

EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFOTMATIKY

Evidenčné číslo: 103003/I/2019/36086129772356612

PRIESTOROVÁ ANALÝZA DÁT V SOFTVÉRI R

Diplomová práca

EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFOTMATIKY

PRIESTOROVÁ ANALÝZA DÁT V SOFTVÉRI R

Diplomová práca

Študijný program: Operačný výskum a ekonometria

Študijný odbor: Kvantitatívne metódy v ekonómii

Školiace pracovisko: Katedra operačného výskumu a ekonometrie

Vedúci záverečnej práce: doc. Ing. Andrea Furková, PhD.

Bratislava 2019

Bc. Emília Šubová



Ekonomická univerzita v Bratislave
Fakulta hospodárskej informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Emília Šubová
Študijný program: operačný výskum a ekonometria (Jednoodborové štúdium, inžiniersky II. st., denná forma)
Študijný odbor: 3.3.24 Kvantitatívne metódy v ekonómii
Typ záverečnej práce: Inžinierska záverečná práca
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický
Názov: Priestorová analýza dát v softvéri R
Anotácia: Regióny, štáty, či iné hospodárske jednotky existujú a fungujú v spoločnom hospodárskom priestore a teda sa často vzájomne ovplyvňujú. Zohľadnenie tejto skutočnosti, t.j. zohľadnenie priestorovej interakcie medzi sledovanými geografickými jednotkami je predmetom priestorovej analýzy dát. Práca je zameraná na vybrané metódy priestorovej analýzy dát v softvéri R.

Vedúci: doc. Ing. Andrea Furková, PhD.
Katedra: KOVE FHI - Katedra operačného výskumu a ekonometrie FHI
Vedúci katedry: prof. Mgr. Juraj Pekár, PhD.

Dátum zadania: 13.10.2017

Dátum schválenia: 05.11.2017

prof. Mgr. Juraj Pekár, PhD.
vedúci katedry

Čestné vyhlásenie

Čestne vyhlasujem, že diplomovú prácu som vypracovala samostatne a že som uviedla všetku použitú literatúru.

Dátum:

.....
Podpis študenta

Pod'akovanie

Chcela by som sa pod'akovať svojej školiteľke doc. Ing. Andrei Furkovej, PhD. za odbornú pomoc, usmerňovanie pri písaní mojej práce, za cenné rady, informácie a v neposlednom rade za ochotu.

ABSTRAKT

ŠUBOVÁ, Emília: Priestorová analýza dát v softvéri R. – Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky, Katedra operačného výskumu a ekonometrie. – Vedúci záverečnej práce: doc. Ing. Andrea Furková, PhD. – Bratislava: FHI EU, 2019, 65 s.

Zámerom prezentovanej diplomovej práce je predstaviť problematiku priestorovej závislosti, históriu a vývoj priestorovej ekonometrie, metodológiu na tvorbu modelov priestorovej ekonometrie a jej následnú aplikáciu pri modelovaní príjmovej konvergenzie členských štátov Európskej únie. V práci uvádzame metodologické aspekty priestorovej ekonometrie, ktoré súvisia aj s konštrukciou matice priestorových váh, testovaním priestorovej autokorelácie, špecifikáciou a odhadom modelov a problematikou heterogenity. Prácu sme rozdelili do štyroch kapitol. Prvú kapitolu sme venovali problematike súčasného stavu, popísaná je história, vývoj a aktuálny stav. Druhá kapitola podrobne opisuje cieľ práce. V tretej kapitole sme sa zaoberali teoretickou stránkou problematiky a metodologickými aspektami popisujúcimi formulovanie a odhad modelov s priestorovo autoregresným procesom. Štvrtá a zároveň záverečná kapitola obsahuje aplikačnú časť, kde sme testovali hypotézy priestorovej príjmovej beta konvergenzie členských štátov Európskej únie s použitím modelov a metodologických aspektov z predošlých častí.

Kľúčové slová: priestorová ekonometria, priestorová autokorelácia, odhad modelov, príjmová beta konvergenzia

ABSTRACT

ŠUBOVÁ, Emília: Spatial analysis of data in software R. - University of Economics in Bratislava. Faculty of Economic Informatics, Department of Operations Research and Econometrics. – Thesis supervisor: doc. Ing. Andrea Furkova, PhD. - Bratislava: FHI EU, 2019, 65 p.

The aim of the presented thesis is to introduce the issue of spatial dependence, history and development of spatial econometrics, methodology for creating models of spatial econometrics and its subsequent application in modeling income convergence of the European Union countries. We present methodological aspects of spatial econometrics, which are also related to the construction of the matrix of spatial weights, testing of spatial autocorrelation, specification and estimation of models and problems of heterogeneity. The work is divided into four chapters. The first chapter is devoted to the issue of the current state, there is describing history, development and the actual state. The second chapter describes the aim of the work in detail. In the third chapter we deal with the theoretical aspect of the issue and methodological aspects describing the formulation and estimation of models with spatially autoregressive process. The fourth and final chapter contains the application part, where we tested the hypotheses of spatial income beta convergence of the European Union countries using models and methodological aspects of previous parts.

Keywords: spatial econometrics, spatial autocorrelation, model estimation, income beta convergence

Obsah

Úvod.....	10
1. Súčasný stav	12
1.1 História vzniku priestorovej závislosti a ekonometrie	12
1.2 Definovanie priestorovej závislosti.....	14
1.3 Vývoj koncepcie priestorovej závislosti a ekonometrie.....	15
1.4 Ekonomické modely rastu a konvergencia	16
1.4.1 Neoklasický model rastu	17
1.4.2 Rozšírený model neoklasického rastu.....	18
1.4.3 Priestorový model neoklasického rastu.....	19
1.4.4 Koncepty konvergenzie	21
1.4.4.1 σ -konvergencia	21
1.4.4.2 β -konvergencia	22
2. Cieľ práce	25
3. Metodika a modely priestorovej ekonometrie	26
3.1 Priestorová závislosť a priestorová ekonometria	26
3.2 Priestorové váhy.....	28
3.2.1 Váhy na základe vzdialenosti	29
3.2.2 Váhy založené na hraniciach.....	29
3.2.3 Kombinované váhy	30
3.2.4 Normalizácia priestorových váh.....	31
3.3 Priestorová autokorelácia.....	33
3.3.1 Moranov I test	33
3.3.2 Gearyho c	34
3.4 Priestorové ekonometrické modely.....	35
3.4.1 Výber vhodného modelu	37
3.4.2 Lineárny regresný model.....	39
3.4.3 Model SAR.....	39
3.4.4 Model SEM.....	40
3.5 Priestorová heterogenita.....	42
4. Výsledky práce a diskusia	43
4.1 Predstavenie softvéru R.....	43
4.1.1 Balíky programu R pre analýzu priestorových údajov.....	43
4.1.2 Čítanie a mapovanie priestorových údajov v R.....	44

4.1.3 Priestorová ekonometria v R	49
4.1.3.1 Model OLS.....	49
4.1.3.2 Modelovanie priestorovej závislosti	50
4.1.3.3 Moranov I test	53
4.1.3.4 Lagrangeov test multiplikátora	53
4.1.4.5 Model SAR.....	54
4.1.4.6 Marginálny efekt.....	57
4.1.4.7 SEM model	58
Záver	62
Zoznam použitej literatúry	64

Úvod

V dnešnom rýchlo sa rozvíjajúcom svete je veľmi dôležité zaoberať sa meraním výkonnosti ekonomiky. Každý rozvoj je ovplyvnený istými parametrami, ktoré môžeme považovať za významné alebo za menej významné. Všetky tieto parametre sa musia dať kvantifikovať, v opačnom prípade nevieme vylúčiť významnosť parametrov. Naším cieľom je zamerať sa na významné miery, ktoré vieme kvantifikovať, a tak určiť závery bez domnienok a s presnými dôkazmi. Vývoj ekonomík vieme merať prostredníctvom hrubého domáceho produktu vyprodukovaného v danej krajine, ktorý odráža výkon danej ekonomiky. Je to presne kvantifikovaná miera, s ktorou ďalej vieme pracovať a použiť ju na rôzne výpočty ukazovateľov. Dôležité je, aby všetky ukazovatele, na ktoré je použitý hrubý domáci produkt, mali ekonomickú výpovednú hodnotu.

Ako je známe prostredníctvom hrubého domáceho produktu sa dá určiť rast ekonomiky, za predpokladu, že máme k dispozícii historické dáta. Takže takzvaný rast ekonomiky je pre našu prácu veľmi dôležitý, pretože v našom záujme je zistiť a skúmať správanie sa ekonomík vzhľadom na ich priestorové umiestnenie a potvrdenie známej hypotézy o príjmovej beta konvergencii, ktorá v skratke hovorí o „*catch-up*“ efekte. Znenie hypotézy je veľmi jednoduché, jej predpokladom je, že príjem chudobnej ekonomiky na obyvateľa má tendenciu rásť a posúvať sa rýchlejšie ako rozvinuté ekonomiky. V dôsledku toho by sa mali nakoniec všetky ekonomiky zbližiť z hľadiska príjmu na obyvateľa. Jedným z predpokladov prečo tento jav nastáva je to, že rozvojová krajina má potenciál rýchlejšie rásť ako rozvinutá krajina, a to z dôvodu znižujúcich sa výnosov (najmä na kapitál), ktoré nie sú také silné ako v krajine bohatej na kapitál. Okrem toho chudobnejšia krajina môže reprodukovať výrobné postupy, technológie a inštitúcie rozvinutej krajiny. Ako bolo spomenuté, zaoberať sa budeme aj priestorovým umiestnením ekonomík, čo zahŕňa vplyv ekonomiky vzhľadom na priestorovú závislosť. Ide o analýzu dát, ktorú budeme skúmať z geodetického hľadiska, na čo budeme potrebovať dáta definované zemepisnou šírkou a dĺžkou. Pomocou nich vieme určiť maticu susednosti a vzájomný vplyv pozorovaní, a tak preskúmať problematiku príjmovej beta konvergenzie vzhľadom na priestorových susedov.

Vybrané ekonomiky sú v rámci Európskej únie, teda dvadsaťosem členských štátov Európskej únie, keďže je nutné vybrať ekonomiky, ktoré majú aspoň jednu spoločnú charakteristiku. V tomto prípade, sú štáty členmi Európskej únie, a tým spĺňajú túto podmienku. Danú problematiku budeme skúmať v softvéri R, ktorý je voľne šíriteľný a okrem štatistických výpočtov vieme v softvéri zobrazit' všetko graficky. Tento si ďalej

predstavíme podrobnejšie v nasledujúcich kapitolách, rovnako ako problematiku priestorovej ekonometrie a metodológiu pre tvorbu a odhad modelov.

1. Súčasný stav

V tejto kapitole sa budeme snažiť priblížiť problematiku priestorovej analýzy dát v ekonometrii. Najskôr predstavíme históriu vzniku a skúmania priestorovej analýzy dát, teda zistíme, kam až siahajú korene pojmu priestorová závislosť. Definujeme podstatu a využitie priestorovej závislosti a uvedieme do povedomia, kde všade je využiteľná táto oblasť.

1.1 História vzniku priestorovej závislosti a ekonometrie

Priestorová autokorelácia je koncept, ktorý pomáha definovať oblasť priestorovej analýzy. Zameriava sa na štúdie využívajúce priestorové štatistiky a priestorovú ekonometriu. V tejto práci budeme sledovať vývoj konceptu a posnažíme sa vysvetliť akademické súvislosti, ktoré tento koncept priniesli do popredia koncom 60-tych rokov. V geografii je zdôraznený význam práce Michaela F. Daceyho, Andrewa D. Cliffa a J. Keitha Orda. Neskôr, s vydaním zväzku *Priestorová ekonometria* (Luc Anselin, 1988), priestorový výskum a použitie konceptu priestorovej autokorelácie získali značný nárast. Tento vývoj sme načrtli spolu s pripomienkami k najnovším a možným budúcim trendom výskumu založeného na priestorovej autokorelácii.

Do roku 1968 sa priestorová autokorelácia nazývala „priestorová závislosť“, „priestorová interakcia“, „priestorová vzájomná závislosť“. V päťdesiatych rokoch minulého storočia prišla motivujúca sila, ktorá oživila geografickú disciplínu, ktorá bola zameraná na blízke susedstvo medzi všetkými druhmi ľudských javov. Sociálny fyzik, John Q. Stewart a jeho kolega a prívrženec na Princetone, geograf William Warntz, urobili veľa preto, aby vysvetlili hypotézu o blízkosti. Po 200 rokoch, od začiatku roku 1800, učitelia identifikovali efekt úpadku vzdialenosti, ktorý George Kingsley Zipf, v roku 1949 spomenul v knihe *Úvod do ekológie človeka*.

V roku 1966 Jean H. P. Paelinck diskutoval o novej oblasti, ktorú v roku 1974 pomenoval priestorová ekonometria (Paelinck a Klaassen 1979). Oblasť priestorovej ekonometrie bola spolu s rôznymi testami priestorovej autokorelácie spomínaná, avšak v danom období nebola známa a často skloňovaná medzi výskumníkmi. Do popredia sa v tom čase dostávala užitočnosť testov typu Moran, ktoré načrtli Cliff a Ord. Paelinck a Klaassen pre zmenu opísali testy priestorových parametrov, ktoré stelesňujú priestorové autokorelačné efekty. Tieto parametre sú základom pre vytvorenie priestorových

autoregresívnych modelov a súčasných základov priestorovej ekonometrie. Pojem priestorová autokorelácia sa však dostala do lexikónu fráz používaných priestorovými analytikmi až v regionálnom vedeckom konferenčnom dokumente Cliffa a Orda z roku 1968. Cliff a Ord (1973, 1981) si uvedomili, tak ako iní pred nimi, že osobitný prípad korelácie vo vesmíre, teda keď existuje vzťah medzi blízkymi priestorovými jednotkami tej istej premennej, by bolo potrebné identifikovať, ak by sa odpovede na výskumné otázky mali detailne zodpovedať. Konferenčný dokument z roku 1968 bol publikovaný v roku 1969 pod názvom „*Problém priestorovej autokorelácie*“. Uviedli pojem *nešpecifikovaná problémová inšpekčná analýza*, ktorý hovorí, že modely, ktoré vyžadovali tradičnú štatistiku pre ich hodnotenie, boli nešpecifikované, ak nebrali v úvahu priestorovú autokoreláciu.

Ďalší z významných prispievateľov bol Luc Anselin, ktorý študoval ekonómiu, ekonometriu, urbanistické a regionálne plánovanie v Belgicku a ktorý bol ovplyvnený spismi Tinbergena a Theila. Prišiel do Spojených štátov v roku 1977, aby vykonal M.A. prácu v regionálnej vede na Cornelli. Jeho dizertačná a výskumná práca na Cornelli na začiatku 80. rokov, za prítomnosti Waltera Isarda a jeho regionálnych vedeckých kolegov, inšpirovala Anselina k príprave knihy „*Priestorová ekonometria: Metódy a modely*“ (1988). Anselin tak priniesol rôzne štúdie priestorových analytikov o skúmaní priestorových ekonometrických modelov. Kniha sa stala základným textom o priestorovej ekonometrii, čiže oblasti, ktorá sa sústreďuje na koncepciu priestorovej autokorelácie. Anselinove neskoršie príspevky do oblasti priestorovej ekonometrie a jeho práce na sofistikovanom softvéri mali obrovský vplyv na teoretických vedcov a odborníkov, ktorí majú záujem o lepšie pochopenie odlišností konceptu. Základným cieľom knihy je vytvoriť komplexný prístup k pochopeniu priestorových efektov obsiahnutých v druhoch ekonometrických modelov, ktoré na ich odhad vyžadujú georeferencované údaje. Anselin vyvinul typológiu priestorových autoregresívnych modelov a potom sa zaoberal požiadavkou, aby boli správne odhadnuté a testované. V procese vysvetľovania charakteristík priestorových ekonometrických modelov skúmal subjekty priestorovej závislosti a priestorovej heterogenity. Vo svojej diskusii načrtáva jasné charakteristiky, ktoré priestorová autokorelácia prináša do priestorového modelovania a zdôraznil dôležitosť matice priestorových váh a rôznych štatistík priestorovej autokorelácie. V knihe Anselin podrobne vysvetľuje ekonometrický pohľad na priestorové efekty.

1.2 Definovanie priestorovej závislosti

Koncepcia priestorovej závislosti, hoci ju možno vnímať ako špeciálnu koncepciu, má svoj vlastný význam. Zatiaľ čo korelačné štatistiky boli rozdelené tak, aby zobrazovali vzťahy medzi premennými, priestorové závislosti ukazujú koreláciu medzi premennými v georeferencovanom priestore. Z niekoľkých definícií priestorovej autokorelácie v literatúre, ktoré Hubert a kol. (1981) uverejnili je možno najpresnejšia práve táto: "Vzhľadom na množinu, ktorá obsahuje N geografických jednotiek, priestorová autokorelácia odkazuje na vzťah medzi niektorými premennými pozorovanými v každej z týchto situácií. Miera geografickej blízkosti je definovaná pre všetky $N(N - 1)$ páry vybrané z N ."

Literatúra na túto tému obsahuje mnoho štatistík, opatrení a parametrov, ktoré výstižne vyjadrujú tento vzťah pre rôzne typy výskumných otázok. Štatistiky boli pôvodne určené na identifikáciu teoretického stavu, v ktorom nie je prítomná žiadna priestorová autokorelácia. V praxi sa však štatistika používa nielen na testovanie hypotéz bez priestorovej autokorelácie, ale aj na meranie stupňa priestorovej autokorelačnej existencie v georeferencovaných údajoch. Pre jednu priestorovo rozdelenú premennú sa tieto štatistiky zvyčajne skladajú z dvoch častí: (1) výraz predstavujúci špecifikovanú, predpokladanú príčinnú súvislosť medzi určenými párami pozorovaní (autokorelácia) a (2) výraz reprezentujúci geometrický (priestorový) vzťah tých istých párov pozorovaní. Čím väčšia je korešpondencia alebo nesúlad medzi oboma časťami, tým väčšia je miera pozitívnej respektíve negatívnej priestorovej autokorelácie. Keď matice predstavujúce dve časti nevykazujú žiadne známky podobnosti, potom nie je dokazovaná žiadna priestorová autokorelácia. Medzi mnohými mierami priestorovej asociácie je *Moranova štatistika I*, ktorá je najrozšírenejším meradlom a testom priestorovej autokorelácie.

Koncept priestorovej autokorelácie zohráva kľúčovú úlohu medzi priestorovými modelármi. Jeho vysvetlenie od začiatku 70. rokov zodpovedá stále rastúcemu počtu a typu modelov, ktoré boli použité na skúmanie javov v mnohých akademických oblastiach. Vzhľadom na to, že využitie georeferencovaných dát sa zvýšilo, tak aj poznanie, že priestorová autokorelácia v jednej alebo druhej forme je základným prvkom všetkých priestorových modelov.

Koncepcia vo svojich mnohých formách poskytuje testy chybného určenia modelu; určuje silu priestorových efektov na ľubovoľnú premennú v modeli; umožňuje testovanie predpokladov priestorovej stacionarity a priestorovej heterogenity; nájde možný závislý vzťah, ktorý môže mať realizácia premennej na iných realizáciách; identifikuje úlohu, ktorú

môže mať distančný rozpad alebo priestorová interakcia na akýkoľvek autoregresný priestorový model; pomáha rozpoznať vplyv, ktorý môže mať geometria priestorových jednotiek, ktoré sú predmetom štúdie, na realizáciu premennej; umožňuje identifikovať silu asociácií medzi realizáciami premennej medzi priestorovými jednotkami; dáva nám prostriedky na testovanie hypotéz o priestorových vzťahoch; dáva nám príležitosť zväžiť dôležitosť časových účinkov; poskytuje zameranie na priestorovú jednotku, aby lepšie porozumel účinku, ktorý by mohol mať na iné jednotky a naopak („lokálna priestorová autokorelácia“); pomáha pri štúdiu odľahlých hodnôt. Žiadna iná koncepcia v empirickom priestorovom výskume nie je taká centrálna pre budovanie modelu, ako je priestorová autokorelácia.

1.3 Vývoj koncepcie priestorovej závislosti a ekonometrie

Predmet priestorovej závislosti sa vyvíjal dlho predtým, ako bol pomenovaný. Štatistická korelácia siaha tak ďaleko ako Sir Francis Galton (bratranec Charlesa Darwina), ktorý bol v roku 1850 zodpovedný za vytvorenie korelačného koeficientu. Matematický rámec pre koreláciu, ktorý dnes používame, poskytol Karl Pearson v 80. rokoch 19. storočia a opäť v 20. rokoch R. A. Fisher. Priestorový zvrät pri korelácii nastal najmä z potreby porovnávania máp a preto, že georeferencované pozorovania nie sú vo všeobecnosti nezávislé. Takéto dobre známe priestorové koncepty ako dištančná a priestorová interakcia majú bohatú históriu, ale nie takú, ktorá môže byť priamo spojená s koncepciou priestorovej autokorelácie. Priestorová autokorelácia je viac prepojená so štatistickou teóriou než s priestorovou teóriou. To neznamená, že tí istí učitelia, ktorí sa zaoberali priestorovou teóriou, si neboli vedomí myšlienky priestorovej korelácie. Práve oni, ktorí sú citliví na zvláštnosti priestorového hľadiska, prijali a rozšírili používanie konceptu tejto teórie.

Sériová autokorelácia je už dlho predmetom štúdia ekonometrie. V skorých 50-tych rokoch Durbin a Watson predstavili svoj test na autokoreláciu časových radov. Je to jednoduchý test, ktorý odpovedá na otázku: „Je pozorovanie na premennej t (kde t je špecifický čas alebo časové obdobie) korelované s pozorovaním $(t-1)$?“ Normálne sa test aplikuje na chyby v modeli časových radov. Je zrejmé, že ide o jednorozmerný test, pohybujúci sa z jedného časového obdobia do druhého. Priestorová autokorelácia sa nesmie považovať za priestorové rozšírenie Durbin-Watsonovho testu. Pre priestorovú autokoreláciu, je potrebné identifikovať koreláciu vo všetkých geografických smeroch, ako protiklad k časovému smeru, toto veľmi komplikuje a robí jeho štúdiu špeciálnou. Rôzne

operácie potrebné na štúdium priestorovej autokorelácie sú však podobné časovej autokorelácii. V obidvoch prípadoch je potrebné identifikovať odchýlky, trendy (časové / priestorové), stupne asociácie, štatistickú významnosť a relevantné modely. Každý z nich je však vypočítaný a chápaný veľmi odlišne a je to priestorová autokorelácia, ktorá je zložitejšia a mnohostrannejšia.

V akademickom svete, nájst' prvú osobu, ktorá objavila a zaoberala sa priestorovými konceptmi je možné hľadať až v renesancii. Niektoré koncepty sa stále používajú a stavia sa na nich a niektoré z nich sú na ceste rozvoja. Aj myšlienky, na ktoré sa na mnoho rokov zabudlo sú pripravené na oživenie, keď sa časom spoločnosť a technológie menia. V prípade priestorových konceptov sa učitelia venujú myšlienke efektov už veľmi dlho.

1.4 Ekonomické modely rastu a konvergencia

Analýza hospodárskeho rastu a konvergenzie je jedným z hlavných cieľov ekonomiky. Preto nie je prekvapujúce, že existuje mnoho modelov rastu a konvergenzie, ktorých štruktúry sa líšia od jednoduchých až po zložité. Slávnym a jednoduchým modelom je rastový model Solowa (1956). Tento prístup vysvetľuje životnú úroveň len s dvoma alebo v rozšírenej verzii troma premennými. Pretože Mankiw a kol. (1992), uverejnili svoj seminárny dokument, rastové modely sa dostávajú aj do zamerania ekonometrie. Autori poskytujú empirické dôkazy o klasickom rastovom modeli Solowa (1956). Okrem toho Mankiw analyzuje rozšírenú verziu modelu pomocou ľudského kapitálu a skúma otázku konvergenzie. V tejto kapitole sú prezentované klasické koncepcie rastu a konvergenzie, ktoré sú dôležité pre nasledujúce časti tejto práce, keďže sa zaoberáme rastom HDP na obyvateľa a vplyvom priestorovej autokorelácie na tento rast.

Solow (1956) navrhuje jeden z najpopulárnejších modelov ekonomického rastu, ktorý vysvetľuje životnú úroveň iba dvomi kovariantmi. Hoci tento jednoduchý model má asi 55 rokov, stále je to téma v súčasnej literatúre. Niekoľko ďalších poskytuje ekonomické rozšírenie modelu a súčasne je jeho empirická platnosť analyzovaná v ekonometrii (napr. Mankiw et al., 1992; Barro, 1991; Barro & Sala-i Martin, 2004). Nasledujúca časť sa teda zaoberá Solowovým modelom, jeho rozšíreniami a empirickým obsahom. V časti 1.4.1 je opísaný Solowov model a jeho odvodenie. Časť 1.4.2 načrtáva rozšírenie Solowovho modelu o ľudský kapitál, ktorý navrhol Mankiw et al. (1992). Nakoniec, v časti 1.4.3 je priestorovo rozšírený Solow prezentovaný model Ertur & Koch (2007), ktorý zachytáva priestorovú závislosť a vedľajšie účinky.

1.4.1 Neoklasický model rastu

Klasický rastový model Solow (1956) môže byť odvodený za predpokladu Cobb-Douglasovej produkčnej funkcie

$$Y = F(K_p, L) = AK_p^\alpha L^{1-\alpha}, \quad 0 < \alpha < 1. \quad (1.1)$$

Celkový výkon Y závisí od faktorov výrobného kapitálu (akumulovaná zásoba kapitálu) K_p a práce (počet obyvateľov v produktívnom veku) L , kde α dáva čiastočnú elasticitu výroby pre kapitál a A je faktor konštantnej mierky merajúci technológie. Pri hodnotení rastu práce sa predpokladá, že rast populácie je exogénny pri konštantnej rýchlosti n , takže L sa môže vyjadriť ako:

$$L(t) = L_0 e^{nt}. \quad (1.2)$$

Ďalej sa predpokladá, že celkový výkon Y možno rozdeliť na hrubé investície a spotrebu. Hrubé investície sa interpretujú ako miera úspor, úspor v tom zmysle, že ide o podiel celkovej úspory na zvyšovaní budúcej produkcie. Čistá investícia vyplýva z rozdielu medzi hrubou investíciou a odpismi

$$\dot{K}_p = s_{k_p} Y - \delta K_p, \quad (1.3)$$

kde δ je odpisová sadzba, ktorá meria podiel na celkovej produkcii, ktorá sa musí investovať na udržanie skutočnej úrovne produkcie. Cieľom Solowovho modelu je vysvetliť životnú úroveň aproximovanú výstupom na obyvateľa

$$y = \frac{Y}{L} = \frac{AK_p^\alpha L^{1-\alpha}}{L^\alpha L^{1-\alpha}} = A \frac{K_p^\alpha}{L^\alpha} = Ak_p^\alpha. \quad (1.4)$$

Produkcia na obyvateľa tak závisí od pomeru kapitálu a pracovnej sily k_p , ktorý meria vzťah medzi nákladmi na kapitál a pracovnú silu výrobných faktorov. Z rovnice (1.2) a (1.3) nasleduje $\dot{K}_p / K_p = s_{k_p} (Y / K_p) - \delta$ respektíve $\dot{L} / L = n$. Tempo rastu pomeru kapitálu a práce je teda dané

$$\frac{\dot{k}_p}{k_p} = s_{k_p} \frac{Y}{K_p} - \delta - n \quad (1.5)$$

a vedie k

$$\dot{k}_p = s_{k_p} y - k_p (\delta + n) \quad (1.6)$$

Z rovnice (1.4) vyplýva, že rovnováha pomeru kapitálu a práce pozitívne súvisí s mierou úspor a negatívne s rastom obyvateľstva v produktívnom veku. Výstup na ustálený stav na obyvateľa možno odvodiť nahradením (1.4) do (1.5),

$$y^* = A \left(\frac{s_{k_p} A}{\delta + n} \right)^{\frac{\alpha}{1-\alpha}} = A^{\frac{\alpha}{1-\alpha}} s_{k_p}^{\frac{\alpha}{1-\alpha}} (\delta + n)^{\frac{-\alpha}{1-\alpha}} \quad (1.7)$$

alebo

$$\log(y^*) = \frac{\log(A)}{1-\alpha} + \frac{\alpha}{1-\alpha} \log(s_{k_p}) - \frac{\alpha}{1-\alpha} \log(\delta + n) \quad (1.8)$$

Pri analýze tohto modelu empiricky odvodíme, že Mankiw zodpovedá logaritmu regresnej formulácie $(y) \simeq X'b$ (X je matica vysvetľujúcich premenných) za predpokladu

$$E(\log(y_i)|X) = \beta_1 + \beta_2 \log(s_{k_p,i}) + \beta_3 \log(\delta + n_i), \quad (1.9)$$

kde $\log(y_i)$ je pozorovaný výstup v ekonomike i , $\beta_1 = \frac{\log(A)}{1-\alpha}$, $\beta_2 = \frac{\alpha}{1-\alpha}$ a $\beta_3 = -\frac{\alpha}{1-\alpha}$.

Pri hodnotení empirickej výkonnosti modelu sa *t-testy* používajú na kontrolu správnych znakov parametrov empirických údajov. Preto prvá kontrola je, či β_2 je významne pozitívny a či je β_3 negatívny. Okrem toho by mali mať obidva parametre približne rovnakú veľkosť a to $\left| \frac{\alpha}{1-\alpha} \right|$. Použitie niekoľkých súborov prierezových údajov a existujúcich empirických dôkazov (napr. Jorgenson et al., 1987) Mankiw naznačuje, že podiel kapitálu na produkcii α je približne konštantný v čase a ekonomiky na úrovni približne jednej tretiny. Predpokladá sa teda, že β_2 a β_3 budú v absolútnej hodnote približne polovičné.

1.4.2 Rozšírený model neoklasického rastu

Rozšírenie Mankiwa spočíva v pridaní ľudského kapitálu do klasického modelu Solowa. Využívanie ľudského kapitálu ako determinant ekonomického rastu je v literatúre skôr bežné (napr. Lucas, 1988). Opäť sa používa a generalizuje Cobb-Douglasova produkčná funkcia

$$Y = AK_p^\alpha K_h^\nu L^{1-\alpha-\nu} \quad 0 < \nu < 1 \quad (1.10)$$

kde K_h označuje nahromadený ľudský kapitál. Analogicky k bodu (1.4) je daná hodnota na obyvateľa

$$y = \frac{Y}{L} = Ak_p^\alpha k_h^\nu \quad (1.11)$$

kde k_h označuje ľudský kapitál na pracovníka a n je podiel ľudského kapitálu na výrobe.

Podobne ako v bode (1.3), čisté investície do ľudského kapitálu vyplývajú z rozdielu medzi hrubými investíciami a odpismi

$$\dot{k}_h = s_{k_h} Y - \delta K_h. \quad (1.12)$$

V upravenom modeli predstavujú dve rôzne miery úspor, napríklad s_{kp} a s_{kh} , hrubú investíciu pre K_p a K_h . Potom, analogicky k (1.6), rast ľudského kapitálu môže byť napísaný ako

$$\dot{k}_h = s_{k_h} y - (\delta + n)k_h. \quad (1.13)$$

Rovnovážna podmienka v tomto modeli je taká, že k_p aj k_h sú konštantné, t.j. $\dot{k}_p = 0$ a $\dot{k}_h = 0$.

Riešenie tohto systému diferenciálnych rovníc pre k_p a k_h je dané

$$k_p^* = \left(\frac{A s_{k_p}^{1-v} s_{k_h}^v}{\delta + n} \right)^{\frac{1}{1-\alpha-v}} \quad (1.14)$$

$$k_h^* = \left(\frac{A s_{k_p}^{1-v} s_{k_h}^v}{\delta + n} \right)^{\frac{1}{1-\alpha-v}} \quad (1.15)$$

Výstup v rovnovážnom stave na obyvateľa sa teda rovná

$$y^* = A^{\frac{1}{1-\alpha-v}} + s_{k_p}^{\frac{\alpha}{1-\alpha-v}} + (\delta + n)^{\frac{\alpha+v}{1-\alpha-v}} + s_{k_h}^{\frac{\alpha}{1-\alpha-v}}, \quad (1.16)$$

alebo

$$\log(y^*) = \frac{\log(A)}{1-\alpha-v} + \frac{\alpha}{1-\alpha-v} \log(s_{k_p}) - \frac{\alpha+v}{1-\alpha-v} \log(\delta + n) + \frac{v}{1-\alpha-v} \log(s_{k_h}). \quad (1.17)$$

Regresná reprezentácia modelu (1.17) je daná logom $(y_i) \simeq X'b$ za predpokladu

$$E(\log(y_i)|X) = \beta_1 + \beta_2 \log(s_{k_p,i}) + \beta_3 \log(\delta + n_i) + \beta_4 \log(s_{k_h,i}). \quad (1.18)$$

Podľa (1.17) sa očakáva, že β_2 a β_4 budú pozitívne, zatiaľ čo β_3 by malo byť negatívne. Veľkosť parametrov samozrejme závisí od empirického ekvivalentu podielu ľudského kapitálu n . Mankiw predpokladá $n = a = 1/3$. β_2 a β_4 by preto mali byť približne rovné jednej a β_3 by mali byť približne rovné mínus dve.

1.4.3 Priestorový model neoklasického rastu

Opäť, na základe Cobb-Douglasovej produkčnej funkcie (1.4) Ertur & Koch (2007) navrhol priestorovo rozšírený model Solowa (1956). Autori predpokladajú, že globálna technologická úroveň A nie je v čase a krajinách konštantná a že medzi krajinami alebo regiónmi sú závislosti. V dôsledku prelievania poznatkov sú tieto vzájomné závislosti

ovplyvnené priestorovým faktorom. Susediace ekonomiky sa navzájom ovplyvňujú viac ako priestorovo vzdialené ekonomiky.

Ertur & Koch (2007) modelujú vzájomné závislosti pomocou tzv. Arrow-Romerových externalít (Arrow, 1962; Romer, 1986). Predpokladá sa úroveň logaritmickej úrovne A

$$\log(A) = \log(\Omega) + \phi \log(k_p) + \gamma \mathbf{W} \log(A). \quad (1.19)$$

A závisí od spoločnej technologickej úrovne \mathbf{W} s konštantným exogénnym rastom, ktorý je dostupný pre všetky ekonomiky. Okrem toho úroveň technológie narastá s kapitálom na pracovníka (prelievanie znalostí), zatiaľ čo ϕ meria veľkosť efektu. Existuje aj priestorový efekt: predpokladá sa, že A závisí od geometricky váženého priemeru technologických zdrojov susedov. Váhy sú uvedené v nesingulárnej váhovej matici \mathbf{W} a γ označuje mieru závislosti od celosvetovej úrovne technológie. Riešenie rovnice (1.19)

$$\log(A) = (I - \gamma \mathbf{W})^{-1} \log(\Omega) + \phi (I - \gamma \mathbf{W})^{-1} \log(k_p). \quad (1.20)$$

Pre vytvorenie rovnovážneho stavu na obyvateľa vytvoríme protokoly (1.4)

$$\log(y) = \log(A) + \alpha \log(k_p) \quad (1.21)$$

a nahradiť $\log(A)$ rovnicou (1.20) takou

$$\log(y) = (I - \gamma \mathbf{W})^{-1} \log(\Omega) + \phi (I - \gamma \mathbf{W})^{-1} \log(k_p) + \alpha \log(k_p) \quad (1.22)$$

$$\log(y) = \log(\Omega) + (\alpha + \phi) k_p - \gamma \alpha \mathbf{W} \log(k_p) + \gamma \mathbf{W} \log(y). \quad (1.23)$$

Analogicky k neoklasickému modelu rastu je možné preukázať, že rovnovážny príjem na obyvateľa je daný

$$\begin{aligned} \log(y^*) = & \frac{1}{1 - \alpha - \phi} \log(\Omega) + \frac{\alpha + \phi}{1 - \alpha - \phi} \log(s_{k_p}) - \\ & \frac{\alpha + \phi}{1 - \alpha - \phi} \log(\delta + n) - \frac{\alpha \gamma}{1 - \alpha - \phi} \sum_{j \neq i} w_{i,j} \log(s_{k_p,j}) + \\ & \frac{\alpha \gamma}{1 - \alpha - \phi} \sum_{j \neq i} w_{i,j} \log(\delta + n_j) + \frac{\gamma(1 - \alpha)}{1 - \alpha - \phi} \sum_{j \neq i} w_{i,j} \log(y_j), \end{aligned} \quad (1.24)$$

kde $w_{i,j}$ je prvok váhovej matice dávajúcej vplyv suseda j na ekonomiku i .

Empiricky sa tento rovnovážny stav odhaduje s regresným modelom za predpokladu, že je to možné

$$\begin{aligned} E(\log(y_i)|X) = & \beta_0 + \beta_1 \log(s_{k_p,i}) - \beta_2 \log(\delta + n_i) + \theta_1 \sum_{j \neq i} w_{i,j} \log(s_{k_p,j}) \\ & + \theta_2 \sum_{j \neq i} w_{i,j} \log(\delta + n_j) + \rho \sum_{j \neq i} w_{i,j} \log(y_j). \end{aligned} \quad (1.25)$$

Výsledkom je priestorová verzia neoklasického rastového modelu s priestorovými oneskoreniami endogénnych a exogénnych premenných. Pozornosť by sa mala venovať skutočnosti, že v tomto prípade X obsahuje endogénnu premennú.

1.4.4 Koncepty konvergenzie

Pojem konvergenzie sa prvýkrát objavil v šesťdesiatych rokoch so spoločnosťou Solowovho modelu neoklasického rastu. V tejto teórii bol rast produkcie determinovaný tromi exogénnymi faktormi: rastom kapitálovej zásoby, rastom obyvateľstva a technologickým pokrokom. Hlavné tvrdenie podľa tejto teórie je, že chudobnejšie regióny sa budú v dlhodobom horizonte približovať k vyspelejším regiónom, pretože oni majú tendenciu rásť rýchlejšie ako bohatšie. (Barro, 1989) Od sedemdesiatych a osemdesiatych rokov minulého storočia bola vyvinutá konvergenčná teória s endogénnym modelom rastu reprezentovaným Lucasom, Romerom, Barrom a Sala-i-Martinom. V tomto modeli zohrávali hlavnú úlohu také endogénne faktory, ako výskum a vývoj, a vzdelávanie. Podľa endogénneho rastu nie je konvergenzia regiónov povinná.

Williamsonova hypotéza v 60-tych rokoch hovorí, že regionálne nerovnosti závisia aj od stavu vývoja. Hospodársky rozvoj spôsobuje v počiatočných fázach regionálne rozdiely a v neskorších fázach vývoja sa objaví regionálna konvergenzia. (Lukovics, 2008) Konceptia klubovej konvergenzie Quah hrá tiež dôležitú úlohu v oblasti konvergenzie. Konvergenčným klubom sa rozumie zoskupenie ekonomík, ktoré sú stabilné a majú podobnosť v niektorých ekonomických alebo geografických ukazovateľoch. (Benedek & Kocziszky, 2014) Quah vyhlásil, že v dlhodobom horizonte 98% chudobných krajín zostáva nadeľ chudobných a 95% bohatých zostáva bohatých; tak vo svetovej ekonomike je možné realizovať len teóriu klubovej konvergenzie.

Proces konvergenzie sa dá merať niekoľkými spôsobmi, ako je sigma, beta, gama alebo vyššie uvedená konvergenzia klubov. V tejto práci sa zameriam na beta konvergenziu.

1.4.4.1 σ -konvergenzia

Takzvaná σ -konvergenzia, v ktorej ide o jednoduché meranie konvergenzie použitím štandardnej odchýlky alebo koeficienta variácie. V štúdiách je často vyjadrená práve koeficientom variácie *HDP na obyvateľa upraveného kúpnu silou*, minimom a maximom *HDP na obyvateľa* v porovnaní s jednoduchým priemerom, ktorý uvádza Ville Kaitila

(2003). Koeficient variácie je počítaný ako štandardná odchýlka delená priemerom. Ak je koeficient variácie záporný, indikuje to konvergenciu. V opačnom prípade, ak je kladný indikuje to divergenciu na *HDP na obyvateľa* v skupine. V duchu konvergenzie sa použijú iba jednoduché priemery, nie vážené, pretože je rovnako neželateľné, aby každá krajina zaostávala napriek veľkosti svojho obyvateľstva.

Analyzovanie konvergenzie prostredníctvom najnižšieho a najvyššieho *HDP na obyvateľa* v porovnaní s priemerom v skupine je tiež dôležitým doplnkom, pretože σ -konvergencia môže ukázať konvergenciu, aj napriek tomu, že jedna krajina je z nejakého dôvodu pozadu. Minimálna hodnota neprehliada túto možnosť. Zmenšujúce sa rozpätie ukazuje, že najchudobnejšie krajiny dobiehajú priemer.

1.4.4.2 β -konvergencia

Ďalšia veľmi známa miera konvergenzie aj β -konvergencia, založená na neoklasickom modeli rastu. Tento model bol predstavený Barro and Sala-i-Martin (1992). Existujú dva druhy β -konvergenzie, a to *absolútna* (nepodmienená) a *podmienená*. Ak predpokladáme, že skúmané krajiny sa zbiehajú do rovnakého koncového bodu alebo bodu ustáleného stavu, hovoríme o konvergencii *absolútnej*. Je to jednoduchá regresná analýza, kde závisle premenná je miera rastu *HDP na obyvateľa* a nezávisle premenná je počiatočná úroveň *HDP na obyvateľa v parite kúpnej sily*.

$$\ln \left(\frac{y_{T,i}}{y_{0,i}} \right) = \alpha + \beta \ln y_{0,i} + \varepsilon_i \quad (1.26)$$

kde

α je konštanta,

β je odhadovaný parameter,

$y_{T,i}$ je regionálne HDP na obyvateľa regiónu i v čase T ,

$y_{0,i}$ je regionálne HDP na obyvateľa regiónu i v čase 0 ,

$\ln \left(\frac{y_{T,i}}{y_{0,i}} \right)$ je rast regionálneho HDP na obyvateľa počas sledovaného obdobia,

ε_i je náhodná zložka.

β koeficient získame bez ďalších premenných, keďže sa predpokladá, že ekonomiky sa výrazne nelíšia v úrovni technológií, investičného pomeru, kvalifikácii ľudského kapitálu,

priemyselnej štruktúry a iných faktorov. Ak majú ekonomiky rozličnú štruktúru, konvergujú k rozličnému konečnému stavu a konvergencia je *podmienená*. Marquezova a Soukiazisova (1998) teória podmienenej konvergenzie je odhadovaná pomocou viacnásobnej regresie s rovnakou závisle premennou, ale zahŕňajúcu rozličné ekonomické, sociálne a politické premenné ako nezávisle premenné, oproti počiatočnej úrovni *HDP na obyvateľa*. V tejto analýze sú zahrnuté ekonomické ukazovatele ako miera inflácie (meraná indexom spotrebiteľských cien), ekonomická otvorenosť, tvorba hrubého fixného kapitálu a sociálno-politické premenné ako miera nezamestnanosti, miera rastu obyvateľstva a vládny dlh. Očakávané znamenie ekonomickej otvorenosti a tvorby hrubého fixného kapitálu je kladné, pričom očakávaná miera inflácie, miera nezamestnanosti, tempo rastu populácie a vládny dlh sú negatívne.

$$\log(y_{i,t}) = \alpha + \beta \log(y_{i,t-1}) + \gamma X_{i,t} + \varepsilon_{i,t} \quad (1.27)$$

pričom $X_{i,t}$ vyjadruje súbor štrukturálnych exogénnych premenných, ktoré môžu ovplyvniť rast *HDP na obyvateľa*.

Koeficient β zachytáva mieru, s akou sa reálny *HDP krajiny na obyvateľa* približuje rýchlosti rastu v rovnovážnom stave, t. j. rýchlosti konvergenzie. Aj keď to môže byť viac ako jedna nezávisle premenná, berie sa v úvahu len koeficient β reálneho *HDP na obyvateľa* a musí byť záporný. Kladná hodnota v tomto prípade indikuje divergenciu. Rýchlosť konvergenzie teda meria, ako rýchlo sa ekonomiky zblížujú smerom k rovnovážnemu stavu a vypočítame ho pomocou nasledujúceho vzorca:

$$s = -\ln(1 + T\beta) / T.$$

kde T je počet období, za ktoré máme údaje o mierach rastu *HDP na obyvateľa* ($T = 10$, pokrývajúce časové rady 2007–2017).

Ďalšia možnosť ako posudzovať rýchlosť konvergenzie je pomocou *polčas procesu* (half-life) konvergenzie. *Polčas* konvergenzie je definovaný ako počet rokov, ktoré sú potrebné na zníženie príjmovej medzery na polovicu. Vypočíta sa ako:

$$h = -\ln(0,5) / s \quad (1.28)$$

Rozlišujeme tri prípady $0 < \beta_I < 1$, $\beta_I = 1$ a $\beta_I > 1$. Pre všetky z nich sa predpokladá β -konvergencia, pretože $\beta_I > 0$.

σ - a β -konvergencia spolu úzko súvisia. β -konvergencia meria rýchlosť akou sa chudobné krajiny vedia priblížiť k bohatým v podmienkach reálneho *HDP na obyvateľa* v stanovenom časovom intervale. Konvergencia σ naznačuje, či sa prierezová variácia

reálneho *HDP na obyvateľa* v skupine krajín časom znižuje. β -konvergenca je nevyhnutnou, ale nie postačujúcou podmienkou konverencie σ .

2. Cieľ práce

V našej práci bude cieľom aplikovať teoretické poznatky priestorovej závislosti na konkrétny príklad. Budeme sa zaoberať jednotlivými ekonometrickými modelmi, vyvinutými pre aplikáciu a zohľadnenie priestorovej závislosti v skúmaných javoch, v našom prípade pôjde o β -konvergenciu medzi štátmi Európskej únie na úrovni NUTS 2. Taktiež sa budeme zaujímať či je nutné a dôležité zohľadniť pri modelovaní priestorovú závislosť alebo bude postačujúce použiť jednoduchý lineárny regresný model, ktorý túto skutočnosť nezohľadňuje. V konečnom dôsledku je taktiež potrebné brať v úvahu aj významnosť parametrov vyskytujúcich sa v modeli, ako aj ich komplexné zhodnotenie.

Ako dáta, na ktorých sme aplikovali teoretické poznatky sme použili hrubý domáci produkt na obyvateľa vyjadrený v parite kúpnej sily a jeho mieru rastu, pričom sme zohľadnili územné celky Európskej únie na úrovni NUTS 2, ako sme spomenuli vyššie. Ukazovatele budeme skúmať z hľadiska konvergenzie regiónov v definovanom priestore. V predchádzajúcej kapitole sme rozvinuli teóriu o konvergencii, kde sme spomenuli, že existuje viacero druhov. My sa však zameriame a budeme zaoberať primárne príjmovou β -konvergenciou.

NUTS je rozdelenie územných celkov Európskej únie do viacerých oblastí, táto metodika bola vytvorená Štatistickým úradom Európskej únie kooperujúcim s národnými inštitúciami pre štatistiku, pre každý štát samostatne. Pre úroveň NUTS 2 sme sa rozhodli hlavne pre rozsiahlosť geografickej oblasti, keďže dôležitou súčasťou nášho výskumu je priestorová analýza a pre lepšiu identifikáciu modelu je vhodné použiť premenné s väčším počtom pozorovaní. Takto definované členenie nám 28 štátov Európskej únie rozdelia na 281 územných celkov, čo je značne výraznejší počet pozorovaní. Pozorovania sme očistili o ostrovné štáty a krajiny, ktoré v pozorovanom období neboli členmi Európskej únie.

Získané dáta budeme ďalej spracovávať a upravovať v štatistickom programe R Studio, ktorý obsahuje balíky na výpočtové a grafické zobrazenie priestorovej ekonometrie, kde vytvoríme program zaoberajúci sa touto problematikou. V programe odhadneme ekonometrické modely a budeme skúmať významnosť parametrov, testovať priestorové efekty a vyberieme model najlepšie popisujúci problematiku. Následne zobrazíme spracované dáta na grafoch a mapách.

3. Metodika a modely priestorovej ekonometrie

Obsahom tretej kapitoly sú teoretické poznatky v oblasti priestorovej ekonometrie, ktoré sú nevyhnutné pre konštrukciu modelov, hypotéz a analýzu výsledkov, s ktorými budeme pracovať v praktickej časti tejto práce.

3.1 Priestorová závislosť a priestorová ekonometria

Aplikovaná práca v regionálnej vede sa vo veľkej miere opiera o údaje o vzorkách, ktoré sa zbierajú s odkazom na miesto merané ako body v priestore. Čo odlišuje priestorovú ekonometriu od tradičnej ekonometrie? Vznikajú dva problémy v prípade, že dáta majú lokálnu zložku, a to priestorovú závislosť medzi pozorovaniami a priestorovú heterogenitu vo vzťahoch, ktoré sa modelujú. Tradičná ekonometria tieto dve otázky do značnej miery ignorovala, možno preto, že porušujú Gauss-Markovove predpoklady použité v regresnom modelovaní. Pokiaľ ide o priestorovú závislosť medzi pozorovaniami, pripomíname, že Gauss-Markovova teória predpokladá, že vysvetľujúce premenné sú stanovené pri opakovanom odbere vzoriek. Priestorová závislosť porušuje tento predpoklad. To vedie k potrebe alternatívnych prístupov odhadovania. Podobne aj priestorová heterogenita porušuje Gauss-Markovov predpoklad, že v pozorovanej vzorke údajov existuje jediný lineárny vzťah s konštantnou odchýlkou. Ak sa vzťah mení, keď sa pohybujeme v rámci vzorky priestorových údajov, alebo zmeny variácií, na úspešné modelovanie tejto variácie a na odvodenie príslušných záverov sú potrebné alternatívne postupy odhadu.

Anselin (1988) poskytuje kompletne riešenie mnohých aspektov priestorovej ekonometrie, z ktorej tiež čerpáme. Okrem diskusie o myšlienkach uvedených v Anseline (1988), táto práca bude obsahovať Bayesovské prístupy, ako aj konvenčné metódy maximálnej vierohodnosti pre všetky priestorové ekonometrické metódy diskutované v práci. Bayesovské metódy sa vyznačujú veľkou výzvou v priestorovej ekonometrii, pretože mnohé z myšlienok používaných v modelovaní regionálnych vied zahŕňajú:

1. rozpad ovplyvnenia dát vzorky so vzdialenosťou
2. podobnosť pozorovaní so susednými pozorovaniami
3. hierarchiu miesta alebo regiónov
4. systematická zmena parametrov s pohybom v priestore

Priestorová závislosť sa vo všeobecnosti zaoberá súbormi dát, ktoré sú založené na pozorovaní prierezových priestorových údajov. Ak uvažujeme prierezový variabilný vektor

reprezentujúci pozorovania s odkazom na body alebo oblasti v priestore. Tieto bodové pozorovania by mohli zahŕňať predajné ceny domov, zamestnanie v rôznych zariadeniach alebo zápis študentov do jednotlivých škôl. Geografické informačné systémy zvyčajne podporujú geokódovanie alebo porovnávanie adries, ktoré umožňujú automatické konvertovanie adries na miestne súradnice. Schopnosť geokódu viedla k obrovskému množstvu priestorovo odkazovaných dát.

Na rozdiel od bodových pozorovaní, pre oblasť sa spoliehame na súradnice vnútorného bodu predstavujúceho stred teda ťažisko. Dôležitým bodom je, že v modeloch priestorovej regresie každé pozorovanie zodpovedá umiestneniu alebo oblasti. Proces generovania dát pre konvenčnú prierezovú vzorku N nezávislých pozorovaní y_i , $i = 1, \dots, N$, ktoré sú lineárne spojené s vysvetľujúcimi premennými v matici X , majú tvar

$$\begin{aligned} y_i &= \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2) \quad i = 1, \dots, N. \end{aligned} \quad (3.1)$$

V $\varepsilon_i \sim N(0, \sigma^2)$ $i = 1, \dots, N$ používame $N(a, b)$ na označenie normálneho rozdelenia s priemerom a a rozptylom b . V rovnici \mathbf{X}_i predstavuje $1 \times k$ vektor vysvetľujúcich premenných, s pridruženými parametrami $\boldsymbol{\beta}$ obsiahnutými v $k \times 1$ vektore. Tento typ procesu generovania údajov sa zvyčajne predpokladá pre lineárne regresné modely. Každé pozorovanie má základný priemer $\mathbf{X}_i \boldsymbol{\beta}$ a náhodnú zložku ε_i . Dôsledkom toho pre situácie, keď pozorovania i predstavujú oblasti alebo body v priestore, je to, že pozorované hodnoty na jednom mieste alebo oblasti sú nezávislé od pozorovaní uskutočnených na iných miestach alebo oblastiach. Štatisticky nezávislé pozorovania naznačujú, že $E(\varepsilon_i \varepsilon_j) = E(\varepsilon_i) E(\varepsilon_j) = 0$, tento predpoklad nezávislosti značne zjednodušuje modely.

Naproti tomu priestorová závislosť odráža situáciu, keď hodnoty pozorované na jednom mieste alebo v oblasti, napríklad pozorovaní i , závisia od hodnôt susedných pozorovaní v blízkych lokalitách. Predpokladajme, že pozorovania $i = 1$ a $j = 2$ predstavujú susedov, možno regióny s hranicami, ktoré sa dotýkajú, potom proces generovania údajov môže mať formu

$$\begin{aligned} y_i &= a_i y_j + \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i \\ y_j &= a_j y_i + \mathbf{X}_j \boldsymbol{\beta} + \varepsilon_j \\ \varepsilon_i &\sim N(0, \sigma^2) \quad i = 1 \\ \varepsilon_j &\sim N(0, \sigma^2) \quad j = 2 \end{aligned} \quad (3.2)$$

Táto situácia naznačuje súčasný proces generovania údajov, kde hodnota y_i závisí od hodnoty y_j a naopak čiže ide o prípad regiónov, kde hodnota závislej premennej v prvom regióne je značne ovplyvnená hodnotami tejto premennej v ďalších, susedných regiónoch.

Intuitívne, keď umožníme vzťah závislostí medzi súborom N pozorovaní / miest, existuje potenciálne $N^2 - N$ vzťahov, ktoré by mohli vzniknúť, odčítame N od potenciálnych vzťahov závislostí N^2 , pretože vylučujeme závislosť pozorovania od seba. Avšak môžu priniesť problém nadmernej parametrizácie. Riešenie tohto problému, ktorý vzniká, keď umožníme, aby každý závislý vzťah mal vzťahovo-špecifické parametre, je uložiť štruktúru na vzťahy priestorovej závislosti. Ord (1975) navrhol parametrizáciu pre vzťahy závislosti medzi pozorovaniami, ktorý staval na ranej práci Whittleho (1954). Táto štruktúra vedie k procesu generovania údajov, ktorý je známy ako priestorový autoregresný proces:

$$y_i = \rho \sum_{j=1}^N w_{ij} y_j + \varepsilon_i \quad (3.3)$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1, \dots, N$$

Termín $\sum_{j=1}^N w_{ij} y_j$ sa nazýva priestorovo oneskorená premenná alebo „spatial lag“, pretože predstavuje lineárnu kombináciu hodnôt oneskorenej závislej premennej y vytvorenej z pozorovaní resp. oblastí, ktoré pozorujú susedov i . To sa dosahuje umiestnením prvkov w_{ij} do $N \times N$ priestorovej váhovej matice \mathbf{W} tak, že $\sum_{j=1}^N w_{ij} y_j$ má za následok skalár, ktorý predstavuje lineárnu kombináciu hodnôt závislých premenných v susedných lokalitách. Tento termín formálne zapíšeme ako

$$[\mathbf{W}y]_i = w_{i1}y_1 + w_{i2}y_2 + \dots + w_{iN}y_N = \sum_{j=1}^N w_{ij}y_j \quad (3.4)$$

Použitie jediného parametra v modeli, priestorového autoregresného parametra ρ reflektuje priemernú úroveň závislostí v priestorovej autoregresnej štruktúre.

3.2 Priestorové váhy

Každá nezáporná matica, $\mathbf{W} = (w_{ij} : i, j = 1, \dots, N)$ je možná priestorová váhová matica, ktorá sumarizuje priestorové vzťahy medzi N priestorovými jednotkami. Tu každá priestorová váha, w_{ij} , zvyčajne odráža „priestorový vplyv“ jednotky j na jednotku i . Podľa štandardnej konvencie tu vylučujeme „vlastný vplyv“ za predpokladu, že $w_{ii} = 0$ pre všetky $i = 1, \dots, N$ (takže \mathbf{W} má nulovú diagonálu).

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \dots & w_{NN} \end{bmatrix}$$

Ďalej uvediem matice váh, ktoré sa často používajú v praxi, a ktoré taktiež budeme aplikovať nasledovne v praktickej časti.

3.2.1 Váhy na základe vzdialenosti

Nasledujúce váhové matice sú založené na *vzdialenosti centroidov* d_{ij} , medzi každou dvojicou priestorových jednotiek i a j .

Váhy na základe k -najbližšieho suseda

Vzdialenosť centroidov od každej priestorovej jednotky i ku všetkým jednotkám j a i sa klasifikuje takto : $d_{ij(1)} \leq d_{ij(2)} \leq \dots \leq d_{ij(n-1)}$. Potom pre každý $k = 1, \dots, N-1$, nastavenie $N_k(i) = \{j(1), j(2), \dots, j(k)\}$ obsahuje k najbližšie jednotky k i (kde pre jednoduchosť ignorujeme väzby). Pre každé dané k , potom má k -najbližšia susedná váhová matica \mathbf{W} , potom priestorové váhy formulované:

$$w_{ij} = \begin{cases} 1, & j \in N_k(i) \\ 0, & \text{ostatné} \end{cases} \quad (3.5)$$

Alternatívne je možné uvažovať o symetrickej verzii, v ktorej sú kladné váhy priradené všetkým ij párom alebo aspoň jedna z nich je medzi k -najbližšími susedmi ostatných:

$$w_{ij} = \begin{cases} 1, & j \in N_k(i) \text{ alebo } i \in N_k(j) \\ 0, & \text{ostatné} \end{cases} \quad (3.6)$$

3.2.2 Váhy založené na hraniciach

Výhodou vyššie uvedených vzdialenostných váh je, že vzdialenosti sa dajú ľahko vypočítať. V mnohých prípadoch však hranica medzi priestorovými jednotkami zohráva dôležitú úlohu pri určovaní stupňa „priestorového vplyvu“.

Váhy priestorových súvislostí

Najjednoduchšie z týchto váh označujú, či priestorové jednotky zdieľajú hranicu alebo nie. Ak je množina hraničných bodov jednotky i označená $bnd(i)$, potom sú takzvané váhové hodnoty „queen“ teda *kráľovnej* definované ako

$$w_{ij} = \begin{cases} 1, & bnd(i) \cap bnd(j) \neq \emptyset \\ 0, & bnd(i) \cap bnd(j) = \emptyset \end{cases} \quad (3.7)$$

To však umožňuje, že priestorové jednotky zdieľajú len jeden hraničný bod, napríklad spoločný rohový bod na mriežke priestorových jednotiek. Silnejšou podmienkou je preto požadovať, aby bola zdieľaná určitá pozitívna časť ich hranice. Ak l_{ij} označuje dĺžku zdieľanej hranice $bnd(i) \cap bnd(j)$, medzi i a j , potom tieto váhové hodnoty nazývame „rook“ teda *veža*.

$$w_{ij} = \begin{cases} 1, & l_{ij} > 0 \\ 0, & l_{ij} = 0 \end{cases} \quad (3.8)$$

Zdieľané hraničné váhy

Ako ostrejšia forma porovnávania, všimnite si, že ak l_i definuje celkovú hraničnú dĺžku $bnd(i)$, ktorá je zdieľaná s inými priestorovými jednotkami, t. j. $\sum_{j \neq i} l_{ij}$ potom je daný zlomok tejto dĺžky zdieľaný s každou konkrétnou jednotkou j s l_{ij} / l_i . Tieto samotné frakcie prinášajú potenciálne relevantný súbor spoločných hraničných hmotností definovaných nasledovne:

$$w_{ij} = \frac{l_{ij}}{l_i} = \frac{l_{ij}}{\sum_{k \neq i} l_{ik}} \quad (3.9)$$

3.2.3 Kombinované váhy

Nakoniec by malo byť zrejmé, že v mnohých situáciách priestorový vplyv môže vykazovať aspekty ako vzdialenostných, tak aj hraničných vzťahov. Jeden klasický príklad je uvedený v pôvodnej štúdii priestorovej autokorelácie Cliffom a Ord (1969). Pri analýze údajov o krvných skupinách Eire zistili, že najlepšia schéma váženia pre zachytávanie priestorových autokorelačných efektov bola daná nasledujúcou kombináciou výkonovej vzdialenosti a hraničných akcií,

$$w_{ij} = \frac{l_{ij}d_{ij}^{-\alpha}}{\sum_{k \neq i} l_{ik}d_{ik}^{-\alpha}} \quad (3.10)$$

S jednoduchou inverznou vzdialenosťou, $\alpha = 1$.

3.2.4 Normalizácia priestorových váh

Vo väčšine prípadov je vhodné normalizovať priestorové váhy, aby sa odstránila závislosť na vonkajších faktoroch stupnice, ako sú konkrétne jednotky použitej vzdialenosti v exponenciálnych a výkonových váhach. Tu sú dva štandardné prístupy:

Riadková normalizácia priestorových váh

Pripomeňme si, že v rade \mathbf{W} sa nachádzajú všetky priestorové váhy ovplyvňujúce priestorovú jednotku i , menovite $(w_{ij} : j \neq i)$ [pripomeňme, že $w_{ij} = 0$]. Ak sú teda (nezáporné) váhy v každom riadku normalizované tak, aby mali jednotkovú sumu, t.j.

$$\sum_{j=1}^N w_{ij} = 1, i = 1, \dots, n \quad (3.11)$$

potom to produkuje to, čo sa nazýva normalizácia riadku \mathbf{W} . Všimnite si, že každá riadkovo normalizovaná váha w_{ij} , môže byť potom interpretovaná ako zlomok všetkých priestorových vplyvov na jednotku i priradenú jednotke j . Odvolanie tejto interpretácie viedlo k súčasnému širokému využívaniu riadkovo normalizovaných váhových matic. Mnohé z vyššie uvedených definícií priestorových váh sa často implicitne definujú ako normalizované riadky. Najzreteľnejším príkladom je podiel spoločných hraničných váh v kombinovaných váhach, ktorý sa podľa definície považuje za normalizovaný riadkom. Ďalší jednoduchý príklad je poskytnutý štandardnou formou váh k -najbližších susedov uvedených vyššie, ktoré sú často definované pomocou váh $1/k$ namiesto 1, aby sa zaistila normalizácia riadkov. Zaujímavejším príkladom sú *váhy výkonových vzdialeností*, ktoré majú riadkovo normalizovaný tvar:

$$w_{ij} = \frac{d_{ij}^{-\alpha}}{\sum_{k \neq j} d_{ik}^{-\alpha}} \quad (3.12)$$

Tieto normalizované váhy sú presne považované za systém inverznej dištančnej váhy (IDW), ktorý sa používa v priestorovom analytickom zariadení na priestorovú interpoláciu. Podobný príklad je poskytnutý *exponenciálnymi váhami*, s normalizovanou formou riadku,

$$w_{ij} = \frac{\exp(-\alpha d_{ij})}{\sum_{k \neq j} \exp(-\alpha d_{ik})} \quad (3.13)$$

Tieto váhy sa tiež používajú pre priestorovú interpoláciu. Okrem toho treba poznamenať, že tieto normalizované váhy sa bežne používajú v modelovaní priestorových interakcií.

Skalárna normalizácia priestorových váh

Napriek svojej popularite majú riadkovo normalizované váhy nevýhody. Normalizácia riadkov mení najmä vnútornú váhovú štruktúru \mathbf{W} tak, že porovnanie medzi radmi sa stáva trochu problematickým. Zvážme napríklad váženie priestorovej súvislosti vzhľadom na jednoduchý príklad troch jednotiek zobrazený dole:

$$i,j,k \quad \mathbf{W} = \begin{pmatrix} 0 & w_{ij} & w_{ik} \\ w_{ji} & 0 & w_{jk} \\ w_{ki} & w_{kj} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Ako je znázornené v matici váh spojitosti \mathbf{W} , jednotka j je ovplyvnená oboma i a k , pričom jednotky i a k sú ovplyvnené iba jednotkou j . Preto by sa dalo argumentovať, že j je vystavený väčšiemu priestorovému vplyvu ako i alebo k . Normalizácia riadkov matice \mathbf{W} však mení tento vzťah, ako je vidieť v riadkovo normalizovanej matici \mathbf{W}_{rn} nižšie:

$$\mathbf{W}_{rn} = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$$

Tu „celkový“ vplyv na každú jednotku je podľa definície rovnaký, takže jednotka i teraz ovplyvňuje j iba „polovične“ ako j ovplyvňuje i . Zatiaľ čo presný význam „vplyvu“ je vo väčšine aplikácií nevyhnutne nejasný, tento účinok normalizácie riadkov sa ťažko dá považovať za neutrálny.

Vzhľadom na toto obmedzenie je prirodzené uvažovať o jednoduchých skalárnych normalizáciách, ktoré vynásobia \mathbf{W} jedným číslom, povedzme $\alpha \cdot \mathbf{W}$, ktorý odstraňuje všetky efekty mernej jednotky, ale zachováva vzťahy medzi všetkými riadkami \mathbf{W} . Napríklad, ak w_{max} označuje najväčší prvok \mathbf{W} , potom voľba,

$$\alpha = \frac{1}{w_{max}} > 0$$

poskytuje jednu takúto normalizáciu, ktorá má tú výhodu, že zabezpečuje, že výsledné priestorové váhy w_{ij} sú všetky medzi 0 a 1, a teda môžu byť stále interpretované ako

intenzity relatívneho vplyvu. Z teoretických dôvodov je však vhodnejšie rozdeliť \mathbf{W} maximálnou vlastnou hodnotou λ_{\max} , z \mathbf{W} , a teda nastaviť

$$\alpha = \frac{1}{\lambda_{\max}} > 0.$$

3.3 Priestorová autokorelácia

V nasledujúcom kroku vysvetlíme pojem autokorelácia a jej prítomnosť v modeli. Testy priestorovej autokorelácie overujú prítomnosť súvislostí medzi polohami regiónov a pozorovanými hodnotami sledovaných premenných. Najpoužívanějšími testami sú Moranov I a Gearyho test, ktoré odhadujú celkovú mieru priestorovej autokorelácie pre dátovú zostavu. Oba testy sú založené na nulovej hypotéze priestorovej nezávislosti.

3.3.1 Moranov I test

V modeli je nutné otestovať prítomnosť autokorelácie, ktorú preveríme pomocou *Moranovho I testu*. Tvorcom autokorelačného testu bol austrálsky štatistik Patrick Alfred Pierce Moran (1998). Priestorová autokorelácia je charakterizovaná koreláciou v signáli medzi blízkymi miestami v priestore. Priestorová autokorelácia je zložitejšia ako jednorozmerná autokorelácia, pretože priestorová korelácia je viacrozmerná (t. j. 2 alebo 3 rozmery priestoru) a viacsmerová. *Moranov I test* je definovaný nasledovne:

$$I = \frac{N}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2} \quad (3.14)$$

kde N je počet priestorových jednotiek indexovaných pomocou i a j ; X je premenná záujmu; \bar{X} je priemerná hodnota skúmanej premennej x ; w_{ij} je matica priestorových váh s nulami na diagonále (t.j. $w_{ii} = 0$) a \mathbf{W} je súčet všetkých w_{ij} . Vzhľadom na riadkovú normalizáciu

matice váh sa vzťah (3.14) z dôvodu $\sum_{i=1}^N \sum_{j=1}^N w_{ij} = N$ redukuje na upravený tvar:

$$I = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2} \quad (3.15)$$

Nulová hypotéza Moranovho testu tvrdí, že nie je prítomná priestorová autokorelácia. Medzi sledovanými jednotkami nie je žiadna priestorová previazanosť. Nemá teda zmysel sa priestorovou analýzou ďalej zaoberať. Za platnosti nulovej hypotézy je očakávaná hodnota *Moranovho I*

$$E(I) = \frac{-1}{N-1}. \quad (3.16)$$

V praxi býva *Moranov index* interpretovaný ako korelačný koeficient, napriek tomu nemusí nadobúdať hodnoty na intervale $\langle -1, 1 \rangle$. *Moranov I test* má asymptotické normované normálne rozdelenie. Dôkaz $I \sim N(0,1)$ by sme mohli nájsť v článku Kelejian a Prucha (2001). Vyššie kladné hodnoty indexu svedčia o pozitívnej priestorovej autokorelácii, jednotka je kladne ovplyvnená svojimi susedmi. Naopak záporné hodnoty vypovedajú o negatívnej priestorovej autokorelácii. Pokiaľ nezamietame nulovú hypotézu, znamená to, že regióny sú priestorovo náhodné. Nemajú na seba žiadny vplyv v dôsledku geografickej polohy.

Testovanie je založené na asymptotickom *z-teste*. V prípade nízkeho počtu pozorovaní nemusí byť tento test spoľahlivý. V článku Cliff a Ord (1972) je navrhnutý simulačný proces Monte Carlo. Generujeme náhodné permutácie za platnosti nulovej hypotézy, kedy pre priestorovú nezávislosť predpokladáme normalitu. Randomizácia prebieha na pozorovaných hodnotách jednotlivých regiónov. Pre každú randomizáciu je počítaný *Moranov I*. Tieto vygenerované hodnoty porovnáme s pozorovaným *Moranovým indexom*. Pokiaľ skutočná hodnota padne medzi najnižších 5% hodnôt alebo medzi najvyšších 95%, tak je významná na hladine $\alpha = 5\%$.

3.3.2 Gearyho c

V podobnom duchu sa nesie aj Gearyho test, kde definujeme index Gearyho *c*:

$$c = \frac{N-1}{2 \sum_{i=1}^N \sum_{j=1}^N w_{ij}} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (X_i - X_j)^2}{\sum_{i=1}^N (X_i - \bar{X})^2}. \quad (3.14)$$

Je zrejmé, že hodnoty Gearyho *c* sú vždy kladné. Za platnosti nulovej hypotézy je očakávaná hodnota Gearyho *c* rovná:

$$E(c) = 1. \quad (3.15)$$

Pokiaľ je $c=1$ nezamietame hypotézu o neprítomnosti priestorovej autokorelácie. Hodnoty na intervale $\langle 0,1 \rangle$ svedčia o pozitívnej závislosti. Čím sú hodnoty bližšie k nule, tým hovoríme o silnejšej autokorelácii.

V prípade Gearyho c sú navrhnuté analogické testy ako pre Moranovo I . Gearyho c je analógiou Durbin-Watsonovho d -testu na $ar(1)$ náhodných zložiek.

3.4 Priestorové ekonometrické modely

V regresnom kontexte sa priestorové efekty vzťahujú na dve kategórie špecifikácií. Jedna sa zaoberá priestorovou závislosťou alebo jej slabšou expresiou, priestorovou autokoreláciou a druhá priestorovou heterogenitou. Heterogenita je jednoducho štrukturálna nestabilita, buď vo forme nekonštantného rozptylu poruchového člena v regresnom modeli (heteroskedasticita) alebo vo forme variabilných regresných koeficientov. Väčšinu metodických otázok týkajúcich sa priestorovej heterogenity možno riešiť pomocou štandardného ekonometrického súboru nástrojov. Formálny rámec používaný na štatistickú analýzu priestorovej autokorelácie je tzv. priestorový stochastický proces často označovaný aj ako priestorové náhodné pole alebo súbor náhodných premenných y , indexovaných podľa polohy i ,

$$\{y_i, i \in D\}, \quad (3.16)$$

kde indexová množina D je buď súvislá plocha alebo konečná množina oddelených miest. Keďže každá náhodná premenná je „označená“ podľa miesta, priestorová autokorelácia môže byť formálne vyjadrená momentovou podmienkou,

$$\text{cov}[y_i, y_j] = E[y_i y_j] - E[y_i] \cdot E[y_j] \neq 0 \quad \text{pre } i \neq j \quad (3.17)$$

kde i, j sa vzťahujú na jednotlivé pozorovania a y_i, y_j je hodnota náhodnej premennej záujmu v danom mieste. Táto kovariancia sa stáva zmysluplnou z priestorového hľadiska, keď konkrétna konfigurácia nenulových párov i, j má interpretáciu z hľadiska priestorovej štruktúry, priestorovej interakcie alebo priestorového usporiadania pozorovaní. Tak by to bolo napríklad keď sa človek zaujíma o modelovanie, do akej miery sa technologické inovácie v kraji šíria do susedných krajín.

Priestorovú kovarianciu možno modelovať tromi základnými spôsobmi. Najprv je možné špecifikovať konkrétnu funkčnú formu pre priestorový stochastický proces generujúci náhodnú premennú, z ktorej by nasledovala kovariančná štruktúra. Po druhé, je možné modelovať štruktúru kovariancie priamo, typicky ako funkciu malého počtu

parametrov s danou štruktúrou kovariancie zodpovedajúcou triede priestorových stochastických procesov. Po tretie, môžeme nechať nešpecifikovanú kovarianciu a odhadnúť ju neparametricky. (LeSage, 1999)

Priestorové modely stochastického procesu

Najčastejšie používaným prístupom k formálnej expresii priestorovej autokorelácie je špecifikácia funkčnej formy pre priestorový stochastický proces, ktorý súvisí s hodnotou náhodnej premennej v danom mieste s jej hodnotou na iných miestach. Štruktúra kovariancie potom vyplýva z povahy procesu. Paralelne s analýzou časových radov sú priestorové stochastické procesy kategorizované ako procesy priestorovej autoregresie (SAR) a priestorového klzavého priemeru (SMA), hoci existuje niekoľko dôležitých rozdielov medzi kontextmi kríženia a časových radov. Napríklad pre vektor $N \times 1$ náhodných premenných je proces simultánneho priestorového autoregresu (SAR) definovaný ako

$$(y - \mu i) = \rho W(y - \mu i) + \varepsilon, \quad \text{alebo} \quad (y - \mu i) = (I - \rho W)^{-1} \varepsilon, \quad (3.18)$$

kde μ je (konštantný) priemer y_i , i je vektor $N \times 1$ vektorov a ρ je parameter priestorového autoregresu. Pred bližším posúdením štruktúry tohto procesu si všimnite prítomnosť $N \times N$ matice W , ktorá sa označuje ako matica priestorových váh. Pre každé umiestnenie v systéme určuje, ktoré z ostatných umiestnení v systéme ovplyvní hodnotu v danom mieste. Je to potrebné, pretože na rozdiel od jednoznačného pojmu „posun“ dlhého času (napríklad $y_t - 1$ v autoregresívnom modeli) v priestorovej doméne neexistuje žiadny zodpovedajúci koncept, najmä ak sa pozorovania nachádzajú nepravidelne v priestore. Namiesto pojmu posun sa používa operátor priestorového oneskorenia, čo je vážený priemer náhodných premenných na „susedných“ miestach.

Priestorové váhy v rozhodujúcej miere závisia od definície susedstva pre každé pozorovanie. Toto sa získa výberom pre každé miesto i (ako riadok) susedov ako stĺpcov, ktoré zodpovedajú nenulovým prvkom w_{ij} v pevnej a kladnej matici priestorových váh $N \times N$ W . Priestorové oneskorenie y pri i potom nasleduje ako

$$[Wy]_i = \sum_{j=1, \dots, N} w_{ij} \cdot y_j, \quad (3.19)$$

alebo v maticovom vyjadrení

$$Wy. \quad (3.20)$$

Vzhľadom k tomu, pre každé i maticové elementy w_{ij} sú len nenulové pre tie $j \in S_i$ (kde S_i je množina susedstva), v oneskorení sú zahrnuté iba zodpovedajúce y_j . Pre jednoduchú

interpretáciu sú prvky matice priestorových váh typicky riadkovo normalizované, takže pre každý i , $\sum_j w_{ij} = 1$. V dôsledku toho môže byť priestorové oneskorenie interpretované ako vážený priemer (s váhou w_{ij}) susedov.

Je dôležité poznamenať, že prvky váhovej matice sú pre model nestochastické a exogénne. Typicky sú založené na geografickom usporiadaní pozorovaní alebo na súvislostiach. Váhy sú nenulové, keď dve miesta zdieľajú spoločnú hranicu, alebo sú v určitej vzájomnej vzdialenosti. Tento pojem je však úplne všeobecný a možno uvažovať aj o alternatívnych špecifikáciách priestorových váh, ako napríklad ekonomická vzdialenosť. (Anselin, 1980).

Obmedzenia uložené váhovou štruktúrou spolu so špecifickou formou priestorového procesu teda autoregresný alebo kĺzavý priemer určujú rozptyl – kovariančnú maticu pre y ako funkciu dvoch parametrov, rozptylu σ^2 a priestorového koeficientu ρ . Pre štruktúru SAR sú tieto výnosy (keďže $E[y - \mu_i] = 0$)

$$\text{cov}[(y - \mu_i), (y - \mu_i)] = E[(y - \mu_i)(y - \mu_i)'] = \sigma^2 \left[(I - \rho \mathbf{W})' (I - \rho \mathbf{W}) \right]^{-1}. \quad (3.21)$$

Toto je úplná matica, ktorá znamená, že otrasy v ktoromkoľvek mieste ovplyvňujú všetky ostatné miesta prostredníctvom tzv. efektu priestorového multiplikátora (alebo globálnej interakcie).

Hlavným rozdielom medzi procesmi v priestore v porovnaní s časovou doménou je to, že aj pri poruchovom člene nie sú diagonálne prvky konštantné. Ďalej heteroskedasticita závisí od susednej štruktúry vlozenej do matice priestorových váh \mathbf{W} . V dôsledku toho proces v y nie je kovariančný a stacionárny. Stacionárnosť sa dosahuje len vo veľmi zriedkavých prípadoch.

3.4.1 Výber vhodného modelu

Keď sa pre modely priestorovej regresie použije odhad maximálnej vierohodnosti, odvodenie parametra priestorovej závislosti ρ môže byť založené na *Waldovom teste* vytvorenom pomocou asymptotického t -testu z odhadovanej variančno-kovariančnej matice alebo testu pravdepodobnostného pomeru. Tieto testy priestorovej závislosti oproti nulovej hypotéze bez závislosti vyžadujú odhad maximálnej vierohodnosti priestorového modelu, ktorý predstavuje alternatívu k nulovej hypotéze bez priestorovej závislosti. Keďže odhad maximálnej vierohodnosti bol v minulosti nepružný, existuje veľa literatúry o štatistike

testovania Lagrangeových multiplikátorov, ktorá vyžaduje len odhad priestorového modelu, ktorý je spojený s nulovou hypotézou.

Test Lagrangeových multiplikátorov porovnávajúci priestorové modely je založený na princípe maximálnej vierohodnosti. Odlišujeme testy s takzvanou „*unidirectional*“ alternatívnou hypotézou, popisuje jednu konkrétnu špecifikáciu modelu. Na druhej strane „*multidirectional*“ alternatíva porovnáva rôzne špecifikácie modelov. Ďalej je možné testovať robustnú variantu, v zmysle možnej prítomnosti druhého typu modelu.

Testy Lagrangeových multiplikátorov LM_λ a LM_δ sú typu *unidirectional* s modelmi SEM a SAR sú ich príslušné alternatívy. Test pre *spatial error* model je daný indexom

$$LM_\lambda = \frac{1}{T} \left(\frac{\varepsilon^T \mathbf{W} \varepsilon}{\sigma^2} \right)^2,$$

kde $T = \text{tr}(\mathbf{W}^T \mathbf{W} + \mathbf{W}^2)$ a tr označuje stopu danej matice a ε je vektor rezíduí lineárneho regresného modelu. Test sa asymptoticky riadi rozdelením χ^2 s jedným stupňom voľnosti. Test pre *spatial lag* model má rovnaké asymptotické rozdelenie.

$$LM_\delta = \frac{1}{NJ_{\delta,\beta}} \left(\frac{\varepsilon^T \mathbf{W} \mathbf{y}}{\sigma^2} \right)^2,$$

kde $J_{\delta,\beta} = \frac{(\mathbf{W} \mathbf{X} \mathbf{b})^T \mathbf{M} (\mathbf{W} \mathbf{X} \mathbf{b}) + T \sigma^2}{N \sigma^2}$. Vektor \mathbf{b} označuje odhady metódou najmenších štvorcov a \mathbf{M} je projekčná matrica $\mathbf{M} = \mathbf{I}_N - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$.

Aby sme vedeli rozlíšiť či sa jedná o *spatial error* alebo *spatial lag* závislosť, boli navrhnuté takzvané robustné testy. Test uvažuje možnú prítomnosť druhého typu závislosti. Robustná štatistika pre závislosť *spatial error* vzhľadom k prítomnosti *spatial lag* je definovaná ako

$$LM_\lambda^* = \frac{1}{T - T^2 (NJ_{\delta\beta})^{-1}} \left(\frac{\varepsilon^T \mathbf{W} \varepsilon}{\sigma^2} - T (NJ_{\delta\beta})^{-1} \frac{\varepsilon^T \mathbf{W} \mathbf{y}}{\sigma^2} \right)^2.$$

Test pre robustný vzťah *spatial lag* vzhľadom k prítomnosti *spatial error* sa definuje ako

$$LM_\delta^* = \frac{1}{NJ_{\delta\beta} - T} \left(\frac{\varepsilon^T \mathbf{W} \mathbf{y}}{\sigma^2} - \frac{\varepsilon^T \mathbf{W} \varepsilon}{\sigma^2} \right)^2.$$

Obe tieto štatistiky majú asymptotické rozdelenie χ^2 s jedným stupňom voľnosti, ako tvrdí článok Florax a Nijkamp (2003), kde je možné dohľadať detailnejšie odvodenie.

3.4.2 Lineárny regresný model

Len v krátkosti si pripomenieme lineárny regresný model

$$Y = \alpha + X\beta + \varepsilon. \quad (3.23)$$

Y je vektor N nezávislých pozorovaní, matica X je typu $(N \times M)$ vysvetľujúcich premenných, kde platí $M < N$. Požadujeme viac pozorovaní než vysvetľujúcich premenných. Predpokladáme, že matica X má plnú hodnotu. Absolútny člen α môže byť vyjadrený zvlášť ako v zápise (3.23) alebo môže byť zahrnutý medzi vektor regresných parametrov β . Tieto parametre nepoznáme a sú modelom odhadované. Pre náhodný vektor ε platí nulovosť strednej hodnoty $E(\varepsilon)=0$ a homoskedasticita, nezávislosť rozptylov $Var(\varepsilon) = \sigma^2 I_N$. Pre smerodajnú odchýlku platí $\sigma^2 > 0$.

Vektor regresných koeficientov β odhadujeme metódou najmenších štvorcov OLS, z anglického *ordinary least squares*. Pre odhadnuté koeficienty platí

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (3.24)$$

Pomocou t -testu overujeme významnosť týchto koeficientov. Ak je p -hodnota nižšia ako 0,05, zamietame nulovú hypotézu i -tého koeficientu $\beta_i=0$ a považujeme ho za štatisticky významný na hladine významnosti 5%. F -štatistiku testujeme, ak je model signifikantný ako celok a taktiež sledujeme koeficient determinácie R^2 . Platí

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y})^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}. \quad (3.25)$$

Jedná sa o doplnok do jednotky pomeru celkového súčtu štvorcov a reziduálneho súčtu štvorcov. Koeficient determinácie hovorí, aká časť variability závisle premennej sa podarila regresiou vysvetliť. Viac o regresných lineárnych modeloch sa dočítame v literatúre Wooldridge (2003).

3.4.3 Model SAR

Model SAR je v angličtine označovaný ako *spatial lag model*, preto sa v literatúre môžeme stretnúť i s označením SLM. Tu sa budeme držať skratky SAR, pretože tento model je tiež nazývaný priestorový autoregresný proces. Súčtom $\sum_{j=1}^N w_{ij} y_j$ označujeme takzvaná priestorovo oneskorená premenná. Ak je matica W riadkovo normalizovaná, suma reprezentuje vážený priemer hodnôt v susedných oblastiach. Koeficient ρ vyjadruje silu

priestorovej závislosti v sledovanej vzorke. Modely SAR vznikajú z plného priestorového modelu tak, že parametre λ a θ položíme rovné nule. To znamená, že neuvažujeme priestorovú závislosť nezávislej premennej ani náhodnej zložky.

$$\begin{aligned} y &= \rho \mathbf{W}y + \mathbf{X}\boldsymbol{\beta} + u \\ u &\sim N(0, \sigma^2 I) \end{aligned} \quad (3.23)$$

kde ρ je priestorový autoregresný parameter a jeho štatistická významnosť, hodnota a matematické znamienko indikujú smer a silu priestorovej závislosti. Prítomnosť priestorovo oneskorenej závislej premennej $\mathbf{W}y$ na pravej strane regresnej rovnice spôsobuje problémy s endogenitou t. j. priestorová oneskorená závislá premenná je korelovaná s poruchovým členom a teda metóda najmenších štvorcov nie je vhodnou odhadovou metódou, odhadnuté parametre by boli skreslené a nekonzistentné. Upravme model (3.23) na redukovanú formu:

$$y = (I - \rho \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + (I - \rho \mathbf{W})^{-1} u \quad (3.24)$$

a po úprave (3.24) môžeme $\mathbf{W}y$ vyjadriť nasledujúcim spôsobom:

$$\mathbf{W}y = \mathbf{W}(I - \rho \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + \mathbf{W}(I - \rho \mathbf{W})^{-1} u \quad (3.25)$$

V tomto vzťahu je porušený predpoklad o nekorelovanosti vysvetľujúcich premenných s poruchovým členom (Anselin a Rey, 2014):

$$\begin{aligned} E\left[(\mathbf{W}y)^T u\right] &= E\left[u^T (I - \rho \mathbf{W}^T)^{-1} \mathbf{W}u\right] \\ &= E\left[tr u^T (I - \rho \mathbf{W}^T)^{-1} \mathbf{W}^T u\right] \\ &= tr \mathbf{W}(I - \rho \mathbf{W}^T)^{-1} E\left[uu^T\right] \neq 0 \end{aligned} \quad (3.26)$$

Výraz (3.26) sa nerovná nule a teda estimátor metódy najmenších štvorcov v tomto prípade nie je adekvátny. Alternatívnou metódou bola v odbornej literatúre navrhnutá priestorová dvojstupňová metóda najmenších štvorcov (S2SLS) (Anselin, 1988c; Kelejian a Robinson, 1993; Lee, 2003; Kelejian a kol. 2004) a priestorová metóda maximálnej vierohodnosti (SML) (Anselin, 1988c; LeSage a Pace, 2009).

3.4.4 Model SEM

Ďalším využívaným modelom je *spatial error model*, SEM. Ako napovedá názov, v tomto modeli sa zaoberáme priestorovou závislosťou náhodnej zložky a nie pozorovanej premennej. Poruchový člen modelu sa riadi modelom SAR, teda obsahuje priestorový autoregresný poruchový člen $\mathbf{W}u$, je to analogický model (3.23). Voľba takéhoto modelu naznačuje, že existuje ďalšia premenná, ktorá nie je v modeli zahrnutá, ale je napriek tomu významná alebo nesprávna špecifikácia funkčného tvaru modelu. Efekt je zachytený v

poruchovom člene a ak sú tieto vynechané premenné priestorovo korelované, tak existuje aj priestorová korelácia v rámci poruchového člena. Formulácia modelu:

$$y = \mathbf{X}\boldsymbol{\beta} + u \quad (3.27)$$

$$u = \lambda + \mathbf{W}u + v \quad (3.28)$$

kde λ je autoregresný parameter priestorového autoregresného poruchového člena $\mathbf{W}u$, vektor v je idiosynkratický poruchový člen. Zložky tohto vektora môžu byť heteroskedastické, čiže $E[v_i^2] = \sigma_i^2$ ale nie sú korelované, $E[v_i v_j] = 0$. Vzťah (3.28) vyjadruje, že náhodná chyba v oblasti i závisí od náhodných chýb v susedných oblastiach a idiosynkratickej zložke v_i . Toto implikuje, že nepozorovateľné chyby u sú predikovateľné z náhodných chýb v susedených lokalitách $\mathbf{W}u$. Poruchový člen z (3.28) môžeme upraviť vzťahom (3.29) a získavame redukovaný tvar pre priestorový autoregresný poruchový proces (3.30):

$$(I - \lambda \mathbf{W})u = v \quad (3.29)$$

$$u = (I - \lambda \mathbf{W})^{-1}v \quad (3.30)$$

Je zrejmé, že variančno-kovariančná matica poruchového člena u nie je diagonálnou maticou ale má nasledujúci tvar $E[uu^T] = (I - \lambda \mathbf{W})^{-1} \sum (I - \lambda \mathbf{W}^T)^{-1}$, kde \sum predstavuje variančno-kovariančnú maticu vektora poruchového člena v . V prípade homoskedastického poruchového člena v sa variančno-kovariančná matica u zjednodušuje na tvar:

$$E[uu^T] = \sigma^2 [(I - \lambda \mathbf{W})^T (I - \lambda \mathbf{W})]^{-1}.$$

Redukovaný tvar z (3.30) môžeme substituovať do základnej špecifikácie modelu vyjadrenej vzťahom (3.27) a získavame transformovaný tvar modelu:

$$y = \mathbf{X}\boldsymbol{\beta} + (I - \lambda \mathbf{W})^{-1}v \quad (3.31)$$

Ďalšiu úpravou, vynásobením oboch strán modelu (3.31) členom $(I - \lambda \mathbf{W})$, tzv. priestorovým filtrom získavame nasledujúci tvar modelu:

$$(I - \lambda \mathbf{W})y = (I - \lambda \mathbf{W})\mathbf{X}\boldsymbol{\beta} + v \quad (3.32)$$

Pomocou tejto úpravy odstránime prítomnosť autokorelácie z poruchového člena na pravej strane rovnice. Tento postup sa nazýva aj ako priestorová Cochrane – Orcuttova transformácia (bližšie pozri Kelejian a Prucha, 1999). V situácii, ak je hodnota priestorového autoregresného parametra λ známa, potom je možné model (3.32) zjednodušiť na lineárnu

regresiu s priestorovo filtrovanou závislou premennou $y_s = y - \lambda \mathbf{W}y$ a priestorovo filtrovanými vysvetľujúcimi premennými $X_s = X - \lambda \mathbf{W}X$:

$$y_s = \mathbf{X}_s \boldsymbol{\beta} + v \quad (3.33)$$

Odhad parametrov tohto modelu je možné prostredníctvom priestorovo váženej metódy najmenších štvorcov.

3.5 Priestorová heterogenita

Ak hovoríme, že vo vzťahu medzi závislou a vysvetľujúcou respektíve vysvetľujúcimi (závislými) premennými môže nastať priestorová odlišnosť, ide o situácie kde je v modeli prítomná priestorová heterogenita. V priestorových režimoch (spatial regime), čo sú vlastne skupiny priestorových jednotiek, namiesto fixných hodnôt regresných parametrov pre všetky priestorové jednotky predpokladáme, že ich hodnoty sa môžu odlišovať. Môžeme uvažovať ešte aj o odlišných regresných parametroch pre každý bod v priestore, t. j. pre všetky priestorové jednotky a takýto vzťah môžeme zapísať formálnym zápisom nasledovne:

$$y_i = x_i^T \boldsymbol{\beta}_i + u_i, \quad i = 1, 2, \dots, N \quad (3.34)$$

kde index i označuje i -te pozorovanie, x_i^T predstavuje $1 \times k$ vektor vysvetľujúcich premenných a zodpovedajúci $k \times 1$ vektor parametrov $\boldsymbol{\beta}_i$. Odhadnúť vektor parametrov $\boldsymbol{\beta}_i$ pre všetkých N pozorovaní však nie je možné, a to z dôvodu, že nastáva problém so stupňami voľnosti. Inak povedané, nedisponujeme dostatočnými údajmi o vzorke dát, takže nemôžeme odhadnúť parametre pre všetky oblasti.

4. Výsledky práce a diskusia

4.1 Predstavenie softvéru R

R je voľne šíriteľný softvér pre štatistické výpočty a grafiku. Je mimoriadne pôsobivý a ako taký je dnes široko používaný pre akademický výskum, ako aj pre obchodný sektor. Na rozdiel od softvéru, akým je napríklad program Excel alebo SPSS, používateľ musí zadávať príkazy na vykonanie úloh, ako je napríklad načítanie do súboru údajov alebo vykonanie výpočtu.

Najväčšou výhodou tohto prístupu je, že môžete vytvoriť dokument alebo skript, ktorý poskytuje prehľad o tom, čo ste urobili, čo zase umožňuje jednoduché opakovanie úloh. Grafika môže byť jednoducho upravená a vylepšená zmenou skriptu alebo prechodom na minulé príkazy a rýchlymi úpravami. Nanešťastie, výpočet príkazového riadku môže byť najskôr vypnutý. Je ľahké urobiť chyby, ktoré nie sú vždy zrejmé na detekciu. Napriek tomu existujú dobré dôvody, aby ste ako výpočtový softvér využili R:

- Je široko intuitívny so silným zameraním na grafiku publikačnej kvality. Je "inteligentný" a má tendenciu držať sa štatistických konvencií a prezentovať dáta v rozumných spôsoboch.
- Je bezplatný, cross-platformový, prispôsobiteľný a rozšíriteľný o celú škálu knižníc vrátane tých, ktoré umožňujú diskretnú voľbu, viacúrovňovú a pozdĺžnu regresiu, mapovanie, priestorovú štatistiku, priestorovú regresiu a geoštatistiku.
- Je uznávaný a využívaný v najväčších svetových technologických spoločnostiach (vrátane spoločností Google, Microsoft a Facebook), najväčších farmaceutických spoločností (vrátane spoločností Johnson & Johnson, Merck a Pfizer) a v stovkách ďalších spoločností.
- Ponúka prenosnú zručnosť, ktorá poukazuje na skúsenosti so štatistikou aj výpočtovou technikou.

4.1.1 Balíky programu R pre analýzu priestorových údajov

V R je základnou jednotkou zdieľaného kódu balík. Ten spája kód, údaje, dokumentáciu a testy, a je ľahko zdieľaný s ostatnými užívateľmi programu. V apríli 2016 bolo k dispozícii viac ako 8 200 balíkov v komplexnej sieti archívov R nazývanej CRAN. Táto obrovská škála balíčkov je jedným z dôvodov, prečo je R tak úspešný, pretože každý užívateľ môže prispieť vlastným vytvoreným balíkom. Čiže je viac ako pravdepodobné, že

problém na ktorom pracujete niekto vyriešil a môžete využiť jeho prácu jednoducho stiahnutím balíka.

V našej práci sa zameriame iba na balíčky, ktoré nám pomôžu spracovať analýzu priestorových dát. Ide o balíky: *maptools*, *spdep*, *leaflet*, *RColorBrewer*. Balík *maptools* má funkcie potrebné na čítanie geografických údajov, najmä ESRI *shapefiles*. Balík *spdep* na druhej strane obsahuje funkcie, ktoré sa použijú pri analýze priestorových údajov. Balíky *leaflet* a *RColorBrewer* budeme používať na dizajn zobrazenia máp a grafov. Pre inštaláciu balíkov zadáme nasledovný príkaz, ktorý automaticky nainštaluje všetky potrebné balíky:

```
install.packages("maptools",dependencies = TRUE)
install.packages("spdep",dependencies=TRUE)
install.packages("leaflet",dependencies = TRUE)
install.packages("RColorBrewer",dependencies = TRUE)
```

Príkaz *dependencies = TRUE* nastaví, aby sa automaticky doinštalovali všetky súvisiace balíky, na ktoré je funkčnosť inštalovaného balíka závislá. Na prístup k funkciám a údajom v balíku musíme najprv „zavolať“. Funkcia volania balíka je *library*:

```
library(maptools)
Loading required package: sp
Checking rgeos availability: TRUE
```

```
library(spdep)
Loading required package: Matrix
Loading required package: spData
```

Teraz máme balíky načítané a môžeme začať prechádzať vstupné dáta.

4.1.2 Čítanie a mapovanie priestorových údajov v R

Priestorové údaje môžu byť uvádzané v mnohých „tvaroch“ a „veľkostiach“, medzi najbežnejšie typy priestorových dát patrí:

- **Body** sú najzákladnejšou formou priestorových údajov. Definujú miesto ako bod napríklad mestá, GPS súradnice alebo akýkoľvek iný diskretný objekt definovaný v priestore.
- **Priamky** sú množinou zoradených bodov spojených úsečkami.
- **Polygóny** označujú oblasť a možno ju považovať za postupnosť spojených bodov, kde prvý a posledný bod sú rovnaké.
- **Raster** je zbierka bodov alebo pravouhlých buniek, usporiadaných v pravidelnej mriežke, mozaike.

Viac informácií o priestorových útvaroch sa môžete dočítať v práci R. S. Bivand, Pebesma a Gomes-Rubio (2008). My sa zameriame na polygóny. Priestorové údaje sú zväčša uvádzané v súbore nazývanom *shapefile* (.shp). Tento typ súboru ukladá topologickú geometriu a informácie o atribútoch priestorových funkcií v súbore údajov. Okrem toho nevyžadujú veľa miesta na disku a dajú sa ľahko čítať a prepísať. (ESRI (1998))

Pre prácu s priestorovými údajmi budeme potrebovať vyššie spomínaný *shapefile* súbor, ktorý bude obsahovať potrebné charakteristiky pre definovanie priestoru. Pre náš zvolený výber dát, teda členské krajiny Európskej únie, stiahneme potrebné dáta na stránke Eurostatu. Avšak tento súbor budeme musieť upraviť, respektíve rozšíriť o ukazovateľ HDP v parite kúpnej sily na obyvateľa, ktorý sme tiež stiahli zo štatistickej stránky. Pre ďalšie spracovanie všetkých dát budeme potrebovať nasledovné súbory:

- Main file: NUTS2.shp
- Index file: NUTS2.shx
- dBase table: NUTS2.dbf

Main file popisuje tvary objektov so zoznamom ich vrcholov. Index file je binárny súbor priestorových indexov. Tabuľka dBase table uchováva atribúty pre každý tvar priestorových dát. Na čítanie údajov z polygónového *shapefile* súboru sa používa funkcia *readShapePoly*, ktorá vytvorí objekt *SpatialPolygonsDataFrame* – dátový rámec priestorových polygónov.

```
nuts.poly<-readShapePoly('NUTS2.shp')
```

Shapefile súbor je teraz nahraný ako *nuts.poly* objekt. Pre uistenie sa, že súbor je naozaj dátový rámec priestorových polygónov využijeme funkciu *class*, ktorá opisuje tvar súborov.

```
class(nuts.poly)
```

```
[1] "SpatialPolygonsDataFrame"
```

```
attr(,"package")
```

```
[1] "sp"
```

Objekt *nuts.poly* má štyri časti, kde prvá časť je dátová, ktorá obsahuje premenné, ktoré použijeme pre našu analýzu. Druhá je časť polygónu a obsahuje informácie o „tvare“. Tretia časť, *bbox*, je ohraničovaci rámček a štvrtá časť je *proj4string* a obsahuje projekcie. Na prístup k dátovej časti môžeme použiť funkciu *slot* alebo symbol @. Pre kompletný vzhľad *nuts.poly* použijeme nasledovnú funkciu:

```
str(slot(nuts.poly,"data"))
```

```
'data.frame': 261 obs. of 7 variables:
 $ NUTS_ID: Factor w/ 261 levels "AT11","AT12",...: 23 37 31 7 24 1 25 8
38 9 ...
 $ NAME : Factor w/ 261 levels "Abruzzo","Alentejo",...: 199 69 200 196
254 26 255 230 233 242 ...
 $ gdp17 : num 11800 35600 22600 45400 13000 27100 23700 40900 40400 4
0600 ...
 $ gdp07 : num 8900 29500 17700 37700 8500 21200 17100 33700 32400 338
00 ...
 $ lngdp17: num 9.38 10.48 10.03 10.72 9.47 ...
 $ lngdp07: num 9.09 10.29 9.78 10.54 9.05 ...
 $ rate : num 0.282 0.188 0.244 0.186 0.425 0.246 0.326 0.194 0.221 0
.183 ...
 - attr(*, "data_types")= chr "C" "C" "N" "N" ...
```

Prvá časť *data.frame* obsahuje dáta pre našu analýzu, ďalej môžeme vidieť premenné:

- gdp17 – Hrubý domáci produkt vyjadrený v parite kúpnej sily na obyvateľa v roku 2017
- gdp07 - Hrubý domáci produkt vyjadrený v parite kúpnej sily na obyvateľa v roku 2007
- lngdp17 – logaritmus hrubého domáceho produktu vyjadrený v parite kúpnej sily na obyvateľa v roku 2017
- lngdp07 - logaritmus hrubého domáceho produktu vyjadrený v parite kúpnej sily na obyvateľa v roku 2007
- rate – logaritmus podielu hrubého domáceho produktu vyjadrený v parite kúpnej sily na obyvateľa v roku 2017 a hrubého domáceho produktu vyjadrený v parite kúpnej sily na obyvateľa v roku 2007

Taktiež môžeme získať súhrnnú štatistiku sledovaných premenných pomocou súhrnu funkcií:

```
summary(nuts.poly@data$lngdp7)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
8.394 9.932 10.442 10.476 11.000 13.376
```

Údaje, s ktorými pracujeme sú stiahnuté z databázy Eurostatu za pozorované obdobie 2007-2017, kde sa v EÚ nachádzalo presne 28 členských štátov, ktoré podľa úrovne NUTS 2 boli rozdelené do 281 územných celkov. Tieto sú presne rozpísané v dole uvedenej tabuľke č. 1. Avšak, tieto územné celky sme museli očistiť o ostrovné štáty, pre ďalšiu prácu s maticou susednosti W, ktorú sme normovali na súčet v riadku 1, čo by nebolo splnené v prípade ostrovov, ktoré nadobúdajú súčet 0. Sú to oblasti, ktoré nemajú susedov. Ďalej sme očisťovali dáta o chýbajúce hodnoty pre jednotlivé pozorovanie a o štáty, ktoré neboli

v oboch pozorovaných obdobiach členmi Európskej únie. Číže konečný počet pozorovaní, s ktorým budeme ďalej pracovať je 261 čo môžeme vidieť aj na výstupe z programu R, kde sú definované jednotlivé vstupné charakteristiky.

Nasledujúci príkaz nám dáva graf vybraných krajín v rámci Európskej únie očistených o ostrovné štáty:

```
plot(nuts.poly)
```

Obr. 1 Mapa členských krajín EÚ



Zdroj: Vlastné spracovanie

Avšak, môžeme ísť o krok ďalej a urobiť reálnejšiu aplikáciu grafov pomocou balíka *leaflet*. Týmto sa vytvorí interaktívna mapa, ktorú možno vykresliť pomocou stránok HTML.

```
library(leaflet)
```

```
> leaflet(nuts.poly) %>%
```

```
+ addPolygons(stroke = FALSE,fillOpacity = 0.5,smoothFactor = 0.5) %>%
```

```
+ addTiles()
```

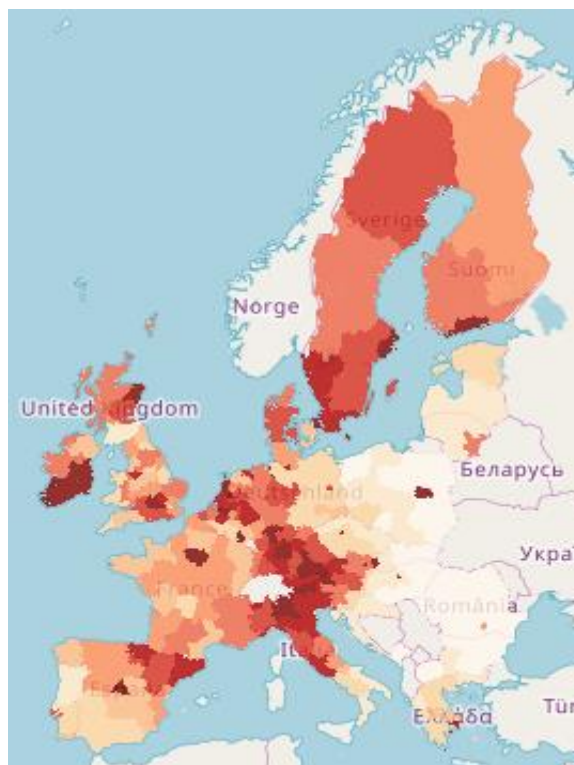
Obr.2 Mapa členských štátov EÚ prepojenie HTML



Zdroj: Vlastné spracovanie

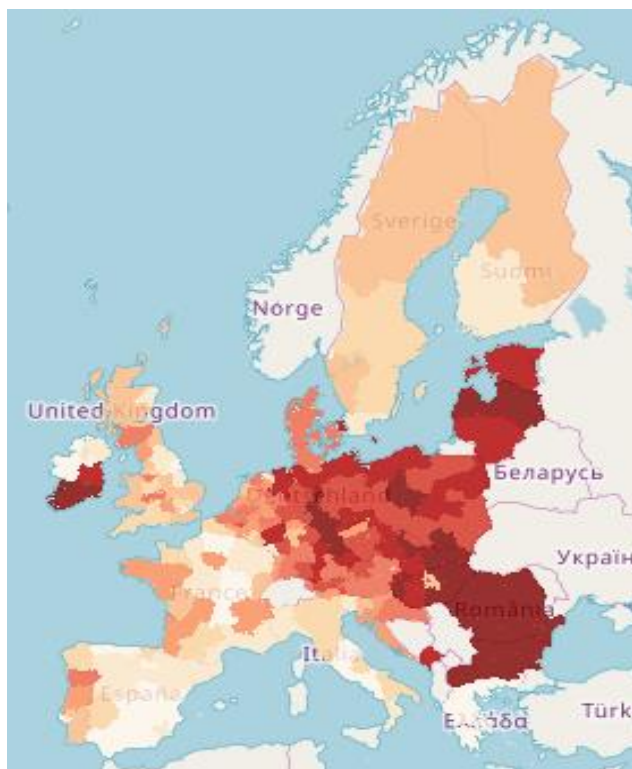
Do grafu môžeme pridať farebnú škálu rozdelenia podľa výšky hodnôt, kde najtmavšie časti sú miesta s najvyššími hodnotami HDP na obyvateľa a naopak.

Obr. 3 Mapa členských štátov Európskej únie HTML podľa výšky HDP na obyv.



Zdroj: Vlastné spracovanie

Obr. 4 Mapa členských štátov Európskej únie HTML podľa miery rastu HDP na obyv.



Zdroj: Vlastné spracovanie

4.1.3 Priestorová ekonometria v R

Ak sa zaoberáme priestorom, musíme mať na pamäti Toblerov prvý zemepisný zákon: „Všetko je spojené so všetkým ostatným, ale blízke veci sú viac ako veci, ktoré sú ďaleko od seba“ (Tobler, 1979). V nasledujúcej časti sa zameriame na špecifikáciu priestorovej závislosti, testov špecifikácií na detekciu priestorovej závislosti v regresných modeloch a na základných regresných modeloch, ktoré zahŕňajú priestorovú závislosť. Toto ukážeme pomocou údajov v súbore *shapefile* načítanom v predchádzajúcej časti.

4.1.3.1 Model OLS

Tradičným prístupom po mnoho rokov bolo ignorovanie priestorovej závislosti údajov a odhadnutie modelu regresie OLS. V programe R odhadneme model pomocou *lm* funkcie, kde závislá premenná je *rate* a vysvetľujúca premenná je *lngdp07*. Tieto premenné sme si vyššie presne špecifikovali. Zvyšok príkazu sa odvoláva na zdroj dát, z ktorých bude model *nuts.ols* vytvorený. Príkaz *summary* nám v tomto prípade zobrazí výstup výsledkov odhadnutého modelu príjmovej beta konvergenencie.

Formulácia modelu: $\ln\left(\frac{y_{T,i}}{y_{0,i}}\right) = \alpha + \beta \ln y_{0,i} + \varepsilon_i$

kde

- α je konštanta,
- β je odhadovaný parameter,
- $y_{T,i}$ je regionálne HDP na obyvateľa regiónu i v čase T (v našom modeli lngdp17),
- $y_{0,i}$ je regionálne HDP na obyvateľa regiónu i v čase 0 (v našom modeli lngdp07),
- $\ln\left(\frac{y_{T,i}}{y_{0,i}}\right)$ je rast regionálneho HDP na obyvateľa počas sledovaného obdobia (v našom modeli premenná rate).

```
nuts.ols<-lm(rate~lngdp07,data=nuts.poly@data)
summary(nuts.ols)
```

Call:

```
lm(formula = rate ~ lngdp07, data = nuts.poly@data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.38817	-0.06957	-0.00447	0.07179	0.49357

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.35411	0.18045	7.504	9.94e-13 ***
lngdp07	-0.12084	0.01792	-6.744	9.99e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1207 on 259 degrees of freedom

Multiple R-squared: 0.1494, Adjusted R-squared: 0.1461

F-statistic: 45.48 on 1 and 259 DF, p-value: 9.989e-11

Problém s ignorovaním priestorovej štruktúry údajov znamená, že odhady OLS v nepriestorovom modeli môžu byť skreslené, nekonzistentné alebo neefektívne v závislosti od toho, čo je skutočnou závislosťou. Z uvedeného modelu vyplýva, že koeficient β je štatisticky významný a zároveň záporný čo indikuje konvergenciu HDP.

4.1.3.2 Modelovanie priestorovej závislosti

Pre otestovanie prítomnosti priestorovej autokorelácie je potrebné pre *Moranovo I* zostaviť maticu priestorových váh, v ktorej sú definované vzťahy susednosti. Špecifikácia susednosti pozorovaní je celkom svojvoľná a v literatúre existuje široká škála návrhov. Jeden

populárny spôsob je použiť jedno z dvoch nasledujúcich kritérií, a to „Rook kritérium“ a „Queen kritérium“, obe sme hlbšie rozvinuli v teórii váh v kapitole (3.2).

Zameriame sa na používanie kritéria „Queen“ a predstavíme aj tvorbu váh s kritériom „Rook“. V R, aby sme získali maticu váh, využívame dve funkcie. V prvom kroku používame *poly2nb*, ktorý vytvára zoznam susedov, ak je zadaná voľba *queen = TRUE*, bude zostavená pomocou kritéria *queen*. Ak je zadaná *FALSE*, použijú sa kritériá veže. Ďalším krokom je doplnenie zoznamu susedov o priestorové váhy. Riadok **W** štandardizuje maticu.

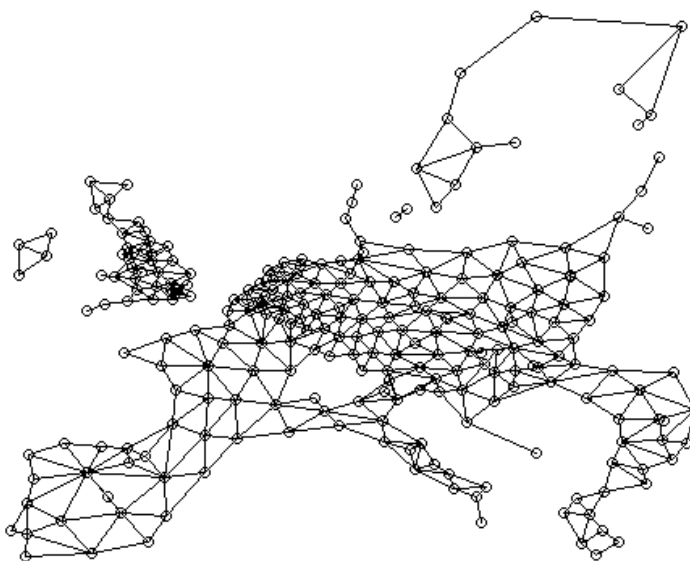
```
list.queen<-poly2nb(nuts.poly, queen=TRUE)
> W<-nb2listw(list.queen, style="W", zero.policy=TRUE)
> W
```

```
Characteristics of weights list object:
Neighbour list object:
Number of regions: 261
Number of nonzero links: 1186
Percentage nonzero weights: 1.74102
Average number of links: 4.544061
weights style: w
weights constants summary:
      n      nn  S0      S1      S2
w 261 68121 261 138.0051 1091.768
```

Matica je typu „Queen“ a teda za susedov považujeme len tie oblasti, ktoré majú spoločnú hranicu so susediacim regiónom. V zhrnutí matice môžeme vidieť, že priemerný počet susedov pre každú oblasť respektíve región je 4 až 5. Teraz môžeme vykresliť distribúciu odkazov s obvyklou grafickou funkciou *plot*.

```
plot(W,coordinates(nuts.poly))
```

Obr. 3 Grafické zobrazenie susedov



Zdroj: Vlastné spracovanie

Na získanie matice váh založenej na vzdialenostiach používame dve funkcie: *coordinates* ktorá získava súradnice centroidov sčítacích úsekov polygónov a *dnearneigh*, ktorá identifikuje susedov medzi dvoma vzdialenosťami v kilometroch meraných pomocou euklidovskej vzdialenosti. Ak napríklad chceme nájsť susedov do 1 km, urobíme:

```
coords<-coordinates(nuts.poly)
```

```
W_dist<-dnearneigh(coords,0,1,longlat = FALSE)
```

```
W_dist
```

```
Neighbour list object:
```

```
Neighbour list object:
```

```
Number of regions: 261
```

```
Number of nonzero links: 432
```

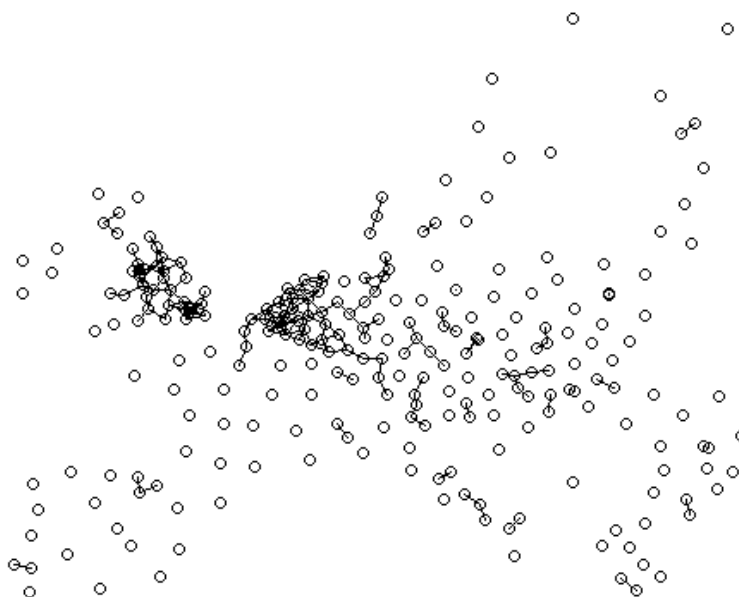
```
Percentage nonzero weights: 0.6341657
```

```
Average number of links: 1.655172
```

```
112 regions with no links:
```

```
1 3 4 5 7 17 18 21 33 39 40 41 43 51 52 54 58 60 61 62 64 74 76 77 78 79
81 82 84 85 90 92 94 100 105 106 108 109 110 111 112 113 115 116 117 118
119 120 126 129 137 138 140 141 143 144 145 146 148 149 150 151 153 154
155 156 157 158 160 161 162 163 165 166 168 169 170 171 173 175 176 177
178 179 183 184 187 189 192 193 194 196 197 198 203 206 207 208 209 216
217 219 224 227 233 237 239 240 245 246 248 254
```

Obr. 4 Váhy na základe vzdialenosti



Zdroj: Vlastné spracovanie

4.1.3.3 Moranov I test

Na vykonanie Moranovho testu na našich údajoch potrebujeme dva vstupy, objekt regresie *lm*, ktorý sme odhadli v časti OLS modelu a maticu **W**, v našom prípade sme zvolili typ kritéria „Queen“. Týmto testom skúmame nulovú hypotézu o neprítomnosti autokorelácie v modeli OLS.

```
moran.lm<-lm.morantest(nuts.ols,W,alternative="two.sided")
print(moran.lm)
```

Global Moran I for regression residuals

```
data:
model: lm(formula = rate ~ lngdp07, data = nuts.poly@data)
weights: W

Moran I statistic standard deviate = 13.855, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:
Observed Moran I      Expectation      Variance
0.610164154          -0.006271037          0.001979621
```

Koeficient indexu je štatisticky významný jeho hodnota je kladná čo dokazuje prítomnosť autokorelácie v modeli, takže je nutné zaoberať sa priestorovou závislosťou v modeloch, a preto odhadneme modely SAR a SEM.

4.1.3.4 Lagrangeov test multiplikátora

Moranov test nám bohužiaľ neodpovie na otázku výberu alternatív modelov, na druhej strane Lagrangeov test multiplikátora špecifikuje alternatívnu hypotézu, ktorá nám pomôže s touto úlohou. Testy Lagrangeových multiplikátorov pre priestorovú závislosť sú zahrnuté do funkcie *lm.LMtests* a zahŕňajú ako alternatívy prítomnosť priestorového oneskorenia a prítomnosť priestorového oneskorenia v poruchovom člene. Oba testy, ako aj ich robustné prevedenie sú zahrnuté do funkcie *lm.LMtests*. Na ich vyvolanie používame voľbu testu = "all". Objekt regresie a objekt priestorového zoznamu listov musia byť znova odovzdané ako argumenty:

```
LM<-lm.LMtests(nuts.ols,W,test = "all")
> print(LM)
```

Lagrange multiplier diagnostics for spatial dependence

```
data:
model: lm(formula = rate ~ lngdp07, data = nuts.poly@data)
weights: W
```

```
LMerr = 183.77, df = 1, p-value < 2.2e-16
```

Lagrange multiplier diagnostics for spatial dependence

```
data:
model: lm(formula = rate ~ lngdp07, data = nuts.poly@data)
weights: w
```

LMlag = 210.51, df = 1, p-value < 2.2e-16

Lagrange multiplier diagnostics for spatial dependence

```
data:
model: lm(formula = rate ~ lngdp07, data = nuts.poly@data)
weights: w
```

RLMerr = 1.3265, df = 1, p-value = 0.2494

Lagrange multiplier diagnostics for spatial dependence

```
data:
model: lm(formula = rate ~ lngdp07, data = nuts.poly@data)
weights: w
```

RLMlag = 28.066, df = 1, p-value = 1.173e-07

Lagrange multiplier diagnostics for spatial dependence

```
data:
model: lm(formula = rate ~ lngdp07, data = nuts.poly@data)
weights: w
```

SARMA = 211.84, df = 2, p-value < 2.2e-16

Keďže LMerr a LMlag sú štatisticky významné takže nevieme rozhodnúť o lepšom formulovaní modelu, a preto prejdeme na ich robustné testy. Tieto sú skutočne odolné voči prítomnosti iného „typu“ autokorelácie. Robustná verzia testov naznačuje, že pravdepodobnejšou alternatívou je opäť model oneskorenia SAR, keďže RLMerr je štatisticky nevýznamný. No odhadneme oba modely pre lepšie znázornenie problematiky a z dôvodu väčšej istoty.

4.1.4.5 Model SAR

Do modelu SAR zahrnieme priestorový autoregresný parameter a použijeme odhad maximálnej vierohodnosti. Toto dosiahneme v R pomocou funkcie *lagsarlm*:

$$\text{Formulácia modelu: } \ln\left(\frac{y_{T,i}}{y_{0,i}}\right) = \alpha + \beta \ln y_{0,i} + \rho \sum_{j=1}^N w_{i,j} \ln\left(\frac{y_{T,j}}{y_{0,j}}\right) + \varepsilon_i$$

kde

- α je konštanta,

- ρ, β sú odhadované parametre,
- $w_{i,j}$ je prvok matice susednosti, i a j sú regióny,
- $y_{T,i}$ je regionálne HDP na obyvateľa regiónu i v čase T (v našom modeli lngdp17),
- $y_{0,i}$ je regionálne HDP na obyvateľa regiónu i v čase 0 (v našom modeli lngdp07),
- $\ln\left(\frac{y_{T,i}}{y_{0,i}}\right)$ je rast regionálneho HDP na obyvateľa počas sledovaného obdobia (v našom modeli premenná rate).

```
sar.nuts<-lagsarlm(rate~lngdp07,data=nuts.poly@data,W)
summary(sar.nuts)
```

```
Call:lagsarlm(formula = Rate ~ lngdp7, data = nuts.poly@data, listw = w)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.35724338	-0.03590845	-0.00057407	0.03748048	0.47268408

Type: lag

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.400415	0.118702	3.3733	0.0007428
lngdp07	-0.036355	0.011624	-3.1276	0.0017623

Rho: 0.74675, LR test value: 196.84, p-value: < 2.22e-16

Asymptotic standard error: 0.038731

z-value: 19.281, p-value: < 2.22e-16

wald statistic: 371.74, p-value: < 2.22e-16

Log likelihood: 280.8859 for lag model

ML residual variance (sigma squared): 0.0056481, (sigma: 0.075154)

Number of observations: 261

Number of parameters estimated: 4

AIC: -553.77, (AIC for lm: -358.93)

LM test for residual autocorrelation

test value: 9.7098, p-value: 0.0018329

Ako možno vidieť parameter ρ , v ktorom je zahrnutá priestorová závislosť modelu má kladnú hodnotu a je významným parametrom v modeli, takže medzi regiónmi EÚ existuje priestorová závislosť. Z modelu tiež vyplýva, že parameter β je záporný a štatisticky významný. Následne môžeme porovnať rezíduá OLS regresie s rezíduami priestorového autoregresívneho modelu. Pre prístup k rezíduám modelu OLS a modelu SAR jednoducho urobíme:

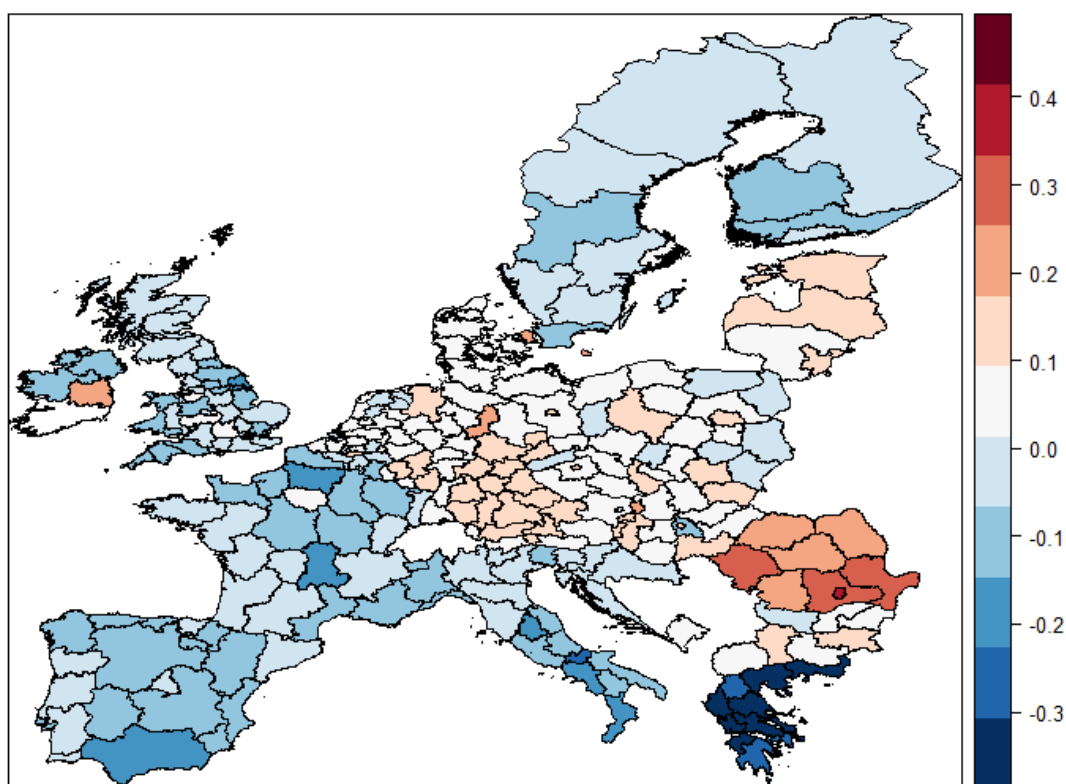
```
nuts.poly@data$nuts.ols.res<-resid(nuts.ols)
```

```
nuts.poly@data$sar.nuts.res<-resid(sar.nuts)
```

V tomto kroku vytvorím novú premennú v údajovej časti súboru *shapefile*, ktorý pomôže pri vykresľovaní rezíduí. Na vykreslenie rezíduí použijeme funkciu *spplot* v balíku *spdep*. Argumenty sú *shapefile*, premenná, ktorú chceme vykresliť, počet prestávok a farby, ktoré budeme používať.

```
spplot(nuts.poly,"nuts.ols.res",
       at=seq(min(nuts.poly@data$nuts.ols.res,na.rm=TRUE),
              max(nuts.poly@data$nuts.ols.res,na.rm=TRUE),length=12),
       col.regions=rev(brewer.pal(11,"RdBu")))
```

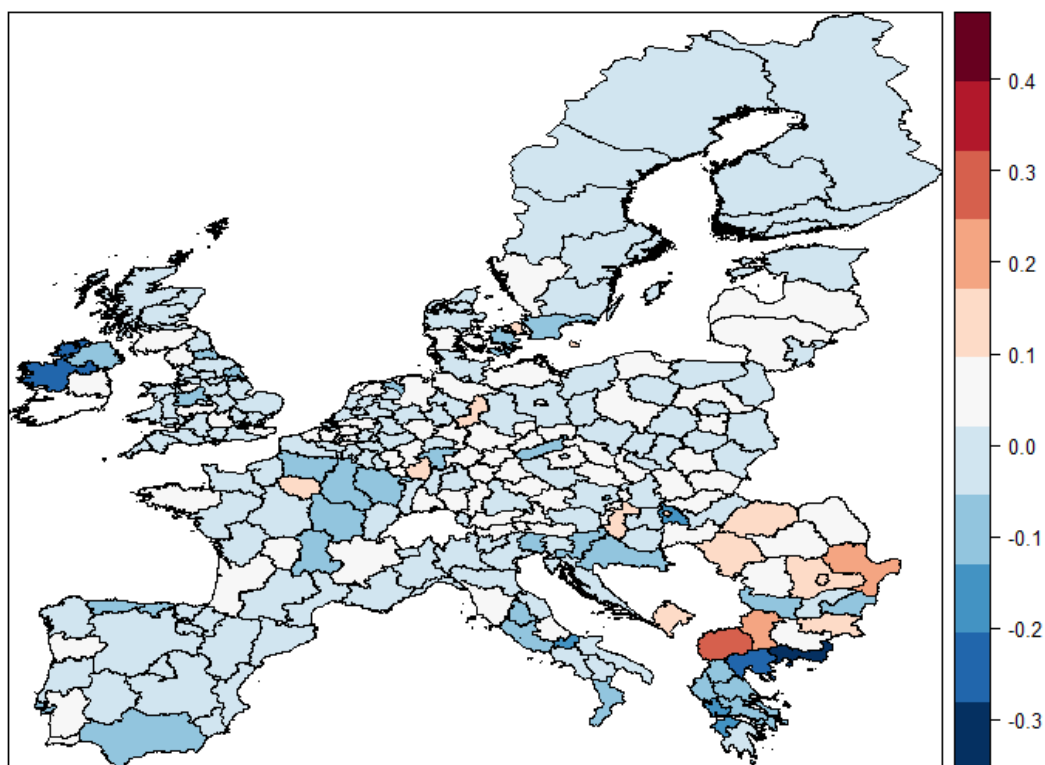
Obr. 5 OLS model – rezíduá



Zdroj: Vlastné spracovanie

```
spplot(nuts.poly,"sar.nuts.res",
       at=seq(min(nuts.poly@data$sar.nuts.res,na.rm=TRUE),
              max(nuts.poly@data$sar.nuts.res,na.rm=TRUE), length=12),
       col.regions=rev(brewer.pal(11,"RdBu")))
```


Obr. 5 SAR model – rezíduá



Zdroj: Vlastné spracovanie

4.1.4.6 Marginálny efekt

Všimnime si, že prítomnosť matice priestorových váh robí marginálne efekty výraznejšími a komplikovanejšími ako v „tradičnom“ modeli OLS. Pace a LeSage (2009) navrhli tri opatrenia na zvýšenie dopadu, ktoré sa vykonajú v R nasledovne:

`impacts(sar.nuts,listw = W)`

Impact measures (lag, exact):

	Direct	Indirect	Total
lnngdp07	-0.04567305	-0.09787968	-0.1435527

Priamy vplyv sa vzťahuje na priemerný celkový vplyv zmeny nezávislej premennej na závislú pre každým pozorovaním t.j. $N^{-1} \sum_{i=1}^n \frac{\partial E(y_i)}{\partial X_i}$, nepriamy vplyv, ktorý je súčtom

dopadu vytvoreného na jedinom pozorovaní všetkými ostatnými pozorovaniami a dopadom jedného pozorovania na všetky ostatné pozorovania. Schopnosť modelov priestorovej ekonometrie zachytiť tieto multiregionálne priestorové interakcie predstavuje dôležitý pozitívny aspekt priestorových ekonometrických modelov.

4.1.4.7 SEM model

Ak chceme odhadnúť model priestorových chýb, opäť pristupujeme k odhadu ako v prvom modeli SAR, použijeme odhad maximálnej virohodnosti.

$$\ln\left(\frac{y_{T,i}}{y_{T,0}}\right) = \alpha + \beta \ln y_{0,i} + \varepsilon_i$$

Formulácia modelu:

$$\varepsilon_i = \lambda \sum_{j=1}^n w_{i,j} \varepsilon_j + \eta_i$$

- α je konštanta,
- $w_{i,j}$ je prvok matice susednosti, i a j sú regióny,
- $y_{T,i}$ je regionálne HDP na obyvateľa regiónu i v čase T (v našom modeli lngdp17),
- $y_{0,i}$ je regionálne HDP na obyvateľa regiónu i v čase 0 (v našom modeli lngdp07),
- $\ln\left(\frac{y_{T,i}}{y_{0,i}}\right)$ je rast regionálneho HDP na obyvateľa počas sledovaného obdobia (v našom modeli premenná rate),
- η_i a ε_i sú rezíduá,
- β a λ sú odhadované parametre.

```
errorsalm.nuts<-errorsarlm(Rate~lngdp7,data=nuts.poly@data,W)
summary(errorsalm.nuts)
```

```
Call:errorsarlm(formula = rate ~ lngdp07, data = nuts.poly@data, listw = w)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.3440412	-0.0353216	-0.0033919	0.0338555	0.4601107

Type: error

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.363746	0.174535	2.0841	0.03715
lngdp07	-0.022395	0.017067	-1.3122	0.18947

Lambda: 0.77148, LR test value: 189.04, p-value: < 2.22e-16

Asymptotic standard error: 0.036958

z-value: 20.874, p-value: < 2.22e-16

wald statistic: 435.74, p-value: < 2.22e-16

Log likelihood: 276.9868 for error model

ML residual variance (sigma squared): 0.0057169, (sigma: 0.07561)

Number of observations: 261

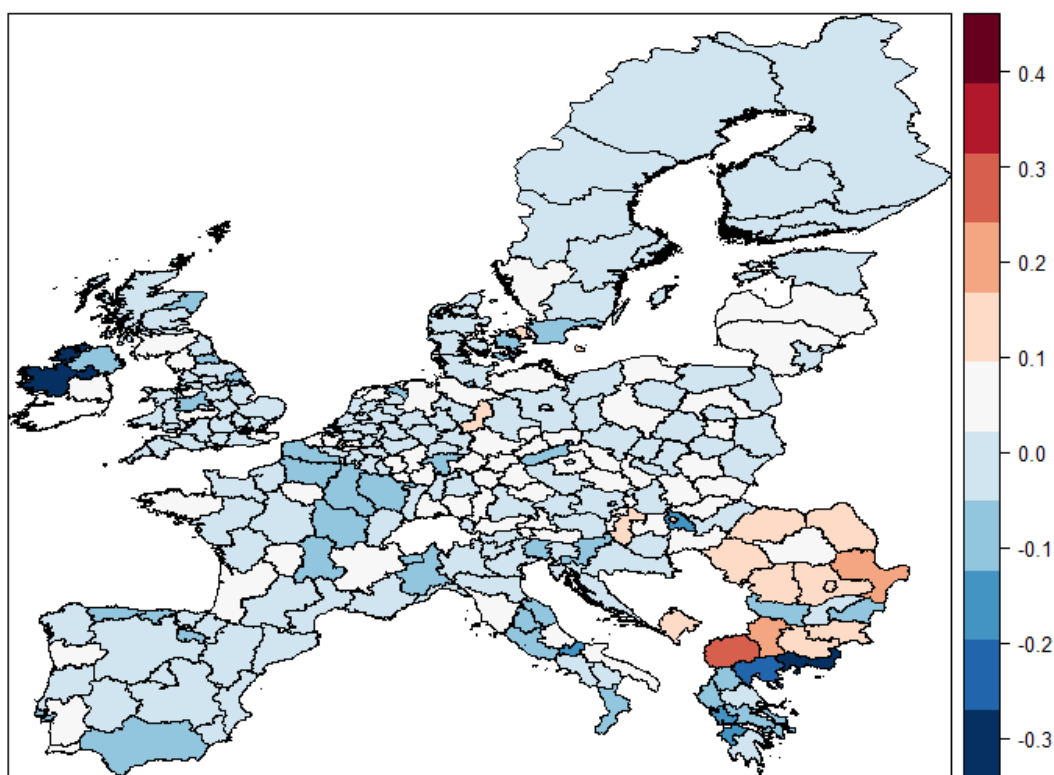
Number of parameters estimated: 4

AIC: -545.97, (AIC for lm: -358.93)

V modeli SEM sa tiež potvrdila významnosť koeficientu, ktorý zachytáva prítomnosť priestorovej závislosti medzi oblasťami λ . Avšak v modeli je záporný no nevýznamný koeficient β z čoho vyvodíme, že v modeli neprijímame hypotézu o prítomnosti príjmovej beta konvergencii. Rovnaký graf pre rezíduá SEM sa môže vykonať ako pri OLS a SAR.

```
nuts.poly@data$nuts.sem.res<-resid(errorsalm.nuts)
spplot(nuts.poly,"nuts.sem.res",at=seq
      (min(nuts.poly@data$nuts.sem.res,na.rm=TRUE),
      max(nuts.poly@data$nuts.sem.res,na.rm=TRUE), length=12),
      col.regions=rev(brewer.pal(11,"RdBu"))))
```

Obr. 6 SEM model – rezíduá



Zdroj: Vlastné spracovanie

Nakoniec, ak sa pozrieme na *likelihood* pre SAR model a SEM model, zistíme, že dosahujeme vyššiu hodnotu pre model SAR, ktorý bol modelom zvýhodneným LM testom vid' tabuľku č. 1. Vyššie uvedený graf rezíduí stále vykazuje určitú prítomnosť priestorovej autokorelácie.

Tabuľka č. 1 Výsledky Lagrangeových multiplikátorov

	value	p- value
LMerr	183,770	2,20E-16
LMlag	210,51	2,20E-16
RLMerr	1,3265	2,49E-01
RLMlag	28,066	1,17E-07

Zdroj: Vlastné spracovanie

Záver nášho modelovania je zhrnutý v tabuľke č. 2, kde sme zosumarizovali výstupy modelov OLS, SAR, SEM. Po potvrdení predpokladu prítomnosti autokorelácie *Moranovou štatistikou I* medzi oblasťami, sme dospeli k záveru, že model OLS odhadnutý metódou najmenších štvorcov nie je vhodným modelom pre odhad príjmovej beta konvergenzie. Jeho odhady môžu byť v našom prípade nekonzistentné a skreslené, takže pokračujeme modelmi kde je zohľadnená prítomnosť priestorovej závislosti. Test Lagrangeových multiplikátorov, ktorého výsledky vidíme v tabuľke č. 1 navrhuje ďalej pokračovať s modelom SAR, no napriek tomu sme odhadli oba modely.

Tabuľka č. 2 Zhrnutie parametrov modelov OLS, SAR, SEM

	OLS model	SAR model	SEM model
<i>Intercept</i>	1,35411	0,400415	0,363746
β	-0,12084	-0,03655	-0,022395
p- value	9,99E-11	0,0017623	0,18947
ρ	-	0,74675	-
p- value	-	2,22E-16	-
λ	-	-	0,77148
p- value	-	-	2,22E-16
R^2	0,1494	-	-
<i>Log likelihood</i>	-	280,8859	276,9868
AIC	-	-553,77	-545,97
<i>s</i>	0,128788373	0,037234687	0,022649576
<i>h</i>	5,382063344	18,61563084	30,60309739

Zdroj: Vlastné spracovanie

Model SAR má koeficient β štatisticky významný, čo vypovedá o prítomnosti príjmovej beta konvergenzie. Koeficient ρ , ktorý vyjadruje v SAR silu priestorovej závislosti, je v modeli štatisticky významný takže autokorelácia je prítomná. Model SEM má koeficient β štatisticky nevýznamný a koeficient λ , vyjadrujúci prítomnosť priestorovej závislosti je významný a kladný, takže v modeli je priestorová autokorelácia, no nie je potvrdená

hypotéza o prítomnosti príjmovej konvergenzie. Z výsledkov vyvodíme závery, že model zvýhodnený testom Lagrangeových multiplikátorov je pre náš odhad lepší, pretože potvrdil prítomnosť priestorovej závislosti a zároveň hypotézu o príjmovej beta konvergencii. Pre výber vhodnejšieho modelu v prípade, že by sme nerozhodli už na základe štatistickej významnosti parametrov, by slúžil na porovnanie modelov aj *Log likelihood*, ktorý je pre model preferovaný testom Lagrangeových multiplikátorov lepší ako pre model SEM. Rýchlosť konvergenzie vypočítaná modelom je rôzna pre všetky tri modely. Koeficient β zachytáva mieru, s akou sa reálny *HDP krajiny na obyvateľa* približuje rýchlosti rastu v rovnovážnom stave, t. j. rýchlosti konvergenzie. V modeli SAR sme vypočítali, že oblasti konvergujú rýchlosťou 3,72 %. Pričom polčas konvergenzie, definovaný ako počet rokov, ktoré sú potrebné na zníženie príjmovej medzery na polovicu je približne 19 rokov.

Záver

Cieľom našej práce bolo dokázať, že medzi členskými štátmi Európskej únie existuje priestorová závislosť, teda korelácia, ktorú môžeme ďalej rozvinúť a skúmať. Na lineárnom regresnom modeli príjmovej beta konvergenzie sme testovali priestorovú závislosť pomocou *Moranovej štatistiky I*, ktorá potvrdila prítomnosť korelácie. To znamená, že členské štáty Európskej únie, sa vzájomne ovplyvňujú, teda existuje medzi nimi závislosť.

Ďalším krokom bolo vytvorenie a odhadnutie správne zvoleného modelu, ktorý zahŕňa predpoklad prítomnosti priestorovej závislosti. Predstavili sme modely, ktoré tieto kritéria spĺňajú a testy, ktoré poukazujú na vhodný výber modelu. Pomocou Lagrangeových multiplikátorov sme zhodnotili výber „lepšieho“ modelu, avšak aj napriek multiplikátorom sme okrem preferovaného modelu SAR odhadli aj model SEM. Popri Lagrangeových multiplikátorov sme model vyberali aj na základe parametra *Log likelihood*. Podľa spomínanej veličiny sme opäť vybrali model SAR, ktorý sa javí ako vhodný model na odhad príjmovej beta konvergenzie zohľadňujúcej priestorovú závislosť.

Naše výpočty na odhadoch dát teda potvrdzujú prítomnosť príjmovej beta konvergenzie v členských štátoch Európskej únie. Čo znamená, že menej rozvinuté krajiny rástli v sledovanom období rýchlejšie ako bohatšie krajiny Európskej únie. Ako uvádzame v tabuľke č. 2, výsledky získané metódou najmenších štvorcov t.j. metódou, ktorá nezohľadňuje špecifiká jednotlivých krajín, naznačujú, že dobiehanie prebieha tempom 12,87 % a polčas konvergenzie je približne 5 rokov. Môžeme teda konštatovať, že rýchlosť konvergenzie podľa „obyčajného“ modelu je značne nadhodnotená, a teda vedie k chybným záverom. Zanedbanie priestorovej autokorelácie vedie aj k skresleným a neefektívnym odhadom. Avšak odhad pomocou metódy maximálnej vierohodnosti hovorí, že tempo dobiehania je 3,70 % pričom polčas konvergenzie je odhadovaný približne na 19 rokov. Táto metóda na rozdiel od metódy najmenších štvorcov zohľadňuje špecifiká jednotlivých krajín, a teda je lepšia pre odhad modelov priestorovej ekonometrie. Je vidieť, že výsledky metódy zohľadňujúcej priestorovú závislosť vypovedajú o reálnych skutočnostiach. Potvrdením predpokladu o konvergencii krajín Európskej únie poukazujú výsledky práce aj na správne fungovanie regionálnej politiky, ktorej hlavným cieľom je konvergenca.

Prostredníctvom programu R Studio sme v našej práci vytvorili *skript*, kde sme zadefinovali použité kódy, pomocou ktorých sme detailne preskúmali teóriu príjmovej beta konvergenzie. V ňom sme sa zamerali na odhad modelov príjmovej beta konvergenzie, overenie predpokladov modelov, preskúmali sme prítomnosť vzťahov medzi oblasťami pomocou

Moranovej I štatistiky, vytvorili potrebnú maticu váh pomocou kritéria „Queen“ a „Rook“, ktoré sme zobrazili graficky. Prostredníctvom kódov sme odhadli konkrétne 3 modely, a to OLS model bez priestorovej závislosti medzi pozorovaniami, ďalej modely SAR a SEM, ktoré zohľadňujú prítomnosť priestorovej závislosti. Graficky sme zobrazili rezíduá všetkých spomínaných modelov. Zobrazili sme, ako sa v programe R vytvorí test Lagrangeových multiplikátorov a následne jeho výstup, ktorý slúži na výber vhodného modelu. Ukázali sme ako dosiahnuť prostredníctvom R interpretáciu modelov pomocou priameho a nepriameho vplyvu. Ak hovoríme o priamom vplyve ide o situáciu, keď zmena hodnoty vysvetľujúcej premennej v jednom regióne ovplyvní samotný región naopak, ak hovoríme o nepriamom vplyve zmena hodnoty vysvetľujúcej premennej v jednom regióne ovplyvní nepriamo aj ďalšie regióny. V prípade použitia iných dát, je tento *skript* vhodný a flexibilný na ďalšie použitie, pre aplikáciu priestorovej ekonometrie pre nezávislého užívateľa. Táto práca mu môže slúžiť ako návod respektíve inšpirácia pre ďalšie použitie s vlastnými údajmi, pričom autor práce bez výhrad poskytne na vyžiadanie kompletný *skript* pre ďalšie využitie.

Zoznam použitej literatúry

ANSELIN, L. 1988c. *Spatial econometrics: methods and models*. Dordrecht: Kluwer Publisher, 1988. 284 p. ISBN 978-94-015-7799-1.

ANSELIN, L. – BERA, K. A. 1998. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In Ullah, A. – Giles, D. E. A. 1998. *Handbook of Applied Economic Statistics*. p. 237 – 289. ISBN-13: 978-0824701291.

ANSELIN, L. – FLORAX, R. J. 1995a. *New directions in spatial econometrics*. Berlin: Springer-Verlag, 1995. 420 p. ISBN 978-3-642-79877-1.

ANSELIN, L. 2006. Spatial Econometrics. In Mills, T. C. –Patterson, K. 2006. *Palgrave Handbook of Econometrics*. Basingstoke: Palgrave Macmillan, p. 901-941. ISBN 978-14055-4155-8.

ANSELIN, L. (2005): Exploring Spatial Data with GeoDaTM: AbWorkbook Center for Spatially Integrated Social Science

ANSELIN, Luc. 2003. “An Introduction to Spatial Regression Analysis in R.” Available at: https://Geodacenter.asu.edu/Drupal_files/Spdepintro.pdf.

ARBIA, Giuseppe. 2014. *A Primer for Spatial Econometrics: With Applications in R*. Palgrave Macmillan.

BIVAND, Roger S, Edzer J PEBESMA, and VIRGILIO Gomez-Rubio. 2008. *Applied Spatial Data Analysis with R*. Springer.

BIVAND, Roger, and Nicholas LEWIN-KOH. 2016. *Maptools: Tools for Reading and Handling Spatial Objects*. <https://CRAN.R-project.org/package=maptools>.

BIVAND, Roger, and Gianfranco PIRAS. 2015. “Comparing Implementations of Estimation Methods for Spatial Econometrics.” *Journal of Statistical Software* 63 (18): 1–36. <http://www.jstatsoft.org/v63/i18/>.

BARRO, R. J. (1989): Economic Growth in a Cross Section of Countries *The Quarterly Journal of Economics*, MIT Press, vol. 106(2), p. 407-43.

BARRO, R. J. – SALA-I-MARTIN, X. 1991. Convergence across States and Regions. In *Brookings Papers on Economic Activity*. 1991, no. 1, p. 107–182.

BENEDEK, J. & KOCZISZKY, Gy. (2014): Paths of economic and social convergence and polarization in the Visegrad-countries – manuscript

CLIFF, A. D. and J. K. ORD. (1969). “The problem of Spatial autocorrelation.” In *London Papers in Regional Science 1, Studies in Regional Science*, 25–55, edited by A. J. Scott, London: Pion.

CHENG, Joe, and Yihui XIE. 2015. *Leaflet: Create Interactive Web Maps with the Javascript 'Leaflet' Library*. <http://rstudio.github.io/leaflet/>.

- DAVIES, S. & HALLET, M. (2002): Interactions between National and Regional Development Hamburgisches Welt-Wirtschafts-Archiv, Discussion Paper, Nr. 207.; 30 p.
- ECKEY, H. F. & TÜRCK, M. (2007): Convergence of EU-Regions. A Literature Report *Investigaciones Regionales*, Nr. 10., pp. 5-32.
- FURKOVÁ, A. 2013. Modely priestorovej ekonometrie - model SAR a model SEM. In *Využitie kvantitatívnych metód vo vedecko-výskumnej činnosti a v praxi X: zborník príspevkov zo seminára : 29. – 31. máj 2013, Čingov – Slovenský raj*. Bratislava: Ekonóm, 2013. ISBN 978-80-225-3646-2, s. 1–6.
- GERKMAN, L. & AHLGREN, N. (2011): Practical Proposals for Specifying k-Nearest Neighbours Weights Matrices Hanken School of Economics, Working Papers 555.; 29. p.
- BARRO, R. J., & SALA-i-MARTIN, X. 1992. Convergence. *Journal of political Economy*, 223-251.
- HUBERT, L. J., R. G. GOLLEDGE, and C. M. COSTANZA. (1981). “Generalized Procedures for Evaluating Spatial Autocorrelation.” *Geographical Analysis* 13, 224–32.
- KELEJIAN, H.H. and PRUCHA, I.R. (2010) “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances”, *Journal of Econometrics*, 157: 53-67.
- LEE L-F, Yu J (2009) Spatial nonstationarity and spurious regression: The case with a row-normalized spatial weights matrix. *Spatial Economic Analysis* 4: 301–327
- LESAGE, J. – PACE, K. 2009. *Introduction to Spatial Econometrics*. Boca Raton: Chapman and Hall/CRC, 2009. 374 p. ISBN 81420064247.
- MANKIW, N. G., ROMER, D. and WEIL, D., 1992. A contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics* 107, 407–37.
- MORAN, P. A. P. 1950. Notes on continuous stochastic phenomena. In *Biometrika*. 1957, vol. 37, no. 1/2, p. 17-23.
- NEURWITH, Erich. 2014. *RColorBrewer: ColorBrewer Palettes*. <https://CRAN.R-project.org/package=RColorBrewer>.
- PACE, R Kelley, and JP LESAGE. 2009. “Introduction to Spatial Econometrics.” *Boca Raton, FL: Chapman & Hall/CRC*.
- PAELINCK J, KLAASEN L (1979) Spatial econometrics. Saxon House, Farnborough
- SOLOW, R. M., 1956. A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics* 70, 65–94.
- TOBLER, WR. 1979. “Cellular Geography.” In *Philosophy in Geography*, 379–86. Springer.
- WHITTLE, P. (1954). “On Stationary Processes in the Plane.” *Biometrika* 41, 434–49.