

TANULMÁNYOK

FÁBIÁN ZSÓFIA

A földrajzilag súlyozott regresszió módszere és alkalmazhatósági példája

Bevezetés

Tanulmányomban egy, a külföldi szakirodalomban az elmúlt időszakban elterjedt, ám Magyarországon csak helyenként megjelenő statisztikai módszerrel, a földrajzilag súlyozott regresszióval foglalkozom. A földrajzilag súlyozott regresszió egy regressziós alapokra épülő, a földrajzi közelséget kiemelten figyelembe vevő lokális modell. Munkám első részében a módszertanról írok, ezt követően egy alkalmazási példán keresztül bemutatom a módszer működését: az egy főre jutó GDP regionális adatsorain végzett számítások segítségével kísérlem meg elemezni a gazdasági fejlettség lokális sajátosságait.

A földrajzilag súlyozott regresszió

A regresszió a társadalomtudományi kutatások egyik legelterjedtebb matematikai-statisztikai vizsgálati eszköze. Közkeletűsége a lényegéből adódik, hiszen ez a kutatások fő célját szolgáló, a jelenségek közötti kapcsolatok feltárására alkalmas módszer. A regressziószámítás alapjában véve két vagy több változó közötti sztochasztikus kapcsolatot leíró függvény megkeresését és leírását jelenti, ami annyiban más, mint a korreláció (hiszen a korreláció is a változók közötti valószínűségi kapcsolatot vizsgálja), hogy míg az utóbbi csak a kapcsolat meglétét jelzi, addig ez részleteiben nem közöl információkat, a kapcsolat további jellemzőinek feltárásához, másik módszert kell használni. E másik módszer a legtöbb esetben a regresszió szokott lenni.

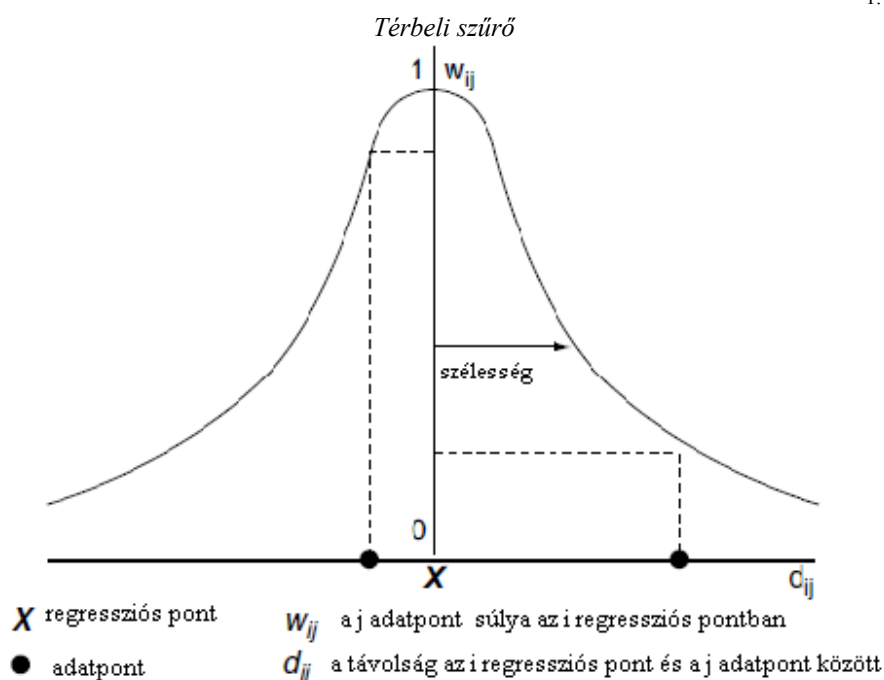
Azonban területi adatok esetében használva jelentős problémával találkozhatunk, ugyanis alapesetben a regresszió az egyes jelenségeket úgy vizsgálja, mintha változatlanok lennének a térben. Ezzel szemben a földrajzilag súlyozott regresszió (*geographically weighted regression*, továbbiakban GWR) a módszertanából adódóan alkalmas arra, hogy a térbeli folyamatokat változóként modellezze, azaz a GWR annyiban több az egyszerű regresszióhoz, hogy megoldja azt a problémát, hogyan lehet vizsgálni a folytonos területi folyamatokat diszkrét súlyozás segítségével (1. ábra).

A GWR logikája a mozgó ablakos regresszióéra (*moving window regression*) hasonlít a legjobban. A mozgó ablakos regresszió segítségével megoldható az a probléma, hogy az egyes folyamatok nem érnek véget a területi egységek határainál. Módszertana a következő. Első lépésként a vizsgált területre a regressziós pontok rácshálóját kell kifeszíteni. Ezáltal minden regressziós pont körül meg lehet határozni egy régiót/ablakot, ami általában négyzet vagy kör alakú, bár elméletileg bármilyen formát felvehetne, ami képes lefedni a teret, abban az értelemben, hogy az összes vizsgálati pontot magába foglalja. A vizsgálati kérdéstől függően tudjuk meghatározni azt, hogy mit értünk az egyes régiók

alatt (például egy regressziós pontot érintő négyzetek). A regressziós modell az egyes, a regressziós pontok körül létrehozott régiókba eső adatpontokon alapul, a folyamat az összes regressziós pont esetében megismétlődik. Az eredményül kapott lokálisparaméterbecsléseket térképezve meg lehet vizsgálni a nem stacionárius feltevést (Fotheringham–Brunsdon–Charlton 2002).

A fő problémája ennek a módszernek az, hogy nem folytonosként kezeli az amúgy annak tekinthető térbeli folyamatokat. Ugyanis a régiókba/ablakba eső adatpontoknak 1 értékű súlyt ad, míg a kívül esőknek 0-t, ami folytonos jelenségeknél önkényesnek tűnhet. Természetesen az eredmény nagyban függ a választott régió/ablak méretétől, valamint a széle felé a kevesebb regressziós pont miatt, ezt kiküszöbölő intézkedések híján, torzít is. A GWR esetében az összes meghatározott távolságon belüli adatpontot súlyozni kell a regressziós ponttól való távolságukkal, s ebből adódóan a regressziós ponthoz közelebbi adatpontok nagyobb súllyal fognak szerepelni a modellben, mint azok, amelyek távolabb vannak (Fotheringham–Brunsdon–Charlton 2002).

1. ábra

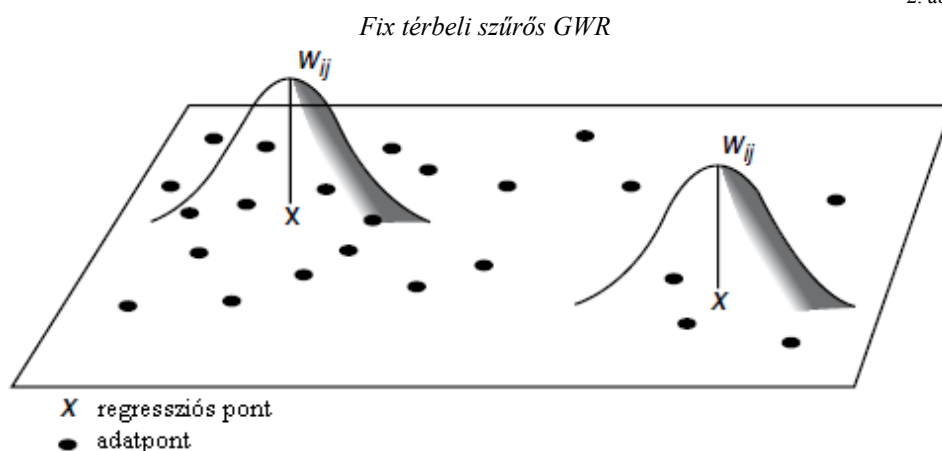


Forrás: Fotheringham–Brunsdon–Charlton 2002, p. 44.

Lehetőség van fix térbeli szűrőt (*fixed spatial kernel*) használni (2. ábra). Ebben az esetben egy adott regressziós pontot tekintve a legnagyobb súllyal az az adatpont fog szerepelni, amely a regressziós ponttal egybeesik. Ez a maximális súly folyamatosan csökken miként a távolság a két pont (a regressziós és az adatpont) között nő. E módszer szerint a regressziós modell úgy lesz lokális, hogy tulajdonképpen a regressziós pont mozog a vizsgált területen. Mivel minden egyes helyen az adatpont másként van súlyoz-

va, a lokális számítások teljesen különbözőek. Letérképezve ezeket a lokális számításokat, a paraméterbecsléseket tartalmazó felületet kapjuk. A GWR eredménye nem érzékeny a súlyozásra, azonban érzékeny a súlyozáshoz használt szélességre/átmérőre. Tehát ennek optimális meghatározása különösen fontos az egyes vizsgálatok esetében. A mozgó ablakos regresszióval összehasonlításként az mondható el, hogy a szűrők különbségéből adódóan a GWR kevésbé ad kiegyenlített képet, jobban kiemeli a lokális különbségeket, tehát folytonos jelenségek esetében használata indokoltabb (Fotheringham–Brunsdon–Charlton 2002).

2. ábra



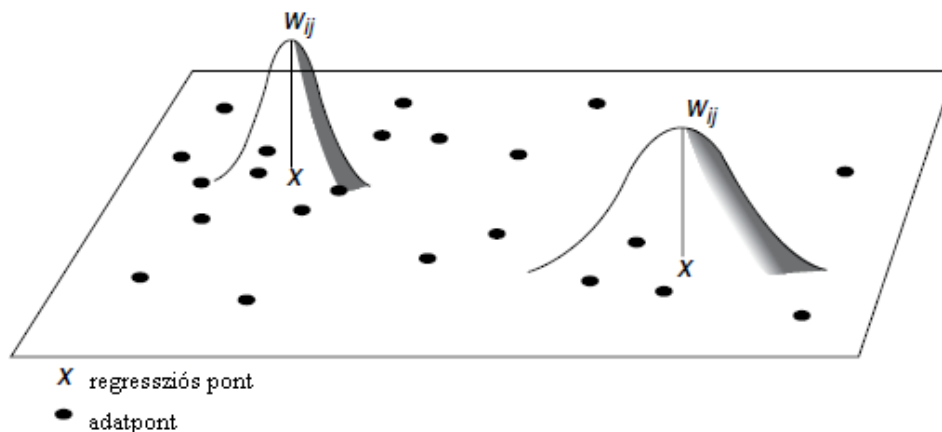
Forrás: Fotheringham–Brunsdon–Charlton 2002, p. 45.

A gyakorlatban ez a vizsgált ponttól egy bizonyos sugarú körön belül levő pontok figyelembevételét jelenti, a kiindulási ponttól távolodva egyre csökkenő súllyal. A kör sugarát a vizsgálat léptéke határozza meg, azonban több lehetőséget ki szoktak próbálni.

A fix térbeli szűrős GWR esetében is felvetődik az a probléma, hogy egyes térrészekben sokkal ritkábban helyezkednek el az adatpontok, így a belőlük becsült lokális modellek hibátényezője is nagyobb. Extrém esetben a kevés adatnak köszönhetően egyes paraméterek becslése nem lehetséges. Megoldása az adaptív szűrő használatával lehetséges (3. ábra). Ennek sajátossága, hogy szélessége alkalmazkodni tud az adatpontok számához, sűrűségéhez, tehát ahol sűrűbben vannak az adatok, ott keskenyebb lesz, ahol ritkábban, ott szélesebb. Magyarország települési hálózatát nézve jó példa az adaptív kernel szükségessége az Alföld és a Dunántúl eltérő településsűrűsége. Egy fix kerneles térképpel összehasonlítva az adaptív kerneles eredményét nagyon hasonló képet láthatunk azzal a különbséggel, hogy az adaptív kerneles kiegyenlítettebb. Ennek oka az, hogy a kevesebb pontból kalibrált lokális modellek esetében nagyobb változatosság várható (Fotheringham–Brunsdon–Charlton 2002).

Az adaptív kerneles vizsgálatot a gyakorlatban legegyszerűbben úgy lehet elvégezni, hogy az egyes pontok bizonyos számú szomszédja alapján állítunk fel regressziós modelleket.

Adaptív szűrős GWR



Forrás: Fotheringham–Brunsdon–Charlton 2002, p. 47.

Egy példán keresztül megvilágítva ezt a módszert úgy kell elképzelni, hogy annyi különböző regressziós egyenlet/beclsés áll rendelkezésünkre, ahány vizsgált területegység van. Ugyanis az egyes területegységekre a meghatározott módon leválogatott szomszédjaikkal együtt állítunk fel egy-egy regressziós modellt, tehát mintegy lokálisan számolunk. A GWR összetett statisztikai módszer, alkalmazásakor az autokorreláció, a regresszió és a területi mozgóátlag számítás egyes elemei is felfedezhetők. A módszer logikája megegyezik a területi mozgóátlag logikájával, a fix térbeli szűrős GWR a változatlan sugarú mozgóátlaghoz, az adaptív szűrős a változó sugarú mozgóátlaghoz hasonló (Dusek 2004).

Habár pontszerű vizsgálati egységekre vonatkozóan írták le ezt a módszert, mint számos szakirodalmi példa bizonyítja, ponttal jellemezhető, nagyobb kiterjedésű területegységek esetén is alkalmazható (például Chasco–García–Vicéns 2007, Eckey–Kosfeld–Türck 2007, Yu 2005).

A földrajzilag súlyozott regresszió egyenlete (Fotheringham–Brunsdon–Charlton 2002):

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i,$$

ahol az (u_i, v_i) az i -edik adatpont földrajzi koordinátáit jelenti, a $\beta_k(u_i, v_i)$ a számított értéke $\beta_k(u, v)$ folytonos függvénynek az i -edik pontban.

A GWR lényege, hogy úgy kezeli a regressziós koefficiens, mint a fekvés függvényét, és nem, mint fix konstans értéket (Yu 2005). A GWR a globális regresszió kereteit úgy bővíti, hogy lehetővé teszi a paraméterek lokális becslését (Bálint 2010).

Súlyozási lehetőségek

Először a fix kernelek súlyozási lehetőségeit tekintem át, e kernelek alakja és kiterjedése nem változik a vizsgálat során:

$$w_{ij} = 1 \text{ minden } i, j,$$

ahol j egy pontot jelent a térben, ahol megfigyelés történt, i pedig egy pontot a térben, aminek a paramétere becsült (Fotheringham–Brunsdon–Charlton 1996).

Ezt a közelítést a globális modellben használják, ott ugyanis minden egyenlő súllyal szerepel.

Egy lehetséges elmozdulás a lokális figyelembevétele felé, amikor a regressziós ponttól nézve egy bizonyos távolságon kívül már nem veszünk figyelembe egyes elemeket. Ez azonos azzal, ha ezeknek 0-s súlyt adunk. A mozgó ablakos regressziós esetben használják ezt a megközelítést:

$$w_{ij} = 1 \text{ ha } d_{ij} < d, \\ w_{ij} = 0 \text{ máskülönben.}$$

Ez a súlyozási módszer egyszerűsíti a számítást, ugyanis az egyes regressziós pontoknál csak az adatpontok egy-egy részhalmazával kell tovább számolni. Ezzel a súlyozással az a probléma, hogy nem folytonosként kezeli az amúgy annak tekinthető térbeli folyamatokat. Ahogy a regressziós pont változik, a becsült koefficiens drasztikusan változhat azáltal, hogy egy adatpont ki- vagy bekerült az „ablakba” (Charlton–Fotheringham–Brunsdon 1997).

Egy lehetséges megoldás arra, hogy a súlyok nem folytonosak, egy olyan w_{ij} mátrix meghatározása, amely a d_{ij} (távolság i és j pont között) folytonos távolság függvényből származik:

$$w_{ij} = \exp [-1/2(d_{ij}/b)^2],$$

ahol b a sáv szélességet jelenti. Ha i és j pont egybeesik, hiszen i pont mérési pont is lehet, az adat súlya ebben a pontban egységnyi lesz, és a többi adatpont súlya csökkenni fog a Gauss-görbe szerint, ahogy a távolság i és j pontok között nő (Fotheringham–Brunsdon–Charlton 1998).

Egy másik lehetőség, amikor egy kernel használja a b^2 függvényt:

$$w_{ij} = \exp [1 - (d_{ij}/b)^2] \text{ ha } d_{ij} < b, \\ w_{ij} = 0 \text{ máskülönben.}$$

Ami azért hasznos, mert folytonos, közel gaussi eloszlású súlyfüggvény, b távolságig a regressziós ponttól és 0 súllyal a b túli adatpontokban (Fotheringham–Brunsdon–Charlton 1998).

Áttérve az adaptív kernelek bemutatására, több érv is szól ezek használatára mellett. Először is, ahol az adatpontok sűrűn helyezkednek el, ott több esély van a kapcsolat változásának vizsgálatára elég kis távolságon belül, ami nagyobb kiterjedésű fix kernel használata esetén figyelmen kívül maradna. Egy olyan terület esetén, ahol az adatpontok elhelyezkedése ritkás, fix kernelt használva a becsült standard hibák értéke magas lehet, mert kevés a felhasznált adatpont.

A GWR számításakor alkalmazott adaptív kerneleknek legalább három fajtája létezik. Az első szerint sorba kell rendezni az adatpontokat aszerint, hogy milyen távol vannak minden egyes i ponttól:

$$w_{ij} = \exp -(R_{ij}/b),$$

ahol R_{ij} a j -edik pont rangszáma i -től, tehát j távolsága i -től. Az i -hez legközelebbi adatpont 1 értékű súlyt kap, és a súlyok csökkennek, ahogy a rangszám nő. Ez automatikusan csökkenti a kernel sáv szélességét azokon a területeken, ahol sok adat van, mert a távolság a 10 legközelebbi adatpontot véve kevesebb lesz, mint amikor a regressziós pont egy

kevés adatpontot magába foglaló régióban van (Fotheringham–Brunsdon–Charlton 2002).

Egy másik komplikáltabb lehetőség adaptív kernel létrehozására az, hogy meghatározzuk azt, hogy a súlyok összege bármely i pontban egy állandóval, C -vel egyenlő. Azokon a területeken, ahol az adatpontok sűrűn helyezkednek el, ott a kernelnek össze kell zsugorodnia, hogy a súlyok összege a meghatározott C érték legyen, míg ahol kevesebb a pont, ott a kernel kiszélesedik:

$$\sum_j w_{ij} = C \text{ az összes } i\text{-re.}$$

E módszer esetében az optimális C érték meghatározása okozhat nehézséget. Ezt a következőképpen lehet megtenni: első lépésként választani kell egy tetszőleges értéket, aztán ezzel az értékkel kell létrehozni a súlyfüggvényt, és egy *goodness-of-fit* illeszkedési tesztet kell futtatni a modellre. Ezt követően egy másik C értéket kell választani, a súlyfüggvényt létrehozni, ismét egy *goodness-of-fit* tesztet futtatni és ezt a két lépést ismételni, amíg meg nem találjuk az optimálisan illeszkedő C értéket (Fotheringham–Brunsdon–Charlton 2002).

Harmadik lehetőségként említeném az N számú legközelebbi szomszéd figyelembevételét:

$$w_{ij} = 1 \text{ ha } j \text{ az egyike } i \text{ } N \text{ darab legközelebbi szomszédjának,}$$

$$w_{ij} = 0 \text{ máskülönben,}$$

vagy

$$w_{ij} = [1 - (d_{ij}/b)^2]^2 \text{ ha } j \text{ az egyike } i \text{ } N \text{ darab legközelebbi szomszédjának és } b \text{ az } n \text{ legközelebbi szomszéd távolsága,}$$

$$w_{ij} = 0 \text{ máskülönben.}$$

Ebben az esetben a modell kalibrálásához hozzátartozik az N értékének meghatározása is. Az N tehát azoknak az adatpontoknak a számát jelenti, amelyek a lokális modell kalibrálásába beleértendőek, és a súlyfüggvény meghatározza minden egyes pont súlyát a N -edik darabig. A súlyok 0 felé tartanak (Fotheringham–Brunsdon–Charlton 2002).

A kernel optimális átmérőjének meghatározása

Az egyik lehetőség az, hogy z értékét minimalizáljuk, a következők szerint:

$$z = \sum [y_i - \hat{y}_i(b)]^2,$$

ahol $\hat{y}_i$ a függő változó becslült értéke, b átmérőt használva. Ahhoz, hogy ezt a becslült értéket megkapjuk, szükség van minden adatpontban a $\beta_k(u_i, v_i)$ értékek becslésére és az x értékekre. Egy általános esetet nézve felléphet az a probléma, hogy ha b túl kicsi értékét vesszük, akkor a súlyozással i -t leszámítva a többi pont értéke elhanyagolhatóvá válik. Ekkor a becslült érték a kiválasztott pontoknál az eredetihez nagyon hasonló lesz, így az egyenlet értéke is 0 lesz. Egy ilyen modell paraméterei tehát bizonyos esetekben nem lesznek meghatározhatók, valamint a becslés változni fog a térben annak érdekében, hogy helyileg megfelelő becslült értékek legyenek minden regressziós pontban (Fotheringham–Brunsdon–Charlton 2002).

Ennek kiküszöbölése érdekében szükség van a fenti összefüggés kiterjesztésére. A lokális regressziók esetén is használt keresztvalidálásra (*cross-validation*):

$$CV = \sum [y_i - \hat{y}_{\hat{i}}(b)]^2,$$

ahol az $\hat{y}_{\#i}$ a függő változó becsült értéke, az i pontot kihagyva a kalibrálási folyamatból. Ez azért jó megoldás, mert ha a b kis értéket vesz fel, akkor a modell az i -hez közeli pontokra fog épülni, és nem magára az i pontra (Fotheringham–Brunsdon–Charlton 2002).

A GWR felhasználása széles körű, ugyanis gazdasági és természetföldrajzi kutatások során is alkalmazzák. Például Yu (2005) tanulmányában a GWR segítségével Peking tágabb körzetének regionális fejlődését vizsgálta területi heterogenitás szempontjából. Lin, Cromely és Yang (2011) területi interpolációs problémák megoldására használták a módszert. Fotheringham, Brunsdon és Charlton (2002) londoni lakásárakra vonatkozóan állítottak fel egy GWR modellt, azzal a céllal, hogy a lokális sajátságokat feltárják. Eckey, Kosfeld és Türck (2007) a német munkaerőpiac regionális konvergenciáját elemezték e módszerrel. Ridefelt, Etzelmüller, Boelhouwers és Jonasson (2011) a hegyi permafroszt kiterjedésének modellezésekor a stacionaritás meglétét vizsgálták a földrajzilag súlyozott regresszió segítségével.

Egyes vélekedések szerint a GWR-módszer az egyik leggyakrabban használt lokális regresszió, amely a térben változó hatások bemutatására kitűnő vizualizációs eszköznek tekinthető (Bálint 2010). Számos kritika fogalmazódott meg ezzel a módszerrel kapcsolatban, amelyek szerint ez nem tekinthető modellnek, mert szerepe inkább illusztratív, és a hangsúly nem a becsléseken, hanem a paraméterbecslések térbeli mintázatán van. További hátránya a módszernek, hogy érzékeny az outlier adatokra, valamint a multikollinearitásból adódó problémák előfordulásának is nagyobb a valószínűsége, mint egy globális regressziós modellben (Lloyd 2007, Wheeler–Páez 2010).

Az európai regionális fejlettségi tagozódás lokális megközelítésű vizsgálata

A következőkben egy alkalmazási példa segítségével azt szeretném bemutatni, hogy a GWR-módszer használata mivel jelent többet a globális regresszió használatánál.

A területi keret meghatározásával kezdve: a számítások elsősorban az Európai Unió tagállamaira vonatkoznak, beleértve a tagjelölt országok közül Horvátországot és Macedóniát is. A vizsgálat minél teljesebb volta miatt bevettem még Svájcot és Norvégiát is. A kontinenstől való nagy távolságuk miatt kihagytam Ciprust és Máltát, valamint Franciaország, Portugália és Spanyolország tengeren túli régióit.

A számításokat először NUTS 2-es szinten kezdtem, azonban az eredmények értelmezésekor nyilvánvalóvá vált, hogy ez nem elég részletes megközelítés, ugyanis Európa alakja miatt a szomszédok kijelölése jelentős esetben nehézségbe ütközött. Emiatt úgy döntöttem, hogy alacsonyabb szintű területegységekre vonatkozóan végzem el a vizsgálatokat. Az európai területi beosztást nézve ez értelemszerűen a NUTS 3-as szint használatát jelentené. Azonban a NUTS-beosztás heterogenitása miatt a NUTS 2-es és 3-as szint vegyes használata mellett döntöttem. Belgium, az Egyesült Királyság, Hollandia, Németország és Svájc esetében maradtam a NUTS 2-es szintnél, a többi ország esetében áttértem a NUTS 3-asra (793 régió). Így homogénebb adatsor állt rendelkezésemre, emiatt a területegységek közötti nagyságkülönbségükből adódó szórás is kisebb lett.

Az adatok nagy része az Eurostat *Regio* adatbázisából származik. Az egyes adatso-rokból hiányzó egy részét a nemzeti statisztikai hivatalok honlapjáról pótoltam, más esetekben pedig hosszabb idősor alapján regresszió segítségével becsültem őket.

A számításokat ismertetve, az egy főre jutó vásárlóerő-paritáson mért GDP-adatokra állítottam fel egy modellt 2009-re vonatkozóan. A számításokat Matlab-programmal végeztem, Fotheringham–Brunsdon–Charlton *Geographically Weighted Regression, the analysis of spatially varying relationships* című módszertani munkája és LeSage *The Theory and Practice of Spatial Econometrics* című, egyes statisztikai módszerek e programmal történő kiszámítását bemutató munkája alapján.

A függő változó:

– egy főre jutó GDP.

A független változók:

– gazdaságilag aktívak aránya,

– munkanélküliségi ráta,

– népsűrűség,

– a tercier szektorban foglalkoztatottak aránya.¹

Az eredmények értelmezésekor érdemes számos lehetőséget figyelembe venni. Van lehetőség az egyes regressziós egyenletek standardizált bétaértékeinek vizsgálatára, ezáltal arra kapunk választ, hogy az egyes területegységekre vonatkozóan mely tényezőknek van nagyobb szerepe a jelenség alakulásában (Eckey–Kosfeld–Türck 2007).

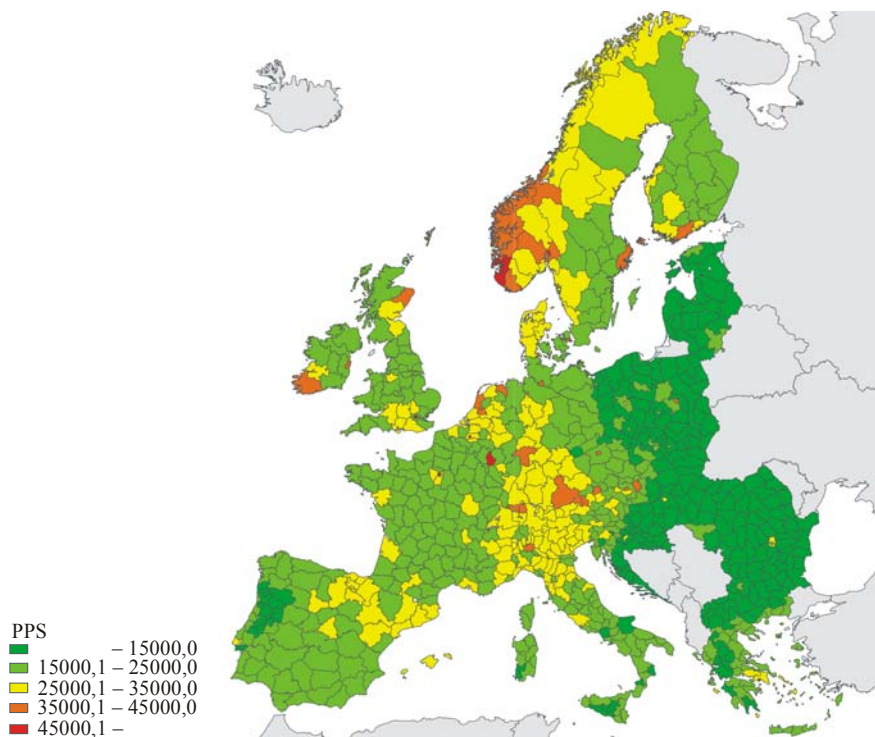
Az R^2 értékeket térképezve láthatjuk, hogy mennyire megbízható a modell az egyes területegységekre vonatkozóan, milyen a modellek térbeli relevanciája (Xiaomin–Shuo-sheng 2011).

Összevethetjük az eredeti és az egyenletek segítségével becsült értékeket.

Először, a kiindulási egy főre jutó GDP értékeket nézve, azt láthatjuk, hogy a fővárosokat magukba foglaló, illetve egyes nagyvárosi régiók vannak kedvezőbb helyzetben (4. ábra). Emellett kirajzolódik egy magas GDP/fő értékekkel rendelkező központi térség, amely elsősorban a dél-német, svájci és észak-olasz térségeket foglalja magába, valamint jelentős kőolajbevételeinek és alacsony lakosságának köszönhetően Norvégia régiói is kiemelkedően magas értékekkel rendelkeznek. Az e tekintetben legelmaradottabb térségek a 2004-ben, illetve a 2007-ben újonnan csatlakozott tagállamok régiói.

¹ Kiindulásként felállítottam egy OLS-modellt, annak érdekében, hogy megnézzem, mely mutatók vannak hatással az egy főre jutó GDP alakulására. A számításokat SPSS-programmal, Regression/Stepwise-módszerrel végeztem. Azért választottam, ezt a módszert, mert a függő és független változók meghatározása után a program a legoptimálisabb regressziós modellt alakítja ki.

4. ábra

Az egy főre jutó GDP értékek alakulása Európa régióiban, 2009

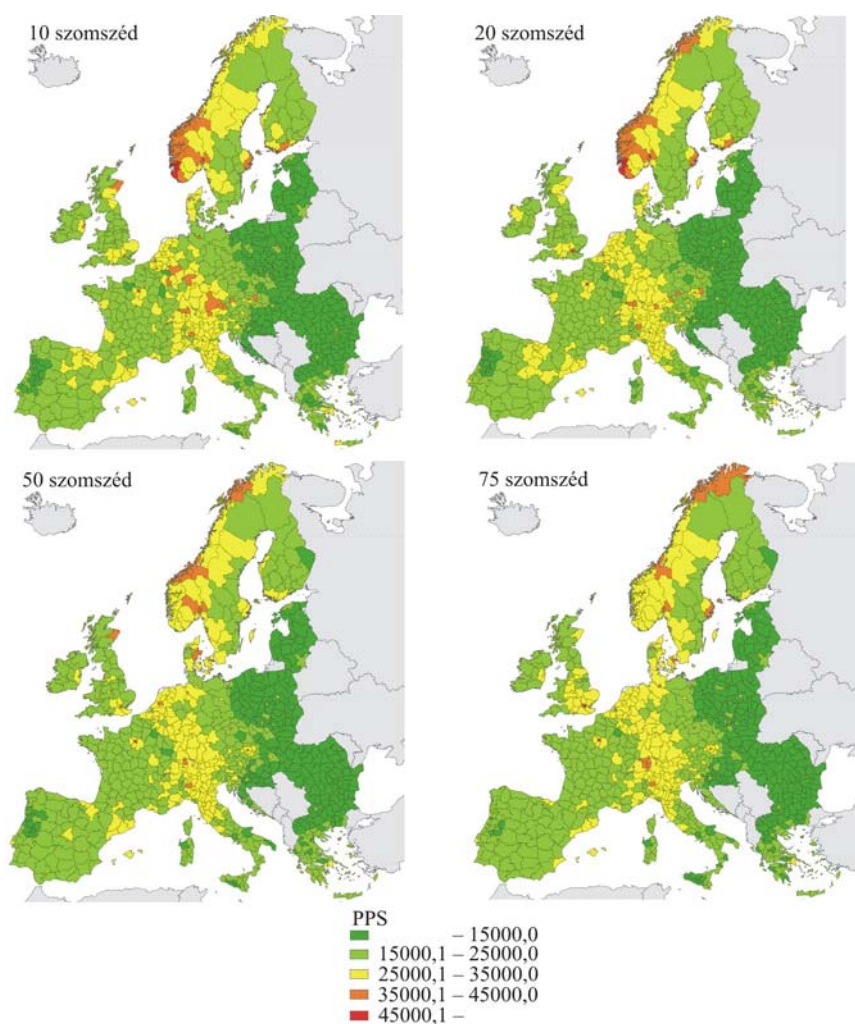
Ezt követően megnéztem, hogy a GWR-modell milyen becslést ad a függő változóra, különböző sávszélességet használva. Mivel a különböző sávszélességgel számított eredményeket is össze szerettem volna hasonlítani, optimális kiszámításától eltekintettem. Mind fix, mind adaptív kernelt használtam, a fix kernelek esetében a b^2 súlyfüggvényt választottam, az adaptív kernelek esetében az N darab legközelebbi szomszéd módszerét alkalmaztam. A fix kernel esetében 75, 100 illetve 150 km sugarú körön belül vettem figyelembe az adatpontokat. Az adaptív kernelnél a 10, 20, 50, illetve 75 legközelebbi szomszédot néztem. Elsőként megállapítottam, hogy Európa „szabálytalan” alakja, valamint a területi beosztás egyenetlen volta miatt ebben az esetben a fix kerneles kevésbé jó közelítés. Ugyanis, ha például 100 km-es távolságon belül keresünk szomszédokat, akkor a kontinens legtöbb régiója esetében több szomszédot is találunk, azonban a távolabb levő, nagyobb területű északskandináv és skót régiók esetében egyet sem. Ebből kifolyólag a fix kerneles módszert tovább nem vittem.

Az adaptív kernellel számolt eredményeket térképen ábrázolva (5. ábra) azt állapíthatjuk meg, hogy a GWR-moddellel végzett becslések az egy főre jutó GDP-re vonatkozóan jó közelítést adnak. Ez azt jelenti, hogy ha az eredeti térképpel összehasonlítjuk a GWR-rel becsült értékeket ábrázolókat, láthatjuk, hogy jelentős a hasonlóság. Mint várható volt, az is megfigyelhető, hogy minél több adatpontot vonunk be az egyes lokális egyen-

letekbe, tehát minél nagyobb a sávszélesség, annál homogénebb képet kapunk. Ugyanis nagyobb mennyiségű pontot figyelembe véve az egyes területek egyedisége jobban háttérbe szorul. A későbbiekben részletesebben a 20 szomszédot figyelembe vevő számítás eredményét fogom elemezni, mivel ez az elemszám tükrözi vissza a regionális viszonyokat, hiszen európai viszonylatban 50 szomszédos régió már a szűkebb értelemben vett regionális lépték fölött van, illetve a 10 elemet tartalmazó regressziós modelleket kevésbé tartom megbízhatónak.

5. ábra

Az egy főre jutó jövedelem GWR-rel meghatározott számú szomszédot figyelembe vevő adaptív kernellel becsült értékei



Ezt követően letérképeztem a lokális egyenletek R^2 értékeit, annak érdekében, hogy megvizsgáljam a földrajzi relevanciájukat (6. ábra). A térképeket nézve elsőként az álla-

pítható meg, hogy a sávszélesség növekedésével csökkent az egyes modellek magyarázóereje. Ez persze megfelel a várt eredményeknek, ugyanis minél több szomszédot/elemet figyelembe véve egyre több olyan is bekerül a modellbe, amelynek nincs tényleges hatása a kiindulási régió egy főre jutó GDP-jének alakulására. Az új elemek mintegy „elrontják” az eredeti modellt. Tehát minél több területegységet bevonva egyre jobban háttérbe szorulnak a lokális sajátosságok és erősödik a globális jelleg. Másrészt minél több szomszédot veszünk alapul az egyes egyenletekhez, annál homogénebben alakul a kép. Már csupán a 20 szomszédot figyelembe vevő számítás R^2 értékeit nézve azt állapíthatjuk meg, hogy az ír, a skót, a nyugatromán, valamint a délgörög régiók esetében a többitől eltérően alacsonyabb a modellek magyarázóereje. Ez elsősorban abból adódhat, hogy az említett területegységek a romániai régiók kivételével periferikusan helyezkednek el, így nem a ténylegesen velük kapcsolatban levő szomszédok alapján állítottam fel a regressziós egyenleteket. Az R^2 értékeket látva fontos megjegyezni, hogy a változók közötti összefüggések nem ugyanannyira igazak az összes régióra, némelyek esetén közelebb van a kapcsolat az egész európai trendhez, míg máshol bizonytalanabb az illeszkedés.

A globális regressziós modell felállításakor a standardizált bétaértékek segítségével az egyes független változók függő változóra való hatásának erősségét szokták meghatározni, tehát hogy melyik magyarázó változó van a legnagyobb hatással az eredményváltozóra. A GWR-modelleknél is alkalmazható ez a közelítés, azzal a különbséggel, hogy mivel itt annyi regressziós egyenlet áll rendelkezésünkre, ahány területegység van, úgy tudjuk megmondani, hogy melyiknek a legnagyobb a hatása, ha megnézzük, hogy az egyes egyenletekben melyik mutató szerepe a legerősebb, és ezt összesítjük. A jelen vizsgálat esetében megnéztem ezt (a 20 szomszédot figyelembe vevő GWR-t), és azt tapasztaltam, hogy a globális modell alapján a várttól eltérően (1. táblázat), a lokális egyenletekben nem a tercier szektorban foglalkoztatottak aránya, hanem a gazdaságilag aktívak aránya befolyásolja legnagyobb mértékben az egy főre jutó jövedelem értékét. A különbség oka az lehet, hogy az OLS-becslés a hibastruktúra függetlenségének hiányában torzított becslésekhez vezethetett (a Moran I értéke 0,380).

A globális regresszió (OLS) és a GWR eredményeit látva azt állapíthatjuk meg, hogy a lokális kapcsolatok esetenként jelentősen különböznek a globálistól. A későbbiekben az is nyilvánvalóan látszani fog, hogy míg globális szinten például csak pozitív egy független változó hatása az egy főre jutó GDP alakulására, addig lokálisan ez lehet pozitív vagy negatív is, tehát eltérő lehet a lokális előjele a globálistól (Shaoming–Huaqun 2010).

1. tábla

A többváltozós globális regresszió eredménytáblája (OLS)

Megnevezés	Standardizálatlan együtthatók		Standardizált együttható, Beta	t	Sig.
	B	standard hiba			
Konstans (<i>Constant</i>)	-17838,8	1824,1		-9,8	0,0
A tercier szektorban foglalkoztatottak aránya	383,3	16,7	0,536	23,0	0,0
Gazdaságilag aktívak aránya	304,1	26,2	0,264	11,6	0,0
Munkanélküliségi ráta	-408,5	40,2	-0,230	-10,2	0,0
Népsűrűség	2,0	0,2	0,237	10,1	0,0

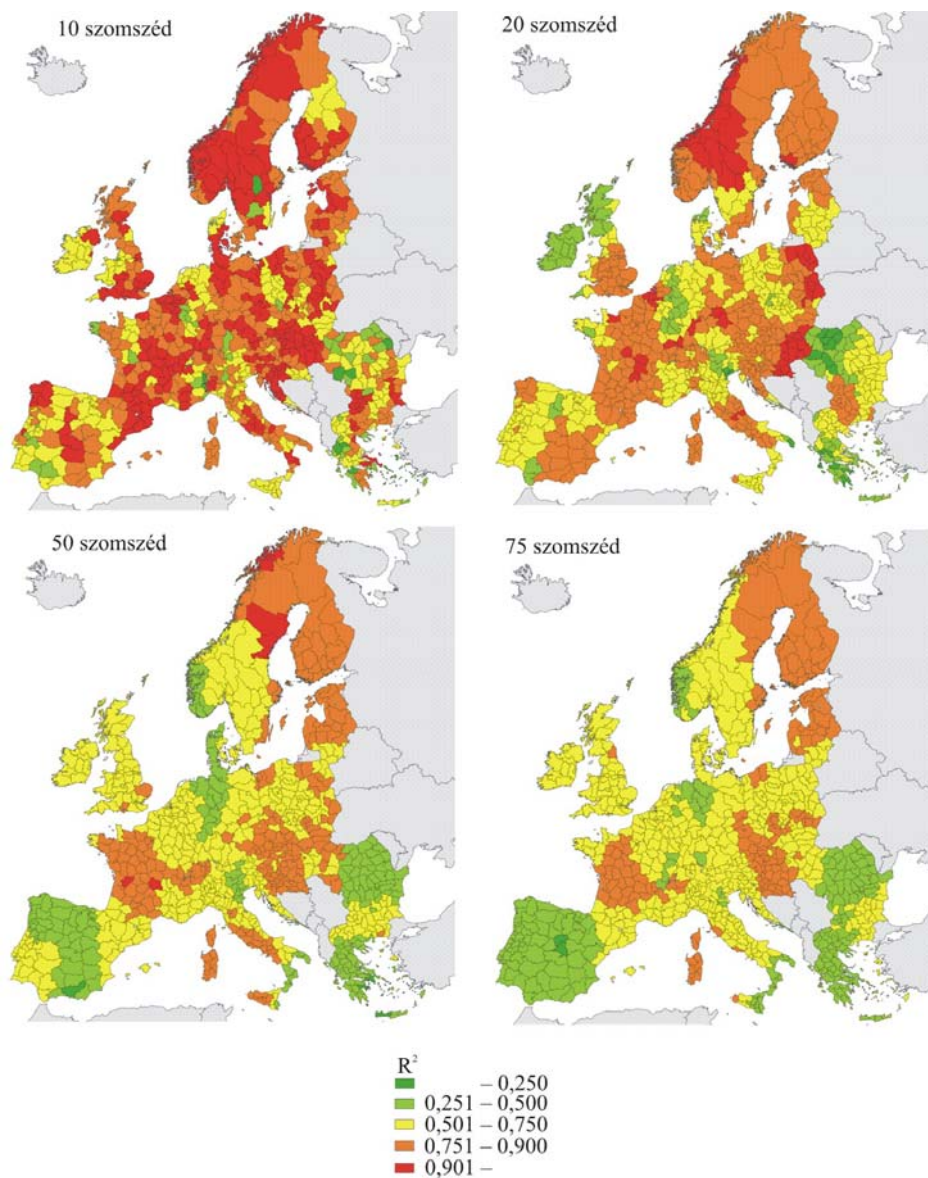
A béta- és standardizált bétaértékeket letérképezve, az egyes magyarázóváltozók függő változóra vonatkozó hatását tudjuk megvizsgálni (7. ábra, a nagymértékű hasonlóság miatt elegendőnek tartom csak a bétaértékek térképeit bemutatni). Tehát a GWR-együtthatók az egyes magyarázóváltozók egy főre jutó GDP-re irányuló befolyásának területi mintázatát mutatják, azzal a különbséggel, hogy a bétaértékek a függő és független változó között fennálló kapcsolat irányát és erősségét is az eredeti mértékegységben, míg a standardizált bétaértékek ennek irányát és erősségét standardizálva, így összehasonlítható módon jelenítik meg (Shaoming–Huaqun 2010).

A különböző standardizált bétaértékeket nézve általánosságban azt mondhatjuk, hogy a magyarázóváltozókat tekintve, a perifériaterületek értékei negatívak, míg a centrumterületek értékei inkább pozitívak. A munkanélküliségi ráta esetében ez ellentétesen jelenik meg. Ennek oka az, hogy a munkanélküliségi ráta változásának jelentéstartalma a többiéhez képest fordított.

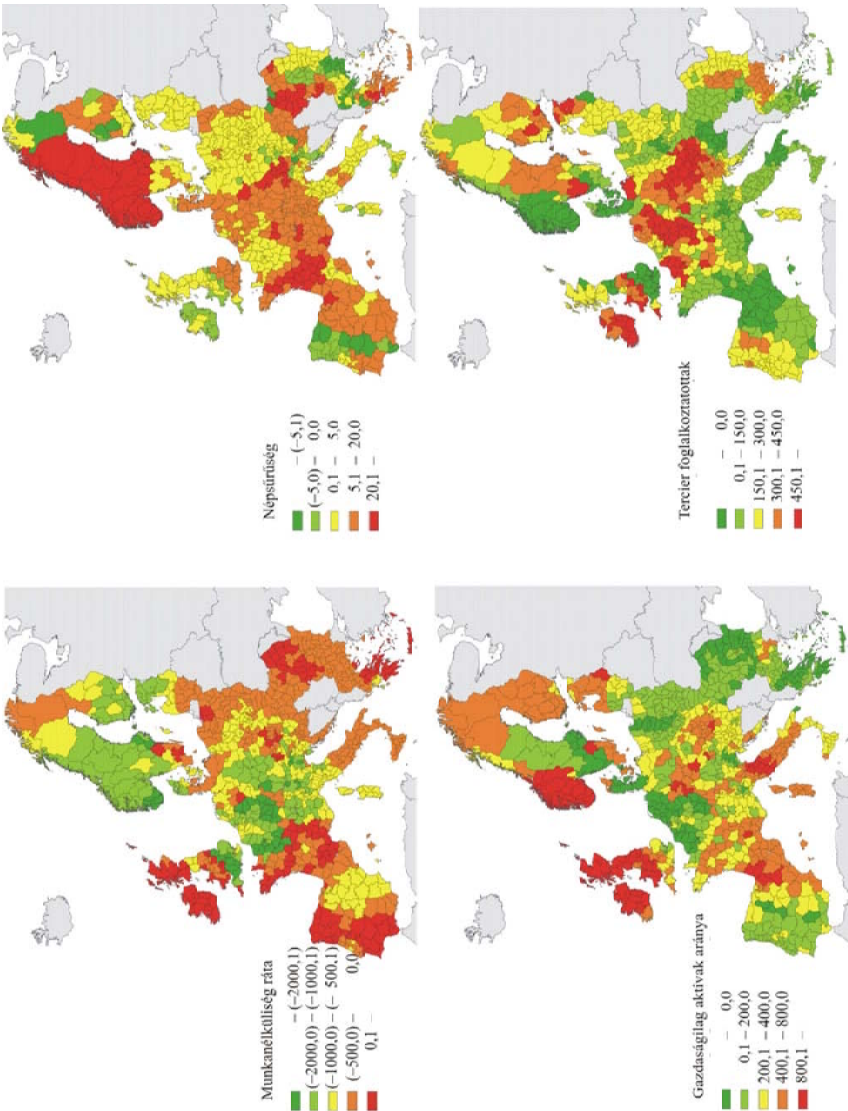
Az egyes regressziós együtthatókat külön nézve a népsűrűség esetében elég homogén képet láthatunk: a legtöbb területegységnél elmondható, hogy a népsűrűség értékének változásával megegyező irányban változik a GDP/fő értéke, míg egyes finn, ír, portugál, bolgár és román régiók esetében a változás ellentétes irányú. A délfancia, a nyugatromán és a svéd régiókban a legerősebb a népsűrűség fejlettséget befolyásoló hatása. A munkanélküliségi ráta bétaértékeit tekintve a periférián levő régiók értéke mutat egyező irányt, míg a központiak ellentétet. Ez alól a skandináv régiók kivételek, ugyanis ott a perifériáktól eltérően inkább ellentétes az irány. A legerősebb befolyásoló hatás Írország, Dél-nyugat-Anglia, Északkelet-Franciaország és Nyugat-Németország esetén látható. A gazdasági aktivitás regressziós együtthatói esetében, a keleti periférikus részekben, illetve a Benelux-államok területén ellentétes irányú, míg a centrumterületeken egyező irányú együttmozgás tapasztalható. Annyit kell kiemelni, hogy az angol és az ír régiók esetében is megegyező irányú együttmozgás látható. A délnorvég, a délnémet és a skót területeken látszik a legjelentősebb, a fejlettséget meghatározó hatás. A tercier foglalkoztatottak arányára vonatkozó bétaértékek az előzőektől eltérően eléggé változatosan alakulnak, ugyanis az ellentétes irányú hatás csak részben jellemző a perifériaterületeken, emellett megfigyelhető Dél-Anglia, Dél-Magyarország és Észak-Spanyolország területén. Dél-Norvégia, Észak-Spanyolország, a Benelux-államok, Csehország és Szlovákia régióira vonatkozóan a legerősebb a fejlettséget befolyásoló hatás.

A regressziós egyenletek által becsült értékeket az eredeti értékekkel összehasonlítva azt láthatjuk, hogy inkább alulbecsültek a modellek, mint felül.

6. ábra

Az adaptív kerneles meghatározott számú szomszédot figyelembe vevő számítás R^2 értékei

7. ábra
A 20 szomszédot figyelembe vevő adaptív kernel GWR-számítás bétaértékei



Összegzés

A GWR tehát egy lokális modell, ugyanis nem a teljes adathalmaz globális összefüggéseit számszerűsíti, hanem lokálisan, meghatározott keresési körzeten belül vizsgálja a függő és a független változók kapcsolatát, így alkalmas térben változó folyamatok figyelembevételére is (Mitchell 2005). Tulajdonképpen a GWR a globális regressziós modell egyfajta javítása, hiszen alkalmazásával figyelembe vehetjük a globális megközelítés által elfedett lokális sajátosságokat. Véleményem szerint használatának legnagyobb eredménye az, hogy az egyes régiókra vonatkozó R^2 értékek megmutatják, hogy a változók közötti összefüggések nem ugyanannyira igazak az összes régióra, némelyek esetén közelebb van a kapcsolat az egész európai trendhez, míg máshol bizonytalanabb az illeszkedés. Ez is megerősíti azt, hogy a globális közelítés jelentősen elfedi/elfedheti a lokális sajátosságokat.

IRODALOM

- Bálint Lajos (2010): A kistérségi életkilátások egy modellje, A GWR alkalmazása hazai példán, kézirat, Pécs, MFT Vándorgyűlése.
- Charlton, M. E. – Fotheringham, A. S. – Brunsdon C. (1997): The geography of relationships: an investigation of spatial non-stationarity. In: Bocquet-Appel, J-P. – Courgeau, D. – Pumain D. (szerk.): *Spatial analysis of biodemographic data*. pp. 23–47. Montrouge, John Libbey Eurotext.
- Chasco, C. – García, I. – Vicéns, J. (2007): *Modeling spatial variations in household disposable income with Geographically Weighted Regression*. MPRA paper No. 9581, Madrid, Universidad Autónoma de Madrid.
- Dusek Tamás (2004): *A területi elemzések alapjai*, Budapest, ELTE Regionális Földrajzi Tanszék.
- Eckey, H-E. – Kosfeld, R. – Türck, M. (2007): Regional Convergence in Germany: a Geographically Weighted Regression Approach. *Spatial Economic Analysis* 2 (1): 45–64.
- Fotheringham, A. S. – Brunsdon, C. – Charlton, M. E. (1998): Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A* 30 (11): 1905–1927.
- Fotheringham, A. S. – Brunsdon, C. – Charlton, M. E. (2002): *Geographically Weighted Regression, the analysis of spatially varying relationships*. Newcastle, University of Newcastle.
- Fotheringham, A. S. – Charlton, M. E. – Brunsdon, C. (1996): The geography of parameter space: an investigation into spatial nonstationarity. *International Journal of GIS*, 10 (5): 605–627.
- LeSage, J. P. (1999): *The Theory and Practice of Spatial Econometrics*. Toledo, Department of Economics, University of Toledo.
- Lin, J. – Cromley, R. – Zhang, C. (2011): Using geographically weighted regression to solve the areal interpolation problem, *Annals of GIS* 17 (1): 1–14.
- Lloyd, C. D. (2007): *Local Models for Spatial Analysis*, London, CRC Press, Taylor & Francis Group.
- Mitchell, A. (2005): *The ESRI guide to GIS analysis*. 2., Spatial measurements and statistics. Redlands, CA: ESRI Press.
- Ridefelt, H. – Etzelmüller, B. – Boelhouwers, J. – Jonasson, C. (2011): Statistic-empirical modelling of mountain permafrost distribution in the Abisko region, sub-Arctic northern Sweden, *Norwegian Journal of Geography* 62 (4): 278–289.
- Shaoming, C. – Huaqun, L. (2010): Spatially Varying Relationships of New Firm Formation in the United States. *Regional Studies* 45 (6): 1–17.
- Xiaomin, Q. – Shuo-sheng, W. (2011): Global and Local Regression Analysis of Factors of American College Test (ACT) Score for Public High Schools in the State of Missouri. *Annals of the Association of American Geographers* 101 (1): 63–83.
- Yu, D. (2005): *GIS and Spatial Modeling in Regional Development Studies: Case Study on the Greater Beijing Area*. PhD dissertation. Milwaukee, Department of Geography University of Wisconsin. <http://www4.uwm.edu/gis/research/upload/danlinyucpetition04.doc> (letöltve: 2012. 12.).

Wheeler, D. C. – Páez, A. (2010): Geographically Weighted Regression. In: Fischer, M. M. – Getis, A. (szerk.): *Handbook of Applied Spatial Analysis*. Berlin, Software Tools, Methods and Applications. Springer-Verlag.

Kulcsszavak: GWR modell, lokális regresszió.

Resume

My research is concerned with a statistical method that has recently become widespread in the international literature, although in Hungary it is present only in few researches. This method is geographically weighted regression (GWR), the use of which is demonstrated through an example of its application. GWR is a local model which is founded on regressional basis, prominently taking into consideration the geographical distance. Since it does not calculate the global relations of the whole data but concentrates on the relationship of the dependent and independent variables locally within a determined search area it makes possible consideration of the spatially varying processes. Basically, the GWR is a kind of developed version of the global regression model since with the use of it we may take into account the local features remaining hidden by the global approach.