjusted
inflationtargeting	inflationary	market	marketes	monetary
policies	policy	rate	rates	rated
remain	remaining	remains	strong	

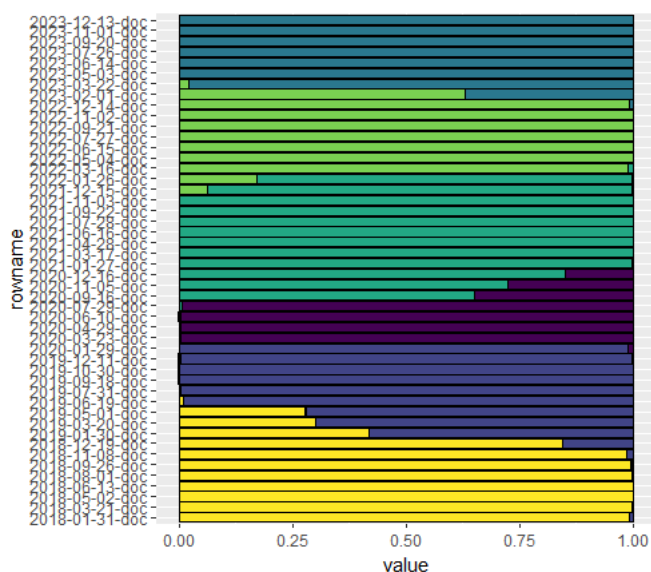
20. táblázat. Nyilatkozatok és beszédek UTC-be átalakított időpontjai

Időpont (helyi)	Időpont (UTC)	Időpont (helyi)	Időpont (UTC)	Időpont (helyi)	Időpont (UTC)
2018.01.31. 14:00 EST	2018.01.31. 19:00	2020.01.29. 14:00 EST	2020.01.29. 19:00	2022.01.26. 14:00 EST	2022.01.26. 19:00
2018.03.21. 14:00 EDT	2018.03.21. 18:00	2020.20.23. 14:00 EDT	2020.20.23. 18:00	2022.03.16. 14:00 EDT	2022.03.16. 18:00
2018.05.02. 14:00 EDT	2018.05.02. 18:00	2020.04.29. 14:00 EDT	2020.04.29. 18:00	2022.05.04. 14:00 EDT	2022.05.04. 18:00
2018.06.13. 14:00 EDT	2018.06.13. 18:00	2020.06.10. 14:00 EDT	2020.06.10. 18:00	2022.06.15. 14:00 EDT	2022.06.15. 18:00
2018.08.01. 14:00 EDT	2018.08.01. 18:00	2020.07.29. 14:00 EDT	2020.07.29. 18:00	2022.07.27. 14:00 EDT	2022.07.27. 18:00
2018.09.26. 14:00 EDT	2018.09.26. 18:00	2020.09.16. 14:00 EDT	2020.09.16. 18:00	2022.09.21. 14:00 EDT	2022.09.21. 18:00
2018.11.08. 14:00 EST	2018.11.08. 19:00	2020.11.05. 14:00 EST	2020.11.05. 19:00	2022.11.02. 14:00 EDT	2022.11.02. 18:00
2018.12.19. 14:00 EST	2018.12.19. 19:00	2020.12.16. 14:00 EST	2020.12.16. 19:00	2022.12.14. 14:00 EST	2022.12.14. 19:00
2019.01.30. 14:00 EST	2019.01.30. 19:00	2021.01.27. 14:00 EST	2021.01.27. 19:00	2023.02.01. 14:00 EST	2023.02.01. 19:00
2019.03.20. 14:00 EDT	2019.03.20. 18:00	2021. 03.17. 14:00 EDT	2021. 03.17. 18:00	2023.03.22. 14:00 EDT	2023.03.22. 18:00
2019.05.01. 14:00 EDT	2019.05.01. 18:00	2021. 04.28. 14:00 EDT	2021. 04.28. 18:00	2023.05.03. 14:00 EDT	2023.05.03. 18:00
2019.06.19. 14:00 EDT	2019.06.19. 18:00	2021. 06.16. 14:00 EDT	2021. 06.16. 18:00	2023.06.14. 14:00 EDT	2023.06.14. 18:00
2019.07.31. 14:00 EDT	2019.07.31. 18:00	2021.07.28. 14:00 EDT	2021.07.28. 18:00	2023. 07.26. 14:00 EDT	2023. 07.26. 18:00
2019.09.18. 14:00 EDT	2019.09.18. 18:00	2021.09.22. 14:00 EDT	2021.09.22. 18:00	2023.09.20. 14:00 EDT	2023.09.20. 18:00

Időpont (helyi)	Időpont (UTC)	Időpont (helyi)	Időpont (UTC)	Időpont (helyi)	Időpont (UTC)
2019.10.30. 24:00 EDT	2019.10.30. 18:00	2021.11.03. 14:00 EDT	2021.11.03. 18:00	2023.11.01. 14:00 EDT	2023.11.01. 18:00
2019.12.11. 14:00 EST	2019.12.11. 19:00	2021.12.15. 14:00 EST	2021.12.15. 19:00	2023.12.13. 14:00 EST	2023.12.13. 19:00

21. táblázat. Struktúrált topik modellek összefoglaló táblázata

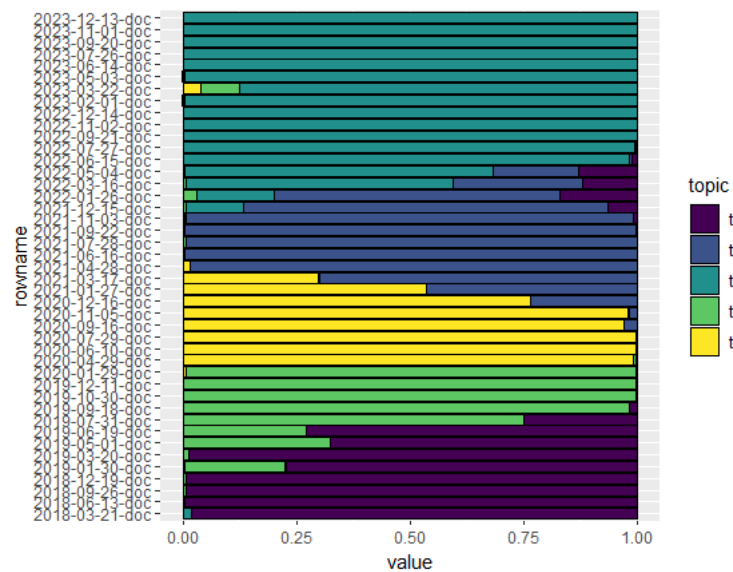
Szöveg	Nyilatkozatok	Sajtótájékoztató teljes	Sajtótájékoztató szűrt
Dokumentum szám	48	43	43
Kifejezés szám	257	1876	792
Token szám	3747	28393	11110
k	6	6	5
LDA modelt tartalmazó ábra megfelelője	1. ábra	2. ábra	3. ábra



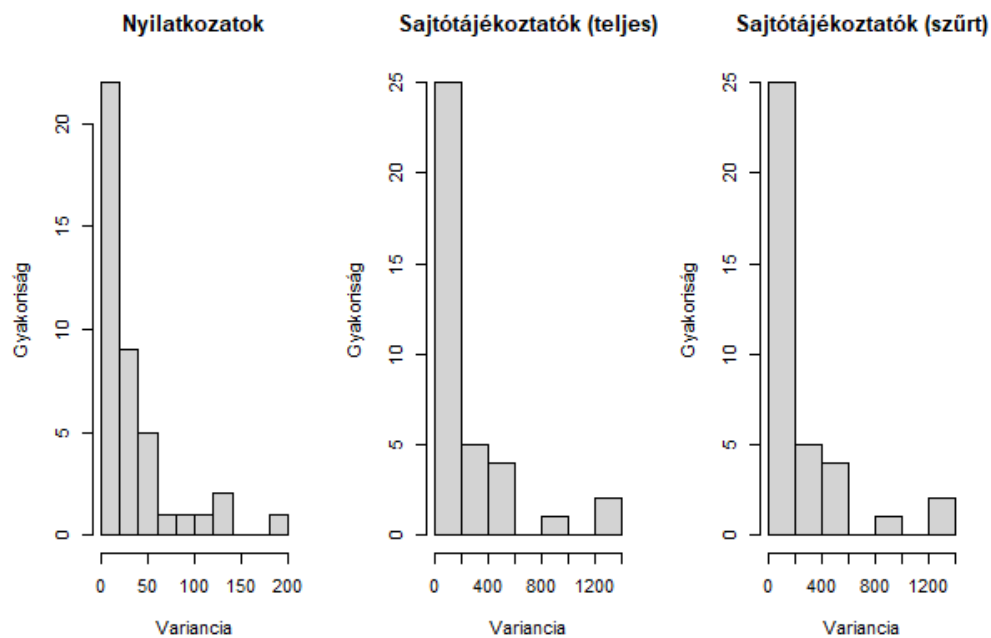
11. ábra. Topik eloszlások strukturált topik modell esetén - nyilatkozatok

Irodalomjegyzék

Benoit, Kenneth, David Muhr, and Kohei Watanabe. 2021. *Stopwords: Multilingual Stopword Lists*. <https://CRAN.R-project.org/package=stopwords>.



12. ábra. Topik eloszlások strukturált topik modell esetén - sajtótájékoztatók szűrt



13. ábra. Egyes szövegek ideje körül márt varianciák eloszlása

- Bernanke, Ben S., and Kenneth N. Kuttner. 2005. „What Explains the Stock Market’s Reaction to Federal Reserve Policy?” *The Journal of Finance* 60 (3): 1221–57. <https://doi.org/10.1111/j.1540-6261.2005.00760.x>.
- Bholat, David M., Stephen Hansen, Pedro M. Santos, and Cheryl Schonhardt-Bailey. 2015. „Text Mining for Central Banks.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2624811>.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. „Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3 (Jan): 993–1022. <https://jmlr.csail.mit.edu/papers/v3/blei03a.html>.
- Blinder, Alan S, Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen. 2008. „Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence.” *Journal of Economic Literature* 46 (4): 910–45. <https://doi.org/10.1257/jel.46.4.910>.
- Cieslak, Anna, and Andreas Schrimpf. 2019. „Non-Monetary News in Central Bank Communication.” *Journal of International Economics* 118 (May): 293–315. <https://doi.org/10.1016/j.jinteco.2019.01.012>.
- „Federal Reserve.” n.d. <https://www.federalreserve.gov/default.htm>.
- Fišer, Radovan, and Roman Horváth. 2010. „Central Bank Communication and Exchange Rate Volatility: A GARCH Analysis.” *Macroeconomics and Finance in Emerging Market Economies* 3 (1): 25–31. <https://doi.org/10.1080/17520840903498099>.
- Hansen, Stephen, and Michael McMahon. 2016. „Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication.” *Journal of International Economics* 99 (March): S114–33. <https://doi.org/10.1016/j.jinteco.2015.12.008>.
- Jansen, David-Jan. 2011. „Does the Clarity of Central Bank Communication Affect Volatility in Financial Markets? Evidence from Humphrey-Hawkins Testimonies.” *Contemporary Economic Policy* 29 (4): 494–509. <https://doi.org/10.1111/j.1465-7287.2010.00238.x>.
- Jegadeesh, Narasimhan, and Di (Andrew) Wu. 2015. „Deciphering FedSpeak: The Information Content of FOMC Meetings.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2939937>.
- Kapadia, Shashank. 2022. „Evaluate Topic Models: Latent Dirichlet Allocation (LDA).” <https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0>.
- Lucca, David O., and Emanuel Moench. 2015. „The Pre-FOMC Announcement Drift.” *The Journal of Finance* 70 (1): 329–71. <https://doi.org/10.1111/jofi.12196>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Reeves, Rachel, and Michael Sawicki. 2007. „Do Financial Markets React to Bank of England Communication?” *European Journal of Political Economy* 23 (1): 207–27. <https://doi.org/10.1016/j.ejpoleco.2006.09.018>.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. „Stm : An R Package for Structural Topic Models.” *Journal of Statistical Software* 91 (2). <https://doi.org/10.18637/jss.v091.i02>.
- Rosa, Carlo. 2011. „Words That Shake Traders.” *Journal of Empirical Finance* 18 (5): 915–34. <https://doi.org/10.1016/j.jempfin.2011.07.005>.
- Sack, Brian P., and Donald L. Kohn. 2003. „Central Bank Talk: Does It Matter and Why?” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.483524>.
- Sebők, Miklós, Orsolya Ring, and Ákos Máté. 2021. *Szövegbányászat És Mesterséges Intelligencia r-Ben*. <https://tankonyv.poltextlab.com/index.html>.
- U. S. Bureau of Labor Statistics. n.d. „Schedule of Releases for the Employment Situation.” https://www.bls.gov/schedule/news_release/empisit.htm.