

OTDK-dolgozat

2023

Budapesti Corvinus Egyetem

Területi autokorrelációs modellek távolságmátrix megválasztására vonatkozó érzékenységvizsgálata

Sensitivity analysis of distance matrix selection for
spatial autocorrelation models

Jakó Fanni

Konzulens: Madari Zoltán
Kézirat lezárásának dátuma: 2022.04.19.

Absztrakt

A tanulmányban a Moran-féle I-próba területi súlymátrix definiálásától való függését vizsgálom. A térbeli adatok kezelésének egyik problémája, hogy a kölcsönös hatások miatt nem egyértelmű az ok-okozati viszony a térségek között. Regressziós modellekben, valamint a statisztikai tesztekben ezeket a hatásokat egy területi súlymátrix segítségével vizsgálják. A térnek már a 2 dimenzióba való sűrítése is problémákat vet fel, azonban a terület-terület párok közötti kapcsolat egy értékkel való jellemzése még nehezebb feladat, mely a modellezőt terheli. A modellek eredményei ezen mátrix definiálásától nagymértékben függenek, ezért fontos a kérdéskör.

Kutatási kérdésem arra vonatkozik, hogy milyen jellegzetességeket kell figyelembe venni a területi súlymátrix megválasztása során, illetve hogy az egyes gazdasági problémák és változók karakterisztikái, illetve a területi szint, melyen a problémát vizsgáljuk, milyen területi súlymátrix alkalmazását indukálja. A kutatás Európa területén, az egyes NUTS szinteken vizsgálja, hogy milyen statisztikai vizsgálatokat javasolt elvégezni az egyes térbeli modellek illesztése előtt.

A tanulmányban 9-9-9 területi súlymátrixot definiálok a NUTS rendszer 3 szintjén, melyek által adott tesztstatisztikák eltéréseit négyféleképpen vizsgálom: matematikai tulajdonságaik alapján, független változókon mutatott eredményeik alapján, valós gazdasági és társadalmi adathalmazon, valamint szimulációk által.

Eredményeim alapján nem elégséges az egyes területi súlymátrixokat a matematikai hasonlóság fogalma alapján összevetni. Területi modellek alkalmazása előtt tesztelni kell, hogy az adott területi súlymátrix független adatokon mekkora elsőfajú hibát vét. Térbeli autokorreláció vizsgálatakor nem lehet csupán egy definícióval adott mátrix segítségével elvégezni a teszteket, többféle területi kapcsolatot reprezentáló mátrix alkalmazásával kell megerősíteni az autokorreláció jelenlétét, kiváltképp akkor, ha az alapsokaság pozitív ferdeségű és csúcsossága 1 feletti. Szimulációkkal tesztelni kell, hogy az adott térbeli jellegzetességek mellett az egyes mátrixok határértékben milyen mintázatok alakítanak ki, majd ennek tudatában kell megválasztani a modellillesztés során használt súlymátrixot.

Kulcsszavak: Területi autokorreláció, területi súlymátrix, NUTS, Moran I-próba

JEL klasszifikáció: C31, R12

Tartalomjegyzék

Ábrák jegyzéke	iii
Táblázatok jegyzéke	v
1. Bevezetés	1
2. Irodalmi áttekintés	4
2.1. Területi modellek a szakirodalomban	4
2.2. Modellspecifikációk a szakirodalomban	7
2.3. Európai körkép	8
3. Adatbázis kialakítása	10
3.1. NUTS rendszer sajátosságai	10
3.2. Felhasznált adatbázisok	12
4. Módszertan	14
4.1. Alapfogalmak	15
4.2. Területi súlymátrix	16
4.3. Területi autokorreláció tesztelése	27
4.4. Területi autokorreláció kezelése	29
5. Eredmények	31
5.1. Alkalmazott területi súlymátrixok	31
5.2. Matematikai összehasonlítás	34
5.3. Autokorrelálatlanság kimutatása	39
5.4. Régiós adatok szemléltetése	41
5.5. Szimulációs eredmények	47
5.6. Javasolt területi súlymátrix választási folyamat	59
6. Összegzés	61
7. Függelék	I

Ábrák jegyzéke

1.	<i>NUTS 3 régiók tipológiai. (Forrás: Saját szerkesztés)</i>	11
2.	<i>NUTS 3 régiók városi csoportosítása, valamint a vizsgált városok.</i> <i>(Forrás: Saját szerkesztés)</i>	12
3.	<i>Szomszédos területek számának eloszlása az egyes NUTS szinteken.</i> <i>(Forrás: Saját szerkesztés)</i>	17
4.	<i>North Yorkshire CC példája a szomszédsági számokra. (Forrás: Saját szerkesztés)</i>	18
5.	<i>Másodszomszédos területek száma NUTS szinteken. (Forrás: Saját szerkesztés)</i>	19
6.	<i>Alsó-Rajna megye, Franciaország és másodszomszédai. (Forrás: Saját szerkesztés)</i>	20
7.	<i>Eltérés a közúti távolság alapú távolságértékekben. (Forrás: Saját szerkesztés)</i>	22
8.	<i>Eltérés a közúti távolság alapú távolságértékek és a középpontok távolsága között. (Forrás:</i> <i>Saját szerkesztés)</i>	23
9.	<i>Mátrix sajátértékek NUTS 1 szinten. (Forrás: Saját szerkesztés)</i>	35
10.	<i>Mátrix sajátértékek NUTS 2 szinten. (Forrás: Saját szerkesztés)</i>	36
11.	<i>Mátrix sajátértékek NUTS 3 szinten. (Forrás: Saját szerkesztés)</i>	37
12.	<i>Független mintán generált p-értékek NUTS 1 szinten $N(100, 10)$ mellett. (Forrás: Saját</i> <i>szerkesztés)</i>	39
13.	<i>Változók alakmutatói a különböző mátrixok által adott eredmények egyezése szerint ki-</i> <i>alakított csoportokban. (Forrás: Saját szerkesztés)</i>	46
14.	<i>Elsőfokú szomszédság mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szerkesztés)</i>	48
15.	<i>Másodfokú szomszédság mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szer-</i> <i>kesztés)</i>	48
16.	<i>Közúti távolság az elsőfokú szomszédok között kilométerben mátrixszal generált adatok</i> <i>NUTS 1 szinten. (Forrás: Saját szerkesztés)</i>	49
17.	<i>Közúti távolság az elsőfokú szomszédok között órában mátrixszal generált adatok NUTS</i> <i>1 szinten. (Forrás: Saját szerkesztés)</i>	50
18.	<i>Közúti távolság a másodfokú szomszédok között kilométerben mátrixszal generált adatok</i> <i>NUTS 1 szinten. (Forrás: Saját szerkesztés)</i>	50
19.	<i>Közúti távolság a másodfokú szomszédok között órában mátrixszal generált adatok NUTS</i> <i>1 szinten. (Forrás: Saját szerkesztés)</i>	51
20.	<i>Régióközpont távolság mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szer-</i> <i>kesztés)</i>	52

21.	<i>Poligonok közötti legrövidebb távolság mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szerkesztés)</i>	52
22.	<i>4-legközelebbi szomszéd mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szerkesztés)</i>	53
23.	<i>Szimulációs eredmények p értékei a generáló és a tesztelő mátrix szerinti bontásban NUTS 1 szinten. (Forrás: Saját szerkesztés)</i>	57
24.	<i>Szimulációs eredmények p értékei a generáló és a tesztelő mátrix szerinti bontásban NUTS 2 szinten. (Forrás: Saját szerkesztés)</i>	58
25.	<i>Szimulációs eredmények p értékei a generáló és a tesztelő mátrix szerinti bontásban NUTS 3 szinten. (Forrás: Saját szerkesztés)</i>	II

Táblázatok jegyzéke

1.	NUTS 1 szinten definiált területi súlymátrixok tulajdonságai. (Forrás: Saját szerkesztés) *: zérus oszlopok és sorok elhagyása után	35
2.	NUTS 2 szinten definiált területi súlymátrixok tulajdonságai. (Forrás: Saját szerkesztés) *: zérus oszlopok és sorok elhagyása után	36
3.	NUTS 3 szinten definiált területi súlymátrixok tulajdonságai. (Forrás: Saját szerkesztés) *: zérus oszlopok és sorok elhagyása után	37
4.	A gazdasági adatokon adott azonos eredmények aránya az egyes mátrixok között (Forrás: Saját szerkesztés)	43



AZ INNOVÁCIÓS ÉS TECHNOLÓGIAI MINISZTERIUM ÚNKP-21-2-I KÓDSZÁMÚ ÚJ NEMZETI KIVÁLÓSÁG PROGRAMJÁNAK A NEMZETI KUTATÁSI, FEJLESZTÉSI ÉS INNOVÁCIÓS ALAPBÓL FINANSZÍROZOTT SZAKMAI TÁMOGATÁSÁVAL KÉSZÜLT.

1. Bevezetés

A térben egymáshoz közel elhelyezkedő egységek - háztartások, vállalatok, városok, megyék, régiók, országok - nagy általánosságban kölcsönösen hatással vannak egymásra. Ezért a regressziós modellekben, legyen az akár keresztmetszeti - akár panel -, ezért az ilyen típusú hatásoktól nem lehet eltekinteni. Ezen problémák megoldására irányuló ökonometriai megoldásokra épül a térökonometria (spatial econometrics), mely két elkülönülő problémával foglalkozik. Egyik ilyen a térbeli interakciók hatásaival, azaz a térbeli autokorrelációval foglalkozik, míg a másik a térbeli struktúra, azaz a térbeli heterogenitás hatásait vizsgálja (Paelinck és tsai. 1979; Anselin 1988; Varga 2002).

A térbeli adatok kezelésének egyik problémája, hogy az idősorelemzéssel ellentétben nem egyértelmű, hogy melyik megfigyelési egység okozza a hatást és melyik egység viseli azt. Az idősorelemzésben az ok-okozat egyértelmű - a későbbi megfigyelési értékre van hatással a korábbi érték. A térbeli kapcsolatok esetében ez az ok-okozati viszony másfajta, az esetek többségében mindkét megfigyelés hatással van egymásra. Amennyiben a regressziós modellekben ezen hatásokat nem kezelik, és a regressziós hibatag területileg autokorreál, úgy a legkisebb négyzetek módszerével kapott paraméterbecslés nem lesz határos, valamint a standard hiba becslése torzított lesz. A területi autokorreláció tesztelésére használt statisztikai tesztek, valamint a térökonometriai modellek egy úgynevezett területi súlymátrix segítségével korrigálják a regressziós modelleket, hogy az általuk adott paraméterbecslések hatásosak legyenek. A modellek eredményei ezen mátrix definiálásától nagyobb mértékben függenek, mint a modellválasztástól (OLS, általános momentum módszer, maximum likelihood becslés), emiatt a területi súlymátrix megválasztása egy kiemelt pontja a térökonometriai modelleknek (Bell és Bockstael 2000).

A területi súlymátrixnak minden területpárra egyértelműen definiálnia kell a kapcsolat szorosságát. Ennek megválasztása azonban nem egyértelmű, hiszen annak eldöntése, hogy mely területek között tételezünk fel kapcsolatot, annak megválasztása, hogy a területek mely pontjai között nézzük a távolságot, valamint hogy magát a távolságot milyen módon definiáljuk mind meghatározzák a kapott eredményt. A mátrix reprezentálhatja csupán a szomszédságot, az egyes területek valamely kiemelt geokoordinátájú pontjainak távolságát vagy valamilyen közgazdasági változó által reprezentált kapcsolatot (például import, export, leányvállalatok száma). Ezeken belül a területek távolságát is többféleképpen definiálhatjuk: vizsgálhatjuk a mértani középpontok távolságát, a legnépesebb vá-

rosok távolságát, a közigazgatási középpontok távolságát vagy akár egyéb módon definiált távolságokat is.

A kutatás Európa területén, az egyes NUTS szinteken (Nomenclature des unités territoriales statistiques – Területi statisztikai rendszer nómenklatúrája) vizsgálja, hogy mely területi súlymátrixok alkalmazása indokolt a térökonometriai modellekben. A NUTS rendszer az EUROSTAT által létrehozott egységes területi rendszer, így a kutatás középpontjában Európa földrajza áll. A kérdést a három NUTS szinten vizsgálom az Európai Unió tagállamainak, valamint az Egyesült Királyságnak és Svájcnak bevonásával. Emellett a Google Maps Distance Matrix API segítségével letöltött közúti adatok segítségével egy új módon is definiálom a NUTS régiók távolságát a kontinens jelentősebb városai közötti közúti távolság átlagos értékének segítségével.

A kutatás középpontjában ezen területi súlymátrixok állnak. Kutatási kérdésem arra vonatkozik, hogy az egyes gazdasági problémák karakterisztikái, illetve a területi szint, melyen a probléma vizsgálva van milyen területi súlymátrix alkalmazását indokolja. A kérdések megválaszolására a Moran-féle globális I-próba eredményeit használom fel, mely globális szinten vizsgálja a területen jelen lévő területi autokorreláció létét.

A tanulmány az alábbi kérdések megválaszolására tér ki. A definiált területi súlymátrixok matematikai tulajdonságainak egyezése és a statisztikai tesztek eredményeinek megegyezése között van-e kapcsolat, azaz a mátrixok hasonlóságának vizsgálatával eldönthető-e, hogy érdemes-e a teszteket különféle területi mátrixokkal elvégezni (5.2). Megvizsgálom, hogy független mintán mekkora elsőfajú hibát vétenek az egyes mátrixok segítségével végzett Moran-féle I-próbák, valamint hogy ezen eredmények milyen módon befolyásolják a valós gazdasági és társadalmi adatokon vett tesztstatisztika értékeket (5.3).

Vizsgálom, hogy mekkora különbségek tapasztalhatóak valós adatokon a különbözőképpen definiált súlymátrixokkal végzett próbák eredményeiben, valamint azt, hogy az adatsokaság karakterisztikái alapján lehet-e következtetni ezen eredmények eltérésének fokára (5.4). Ezt követően szimulációk segítségével elemzem, hogy az adott területi súlymátrix által generált adatok határértékben milyen mintázatot mutatnak a NUTS régiókon, valamint bemutatom, hogy az egyes súlymátrixok által generált területi autokorrelációkat a másik módon definiált súlymátrixok segítségével végzett Moran-féle I-próba milyen arányban mutatja ki (5.5). Végezetül az eredmények alapján javaslatot teszek arra, hogy a területi autokorreláció korrigálása során milyen lépéseket kell tenni ahhoz, hogy a földrajzi

struktúrát és a gazdasági adatok jellemzőit is figyelembe vevő területi mátrix kerüljön a modellekbe (5.6).

A tanulmány felépítése a következő. A 2. fejezetben bemutatom a térökonometria eddigi alkalmazási területeit, majd a 3. fejezetben ismertetem az EUROSTAT által kidolgozott NUTS rendszer sajátosságait, az adattisztítási lépéseket, valamint a városi adatbázis letöltésének és felhasználásának formáját. A 4. fejezetben részletesen ismertetem az alapvető fogalmakat, az egyes súlymátrixok definiálásának lehetőségeit, valamint a Moran-féle I-próbát és a területi autokorrelációs modelleket. Az 5. fejezetben bemutatom a korábban feltett kérdésekre adott válaszokat, melyeket a 6. fejezetben összegzek.

2. Irodalmi áttekintés

Az alábbi fejezetben bemutatom a területi autokorrelációs modellek széles alkalmazási területeit (2.1. alfejezet), a szakirodalomban található módszereket a modellekspecifikációk, tesztstatisztikák és becslési eljárások alkalmazására (2.2. alfejezet), valamint az Európai vizsgálatok eredményeit a NUTS szinteken (2.3. alfejezet).

2.1. Területi modellek a szakirodalomban

A területi kapcsolatok gazdasági modellekben való alkalmazásának kiinduló tanulmányának Paelinck és tsai. (1979) munkája tekinthető, mely regionális és multiregionális szinteken vizsgálta a területi hatásokat és azok korrigálását a keresztmetszeti regressziós modellekben. A szerzőpáros az európai városok urbanizációjának témakörében írt több tanulmányt is, ezek vizsgálata során figyelték meg és alkalmazták később a térökonometriai modelleket. Azon kérdéskört vizsgálták, hogy miért virágzik az egyik város, miért stagnál a másik és ezek a típusok miért csoportosulnak a térben. Egyik 1979-es cikkükben a városok növekedésének és stagnálásának dinamikai modelljét kívánták felállítani. A céljuk az volt, hogy megmutassák, nem lehet csak a városon belüli szabályozásokat figyelembe venni a szabályozóknak, a külső dinamika is fontos, s ezen keresztül figyelmet kell fordítani a környező települések fejlődési görbéire is (Klaassen és Paelinck 1979).

A cikk hatására kialakult az urbanizáció hatásait vizsgáló tanulmánykör, például a Tokioi Chuo-line régiót (Fukatsu és Kawashima 1999), a finn régiókat (Hirvonen, Hiraoka és Kawashima 1999) illetve Iránt vizsgálva (Laudor és Tabriztchi 1975).

Az urbanizációt vizsgáló tanulmányok mellett a területi hatásokat figyelembe vevő modellek megjelentek a tudásexternáliák területén is. Ilyen volt például Acs, Anselin és Varga (1997) kutatása, melyben területi spillover hatásokat vizsgáltak az egyetemi kutatások és a technológiai innovációk között. Az endogén gazdasági növekedés elmélet alapján a tudás felhalmozódik, és áttérjed a termelési kapacitásba a technológiai változásokon keresztül. A tanulmányban megfigyelt direkt és indirekt hatások nem csupán megyei szinten maradtak meg, hanem azok határain túl is felismerhetőek voltak. Az egyetemek hatása ez alapján kettős, egyrészt az egyetemi kutatásokkal alátámasztják a technológiai kutatásokat, másrészt magasan képzett szakemberek oktatása által a humán tőkét erősítik az adott régiókban. Varga (2000) tanulmánya kiegészíti ezt az eredményt azzal a megfigyeléssel,

hogy a tudástranszferek esetében is megfigyelhető az agglomerációs hatás, azaz a fejlett technológiai foglalkoztatási lehetőségek számát is figyelembe kell venni a tudásexternáliák vizsgálata során. Az agglomerációban található ilyen típusú lehetőségeknek ell kell érniük egy minimális szintet ahhoz, hogy a korábban megfigyelt hatások érvényesülni tudjanak. A területileg összefüggő tudásexternáliák kutatásának egy másik eleme Acs, Anselin és Varga (2002) tanulmánya, mely a közgazdaságilag hasznosítható új tudás jelenlétét és területi eloszlását vizsgálja Amerikában. Az innovációk és a szabadalmak területi eloszlásában pozitív autokorrelációt mutattak ki az 1982-es adatokon, melyek főleg az Egyesült Államok keleti és dél-nyugati partvidékén összpontosulnak. Eredményeik alapján a szabadalmak száma megbízható mutatója az innovációs tevékenységnek mind a regresszió illeszkedési pontosságában, mind a becsült paraméterek struktúrájában bekövetkezett változásokban, mind a területi függőség típusában. Megegyező évben mutatták ki a kapcsolatot Ausztria területén is Fischer és Varga (2003) tanulmányában. Az eredmények összeegyeztethetők a korábban bemutatott hatásokkal, kiegészítve azzal, hogy a spillover hatások a távolság növekedésével csökkennek.

A területi modellek azonban nem maradnak az urbanizáció és a tudásexternáliák vizsgálatánál, széles skálán figyeltek már meg területi hatásokat az elmúlt évtizedekben, melyek alkalmazása is széleskörű. Ezekből mutatok be most néhányat. Aten (1996) az országok között vizsgált ár-mennyiség vizsgálatokat egészítette ki a területi autokorrelációs vizsgálatokkal. Ebben az esetben a megfigyelések közötti távolságot nemcsak fizikai formában mérik, hanem a két ország közötti kereskedelem nagyságával arányos módon meghatározott területi súlymátrix segítségével is. Case (1991) az indonéz piacokon fogyasztott rizs mennyiségét vizsgálta, azon belül pedig azt, hogy mennyire klasztereződnek területileg azon kerületek, ahol a lakosság nem otthon termeszti a terméket, hanem azt a piacon vásárolja.

Gyódi és Nawaro (2021) az európai nagyvárosok esetében mutatta ki, hogy az Airbnb árak vizsgálata során nem lehet figyelmen kívül hagyni a területi függéseket, hiszen a városközponttól, a látnivalóktól és a gazdasági negyedről való távolság, valamint a tömegközlekedési eszközök elérhetősége mind olyan területileg összefüggő változók, melyek befolyásolják az adott lakás értékét. A területi modellekkel kapott eredmények nem csupán a turizmusnak adnak értékes információkat, hanem a várostervezésnek is, hiszen az Airbnb térnyerésével egyre több kiadó lakást alakítottak át rövidtávú kiadásokra, mely

átalakította a lakáspiacot a legtöbb nagyvárosban. Ezen hatások vizsgálata és az ez ellen történő szabályozások során fontos inputot tudtak az ilyen típusú modellek adni.

Zhang, Wang és Yao (2021) kínai városok esetében mutattak ki negatív kapcsolatot a nagysebességű vasúti közlekedés elérhetősége és a ködszennyezés mértéke között. A hatás az erőforráselosztás hatékonyságának növelésén, az ipari struktúraváltás elősegítésén, valamint a technológiai innováción keresztül érvényesül. Emellett kimutatták, hogy a szennyezés csökkentésének hatása eltér az egyes városokban. A tanulmány rámutat arra, hogy a helyközi közlekedés fejlesztésével hogyan lehet a környezetszennyezést csökkenteni, amely a világ minden táján, de Kínában különösen sürgető probléma.

A tanulmányok sorából természetesen az egészségügy, és azon belül a koronavírus-járvány okozta lezárások, a halálozási ráták és az egyes betegségek okainak feltárásai sem maradhattak ki. Alexakis, Eleftheriou és Patsoulis (2021) tanulmányában a lezárások és a piaci indexek közötti kapcsolatot vizsgálták 45 piaci index bevonásával. Eredményeik alapján kimutatható a negatív kapcsolat a lezárások súlyossága és a piaci index hozama között. Spijker és tsai. (2021) Kolumbia példáján keresztül vizsgálja a területi korrelációt a halálozási rátákban. Eredményeik rámutatnak arra, hogy az ország mely területein van a legnagyobb szükség a beavatkozásra. Hu, Chun és Griffith (2020) tanulmányában 10 év mintáján keresztül vizsgálták a mellrák kialakulását, mely változó esetében kimutatják a pozitív autokorrelációt, valamint azt, hogy a modellek pontosabb előrejelzést adnak, amennyiben ezt a tulajdonságot figyelembe veszik.

Alkalmazási területek még - a teljesség igénye nélkül - az innovációs hatékonyság (Ribeiro és Bao 2021) és a piaci együttmozgások vizsgálata. Piaci indexek esetében sem lehet eltekinteni a területi hatásoktól, hiszen a kereskedelem által az egyes országok összekötöttsége nem elhanyagolható. Különösen igaz ez válságok idejében, melyre példa Thaiföld szerepe az ázsiai válság továbbgyűrűződésében (Asgharian, Hess és Liu 2013).

2.2. Modellspecifikációk a szakirodalomban

A tanulmány módszertani alapját a globális Moran-féle I-próba adja, melyet egyszerűsége és független mintán vett ismert eloszlása miatt széles körben alkalmaznak. Ennek ellenére a limitációit ritkán emelték ki a szakirodalomban. Ennek egyik példája Li, Calder és Cressie (2007) tanulmánya, melyben a szerzők a területi autoregresszív modell esetében mutatják be, hogy a Moran I-próba által adott pontbecslés csak abban az esetben elfogadható, ha a paraméter értéke 0 közeli. Emiatt definiálják az APLE tesztet, mely az autokorreláció tesztelésére és a területi autoregresszív modell paraméterbecslésére is alkalmazható abban az esetben is, ha a regressziós paraméter értéke nem 0 közeli.

Tiefelsdorf és Boots (1995) tanulmányukban a Moran I-próba kismintás adatokon vett eloszlását határozták meg numerikus eljárások segítségével. A Moran-féle I-próba elméleti megalapozásakor csupán az aszimptotikus tulajdonságait vizsgálták a tesztstatisztika értékének, annak kismintás adatokon vett eloszlása nem volt ismert.

Jackson és tsai. (2010) tanulmányukban a Moran-féle I-próba pontosságát vizsgálták abban az esetben ha a vizsgált alapváltozó homogén. Bevezették egy módosított változatát a tesztstatisztikának, mely homogén adaton a szimulációs eredményeik alapján jobb eredményeket ért el globális és lokális szinten is. A tanulmány 2010-es megjelenése óta mégsem csökkent a területi vizsgálatok során az eredeti tesztstatisztika alkalmazási gyakorisága. Szintén a tesztstatisztika egy korrigált értékét vezette be a szakirodalomba Oden (1995), mely a népességeloszlással korrigálja a mutató értékét. Ezen mutatót több esetben hivatkozták, mint a másik korrekciót, azonban általános elterjedésben nem tudta az eredeti próbát utolérni.

Jelen tanulmányhoz hasonlóan Bivand, Müller és Reder (2009) is a Moran-féle I-próba értékenységvizsgálatát végezték el. A szerzők azt vizsgálták, hogy a teszt által felállított feltételezések nem teljesülése esetén hogyan teljesít a mutató. Eredményeik alapján ezen feltételek megsértése esetén a teszt túlbecsüli a területi autokorrelációt.

A legújabb modellezési tendenciák sem függetlenek a területi adatok által okozott problémáktól. Gépi tanulási modellek esetében a felhasznált változósám jóval nagyobb, mint a hagyományos ökonometriai modellekben. Ennek okán a területi adatok okozta problémák is gyakran előfordulnak ezen modellekben, melyek korrigálására ebben az esetben is szükség van. Amennyiben a gépi tanulási modellekben nem korrigálják a területi autokorrelációkat, akkor az túltanuláshoz vezethet, azaz a modell magasabb pontosságot fog

mutatni a tanuló adathalmazon, mint amekkorát egy független adathalmazon az alkalmazása során el fog tudni érni. Salazar és tsai. (2022) a szakirodalomban alkalmazott különböző adatfelosztás által korrigáló módszerek (validációs halmaz képzése, keresztvalidáció, illetve utóbbi különböző struktúrák figyelembe vételével történő alkalmazása) rugalmatlanságát és túlzott paraméterillesztési nehézségeit emelik ki tanulmányukban. Egy olyan 8 lépésből álló keresztvalidációt javasolnak, mely ezen problémák megoldására hivatott, s mely figyelembe veszi a modell valós alkalmazása során feldolgozott adatok vélt varianciáját is.

2.3. Európai körkép

Az alábbi alfejezetben bemutatom, hogy a NUTS szinteket milyen gazdasági kérdések megválaszolására alkalmazták a szakirodalomban, ezáltal szemléltetve, hogy Európa így módon történő felosztásának vizsgálata releváns és megalapozott. A bemutatott szakirodalmi példák szemléltetést nyújtanak, jelen tanulmány kutatási kérdéseihez nem nyújtanak támpontot.

Benedek és Veress (2013) a romániai NUTS 2 és NUTS 3 szinteken vizsgálták, hogy a román GDP/fő alapján az egyes területek egymáshoz konvergálnak-e, valamint hogy Románia az Európai Unió tagállamaihoz tud-e közeledni. Eredményeik alapján a 2000-2010-es időszakban az ország és az EU közötti konvergencia bizonyítható, mely főleg a 2008 előtti időszaknak köszönhető, míg a NUTS régiók között a gazdasági fejlettségbeli eltérés növekedett.

Auzina-Emsina és Ozolina (2017) Litvánia 6 db NUTS 3 régióján keresztül vizsgálta az ország versenyképességét a csúcstechnológiai iparágakon (gyógyszergyártás, számítógépgyártás) keresztül. Eredményeik rámutatnak arra, hogy ezen litván iparágak a globális szintén is versenyképesek, a versenyképességük stabil, mely jó indikátora a jövőbeni fejlődésnek.

Paluchowski (2012) Németország és az Egyesült Királyság példáján keresztül kutatja a régiós egyenlőtlenségek kérdéskörét, valamint a regionális népességszám előrejelzéseket. Kiemeli, hogy a társadalmi folyamatok nemcsak a társadalmi, hanem a gazdasági változások mozgatórugói is, így a NUTS szinteken vizsgálja a népességnövekedést, a népességszámot és a korfát.

Paas és tsai. (2007) jövedelemkonvergenciát mutattak ki NUTS 3 szinten az Európai Unió

legnagyobb növekedési hulláma előtt 1995 és 2002 között az új tagok és a régi tagok között is.

Dauvin és Grzybowski (2014) paneladatok segítségével vizsgálták 2006 és 2010 között a szélessávú hálózat kiépítettségét 27 európai uniós tagállamban. Eredményeik alapján a DSL és a kábelmodem NUTS 1 szintű részaránya jól mutatja az egész terület hozzáférését a szélessávú internethez.

Bertocchi és tsai. (2021) javaslata alapján a turisztikai fenntarthatóság vizsgálata során NUTS 2 szintet kell alkalmazni. Egy regionális fenntarthatósági indikátort definiálnak mind a 281 NUTS 2 régióra. A tanulmány ezáltal hozzájárul ahhoz, hogy a NUTS beosztások által Európa területei minél több szempontból egységesen vizsgálhatóak legyenek.

Mura és tsai. (2021) az EU ETS (szén-dioxid kvótakereskedelmi rendszer) szabályozásainak hatásait vizsgálták a NUTS szinteken, az egyes iparágak szén-dioxid egyenérték kibocsátásának és az iparág GDP értékének hányadosának vizsgálata által.

A fenti szemléltető példák is mutatják, hogy Európa, és legfőképp az Európai Unió tagállamainak vizsgálata során kézenfekő területi aggregálási szint a NUTS rendszer. A következő fejezetben ezen rendszer sajátosságait mutatom be, az adatbázis kialakításával egyetemben.

3. Adatbázis kialakítása

Az alábbi fejezetben bemutatom a NUTS rendszer sajátosságait (3.1. alfejezet), valamint az adattisztítási lépéseket, a városi adatok letöltésének folyamatát, valamint az általuk kialakított adatbázist (3.2. alfejezet).

3.1. NUTS rendszer sajátosságai

A NUTS rendszer (Nomenclature des unités territoriales statistiques – Területi statisztikai rendszer nómenklatúrája) az EUROSTAT által létrehozott egységes területi rendszer, mely Európa területét osztja régiókra 3 különböző szinten. A jelenleg hatályos beosztás 2021. 01. 01. óta él és az egyes szinteken az Európai Unió tagállamaiban 92-242-1166 régiót különít el. A beosztás hierarchikus módon tartalmazza az Európai Unió tagállamai mellett az Egyesült Királyság, Izland, Törökország, Norvégia és Svájc területét.

A rendszer célja, hogy az európai régiós statisztikák gyűjtését egységesítse. A legfelső szinten az úgynevezett fő társadalmi-gazdasági régiók vannak (3 000 000 - 7 000 000 lakos), mint például hazánk esetében az ország Duna és Tisza által harmadolt régiói (Közép-Magyarország, Dunántúl, Alföld és Észak-Magyarország). A középső szinten az úgynevezett alaprégiók találhatók (800 000 - 3 000 000 lakos), melyek a regionális politikák alkalmazásához igazodnak. Hazánkat ez a NUTS 2 beosztás 8 részre osztja (Budapest, Pest, Közép-Dunántúl, Nyugat-Dunántúl, Dél-Dunántúl, Észak-Magyarország, Észak-Alföld, Dél-Alföld). A legalsó szinten a kisrégiók találhatók (150 000 - 800 000 lakos), mely esetünkben a 19 megyét és Budapestet jelenti (EUROSTAT 2022b).

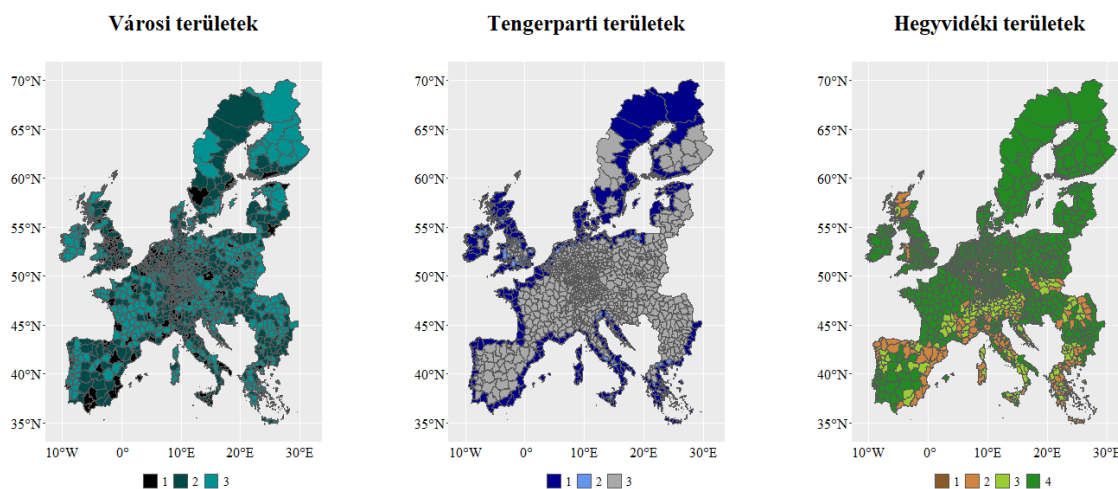
Bár hazánk esetében a régiók kialakítása a megszokott képet mutatja, néhány terület esetében láthatók szokatlan megoldások a felosztásban. Ennek egyik legfőbb oka, hogy a régiókat elsődlegesen a lakosságszám figyelembevételével alakították ki, mely több esetben nem esik egybe a földrajzi adottságok által kialakított területekkel. Ennek egyik nyilvánvaló esete a kontinens külső, földrajzilag tagolt határainál lévő szigetcsoportok összevonása, azonban az érdekesebb eseteket más helyen kell keresnünk. Több ízben előfordul, hogy egy régió nem összefüggő területileg, több kisebb részből áll, s ezen részek egyikét egy másik régió közrefogja. Ilyenre számtalan példát látunk Spanyolországban (például Kasztília és León autonóm közössége, Navarra tartománya és Valencia).

A rendszer minden régióról egységes módon gyűjti össze az alapvető statisztikákat (né-

pesség, születéskor várható élettartam, foglalkoztatási ráta, munkanélküliségi ráta, oktatásban való részvétel aránya, GDP mutatók stb.), készít egységes összefoglalókat a területekről, valamint az adatszolgáltatása által lehetővé teszi a kontinens egységes empirikus elemzését.

Az adatbázis kialakításához az EUROSTAT (2022a) által szolgáltatott *.shp* kiterjesztésű vektorgrafikus reprezentációt használtam, mely 0.1M-es felbontásban, EPSG:4326 kódolásban tartalmazza a 2021-es NUTS területi adatokat. A NUTS szintek területi adatait úgynevezett multipolygonok segítségével tárolja, valamint ezen túl információt ad az egyes területek típusáról is a NUTS 3-as szinten (hegyvidéki, városi vagy tengerparti). Ennek bemutatását a 1. ábra tartalmazza. Az ábra a földrajzilag távol eső területeken kívül tartalmazza az egész adatsokaságot, melyet az EUROSTAT szolgáltat. Az ábra célja az adatbázis bemutatása, a későbbi területi szűréseket, valamint azok okait a következő alfejezet tartalmazza.

NUTS 3 szint területi osztályozása és tipológiai

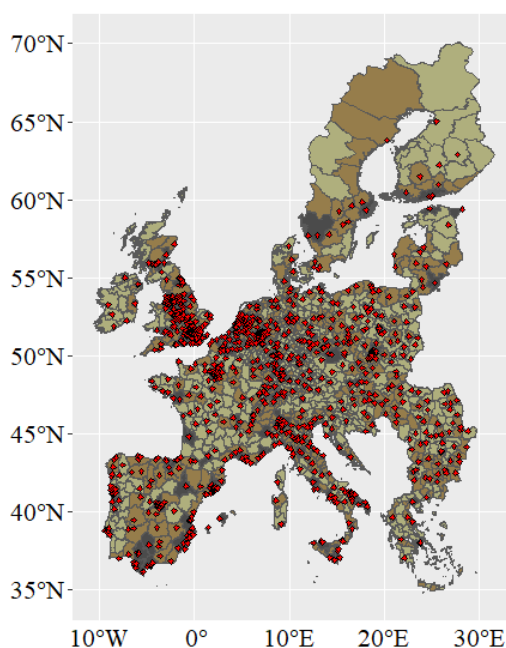


1. ábra. NUTS 3 régiók tipológiai. (Forrás: Saját szerkesztés)

3.2. Felhasznált adatbázisok

Az adatbázis tisztítása során a külterületek régióitól (például Madeira), valamint néhány országtól megváltam, hogy a szomszédsági kapcsolatok, valamint a közúti távolság értelmezése (Izland, Ciprus, Málta) ne okozzon nehézséget, valamint az átlagosnál nagyobb belső távolságok (Norvégia) és az európai uniós tagállamoktól való jelentős gazdasági eltérés (Törökország) ezen területek esetében ne torzítsák az eredményeket. A felsorolt országok szűrésén túl egyes szinteken tisztított területek kódjait a Függelék tartalmazza. A területi adatok mellett a terület nagyvárosainak listáját is letöltöttem az EUROSTAT honlapjáról, hogy az egyes régiók közötti kapcsolatokat a legfőbb városok közötti átlagos közúti távolsággal is reprezentálni tudjam. Az egyes városok pontos geokoordinátáit a *ggmap* R-package, valamint a Google API segítségével töltöttem le a Google Maps felületéről (Kahle és Wickham 2013). A vizsgált városok elhelyezkedését a 2. ábra szemlélteti.

Városi területek és a terület fő városai



2. ábra. NUTS 3 régiók városi csoportosítása, valamint a vizsgált városok.

(Forrás: Saját szerkesztés)

A pontos geokoordináták, valamint a régiókat reprezentáló multipolygonok koordinátái alapján egyértelműen meghatározható, hogy mely város mely régió területén található. Annak ellenére, hogy a területek kialakításánál a lakosságszám volt a meghatározó szempont, jelentős eltéréseket tapasztalhatunk abban, hogy az egyes területeken hány nagyobb

vagy jelentősebb város helyezkedik el. Ez egy korlátot jelent majd a későbbiekben a közúti távolság alapján definiált területi súlymátrixok alkalmazásánál.

NUTS 1-es szinten a város/régió arány a $[0; 41]$ intervallumon oszlik el. Nem található az EUROSTAT besorolása szerint jelentősebb település Kréta szigetén, a Finn szigetvilágban és Lichtenstein területén, míg a spanyol Este régióban (Spanyolország keleti tengerparti szakasza) 41 várost is kiemel (Barcelona, Tarregona, Valencia).

A területek tagozásával, azaz az alacsonyabb NUTS szintek létrehozásával egyre több olyan terület alakul ki, melynek nem része jelentősebb város. NUTS 2-es szinten ezek száma 23 (például Sjælland - Dánia legnagyobb szigete, Ionia Nisia - Jón szigetek Görögország, Voreio Aigaio - Görögország szigetvilágának a török partszakasz mellett fekvő része, Åland - Finn szigetvilág, Pest - Pest megye, Mellersta Norrland - Svédország), míg a legtöbb várost tartalmazó terület 27 várost jelöl ki (Ile-de-France - Párizs és környéke). Nagyobb városok esetében megfigyelhető a városészek kiemelése is (például Párizs, Barcelona és London esetében).

Pest megye példáján jól látszik az egyik kiváltó ok, a terület földrajzi értelemben vett középpontját adó nagyváros (ebben az esetben Budapest) a lakosságszám alapján adott besorolásban egy külön kategóriába esik, így az azt körülvevő régióban (ebben az esetben Pest megyében) nem található jelentős város. Másik kiváltó okra jó példa akár a görög szigetcsoport, akár a Svédország és Finnország között elhelyezkedő Finn szigetvilág. Ezen esetekben a földrajzi tagoltság okán nem tudtak kialakulni olyan nagyvárosok, melyek az apró szigetek gazdaságában jelentős befolyással tudnának lenni. A harmadik kiváltó ok szintén természeti, a nemzeti parki területeken, mint például Mellersta Norrland területén a természeti adottságok és azok védelme miatt nem alakultak ki nagyvárosok.

NUTS 3 szinten kettő kivétellel minden régió 10 vagy annál kevesebb várost foglal magában, míg 657 esetben nem található jelentős város az adott területen. Ez a kimagasló szám a NUTS 3-as szint vizsgálatának korlátaiban nagy szerepet játszik.

4. Módszertan

A térökonometria módszertanának közgazdasági alkalmazásokban történő elterjedésének egyik okaként az egyének közötti interakciók hatásainak közgazdasági modellekben (Romer-Lucas endogén növekedésemélet, Krugman gazdaságföldrajz, Aoki makroökonómia) történő egyre gyakoribb megjelenését azonosították (Anselin 2001). Az elméletek alapján a gazdasági fejlődés és a gazdasági folyamatok egy része a szereplők térbeli tudásátadásának köszönhető, így ezen hatásoktól az új típusú modellekben nem lehet eltekinteni. A térbeli hatásokat reprezentáló területi súlymátrix meghatározásánál ezért ezen kapcsolatokat kell figyelembe venni.

Két földrajzi terület kapcsolati fokának meghatározása egy értékkel nehézkes feladat, hiszen figyelembe vehetjük a közös határszakasz jelenlétét, a városok távolságát vagy akár a területek mértani középpontjainak távolságát is. A területi súlymátrix definiálásánál ezen definíciók bármelyikét alkalmazhatjuk a térbeli közelség reprezentálására. A globalizáció egyre nagyobb mértékével azonban nem csupán a fizikai térben közel lévő emberek között alakulhatnak ki kapcsolatok, ezért a területi súlymátrix létrehozását nem lehet a fizikai térre korlátozni, hanem lehetőséget kell adni az egyes gazdasági kapcsolatokat leíró változók (pl.: kereskedési volumen, politikai változók, vállalatok leányvállalatainak száma) alapján készített reprezentációknak is. A fejezet ezen definíciók által létrehozott területi súlymátrixok bemutatásával és alkalmazási lehetőségeivel foglalkozik.

A fejezet a térökonometria alapfogalmait (4.1. alfejezet), a térbeli területi súlymátrix meghatározásának módjait, azok NUTS szintekre vonatkoztatott előnyeit és hátrányait (4.2. alfejezet), a térbeli autokorreláció tesztelésére alkalmas mutatókat (4.3. alfejezet), valamint annak kezelésére alkalmas módszereket (4.4. alfejezet) mutatja be.

A módszertani blokk Varga (2002) tanulmánya alapján készült.

4.1. Alapfogalmak

1. Definíció. *Térbeli heterogenitás (spatial heterogeneity)*

A megfigyelési egységek egyes változói között tapasztalható pozitív vagy negatív korreláció. Térbeli heterogenitás esetében a térbeli kapcsolatok kezelésének elhanyagolása heteroszkedaszticitást eredményez a regressziós modellben, így a probléma kezelése a heteroszkedaszticitás kezelésévé egyszerűsödik, melyre számos ökonometriai megoldás alkalmazható.

A globális folyamatok sok esetben természeti erőforrások mentén alakulnak ki (pl.: folyók mentén, hegységek lábánál). Az ezen területen kialakult gazdasági folyamatokat a politika által húzott országhatárok vagy a közigazgatási régióhatárok több részre oszthatják, mely aggregációs probléma egy magyarázata a térbeli összefüggőség kialakulásának. A történelem során dinamikusan változó országhatárok is okozhatnak ilyen típusú kapcsolatot, hiszen a politikai váltás nem feltétlenül eredményez gazdasági változást az adott területen, kiváltképp abban az esetben, ha a gazdasági tevékenység a terület földrajzi adottságai által adott. Mindezek indukálják, hogy a két szomszédos ország között az adott változóban kapcsolat lesz felfedezhető. A térbeli adatok kapcsolatának egy másik oka lehet a szomszédos országok közötti gazdasági együttműködés, a kitelepített leányvállalatok általi kapcsolat vagy a kereskedelmi, politikai viszony kialakítása is (Varga 2002).

2. Definíció. *Térbeli autokorreláció (spatial autocorrelation)*

A megfigyelési egységek vizsgált változójának függvényyszerű összefüggése az egyes térbeli helyeken mért értékei között. Pozitív autokorreláció jelenlétének egyik ismertetőjele a térbeli klaszterek kialakulása (pl.: Szilícium-völgy, Neza-Chalco-Itza nyomornegyed). Negatív autokorreláció által kialakított sakktáblaszerű elrendezésre intuitív magyarázat és ismertetőjel nem adható.

4.2. Területi súlymátrix

A térbeli adatok kétdimenziós volta miatt a kapcsolatokat nem lehet egydimenziós kapcsolattá alakítani és pusztán lexikografikus módon sorrendet felállítani közöttük, kivéve ha az adatok térben homogének vagy függetlenek (Varga 2002). A térökonometriai modellek alkalmazásának egyik sarokköve a vizsgált megfigyelési egységek közötti térbeli egységek megfelelő módon való tömörítése. Két térbeli város, régió, ország közötti távolság vagy akár a közöttük lévő kapcsolat egy számértékkel való leírása nehézkes és nagy felelősséget ró a modellezőre annak megválasztásával. A területi súlymátrixok minden esetben a megfigyelési egységek számával megegyező dimenziójú mátrix, így 1-1 egység közötti kapcsolat reprezentálására mindössze egy érték rendelhető. Definíciószerűen minden területi súlymátrix minden eleme egy nemnegatív valós szám, valamint a mátrix nyoma 0. Az önnön hatások statisztikai értelmezhetőségének hiánya miatt minden súlymátrix diagonálisának összes eleme zérus. A területi hatásokat reprezentáló területi súlymátrix definiálására az alábbi módszerek találhatók a szakirodalomban.

3. Definíció. *Bináris szomszédságmátrix (binary contiguity matrix) $W = [\omega_{ij}]$*

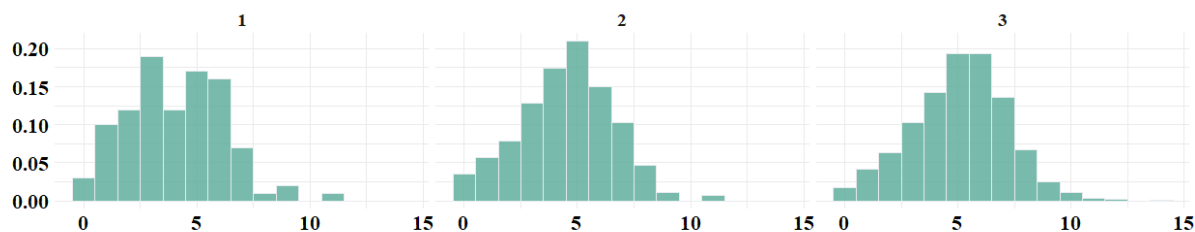
$$\omega_{ij} = \begin{cases} 1 & \text{ha az } i. \text{ és a } j. \text{ megfigyelési egység} \\ & \text{rendelkezik közös határral} \\ & \text{és } i \neq j \\ 0 & \text{egyébként} \end{cases}$$

A legegyszerűbb modell a kapcsolat leírására, szimmetrikus mátrixot eredményez. Ellenérv a használata mellett például Szlovénia, melynek 4 szomszédja van (Ausztria, Olaszország, Horvátország és Magyarország), viszont a gazdasági kapcsolatainak mértéke, valamint a közös határainak hossza jelentősen eltér a 4 szomszédja esetében, míg a bináris szomszédságmátrix mindegyiknek ugyanakkora súlyt ad.

Szintén problémásnak tekinthető az exklávék általi szomszédságok kérdésköre. Erre példa lehet Oroszország esetében a Kalanyingrádi terület, a NUTS szinteknél maradva a görög szigetvilág területeinek viszonya vagy Portugália és a Kanári-szigetek kapcsolata. Az első példánál maradva felmerülhet ilyen esetekben a kérdés, hogy ha egy régió, mely Oroszország teljes területének kevesebb mint 0,1%-t teszi ki, mégis szomszédossá teszi azt két

új állammal (Litvánia, Lengyelország), mennyire ad reális képet az egyes területek, jelen esetben országok kapcsolatáról és az egymás gazdasági mutatóira való hatásairól.

Szomszédos területek aránya NUTS szinteken



3. ábra. Szomszédos területek számának eloszlása az egyes NUTS szinteken.

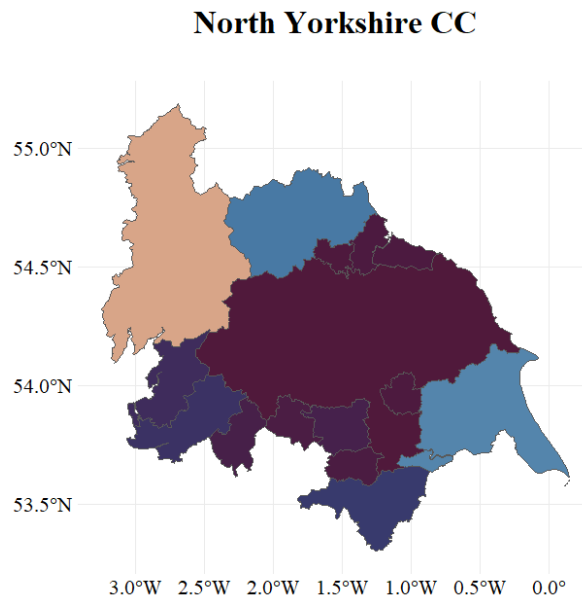
(Forrás: Saját szerkesztés)

Ahogy a 3. ábrán is láthatjuk a NUTS régiók esetében is jelentős eltéréseket tapasztalhatunk abban, hogy az egyes területek hány szomszéddal rendelkeznek abban az esetben, ha a szomszédságot a területeket leíró poligonok érintkezése alapján definiáljuk. Az elemzés során ezt a definíciót 3 különleges kapcsolattal egészítettem ki: Franciaország és az Egyesült Királyság megfelelő területeit összekötő Csalagúttal, a Svédország és Dánia területeit összekötő Øresund híddal, valamint a Szicíliát és Dél-Olaszországot összekötő Messina-híddal. Ezen három területpár esetében okozott problémát az, hogy bár a közúti összeköttetés a területek között megtalálható, és a közlekedési lehetőségek jól kihasználtak, a poligon alapú reprezentáció nem tudja a kapcsolatot megfelelően leírni.

Ezek alapján a vizsgált területek elsőfokú szomszédai alapján kialakított adattáblában az európai régiók alábbi tulajdonságai láthatóak. Legtöbb szomszéddal az Egyesült Királyság középső részén, York városától északra elhelyezkedő North Yorkshire CC rendelkezik, összesen 14-gyel. A régió NUTS 3 szinten helyezkedik el a hierarchiában. Nagyobb lélekszámú vagy jelentősebb település nem található a területén, ez indokolja a régiósan kiugró nagyságú területét, mely a szomszédjainak kimagasló számát is indokolja.

A lakosságszám alapján meghatározott területi csoportok indukálják, hogy az egyes régiók területi méreteiben nagyon eltérnek egymástól. Egy kietlenebb, kevés nagyvárost tartalmazó, területileg nagy régió mellett sok esetben található több kisebb régió, mely a szomszédok számát jelentősen megnöveli. Ilyen például a már előbb említett North Yorkshire CC (ahogy a 4. ábrán is látható), vagy többek között a Rhein-Pfalz-Kreis régió Németországban és a South and West Derbyshire és Staffordshire CC az Egyesült Király-

ságban. A legmagasabb szinten (NUTS 1) a kiugró - 11 szomszédval rendelkező terület - Niedersachsen régiója Németországban, mely többek között Hannover, Göttingen és Celle városait foglalja magába.



4. ábra. North Yorkshire CC példája a szomszédsági számokra. (Forrás: Saját szerkesztés)

A skála másik végén azon régiók találhatók, melyek 1-1 nagyváros köré rajzolódtak, s ennek következtében egy nagyobb régió belül kerültek kialakításra, s így egyetlen szomszédjuk az őket körülölelő nagyrégió. Ilyenre példa Berlin, Bremen és Madrid is. A szomszéd nélküli területeket a szigetek alkotják, mint Kréta, Korzika, vagy Åland, a Finnország és Svédország között elhelyezkedő körülbelül 6000 szigetből álló szigetcsoport.

Területi méretekben nagyságrendi eltéréseket tapasztalhatunk az egyes régiókban. NUTS 3 szinten a legnagyobb területtel Svédország Norrbotten megyéje (Norrbottens Län) rendelkezik ($105\,180,6\text{ km}^2$), míg ezen a szinten a legkisebb terület csupán $19,8\text{ km}^2$ (Tower Hamlets kerület - London). A lakosság szerinti határok által megszabott területi régiók így tényleges földfelszíni méretben jelentősen eltérnek egymástól. A területi méretek eloszlása jobbra elnyúló, balra ferde, a kigróan magas értékek ($20\,000\text{ km}^2$ -t meghaladó méretek) 7 esetben találhatók Svédországban, 5 esetben Finnországban és egy esetben Spanyolországban.

A példák mutatják, hogy milyen eltéréseket eredményezhet az, ha csupán a közös határ jelenlétével próbáljuk meg leírni azt, hogy két terület hatással van-e egymásra, valamint az, hogy minden szomszéd hatását ugyanakkorának értékeljük. A továbbiakban olyan te-

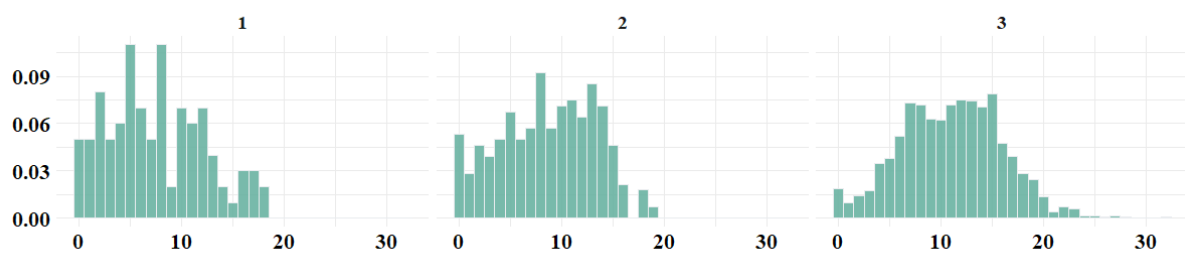
rületi súlymátrixokat mutatok be, melyek ezen ellentmondások feloldására törekednek. Fontos megjegyezni, hogy a térbeli kapcsolatok ilyen módon történő tömörítése sosem eredményez tökéletes megoldást. A földrajzi kapcsolatok leírása minden esetben nehézkes feladat, gondoljunk csupán a földfelszín kétdimenziós térképekre való vetítésének nehézségeire. Következésképp a területi hatások leírására is számtalan megoldás alkalmazható, s ezek között nem lehet rangsort felállítani, csupán kiemelni lehet az egyes megoldások előnyeit és hátrányait.

4. Definíció. Bináris szomszédságmátrix n fok szerint $W = [\omega_{ij}]$

Tegyük fel, hogy az $(n - 1)$ fok szerinti bináris szomszédságmátrix már definiált, jelölje ezt $\hat{W} = [\hat{\omega}_{ij}]$, valamint jelölje $\hat{\hat{W}} = [\hat{\hat{\omega}}_{ij}]$ a 3. definíció során definiált bináris szomszédságmátrixot ($n = 2$ esetén $\hat{W} = \hat{\hat{W}}$). Ekkor az n fok szerinti bináris szomszédságmátrix az alábbi módon definiált:

$$\omega_{ij} = \begin{cases} 1 & \text{ha } \hat{\omega}_{ij} = 1 \\ 1 & \text{ha } \exists \hat{j} : \hat{\omega}_{i\hat{j}} = 1 \text{ és } \hat{\hat{\omega}}_{\hat{j}j} = 1 \\ 0 & \text{egyébként} \end{cases}$$

Másodszomszédos területek aránya NUTS szinteken



5. ábra. Másodszomszédos területek száma NUTS szinteken. (Forrás: Saját szerkesztés)

Egy régió tehát két esetben szomszédos n fok szerint egy másik régióval, vagy $n - 1$ -ed fokban szomszédos vele, vagy létezik olyan régió, mely az egyikkel $n - 1$ -ed fokban szomszédos, míg a másikkal első fokban szomszédos.

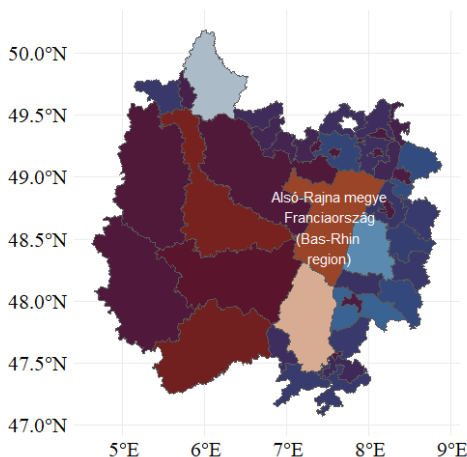
Ezen meghatározással egy probléma lehet például, hogy a NUTS szintek lakosságszám alapján lettek meghatározva, így területi méretekben jelentősen eltérnek. Minél magasabb

fokú szomszédságot vesszünk figyelembe, annál nagyobb eltérések lehetnek a tényleges távolságok között.

Az eltérések tovább erősödnek, ha nemcsak az elsőfokú szomszédságot, hanem a másodfokú szomszédságot is figyelembe vesszük. A másodfokú szomszédok száma ebben az esetben már a $[0; 32]$ intervallumon helyezkedik el. Legtöbb szomszéddal Franciaország legkeletibb régiója, Alsó-Rajna megye rendelkezik, összesen 32-vel. A régió NUTS 3 szinten helyezkedik el a hierarchiában, jelentősebb települése Strasbourg. Szomszédai között francia, német, svájci és luxembourgi területek is megjelennek.

A skála másik végén a korábban már említett régiók mellett olyan területek is megjelennek, melyek földrajzilag vannak jobban elszigetelve a kontinens többi részétől, mint például a Théba városa köré húzott Kentriki Elláda régió vagy az Athéni régió, Dánia Midtjylland területe, Szicília, Észtország vagy Svédország NUTS 1-es régiója. Ezen területek egy része azért rendelkezik kevés szomszéddal, mert a vizsgált terület „szélén” helyezkedik el, azaz első és másodszomszédainak egy része már nem esik a vizsgált területi részbe, míg a másik része a kontinens földtörténeti tagozódásának következtében került egy olyan helyre, ahol szárazföldi kapcsolatai limitálva vannak.

Bas-Rhin régió szomszédai és másodszomszédai



6. ábra. Alsó-Rajna megye, Franciaország és másodszomszédai. (Forrás: Saját szerkesztés)

Alsó-Rajna régió példáján keresztül is jól látható a 6. ábrán, hogy a másodszomszédos területek kiterjedése kelet-nyugati és észak-déli irányban sem vethető össze. Horizontálisan és vertikálisan is nagyságrendi különbség van a régiók tényleges távolságában. Ez mutatja, hogy milyen eltérések lehetnek a területi súlymátrix által leírt kapcsolatokban,

ha a szomszédságot binárisan vagy valamely távolságdefiníció alapján definiáljuk.

A vizsgált területeken NUTS 1-es szinten átlagban 1 152 km-re találhatók egymástól a régiók középpontjai, ha az összes régió-regió kapcsolatot figyelembe vesszük. Ezen érték a NUTS 2-es szinten 1 147 km, míg a NUTS 3-as szinten 1 034 km. Átlagos értékben tehát nincs jelentős eltérés a szintek között. Amennyiben minden régió esetében csak a legközelebbi régiós középpont értékét vesszük figyelembe, akkor ezen átlagos régió-regió távolsági értékek 159 - 98 - 38 km-re változnak.

A szakirodalomban a korai alkalmazások legnagyobb része szomszédsági mátrixot használ, mivel megyei, régiós vagy országos szintű adatokkal dolgozik, ahol a tudásátadások ilyen szinteken mennek végbe, valamint ahol a határok tiszta volta miatt könnyű definiálni a szomszédságot. Ezzel ellentétben a háztartások és a vállalatok szintjén, valamint a globalizáció miatt egyes gazdasági és társadalmi változók esetében a szomszédságot definiálni nehezebb. Ennek oka, hogy szabálytalanul vannak a térben elosztva a megfigyelési egységek, valamint az adott egységre hatással lévő egységek nem feltétlen egyeznek meg a fizikai szomszédokkal. Emiatt ilyen szinten a távolsággal arányosan csökkenő súlyokat adó mátrixok használhatóbbnak bizonyulnak ilyen esetekben (Bell és Bockstael 2000).

A NUTS rendszer példáin keresztül is láthattuk, hogy milyen ellentmondásokat és problémákat szül a bináris szomszédság mátrix alkalmazása. A következőkben olyan területi súlymátrixok bemutatásával foglalkozom, melyek valamely távolságdefiníció inverzének segítségével írják le a területek közötti kapcsolatokat fokát.

5. Definíció. *Inverz távolság alapú területi súlymátrix (inverse distance weight matrix)*

$$W = [\omega_{ij}]$$

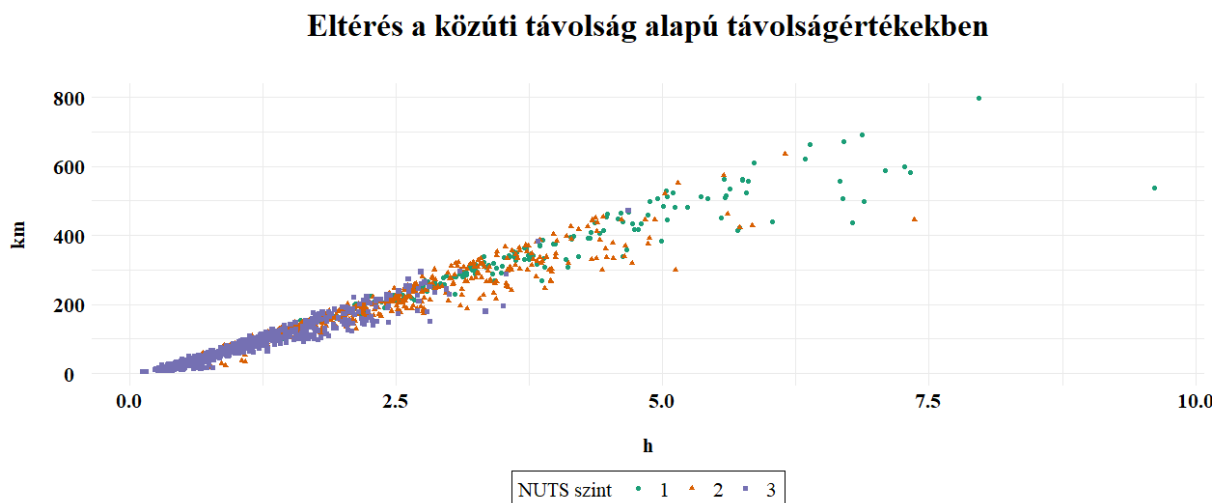
Tegyük fel, hogy két régió között a távolság definíciója adott, jelölje ezt d_{ij} , valamint legyen $c \in \mathbf{R}^+$ küszöbtávolság és $n \in \mathbf{N}$ súlyozási érték adott. Ebben az esetben legyen az inverz távolság alapú területi súlymátrix az alábbi módon definiált:

$$\omega_{ij} = \begin{cases} \left(\frac{1}{d_{i,j}}\right)^n & \text{ha } i \neq j \text{ és } d_{ij} \leq c \\ 0 & \text{ha } i \neq j \text{ és } d_{ij} > c \\ 0 & \text{ha } i = j \end{cases}$$

A távolság ($d_{i,j}$) definiálása ebben az esetben még utat nyit többféle alkategóriának. Két

területi egység között a távolságot lehet mérni a területek mértani középpontjainak távolságával, a két terület pontjai között mérhető legkisebb távolsággal (matematikai értelemben vett távolságdefiníció két poligon között), a legnagyobb városok távolságával, a közigazgatási középpontok távolságával, a város-város párok közti átlagos távolsággal (ahol szintén lehet többféleképpen értelmezni két város távolságát) vagy súlyozott város-város távolsággal (ahol súly lehet a város népessége, területe vagy valamely gazdasági mutatója).

Az elemzés során a mértani középpontok távolságát, a két terület pontjai között mérhető legkisebb távolságot, a város-város párok közötti átlagos közúti távolságot órában, valamint kilométerben mérve alkalmaztam mint távolságdefiníciókat a területi súlymátrixok létrehozásához. A súlyozott távolság, valamint a legnagyobb városok, illetve közigazgatási középpontok távolságának mellőzésének oka, hogy nem érhető el olyan adat, mely alkalmazható lenne a teljes vizsgált területre. A francia városok esetében 2017-ben, a lengyel, román, svéd és az Egyesült Királyságbeli városok esetében pedig 2018-ban megszűnt az adatszolgáltatás, 5 évvel korábbi adatok felhasználása pedig torzítaná az eredményeket.

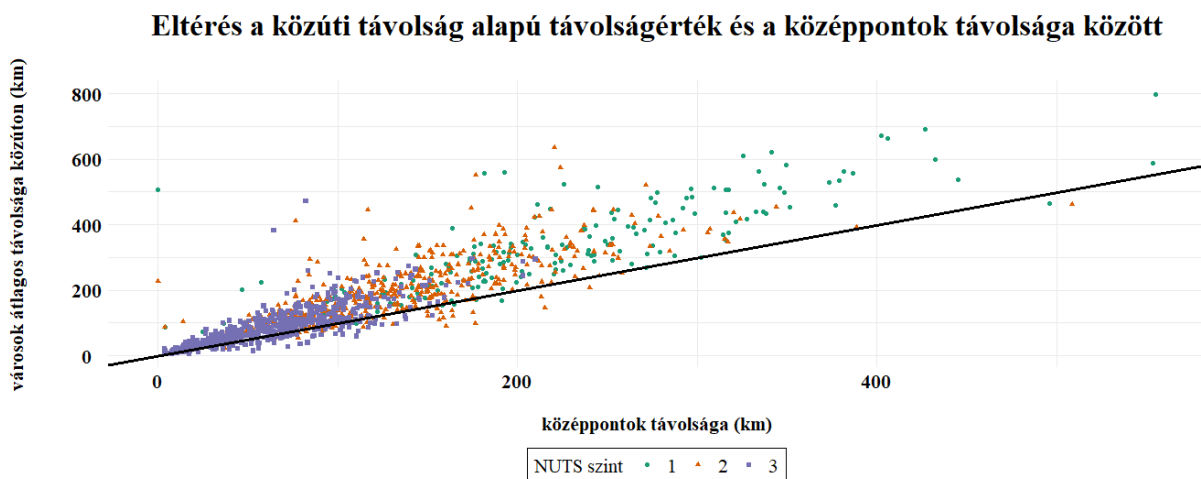


7. ábra. *Eltérés a közúti távolság alapú távolságértékekben.* (Forrás: Saját szerkesztés)

A 7. ábra mutatja, hogy a NUTS 1-2-3 szinteken a bináris szomszédságmátrix alapján szomszédosnak ítélt területek között mekkora az átlagos város-város távolság közúton kilométerben és órában. Az ábrán az adatpontok lineáris elhelyezkedése mutatná azt, hogy a közúton kilométerben és órában mért távolságdefiníciók ugyanazt a kapcsolatot írják le, a két külön mérés közötti átváltás csak egy affin transzformáció. A közúti távolság alapú megközelítések már az idő - távolság értelmezés cseréjével is változtatnak a területi

súlymátrix által reprezentált kapcsolati rendszeren, hiszen a két érték közötti átváltás nem konstansszoros. Az eltérés legfőképp a NUTS 1-es szinten figyelhető meg, azonban kiugróan eltérő értékek mindhárom szinten megtalálhatóak.

A 8. ábra mutatja, hogy NUTS 1-2-3 szinteken a bináris szomszédsságmatrix alapján szomszédosnak ítélt területek között mekkora az átlagos város-város távolság közúton kilométerben mérve, valamint mekkora a területek mértani középpontjainak légvonalban mért távolsága kilométerben. Az ábrán az adatok lineáris elhelyezkedése mutatná azt, hogy a két távolságdefiníció ugyanazt a kapcsolatot írja le. A régiók mértani középpontjainak légvonalbeli távolsága és a városok átlagos közúti távolsága között sem lineáris a kapcsolat, azaz a kétféleképpen definiált területi súlymátrix közötti eltérés nem írható le egy konstans szorzóval. Ez garantálja, hogy a két mátrixszal végzett statisztikai tesztek nem ugyanazt az eredményt fogják adni.



8. ábra. *Eltérés a közúti távolság alapú távolságértékek és a középpontok távolsága között.* (Forrás: Saját szerkesztés)

Elsőre talán meglepő módon néhány esetben a közúti átlagtávolság kisebb, mint a légvonalbeli távolság. Ennek magyarázata, hogy nem ugyanazoknak a pontoknak a távolságát vizsgáltam légvonalban és közúton. Abban az esetben, ha a jelentős városok mind a két terület esetében a közös határhoz közelebb helyezkednek el, a város-város átlagos távolság adhat kisebb értéket, mint a mértani középpontok légvonalbeli távolsága.

6. Definíció. *K legközelebbi szomszéd alapú területi súlymátrix* (*k nearest neighbor distance weight matrix*) $W = [\omega_{ij}]$

Tegyük fel, hogy két régió között a távolság definíciója adott, jelölje ezt d_{ij} , valamint

legyen $k \in \mathbf{N}^+$ paraméter adott. Ebben az esetben legyen a K legközelebbi szomszéd alapú területi súlymátrix az alábbi módon definiált:

$$\omega_{ij} = \begin{cases} \frac{1}{k} & \text{ha } |\{\hat{j} \mid d(i, \hat{j}) \leq d(i, j); i \neq j\}| \leq k \\ 0 & \text{egyébként} \end{cases}$$

A $|\{\hat{j} \mid d(i, \hat{j}) \leq d(i, j); i \neq j\}|$ azon halmaz számosságát jelöli, melynek elemei azon i sorindextől különböző indexek, melyek esetében az i . és j . megfigyelési egység távolsága kisebb, mint az épp figyelembe vett távolságérték. Amennyiben ennek száma kisebb vagy egyenlő, mint a k paraméter értéke, úgy a területi súlymátrix értéke 1, ellenkező esetben 0.

Ezen területi súlymátrix azon régióknak ad pozitív súlyt, melyek egy adott távolságdefiníció alapján a legközelebb helyezkednek el a referencia régióhoz. Ez garantálja, hogy minden régió ugyanannyi szomszéddal rendelkezzen, azaz minden régióra k másik régió befolyásoló hatását vizsgáljuk. A NUTS szinteken korábban már bemutatott szomszédságszám eltéréseket ezen meghatározás korrigálja. Ennek következtében előfordulhat olyan eset, hogy két régió ugyan rendelkezik közös határral, jelentősebb városaik is közel helyezkednek el egymáshoz, az adott távolságdefinícióban mégsem tekinti őket ez a fajta területi súlymátrix szomszédosnak, s ezáltal a közöttük lévő kapcsolat nem lesz reprezentálva.

Természetes módon a fent definiált területi súlymátrix esetében is felmerül a lehetőség, hogy a $d(i, j)$ módon definiált értékeket különböző módokon értelmezzük. Ennek vizsgálatába ezen elemzés keretein belül nem térek ki. Az elemzés során ebben az esetben a távolságot, mint a két terület mértani középpontjainak légvonalbeli távolságát értem. Szintén nem térek ki arra, hogy k mely értékét célszerű választani, az elemzés során a $k = 4$ értéket alkalmazom.

$k < 14$ esetén a 5. ábrán is bemutatott North Yorkshire CC esetében lenne olyan terület, mely ugyan szomszédos az adott régióval, mégsem tekinti ezen mátrix őket úgy, mint egymásra hatással lévő területek. A 6. ábrán bemutatott példában pedig egy földrajzilag nem határos, de középpont-középpont távolságban kisebb értéket felvevő terület lesz Alsó-Rajna megye szomszédja, míg egy nagyobb területű és ennek következtében távolabbi mértani középponttal rendelkező régió hatásait nem fogja reprezentálni.

A korábban már említett területi méret - lakosságszám - jelentősebb városok száma át-

váltás ezen esetben is megjelenik a mátrix definiálása következtében a reprezentációban. Egy kevésbé lakott és ezáltal nagyobb területű régiónak a mértani középpontja nagyobb távolságra lesz a szomszédjaitól, mint annak kisebb területű szomszédjai. Ahogy korábban már bemutattam, nagyvárosok több esetben alkotnak önálló régiót, melyek területe jóval kisebb, mint az átlagos területi érték az adott NUTS szinteken. Ennek következtében a nagyvárosok alkotta régiók ezen definíció által nagyobb arányban lesznek szomszédként definiálva, mint a nagyobb területű, kevésbé sűrűn lakott régiók. Gazdasági hatásokat vizsgálva azonban ezen tulajdonság nem hátrányos, hiszen a nagyvárosok vonzáskörzetükre gyakorolt hatásait meg tudja ragadni.

A mátrix definíciójából adódóan nem szimmetrikus, ellentétben az összes eddig definiált területi súlymátrixszal. Ezen tulajdonság is hozzájárul ahhoz, hogy a nagyvárosok környezetében előnyösnek mondhassuk alkalmazását, hiszen ezen területek több régióra vannak hatással, mint amennyi az ő gazdasági viselkedésüket befolyásolja.

7. Definíció. *Gazdasági mutató ($\Delta = [\delta_{ij}]$) alapú területi súlymátrix $W = [\omega_{ij}]$*

Legyen $\Delta = [\delta_{ij}]$ valamely gazdasági mutató (pl.: i országban bejegyzett cégek j országban található leányvállalatainak száma) definiált és legyen $c \in \mathbf{R}$ küszöbszám rögzített. Ekkor a területi súlymátrix értékeit definiáljuk az alábbiak szerint:

$$\omega_{ij} = \begin{cases} \delta_{ij} & \text{ha } \delta_{ij} \geq c \text{ és } i \neq j \\ 0 & \text{egyébként} \end{cases}$$

Az elemzés során ilyen típusú területi súlymátrix definíciót nem vizsgállok, mivel ilyen aggregációs szinten nincs olyan elérhető adatbázis, mely reálisan mutatná a régiók közötti gazdasági kapcsolatokat. Ezt a típusú definíciót országok közötti kapcsolatok vizsgálatakor érdemes alkalmazni, hiszen ilyen szinten az adatok aggregálásával és azok elérhetőségével nem szokott probléma lenni.

8. Definíció. *Sorstandardizált területi súlymátrix (row standardized weight matrix)*

Legyen $W = [w_{ij}]$ NxN-es mátrix a korábbiak alapján definiált. Ekkor jelölje $W' = [\omega'_{ij}]$ a sorstandardizált területi súlymátrixot, ahol $\forall i, j$

$$\omega'_{ij} = \frac{\omega_{ij}}{\sum_{j=1}^N \omega_{ij}}$$

A definíció alapján az alábbi két állítás minden sorstandardizált területi súlymátrixra igaz:

1. Állítás.

$$\forall i, j : 0 \leq \omega'_{i,j} \leq 1$$

2. Állítás.

$$\forall i : \sum_{j=1}^N \omega'_{i,j} = 1$$

A sorstandardizált területi súlymátrixok használatának egyik előnye, hogy a maximum likelihood becslések csak ezen mátrix használata esetén megoldhatóak (Varga 2002). Ezzel ellentétben hátrányuk, hogy a súlymátrix minden sorát különböző értékkel osztja, ezáltal a mátrixon nem lineáris transzformációt hajt végre. A transzformáció célja, hogy minden megfigyelési egységre ugyanakkora összsúlyt adjon, viszont ezáltal a mátrix elveszti szimmetrikus voltát, valamint értelmezhetőségét is (Bell és Bockstael 2000).

4.3. Területi autokorreláció tesztelése

A teljes adatsor figyelembevételével készített globális területi autokorreláció tesztelésének egyik lehetséges próbája a *Moran-féle I-próba* (Moran 1948; Cliff és Ord 1981). Legyen a $W = [\omega_{ij}]$ területi súlymátrix adott, ekkor a Moran-féle I-próba tesztstatisztikája a következő alakot ölti:

$$I \doteq \left(\frac{N}{\sum_{i=1, j=1}^N \omega_{ij}} \right) \left(\frac{\sum_{i=1, j=1}^N \omega_{ij} (x_i - E(x))(x_j - E(x))}{\sum_{i=1}^N (x_i - E(x))^2} \right)$$

A tesztstatisztika első tagja $\left(\frac{N}{\sum_{i=1, j=1}^N \omega_{ij}} \right)$ azt garantálja, hogy a területi súlymátrix konstans értékkel történő szorzása ne változtassa meg a tesztstatisztika értékét. A megelőző alfejezetben ennek okán tértem ki külön arra, hogy az egyes távolságdefiníciókkal adott inverz távolság alapú területi súlymátrixok konstansszoros eltérést adnak-e vagy sem.

Legyen $k \neq 0$ tetszőleges valós szám. Legyen $W^* = [\omega_{ij}^*] = k * [\omega_{ij}]$. Ekkor az ez alapján számolt I-próbastatisztika értéke egybeesik az eredeti tesztstatisztika értékével:

$$\begin{aligned} I([\omega_{ij}^*]) &= \left(\frac{N}{\sum_{i=1, j=1}^N \omega_{ij}^*} \right) \left(\frac{\sum_{i=1, j=1}^N \omega_{ij}^* (x_i - E(x))(x_j - E(x))}{\sum_{i=1}^N (x_i - E(x))^2} \right) = \\ &= \left(\frac{N}{\sum_{i=1, j=1}^N k * \omega_{ij}} \right) \left(\frac{\sum_{i=1, j=1}^N k * \omega_{ij} (x_i - E(x))(x_j - E(x))}{\sum_{i=1}^N (x_i - E(x))^2} \right) = \\ &= \left(\frac{N}{k * \sum_{i=1, j=1}^N \omega_{ij}} \right) \left(\frac{k * \sum_{i=1, j=1}^N \omega_{ij} (x_i - E(x))(x_j - E(x))}{\sum_{i=1}^N (x_i - E(x))^2} \right) = \\ &= \left(\frac{N}{\sum_{i=1, j=1}^N \omega_{ij}} \right) \left(\frac{\sum_{i=1, j=1}^N \omega_{ij} (x_i - E(x))(x_j - E(x))}{\sum_{i=1}^N (x_i - E(x))^2} \right) = \\ &= I([\omega_{ij}]) \end{aligned}$$

A tesztstatisztika második tagja a területi súlymátrix által szomszédosnak ítélt területeken (pozitív súllyal ellátott értékek esetében) vizsgálja, hogy ezen területeken a megfigyelt változó értékei együttesen átlagon felüliek vagy átlagon aluliak-e (pozitív területi autokorreláció) vagy az átlagon felüli és az átlagon aluli értékek szomszédsága a mérvadó a területen (negatív területi autokorreláció).

Legyen a $W' = [\omega'_{ij}]$ sorstandardizált területi súlymátrix adott, ekkor a Moran-féle I-próba tesztstatisztikája a következő alakot ölti:

$$I' \doteq \frac{\sum_{i=1, j=1}^N \omega'_{ij} (x_i - E(x))(x_j - E(x))}{\sum_{i=1}^N (x_i - E(x))^2}$$

$$E(I') = \frac{-1}{N-1}$$

$$\lim_{n \rightarrow \infty} E(I') = 0$$

Döntési szabály a területi autokorreláció globális jelenlétéről:

$$\text{Területi autokorreláció} \doteq \begin{cases} \text{pozitív} & \text{ha } I' > \frac{-1}{N-1} \\ \text{negatív} & \text{ha } I' < \frac{-1}{N-1} \end{cases}$$

A Moran-féle I-próba szignifikanciája normális eloszlás feltételezése mellett:

$$Z_i \doteq \frac{I - E(I)}{\sigma(I)}$$

z-próbája alapján adódik. A próba szignifikanciájának tesztelése a normális eloszlás feltételezése nélkül, a permutációs módszerrel, azaz a változók értékeinek lokációkra mért véletlen permutációinak létrehozásával ($[x_{ij}^*]$):

$$\frac{|I([x_{ij}^*])| \geq |I| + 1}{|I([x_{ij}^*])| + 1}$$

A vizsgált érték azon permutációk száma + 1, melyek esetében a képzett teststatisztika értéke nagyobb vagy egyenlő, mint a kapott teststatisztika érték, osztva az összes permutáció számával.

A globális próbán túl lehetőség van a lokális csoportosulás tesztelésére is, melyre lokális próbák (Lokális Moran próba) állnak rendelkezésre. A tanulmány a globális próbákra összpontosít, így a lokális próbákat annak keretében nem részletezem.

4.4. Területi autokorreláció kezelése

A területi autokorreláció kezelésére az elemzés nem tér ki, azonban fontos megemlíteni, hogy nem csupán a területi autokorreláció globális jelenlétének kimutatása során szükséges a területi súlymátrix alkalmazása, hanem azt követően a regressziós modellek illesztése során is. Az alfejezet röviden bemutatja a területi autokorrelációs modelleket.

A területi autokorreláció kezelésére használható módszerek például a térbeli késleltetés modellje és a térbeli hiba autokorreláció modellje.

A térbeli késleltetés modellje a késleltetési operátor definiálása által adott. Legyen $W = [\omega_{ij}]$ sorstandardizált területi súlymátrix adott

$$L^s(x_i) \doteq \sum_{j=1}^N \omega_{ij} x_j$$

azaz a területi súlymátrix értékei által súlyozott átlaga a többi megfigyelésnek. Ekkor a regressziós egyenlet

$$y = \rho W y + X\beta + \varepsilon$$

Hipotéziseket a térbeli késleltetések hatásaira (ρ) és az exogén változók hatásaira (β) állíthatunk. Megfelelő modellspecifikáció esetén a hibatagok (ε) független, azonos eloszlású, 0 várható értékű, σ^2 szórású és területileg autokorrelálatlan változók. $W y$ endogenitása miatt az OLS módszer nem alkalmazható a paraméterbecslésre, hiszen az torzított és inkonzisztens lenne (Anselin 1988). A paraméterbecslésre a maximum likelihood módszer, az instrumentális változók módszere és az általános momentum módszer alkalmazható.

A maximum likelihood becslés alkalmazásának feltétele a reziduum normalitása. A logaritmusos likelihood függvény alakja és az általa adott paraméterbecslések

$$L = \ln|I - \rho W| - \frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{(y - \rho W y - X\beta)'(y - \rho W y - X\beta)}{2} \sigma^2$$
$$(\beta, \rho, \sigma^2) = \arg \max_{\beta, \rho, \sigma^2} L$$

ahol I ebben az esetben az $N \times N$ -es egységmátrix, $||$ a determináns jele. Ezen változók mellett y mintájának valószínűsége maximális.

A GIS (Geographical Information System - egyes egységek térbeli elhelyezkedésének adatait tároló adatbázisok) terjedésével indult a nagymintás adatok elemzésének igénye. A

regressziós egyenletek illesztése a maximum likelihood becslés esetében a területi súlymátrix determinánsának becslését indukálja, ami N növelésével egyre számításigényesebb. A regressziós paraméterbecslés során a determinánsbecslést több alkalommal kell elvégezni, mely nagyméretű mátrixok esetében problémát okoz. A determináns meghatározása helyett a mátrix kifejtése megoldaná a számításigény problémáját, azonban a sajátértékek numerikus megközelítésének pontossága N növelésével csökken, így óva intenek 400 megfigyelés felett a maximum likelihood becsléstől (Kelejian és Prucha 1999; Bell és Bockstael 2000).

Ezzel szemben az általános momentum módszer alkalmazása esetén csupán mátrixszorzások és mátrixnyomok számításigényével kell kalkulálni, mely kisebb nagyságrendű, mint a maximum likelihood becslés esetén fellépő determináns és sajátértékmeghatározás. A módszer másik előnye, hogy nem szükséges a konzisztens paraméterbecsléshez a hibatagok normalitása. A standard hibák nem becsülhetők vele a térbeli autokorrelációt jelölő paraméterre, azonban amíg a becslés konzisztens, addig béta és sigma becslései megtartják a nagymintás tulajdonságaikat, így a modell alkalmazható és a standard közgazdasági modellek hipotézisei felállíthatóak és tesztelhetőek (Kelejian és Prucha 1999; Bell és Bockstael 2000).

5. Eredmények

Az alábbi fejezetben 9 különböző módon definiált területi súlymátrix (5.1) segítségével mutatom be, hogy mennyiben és milyen módon térnek el a Moran-féle tesztstatisztika értékek a súlymátrix definiálása által. Ahogy az előző fejezetben már bemutattam, a súlymátrix definiálása nemcsak a területi autokorreláció kimutatására, hanem a regressziós egyenletek illesztésére és a paraméterbecslések hatásosságára is kihat. Az elemzés keretein belül ennek tesztelésére nem térek ki.

A mátrix definiálása által keltett hatásokat 4 módon szemléltetem. Első körben a definiált mátrixok determinánsának és sajátértékeinek eltérése által (5.2), majd a területileg autokorrelálatlan változók kimutatása által (5.3), ezt követően gazdasági és társadalmi adatokon szemléltetve (5.4), végezetül pedig szimulációk segítségével (5.5) mutatom be, milyen módon befolyásolja az egyes NUTS szinteken a Moran-féle I tesztstatisztika értékét a területi súlymátrix megválasztása.

5.1. Alkalmazott területi súlymátrixok

Az alábbi alfejezetben pontosan definiálom, hogy az általam vizsgált 9 súlymátrix mely bemutatott nagy kategóriába sorolható, illetve hogy milyen paraméterbeállításokkal alkalmaztam azokat. A definíciók által adott területi súlymátrixok NUTS 1, NUTS 2 és NUTS 3 szintekre vonatkoztatott adatai az alábbi GitHub linken elérhetőek.

1. Területi súlymátrix. *Elsőfokú szomszédság*

A 3. definíció által adott szomszédsági mátrix. A közös határ definícióját a régiókat leíró multipoligonok közös pontja mellett a Csalagút, az Øresund híd és a Messina-híd adja. Abban az esetben, ha a régió területe nem összefüggő, azaz leírásához egynél több poligonra van szükség, minden poligon által alkotott közös határszakasz számít.

2. Területi súlymátrix. *Másodfokú szomszédság*

A 4. definíció által adott szomszédsági mátrix $n = 2$ esetén, ahol a referencia bináris szomszédságmátrixot az előző területi súlymátrix adja.

3. Területi súlymátrix. Régióközpont távolság

Az 5. definíció által adott inverz távolságalapú súlymátrix, ahol a távolságdefiníció a régiókat reprezentáló multipolygonok mértani középpontjainak légvonalbeli távolsága (km) alapján definiált, $c = \infty$ és $n = 1$. Abban az esetben, ha a régió területe nem összefüggő, azaz leírásához egynél több poligonra van szükség, a mértani középpontot a legnagyobb területű poligon mértani középpontjaként definiáltam.

4. Területi súlymátrix. Közúti távolság az elsőfokú szomszédok között kilométerben

Az 5. definíció által adott inverz távolságalapú súlymátrix, ahol a távolságdefiníció a régiók városai között létrejövő összes párosítás átlagos közúti távolsága kilométerben, valamint $n = 1$. A küszöbtávolság ebben az esetben nem egy $c \in \mathbf{R}^+$ értékkel, hanem az elsőfokú szomszédság mátrix által adott. A közúti távolság mérésére a Google Maps adatai az irányadók, melynek letöltését a Google Distance Matrix API segítségével értem el. Abban az esetben, ha valamely régióban nem volt jelentősnek tekinthető város az EUROSTAT adatbázisa alapján, a közúti távolságot a régióközpont távolsággal helyettesítettem.

5. Területi súlymátrix. Közúti távolság az elsőfokú szomszédok között órában

Az 5. definíció által adott inverz távolságalapú súlymátrix, ahol a távolságdefiníció a régiók városai között létrejövő összes párosítás átlagos közúti távolsága órában, valamint $n = 1$. A küszöbtávolság ebben az esetben nem egy $c \in \mathbf{R}^+$ értékkel, hanem az elsőfokú szomszédság mátrix által adott. A közúti távolság mérésére a Google Maps adatai az irányadók, melynek letöltését a Google Distance Matrix API segítségével értem el. Abban az esetben, ha valamely régióban nem volt jelentősnek tekinthető város az EUROSTAT adatbázisa alapján, a közúti távolságot a régióközpont távolság órára átváltott értékével helyettesítettem. A közúti adatok alapján az átváltási szorzó 0,01103 NUTS 1-es szinten, 0,01152 NUTS 2-es szinten és 0,0124 NUTS 3-as szinten, mely egybeesik egy 90, 87, valamint 80 km/h-s átlagsebességgel.

6. Területi súlymátrix. Közúti távolság a másodfokú szomszédok között kilométerben

Az 5. definíció által adott inverz távolságalapú súlymátrix, ahol a távolságdefiníció a régiók városai között létrejövő összes párosítás átlagos közúti távolsága kilométerben, valamint $n = 1$. A küszöbtávolság ebben az esetben nem egy $c \in \mathbf{R}^+$ értékkel, hanem a másodfokú szomszédság mátrix által adott. A közúti távolság mérésére a Google Maps adatai az

irányadóak, melynek letöltését a Google Distance Matrix API segítségével értem el. Abban az esetben, ha valamely régióban nem volt jelentősnek tekinthető város az EUROSTAT adatbázisa alapján, a közúti távolságot a régióközéppont távolsággal helyettesítettem.

7. Területi súlymátrix. *Közúti távolság a másodfokú szomszédok között órában*

Az 5. definíció által adott inverz távolságalapú súlymátrix, ahol a távolságdefiníció a régiók városai között létrejövő összes párosítás átlagos közúti távolsága órában, valamint $n = 1$. A küszöbtávolság ebben az esetben nem egy $c \in \mathbf{R}^+$ értékkel, hanem a másodfokú szomszédság mátrix által adott. A közúti távolság mérésére a Google Maps adatai az irányadóak, melynek letöltését a Google Distance Matrix API segítségével értem el. Abban az esetben, ha valamely régióban nem volt jelentősnek tekinthető város az EUROSTAT adatbázisa alapján, a közúti távolságot a régióközéppont távolsággal órára átváltott értékével helyettesítettem.

8. Területi súlymátrix. *Poligonok közötti legrövidebb távolság*

Az 5. definíció által adott inverz távolságalapú súlymátrix, ahol a távolságdefiníció a régiókat leíró multipoligonok pont-pont (standard euklideszi értelemben vett) légvonalbeli távolságainak minimuma, valamint $n = 1$. A küszöbtávolság ebben az esetben $c = 500 \text{ km}$ értékkel adott. Ebben az esetben a közös határral rendelkező területek távolsága definíciószerűen 0, ezért ezen értékeket 0,5-es érték helyettesíti a súlymátrixban (az inverz távolságok maximális értéke 0,4281113, ezért esett a választás a 0,5-es értékre).

9. Területi súlymátrix. *4-legközelebbi szomszéd*

A 6. definíció által adott K legközelebbi szomszéd alapú területi súlymátrix, ahol $k = 4$, valamint a két régió közötti távolságot a mértani középpontjuk légvonalbeli távolsága adja. Abban az esetben, ha a régió területe nem összefüggő, azaz leírásához egynél több poligonra van szükség, a mértani középpontot a legnagyobb területű poligon mértani középpontjaként definiáltam.

5.2. Matematikai összehasonlítás

Az alábbi alfejezetben a három NUTS szinten kialakított 9-9-9 mátrix tulajdonságait hasonlítom össze. Az alfejezet célja, hogy bemutassa, a kilencféleképpen definiált mátrixok matematikai értelemben nem hasonlók, azaz nem ugyanazt a lineáris kapcsolatot reprezentálják.

Az alfejezet során a mátrixok determinánsát, rangját, sajátértékeinek típusát, a mátrixban található nemnulla elemek arányát mutatom be táblázatos formában. A tisztán zérus elemeket tartalmazó oszlopok számát, valamint az ezek eltávolítása után kapott determináns értéket is bemutatom a táblázatokban (1, 2, 3. táblázat). Ezt követően a valós sajátértékekkel rendelkező mátrixok sajátértékeit 1-1 ábrán szemléltetem. A determináns, a rang, valamint a sajátérték a matematikai értelemben vett hasonlóság eldöntéséhez szükséges. A nemnulla elemek aránya, a nulloszlopok száma, valamint az ennek eltávolítása után kapott determináns az értelmezések során fog előkerülni a következő alfejezetekben.

9. Definíció. Hasonlóság

Azt mondjuk, hogy az $N \times N$ -es A és B mátrixok hasonlóak, ha létezik olyan invertálható C mátrix, hogy $A = C^{-1}BC$.

3. Állítás. Ha A és B hasonlóak, akkor

- $|A| = |B|$, (azaz $\det(A) = \det(B)$)
- $\text{rang}(A) = \text{rang}(B)$
- A és B sajátértékei megegyeznek.

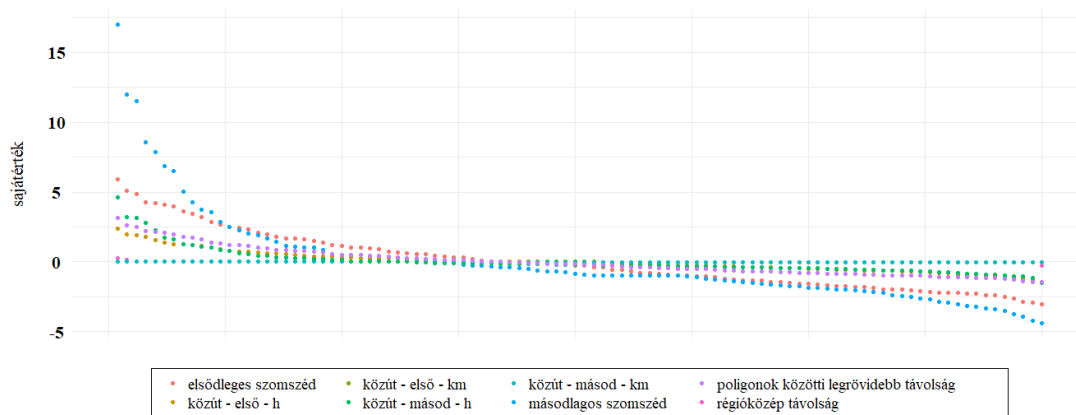
Amennyiben két mátrix ebben az értelemben hasonló lenne, a segítségükkel számított Moran-féle I tesztstatisztika megegyezne, azaz a későbbiekben alkalmazott eredmények és szimulációk megegyeznének.

	Determináns	Rang	Sajátérték	Determináns*	Nemnulla elemek aránya	Zérus oszlopok száma
Elsőfokú szomszédság	0	95	valós	0	4,3%	3
Másodfokú szomszédság	0	96	valós	-9,5 e-6	12,3%	3
Régióközpont távolság	-2,3 e-223	100	valós	-2,3 e-233	99%	0
Közúti távolság az elsőfokú szomszédok között km-ben	0	96	valós	0	4,3%	3
Közúti távolság az elsőfokú szomszédok között h-ban	0	97	valós	0	4,3%	3
Közúti távolság a másodfokú szomszédok között km-ben	0	97	valós	8,87 e-238	12,3%	3
Közúti távolság a másodfokú szomszédok között h-ban	0	97	valós	4,1 e-48	12,3%	3
Legrövidebb távolság	4,6 e-29	100	valós	4,6 e-29	29,4%	0
4-legközelebbi szomszéd	0	99	komplex	0	4%	0

1. táblázat. NUTS 1 szinten definiált területi súlymátrixok tulajdonságai.

(Forrás: Saját szerkesztés) *: zérus oszlopok és sorok elhagyása után

Mátrix sajátértékek NUTS 1 szinten valós sajátértékek esetén



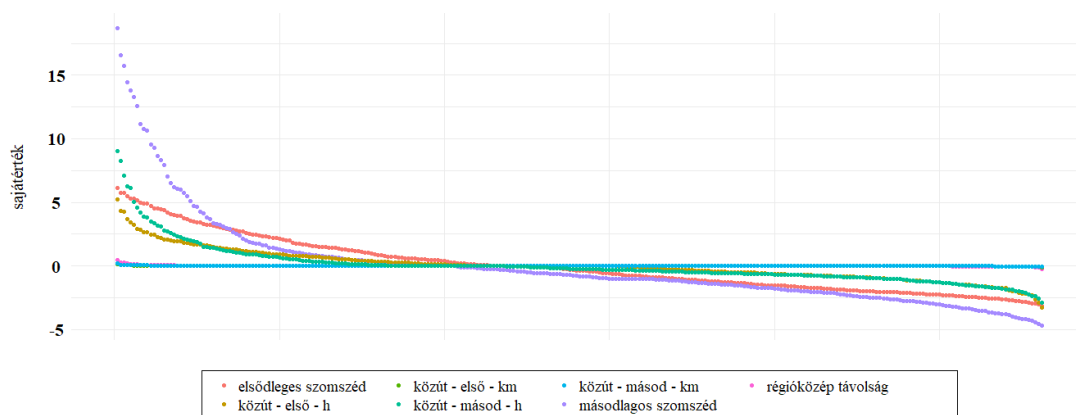
9. ábra. Mátrix sajátértékek NUTS 1 szinten. (Forrás: Saját szerkesztés)

A NUTS 1-es adatokon látható az 9. ábrán, hogy a sajátérték egyezése csak a *Közúti távolság az elsőfokú szomszédok között kilométerben*, a *Közúti távolság a másodfokú szomszédok között kilométerben*, illetve a *Régióközpont távolság* szerint definiált mátrixok között merülhet fel. A 1. táblázat megfelelő értékei segítségével azonban elvethető az a hipotézis, hogy ezen három definíció által felírt mátrixok egymáshoz páronként hasonlóak lennének.

	Determináns	Rang	Sajátérték	Determináns*	Nemnulla elemek aránya	Zérus oszlopok száma
Elsőfokú szomszédság	0	270	valós	0	1,7%	10
Másodfokú szomszédság	0	271	valós	-1,6 e+37	5,05%	10
Régióközpont távolság	0	281	valós	0	99,6%	0
Közúti távolság az elsőfokú szomszédok között km-ben	0	271	valós	0	1,7%	10
Közúti távolság az elsőfokú szomszédok között h-ban	0	271	valós	5 e-161	5,05%	10
Közúti távolság a másodfokú szomszédok között km-ben	0	271	valós	0	5,05%	10
Közúti távolság a másodfokú szomszédok között h-ban	0	271	valós	-6,6 e-129	5,05%	10
Legrövidebb távolság	-5,0 e-74	281	valós	-5,0 e-74	24,4%	0
4-legközelebbi szomszéd	0	277	komplex	0	1,4%	0

2. táblázat. NUTS 2 szinten definiált területi súlymátrixok tulajdonságai. (Forrás: Saját szerkesztés) *: zérus oszlopok és sorok elhagyása után

Mátrix sajátértékek NUTS 2 szinten valós sajátértékek esetén



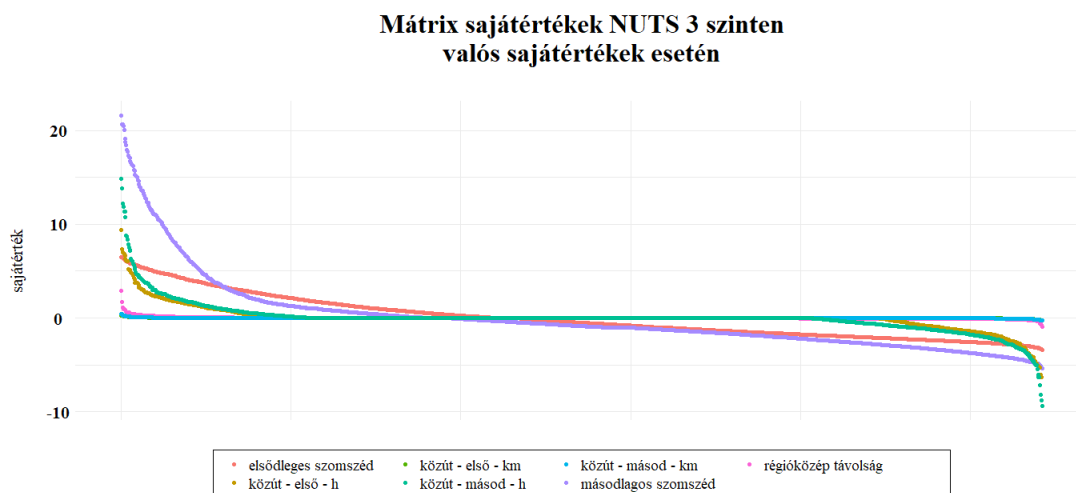
10. ábra. Mátrix sajátértékek NUTS 2 szinten. (Forrás: Saját szerkesztés)

A NUTS 2-es adatokon látható a 10. ábrán, hogy a sajátérték egyezése csak a *Közúti távolság az elsőfokú szomszédok között kilométerben*, a *Közúti távolság a másodfokú szomszédok között kilométerben*, illetve a *Régióközpont távolság* szerint definiált mátrixok között merülhet fel. A 2. táblázat megfelelő értékei segítségével azonban elvethető az a hipotézis, hogy ezen három definíció által felírt mátrixok egymáshoz páronként hasonlóak lennének.

	Determináns	Rang	Sajátérték	Determináns*	Nemnulla elemek aránya	Zérus oszlopok száma
Elsőfokú szomszédság	0	1327	valós	0	0,4%	23
Másodfokú szomszédság	0	1330	valós	0	1,2%	23
Régióközpont távolság	0	1355	valós	0	99,9%	0
Közúti távolság az elsőfokú szomszédok között km-ben	0	1330	valós	0	0,4%	23
Közúti távolság az elsőfokú szomszédok között h-ban	0	1332	valós	0	0,4%	23
Közúti távolság a másodfokú szomszédok között km-ben	0	1332	valós	0	1,2%	23
Közúti távolság a másodfokú szomszédok között h-ban	0	1330	valós	0	1,2%	23
Legrövidebb távolság	3,5 e-267	1355	valós	3,5 e-267	25,4%	0
4-legközelebbi szomszéd	0	1340	komplex	0	0,3%	0

3. táblázat. NUTS 3 szinten definiált területi súlymátrixok tulajdonságai.

(Forrás: Saját szerkesztés) *: zérus oszlopok és sorok elhagyása után



11. ábra. Mátrix sajátértékek NUTS 3 szinten. (Forrás: Saját szerkesztés)

A NUTS 3-es adatokon látható a 11. ábrán, hogy a sajátérték egyezése csak a *Közúti távolság az elsőfokú szomszédok között kilométerben*, a *Közúti távolság a másodfokú szomszédok között kilométerben*, illetve a *Régióközpont távolság* szerint definiált mátrixok között merülhet fel. A 3. táblázat megfelelő értékei segítségével azonban elvethető az a hipotézis, hogy ezen három definíció által felírt mátrixok egymáshoz páronként hasonlóak lennének.

Fontos megemlíteni, hogy a három NUTS szinten nézett mátrixok karakterisztikái a szintek között nem tértek el jelentősen. Ez arra utal, hogy a mátrixok megegyező definiálása több hasonlóságot eredményez, mint amennyi különbséget indukál a területi eltérés a szintek között.

A legszembevetőbb egyezések, hogy minden esetben a 4-legközelebbi szomszéd mátrixnak vannak komplex sajátértékei, míg a másik 8 mátrixnak a sajátértékei valósak. Mindhárom szinten ugyanaz a 3 mátrix eredményezte a 0-hoz legközelebbi és ezáltal az egymástól legkisebb mértékben eltérő sajátértékeket. A legrövidebb távolságot leíró mátrix mindhárom szinten lineárisan független mátrixot eredményezett.

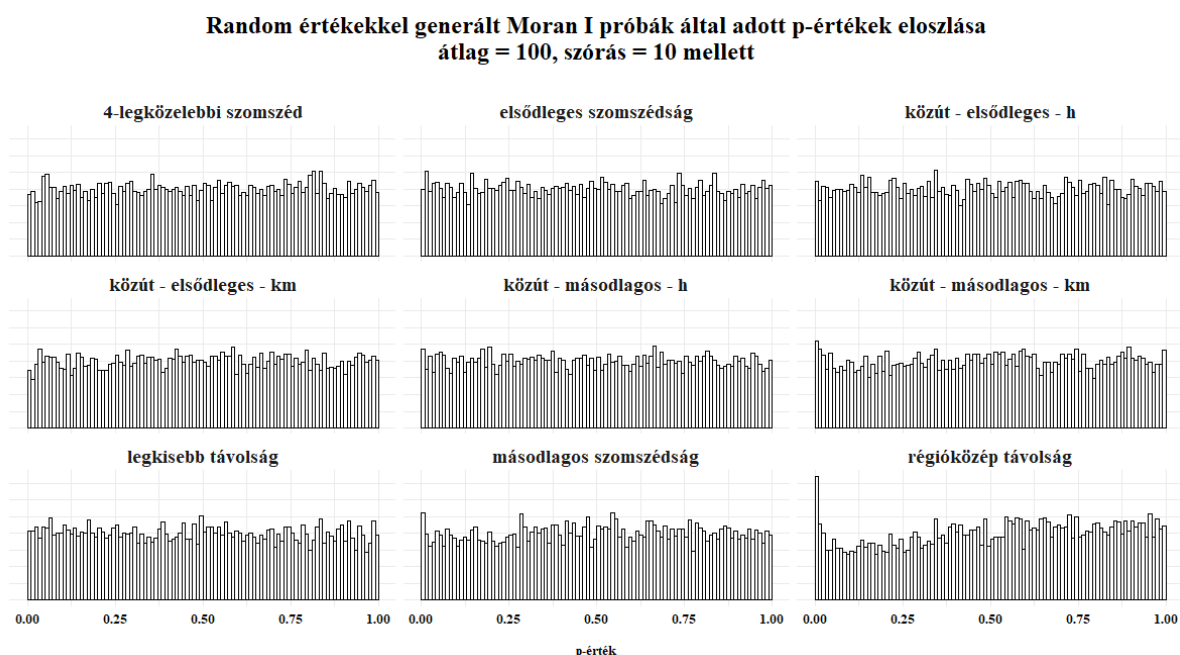
Összességében ezek alapján elmondható, hogy mindhárom szinten mind a 9 mátrix különböző kapcsolati hálót ír fel, ezáltal van alapja a további összehasonlításoknak. Fontos megjegyezni, hogy a definíciószerű eltérések nagyobb eltéréseket eredményeztek, mint a NUTS szintek eltérése.

5.3. Autokorrelálatlanság kimutatása

Az alábbi alfejezet bemutatja, hogy az egyes területi súlymátrixok segítségével végzett Moran-féle I-próbáknak mekkora az elsőfajú hibája. Ennek tesztelésére NUTS 1 és 2 szinten 10 000 tesztet, valamint NUTS 3 szinten számítási kapacitások miatt 2 500 tesztet végeztem el, melynek során független adatokra illesztettem Moran I-próbát a 9-9-9 mátrix segítségével.

A független mintát két paraméterbeállítással futtattam le. Első esetben 100-as várható értékű és 10-es szórású normális eloszlású változót generáltam az adott NUTS szint eleméhez. Ezt követően az illesztést elvégeztem 2-es szórás beállításával is. A tesztek p-értékeit hisztogramon ábrázoltam, valamint megvizsgáltam, adott szignifikanciaszint mellett hány %-ban utasította volna el az egyes mátrix segítségével specifikált Moran-féle I-próba a függetlenség hipotézisét.

A 12. ábrán látható a NUTS 1 szinten $N(100, 10)$ alapján generált független adatok segítségével az egyes területi súlymátrixokkal számított Moran-féle I tesztstatisztika p-értékeinek eloszlása 10 000 generálás alapján.



12. ábra. Független mintán generált p-értékek NUTS 1 szinten $N(100, 10)$ mellett. (Forrás: Saját szerkesztés)

Az ábrán látható, hogy NUTS 1 szinten $N(100, 10)$ generálás során a régióközéppont mátrix segítségével végzett Moran I-próba az elvártnál nagyobb mértékben ad alacsony p-értékeket. Ez a tulajdonság NUTS 1 és 2 szinten mindkét paraméterbeállítás mellett megfigyelhető, azonban NUTS 3 szinten már nem.

NUTS 3 szinten $N(100, 2)$ generálás során a másodfokú szomszédság alapján definiált mátrix esetében mutatott az adatok 6%-a 5%-on szignifikáns autokorrelációt a független adatokon, mely 92 esetben negatív, míg 57 esetben pozitív autokorrelációra utal. A megegyező adat 1%-os szignifikancia esetén 1,24%, míg 10% esetén 10,6% (mely megegyezik a poligonok közötti legkisebb távolság alapú értékkel).

NUTS 3 szinten $N(100, 10)$ generálás során a 4-legközelebbi szomszéd mátrix esetében mutatott az adatok 11,24%-a 10%-on szignifikáns autokorrelációt független adatokon, mely ugyanannyi esetben mutatott pozitív, valamint negatív területi autokorrelációt. A megegyező adat 5%-os szignifikancia esetén 5,44%, míg 1% esetén 1,48%. Ugyanezen beállítások mellett a másodfokú szomszédság a generált adatok 6,36% során mutatott 5%-on szignifikáns eredményt.

NUTS 3 szinten 2-es szórás mellett a generált adatok 485 esetben mutattak 5%-on szignifikánsan negatív, míg 600 esetben szignifikánsan pozitív autokorrelációt a 22 500 szimulációból. Összességében megfigyelhető, hogy általában a szignifikáns autokorreláció kimutatása egyenlően oszlik meg a skála pozitív és a negatív végén. Kivételt képez ez alól NUTS 1 és 2 szinten a másodfokú szomszédság mátrix, míg NUTS 3 szinten a régióközép mátrix. Mindegyik esetben a negatív autokorrelációt mutatja ki nagyobb mértékben a teszt.

Összegzőként tehát elmondható, hogy a területi súlymátrixszal végzett statisztikai tesztek használata előtt fontos ellenőrizni, hogy azonos várható érték és szórás mellett független mintán az adott területi viszonyok mellett megfelelő-e az elsőfajú hiba mértéke. NUTS 1 és 2 szinten a régióközéppont mátrix, míg NUTS 3 szinten a 4-legközelebbi szomszéd mátrix alkalmazásától óva intem a kutatókat.

5.4. Régiós adatok szemléltetése

Az alfejezetben 868 gazdasági és társadalmi adathalmazon hasonlítom össze a 9 definiált területi súlymátrixszal számított Moran-féle I-próba eredményeit. Az alfejezet arra a kérdésre keresi a választ, hogy mekkora különbségek tapasztalhatóak valós adatokon a másképpen definiált súlymátrixokkal végzett Moran I-próbák eredményeiben, valamint arra, hogy az adatsokaság karakterisztikái (leíró statisztikai mutatói) alapján lehet-e következtetni ezen eredmények eltérésének fokára.

5.4.1. Gazdasági és társadalmi adatok

Valós adatokon való szemléltetés céljából letöltöttem 78 gazdasági vagy társadalmi változó NUTS szinteken aggregált értékeit az EUROSTAT adatbázisából. A változók pontos azonosításához használható kódok a Függelékben találhatóak. Minden változó több csoport szerinti bontásban (pl.: nemi vagy korosztályi, összesen 125 változó), több időpontban és több NUTS szinten aggregálva volt elérhető. Az adattisztítás során megváltam azon változóktól, melyek esetében a hiányzó adatok aránya a 25%-t elérte, így összesen 868 vizsgálható adatsort kaptam, melyből 141 db NUTS 1, 652 db NUTS 2 és 75 db NUTS 3 szinten került aggregálásra. A változók az alábbi kategóriákba sorolhatóak.

- Mezőgazdasági változók: talajerózió, mezőgazdasági üzemek szerkezete, földterület, állatpopuláció, trágyatárolók száma
- Népeségre vonatkozó adatok: lakosságszám, lakossűrűség, termékenységi indikátorok, halálozás (összes, rákos, ischaemiás szívbetegség, baleset, közlekedési baleset), orvosok, fogászok, kórházi ágyak száma, élveszületés, teljes termékenységi arányszám, várható élettartam,
- Alapvető gazdasági adatok: GDP, árindex, foglalkoztatási ráta (15-64, 20-64, 55-64), munkanélküliségi ráta, magánháztartások jövedelme, regionális bruttó hozzáadott érték reálnövekedési üteme,
- Humán tőkére vonatkozó változók: felsőfokú iskolai végzettséggel rendelkezők, korai iskolaelhagyók száma, humán tőke a tudományban és a technológiában, high-tech szektor foglalkoztatása, (high-tech) szabadalmi bejelentések az Európai Szabadalmi Hivatalhoz, intramurális K+F ráfordítások, kutatók száma, internetelérés,

- Transzporthoz kapcsolódó változók (tengeri, légi, vasúti, közúti),
- Turisztikai változók: turisztikai szálláshelyeken eltöltött éjszakák száma, létesítmények és férőhelyek száma.

5.4.2. Eredmények összehasonlítása

NUTS 1 szinten a 141 változón elvégzett 1 269 db Moran-féle I-próba 10 esetben mutatott 5%-on szignifikáns pozitív, míg 40 esetben szignifikánsan negatív autokorrelációt. A 10 pozitív eredmény mindegyike az általam definiált Közúti távolság a másodfokú szomszédok között kilométerben, valamint a Közúti távolság a másodfokú szomszédok között órában területi mátrixhoz köthető. 4 változó esetében mindkét általam definiált mátrixszal számított teszt szignifikáns eredményt adott, míg 2 esetben csak a kilométerben mért távolság által mátrix adott szignifikánsan pozitív eredményt. A 6 változó ugyanazon népességi struktúrát leíró minta 2014-2019 közötti realizációja (női medián életkor).

A NUTS 1 szinten szignifikánsan negatív területi autokorreláltnak ítélt 36 változónál 20 esetben csak 1 mátrix mutatott szignifikáns eredményt, mely 5 esetben volt a Másodfokú szomszédság mátrix (Bruttó hozzáadott érték 2016 - 2020), míg 15 esetben a 4-legközelebbi szomszéd mátrix (Átlagos éves népesség, Felsőoktatásban való részvétel, Korai iskolaelhagyók). 16 változó esetében 3 vagy 4 mátrix is szignifikáns kapcsolatot mutatott (Korai iskolaelhagyáshoz kapcsolódóan), melyek az Elsőfokú szomszédság mátrix, a Közúti távolság az elsőfokú szomszédok között kilométerben, a Közúti távolság az elsőfokú szomszédok között órában mátrix, valamint a 4-legközelebbi szomszéd mátrix voltak.

A NUTS 2 szinten elvégzett 5 868 próba 2 782 esetben mutatott 5%-on szignifikáns pozitív területi autokorrelációt, míg 0 esetben negatív. A pozitív eredmények esetében 164 változónál mutatott mind a 9 mátrix szignifikáns autokorrelációt, míg 31 esetben csak 1 (ebből legnagyobb arányban a 4-legközelebbi szomszéd és a Régióközpont távolság mátrixok mutattak a többi mátrixszal nem egyező eredményt).

A NUTS 3 szinten elvégzett 675 próba 631 esetben mutatott 5%-on szignifikáns pozitív területi autokorrelációt, míg egy esetben sem mutatott negatív. A pozitív eredmények 71 változóhoz köthetők, ebből 70 esetben mind a 9 mátrix szignifikáns eredményt mutatott, egy esetben 1 (a Régióközpont távolság mátrix).

A mátrixokról elmondható, hogy a Közúti távolság másodfokú szomszédok között km-ben és órában, a Régióközpont távolság és a Légrövidebb távolság mátrix egy esetben sem

adott szignifikánsan negatív autokorrelációt, a 4-legközelebbi mátrix az átlagnál (380) 45-tel több esetben mutatott szignifikáns pozitív autokorrelációt, míg az elsőfokú szomszédság mátrix az átlagnál 34-gyel kevesebbszer mutatott szignifikánsan pozitív eredményt.

	Elsőfokú szomszédság	Másodfokú szomszédság	Közúti távolság az elsőfokú szomszédok között km-ben	Közúti távolság az elsőfokú szomszédok között h-ban	Közúti távolság a másodfokú szomszédok között km-ben	Közúti távolság a másodfokú szomszédok között h-ban	Régió-középpont távolság	Legrövidebb távolság	4-legközelebbi szomszéd
Elsőfokú szomszédság	1	0,85	0,94	0,94	0,86	0,86	0,75	0,91	0,79
Másodfokú szomszédság	0,85	1	0,84	0,84	0,91	0,91	0,81	0,85	0,80
Közúti távolság az elsőfokú szomszédok között km-ben	0,94	0,84	1	0,99	0,88	0,88	0,77	0,91	0,81
Közúti távolság az elsőfokú szomszédok között h-ban	0,94	0,84	0,99	1	0,88	0,88	0,76	0,90	0,80
Közúti távolság a másodfokú szomszédok között km-ben	0,86	0,91	0,88	0,88	1	0,98	0,81	0,86	0,83
Közúti távolság a másodfokú szomszédok között h-ban	0,86	0,91	0,88	0,88	0,98	1	0,81	0,86	0,83
Régióközpont távolság	0,75	0,81	0,77	0,76	0,81	0,81	1	0,81	0,88
Legrövidebb távolság	0,91	0,85	0,91	0,90	0,86	0,86	0,81	1	0,85
4-legközelebbi szomszéd	0,79	0,80	0,81	0,80	0,83	0,83	0,88	0,85	1

4. táblázat. A gazdasági adatokon adott azonos eredmények aránya az egyes mátrixok között (Forrás: Saját szerkesztés)

A 4. táblázat bemutatja, hogy az összességében vett 9*868 db Moran-féle I-próba esetében mekkora arányban adott megegyező eredményt az egyes mátrixokkal definiált tesztstatisztika (5%-on szignifikánsan negatív, nincs kapcsolat, 5%-on szignifikánsan pozitív). Fontos megjegyezni, hogy bár a korábbi fejezetekben bemutattam, hogy a közúton órában és kilométerben mért távolságok eltérnek, a tesztstatisztika eredmények az elsőfokú szomszédok között 99%-ban, míg a másodfokú szomszédok esetében 98%-ban ugyanazt mutatták. Szintén megfigyelhető, hogy a Régióközpont távolság alapú és az Elsőfokú szomszédság mátrixok által végzett I-próbák adták a leginkább eltérő eredményeket.

Összegzésképp elmondható, hogy a vizsgált 868 adathalmazon nem volt olyan eset, amikor két mátrix ellentétes eredményt adott volna olyan szinten, hogy az egyik szignifikánsan negatív, míg a másik szignifikánsan pozitív autokorrelációt mutatott volna. Az előző alfejezetben kiemelt Régióközpont távolság mátrix ebben a megvilágításban is több esetben

kiemelkedett, NUTS 2 és 3 szinten is megfigyelhető volt, hogy több esetben ad úgy szignifikáns eredményt, hogy azt a másképpen definiált mátrixok nem erősítik meg. Ez az eredmény tovább erősíti azt a korábban megfogalmazott tanácsot, hogy egy távolságmátrix használata előtt meg kell bizonyosodni arról, milyen arányban mutat az adott területi viszonyok mellett független adaton szignifikáns eredményt az adott mátrix.

Szintén megfigyelhető volt annak a többszöri előfordulása, hogy mind a 9 mátrix szignifikáns eredményt adjon, ezáltal megfogalmazható az a tanács is, hogy területi autokorreláció vizsgálatakor ne csak 1 területi súlymátrix segítségével vizsgáljuk meg az adatokat, hanem támaszkodjunk legalább 3, de inkább 4 különbözőképpen definiált súlymátrixra.

Fontos kiemelni, hogy az előző fejezetben bemutatott sajátértékek nem azon mátrixok esetében estek közel egymáshoz, melyek aztán hasonló eredményeket adtak a valós adatokon, sokkal inkább azon értékek voltak egymáshoz hasonlóak, melyek ugyanazon szinten (elsőfokú kapcsolatok, másodfokú kapcsolatok) lévő kapcsolatokat írták le. A megerősítő tesztek esetében ezért ily módon különböző mátrixokkal érdemes újabb teszteket futtatni.

5.4.3. Leíró statisztikák és eredmények kapcsolata

Az alábbi részben bemutatom, hogy az alapsokaság leíró statisztikai mutatói (csúcsosság, ferdeség, relatív szórás, relatív terjedelem) és a Moran-féle I-próba eredményeinek egyezősége között megfigyelhető-e kapcsolat. Ennek vizsgálatához a tesztek eredményeit 3 csoportba soroltam:

- 2-es csoport: mind a 9 mátrixszal végzett Moran-féle I-próba alapján 5%-on szignifikáns pozitív területi autokorreláció figyelhető meg
- 1-es csoport: a 9 mátrix közül legalább 1 db 5%-on szignifikáns területi autokorrelációt mutat (pozitív vagy negatív) és legalább 1 mátrix alapján a kapcsolat nem szignifikáns (azaz a mátrixok eltérő eredményt adnak)
- 0-s csoport: a 9 mátrixszal végzett Moran-féle I-próba alapján nincs 5%-on szignifikáns területi autokorreláció (azaz a 9 mátrix eredménye egyezik a területi kapcsolat hiányában)

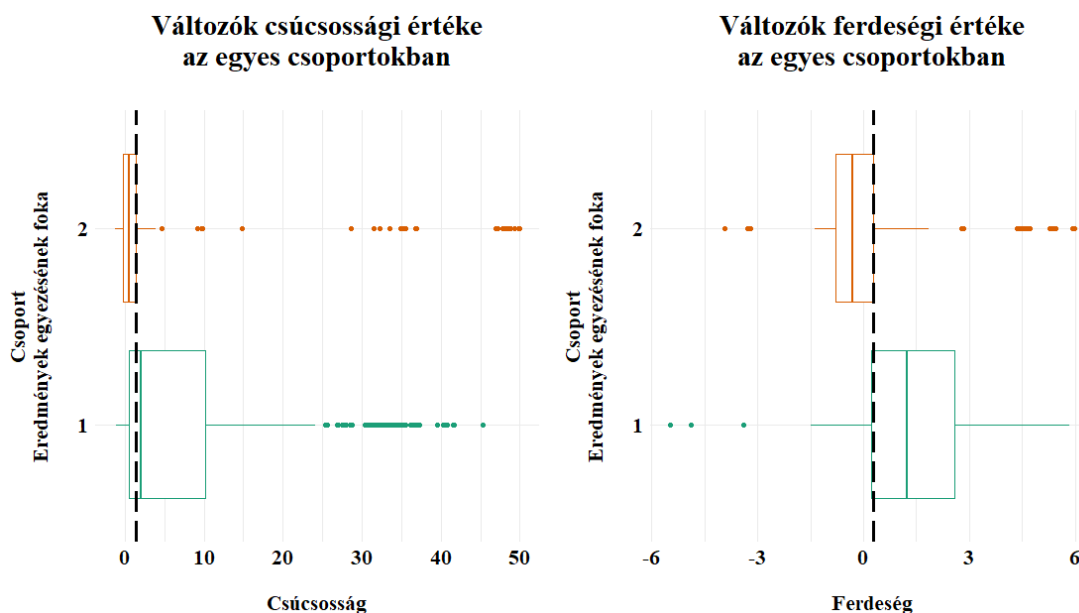
A csoportazonosítók faktorok, nem jeleznek sorrendiséget vagy arányt, a könnyebb leírás és az ábrázolhatóság miatt kerültek így kialakításra. Az adatok tisztítása során megváltam 26 változótól, melyek relatív szórása meghaladta a 10-et, csúcsossága meghaladta az 50-et vagy relatív terjedelme meghaladta a 200-at. Az egyes csoportokban így 230-310-302 megfigyelés maradt.

A vizsgálatot varianciaanalízissel végeztem el, így vizsgálva a csoportok közötti átlagos értékek eltérését. A vizsgált változók között abszolút mutatók nincsenek, mivel a változók mértékegységei nem egyeztek meg, így csak relatív mutatókat (relatív szórás és terjedelem), valamint alakmutatókat alkalmaztam (csúcsosság és ferdeség).

A varianciaanalízis alapján mind a 4 változó csoportokon belüli átlaga 5%-os szignifikanciaszinten eltér, azonban a csoportok megegyező varianciájának nullhipotézisét a Levene teszt alapján el kellett vetnem. A későbbiekben ezért egy másik csoportosítással is elvégeztem a varianciaanalízist, mely esetben a nullhipotézis elvetésének szükségességére nem volt bizonyíték.

Az új csoportosítás az 1-es, illetve 0-s jelű csoport összevonásával alakult ki. Ennek következtében a csoportosítás azt mutatja, hogy a 9 mátrix mindegyike kimutatta-e 5%-os szignifikanciaszinten a pozitív területi autokorrelációt vagy sem. Ebben az esetben a 2-es csoportban 230, míg az egyes csoportban 612 megfigyelést találunk. A Levene teszt

az alakmutatók esetében nem vetette el a megegyező variancia nullhipotézisét, így ezen változókra alkalmaztam a varianciaanalízist. Mindkét esetben 1%-os szignifikanciaszinten vetettem el a csoportok közötti megegyező várható érték nullhipotézisét.



13. ábra. Változók alakmutatói a különböző mátrixok által adott eredmények egyezése szerint kialakított csoportokban. (Forrás: Saját szerkesztés)

A 13. ábrán látható, hogy a 2-es számmal azonosított, azaz a minden súlymátrix által pozitív területi autokorreláló változó csoport értékei alacsonyabb értékeket mutatnak. Abban az esetben mutatott mind a 9 mátrix megegyezően pozitív eredményt, ha a változó csúcsossága és ferdesége alacsony volt. A 2-es csoporton belül a csúcsossági mutatók 75%-os percentilise (1,45) alacsonyabb, mint a másik csoport medián értéke (az ábrán szaggatott vonallal jelölve). A ferdeségi mutatók esetében még szembetűnőbb az eltérés, a 2-es csoport 75%-os percentilise (0,28) csupán 6 századdal haladja meg az 1-es csoport 25%-os percentilisét (az ábrán szaggatott vonallal jelölve).

A korábban megfogalmazott ajánlás a több területi súlymátrix alkalmazására tehát akkor igazán fontos, ha az alapsokaság pozitív ferdeségű, azaz jobbra elnyúló, valamint csúcsossága 1 feletti.

5.5. Szimulációs eredmények

Az alábbi fejezetben szimulációk során bemutatom, hogy az egyes területi szinteken a súlymátrixok mennyire helyettesíthetők egymással, ezáltal adok egy áttekintést arra vonatkozóan, hogy a térbeli függés kimutatása mennyire érzékeny a súlymátrix megválasztására. Kétféle eredményt közlök az alfejezetben. Egyrészt bemutatom, hogy az adott területi súlymátrix által generált adatok határértékben milyen mintázatot mutatnak a NUTS régiókon, valamint bemutatom, hogy az egyes súlymátrixok által generált területi autokorrelációkat a másik módon definiált súlymátrixok segítségével végzett Moran-féle I-próba milyen arányban mutatja ki.

5.5.1. Határmintázatok

Az alfejezet célja, hogy bemutassa, az egyes távolságdefiníciók által adott területi mátrixok milyen különböző eredményeket adnak, s hogy az eredmények mennyire függnak az adott terület földrajzi adottságaitól. Ezáltal adok szemléltetést arra, hogy miért szükséges az egyes területi mintázatok figyelembe venni a területi súlymátrix megválasztása előtt. A szimuláció során a területileg autokorrelált adatokat egy több időszakos „tanulási folyamat” segítségével hozom létre. Ennek során minden terület értékét frissítem az alábbi egyenlet által:

$$x_{t+1} = x_t + 0,4 * W * x_t + \varepsilon$$

$$x_0 \in \mathbf{N}(100, 10)$$

$$\varepsilon \in \mathbf{N}(10, 2)$$

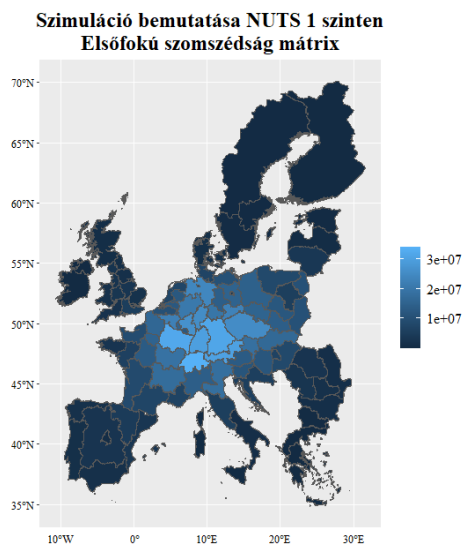
$$y = x_p$$

$$p \in [10; 500]; p \in \mathbf{N},$$

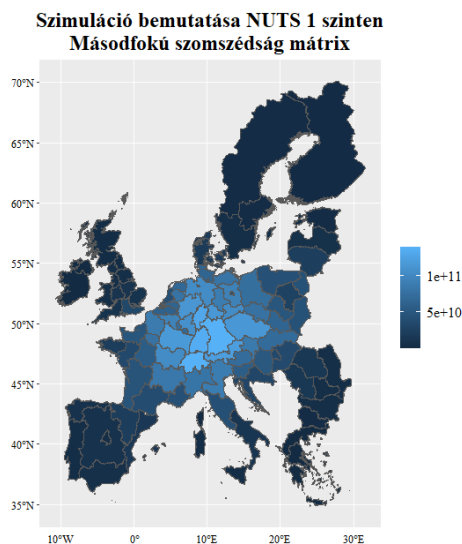
ahol W jelöli a korábban már definiált területi súlymátrixot, p jelöli a „tudásátadási folyamat” hosszát, melynek értéke a W súlymátrix átlagos értékétől függ és y jelöli a változót, melyet a területi fájlokon ábrázolni fogok.

Az *elsőfokú szomszédság mátrix* segítségével generált szimuláció egy NUTS 1 szinten vett realizációját a 14. ábra mutatja. A szimulációt $p = 10$ mellett végeztem el, azonban a mintázatok nagyobb p értékek mellett is megegyezők. Az alábbi részben a változók területi mintázatának ábrázolása a cél, így p értéke azt a szintet jelképezi, ahonnan már a fő

területi karakterisztikák nem változnak, csupán az y változó nagyságrendje. Az ábrákon ezért a változónagyság nem releváns, csupán a szemléltetést segíti. A 14. ábrán jól látható a Svájc, Németország, valamint Franciaország területén központosuló adat. Szintén megfigyelhető, hogy mivel a generálás az elsőfokú szomszédsági mátrix alapján történt, ezért a közrefogott régiók (általában a nagyvárosok, pl.: Madrid) az alacsonyabb értékek közé sorolódnak, csakúgy mint a földrajzi okok miatt elszigetelt területek (Görögország, Svédország, Finnország). A fő karakterisztikák NUTS 2 és NUTS 3 szinten nem változnak.



14. ábra. *Elsőfokú szomszédság mátrixszal generált adatok NUTS 1 szinten.*(Forrás: Saját szerkesztés)

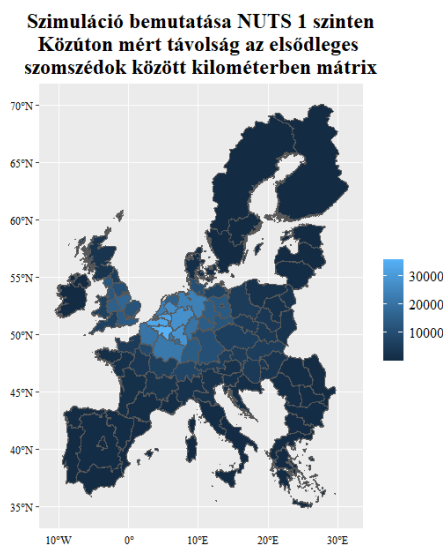


15. ábra. *Másodfokú szomszédság mátrixszal generált adatok NUTS 1 szinten.*(Forrás: Saját szerkesztés)

A másodfokú szomszédság mátrixszal generált adatok (15. ábra) hasonló képet mutatnak,

mint a 14. ábra, azonban a közrefogott régiók lemaradása kisebb a környezetükhöz képest, mint amit az elsőfokú szomszédság esetében láttunk. Emellett megfigyelhető, hogy a kontinens közepén összpontosuló „fejlettség” kevésbé centralizált, mint az elsőfokú szomszédság esetében. A fő karakterisztikák NUTS 2 és NUTS 3 szinten nem változnak.

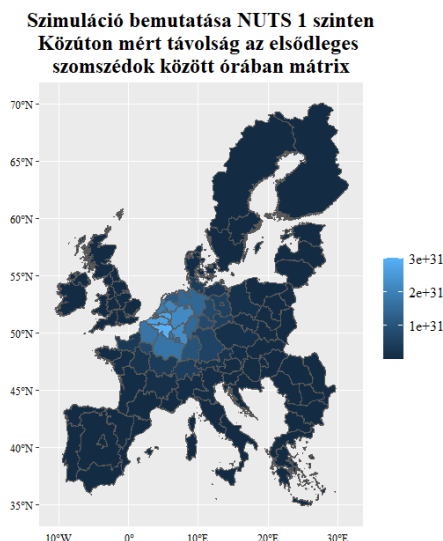
A *Közúti távolság az elsőfokú szomszédok között kilométerben mátrix* által kialakított adatok (16. ábra) sokkal kevésbé mutatnak a kontinens közepén összpontosuló adatokat, mint ahogy azt az elsőfokú szomszédság mátrix esetében láthattuk annak ellenére, hogy ugyanúgy a közvetlen szomszédságot jelöli, mint az. Ebben az esetben a magas értékek a BeNeLux államok, Nyugat-Németország, Kelet-Franciaország és London körzetében összpontosulnak. Ennek oka, hogy a lakosságszám alapján kialakított régiók miatt ezen területek rendelkeznek a legkisebb területtel, így ezen területeken jóval kisebbek a közúti távolságok, mint a kontinens többi részén. NUTS 2 szinten ez csak tovább erősödik, az adatok nagy része London kerületeiben centralizálódik, míg NUTS 3 szinten a nagyobb adatok London Lambeth városrészébe összpontosulnak.



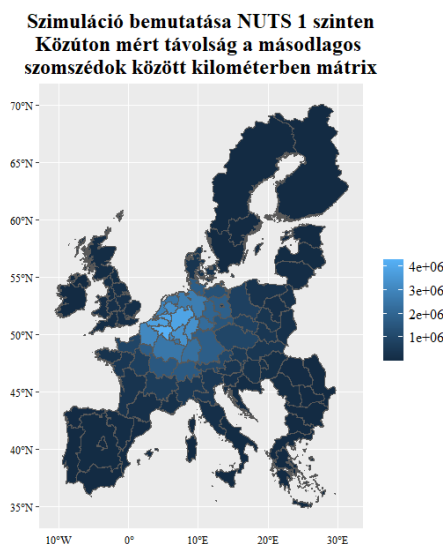
16. ábra. *Közúti távolság az elsőfokú szomszédok között kilométerben mátrixszal generált adatok NUTS 1 szinten.* (Forrás: Saját szerkesztés)

A *Közúti távolság az elsőfokú szomszédok között órában mátrix* által kialakított adatok (17. ábra) hasonló képet mutatnak, mint a kilométerben mért megfelelőik. Ebben az esetben a magas értékek sokkal jobban centralizálódnak a BeNeLux államok területén, azok nem terjednek ki az Egyesült Királyságra és Nyugat-Németországra, valamint Kelet-Franciaországra. Ennek oka lehet, hogy a határátlépések és a Csalagút csökkenti

a közúton elérhető átlagsebességet, ezáltal növeli az Egyesült Királyság és a kontinens területei közötti távolságot a kilométer - óra átváltás. NUTS 2 szinten az adatok már London kerületeiben és annak szomszédaiban összepontosulnak. NUTS 3 szinten az adatok Mannheim, Heidelberg, Frankenthal és Ludwigshafen am Rhein német városokban koncentrálódnak.



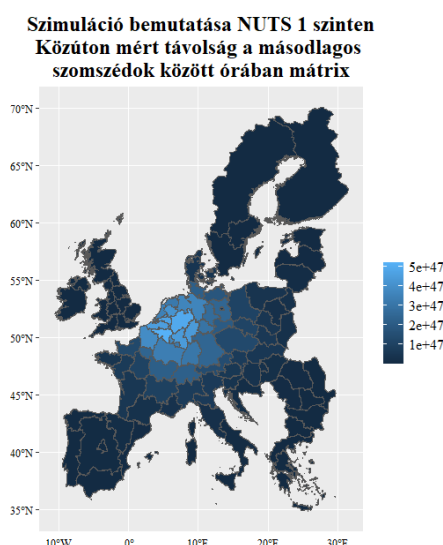
17. ábra. Közúti távolság az elsőfokú szomszédok között órában mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szerkesztés)



18. ábra. Közúti távolság a másodfokú szomszédok között kilométerben mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szerkesztés)

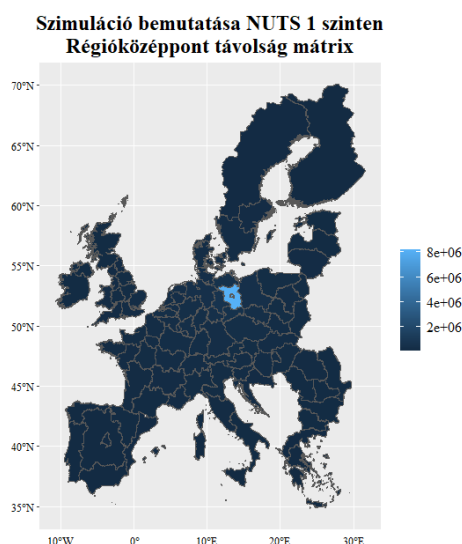
A *Közúti távolság a másodfokú szomszédok között kilométerben mátrix* által kialakított adatok (18. ábra) hasonló képet mutatnak, mint az elsőfokú szomszédság alapú megfelelőik. A különbséget itt is az Egyesült Királyság területén látjuk, melyre intuitív magyarázatot nem találtam. NUTS 2 és 3 szinten itt is London és a környező kerületek emelkednek ki.

A *Közúti távolság a másodfokú szomszédok között órában mátrix* által kialakított adatok (19. ábra) megegyező képet mutatnak a kilométerben kifejezett megfelelőikkel. Ebben az esetben a kilométer - óra átváltás a közúti adatokon ebben a határesetben nem meghatározó. NUTS 2 és 3 szinten itt is London és a környező kerületek emelkednek ki.



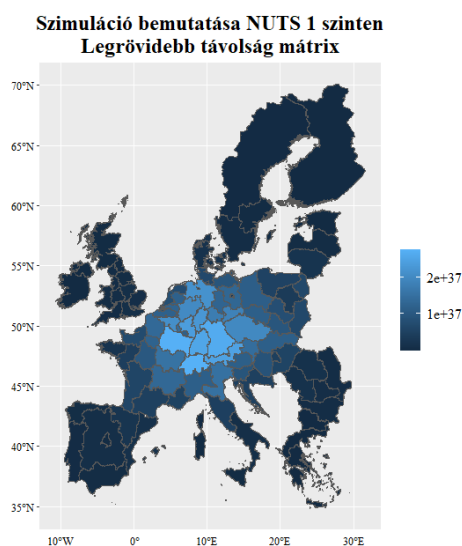
19. ábra. *Közúti távolság a másodfokú szomszédok között órában mátrixszal generált adatok NUTS 1 szinten.* (Forrás: Saját szerkesztés)

A *Régióközéppont távolság mátrix* által kialakított adatok (20. ábra) a teljes centralizációt, s ezáltal Berlin és Brandenburg tartományának mindenek feletti kiemelését eredményezik. Hasonló eredményeket ad a mátrix alkalmazása NUTS 2 szinten is, azonban itt már London és Brüsszel, valamint az őket körülvevő régiók is kiemelkednek a kontinens többi régiója közül. NUTS 3 szinten ezekhez még csatlakozik a teljes Németország, valamint a BeNeLux államok területe is, a népsűrűség miatt kis területi egységekre bontott régiók miatt.



20. ábra. Régióközpont távolság mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szerkesztés)

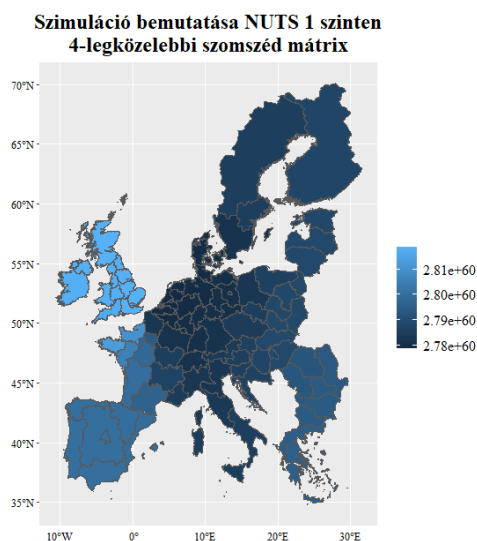
A *Poligonok közötti legrövidebb távolság mátrix* által kialakított adatok (21. ábra) az elsőfokú és másodfokú szomszédság mátrixhoz hasonlóan a kontinens középső területeit emelik ki NUTS 1 szinten, míg NUTS 2 szinten már a görög szárazföldi régiókat, NUTS 3 szinten pedig Londontól északra fekvő nem különösebben jelentős régiókat.



21. ábra. *Poligonok közötti legrövidebb távolság mátrixszal generált adatok NUTS 1 szinten.* (Forrás: Saját szerkesztés)

A *4-legközelebbi szomszéd mátrix* által kialakított adatok (22. ábra) tudják egyedül bemutatni a skandináv országok fejlettségét. Ennek eredményei észak-déli irányban adnak

azonos fejlettségű sávokat a kontinensen, valamint nemcsak földrajzilag, hanem az adatok eltérésében is kiemelik az Egyesült Királyságot a kontinensből. Ezen mátrix esetében a karakterisztikák nem térnek el a különböző NUTS szinteken.



22. ábra. 4-legközelebbi szomszéd mátrixszal generált adatok NUTS 1 szinten. (Forrás: Saját szerkesztés)

A különböző futtatások ebben az esetben eltérő eredményeket adtak abban a tekintetben, hogy az adatok nagyságrendjében melyik „sáv” emelkedett ki, melyre a többi mátrix esetében nem volt példa.

A szemléltető ábrák is jól mutatják, hogy a területi súlyok meghatározása mennyire eltérő eredményeket adhat Európa kontinensén, valamint a különböző NUTS szinteken is. Emiatt nem elegendő pusztán a definíció alapján megválasztani a teszteléshez és modellezéshez használt területi súlymátrixot, azt igazítani kell nem csupán az alapsokaság jellemzőihez, hanem a területi sajátosságokhoz is, amelyen az adatok aggregálva lettek. A modellezőknek javasolt ezen karakterisztikákat megfigyelni a távolságdefiníciók alkalmazása előtt, majd ennek tudatában kiválasztani azon mátrixokat melyekkel az autokorreláció tesztelését elvégzik. A továbbiakban a Moran-féle I statisztika segítségével fogom igazolni a területi súlymátrix megválasztási módszerének fontosságát.

5.5.2. Eredmények összevetése

Az alfejezet célja, hogy bemutassa, az egyes súlymátrixok által generált területi autokorrelációkat a másik módon definiált súlymátrixok segítségével végzett Moran-féle I-próbák milyen arányban mutatják ki. Ezáltal adok szemléltetést arra, hogy miért szükséges több távolságdefiníció által adott mátrix segítségével tesztelni a területi autokorrelációt. A szimuláció során a területileg autokorrelált adatokat egy több időszakos „tanulási folyamat” segítségével hozom létre. Ennek során minden terület értékét frissítem az alábbi egyenlet által:

$$x_{t+1} = x_t + c * W * x_t + \varepsilon$$

$$x_0 \in \mathbf{N}(100, 10)$$

$$\varepsilon \in \mathbf{N}(10, 2)$$

$$y = x_p$$

$$c \in [0, 01; 1]$$

$$p \in [1; 5]; p \in \mathbf{N},$$

ahol W jelöli a korábban már definiált területi súlymátrixot, p jelöli a „tudásátadási folyamat” hosszát, c jelöli az időszakonkénti átadás mértékét, melynek értéke 0,01 és 1 között 0,05-ös lépésközökkel változik és y jelöli a változót, melyen a területi autokorreláció jelenlétét a 9 definiált mátrixszal tesztelni fogom. Minden p - c - W paraméterhármassal 100 szimulációt készítettem NUTS 1 és 2 szinten, ezáltal szintenként 90 000 szimulációs eredményen tudom összevetni a 9 mátrix által adott összesen 810 000 Moran I tesztstatisztikát. NUTS 3 szinten a Moran tesztek számítási időigényének növekedése miatt p értékét csak $p = 4$ -ig, c értékét $c = 0,5$ -ig teszteltem, míg minden p - c - W paraméterhármassal 50 szimulációt készítettem. NUTS 3 szinten így összesen 18 000 szimulációs eredményen készített 162 000 megfigyelést hasonlítottam össze. A szimulációs eredmények elérhetőek az alábbi GitHub linken.

NUTS 1 szinten a 810 000 teszt 37,9%-a (307 136) mutatott 5%-on szignifikáns autokorrelációt, azonban ebből 1 092 esetben szignifikánsan negatív autokorrelációt mutattak ki a tesztek, mely a teljes adatsokaság 0,135%-a. NUTS 2 szinten a 810 000 teszt esetében már 71,4% mutatott szignifikáns autokorrelációt, melyből 785 volt negatív, mely a teljes minta 0,09%-a. NUTS 3 szinten a 162 000 teszt esetében 75,9% mutatott szignifikáns

autokorrelációt, melyből 233 volt negatív, mely a teljes minta 0,14%-a. A negatív esetek problémákat vehetnének fel, azonban ezek közül csak elhanyagolható esetben adott a teszt 1%-os szignifikanciaszinten is szignifikáns eredményt, valamint ezen szimulációk többsége $c = 0,01$, $p = 1$ és eltérő generáló és tesztelő mátrix mellett alakultak ki, így ezen problémát elhanyagolhatónak tekintettem. A továbbiakban a szignifikánsan pozitív eredmények bemutatására szorítkozom.

Azonos mátrixszal tesztelt és szimulált adatok esetében a szimulált adatok 45,52%-a mutatott szignifikáns eredményt NUTS 1, 71,0%-a NUTS 2 és 75,1%-a NUTS 3 szinten, míg ugyanez az arány eltérő tesztelési mátrixok esetében NUTS 1 szinten 36,8%, NUTS 2 szinten 73,7%, míg NUTS 3 szinten 80,9% volt. Nagyobb eltérések tapasztalhatóak abban a tekintetben, hogy mely mátrix segítségével generált adatok adták a pozitív eredményeket. NUTS 1 szinten a *Másodfokú szomszédság mátrix* esetében a teszteknek csupán a 17%-a, míg a *Poligonok közötti legrövidebb távolság mátrix* esetében a tesztek 81,5%-a mutatott szignifikánsan pozitív autokorrelációt. NUTS 2 szinten ezen értékek a *Régióközpont mátrix* által adott 44,8% és a *4-legközelebbi szomszéd mátrix* által adott 87,1% között szóródtak. NUTS 3 szinten ezen értékek a *Régióközpont távolság mátrix* által adott 18,8% és a *Közúti távolság az elsőfokú szomszédok között kilométerben mátrix* által adott 87,0% között változtak, azonban a Régióközpont mátrix által adott érték jelentősen kilóg a többi eredmény közül.

Ezen eredmények mutatják, hogy a teszteléshez használt mátrix megválasztása során kiemelt figyelmet kell fordítani arra, hogy a területek közötti tudásátadás milyen csatornákon keresztül történhet, s ennek megfelelően kell definiálni a teszteléshez használt mátrixot. Szintén figyelembe kell venni a terület földrajzi karakterisztikáit is, hiszen ahogy láthattuk voltak olyan tudásátadási folyamatok, melyek esetében az autokorreláció kimutatása nehézkes, valamint eltérő eredmények születtek az egyes NUTS szinteken.

Még alacsonyabb arányok tapasztalhatóak abban az esetben, ha a tesztelő mátrixot is megkülönböztetjük. Ebben az esetben már 15% alatti kimutatási arányokat is láthatunk az alábbi párok között (első mátrix a generáló, második a tesztelő mátrix)

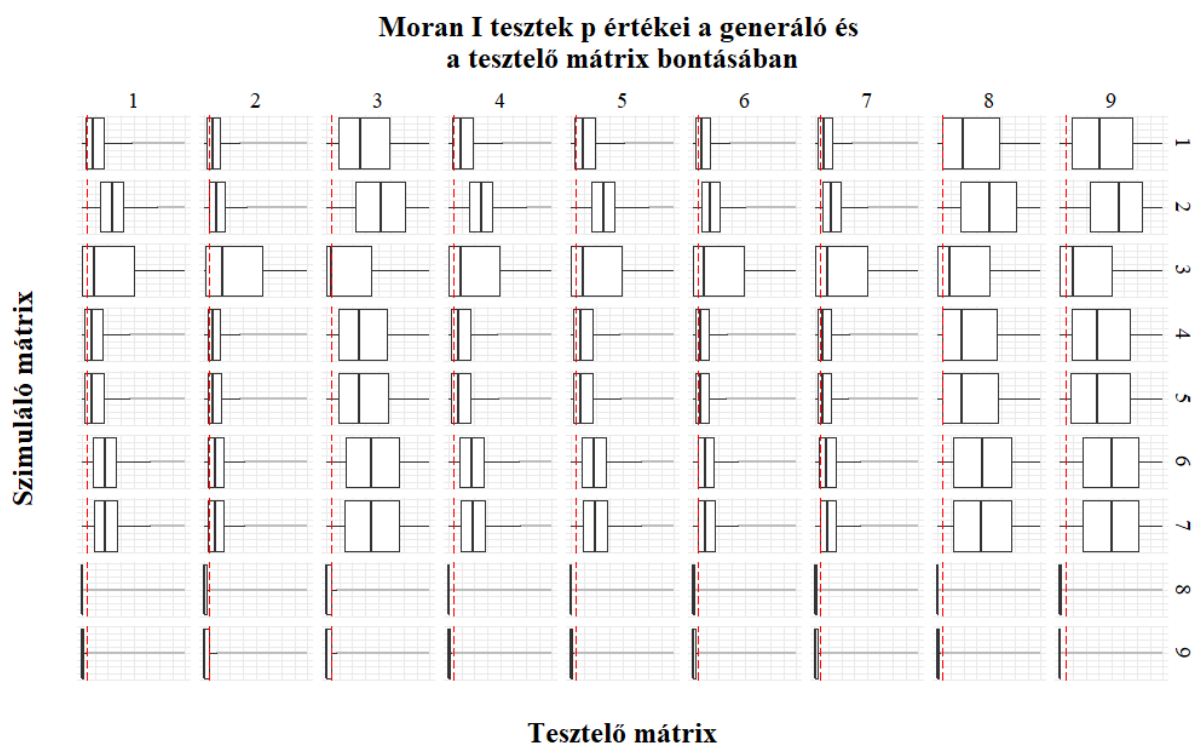
- Közúti távolság a másodfokú szomszédok között kilométerben - 4-legközelebbi szomszéd (8,82%) - NUTS 1
- Közúti távolság a másodfokú szomszédok között kilométerben - Régióközpont távolság (9,26%) - NUTS 1

- Közúti távolság a másodfokú szomszédok között órában - 4-legközelebbi szomszéd (8,85%) - NUTS 1
- Közúti távolság a másodfokú szomszédok között órában - Régióközpont távolság (9,43%) - NUTS 1
- Másodfokú szomszédság - Elsőfokú szomszédság (6,27%) - NUTS 1
- Másodfokú szomszédság - Közúti távolság az elsőfokú szomszédok között kilométerben (5,3%) - NUTS 1
- Másodfokú szomszédság - Közúti távolság az elsőfokú szomszédok között órában (5,38%) - NUTS 1
- Másodfokú szomszédság - 4-legközelebbi szomszéd (6,14%) - NUTS 1
- Másodfokú szomszédság - Poligonok közötti legrövidebb távolság (8,56%) - NUTS 1
- Másodfokú szomszédság - Régióközpont távolság (5,61%) - NUTS 1
- Másodfokú szomszédság - Közúti távolság az elsőfokú szomszédok között kilométerben (10,24%) - NUTS 2
- Másodfokú szomszédság - Közúti távolság az elsőfokú szomszédok között órában (10,51%) - NUTS 2

A tesztelésre használt mátrixok közül NUTS 1 szinten a legmagasabb kimutatási arányt (44,35%) a Közúti távolság a másodfokú szomszédok között kilométerben, valamint órában (44,29%) mátrixok adták. NUTS 2 szinten a legmagasabb kimutatási aránnyal a Poligonok közötti legrövidebb távolság mátrix (84,9%), valamint a 4-legközelebbi szomszéd mátrix (83,5%) rendelkezik. NUTS 3 szinten minden mátrix kimutatási aránya 74,8% és 77,2% között volt. Ezek alapján a tesztek abban az esetben adják a legpontosabb eredményt a tudásátadási folyamat figyelembevétele nélkül, ha ezen mátrixok segítségével végezzük el a Moran-féle I-próbákat. Mivel azonban az egyes szinteken eltértek, hogy melyek ezek a mátrixok, illetve hogy van-e jelentős különbség közöttük, ezért nem lehet általános következtetést levonni arra vonatkozóan, hogy a földrajzi struktúrától függetlenül mely

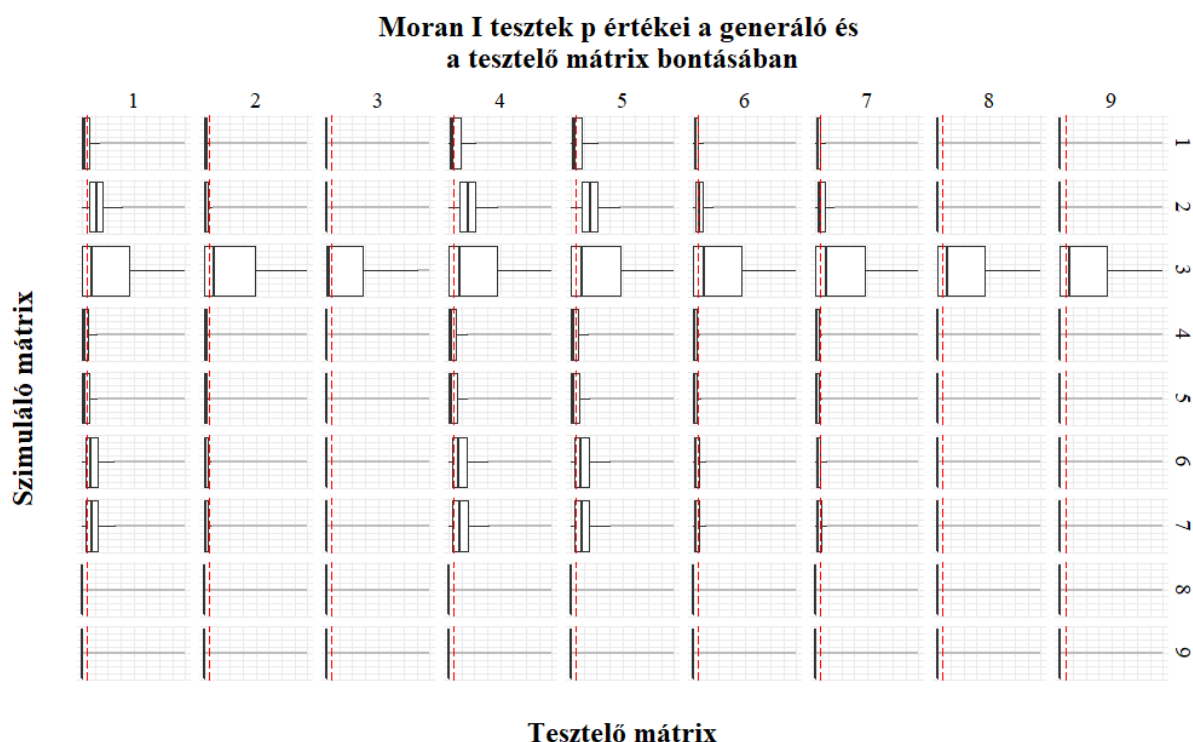
mátrixot érdemes használni, annak megválasztását mindig a földrajzi adottságokhoz és a közgazdasági elmélethez kell igazítani.

Az előzetes várakozásaimmal ellentétes módon a pozitív autokorreláció kimutatási aránya nem növekedett, sőt néhány mátrix esetében jelentősen csökkent a tanulási folyamat hosszának növelése által, mely NUTS 1 és NUTS 2 szinten is megfigyelhető volt. NUTS 1 szinten például a másodfokú szomszédság mátrix esetében az 5 időszak alatt az arány 19%-ról 4,9%-ra csökkent, míg a Régióközéppont mátrix esetében 11,5%-ról 67%-ra emelkedett. A c paraméter esetében a kimutatási arány mindhárom szinten a várakozásoknak megfelelően emelkedő tendenciát mutatott, NUTS 1 szinten például $c = 0,01$ esetében 4,3%, míg $c = 0,21$ esetében eléri a 40%-ot, ahonnan nagyobb növekedés már nem tapasztalható.



23. ábra. Szimulációs eredmények p értékei a generáló és a tesztelő mátrix szerinti bontásban NUTS 1 szinten. (Forrás: Saját szerkesztés)

A 23. ábrán látható, hogy az egyes mátrixok által generált és a másik mátrix alapján tesztelt Moran I-próba p értékei NUTS 1 szinten milyen módon térnek el a generáló mátrixok által. Megegyező módon a NUTS 2 szinten kapott eredményeket a 24. ábra, míg NUTS 3 szinten a Függelékben található 25. ábra tartalmazza. Az egyes sorszámok megegyeznek a 5.1. alfejezetben definiált mátrixok sorszámaival.



24. ábra. Szimulációs eredmények p értékei a generáló és a tesztelő mátrix szerinti bontásban NUTS 2 szinten. (Forrás: Saját szerkesztés)

Ezek az eredmények tovább erősítik azt a megfigyelést, hogy a teszteléshez használt mátrix megválasztása során kiemelt figyelmet kell fordítani arra, hogy a területek közötti tudásátadás milyen csatornákon keresztül történhet. Ennek megfelelően kell definiálni a teszteléshez használt mátrixot, hiszen bár vannak olyan szimuláló mátrixok melyek által generált autokorrelációt szinte bármelyik mátrix kimutatja, több esetben nagyon alacsony kimutatási arányt figyelhetünk meg. NUTS 2 és 3 szinten kevésbé nehézkes az autokorreláció kimutatása, ezen esetekben csupán a 3-as számmal jelölt Régióközpont távolság mátrixot tudom kiemelni. Az eredmény nem meglepő, hiszen ezen mátrix fő karakteristikái térnek el leginkább a többitől, valamint több esetben is kiemelkedett már az eredmények során ez a távolságdefiníció.

5.6. Javasolt területi súlymátrix választási folyamat

Az 5. fejezetben részletesen bemutatott problémák feloldására ezen alfejezetben összefoglalom a javaslataimat.

Elsődlegesen a vizsgált közgazdasági probléma esetében felmerülő területek közötti tudásátadási folyamatok szintjét szükséges meghatározni. A teljesség igénye nélkül erre 3 példát adok. Abban az esetben, ha a területi autokorreláció létét a közigazgatási régiókon átívelő földrajzi területek által nyújtott természetbeni lehetőségek aggregálási pontatlansága okozza, valamely elsőfokú szomszédsági szinten értelmezett területi súlymátrix alkalmazása a javasolt. Ebben az esetben egy olyan mátrix alkalmazása, mely másodfokú szomszédokhoz pozitív értéket rendel vagy minden területi értékhez rendel értéket, félrevezető eredményeket adhat.

Abban az esetben, ha a területek kapcsolatának oka nem az aggregálási pontatlanság, hanem a tapasztalat és a tudásátadás, valamint a gazdasági adat forrása a városokhoz köthető, akkor a városok távolsággal arányos területi súlymátrix alkalmazása a célravezető. Magyar példánál maradva gondolhatunk itt például a külföldi nagyvállalatok magyarországi telephelyeinek környezetében lévő régiók esetében a háztartások jövedelmére. Fontos figyelembe venni ilyen esetekben a legnagyobb távolságot is, melyen a hatást értelmezni tudjuk, hiszen a Régióközéppont távolság mátrix és a Poligonok közötti legrövidebb távolság mátrix eredményein láthattuk, hogy az ilyen típusú határ beállításának hiánya milyen módon okozott problémákat. Ebben az esetben, ha az adott gyár 100 km-es körzetben indít rendszeres, ingyenes buszjáratokat a munkavállalók közlekedési problémáinak megoldására, a súlymátrix minden olyan értékének 0 értéket kell adni, mely ezen határon felül esik.

Abban az esetben, ha a vizsgált változó politikai vonatkozású, érdemes lehet a 4-legközelebbi szomszéd mátrix alkalmazása, hiszen ebben az esetben természetes módon értelmezhető, hogy az egyes régiók vezetői a szövetségeseket az ilyen típusú közelség alapján keresik. Szintén erősíti ezt a hatást, hogy általánosságban az ideológiai csoportosulások ilyen szintén jöttek létre a történelmünk során.

A területi kapcsolódási szintek meghatározása után szükséges a vizsgált változó eloszlásának vizsgálata is, hiszen az 5.4. alfejezetben láthattuk, hogy abban az esetben, ha az alapsokaság pozitív ferdeségű, azaz jobbra elnyúló, valamint csúcsossága 1 feletti, a különböző mátrixszal specifikált Moran-féle I-próbák eltérő eredményeket adhatnak. Az

alakmutatók vizsgálata megadja számunkra, hogy hány eredmény szükséges ahhoz, hogy az autokorreláció léte legyen erősítve.

Az így kiválasztott területi súlymátrixokon ezt követően független mintán vizsgálni kell, hogy az adott földrajzi viszonyok mellett mekkora elsőfajú hibát ad az adott specifikáció. Az 5.3. alfejezetben láthattuk, hogy egy rosszul specifikált távolságmátrix, jelen tanulmány esetében a Régióközpont távolság mátrix, mely ezen a teszten elbukik, hogyan emelkedik ki negatívan a többi vizsgálat során és hogy ad fals eredményeket a területi kapcsolatokról. Szintén vizsgálni kell, hogy az adott földrajzi viszonyok mellett az egyes mátrix specifikációk határértékben milyen mintázatokat rajzolnak ki az adott földrajzi régión. Az így kapott eredmények és a valós változó területi mintának a kapcsolata segítséget nyújthat abban, hogy a modellekben mely specifikáció alkalmazása javasolt.

A területi autokorrelációs modellek lépéseinek általam javasolt sorrendje:

1. Területi kapcsolatok tudásátadási szintjének meghatározása: szomszédsági, távolság alapú, ideológiai.
2. Gazdasági változó eloszlásának vizsgálata, ez alapján a szükséges tesztelési szám meghatározása.
3. Független mintán való vizsgálat a kiválasztott területi súlymátrixokkal, valamint az elsőfajú hiba vizsgálata.
4. A valós földrajzi viszonyok feltérképezése a súlymátrixok által generált területi mintázatok határértékéről, valamint ezen mintázatok összevetése az adat mintázatával.
5. Azon mátrixok kiválasztása, melyek a választott tudásátadási szinten vannak értelmezve, melyek független mintán adott elsőfajú hibája elfogadható, s melyek az adott földrajzi viszonyok mellett határértékben hasonló mintázatot adnak, mint a vizsgált változó, majd ezekkel a Moran-féle I-próbával a területi autokorreláció vizsgálata.
6. Amennyiben ezen vizsgálatok megerősítik a területi autokorreláció létét, a modellezés során ezen területi mátrixok közül a közgazdaságilag leginkább összeegyeztethető alkalmazása.

6. Összegzés

A tanulmány célja, hogy betekintést nyújtson a területi autokorrelációs modellek területi súlymátrix megválasztására vonatkozó érzékenységebe. Területi modellek esetében a régiók kölcsönös függését egy területi súlymátrix segítségével vizsgálják, valamint korrigálják. A földrajzi adatok egyedisége miatt ezen kapcsolatokat egy dimenzióba sűríteni nehézkes és nagy felelősséget ró a modellezőre. Szintén megfigyelt tény a szakirodalomban, hogy a modellek eredményei ezen mátrix megválasztásától jobban függnék, mint a modellválasztástól (Bell és Bockstael 2000). A tanulmány ezen választás megkönnyítése céljából javasol egy választási mechanizmust.

Két közigazgatási terület közötti kapcsolatot lehet értékelni pusztán szomszédsági szinten, bizonyos speciálisan kiemelt geokoordináták közötti távolsági szinten, a legközelebbi területek kiemelése által, illetve akár valamely gazdasági vagy politikai kapcsolatot leíró változó értékei alapján is. Ezekben belül a távolságot is többféleképpen definiálhatjuk (légvonalban, közúton, légi közlekedés által kilométerben vagy a közlekedési változók által akár órában), illetve a speciálisan kiemelt geokoordináták megválasztására is több lehetőség nyílik (mértani középpont, közigazgatási középpont, legnagyobb népességű város, az összes város súlyozott középpontja stb.). Minden ily módon definiált mátrix esetében van lehetőség egy felső határ megválasztására is, melyen felül a kapcsolatot leíró változó értéke automatikusan 0. Ily módon a definiálható területi mátrixok száma végtelen, melyből a modellezőnek kell egyet választania.

A tanulmányban az Európa területeit 3 szinten aggregáló NUTS rendszer adatain definiáltam 9 mátrixot, melyeken a Moran-féle I-próbának ezen mátrixok megválasztására vonatkozó érzékenységvizsgálatát elvégeztem. Az 5.1. alfejezetben a mátrixok pontos definiálása megtalálható. A vizsgálathoz a Google Maps Distance Matrix API segítségével letöltöttem az EUROSTAT által jelentősnek ítélt európai városokat is, így 4 olyan távolságmátrix definiálására is lehetőségem volt, melyre a szakirodalomban még nem volt példa, azonban melyek intuitív jelentése a tudásátadási folyamatok jelenlétéről magasabb. Az 5.2. alfejezetben a definiált mátrixok matematikai hasonlóság definíciója alapján kapott összehasonlítását végeztem el. Ezek alapján egyik mátrix páros sem hasonló egymáshoz matematikai értelemben. A későbbi eredmények alapján azonban valós gazdasági és társadalmi adatokon (5.4. alfejezet) nem azon mátrixok adnak hasonló eredményt a Moran I-próbán, melyek matematikai értelemben hasonlóak, így ennek alkalmazását öko-

nometriai környezetben nem ajánlom.

Az 5.3. alfejezetben megvizsgáltam, hogy független adatokon mekkora az egyes mátrixok elsőfajú hibája. Ennek során egy olyan mátrixot találtam, mely rosszul specifikált volt, mely a kapcsolatok létre nem állított felső korlátot. Fontos megjegyezni, hogy ezen eredmények nem csupán a mátrix definiálásától függnék, hanem az adott terület földrajzi tulajdonságaitól is, hiszen a kialakított régiók közigazgatási és politikai határok mentén húzódnak vagy speciálisan a NUTS szinteknél maradva lakosságszám alapú csoportokba lettek osztva. A különböző területeken ugyanúgy definiált mátrixok eltérő eredményeket adhatnak ezen a teszten. NUTS szinteken haladva is eltérő eredmények születtek, hiába volt a felosztott földrajzi terület megegyező. A területi súlymátrixszal végzett statisztikai tesztek használata előtt tehát fontos ellenőrizni, hogy azonos várható érték és szórás mellett független mintán az adott területi viszonyok mellett megfelelő-e az elsőfajú hiba mértéke. NUTS 1 és 2 szinten a régióközpont mátrix, míg NUTS 3 szinten a 4-legközelebbi szomszéd mátrix alkalmazásától óva intem a kutatókat.

Az 5.4. alfejezetben megvizsgáltam, hogy 868 valós gazdasági és társadalmi adaton az általam definiált 9 mátrixszal végzett Moran-féle I-próba milyen arányban ad azonos eredményt, valamint hogy az alapsokaság leíró statisztikai mutatói alapján lehet-e következtetni ezen eredmények eltérésének fokára. Az alfejezet megerősítette az előző bekezdésben kapott eredményeket, hiszen az elsőfajú hiba teszten elbukott mátrix mutatott több esetben is úgy pozitív területi autokorrelációt, hogy a többi mátrix ezt nem tudta megerősíteni. Szintén megfigyelhető volt, hogy nem ritka eset, hogy mind a 9 mátrix szignifikáns eredményt adjon, emiatt tanácsos területi autokorreláció vizsgálatokor nemcsak 1, hanem 3-4 területi súlymátrix segítségével megvizsgálni az adatokat.

A valós adatokon végzett varianciaanalízis alapján az alapsokaság alakmutatói (ferdeség, csúcosság) alapján lehet következtetéseket levonni arra, hogy a különféleképpen definiált mátrixok azonos eredményt fognak-e adni. Ennek következtében elmondható, hogy abban az esetben van szükség megerősítő tesztekre is az autokorreláltság kimutatása során, ha az alapsokaság pozitív ferdeségű, azaz jobbra elnyúló és csúcossága 1 feletti.

Az 5.5. alfejezetben szimulációk során bemutattam, hogy az adott területi súlymátrix által generált adatok határértékben milyen mintázatot mutatnak a NUTS régiókon, valamint bemutattam, hogy az egyes súlymátrixok által generált területi autokorrelációkat a másik módon definiált súlymátrixok segítségével végzett Moran-féle I-próba milyen arányban

mutatja ki. Az általam definiált 9 területi súlymátrix határértékben különböző mintázatokat mutatott a NUTS régiókon, néhány esetben az adatok a kontinens mértani közép-pontjának környezetében, néhány esetben a kisebb területű régióknál (BeNeLux államok, London) összpontosultak, néhány esetben 1-1 régió teljes centralizációjához vezettek, míg a 4-legközelebbi szomszéd esetében egy hosszúsági fokhoz hasonlóan húzódó sávokat alakítottak ki. Emiatt nem elegendő pusztán a definíció alapján megválasztani a teszteléshez és modellezéshez használt területi súlymátrixot, azt igazítani kell nem csupán az alapsokaság jellemzőihez, hanem a területi sajátosságokhoz is, amelyen az adatok aggregálva lettek. A modellezőknek javasolt ezen karakterisztikákat megfigyelni a távolságdefiníciók alkalmazása előtt, majd ennek tudatában kiválasztani azon mátrixokat melyekkel az autokorreláció tesztelését elvégzik.

A szimulációs adatok eredményei tovább erősítik azt a megfigyelést, hogy a teszteléshez használt mátrix megválasztása során kiemelt figyelmet kell fordítani arra, hogy a területek közötti tudásátadás milyen csatornákon keresztül történhet, s ennek megfelelően kell definiálni a teszteléshez használt mátrixot, hiszen bár voltak olyan szimuláló mátrixok melyek által generált autokorrelációt szinte bármelyik mátrix kimutatta, több esetben nagyon alacsony kimutatási arányt figyelhetünk meg. NUTS 2 és 3 szinten kevésbé nehézkes az autokorreláció kimutatása, mint NUTS 1 szinten, ezen esetekben csupán a szokszor kiemelten rosszul specifikált Régióközpont távolság mátrix emelkedik ki negatív irányban.

Az eredmények alapján az alábbi lépéseket javaslom a területi autokorrelációs modellek illesztése során:

1. Területi kapcsolatok tudásátadási szintjének meghatározása: szomszédsági, távolság alapú, ideológiai.
2. Gazdasági változó eloszlásának vizsgálata, ez alapján a szükséges tesztelési szám meghatározása.
3. Független mintán való vizsgálat a kiválasztott területi súlymátrixokkal, valamint az elsőfajú hiba vizsgálata.
4. A valós földrajzi viszonyok feltérképezése a súlymátrixok által generált területi mintázatok határértékéről, valamint ezen mintázatok összevetése az adat mintázatával.

5. Azon mátrixok kiválasztása, melyek a választott tudásátadási szinten vannak értelmezve, melyek független mintán adott elsőfajú hibája elfogadható, s melyek az adott földrajzi viszonyok mellett határértékben hasonló mintázatot adnak, mint a vizsgált változó, majd ezekkel a Moran-féle I-próbával a területi autokorreláció vizsgálata.
6. Amennyiben ezen vizsgálatok megerősítik a területi autokorreláció létét, a modellezés során ezen területi mátrixok közül a közgazdaságilag leginkább összeegyeztethető alkalmazása.

A tanulmány betekintést nyújt abba, hogy a területi autokorrelációs modellek mennyire érzékenyek a területi súlymátrix megválasztására. A vizsgálatához Európa területének NUTS rendszer alapján történő regionális csoportosítását használok, azonban többször kiemelt szerepet kap az a tény, hogy a különböző földrajzi specifikációk mellett az egyes területi súlymátrixok eltérő eredményeket adhatnak. Területi modellek során a vizsgált terület földrajzi tulajdonságait is figyelembe kell venni, ezért egy lehetséges folytatása a kutatásnak a más területi egységek vizsgálata akár Európában, akár más kontinenseken. Szintén folytatásra ad lehetőséget több területi mátrix definiálása, hiszen ahogy kiemeltem, ezek specifikálásának száma végtelen, s ezen tanulmány csak 9 változattal foglalkozik. Egy újabb egyéni távolságmátrix definiálására adna lehetőséget az egyes régiók közötti légi közlekedési lehetőségek vizsgálata (repülési idő a régiók között, légitársaságok száma, melyek indítanak járatokat az adott területek között, valamint ezek gyakorisága). Az ilyen típusú távolságértelmezés a turisztikai adatok vizsgálata során lenne alkalmazható. Szintén kiemelendő, hogy ezen tanulmány csak az autokorreláció Moran-féle I-próbával való tesztelésére terjed ki, az egyes modellek eredményeinek ilyen típusú összehasonlítása is teret enged még jövőbeni kutatásoknak.

Hivatkozások

- Acs, Z., L. Anselin és A. Varga (1997). „Local Geographic Spillovers Between University Research and High Technology Innovations”. *Journal of Urban Economics* 42.3, 422–448. old. DOI: 10.1006/juec.1997.2032.
- (2002). „Patents and Innovation Counts as Measures of Regional Production of New Knowledge”. *Research Policy* 31.7, 1069–1085. old. DOI: 10.1016/S0048-7333(01)00184-6.
- Alexakis, C., K. Eleftheriou és P. Patsoulis (2021). „COVID-19 Containment Measures and Stock Market Returns: An International Spatial Econometrics Investigation”. *Journal of Behavioral and Experimental Finance* 29, 100428. old. DOI: 10.1016/j.jbef.2020.100428.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. 4. köt. Springer Science & Business Media.
- (2001). „Spatial Econometrics”. *A Companion to Theoretical Econometrics* 310330.
- Asgharian, H., W. Hess és L. Liu (2013). „A Spatial Analysis of International Stock Market Linkages”. *Journal of Banking & Finance* 37.12, 4738–4754. old. DOI: 10.1016/j.jbankfin.2013.08.015.
- Aten, B. (1996). „Evidence of Spatial Autocorrelation in International Prices”. *Review of Income and Wealth* 42.2, 149–163. old. DOI: 10.1111/j.1475-4991.1996.tb00163.x.
- Auzina-Emsina, A. és V. Ozolina (2017). „High-technology Industries Competitiveness and Regional Allocation by NUTS 3 Regions in Latvia”. *Research for Rural Development* 2, 241–248. old. DOI: 10.22616/rrd.23.2017.074.
- Bell, K. P. és N. E. Bockstael (2000). „Applying the generalized-moments estimation approach to spatial problems involving micro-level data”. *Review of Economics and Statistics* 82.1, 72–82. old. DOI: 10.1162/003465300558641.
- Benedek, J. és N. Cs. Veress (2013). „Economic Disparities and Changes in the Convergence of the Romanian NUTS 2 and NUTS 3 Regions”. *Romanian Review of Regional Studies* 9.1, 85. old.
- Bertocchi, D. és tsai. (2021). „Assessing the tourism sustainability of EU regions at the NUTS-2 level with a composite and regionalised indicator”. *Journal of Sustainable Tourism* 0.0, 1–18. old. DOI: 10.1080/09669582.2021.2000993.

- Bivand, R., W. G. Müller és M. Reder (2009). „Power calculations for global and local Moran’s I”. *Computational Statistics Data Analysis* 53.8, 2859–2872. old. DOI: 10.1016/j.csda.2008.07.021.
- Case, A. C. (1991). „Spatial Patterns in Household Demand”. *Econometrica: Journal of the Econometric Society*, 953–965. old. DOI: 10.2307/2938168.
- Cliff, A. D. és J. K. Ord (1981). *Spatial Processes: Models & Applications*. Taylor & Francis.
- Dauvin, M. és L. Grzybowski (2014). „Estimating Broadband Diffusion in the EU Using NUTS 1 Regional Data”. *Telecommunications Policy* 38.1, 96–104. old. DOI: 10.1016/j.telpol.2013.05.006.
- EUROSTAT (2022a). *"GISCO: Geographical Information and Maps"*. Letöltve: Marc. 17, 2022. URL: <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts>.
- (2022b). *"Regions and cities — Overview"*. Letöltve: Marc. 15, 2022. URL: <https://ec.europa.eu/eurostat/web/regions-and-cities/overview>.
- Fischer, M. M. és A. Varga (2003). „Spatial Knowledge Spillovers and University Research: Evidence from Austria”. *The Annals of Regional Science* 37.2, 303–322. old. DOI: 10.1007/s001680200115.
- Fukatsu, A. és T. Kawashima (1999). „Urbanization, Suburbanization and Revived-urbanization: ROXY-index Analysis for the Chuo-line Region of Tokyo”. *Gakushuin Economic Papers: The Journal of Faculty of Economics, Gakushuin University* 36.3, 389–414. old.
- Gyódi, K. és L. Nawaro (2021). „Determinants of Airbnb Prices in European Cities: A Spatial Econometrics Approach”. *Tourism Management* 86, 104319. old. DOI: 10.1016/j.tourman.2021.104319.
- Hirvonen, M., N. Hiraoka és T. Kawashima (1999). „Long-term Urban Development of the Finnish Population: Application of the ROXY-index Analytical Method”. *Gakushuin Economic Papers (Gakushuin Daigaku Keizai Ronshuh)* 36.2, 243–263. old.
- Hu, L., Y. Chun és D. A. Griffith (2020). „Uncovering a Positive and Negative Spatial Autocorrelation Mixture Pattern: A Spatial Analysis of Breast Cancer Incidences in Broward County, Florida, 2000–2010”. *Journal of Geographical Systems* 22 (3), 291–308. old. DOI: 10.1007/s10109-020-00323-5.

- Jackson, M. C. és tsai. (2010). „A modified version of Moran’s I”. *International journal of health geographics* 9.1, 1–10. old. DOI: 10.1186/1476-072X-9-33.
- Kahle, D. és H. Wickham (2013). „ggmap: Spatial Visualization with ggplot2”. *The R Journal* 5.1, 144–161. old. URL: <https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>.
- Kelejian, H. H. és I. R. Prucha (1999). „A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model”. *International Economic Review* 40.2, 509–533. old. DOI: 10.1111/1468-2354.00027.
- Klaassen, L. és J. Paelinck (1979). „The Future of Large Towns”. *Environment and Planning A* 11.10, 1095–1104. old. DOI: 10.1068/a111095.
- Laudor, C. R. és S. Tabriztchi (1975). „Predicting Urban Growth in Iran: A Dynamic Model”. *Iranian Studies* 8.3, 124–133. old. DOI: 10.1080/00210867508701494.
- Li, H., C. A. Calder és N. Cressie (2007). „Beyond Moran’s I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model”. *Geographical Analysis* 39.4, 357–375. old. DOI: 10.1111/j.1538-4632.2007.00708.x.
- Moran, P. A. (1948). „The Interpretation of Statistical Maps”. *Journal of the Royal Statistical Society. Series B (Methodological)* 10.2, 243–251. old. URL: <http://www.jstor.org/stable/2983777>.
- Mura, M. és tsai. (2021). „Industrial carbon emission intensity: A comprehensive dataset of European regions”. *Data in Brief* 36, 107046. old. DOI: <https://doi.org/10.1016/j.dib.2021.107046>.
- Oden, N. (1995). „Adjusting Moran’s I for population density”. *Statistics in Medicine* 14.1, 17–26. old. DOI: <https://doi.org/10.1002/sim.4780140104>.
- Paas, T. és tsai. (2007). „Econometric Analysis of Income Convergence in Selected EU Countries and their NUTS 3 Level Regions”. *The University of Tartu Faculty of Economics and Business Administration Working Paper* 60-2007.
- Paelinck, J. és tsai. (1979). *Spatial Econometrics*. Saxon Hause.
- Paluchowski, P. (2012. júl.). „Demographic and Economic Disparities for NUTS Regions in Germany and the UK in the Recent Past (1990–2010) and Future”. *Cambridge Journal of Regions, Economy and Society* 5.2, 239–254. old. DOI: 10.1093/cjres/rsr032.

- Ribeiro, V. M. és L. Bao (2021). „Impact of Next-generation Access Networks on the Innovation Efficiency of Portuguese Municipalities: A Spatial Econometrics Approach”. *Regional Science Policy & Practice* 13.5, 1615–1637. old. DOI: 10.1111/rsp3.12436.
- Salazar, J. J. és tsai. (2022). „Fair Train-Test Split in Machine Learning: Mitigating Spatial Autocorrelation for Improved Prediction Accuracy”. *Journal of Petroleum Science and Engineering* 209, 109885. old. DOI: 10.1016/j.petrol.2021.109885.
- Spijker, J. és tsai. (2021). „Mortality by Cause of Death in Colombia: A Local Analysis Using Spatial Econometrics”. *Journal of Geographical Systems* 23.2, 161–207. old. DOI: 10.1007/s10109-020-00335-1.
- Tiefelsdorf, M. és B. Boots (1995). „The Exact Distribution of Moran’s I”. *Environment and Planning A: Economy and Space* 27.6, 985–999. old. DOI: 10.1068/a270985.
- Varga, A. (2000). „Local Academic Knowledge Transfers and the Concentration of Economic Activity”. *Journal of Regional Science* 40.2, 289–309. old. DOI: 10.1111/0022-4146.00175.
- (2002). „Térökonometria”. *Statistikai Szemle* 80, 354–370. old.
- Zhang, F., F. Wang és S. Yao (2021). „High-speed Rail Accessibility and Haze Pollution in China: A Spatial Econometrics Perspective”. *Transportation Research Part D: Transport and Environment* 94, 102802. old. DOI: 10.1016/j.trd.2021.102802.

7. Függelék

Tisztított területek:

- *NUTS 1 szinten:*

Região Autónoma dos Açores (Portugália) - PT2, Região Autónoma da Madeira (Portugália) - PT3, RUP FR — Régions Ultrapériphériques Françaises (Franciaország) - FRY, Canarias (Spanyolország) - ES7

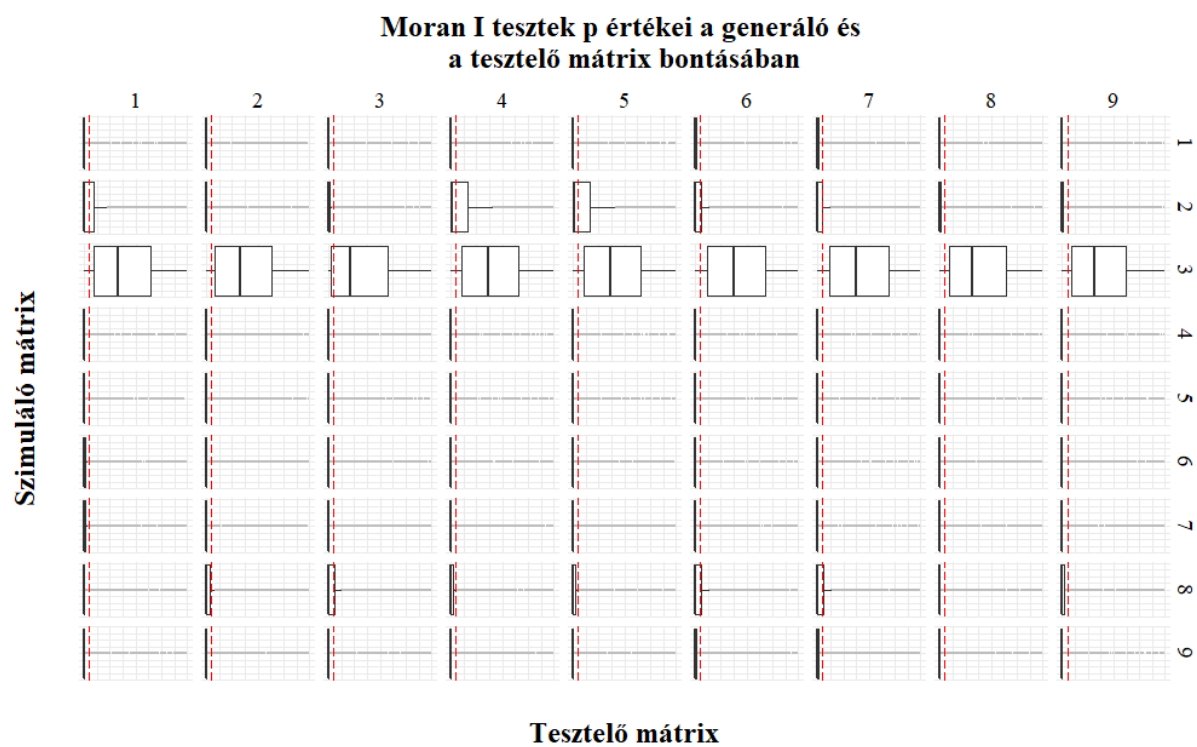
- *NUTS 2 szinten:*

Guadeloupe (Franciaország) - FRY1, Martinique (Franciaország) - FRY2, Guyane (Franciaország) - FRY3, La Réunion (Franciaország) - FRY4, Mayotte (Franciaország) - FRY5, Canarias (Spanyolország) - ES70, Região Autónoma dos Açores (Portugália) - PT20, Região Autónoma da Madeira (Portugália) - PT30

- *NUTS 3 szinten:*

El Hierro (Spanyolország) - ES703, Fuerteventura (Spanyolország) - ES704, Gran Canaria (Spanyolország) - ES705, La Gomera (Spanyolország) - ES706, La Palma (Spanyolország) - ES707, Lanzarote (Spanyolország) - ES708, Tenerife (Spanyolország) - ES709, Guadeloupe (Franciaország) - FRY10, Martinique (Franciaország) - FRY20, Guyane (Franciaország) - FRY30, La Réunion (Franciaország) - FRY40, Mayotte (Franciaország) - FRY50, Região Autónoma dos Açores (Portugália) - PT200, Região Autónoma da Madeira (Portugália) - PT300

EUROSTAT adatletöltési kódok: AEI_PR_SOILER, AEI_FM_MS, CENS_11AG_R3, CENS_11DWOB_R3, CENS_11MS_R3, DEMO_R_BIRTHS, DEMO_R_D3DENS, DEMO_R_DEATHS, DEMO_R_FIND3, DEMO_R_PJANAGGR3, DEMO_R_PJANIND3, EF_R_NUTS, NAMA_10R_3GVA, NAMA_10R_3POPGDP, TGS00002-07, TGS00010, TGS00024, TGS00026, TGS00036-43, TGS00045-52, TGS00054, TGS00057-64, TGS00075-78, TGS00096-109, TGS00111-114



25. ábra. Szimulációs eredmények p értékei a generáló és a tesztelő mátrix szerinti bontásban NUTS 3 szinten. (Forrás: Saját szerkesztés)