

2.3. MEKKORÁRA NŐHETETT A MUNKANÉLKÜLISÉG – BECSLÉS INTERNETES KERESÉSEK ALAPJÁN

KÓNYA ISTVÁN & KÖLLŐ JÁNOS

Bevezetés

A koronavírus-járvány első hulláma 2020 tavaszán érte el Magyarországot és a magyar gazdaságot. A járvány okozta kínálati és keresleti sokkok miatt állások sokasága szűnt meg ebben az időszakban. Az erre vonatkozó statisztikai adatok csak jelentős késéssel és máig nehezen becsülhető pontossággal kerültek nyilvánosságra. A Nemzeti Foglalkoztatási Szolgálatnak (NFSZ) a 2020. március 20-i „zárónapra” vonatkozó, ellenőrzött adatai például csak május elején jelentek meg. A KSH Munkaerő-felmérés (MEF) második negyedévi adatait – amelyek a munkaidőre, az állásvesztésre és az álláskeresésre vonatkozó információkat is tartalmaznak – pedig csak júniusban, a szokásoshoz képest jóval nagyobb adatbizonytalansággal hozták nyilvánosságra.¹

A munkapiac állapotáról – és ezen belül a munkanélküliség nagyságáról – ugyanakkor fontos minél hamarabb képet kapnia a döntéshozónak, hogy a megfelelő szakpolitikai intézkedéseket időben meg tudja hozni. Ehhez nyújthat segítséget egy olyan indikátor, amely gyakorlatilag valós időben rendelkezésre áll, és a munkanélküliséghez szorosan kapcsolódó internetes keresési gyakoriságon alapul. Az alábbiakban egy ilyen módszert mutatunk be, elsősorban a 2020. tavaszi munkapiaci folyamatok előrejelzésének kontextusában, az azóta beérkezett tényadatokat használva külső validálásra. Számításunk kapcsolódik hasonló, a Google-kereséseken nyugvó kísérletekhez (például *Kong–Princz*, 2020, valamint *Tóth és szerzőtársai*, 2020). A fejezetben leírt módszer közvetlen előzménye pedig a szerzők 2020. május 15-én megjelent blogbejegyzése (*Kónya–Köllő*, 2020).

Havi adatok

A munkapiaci vizsgálatok alapadatai havi és negyedéves frekvencián állnak rendelkezésre a KSH honlapján.² Először a havi adatok előrejelzését vizsgáljuk meg a Google keresési adatai segítségével. Ehhez letöltöttük a Google Trends honlapról³ a Magyarország területén a „munkanélküli-segély” kifejezésre keresők idősorát. Az adatok szintén havi frekvencián érhetőek el. A KSH frissített módszertana⁴ szerinti munkapiaci állományi idősorok kezdeti dátuma 2009. január, így a Google adatait is ettől a hónaptól tudjuk használni. A Google-adatok egy adott időszakra relatív skálán vesznek fel értékeket, ahol a legmagasabb keresési intenzitású hónap értéke 100-ra van normalizálva.

A keresési adatok használatánál érdemes figyelembe venni azt a tényt is, hogy a mintaidőszakban a magyar háztartások internet-hozzáférési lehető-

¹ Lásd a K2.1. keretes írását erről a témáról.

² Lásd: [KSH](#).

³ Lásd: [Google Trends](#).

⁴ Lásd a K2.2. keretes írását a foglalkoztatás megváltozott definíciójáról.

ségei érdemben nőttek. Különösen igaz lehet ez a munkanélküliekre, akik között az alacsonyabban képzettek és alacsony jövedelműek felül vannak reprezentálva. Vagyis az időszak első felében valószínűleg a munkanélküliek kisebb része végzett internetes keresést, mint a minta második felében. Ezt a következőképpen korrigáljuk. Az Eurostat honlapjáról letöltöttük a „Individuals – internet use” táblát,⁵ amelyből a munkanélküliek azon hányadát vesszük, akik a kérdéses előtti három hónapban használtak internetet. Az adatok éves frekvencián állnak rendelkezésre, ezért a havi értékeket lineáris interpolációval állítjuk elő. A kapott idősor 52 százalék (2009) és 81 százalék (2021) közötti értékeket vesz fel. A korrigált keresési adatokat úgy kapjuk meg, hogy a nyers adatokat az ily módon előállított havi internethasználati adatokkal osztjuk el.

A Google-adatok előrejelző (illetve *nowcasting*) képességét a következőképpen vizsgáljuk. Induljunk ki egy egyszerű idősoros regresszióból, ahol a munkapiaci változót annak késleltetett értékeivel magyarázunk, illetve jelezzük előre. Ehhez adjuk hozzá a (korrigált) keresési gyakoriságot:

$$x_t = \alpha + \sum_{k=1}^K \beta_k x_{t-k} + \delta z_t + \varepsilon_t, \quad (1)$$

ahol x_t a munkapiaci változó (munkanélküliség vagy foglalkoztatottság), z_t a korrigált keresési mutató, K pedig a késleltetések száma. Ez utóbbit a havi adatok esetében $K = 3$ -nak vesszük.

A keresési változó előrejelző erejét egyrészt mutatja a δ együttható szignifikanciája. Minél erősebb a szignifikancia, annál nagyobb a mutató magyarázó ereje a munkapiaci változó varianciájában. Ez az úgynevezett mintabeli (*in sample*) magyarázó erő. Ugyanakkor az általunk vizsgált kérdéshez fontosabb a mintán kívüli (*out of sample*) előrejelzési képesség. Ezt a következőképpen teszteljük.

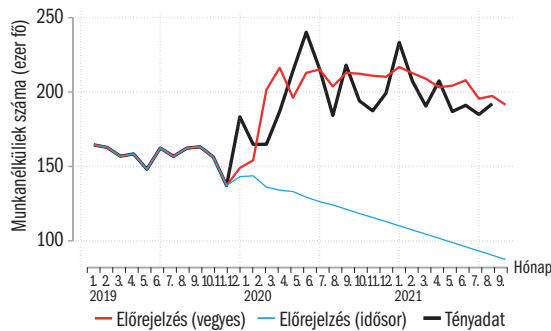
A Covid-járvány kezdetekor nagy bizonytalanság övezte a munkapiac alakulását. Azt nézzük meg, hogy a *2020 elején rendelkezésre álló információs halmazon* mennyire segít a keresési gyakoriság felhasználása az előrejelzésben. Ehhez először is megbecsüljük az (1) összefüggést 2009 és 2019 közötti adatokon. Majd felhasználva a becsült együtthatókat, illetve a gyakorlatilag azonnal rendelkezésre álló keresési adatokat, előrejelezzük a munkanélküliség, illetve a foglalkoztatás alakulását a következő hónapokban. Összehasonlítási alapként az (1) egyenlet exogén változó nélküli változatát használjuk, vagyis azt az esetet, amikor $\delta = 0$.

Mivel a becslést 2019. decemberrel bezárólag végezzük, az előrejelzést 2020. januárban kezdjük. Az előrejelzés készítésekor a munkapiaci változó utolsó tényadatok 2019. december, a Google-keresési mutató viszont az aktuális havi adat. Vagyis arra vagyunk kíváncsiak, hogy a Covid-válság alatt a korábbi munkapiaci folyamatok ismeretében, de *csak a Google keresési adataival frissítve az információs halmazunkat*, milyen előrejelzési teljesítményt látunk.

⁵ Lásd: Eurostat.

A munkanélküliekre vonatkozó eredményeket a 2.3.1. ábra illusztrálja. A fő észrevétel az ábra alapján az, hogy a (korrigált) keresési változó érdemben és jelentősen javítja a munkanélküliség előrejelzését. Ez persze részben nem meglepő, mert a tisztán idősoros modell egyre inkább „lemarad” a valós folyamatokról, hiszen abban az utolsó információ időpontja 2019 vége. Ugyanakkor figyelemre méltó az, hogy a keresési gyakoriság önmagában mennyire jól követi a tényleges munkanélküliség alakulását. Ha a havi ingadozásokat nem is tökéletesen adja vissza az előrejelzés, a munkanélküliség középtávú folyamatait rendkívül pontosan jelzi előre a kibővített modell. Egy további észrevétel az, hogy a keresések figyelembevételével rövid távon is lényegesen pontosabb becslést tudunk adni a tényleges folyamatra – ezt támasztják alá a 2020 első néhány hónapjában látható különbségek az ábrán.

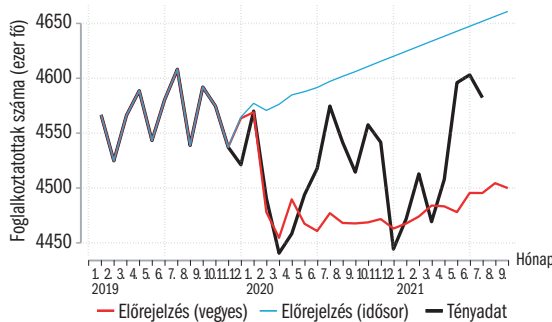
2.3.1. ábra: A munkanélküliség előrejelzése



Forrás: KSH Munkaerő-felmérés és Google Trends.

Az előrejelzési gyakorlatot elvégeztük a foglalkoztatási adatokra is (2.3.2. ábra). Rövid távon talán még látványosabb az eredmény, mint a munkanélküliség esetében.

2.3.2. ábra: A foglalkoztatottak előrejelzése



Forrás: KSH Munkaerő-felmérés és Google Trends.

A keresési adatokra támaszkodó modell szinte tökéletesen eltalálja a 2020. tavaszi foglalkoztatási folyamatokat: mind a jelentős csökkenést, mind pedig az azt követő kezdődő helyreállást. Ezzel szemben a csak 2019. végi adatokból

extrapoláló idősoros modell – teljesen természetes módon – nem képes „előrelátni” a Covid-sokkot. Érdekes, hogy a foglalkoztatás jelentős „visszapattanását” a keresési adatok nem követték sem 2020 nyarán, sem 2021 tavaszán. Ennek oka lehet az, hogy a foglalkoztatás változásában ekkor már elsősorban az állászerzés és nem az állásvesztés dominált. Erre a kérdésre a következő részben térünk még vissza.

Negyedéves adatok

A havi adatok meglehetősen zajosak, és vélhetően több mérési probléma jelentkezik, mint negyedéves frekvencián. Ezért megismételjük az előző rész számításait negyedéves adatokkal is, helyszűke miatt csak a munkanélküliségre alkalmazva. Egy további előnye a negyedéves frekvenciának, hogy az álmányi adatok mellett áramlási adatok is elérhetők. Az Eurostat honlapjáról letöltöttük a különböző munkapiaci állapotok közötti átmenetek idősorait.⁶ Ezek közül a foglalkoztatás → munkanélküliség áramlás előrejelzését vizsgáljuk, mivel vélhetően (és gyakorlatilag is) ez áll a legszorosabb kapcsolatban a keresésben használt kifejezéssel. A Google-adatsort egyszerű átlagolással aggregáltuk negyedéves frekvenciára, az internethasználat esetében pedig az éves adatokat a korábbihoz hasonlóan lineáris interpolációval tesszük negyedévéssé. Az előrejelzéshez továbbra is az (1) egyenletet használjuk, illetve annak $\delta = 0$ feltevessel korlátozott változata a viszonyítási alapunk. A negyedéves frekvencia miatt a késleltetések számát 1-nek vesszük ($K = 1$).

A 2.3.3. *ábra* ismerteti az eredményeket. A havi adatokhoz hasonlóan a keresési információ figyelembevétele drámaian javítja az előrejelző képességet. Bár a modell kissé túlbecsüli a munkanélküliség emelkedését, különösen az első két negyedévben kifejezetten jól teljesít. A tisztán idősoros megközelítés most sem tudja előrejelezni a Covid-válságot, ezért nagyon hasznos az azonnal rendelkezésre álló keresési információ figyelembevétele. Érdekes módon a bővített modell – a havi adatok esetén a foglalkoztatáshoz hasonlóan – 2020 őszén szakad el leginkább a tényadatoktól. Érdekes lenne megvizsgálni, hogy mi lehet ennek a magyarázata – rövid elemzésünkben azonban erre nincsen mód.

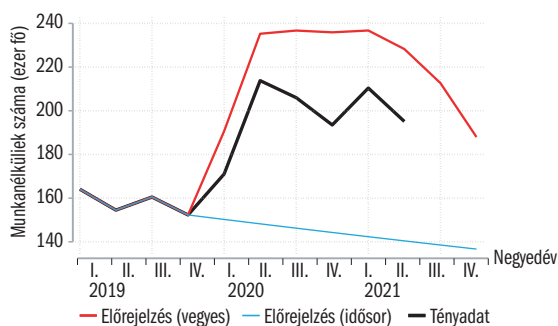
Utolsó vizsgálatunkban a már említett munkapiaci áramlást próbáljuk előrejelezni, vagyis a foglalkoztatottból munkanélkülivé válók számát. Ez egyrészt azért érdekes, mert a válságok kezdetén a foglalkoztatottságból történő kiáramlás hatására ugrik meg a munkanélküliség, tehát a váratlan recessziókban különösen fontos a mutató előrejelzése. Másrészt pedig az „munkanélküli segély” keresési kifejezés vélhetőleg ezzel az áramlással van a legszorosabb kapcsolatban, mivel ilyen keresést valószínűleg a frissen állásukat veszített emberek végeznek.

A 2.3.4. *ábra* mutatja az eredményeket. A vártnak megfelelően a modell rendkívül jó teljesít. Az egyetlen jelentősebb eltérés a tényadatokhoz képest

⁶ Lásd: Eurostat.

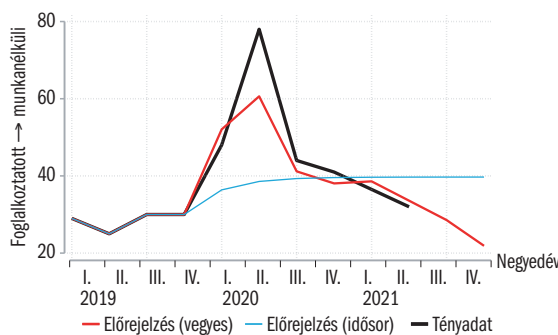
2020 második negyedében van, ahol a kereséssel kibővített előrejelzés is alábecsüli az áramlás tényleges megemelkedését. A többi negyedében azonban az előrejelzés nagyon pontos. Ez különösen figyelemre méltó 2020 második felében és 2021 elején, hiszen itt már szinte kizárólag a keresési adatokra – illetve a válság előtt adatokon becült dinamikára – támaszkodik a modell.

2.3.3. ábra: Munkanélküliség előrejelzése (negyedéves frekvencia)



Forrás: KSH Munkaerő-felmérés és Google Trends.

2.3.4. ábra: Foglalkoztatottból munkanélkülivé válók előrejelzése



Forrás: KSH Munkaerő-felmérés és Google Trends.

Összegzés

Ebben az alfejezetben munkanélküliségi és foglalkoztatási adatok előrejelzését végeztük el Google-keresési intenzitásadatok felhasználásával. Fő eredményünk az, hogy a késéssel elérhető tényadatok jól előrejelezhetők ezzel az egyszerű és könnyen elérhető többletinformációval. Különösen igaz ez olyan válságok idején, mint a Covid-járvány, amikor a munkapiac állapota különösen gyorsan és radikálisan változik. Bár a keresési mutató jól teljesített mind a munkanélküliség, mind pedig a foglalkoztatottság előrejelzésében, a legpontosabb eredményeket a munkájukat frissen elvesztettek áramlási adatainál láttunk. Ez megerősíti azt a következtetést, hogy különösen gyors változások esetén érdemes kibővíteni a munkapiacra vonatkozó információs halmazt.

Hivatkozások

KONG, E.–PRINZ DÁNIEL (2020): The Impact of Shutdown Policies on Unemployment During a Pandemic, Harvard University, nyers kézirat, május 4.

KÓNYA ISTVÁN–KÖLLŐ JÁNOS (2020): [Mekkorára nőhetett a munkanélküliség – Becslés internetes keresések alapján](#). KRTK KTI, Koronavírus-blog, május 15.

TÓTH GERGŐ–CZALLER LÁSZLÓ–ELEKES ZOLTÁN (2020): [Mit árulnak el az internetes keresések a munkaerőpiaci kilátások területi megoszlásáról?](#) KRTK KTI, Koronavírus-blog, május 4.

2.3. Függelék – regressziók

F2.3.1. táblázat: Havi adatok

	Munkanélküliség		Foglalkoztatottság	
	(1)	(2)	(3)	(4)
Google-keresés	0,533*** (0,0911)		-0,758*** (0,208)	
1. késleltetés	0,459*** (0,0744)	0,585*** (0,0801)	0,573*** (0,0828)	0,654*** (0,0835)
2. késleltetés	-0,000888 (0,0850)	0,00428 (0,0956)	-0,0493 (0,0975)	-0,0292 (0,102)
3. késleltetés	0,473*** (0,0719)	0,417*** (0,0802)	0,422*** (0,0811)	0,371*** (0,0838)
Konstans	-7,204 (4,801)	-5,783 (5,394)	275,5** (84,54)	23,75 (51,11)
<i>N</i>	129	129	129	129
<i>R</i> ²	0,979	0,973	0,983	0,982

Megjegyzés: A Google-keresésnél a „munkanélküli-segély” kifejezést használtuk. A zárójelben standard hibák szerepelnek.

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$.

F2.3.2. táblázat: Negyedéves adatok

	Munkanélküliség		Foglalkoztatott → munkanélküli	
	(1)	(2)	(3)	(4)
Google keresés	0,336*** (0,0556)		0,184*** (0,0438)	
1. késleltetés	0,864*** (0,0321)	0,987*** (0,0278)	0,102 (0,142)	0,343* (0,157)
Konstans	-9,446 (7,966)	-0,0343 (9,528)	11,44 (6,630)	26,10*** (6,822)
<i>N</i>	43	49	38	38
<i>R</i> ²	0,981	0,964	0,413	0,116

Megjegyzés: A Google-keresésnél a „munkanélküli-segély” kifejezést használtuk. A zárójelben standard hibák szerepelnek.

*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$.