

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

**Využitie priestorovej ekonometrie na modelovanie
hrubého domáceho produktu SR**

2010

Bc. Katarína Kuricová

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Využitie priestorovej ekonometrie na modelovanie
hrubého domáceho produktu SR

Diplomová práca

Študijný program : Ekonomická a finančná matematika

Študijný odbor: 9.1.9 Aplikovaná matematika

Školiteľ: Mgr. Ing. Michal Páleník

Bratislava 2010

Bc. Katarína Kuricová

ČESTNÉ PREHLÁSENIE

Čestne prehlasujem, že som diplomovú prácu vypracovala samostatne s použitím uvedenej literatúry.

V Bratislave dňa 2.8.2010

.....
Katarína Kuricová

POĎAKOVANIE

Touto cestou by som chcela poďakovať svojmu diplomovému vedúcemu Mgr. Ing. Michalovi Páleníkovi za cenné rady, trpezlivosť, ochotu a čas strávený konzultáciami.

ABSTRAKT

KURICOVÁ, Katarína: Využitie priestorovej ekonometrie na modelovanie hrubého domáceho produktu SR [Diplomová práca] – Univerzita Komenského v Bratislave. Fakulta matematiky, fyziky a informatiky. Katedra aplikovanej matematiky a štatistiky. Diplomový študijný odbor: aplikovaná matematika – ekonomická a finančná matematika. – Vedúci diplomovej práce: Mgr. Ing. Michal Páleník. – Bratislava: FMFI UK, 2010. – 56 s.

Hlavným cieľom tejto práce je zadefinovanie a vysvetlenie metodológie priestorovej ekonometrie. Vysvetliť pojmy priestorový susedia a priestorové váhy a aké typy modelov sa v priestorovej ekonometrii používajú. Základným predpokladom využitia priestorovej ekonometrie je prítomnosť priestorovej autokorelácie medzi jednotlivými územnými jednotkami. Keďže je tento predpoklad splnený pre okresy Slovenskej republiky, budeme sa v praktickej časti venovať práci s reálnymi dátami. Budeme odhadovať hrubý domáci produkt v jednotlivých okresoch použitím štandardnej (nepriestorovej) a priestorovej ekonometrie. Na záver urobíme predikciu.

Kľúčové slová: HDP, priestorová autokorelácia, panelové dáta, priestorová ekonometria, SAR model

ABSTRACT

KURICOVÁ, Katarína: Application of spatial econometrics in modeling of gross domestic product of Slovak Republic [Master thesis] – Comenius University in Bratislava. Faculty of Mathematics, Physics and Informatics. Department of Applied Mathematics and Statistics. Branch of study: Applied Mathematics – Economic and Financial Mathematics. – Supervisor: Ing. Mgr. Michal Páleník. – Bratislava: FMFI UK, 2010. – 56p.

The principal aim of this paper is to define and explain the methodology of spatial econometrics. To explain concepts of spatial neighbours and spatial weights, and what types of models are used in spatial econometrics. The underlying assumption of using spatial econometrics is presence of spatial autocorrelation between individual units. Since this assumption is fulfilled for districts of the Slovak Republic we will work in the practical section with real data. We estimate the gross domestic product in single districts using standard (non-spatial) and spatial econometrics. In the end we make prediction.

Keywords: GDP, spatial autocorrelation, panel data, spatial econometrics, SAR model

OBSAH

Úvod	10
1. Teoretický základ a terminológia	12
1.1 Hrubý domáci produkt	12
1.2 Trh práce	12
1.3 Rómska otázka	14
2. Základy priestorovej ekonometrie	17
2.1 Priestorový susedia (Spatial Neighbours) a priestorové váhy (Spatial Weights)	18
2.1.1 Priestorový susedia (Spatial Neighbours).....	18
2.1.1.1 Konštrukcia matice susednosti podľa spoločnej hranice.....	19
2.1.1.2 Konštrukcia matice susednosti na základe vzdialenosti medzi regiónmi	20
2.1.2 Priestorové váhy (Spatial Weights)	20
2.1.2.1 Transformácia nazývaná „riadková normalizácia“ (row standardisation).....	21
2.2 Priestorová autokorelácia a jej testovanie	22
2.3 Modelovanie v priestorovej ekonometrii	23
2.3.1 Priestorové modely	23
2.3.1.1 SAR model.....	24
2.3.1.2 SEM model.....	25
3. Priestorová ekonometria pre modely s panelovými dátami	27
3.1 Panelové dáta	27
3.2 Jednoduché modely pre panelové dáta	28
3.2.1 Model fixných efektov (Fixed Effects Model).....	29
3.2.2 Model náhodných efektov (Random Effects Model)	29
3.2.3 Hausmanov test.....	30
3.3 Modely pre panelové dáta s priestorovou závislosťou	31
3.3.1 SAR model s fixnými a náhodnými efektmi pre panelové dáta.....	31
3.3.2 SEM model s fixnými a náhodnými efektmi pre panelové dáta	31
4. Praktická časť	33
4.1 Dáta	33
4.2 Model	34
4.2.1 Nepriestorový model s fixnými a náhodnými efektmi	34
4.2.1.1 Model s fixnými efektmi	35
4.2.1.2 Model s náhodnými efektmi.....	36
4.2.2 Priestorový model	38
4.2.2.1 Testovanie výskytu priestorovej autokorelácie v dátach	38
4.2.3 Zhrnutie a porovnanie výsledkov modelov	40
4.2.3.1 Lineárny model	40
4.2.3.2 Logit model.....	41
4.3 Predikcie	41
4.3.1 „Ex- post“ predikcia	42

4.3.2 Predikcia pre zvýšenie produktivity rómskeho obyvateľstva	44
4.4 Modelovanie v softvéri R	45
Záver	48
Zoznam použitej literatúry	50
Príloha 1 : Transformácia prierezových dát na panelové urobená v softvéri STATA..	53
Príloha 2: Odhadovanie modelov – zdrojový kód	54
Príloha 3: Zdrojové kódy predikcií pre priestorové modely	55

ZOZNAM OBRÁZKOV, GRAFOV A TABULIEK

Obrázok 1: <i>Odhad nepriestorového modelu s fixnými efektmi</i>	35
Obrázok 2: <i>Odhad nepriestorového modelu s náhodnými efektmi</i>	36
Obrázok 3: <i>Výsledok Hausmanovho testu</i>	37
Obrázok 4: <i>Odhad nepriestorového modelu s fixnými efektmi po vynechaní okresov</i> ..	37
Obrázok 5: <i>Odhad nepriestorového logit modelu</i>	38
Obrázok 6: <i>Odhad pre priestorový model s vynechaním okresov</i>	39
Obrázok 7: <i>Odhad priestorového logit modelu</i>	39
Obrázok 8: <i>Odhad nepriestorového modelu rokov 2001-2006</i>	42
Obrázok 9: <i>Odhad priestorového modelu rokov 2001- 2006</i>	42
Obrázok 10: <i>Graf okresov SR z načítaných dát zo súboru typu “shapefile“</i>	46
Graf 1: <i>Percentuálne odchýlky odhadu HDP v roku 2007 od reálnych dát pre obidva typy modelov</i>	43
Graf 2: <i>Percentuálny nárast HDP modelu s vyššou produktivitou oproti modelu s pôvodnou produktivitou minoritnej skupiny obyvateľstva</i>	44
Tabuľka 1: <i>Národnostné zloženie obyvateľov</i>	15
Tabuľka 2: <i>Národnostné zloženie obyvateľov</i>	15
Tabuľka 3: <i>Koeficienty z odhadov lineárnych modelov</i>	40
Tabuľka 4: <i>Koeficienty z odhadov logit modelov</i>	41

Úvod

História priestorovej štatistickej a ekonomickej analýzy dát siaha až na koniec 40.-tych a začiatok 50.-tych rokov 20. storočia. Pojem priestorová ekonometria vymyslel začiatkom 70.-tych rokov holandsko-belgický ekonóm Jean Paelinck. Tento ekonóm, spoločne s Klaassenom napísali prvú publikáciu o priestorovej ekonometrii v roku 1979, v ktorej poukazujú na priestorové vzťahy a priestorovú závislosť. Hoci je tento vedný odbor vo svete známy už takmer tretinu storočia, u nás sa o ňom v súčasnosti stále veľa nevie alebo lepšie povedané, nevyužíva sa. Keďže okresy Slovenska spĺňajú všetky predpoklady na to, aby medzi nimi existovala priestorová závislosť, za hlavný cieľ práce sme zvolili priblíženie metodológie priestorovej ekonometrie a jej využívanie.

Hrubý domáci produkt je jedným z najhlavnejších makroekonomických ukazovateľov. Mnoho svetových organizácií ho používa ako meradlo toho, aká vyspelá krajina je. Dalo by sa povedať, že úroveň HDP dáva „meno“ a určuje úroveň svojej krajiny. Týmto makroekonomickým ukazovateľom sa budeme zaoberať aj my v diplomovej práci. Pokúsime sa aplikovať metodológiu priestorovej ekonometrie na modelovanie tejto veličiny v okresoch SR v rokoch 2001 až 2007, a teda ukázať priestorovú závislosť medzi okresmi SR.

Naša práca je rozdelená do štyroch jednotlivých kapitol. Prvé tri kapitoly sú zamerané na teoretickú stránku a poslednou kapitolou je praktická časť.

Prvá kapitola sa týka definícií makroekonomického ukazovateľa HDP a vybraných ekonomických ukazovateľov trhu práce a neskôr sa týka pojmov a štatistik minoritnej skupiny obyvateľstva rómskej národnosti žijúcej v Slovenskej republike. V úvode si povieme, čo rozumieme pod pojmom HDP a ako sme sa dostali k pojmu priestorová ekonometria. Ďalej sa venujeme pojmom ako zamestnanosť (zamestnaní) a nezamestnanosť (nezamestnaní), z ktorých sú zložené pracovné sily štátu alebo ekonomicky aktívne obyvateľstvo. Venujeme pozornosť aj veľmi dôležitému ukazovateľu trhu práce, a tým je miera nezamestnanosti. V závere druhej podkapitoly

venujeme pár slov produktivite práce. Koniec tejto kapitoly je venovaný demografii a dostupným štatistikám minoritnej skupiny, ktorou sú Rómovia.

Nadpis druhej kapitoly „Základy priestorovej ekonometrie.“ presne vystihuje to, čomu sa v nej budeme venovať. Na začiatok si vysvetlíme, čo to tá priestorová ekonometria vlastne je a aké druhy existujú. Povieme si, čo rozumieme pod pojmami priestoroví susedia (spatial neighbours) a priestorové váhy (spatial weights), a tiež, čo rozumieme pod priestorovou autokoreláciou, ktorá musí byť prítomná, aby sme priestorovú ekonometriu mohli využívať. Na záver si definujeme základné typy modelov, ktoré sa v tomto odbore používajú.

V ďalšej kapitole je zaznamenaný prechod medzi priestorovou ekonometriou pre priestorové modely a modelmi s panelovými dátami. Je tu povedané, čo panelové dáta sú a akú majú štruktúru. Nasledujúcim krokom tohto prechodu je podanie jednoduchých modelov pre panelové dáta, kde sú zahrnuté pojmy fixný a náhodný efekt. A prepojením základných typov modelov v priestorovej ekonometrii a jednoduchých modelov pre panelové dáta sa dostávame k cieľu tejto kapitoly, k modelom s panelovými dátami s priestorovou závislosťou.

Poslednou časťou tejto práce je praktická časť, v ktorej budeme pracovať s reálnymi dátami a využívať modely opísané v predošlých kapitolách. Cieľom je definovať model, pre ktorý sú dostupné dáta, urobiť odhady a prezentovať výsledky. Predposlednou súčasťou bude urobenie dvoch typov predikcií a na záver krátky manuál pre priestorovú ekonometriu v softvéri R, ktorý budeme používať.

1. Teoretický základ a terminológia

V tejto kapitole sa budeme venovať vysvetleniu pojmov, ktoré sú používané v práci alebo sa vyskytujú v odhadovaných modeloch a týkajú sa trhu práce a HDP. Vysvetlíme, prečo sme sa rozhodli v práci využívať priestorovú ekonometriu. Takisto sa budeme zaoberať rómskou otázkou, merateľnosťou dát, ich demografiou, zamestnanosťou.

1.1 Hrubý domáci produkt

HDP vyjadruje trhovú hodnotu výrobkov a služieb vyprodukovaných v krajine.¹ Zahŕňa príjmy zarobené na území danej krajiny, nie je však podstatné, kto je vlastníkom. ŠÚSR eviduje tieto hodnoty podľa regiónov na úrovni NUTS2 a NUTS3. Aby sme sa dostali do úrovne LAU bolo potrebné previezť niekoľko výpočtov.

Porovnávanie HDP na úrovni okresov je zložité. Príčinou je to, že tieto územné jednotky majú malú rozlohu a navzájom svoje HDP ovplyvňujú. Je to spôsobené tým, že obyvatelia jedného okresu nemajú problém cestovať za prácou do susedného, a tým prispievať k tvorbe HDP okresu susedného a nie tomu, v ktorom sú evidovaní. Toto zistenie nás vedie v práci k tomu, aby sme využili **priestorovú ekonometriu**. Čo znamená, že do odhadu modelu zahrnieme vzájomné ovplyvňovanie sa susednými okresmi. Teoretickej stránke priestorovej ekonometrie sa budeme venovať v ďalších kapitolách.

1.2 Trh práce

V ekonómii, ekonometrii, politike je odjakživa často riešenou a rozoberanou témou zamestnanosť a nezamestnanosť. Aj my sa týmito elementmi trhu práce budeme

¹ Podľa ŠÚSR.

zaoberať. Preto je nutné si definovať podľa akých definícií² a kritérií ich budeme skúmať a merať.

Počet pracovných síl, ktorý budeme využívať aj ďalej v práci pri definovaní miery nezamestnanosti, sa skladá z počtu zamestnaných a nezamestnaných obyvateľov.

Do počtu nezamestnaných sa zahŕňajú osoby v predpísanom veku (od 15 rokov), ak spĺňajú počas určitej časovej periódy aspoň jednu z týchto troch podmienok:

- sú bez práce, t.j. neboli v platenej činnosti alebo samostatnej zárobkovej činnosti.
- sú prístupní pre prácu, t.j. k dispozícii pre platenú činnosť alebo samostatnú zárobkovú činnosť.
- hľadajú si prácu, t.j. sú aktívni v hľadaní si plateného zamestnania alebo samostatnej zárobkovej činnosti. Táto aktivita znamená, že sa registrovali na súkromnom alebo verejnom úrade práce; kontrolovanie pracovných miest; vypisovanie alebo odpovedanie na inzeráty; atď.

Do počtu zamestnaných zarátavame všetky osoby od určitého veku (od 15 rokov), ktoré počas určitého krátkeho obdobia (týždeň alebo aj deň) patrili do jednej z týchto kategórií:

- majú platenú prácu alebo samostatnú zárobkovú činnosť, t.j. sú to osoby, ktoré pracujú, resp. vykonávajú samostatnú zárobkovú činnosť, za plat alebo mzdu, resp. pre zisk, v hotovosti alebo v naturáliách.
- majú prácu alebo samostatnú zárobkovú činnosť, ale momentálne nepracujú, t.j. osoby, ktoré už pracovali vo svojej práci alebo podniku, ale momentálne z nejakého dôvodu nepracujú (choroba, štrajk, dovolenka), no vyplácanie mzdy alebo jej časti pokračuje.

Súčet počtu zamestnaných a nezamestnaných sa dá okrem počtu pracovných síl nazvať ešte ako ekonomicky aktívne obyvateľstvo. Pod *ekonomicky aktívnym obyvateľstvom*³ rozumieme osoby vo veku od 15 rokov, ktoré patria medzi

² Definície platné podľa Medzinárodnej organizácie práce (International Labour Organization- ILO) od roku 1982. S týmito definíciami sa stotožňuje aj Štatistický úrad Slovenskej republiky

³ Podľa ŠÚSR

zamestnaných v civilnom sektore, nezamestnaných alebo príslušníkov ozbrojených zložiek.

Najčastejším ukazovateľom, ktorým sledujeme nezamestnanosť, je *miera nezamestnanosti*. Táto miera, označovaná ako u , má vzťah na výpočet v percentách daný ako: $u = \frac{U}{L} \times 100$, kde U je počet nezamestnaných a L je počet pracovných síl. Na Slovensku existujú dva prístupy sledovania miery nezamestnanosti. Jedným je Výberové zisťovanie pracovných síl ŠÚSR (Štatistického úradu Slovenskej republiky) a druhým sú údaje evidované na úradoch práce, ktoré vedie UPSVAR (Ústredie práce, sociálnych vecí a rodiny). Tieto štatistiky sa navzájom líšia ako na Slovensku, tak aj v zahraničí. My sme využívali obe, aby sme plne využili všetky dostupné dáta. V našom prípade sme do miery nezamestnanosti zahrnuli aj tých, ktorí rezignovali na to, aby si našli zamestnanie.

*Produktivita práce*⁴ je celkový výstup delený pracovný vstupmi. Zvyšuje sa v dôsledku dokonalejších technológií, vyššou pracovnou zručnosťou a prehlbovaním kapitálu. My sme si v našej práci zvolili vlastnú produktivitu, a to produktivitu, ktorou každý zamestnaný v príslušnom okrese prispieva k tvorbe HDP. Produktivita závisí od viacerých faktorov: od prírodných a klimatických podmienok, použitých technológií, kvalifikácie a motivácie pracovníkov. Od týchto faktorov závisí aj veľkosť produktivity pre jednotlivé okresy.

1.3 Rómska otázka⁵

V súčasnosti sa veľké množstvo organizácií a s tým spojených výskumov vo väčšine krajín Európy zaoberá riešením rómskej otázky. Čo sa týka vzdelávania detí tejto menšiny v krajinách severnej Európy sa zdá byť v súčasnosti tento problém vyriešený a o pár rokov neskôr sa uvidí, čo sa stane z ich zamestnanosťou, pričom u nás je to ešte hlavnou témou debát. My sa v našej práci budeme tiež venovať tejto minoritnej skupine obyvateľstva, a to z hľadiska zamestnanosti. Nebudeme však hľadať

⁴ Prebrané z definície zo slobodnej encyklopédie Wikipédia

⁵ Zdrojom a námetom k tejto kapitole bola práca Marcinčin, Marcinčinová (2009) a Súhrnná správa o Rómoch na Slovensku, Vašečka(2002)

riešenie ako zvýšiť vzdelanosť alebo zamestnanosť, ale ako vplýva produktivita práce tejto menšiny v jednotlivých okresoch na HDP Slovenskej republiky a následne, čo sa stane s veľkosťou HDP, keby táto veličina mala tendenciu rasti. V tejto podkapitole si položíme zopár otázok k vymedzeniu pojmov, následne k demografii a na záver k zamestnanosti, resp. nezamestnanosti.

V súčasnosti je veľký problém v tom, že na Slovensku neexistujú oficiálne štatistiky o etnickom zložení obyvateľstva. Sú k dispozícii len do roku 1989, resp. 1991. Po týchto rokoch sú takéto údaje v rozpore s ľudskými právami. Tieto štatistické údaje sú momentálne najčastejšie nahrádzané údajmi o národnostnom zložení obyvateľstva. Tu si ukážeme prípad, kde nastáva rozdiel. Porovnáme údaje v Tab.1 a 2 na okrajoch nášho sledovaného obdobia, ktoré uvádza ŠÚSR a odhady a výsledky výskumov v rómskych osadách pre počet obyvateľov slovenskej a rómskej národnosti.

Tabuľka 1: Národnostné zloženie obyvateľov

rok\národnosť	slovenská	rómska
2001	4613747	90944
2007	4613875	101960

Zdroj: ŠÚSR

Tabuľka 2: Národnostné zloženie obyvateľov

rok\národnosť	slovenská	rómska
2001	4344691	360000
2007	4295835	420000

Zdroj: odhady M.Šprochu a odhady autora

Vidíme, že údaje, ktoré uvádza štatistický úrad sú zhruba štvrtinou z toho, aké sú odhady. Dôvodom týchto rozdielov je spôsob sčítavania ľudu pred a po roku 1989 a tiež rozdiel medzi pojmom etnicita a národnosť. Náležia otázky: Ako je chápaná etnicita a národnosť? Kto je slovenskej národnosti a kto rómskej?

Dôležitým faktorom rozdielnosti týchto štatistík je fakt, že pred rokom 1989, keď boli evidované etnické štatistiky bol občan priradený k etniku na základe inej osoby. Pričom národnostné štatistiky po tomto zlomovom roku uprednostňujú dobrovoľné zaradenie sa občana k národnosti, aby sa vyhlo rasovej diskriminácii. Aby sme sa dopracovali k reálnym demografickým dátam, a teda aj dátam o ekonomicky

aktívnom a zamestnanom obyvateľstve v národnostnom zložení sú potrebné štatistické odhady.

Čo sa týka štatistík o zamestnanosti, resp. nezamestnanosti rómskej menšiny na Slovensku toho nevieme veľa povedať. Z rôznych zdrojov si však aj s týmito údajmi poradíme. Počet nezamestnaných tejto menšiny sa výrazne líši od majoritnej skupiny obyvateľstva. Až dve tretiny Rómov sú bez zamestnania, pričom sa jedná vo väčšine prípadov o dlhodobú nezamestnanosť s ukončeným základným vzdelaním.

2. Základy priestorovej ekonometrie

V tejto kapitole sa budeme venovať úvodu k priestorovej ekonometrii, základným pojmom a modelom pre jednoduché dátové štruktúry, ktoré ďalej prevedieme na zložitejšie, panelové dáta.

Rozlišujeme tri druhy priestorovej analýzy : priestorový bodový proces (Spatial Point Pattern Analysis), geoštatistiku (Geostatistics) a analýzu plošných dát (Analysing Areal Data)⁶.

Pre priestorový bodový proces platí, že je to stochastický proces, v ktorom sledujeme umiestnenie nejakých udalostí vo vymedzenej oblasti, pozorujeme množinu bodových dát v priestore. Poznáme rôzne typy bodových procesov a ich vlastnosti. Bežne sa používa v priestorovej epidemiológii, krajinnej ekológii, na analýzu trestnej činnosti a mnohých ďalších oblastiach, ale základné princípy sú rovnaké. Jeho prvý popis bol urobený na základe odhadu priestorovej štatistickej hustoty z pozorovaných dát. Pýtame sa, či dáta vykazujú nejaké priestorové delenie modelu (Spatial Clustering Pattern) alebo priestorovú náhodnosť (Complete Spatial Randomness).

Geoštatistické dáta sú dáta, ktoré sa môžu merať kdekoľvek, ale zvyčajne sú to merania na obmedzenom počte pozorovaní. Na základe týchto údajov sa pokúšame o interpoláciu dát na nepozorované body. Napríklad, pozorujeme kvalitu ovzdušia na množine monitorovacích staníc. Potom sa zaoberáme otázkou, čo nám to hovorí o kvalite ovzdušia pre celú oblasť.

Plošné dáta sú pozorované na oblastiach s definovanými hranicami. Hranice si vytvára každý sám podľa toho, čo chce pozorovať. Typickou otázkou, ktorú si pri tejto analýze kladieme je, že či to, čo sa stane v jednom regióne je ovplyvnené tým, čo sa deje v ostatných. Napríklad, či miera kriminality v jednom územnom celku závisí od miery v inom.

V našej štúdii budeme používať tretí typ analýzy, a teda budeme pracovať s plošnými dátami. Oblasti s definovanými hranicami môžu byť štáty, regióny, oblasti, ktoré si sami určíme, ale musia mať nejakú stanovenú hranicu. V našom prípade to budú

⁶ Podľa Bivand, Pebesma a Gomez- Rubio (2008).

okresy Slovenskej republiky. Predpokladáme, že v množine máme N územných celkov (N okresov).

2.1 Priestorový susedia (Spatial Neighbours) a priestorové váhy (Spatial Weights)

V tejto kapitole sa budeme zaoberať priestorovými subjektmi ako priestorovými susedmi. Základom väčšiny modelov je určiť si, či jeden región je priestorovým susedom iného. Na súvislej ploche sú všetky body susedmi ostatných, hoci niektoré môžu mať menšiu váhu, lebo sú príliš ďaleko. Existuje viacero definícií na „susedstvo“, ktoré budú rozobrané v nasledujúcej podkapitole. Tento indikátor susednosti bude matica $N \times N$ zostavená podľa pravidiel uvedených neskôr. V praxi sa však používa mierne transformovaná matica, ktorej sa budeme venovať v podkapitole 4.1.2.

2.1.1 Priestorový susedia (Spatial Neighbours)

Určenie priestorových susedov je nevyhnutným krokom k používaniu plošných dát. Najprv si definujeme vzťahy medzi pozorovaniami, ktoré majú nenulové váhy; to znamená vybrať si kritérium pre susedstvo dvoch regiónov, ktoré budeme používať.

Poznáme viacero druhov „susedstva“ dvoch regiónov. Či dva regióny susedia vieme povedať podľa toho, či majú spoločnú hranicu, vyhodnotenie založené na grafe, na vzdialenosti regiónov, podľa významnosti. My budeme uvažovať len dva z nich. Budeme zostavovať maticu susednosti na základe toho, či majú regióny nejakú spoločnú hranicu a druhý prípad susedstvo založené na vzdialenosti regiónov.

To, či regióny so sebou susedia alebo nie, zapíšeme do matice susednosti, ktorá je symetrická $N \times N$ matica (kde N je počet regiónov) s elementmi (i, j) . Základným typom takejto matice je matica, ktorej prvky sú 0 a 1⁷. Ak región i susedí s regiónom j

⁷ Mohli by sme, napríklad, použiť maticu, do ktorej by sme zapísali vzdialenosti regiónov namiesto 0 a 1.

na základe stanovených podmienok, element matice (i, j) sa rovná 1, ak nie, rovná sa 0. Diagonálne elementy tejto matice budú 0.

2.1.1.1 Konštrukcia matice susednosti podľa spoločnej hranice

Existuje veľké množstvo pravidiel, podľa ktorých sa konštruujú tieto matice. My si uvedieme dva základné⁸: susedstvo typu Rook (Rook contiguity) a susedstvo typu Queen (Queen contiguity).

Rook contiguity- dva regióny sú susedné, ak majú spoločnú nejakú časť hranice na ľubovoľnej strane. V tomto prípade je dôležité, akú dlhú tú spoločnú hranicu majú. Je daná nejaká „snap distance“ a keď je spoločná hranica dlhšia ako táto vzdialenosť, vtedy sú regióny susedmi.

Queen contiguity- je to podobný prípad ako predošlý, ale nezáleží na dĺžke hranice. Čiže dva regióny sú susedia, keď majú nejakú spoločnú hranicu. Na rozdiel od Rook contiguity, v tomto prípade budú dva regióny susedmi, aj keď budú mať spoločný len jeden bod hranice.

Môže nastať prípad, že v skúmanej oblasti sa nachádza ostrov, to znamená, že nemá so žiadnym regiónom spoločnú hranicu. V tomto prípade je najlepšie (ak je to možné) tieto regióny z dátovej štruktúry vynechať. Napríklad, v prípade Talianska vynechať ostrov Sicília a nebrať ho ako priestorového suseda. Tak isto v prípade niektorých štátov, Anglicko nie je sused s Holandskom.

Pri vytváraní matice susednosti netreba postupovať mechanicky, ale treba si zvoliť podmienky „susedstva“ na základe toho, čo ideme skúmať. Napríklad, zvolíme, že ideme používať pravidlo Rook contiguity a ideme určovať priestorových susedov. Môže sa nám stať, že podľa tohto pravidla sa nám vykreslia ako susedia regióny, ktoré nimi nie sú a naopak.

⁸ Tieto dva spôsoby používajú v práci aj Bivand, Pebesma a Gomez- Rubio (2008)

2.1.1.2 Konštrukcia matice susednosti na základe vzdialenosti medzi regiónmi

Naskytuje sa otázka, ako určíme vzdialenosť regiónov? Najprv si určíme ťažiská. Najlepšie by bolo za ťažisko regiónu, oblasti stanoviť hlavné mesto a vzdialenosti medzi nimi určiť po cestných komunikáciách v kilometroch alebo hodinách. To nám však dostupné dáta a softvér neumožňujú, a tak budeme merať vzdialenosť medzi regiónmi vzdušnou čiarou z ťažísk. Tento spôsob hľadania susedov má tiež viacero variantov.

Jedným z nich môže byť nájdenie k najbližších susedov, kde k je od 1..N-1. Vo väčšine prípadov to vedie k nesymetrickej matici, ale je isté, že každý región bude mať suseda. Určíme koľko najbližších susedov chceme k regiónu nájsť a hľadáme minimálne vzdialenosti k ostatným.

Ďalšou možnosťou je určenie si vzdialenosti, do ktorej chceme nájsť susedov každému regiónu. Vieme si vypočítať takú vzdialenosť, aby každý z regiónov mal aspoň jedného suseda. Keď budeme túto hodnotu zväčšovať, bude nám rásť aj počet susedov. Môže nastať prípad, že regióny sú nerovnomerne rozložené a musela by sa použiť taká veľká minimálna vzdialenosť, že ostatné regióny by mali príliš veľký počet susedov. Tak použijeme menšiu a vznikne nám región, ktorý žiadneho suseda nemá, to však nie je problém. V prípade, že sú objekty približne rovnomerne rozložené, nájdeme suseda pre každý región.

2.1.2 Priestorové váhy (Spatial Weights)

Na modelovanie závislostí medzi regiónmi je možné priamo použiť maticu, ktorá nám vznikla pri hľadaní priestorových susedov, no praxi sa skôr používa mierne transformovaná matica. Poznáme viacero takýchto transformácií, my si uvedieme najbežnejšie používanú transformáciu.

2.1.2.1 Transformácia nazývaná „riadková normalizácia“ (row standardisation)⁹

Je to najbežnejšia transformácia, v ktorej má každý riadok matice priestorových váh súčet rovný 1. Nech matica \tilde{W} s elementmi \tilde{w}_{ij} je matica priestorových susedov. Potom keď riadkovo normalizujem, delíme každý element v riadku súčtom elementov v riadku. To znamená, že matica váh priestorových susedov W s elementmi w_{ij} je definovaná ako

$$w_{ij} = \frac{\tilde{w}_{ij}}{\sum_j \tilde{w}_{ij}}$$

Táto matica na rozdiel od matice priestorových susedov už nie je symetrická.

Keď sa pozrieme na prípad, že sa v regiónoch, oblastiach, ktoré ideme skúmať nachádza nejaký ostrov a nemôžeme ho vynechať, upravíme vzorec pre prvky matice priestorových váh takto:

$$w_{ij} = \frac{\tilde{w}_{ij}}{\max(1, \sum_j \tilde{w}_{ij})}$$

Ďalšou možnosťou je vychádzať z konštrukcie matice zo vzdialeností medzi regiónmi. Nech d_{ij} je vzdialenosť medzi regiónmi i a j . Potom definujeme

$$\tilde{w}_{ij} = \begin{cases} 1 & d_{ij} < d \\ 0 & \text{inak} \end{cases}$$

pre preddefinované d . Znamená to, že regióny i a j sú priestorový susedia, keď sú navzájom vzdialené d jednotiek.

Ešte ďalšou možnosťou je definovanie b_{ij} ako pomer dĺžky hranice regiónu i , ktorú má spoločnú s regiónom j . Potom dostávame, že

$$\tilde{w}_{ij} = \frac{b_{ij}^\alpha}{d_{ij}^\beta}$$

kde α a β sú buď špecifikované alebo odhadované konštanty.

⁹ Podľa Viton (2010)

2.2 Priestorová autokorelácia a jej testovanie

Podstatou priestorovej analýzy je to, že čo sa deje v jednom regióne, je ovplyvnené dianím v iných regiónoch. Na toto poukazuje aj Prvý zákon geografie: „Všetko súvisí so všetkým ostatným, ale bližšie veci ešte viac“¹⁰. K tomuto môžeme pristúpiť cez pojem priestorová autokorelácia.

Priestorová autokorelácia sa vyskytuje vtedy, keď hodnoty pozorovaného znaku pre dané dve oblasti sú korelované. Ak nie sú, zamietame výskyt priestorovej korelácie. Rozlišujeme kladnú a zápornú priestorovú autokoreláciu. Kladná priestorová autokorelácia znamená, že podobné hodnoty skúmanej premennej majú tendenciu zhlukovania sa v priestore. A naopak záporná nám hovorí, že sa zhlukujú v priestore odlišné hodnoty.

Pre zistenie priestorovej autokorelácie v modeli sa zoznámime s testom, ktorý pochádza od Morana. Koeficient, ktorým určujeme prítomnosť priestorovej autokorelácie sa nazýva Moranov koeficient I , ktorý je definovaný pre konkrétny vektor x a má tvar:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

kde N je počet regiónov a \bar{x} je priemer x . Pričom očakávame, že stredná hodnota koeficientu bude $E(I) = \frac{-1}{N-1}$ a variácia má tiež určitý tvar, ale nebudeme si ho uvádzať z dôvodu zložitosti. Ak sa hodnota blíži k 1, potom hovoríme o kladnej a keď sa blíži k -1 o zápornej priestorovej autokoreláci. Ak $I=0$, tak autokorelácia nie je prítomná.

Čo sa deje, keď dáta nevykazujú priestorovú autokoreláciu? Ďalej pokračujeme s modelom bez použitia priestorovej ekonometrie. Ak je prítomná, budeme formulovať omnoho zložitejší model, model priestorový.

¹⁰ First Law of Geography : „Everything is related to everything else, but closer things more so.“ Idea W.Tobblera

2.3 Modelovanie v priestorovej ekonometrii

Priestorovú závislosť môžeme modelovať rôznymi štatistickými modelmi. V úvode si uvedieme jednoduchý lineárny model, ktorý v nasledujúcich kapitolách rozvieme. Ďalej sa budeme v kapitole venovať aj odhadom modelov pomocou maximálnej vierohodnosti.

Položíme základný model, ktorý má tvar :

$$y = X\beta + u \quad (1)$$

kde X je vektor vysvetľujúcich premenných a u je vektor náhodných chýb po dvojiciach nekorelovaných, ktorý je z normálneho rozdelenia so strednou hodnotou 0 a známou varianciou. V našej práci sa budeme zaoberať všeobecným priestorovým modelom (General Spatial Model) a od neho odvodenými modelmi (podľa toho akú závislosť medzi premennými chceme vyjadriť).

2.3.1 Priestorové modely

Všeobecný priestorový model, ktorý je uvádzaný vo veľa prácach má tvar:

$$y = \lambda Wy + X\beta + u \quad (2)$$

$$u = \rho Mu + \varepsilon$$

kde y je vektor pozorovaní vysvetľovanej premennej, X je matica pozorovaní vysvetľujúcich premenných, W a M sú matice priestorových váh (viď kapitola 2.1.2 Priestorové váhy (Spatial Weights)), λ a ρ sú koeficienty priestorovej autokorelácie, u sú priestorovo korelované reziduá a ε sú nezávislé a identicky rozdelené chyby.

Teraz si definujeme dva typy modelov: SAR (Spatial Autoregressive Model) a SEM (Spatial Error Model).

2.3.1.1 SAR model

Podstata tohto modelu je, že hodnota vysvetľovanej premennej y v regióne je ovplyvnená hodnotami v susedných regiónoch. Priestorová závislosť premennej je zahrnutá na pravej strane modelu.

Potom SAR model má tvar:

$$y = \lambda W y + X \beta + u \quad (3)$$

kde u sú priestorovo korelované reziduá a $\lambda W y$ má zmysel vtedy, keď diagonálne prvky matice W sú nulové, y na ľavej strane nebol ovplyvnený rovnakým y na strane pravej. A ľahko z tohto tvaru dostaneme redukovaný tvar, aby výsledky odhadovania neboli skreslené výskytom vysvetľovanej premennej na oboch stranách modelu :

$$y = (I - \lambda W)^{-1} X \beta + (I - \lambda W)^{-1} u .$$

Tento tvar nám môže pomôcť pri interpretácii.

Na odhad tohto modelu použijeme ML odhad (odhad maximálnou vierohodnosťou)¹¹. Model si prepíšeme do tvaru :

$$y - \lambda W y = X \beta + u ,$$

kde W je priestorová matica váh a $u \sim N(0, \sigma^2 I)$. Potom logaritmická funkcia vierohodnosti je:

$$\ln L = -\frac{N}{2} (\ln 2\pi + \ln \sigma^2) + \ln \|I - \lambda W\| - \frac{1}{2\sigma^2} (y - \lambda W y - X \beta)' (y - \lambda W y - X \beta)$$

Z nutných podmienok prvého rádu dostaneme odhady pre $\hat{\beta}$ a $\hat{\sigma}^2$:

$$\hat{\beta} = (X'X)^{-1} X' (y - \lambda W y)$$

alebo

$$\hat{\beta} = \hat{\beta}_0 - \lambda \hat{\beta}_L ,$$

kde $\hat{\beta}_0 = (X'X)^{-1} X' y$, $\hat{\beta}_L = (X'X)^{-1} X' W y$

a

¹¹ Ako sa uvádza vo Anselin (1999)

$$\hat{\sigma}^2 = \frac{1}{N} (e_0 - \lambda e_L)' (e_0 - \lambda e_L),$$

kde $e_0 = y - X\hat{\beta}_0$, $e_L = y - X\hat{\beta}_L$. Dosadením naspäť do $\ln L$ a následným maximalizovaním s ohľadom na λ dostaneme odhad metódou maximálnej vierohodnosti tohto parametra.

2.3.1.2 SEM model

V tomto modeli ovplyvnenie regiónu od jeho susedných vstupuje len cez chybu reziduí.

SEM model má tvar:

$$y = X\beta + u \quad (4)$$

$$u = \rho M u + \varepsilon$$

kde ε je z normálneho rozdelenia so strednou hodnotou 0 a varianciou $\sigma^2 I$. Riešením pre u potom dostaneme $u = (I - \rho M)^{-1} \varepsilon$ a teda transformovaný model má tvar:

$$y = X\beta + (I - \rho M)^{-1} \varepsilon.$$

Tento odhad je založený na tom, že $u \sim N(0, \sigma^2 [(I - \rho M)'(I - \rho M)]^{-1})$, potom logaritmickej vierohodnostná funkcia je:

$$\ln L = -\frac{N}{2} (\ln(2\pi) + \ln \sigma^2) + \ln \|I - \rho M\| - \frac{1}{2\sigma^2} (y - X\beta)' (I - \rho M)' (I - \rho M) (y - X\beta).$$

Skúmaním tejto funkcie vidíme, že maximalizovanie tejto funkcie je ekvivalentné minimalizácii sumy štvorcov reziduí v regresii, kde priestorová vysvetľovaná premenná je $y^* = y - \rho M y$ a priestorové vysvetľujúce premenné sú $X^* = X - \rho M X$. Nutné podmienky prvého rádu nám dávajú:

$$\hat{\beta} = [(X - \rho M X)'(X - \rho M X)]^{-1} (X - \rho M X)' (y - \rho M y)$$

a

$$\hat{\sigma}^2 = \frac{1}{N} (e - \rho M e)' (e - \rho M e),$$

kde $e = y - X\hat{\beta}$. Na dopočítanie použijeme iteratívnu metódu, ktorú navrhuje Anselin (1988):

1. Zavolať koeficient $\hat{\beta}$ a vypočítať vektor reziduí $e = y - X\hat{\beta}$.
2. Použiť toto e v logaritmickú funkciu vierohodnosti a optimalizovaním nájsť $\hat{\rho}$.
3. Použiť $\hat{\rho}$ na výpočet β_{GLS} odhadom GLS (zovšeobecnená metóda najmenších štvorcov) a potom nový vektor $e = y - X * \beta_{GLS}$.
4. Prvý raz alebo ak reziduá nie sú konvergentné, vrátiť sa na krok 2 a odhadnúť $\hat{\rho}$. Inak pokračovať na krok 5.
5. Máme odhad $\hat{\rho}$ a reziduálny vektor e , a GLS odhad beta. Môžeme teraz odhadnúť σ^2 ako $\hat{\sigma}^2 = \frac{1}{N}(e - \rho Me)'(e - \rho Me)$.

3. Priestorová ekonometria pre modely s panelovými dátami

Úvodné pojmy a základné modely k priestorovej ekonometrii potrebné k pochopeniu, o čo v priestorovej ekonometrii ide, boli uvedené v kapitolách vyššie. V tejto kapitole sa budeme venovať aplikácii priestorovej ekonometrie na panelové dáta. Na úvod si vysvetlíme, čo panelové dáta sú, prečo sa používajú, ich výhody oproti prierezovým dátam. V ďalšej podkapitole si uvedieme nepriestorové a priestorové modely s použitím panelových dát, ktorých odhad je aj cieľom tejto diplomovej práce.

3.1 Panelové dáta

Panelové dáta majú na rozdiel od dát prierezových alebo časových radov viac dimenzií. Môžu byť spojením týchto dvoch prípadov, a to tak, že jedna dimenzia je časová označovaná ako t a druhou prierezová dimenzia označovaná i a majú zložitejšiu štruktúru. Napríklad, ako v našom prípade máme pozorovanie v okresoch SR, čo je prierezová dimenzia $i = 1, \dots, N$ (N je 38 okresov) a máme tieto pozorovania pre určité časové obdobie, pre $t = 1, \dots, T$ rokov (roky 2001-2007).

Dáta v tomto tvare majú oproti časovým radom alebo prierezovým dátam niekoľko výhod¹².

1. Jednou z nich je, že nimi môžeme získať presnejší odhad modelovaných parametrov. Panelové dáta majú viac stupňov voľnosti a menšiu multikolinearitu ako časový rad ako panel s $N=1$ a prierezové dáta ako panel s $T=1$.

2. Konštruovanie a testovanie zložitejších hypotéz; kontrola vplyvu vynechaných premenných, ktoré sú korelované inými; tiež odhalenie dynamických vzťahov; odhadovanie presnejších predikcií.

3. Zjednodušenie výpočtov a štatistických záverov. Napríklad, pri analýze nestacionárnych časových radov.

¹² Ako uvádza vo svojom článku Hsiao (2005)

Pri modelovaní s panelovými dátami je dôležité vnímať rozdiel medzi „balanced“ (vyváženými) a „unbalanced“ (nevyváženými) panelovými dátami. Balanced panel je taký, v ktorom máme pre každú prierezovú jednotku rovnaký počet pozorovaní (pre každé $i = 1, \dots, N$ je T rovnaké). Pri unbalanced to tak nie je. Riešením je, že sa z unbalanced panelu urobí balanced tak, že sa vynechajú tie prierezové jednotky, pre ktoré dáta nie sú úplné. V našom prípade máme balanced panel už na začiatku.

3.2 Jednoduché modely pre panelové dáta

Základným modelom pre panelové dáta používaným v ekonometrii je lineárny model zapísaný¹³:

$$y_{it} = \alpha_{it} + \beta_{it}^T x_{it} + u_{it} \quad (5)$$

kde $i = 1, \dots, N$ je prierezový index, $t = 1, \dots, T$ je časový index. y_{it} je vysvetľovaná premenná i -tého pozorovania v čase t , x_{it} sú vysvetľujúce premenné a u_{it} je náhodná chyba so strednou hodnotou 0. Najbežnejšie používaným je, že $\alpha_{it} = \alpha$ pre všetky i a t , a $\beta_{it} = \beta$ pre všetky i a t . Potom tvar štandardného lineárneho modelu je:

$$y_{it} = \alpha + \beta^T x_{it} + u_{it} \quad (6)$$

kde α zahŕňa vynechané vplyvy jednotlivých pozorovaní.

Často predpokladáme, že chybový člen sa skladá z dvoch komponentov, ktorých jeden je špecifický len pre prierezovú zložku a nemení sa v čase. Tento model je nazývaný model “nepozorovaných efektov“ (Unobserved Effects Model) :

$$y_{it} = \alpha + \beta^T x_{it} + \mu_i + \varepsilon_{it} \quad (7)$$

Vhodné metódy odhadu pre tento model vyplývajú z vlastností chybových zložiek.

Na riešenie problému heterogenity v modeloch s panelovými dátami poznáme dva typy modelov. Tieto modely sú typmi modelov nepozorovaných efektov: Model

¹³ Ako definuje vo svojej práci Croissant a Milo (2008)

fixných efektov (Fixed Effects Model) a Model náhodných efektov (Random Effects Model).

3.2.1 Model fixných efektov (Fixed Effects Model)

Model fixných efektov má tvar:

$$y_{it} = \beta^T x_{it} + \mu_i + \varepsilon_{it} \quad (8)$$

v ktorom v konštante μ_i sú zachytené efekty premenných charakteristické pre i -te pozorovanie (i -ty okres). Chybová zložka ε_{it} zachytáva efekty nesignifikantných premenných pre i -te pozorovanie v danom čase t a je nekorelovaná s vysvetľujúcimi premennými x_{it} pre všetky i a t . A platí, že $\mu_i \sim (0, \sigma_\mu^2)$, $\varepsilon_{it} \sim (0, \sigma_\varepsilon^2 I)$, $E(\mu_i \varepsilon_{it}) = 0$.

3.2.2 Model náhodných efektov (Random Effects Model)

Štandardne predpokladáme, že efekty, ktoré ovplyvňujú vysvetľovanú premennú nie sú priamo zahrnuté do modelu, ale zahrňame ich do premennej u_{it} vyjadrujúcej náhodné vplyvy. Pre model náhodných efektov máme tvar:

$$y_{it} = \alpha + \beta^T x_{it} + u_{it} \quad (9)$$

$$u_{it} = \mu_i + \varepsilon_{it}$$

pričom zložka μ_i vyjadruje prierezový efekt a platí, že : $\mu_i \sim (0, \sigma_\mu^2)$, $\varepsilon_{it} \sim (0, \sigma_\varepsilon^2 I)$, $E(\mu_i \varepsilon_{it}) = 0$.

Tieto dva typy modelov sú tiež nazývané „one-way“ model fixných a náhodných efektov. Ďalej sa dajú rozšíriť na „two-way“ modely pridaním efektu λ_t , čo je časový efekt.

Rozdiel medzi fixným a náhodným efektom je v tom, že pre fixný efekt predpokladáme, že nie je nezávislý od vysvetľujúcich premenných a pre efekt náhodný

predpokladáme, že je nezávislý. V literatúrach sa uvádza viac kritérií rozhodovania, či použijeme fixný alebo náhodný efekt¹⁴.

- Jedným z nich je, či používame mikro-panely alebo makro-panely. Mikro-panely sú také, kde máme mnoho jednotiek, ale relatívne málo pozorovaní na jednotku. Makro-panely sú zase také, kde máme málo pozorovaných jednotiek, ale veľa pozorovaní pre každú z nich. Pre mikro-panely je zvyčajne vhodnejší náhodný efekt a pre makro fixný.
- Pri rozhodovaní, ktorý efekt použiť pri odhadovaní modelu, nám môže pomôcť Hausmanov test, ktorý je opísaný v ďalšej kapitole.
- Ďalšia možnosť je, že fixný efekt môže byť použitý, keď robíme analýzu na presne stanovenom počte pozorovaných jednotiek (napr. krajín, okresov, firiem). A naopak náhodný efekt, keď máme náhodný počet pozorovaných jednotiek.

3.2.3 Hausmanov test

Tento test je založený na porovnávaní dvoch odhadov. Klasické využitie Hausmanovho testu je rovnaké ako v našom prípade, a je to porovnanie odhadov dvoch modelov, jeden používa fixný efekt a druhý používa náhodný.

Máme dva odhady $\hat{\beta}_F$ pre fixný efekt a $\hat{\beta}_R$ pre náhodný efekt. Nulová hypotéza je, že obidva odhady sú konzistentné oproti alternatívnej hypotéze, že jeden alebo oba odhady sú nekonzistentné. Potom štatistika pre Hausmanov test má tvar:

$$H = (\hat{\beta}_F - \hat{\beta}_R)^T (Var(\hat{\beta}_F) - Var(\hat{\beta}_R))^{-1} (\hat{\beta}_F - \hat{\beta}_R),$$

kde táto štatistika je z χ^2 rozdelenia a má k stupňov voľnosti (k je dĺžka vektora $\hat{\beta}$). Tento test nám pomôže zistiť, ktorý efekt je vhodnejší použiť.

¹⁴ Niekoľko príkladov je uvedených v Baltagi (2001)

3.3 Modely pre panelové dáta s priestorovou závislosťou

Spojením kapitoly o priestorovej ekonometrii pre jednoduché priestorové modely a kapitoly o modeloch pre panelové dáta sa dostávame k ďalšej kapitole, ktorá sa bude zaoberať priestorovou ekonometriou s použitím panelových dát. V kapitole 2.3 Modelovanie v priestorovej ekonometrii sme si ukázali dva typy priestorových modelov. Tie teraz aplikujeme na modely s fixnými a náhodnými efektmi.

3.3.1 SAR model s fixnými a náhodnými efektmi pre panelové dáta

Spojením modelu (3) a (8) dostávame SAR model s fixnými efektmi a môžeme ho zapísať v tvare:

$$y_{it} = \lambda W y_{it} + X\beta + \mu_i + \varepsilon_{it} \quad (10)$$

kde W je matica priestorových váh, λ priestorový autoregresný koeficient a μ_i predstavuje fixný efekt.

Spojením modelu (3) a (9) dostaneme SAR model s náhodnými efektmi a zapíšeme ho v tvare:

$$y_{it} = \alpha + \lambda W y_{it} + X\beta + u_{it} \quad (11)$$

$$u_{it} = \mu_i + \varepsilon_{it}$$

kde W , λ sú definované rovnako ako v predošlom prípade a zložky μ_i predstavuje náhodný efekt.

3.3.2 SEM model s fixnými a náhodnými efektmi pre panelové dáta

Spojením modelov (4) a (8) dostanem SEM model s fixnými efektmi v tvare:

$$y_{it} = X\beta + \mu_i + \varepsilon_{it} \quad (12)$$

$$\varepsilon_{it} = \rho W \varepsilon_{it} + \eta_i$$

kde W je matica priestorových váh, ρ je priestorový autokorelačný koeficient a η_i je nezávislé od vysvetľujúcich premenných so strednou hodnotou 0 a známou varianciou.

Priestorový error model s náhodnými efektmi získame spojením modelov (4) a (9) a má tvar:

$$y_{it} = \alpha + X\beta + u_{it} \quad (13)$$

$$u_{it} = \mu_i + \varepsilon_{it}$$

$$\varepsilon_{it} = \rho W \varepsilon_{it} + \eta_i$$

kde W je matica priestorových váh, ρ je priestorový autokorelačný koeficient a η_i je nezávislé od vysvetľujúcich premenných so strednou hodnotou 0 a známou varianciou.

4. Praktická časť

Pojem priestorová ekonometria je veľmi obšírny. Venovanie sa v práci všetkým typom susedstiev a modelom je pre rozsah nemožné. Preto sme sa rozhodli použiť pri práci s dátami susedstvo typu 0,1, matica váh bude riadkovo normalizovaná a základným modelom bude pre nás priestorový autoregresívny model s fixnými alebo náhodnými efektmi.

V tejto kapitole sa budeme venovať práci s reálnymi dátami. Na úvod opíšeme dáta, ktoré budeme používať. Neskôr si zostavíme model, ktorý budeme testovať bez priestorovej závislosti a následne s ňou, a porovnáme výsledky. V záverečnej časti urobíme predikcie a predstavíme si odhadovanie priestorových modelov v softvéri R.

4.1 Dáta

Zdrojom dát boli rôzne štatistiky a odhady, od celoslovenských po regionálne. Je dôležité podotknúť, že v súčasnosti na skúmanie regionálnych závislostí nie je dostupné dostatočné množstvo údajov na regionálnej, okresnej úrovni, a tak si musíme vystačiť s tým, čo je voľne prístupné. Je možné, že touto skutočnosťou budú ovplyvnené aj výsledky odhadov.

Budeme používať okresné dáta v rozpätí rokov 2001-2007, pričom budeme brať do úvahy územné členenie okresov SR do roku 1996, kde je namiesto terajších 79 okresov, okresov 38, z ktorých si zostrojíme súbor dát s panelovou štruktúrou. Prierezových jednotiek bude podľa počtu okresov $N=38$ ($i= 1, \dots, 38$) a časových jednotiek bude podľa rokov 2001-2007 $T=7$ ($t= 1, \dots, 7$). Pre každý okres a príslušný rok budeme mať údaje o výške HDP v mil. eur, o počte zamestnaných, rozdelených na zamestnaných majoritnej a minoritnej rómskej skupiny obyvateľstva a údaje o produktivite, tak isto pre obe skupiny obyvateľstva. Tento súbor dát je „balanced“ panel.

4.2 Model

V našej práci sa budeme zaoberať dvoma modelmi. V prvej podkapitole sa budeme venovať modelu bez priestorovej závislosti, v druhej s priestorovou závislosťou, ktorá bude orientovaná na závislosť výšky HDP v danom okrese od HDP v susedných okresoch. Budeme odhadovať model, v ktorom budeme skúmať ako závisí hrubý domáci produkt pre okresy Slovenskej republiky s produkciou zamestnaných ľudí v okrese rozdelených na majoritnú a minoritnú skupinu obyvateľov.

Odhadneme priestorový autoregresný model s použitím fixných a náhodných efektov. Potom pomocou Hausmanovho testu rozhodneme, ktoré efekty je potrebné v prípade nášho modelu používať.

Hneď na začiatku predpokladáme, že okresy na vyššej úrovni, ako sú Bratislava-mesto a Košice- mesto, ktoré svojou produkciou a vyspelosťou vyčnievajú zo slovenského priemeru, by mohli aj v modeli spôsobovať výkyvy. Tento predpoklad si tiež otestujeme.

Predpokladáme, že pri odhadovaní modelu by nám mali vyjsť pri každej premennej kladné znamienka. Zdôvodnené je to tým, že čím viac sa produkuje, tým je väčší hrubý domáci produkt.

Po odhadnutí modelov sa dostaneme k cieľu práce, a tým je zistiť, ako sa bude správať veľkosť HDP, keď zvýšime produktivitu na strane minoritnej skupiny obyvateľstva.

Zhrnutie predpokladov:

1. Odhadované koeficienty by mali byť kladné.
2. Po predikcii by sme mali dospieť k záveru, že po zvýšení produktivity minority, nám narastie aj HDP.

4.2.1 Nepriestorový model s fixnými a náhodnými efektmi

Budeme testovať základný model, ktorého vyplýva z modelu (6):

$$HDP_{it} = \alpha + \beta * P_{it} + u_{it} \quad (14)$$

kde i označuje okresy Slovenskej republiky (38 okresov),

t označuje časovú dimenziu (roky 2001-2007),

HDP_{it} označuje hrubý domáci produkt pre určité okresy v každom roku,

P_{it} označuje produkciu majority (nerómskeho obyvateľstva) a minority (rómskeho obyvateľstva) s príslušnou produktivitou v daných okresoch a rokoch

Najprv odhadneme tento model so zahrnutím všetkých okresov a rozhodneme, ktoré efekty je potrebné používať pri odhade. Ďalej sa budeme pokúšať tento model vylepšiť a otestovať predpoklad, že vynechanie ekonomicky vyspelejších okresov by mohlo výsledky zmeniť. Tak isto urobíme odhad pre logit model, ktorý má iné vysvetlenie odhadnutých koeficientov ako model lineárny.

4.2.1.1 Model s fixnými efektmi

Do všeobecne zapísaného modelu (8) dosadíme naše skúmané premenné a dostávame model v tvare:

$$HDP_{it} = \beta * P_{it} + \mu_i + \varepsilon_{it} \quad (15)$$

```
Call:
plm(formula = model, data = data, model = "within")

Balanced Panel: n=38, T=7, N=266

Residuals :
  Min. 1st Qu.  Median 3rd Qu.  Max.
-3480.00 -124.00   -4.26   81.00  3600.00

Coefficients :
  Estimate Std. Error t-value Pr(>|t|)
P 0.0148498  0.0036321  4.0885 6.028e-05

Total sum of squares: 48939000
Residual sum of squares: 45582000
F-statistic: 16.7158 on 1 and 227 DF, p-value: 6.0284e-05
```

Obrázok 1: Odhad nepriestorového modelu s fixnými efektmi

Výsledky pre odhad tohto modelu sú v Obr.1, pričom sme do modelu zahrnuli všetky okresy Slovenska. Z výstupu vidíme, že naše dáta sú balanced panel.

4.2.1.2 Model s náhodnými efektmi

Po dosadení premenných do všeobecného modelu s náhodnými efektmi (9) dostaneme tvar nepriestorového modelu:

$$HDP_{it} = \alpha + \beta * P_{it} + u_{it} \quad (16)$$

$$u_{it} = \mu_i + \varepsilon_{it}$$

Odhadneme aj tento model so zahrnutím všetkých okresov a následne ho porovnáme s modelom s fixnými efektmi. Výsledky pre tento model sú uvedené v Obr.2.

```
Call:
plm(formula = model, data = data, model = "random")

Balanced Panel: n=38, T=7, N=266

Residuals :
  Min. 1st Qu.  Median 3rd Qu.    Max.
-2690.0 -129.0   -38.6    97.4   3930.0

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  2.0408e+02  7.5494e+01  2.7032  0.007312
P              5.4886e-03  3.4025e-04 16.1312 < 2.2e-16

Total sum of squares:    107540000
Residual sum of squares: 54156000
F-statistic: 260.215 on 1 and 264 DF, p-value: < 2.22e-16
```

Obrázok 2: Odhad nepriestorového modelu s náhodnými efektmi

Po odhadnutí modelov s oboma typmi efektov sme z Hausmanovho testu, ktorého výsledky sú v Obr.3, zistili, že pre našu dátovú štruktúru je vhodný model s použitím fixných efektov. A teda sa ďalej budeme zaoberať len odhadmi modelov s týmito efektmi.

```

Hausman Test
data: model
chisq = 6.7016, df = 1, p-value = 0.009633
alternative hypothesis: one model is inconsistent

```

Obrázok 3: Výsledok Hausmanovho testu

Pozrieme sa na výsledky po vynechaní niektorých okresov, menovite Bratislava-mesto a Košice- mesto. Výsledky sú uvedené v Obr.4. Po porovnaní koeficientu determinácie s modelom, kde sú zahrnuté všetky okresy vidíme, že model s vynechanými okresmi má túto hodnotu podstatne lepšiu. Budeme teda ďalej tieto dva okresy vynechávať.

```

Call:
plm(formula = model, data = panel, model = "within")
Balanced Panel: n=36, T=7, N=252
Residuals :
  Min. 1st Qu.  Median 3rd Qu.    Max.
-764.00 -104.00   2.21   84.80 1330.00
Coefficients :
  Estimate Std. Error t-value Pr(>|t|)
P 0.0321992  0.0028047  11.480 < 2.2e-16
Total sum of squares:    17199000
Residual sum of squares: 10663000
F-statistic: 131.799 on 1 and 215 DF, p-value: < 2.22e-16

```

Obrázok 4: Odhad nepriestorového modelu s fixnými efektmi po vynechaní okresov

Teda budeme používať model v tvare:

$$HDP_{it} = 0.0321992 * P_{it} + \mu_i + \varepsilon_{it}. \quad (17)$$

Na záver otestujeme logit model pre rovnaké okresy ako v predošlom odhade, ktorý ma odlišné vysvetlenie koeficientov ako lineárny model.

Keď porovnáme modely, ktorých výstupy sú v Obr. 4 a 5, vidíme, že znamienka koeficientov sú kladné, premenné aj model ako celok sú signifikantné a majú takmer totožné koeficienty determinácie R^2 . Rozdiel pri nich je ten, že majú iné vysvetlenie koeficientov.

```

Call:
plm(formula = logmodel, data = panel, model = "within")

Balanced Panel: n=36, T=7, N=252

Residuals :
  Min. 1st Qu.  Median 3rd Qu.  Max.
-0.5400 -0.1620  0.0197  0.1720  0.4090

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
log(P)    3.57719      0.32842  10.892 < 2.2e-16

Total sum of squares:    17.306
Residual sum of squares: 11.152
F-statistic: 118.639 on 1 and 215 DF, p-value: < 2.22e-16

```

Obrázok 5: Odhad nepriestorového logit modelu

4.2.2 Priestorový model

Ak chceme začať testovať model s použitím priestorovej ekonometrie, a teda použiť maticu susednosti okresov, je potrebné najprv otestovať, či sa medzi okresmi nachádza priestorová autokorelácia.

4.2.2.1 Testovanie výskytu priestorovej autokorelácie v dátach

Ukážeme si to na príklade dát z roku 2001 a na identifikáciu prítomnosti priestorovej autokorelácie použijeme Moranov koeficient, ktorý bol opísaný v kapitole 2.2. V tomto prípade je hodnota koeficientu I rovná 0,33, čo znamená že autokorelácia je prítomná. Rovnako to je aj pre ostatné dáta.

Teraz si môžeme zdefinovať model, kde budeme vychádzať z modelu (10), ktorý budeme odhadovať:

$$HDP_{it} = \lambda * W * HDP_{it} + \beta * P_{it} + \mu_i + \varepsilon_{it} \quad (18)$$

kde λ je priestorový autoregresný koeficient,

W je normalizovaná matica priestorových váh susednosti

a ostatné premenné sú definované ako v nepriestorovom modeli.

Budeme testovať model s použitím fixných efektov. Modelu so zahrnutím všetkých okresov sa venovať nebudeme, pretože sme si v prípade nepriestorového modelu mohli všimnúť, že výsledok pre model s vynechanými okresmi ma významne lepšiu hodnotu koeficientu determinácie. A tak sa budeme ďalej zaoberať týmto modelom aj v prípade ďalšieho odhadovania, aj pri predikciách.

Výsledky modelu s vynechanými okresmi Bratislava- mesto a Košice- mesto sú uvedené v Obrázku 6. Dosadením odhadnutých koeficientov dostaneme tvar modelu :

$$HDP_{it} = 0.76417 * W * HDP_{it} + 0.01148 * P_{it} + \mu_i + \varepsilon_{it} \quad (19)$$

```
Call:
spfeml(formula = model, data = data, listw = mat2listw(SRmat),
        model = "lag", effects = "spfe", method = "eigen")

Residuals:
    Min.   1st Qu.   Median     3rd Qu.    Max.
-620.0000  -27.6000    0.0379   32.3000   837.0000

Coefficients:
      Estimate Std. Error t-value Pr(>|t|)
rho  0.764174   0.032017  23.8679 < 2.2e-16
P    0.011484   0.014646   7.8411 4.467e-15
```

Obrázok 6: Odhad priestorový model s vynechaním okresov

Výsledky pre odhad Logit modelu sú uvedené v Obrázku 7.

```
Call:
spfeml(formula = logmodel, data = data, listw = mat2listw(SRmat),
        |model = "lag", effects = "spfe", method = "eigen")

Residuals:
    Min.   1st Qu.   Median     3rd Qu.    Max.
-0.098800 -0.019000    0.000276    0.018200    0.093900

Coefficients:
      Estimate Std. Error t-value Pr(>|t|)
rho    0.9343316  0.0096449  96.8732 < 2.2e-16
log(P) 0.5294280  0.0541765   9.7723 < 2.2e-16
```

Obrázok 7: Odhad priestorového logit modelu

V týchto dvoch modeloch sú tak isto premenné signifikantné a koeficienty kladné.

4.2.3 Zhrnutie a porovnanie výsledkov modelov

4.2.3.1 Lineárny model

Najprv sa pozrieme na výsledky odhadov nepriestorového a priestorového modelu, v ktorých sme vylúčili okresy Bratislava- mesto a Košice- mesto, ktoré sú uvedené v Obrázkoch 4 a 6. Z Obrázku 4 a hodnoty p-value, ktoré hovoria o signifikancii parametrov je možné vidieť, že premenná zohráva v modeli významnú úlohu, pretože jej hodnota je ďaleko pod 5%. To isté platí pre p-value o signifikancii pre celý model. Pre priestorový model sú tieto hodnoty porovnateľné, a tak isto je premenná aj s premennou, ktorá vyjadruje priestorovú závislosť signifikantná.

Pre nepriestorový model máme k dispozícii ešte údaje ku koeficientu determinácie, ktorý nám hovorí, na koľko percent závisí odhadovaná premenná vysvetľujúcimi . HDP na 38% vysvetlené premennou P ($R^2 = 1 - \frac{RSS}{TSS} = 0,38$).

Ako môžeme vidieť z Tabuľky 3 pri oboch modeloch je odhad pre znamienka koeficientov rovnaký. V oboch prípadoch nám vyšlo, že keď zväčšíme hodnotu vysvetľujúcej premennej, zvýši sa nám HDP. V lineárnom modeli nám veľkosť koeficientov hovorí, že ak nastane zmena vysvetľujúcej premennej o jednotku, vysvetľovaná sa zmení o koeficient. V prípade nepriestorového modelu má koeficient pri premennej P väčšie hodnoty. Je to zapríčinené tým, že do priestorového modelu je zahrnutá ešte ďalšia premenná, ktorá hovorí o priestorovej závislosti okresov. Význam koeficientu rho je teda taký, že keď sa v niektorom zo susediacich okresov zvýši HDP o jednotku, v tom druhom sa zvýši o 0,76.

Tabuľka 3: Koeficienty z odhadov lineárnych modelov

	nepriestorový model	priestorový model
priestorový koeficient rho	---	0.764174
koeficient pre P	0.0321992	0.011484

4.3.2.2 Logit model

Pre porovnanie výsledkov logit nepriestorového a priestorového modelu použijeme výstupy odhadov z Obrázkov 5 a 7. Keď sa pozrieme na p-hodnoty, ktorých hodnotami určujeme signifikanciu parametrov a modelu, všetky premenné sú bez problémov významné.

Koeficient determinácie pre nepriestorový logit model je 0,35 ($R^2 = 1 - \frac{RSS}{TSS} = 0,35$). Premenná P vysvetľuje HDP na 35%.

Čo sa týka odhadov koeficientov pre tieto modely všetky majú opäť kladnú hodnotu. Ak zväčšíme hodnotu vysvetľujúcich premenných, bude rásť aj HDP. V tomto prípade však nejde o zmenu absolútnu, ale zmenu percentuálnu. Napríklad, keď sa zvýši HDP v niektorom zo susedných okresov o jedno percento, HDP v okrese, ktoré je jeho priestorovým susedom sa zvýši o 0,84%.

Tabuľka 4: Koeficienty z odhadov logit modelov

	nepriestorový model	priestorový model
priestorový koeficient rho	---	0.9343316
koeficient pre P	3.57719	0.529428

4.3 Predikcie

V tejto podkapitole budeme robiť dve typy predikcií a každú z nich urobíme aj pre prípad nepriestorovej, aj priestorovej ekonometrie. Jedna predikcia bude, tzv. „ex-post“ predikcia a druhou bude odhad toho, ako sa bude správať hodnota HDP, keby produktivita minoritného obyvateľstva bola vyššia.

4.3.1 „Ex- post“ predikcia

Tento druh predikcie je založený na odhade hodnoty vysvetľovanej premennej a následne na porovnaní so skutočnými hodnotami. Keďže naše sledované obdobie rokov 2001- 2007 je časovo krátke, rozhodli sme sa urobiť predikciu len pre rok 2007. Z dát z rokov 2001- 2006 si odhadneme model a koeficienty využijeme na odhad veľkosti HDP na konci roku 2007. Výsledné hodnoty porovnáme s reálnymi dátami.

Výsledok odhadovania nepriestorového modelu na dátach z rokov 2001- 2006, v ktorom sme opäť vylúčili z dátovej štruktúry okres Bratislava- mesto a okres Košice- mesto, je v Obr. 8. A odhad modelu s priestorovou závislosťou je v Obr. 9. Koeficienty z týchto výstupov budeme používať pri predikcii a vlastne aj zisťovaní toho, ako presné sú odhady týchto koeficientov.

```
Call:
plm(formula = model, data = vyber, model = "within")

Balanced Panel: n=36, T=6, N=216

Residuals :
  Min. 1st Qu.  Median 3rd Qu.    Max.
-650.00 -84.70   -0.78   71.90   830.00

Coefficients :
  Estimate Std. Error t-value Pr(>|t|)
P 0.0203171  0.0027586   7.365 6.244e-12

Total sum of squares:    7485300
Residual sum of squares: 5744500
F-statistic: 54.2432 on 1 and 179 DF, p-value: 6.2443e-12
```

Obrázok 8: Odhad nepriestorového modelu rokov 2001-2006

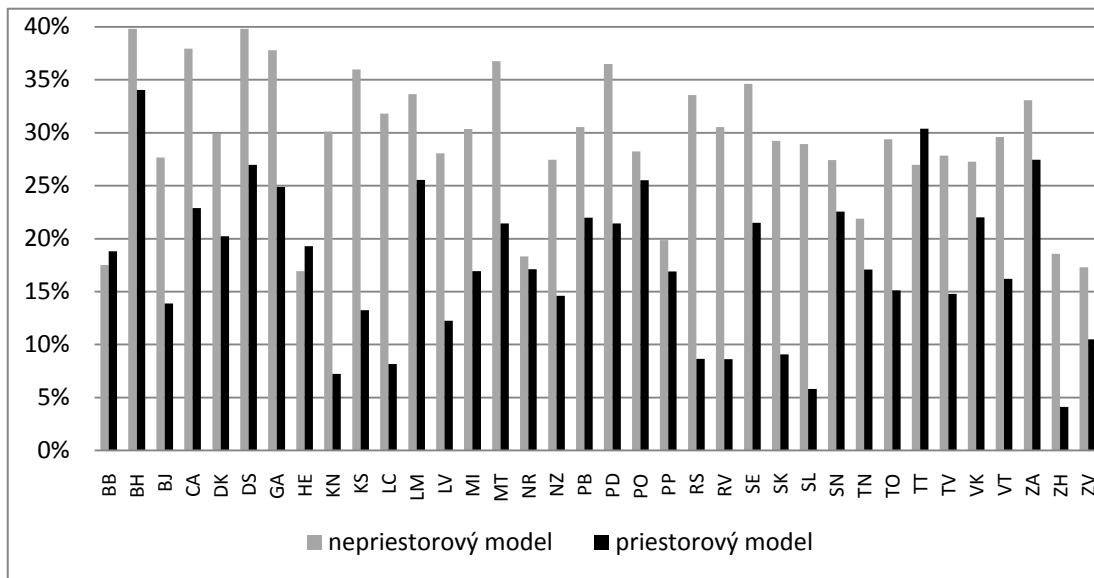
```
Call:
spfeml(formula = model, data = vyber, listw = mat2listw(SRmat),
  effects = "spfe", method = "eigen")

Residuals:
  Min. 1st Qu.  Median 3rd Qu.    Max.
-507.00 -22.40    1.59   21.20   443.00

Coefficients:
  Estimate Std. Error t-value Pr(>|t|)
rho 0.792527  0.034786  22.7829 < 2.2e-16 ***
P 0.0075175  0.012779  5.8827 4.037e-09 ***
```

Obrázok 9: Odhad priestorového modelu rokov 2001- 2006

Z odhadov vidíme, že všetky koeficienty sú významné, tak ich môžeme použiť pre predpoveď veľkosti HDP. Kód pre predikcie je v Prílohe 3. V Grafe 1 sú znázornené odchýlky odhadu od reálnych dát pre obidva typy modelov.



Graf 1: Percentuálne odchýlky odhadu HDP v roku 2007 od reálnych dát pre obidva typy modelov

Z Grafu 1 si je možné všimnúť, že odchýlky sa pohybujú v nepriestorovom modeli od 17% v okrese Humenné po 39,8% v okrese Čadca. Do odchýlky 20% sa zmestilo 6 okresov (HE, ZV, BB, NR, ZH, PP). Najväčšie odchýlky, a to nad 35%, boli zistené v 7 okresoch (KS, PD, MT, GA, CA, BH, DS).

Pre predikciu z priestorového modelu sa odchýlky pohybujú od 4,1% v okrese Komárno po 34% v okrese Bratislava-vidiek. Medzi najpresnejšie odhady, ktorých odchýlky nepresiahli 10% patria okresy ZH, SL, KN, LC, RV, RS, SK. Naopak okresy, v ktorých odhady sa líšia od reálnych dát o viac ako 30%, sú len 2 (TT a BH).

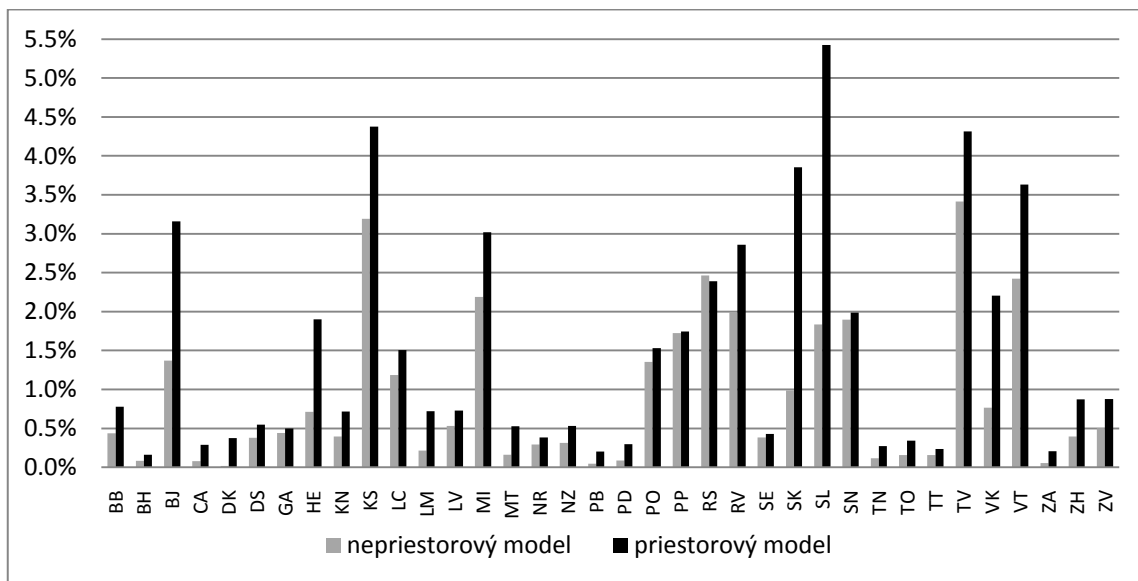
Pre porovnanie priemernej odchýlky máme pre nepriestorový model 29,2% a pre priestorový model 17,7%. Na záver môžeme zhodnotiť, že presnejší odhad HDP nám dal model s použitím priestorovej závislosti, a to dosť veľkým rozdielom 11,5%.

4.3.2 Predikcia pre zvýšenie produktivity rómskeho obyvateľstva

V tejto predikcii budeme používať odhadnuté koeficienty z nepriestorového a priestorového modelu, kde boli vylúčené okresy a sú uvedené v Tabuľke 3.

Hodnotu produktivity minoritnej skupiny obyvateľstva zvýšime o 10 percent a budeme sledovať, čo sa bude diať s hodnotami HDP. Z koeficientu pre P by už teraz malo byť jasné, že HDP narastie, ale budeme sledovať o koľko, v ktorom okrese a porovnáme výsledky z oboch modelov.

Pre jednoduchšie porovnávanie a z poznania, že výsledky percentuálneho rastu HDP pre určitý okres v každom roku sú takmer rovnaké, budeme porovnávať priemerné hodnoty pre každý okres nášho výskumu. V Grafe 2 sú znázornené percentuálne zmeny HDP s vyššou produktivitou minoritnej skupiny oproti HDP s produktivitou pôvodnou.



Graf 2: Percentuálny nárast HDP modelu s vyššou produktivitou oproti modelu s pôvodnou produktivitou minoritnej skupiny obyvateľstva

Pomocou Grafu 2 môžeme s určitosťou prehlásiť, že keď zvýšime produktivitu pre obyvateľov rómskej národnosti, HDP bude mať rastúcu tendenciu. Túto veličinu sme v našom prípade zvýšili o 10%, čo pri produktivite minoritného obyvateľstva v súčasnosti je úplne malé číslo. Hodnoty percentuálneho nárastu HDP sa pre nepriestorový model pohybujú v rozmedzí od 0,02% v okrese Dolný Kubín po 3,4% v okrese Trebišov. Najnižší nárast sme zaznamenali v okresoch DK, PB, ZA, CA, BH,

PD, kde sa HDP zvýši do 0,1%, naopak je to v okresoch PP, SL, SN, RV, MI, VT, RS, KS, TV, kde je nárast o vyše 1,5% a v posledných dvoch prípadoch o viac než 3%. Tieto značné rozdiely sú spôsobené rozmiestnením rómskeho obyvateľstva v okresoch Slovenska.

Podobné výsledky máme aj pre priestorový model. Hodnoty priemerného rastu HDP sa pohybujú v intervale 0,16% v okrese Bratislava- vidiek a 5,42% v okrese Stará Ľubovňa. Nárast v okresoch, kde demografia Rómov nemá vysoké hodnoty, je opäť nižší do hodnoty 0,3%, pre okresy BH, PB, ZA, TT, TN, CA, PD. V okresoch (MI, BJ, VT, SK, TV, KS, SL), kde je počet obyvateľov tejto národnosti väčší, je aj nárast HDP vyšší, nad úrovňou 3% a posledný a nich nad 5%.

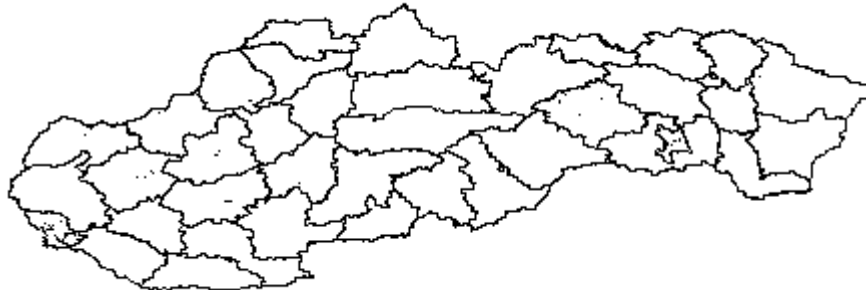
Keď sa dostanem k celkovým priemerným hodnotám pre Slovensko ako celok, prichádzame k záveru, že obidva typy modelov nám dali podobné výsledky. Pre nepriestorový model je hodnota priemerného percentuálneho nárastu HDP po zvýšení produktivity na úrovni 1% a pre model priestorový je to 1,5%. Túto percentuálnu zmenu pokladáme za celkom významné zistenie. Keby sa nám podarilo skupinu minoritného obyvateľstva efektívnejšie zapojiť do chodu ekonomiky Slovenska, podľa našich odhadov by sa dosiahla výrazná zmena HDP.

4.4 Modelovanie v softvéri R

V našej práci sme pracovali so štatistickým softvérom R. V tejto podkapitole si zhrnieme základné príkazy pre odhadovanie modelov s použitím priestorovej ekonometrie. Na úvod sa zbežne pozrieme na tvorbu matice susednosti a v ďalšej časti na modely. Základom celej práce s priestorovými modelmi pre panelové dáta, ktoré sme si uviedli v kapitole 3.3 Modely pre panelové dáta s priestorovou závislosťou, je mať verziu 2.11.1 softvéru R, ktorá vyšla 31. 5. 2010, pretože len v nej funguje knižnica (alebo balík) *splm* pre priestorové modely správne.

Na vytváranie matíc susednosti je potrebný štatistický balík *sp*, ktorý vie konštruovať priestorových susedov a matice váh susednosti, a tiež vie pracovať so súbormi “shapefile“, z ktorých sa načítajú geografické údaje plošných dát (pomocou ktorých sa potom dajú vytvárať mapy sledovaného územia). Načítanie súboru typu

“shapefile“ sa vykonáva pomocou funkcie *readShapePoly*. V našom prípade to bol súbor okresov podľa členenia spred roku 1996. A použitím funkcie *plot* na týchto načítaných dátach vykreslíme mapu ako je v Obrázku 10. Z tohto súboru sa dajú tvoriť



Obrázok 10: Graf okresov SR z načítaných dát zo súboru typu “shapefile“.

matice susednosti. My používame maticu susednosti týchto okresov typu 0,1 (0 nie je sused, 1 je sused), ktorú je možno načítať zo súboru, v ktorom je vypísané, koľko susedov má okres a ktoré sú to. Je to súbor typu .GAL, ktorý má predpísaný tvar. Na normalizovanie tvaru tejto matice sa používa príkaz *nb2mat* (vie urobiť rôzne typy matic, podľa toho, aký typ chceme ďalej používať).

Pre odhady modelov s fixnými efektmi sa používa funkcia *spfeml* (formula, data, listw, model, effects), kde

- *formula* je definovanie modelu,
- *data* sú dáta potrebné pre odhad,
- *listw* matica susednosti typu, akú si zvolíte, keď ju vyrábate,
- v podmienke *model* si môžete určiť, aký typ modelu budete odhadovať, “lag“ alebo “error“ model
- pri *effects* je viacero možností, podľa toho aké efekty chcete do modelu pridať, konštantné (“pooled“), prierezové (“spfe“), časové (“tpfe“), oboje (“sptpfe“)

Pre modely s efektmi náhodnými je v tejto knižnici dostupná funkcia *spreml* (formula, data, w, errors), kde prvé tri zložky sú súhlasné ako pre modely s fixnými efektmi a errors je typ chyby.

Pre predikciu v priestorovej ekonometrii zatiaľ funkcia v tomto štatistickom softvéri nie je dostupná. Používali sme vlastný postup, ktorý vyplýva z teórie o priestorových modeloch.

Záver

V prvých troch kapitolách sme sa venovali metodológii priestorovej ekonometrie. Od hľadania priestorových susedov, zostavovania matíc susednosti až po definovanie modelov. Spojením modelov priestorových pre prierezné dáta a štandardných (nepriestorových) pre dáta panelové sme získali priestorové modely pre panelové dáta. V poslednej kapitole sme sa venovali využitiu štandardnej a priestorovej ekonometrie na reálnych dátach.

Prvým zistením, a teda aj jedným zo záverov tejto práce je skutočnosť, že na Slovensku sú regionálne, okresné dáta na veľmi nízkej úrovni. Nemerajú a neevidujú sa aj kvôli tomu, že tieto územné jednotky majú malú rozlohu a to, čo sa v okrese deje, napríklad, pre tvorbu HDP, je ovplyvňované aj obyvateľmi, ktorí v tomto okrese nebývajú a sú evidovaní v inom, a tým by pre príslušný okres nemali tieto dáta veľkú výpovednú hodnotu. Práve táto skutočnosť vzájomného vplyvu medzi okresmi je základom pre priestorovú ekonometriu.

Keď sme aplikovali teóriu o priestorovej závislosti na okresy Slovenska, výsledok bol taký, ako sme predpokladali: medzi okresmi existuje priestorová autokorelácia. Preto je pri odhadovaní vhodné použitie priestorových modelov, a teda využívanie priestorovej ekonometrie aj u nás.

Skúmali sme model pre 38 okresov Slovenska v územnom členení spred roku 1996 v časovom horizonte 7 rokov v období 2001- 2007. Základnou otázkou priestorového modelu bolo, ako závisí HDP okresu od HDP v okresoch susedných a od produktivity zamestnaných ľudí v okrese, pričom toto množstvo bolo rozdelené na množstvo vyprodukované majoritnou a minoritnou skupinou obyvateľstva s vlastnou produktivitou.

Odhadovali sme lineárny a logit priestorový autoregresívny model a skúmaním odhadnutých koeficientov zistili, že všetky premenné sú signifikantné a majú význam pre modelovanie. Pre lineárny model bol výsledok priestorovej závislosti taký, že keď sa zmení HDP v niektorom zo susedných okresov o jednotku, HDP v danom skúmanom okrese sa zmení o 0.76 rovnakým smerom. V prípade logit modelu, ak sa HDP v susednom okrese zmení o jedno percento, HDP v skúmanom okrese sa zmení o 0,93%

takisto v rovnakom smere. Pri analyzovaní toho, čo by sa dialo s veľkosťou HDP, keby produktivita minoritnej skupiny obyvateľstva bola väčšia, sme prišli k záveru, že úroveň HDP by bola vyššia. Odhad modelu nám dal výsledok, že zvýšenie produktivity o 10%, by malo za následok zvýšenie HDP približne 1,5%, čo je celkom významné číslo. To je odhad nášho modelu, pričom skutočnosť by mohla byť iná. Treba si uvedomiť, že na aktivity, ktoré by podporili nárast produktivity sú potrebné finančné prostriedky, ktoré by zase ovplyvnili veľkosť HDP opačným smerom. Na odhadnutie presnejšieho čísla by boli potrebné štatistiky so širším záberom.

Zoznam použitej literatúry

Knižné zdroje:

1. BIVAND, R.S. - PEBESMA, E.J. - GÓMEZ-RUBIO, V. 2008. *Applied Spatial Data Analysis with R (Use R!)*. Secaucus, N.J.: Springer, 2008. ISBN 978-0-387-78170-9.
2. DIGGLE, P.J. 2003. *Statistical Analysis of Spatial Point Patterns*. London: Arnold 2003. ISBN 0-340-74070-1.
3. LORAN, T. 2002. Sociálna politika a zamestnanosť. In: In: VAŠEČKA, M.: *ČAČIPEN PAL O ROMA– Súhrnná správa o Rómoch na Slovensku*. Bratislava: Inštitút pre verejné otázky, 2002. ISBN 80-88935-41-5, str. 565-586.
4. VAŇO, B. - HAVIAROVÁ, E. 2002. Demografické trendy rómskej populácie . In: VAŠEČKA, M.: *ČAČIPEN PAL O ROMA– Súhrnná správa o Rómoch na Slovensku*. Bratislava: Inštitút pre verejné otázky, 2002. ISBN 80-88935-41-5, str. 475-502.
5. VAŠEČKA, M. 2002. *ČAČIPEN PAL O ROMA– Súhrnná správa o Rómoch na Slovensku*. Bratislava: Inštitút pre verejné otázky, 2002. ISBN 80-88935-41-5.

Elektronické zdroje:

1. ANSELIN, L. 1999. *Spatial Econometrics*. [online]. Richardson: University of Texas at Dallas, 1999. Dostupné na internete: <<http://www.scribd.com/doc/6917442/spatial-econometricsanselin>>
2. ARBIA, G. – BASILE, R. – PIRAS, G. 2005. Using Spatial Panel Data in Modelling Regional Growth and Convergence. [online]. In: *ISAE Working Paper No.55*. Roma: Instituto di Studi e Analisi Economica, 2005. Dostupné na internete: <http://www.isae.it/Working_Papers/WP_55_2005_Arbia_Piras_Basile.pdf>
3. BALTAGI, B.H. 2001. *Ecometric Analysis of Panel Data*. [online]. West Sussex: John Wiley & Sons, Ltd., 2001. Dostupné na internete: <<http://www.filestube.com/ae2cfb307525fedb03ea,g/Baltagi-2001-Econometric-Analysis-Of-Panel-Data.html>>
4. BIVAND, R. 2002. *Spatial econometrics functions in R: Classes and methods*. [online]. Berlin: Springer, 2002. Dostupné na internete: <spatial.nhh.no/papers/bivand-jgs03-rev.pdf>
5. CROISSANT, Y. – MILLO, G. 2008. Panel Data Econometrics in R: The plm Package. [online]. In: *Journal of Statistical Software*, vol.17. Alexandria: American Statistical Association, 2008. Dostupné na internete: <<http://www.jstatsoft.org/v27/i02/paper>>

6. DRUKKER, D.M. 2008. *Analyzing spatial autoregressive models using Stata*. [online]. Chicago: Summer North American Stata Users Group meeting, 2008. Dostupné na internete: <http://www.stata.com/meeting/snasug08/drukker_spatial.pdf>
7. ELHORST, J.P. 2010. Spatial Panel Data Models. [online]. In: FISCHER, M. M., GETIS, A.: *Handbook of Applied Spatial Analysis*. Berlin: Springer, 2010. Dostupné na internete: <<http://www.regroningen.nl/elhorst/doc/Spatial%20Panel%20Data%20Models.pdf>>
8. FILADELFOVÁ, J. - GERBERY, D. - ŠKOBLA, D. 2006. *Správa o životných podmienkach rómskych domácností na Slovensku*. [online]. Bratislava: UNDP, 2006. Dostupné na internete: <http://library.fes.de/pdf-files/bueros/slowakei/04679.pdf>
9. FLORAX, R.J.G.M. – VAN DER VLIST, A.J. 2003. Spatial Econometrics Data Analysis: Moving Beyond Traditional Models. [online] In: *International Regional Science Review*26(3). SAGE Publications, 2003. Dostupné na internete: <http://brandenburg.geoecology.uni-potsdam.de/users/schroeder/download/spatialstats_workshop/Florax2003IntRegSciRev.pdf>
10. FRAZIER, CH. - KOCKELMAN, K.M. 2005. Spatial Econometric Models for Panel Data: Incorporating Spatial and Temporal Data. [online] In: *Transportation Research Record: Journal of the Transportation Research Board* 1902. Washington: Transportation Research Board Business Office, 2005. Dostupné na internete: <http://www.ce.utexas.edu/prof/kockelman/public_html/TRB05satdata.pdf>
11. HSIAO, Ch. 2005. Why Panel Data?. [online]. In: *IEPR Working Paper 05.33*. California: Institute of Economic Policy Research, 2005. Dostupné na internete: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=820204>
12. LeSAGE, J.P. 1998. Spatial Econometrics. [online]. Toledo: University of Toledo, 1998. Dostupné na internete: <<http://www.spatial-econometrics.com/html/wbook.pdf>>
13. MARCINČIN, A. - MARCINČINOVÁ, Ľ. 2009. *Straty z vylúčenia Rómov. Kľúčom k integrácii je rešpektovanie inakosti*. [online]. Bratislava: Nadácia otvorenej spoločnosti, 2009. Dostupné na internete: <<http://www.iz.sk/download-files/sk/osf-straty-z-vylucenia-romov.pdf>>
14. NOVÁK, P. 2007. Analýza panelových dát. [online]. In: *Acta Oeconomica Pragensia* 1/2007. Praha: Vysoká škola ekonomická v Prahe, 2007. Dostupné na internete: <<http://www.vse.cz/aop/pdf/40.pdf>>

15. PIRAS, G. 2009. *Splm, Spatial Panel Data Models in R*. [online]. Morgentown: West Virginia University, 2009. Dostupné na internete: <http://www.rri.wvu.edu/pdffiles/splm_10152009.pdf>
16. ŠPROCHA, B. *Dáta k demografii Rómov v okresoch*. Mailová komunikácia.
17. VAŇO, B. 2001: *Demografická charakteristika rómskej populácie v SR*. [online]. Bratislava: Infostat, 2001. Dostupné na internete: <<http://www.infostat.sk/vdc/pdf/rom.pdf>>
18. VITTON, P.A. 2010. *Notes on Spatial Econometric Models*. Columbus: The Ohio State University, 2010. Dostupné na internete: <<http://facweb.knowlton.ohio-state.edu/pviton/courses2/crp8703/spatial.pdf>>

Odkazy na použité stránky:

1. Štatistický úradu Slovenskej republiky- www.statistics.sk
2. Úrad práce, sociálnych vecí a rodiny- www.upsvar.sk
3. Slobodná encyklopédia Wikipedia- www.wikipedia.org
4. National Human Development Report Slovak Republic 2001-2002- http://planipolis.iiep.unesco.org/upload/Slovakia/Slovak_Republic_2002_en.pdf
5. International Labour Organization- <http://www.ilo.org/global/lang--en/index.htm>
6. Komplexný archív softvéru R- <http://cran.r-project.org/>
7. Dáta pre HDP- <http://www.iz.sk/sk/projekty/ukazovatele-za-EU/hdp-na-obyvatela-v-pps-ako-priemer-eu-nuts-3>
8. Hranice okresov SR- <http://wiki.freemap.sk/Regiony>

Príloha 1 : Transformácia prierezoých dát na panelové urobená v softvéri STATA.

```
use "I:\data.dta"
save "I:\datatemp.dta", replace
gen year=2001
gen newHDP=HDP_2001
gen newZR=ZR_2001
gen newZM=ZM_2001

save "I:\datapanel.dta", replace

use "I:\data.dta"
gen year=2002

gen newHDP=HDP_2002
gen newZR=ZR_2002
gen newZM=ZM_2002

save "I:\datatemp.dta", replace
use "I:\datapanel.dta", replace
append using "I:\datatemp.dta"
save "I:\datapanel.dta", replace

forval j=2003(1)2007 {
use "I:\data.dta"
gen year=`j'

gen newHDP=HDP_`j'
gen newZR=ZR_`j'
gen newZM=ZM_`j'

save "I:\datatemp.dta", replace
use "I:\datapanel.dta", replace
append using "I:\datatemp.dta"
save "I:\KKdatapanel.dta", replace
}
drop HDP_* ZR_* ZM_*
sort Mark year
rename newHDP HD
rename newZM ZM
rename newZR ZR

save "I:\datapanel.dta", replace
```

Príloha 2: Odhadovanie modelov – zdrojový kód

```
library(splm)

#matica váh (susedstvo 0 a 1) bez BA, BH, KE
SR<- read.geoda("okresy.txt" [1])
SRgal<-read.gal("susedia.GAL" [1],SR$okres)
SRmat <- nb2mat(SRgal, zero.policy=TRUE)
SRw <- mat2listw(SRmat)

#panelové dáta bez BA, BH, KE
data<-read.table("Data.txt", header=TRUE)
head(data)
panel<-pdata.frame(data,index=c("okres","rok"),drop.index=TRUE,row.names=TRUE)

#definovanie modelu bez BA, BH, KE
model<-HDP~P
logmodel<-log(HDP)~log(P)

#fe a re one-way nepriestorovy
panel.fe<-plm(model, data=panel, model="within")
summary(panel.fe)
panel.re<-plm(model, data=panel, model="random")
summary(panel.re)
phtest(panel.fe,panel.re)

logpanel.fe<-plm(logmodel, data=panel, model="within")
summary(logpanel.fe)
logpanel.re<-plm(logmodel, data=panel, model="random")
summary(logpanel.re)
phtest(logpanel.fe,logpanel.re)

#spatial lag fe priestorovy
spatial.fe<-spfeml(model, data=data, listw=mat2listw(SRmat), model="lag",
                  effects = "spfe",method = "eigen")
summary(spatial.fe)

logspatial.fe<-spfeml(logmodel, data=data, listw=mat2listw(SRmat), model="lag",
                    effects = "spfe",method = "eigen")
summary(logspatial.fe)
```

Príloha 3: Zdrojové kódy predikcií pre priestorové modely

Predikcia pre rok 2007:

```
library(splm)
library(car)
SR<- read.geoda("okresy.txt" [1])
SRgal<-read.gal("susedia.GAL" [1],SR$okres)
SRmat <- nb2mat(SRgal, zero.policy=TRUE)

data<-read.table("Data.txt", header=TRUE)
vyber <- data[data$rok %in% 2001:2006, ]

model<-HDP~P
odhad<-spfeml(model, data=vyber, listw=mat2listw(SRmat), effects = "spfe",
              method = "eigen")
summary(odhad)
eff<-effects(odhad)
write.effects.splm(eff, "effects",1)

panel<-pdata.frame(vyber, index=c("okres","rok"),
drop.index=TRUE,row.names=TRUE)
HDP<-as.matrix(panel$HDP)
P<-as.matrix(panel$QM)

FE<-read.table("effects", header=TRUE)
e<-read.table("rez.txt", header=TRUE)
pe<-rep(0,time=length(e[,1]))
for (i in 1:length(e[,1])){
  pe[i]<-sum(e[i,])/length(e[1,])
}
pe
new <- data[data$rok %in% 2007, ]

W=SRmat
I=diag(rep(1,time=length(W[,1])))
rho= 0.792527
betaP=0.0075175
M=ginv(I-rho*W)

newHDP=M%*(betaP*new$P)+M%*(FE+pe)
predikcia_pr<-data.frame(newHDP,new$HDP)
predikcia_pr_chyba<-abs(new$HDP-newHDP)/new$HDP
```

Predikcia pre zvýšenú produktivitu:

```
library(splm)
library(car)
SR<- read.geoda("okresy.txt" [1])
SRgal<-read.gal("susedia.GAL" [1],SR$okres)
SRmat <- nb2mat(SRgal, zero.policy=TRUE)

data<-read.table("Data.txt", header=TRUE)
model<-HDP~P
odhad<-spfeml(model, data=data, listw=mat2listw(SRmat), model="lag",
              effects = "spfe",method = "eigen")
summary(odhad)
eff<-effects(odhad)
print(eff)
write.effects.splm(eff, "effects",1)

panel<-pdata.frame(data,index=c("okres","rok"),drop.index=TRUE,row.names=TRUE)
HDP<-as.matrix(panel$HDP)
P<-as.matrix(panel$P)
Pnove<-as.matrix(panel$Pnove)

W=SRmat
I=diag(rep(1,time=length(W[,1])))
rho= 0.764174
betaP<-0.011484
FE<-read.table("FEpr.txt", header=TRUE)
FE<-as.matrix(FE)
e<-read.table("rez_pr.txt", header=TRUE)
e<-as.matrix(e)

M=ginv(I-rho*W)
HDPodhad=M%*(betaP*Pnove)+M%*(FE+e)
predikcia<-data.frame(HDPodhad,HDP)
predikcia_p<-data.frame((HDPodhad-HDP)/HDP)
```