UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MODELOVANIE NEZAMESTNANOSTI PRE REGIÓNY NUTS 2 KRAJÍN EÚ

DIPLOMOVÁ PRÁCA

2017 Bc. Bystrík KUBALA

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MODELOVANIE NEZAMESTNANOSTI PRE REGIÓNY NUTS 2 KRAJÍN EÚ

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	1114 Aplikovaná matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	Ing. Marek Radvanský, PhD.

Bratislava 2017

Bc. Bystrík KUBALA

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

35857081

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	Bc. Bystrík Kubala
Študijný program:	ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor:	aplikovaná matematika
Typ záverečnej práce:	diplomová
Jazyk záverečnej práce:	slovenský
Sekundárny jazyk:	anglický

Názov: Modelovanie nezamestnanosti pre regióny NUTS 2 krajín EÚ
Unemployment modeling for NUTS 2 regions in the European union countries

Cieľ: Priestorová ekonometria (z ang. Spatial Econometrics) je nástroj na analyzovanie dát, ktoré v sebe majú zahrnutú priestorovú informáciu. Diplomová práca bude zameraná na: 1.) zavedenie pojmov, modelov a odhadov parametrov týchto modelov pomocou MLE, 2.) spracovanie reálnych makroekonomických panelových dát regiónov NUTS 2 krajín EÚ a rôznych prístupov k tvorbe váhovej matice.

Vedúci:	Ing. Marek Radvanský, PhD.
Katedra:	FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry:	prof. RNDr. Daniel Ševčovič, DrSc.
Dátum zadania:	10.02.2015

Dátum schválenia: 11.02.2015

prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

....
študent

....
vedúci práce

Poďakovanie

Touto cestou sa chcem poďakovať všetkým, ktorý mi poskytli odborné rady a čas, ktoré boli potrebné pri písaní diplomovej práce.

Abstrakt

KUBALA, Bystrík: Modelovanie nezamestnanosti pre regióny NUTS 2 krajín EÚ [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Ing. Marek Radvanský, PhD., Bratislava, 2017, 58 s.

Hlavným cieľom diplomovej práce je predstaviť teóriu priestorovej ekonometrie a túto teóriu aplikovať na makroekonomické dáta, ktoré v sebe zahŕňajú priestorový efekt.
V teoretickej časti sa zaoberáme základnými pojmami, akými sú priestorová závislosť a priestorová rôznorodosť, autoregresný proces, maticou priestorových váh a jej zhotovením. Predstavíme si tri základné modely priestorovej ekonometrie, ktorými sú priestorový autoregresný model, model s priestorovými chybami a Durbinov model . Odvodíme odhady parametrov pre jednotlivé modely pomocou metódy maximálnej vierohodnosti. V praktickej časti je našou úlohou zostrojiť maticu vzdialenosti pomocou Haversínyho formuly a rôzne typy matice priestorových váh. Nakoniec práce modelujeme nezamestnanosť v rámci NUTS 2 regiónov a podrobne opíšeme dosiahnuté výsledky.

Kľúčové slová: autoregresný proces, matica priestorových váh, priestorový autoregresný model, model s priestorovými chybami, Durbinov model, metóda maximálnej vierohodnosti

Abstract

KUBALA, Bystrík: Unemployment modeling for NUTS 2 regions in the European union countries [Master Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Ing. Marek Radvanský, PhD., Bratislava, 2017, 58 p.

The main object of the master thesis is to introduce the theory of spatial econometrics and apply this theory to macroeconomic data, which include spatial effect.
In the theoretical part we deal with basic terms, which are spatial dependence and spatial heterogenity, autoregression process, spatial weight matrix and the construction of this matrix. We present three basic models of spatial econometrics, which are spatial autoregressive model, spatial error model and Durbin model. We estimate the parameters of individual models using the maximum likelihood approach. In the practical part, we construct a matrix of distance by using Haversine formula and the different forms of spatial weight matrices. At last, we model unemployment in the NUTS 2 regions and describe the results achieved in detail.

Keywords: autoregressive process, spatial weight matrix, spatial autoregressive model, spatial error model, Durbin model, maximum likelihood method

Obsah

Úvod

Priestorová ekonometria je z historického hľadiska pomerne mladá disciplína. Prvýkrát bol tento pojem použitý v sedemdesiatych rokoch dvadsiateho storočia. Záznamy sú však už o nej oveľa skôr, a to z 19. storočia. Nástroje priestorovej ekonometrie sa využívajú prevažne v zahraničí, na Slovensku je ešte stále málo známa. Hlavným z dôvodov prečo vznikla táto časť ekonometrie je, že sa začali zhromažďovať dáta, ktoré majú v sebe zakomponovaný priestor. Pri použití klasickej regresie sa neberie do úvahy tento efekt priestoru a tým pádom nie sú výsledky úplné.

V prvej kapitole si bližšie predstavíme priestorovú ekonometriu. Pripomenieme si klasický lineárny regresný model, jeho predpoklady a odhad pomocou metódy najmenších štvorcov. Ukážeme si, prečo sa nedá aplikovať na dáta, na ktoré sa využíva priestorová ekonometria. Vysvetlíme si pojem priestor a predstavíme si priestorové efekty, ktorými sú priestorová závislosť a priestorová rôznorodosť. Neskôr prejdeme k autoregresnému procesu a s ním súvisiacej váhovej matici. Predstavíme si rôzne spôsoby, ako ju môžeme zostrojiť. Nakoniec sa dopracujeme k základným modelom priestorovej ekonometrie, ktorými sú priestorový autoregresný model, model s priestorovými chybami a Durbinov model. Na záver tejto kapitoly si na ilustračnom príklade bližšie priblížime danú teóriu priestorovej ekonometrie.

V druhej kapitole sa budeme zaoberať odhadom parametrov pre jednotlivé modely, ktoré sme predstavili v prvej kapitole. Parametre budeme odhadovať pomocou metódy maximálnej vierohodnosti. Na záver si by sme chceli zadefinovať formulu, ktorá nám bude slúžiť na overenie priestorovej interakcie v dátach. Obe kapitoly sú prevažne spracovaná podľa kníh [2] a [7].

V tretej kapitole si popíšeme dáta a priestor, s ktorými budeme pracovať. Dáta budeme brať s ohľadom na regióny NUTS 2. Modelovať budeme nezamestnanosť v rámci jednotlivých regiónov s ohľadom na infláciu, HDP a hustotu obyvateľstva v rámci regiónu. V tejto kapitole sa budeme venovať i zostrojeniu váhovej matici. Táto časť bude pomerne komplikovaná, pretože by sme radi získali váhovú maticu vzdialeností pre jednotlivé regióny. Tento časovo náročný proces sa však dá zjednodušiť použitím Haversínyho formuly, ktorú si predstavíme.

V poslednej, štvrtej kapitole sa budeme venovať zostrojeniu rôznych modelov pre naše dáta. Tieto modely sa budú líšiť hlavne výberom rôznych váhových matíc. Pracovať budeme v systéme `R`, kde je pripravený balík na spracovanie dát pre priestorovú ekonometriu `splm`. Tento balíček si i popíšeme. Na záver tejto kapitoly si interpretujeme dosiahnuté výsledky.

1 Úvod do priestorovej ekonometrie

Z historického hľadiska bol pojem priestorová ekonometria (z angl. spatial econometrics) prvýkrát použitý v sedemdesiatych rokoch dvadsiateho storočia. Zakľadatelom bol Belgičan Jean Paelinck, ktorý v roku 1979 spolu so spoluautorom L. H. Klaassenom vydali knihu *Spatial Econometrics*. V tejto publikácií poukázali na viacero charakteristík priestorovej ekonometrie, ako sú úloha priestorovej vzájomnej závislosti v priestorových modeloch, nesymetrickosť v priestorových vzťahoch, a.i.

Súčasným najväčším priekopníkom v tejto oblasti je tiež rodený Belgičan Luc Anselin, ktorý spravil pokrok v priestorovej ekonometrii, či už po teoretickej, ale i po praktickej stránke. Táto kapitola je spracovaná z veľkej časti podľa jeho knihy [2] a podľa knihy [7] od autorov J. LeSage a R.K.Pace.

1.1 Klasický lineárna regresný model

Najviac používaná štatistická metóda, ktorá je využívaná v ekonometrii je bezpochyby klasický lineárny regresný model (LRM). Tento model môžeme napísať v maticovom tvare nasledovne:

$$y = X\beta + \epsilon, \tag{1}$$

kde:

- y označuje $n \times 1$ rozmerný vektor závislej premennej
- X označuje $n \times k$ rozmernú maticu vysvetľujúcich premenných, pričom $n > k$
- β je $k \times 1$ rozmerný neznámy vektor parametrov
- ϵ je $n \times 1$ rozmerný vektor rezíduí alebo i vektor chybových členov.

Našou úlohou je nájsť čo najlepší odhad parametrov $\hat{\beta}$ pomocou dostupných dát X a y, pričom získaná regresná krivka by mala čo najlepšie vystihovať lineárny charakter dát.

Uvažujme, že pre daný klasický lineárny regresný model platia predpoklady:

1. $X_0, \ldots, X_{k-1}$ sú lineárne nezávislé, t.j. hodnosť matice X je $h(X) = k$;

2. $E[\epsilon|X] = 0$;

3. $Var[\epsilon_i|X] = \sigma^2$, kde $i \in \langle 1, n \rangle$- homeskedasticita rezíduí;

4. $Cov[\epsilon_i, \epsilon_j|X] = 0$, pre $i \neq j$ - párová nekorelovanosť rezíduí;

5. matica X je nenáhodná.

Veta 1.1. *Gauss - Markovova veta*
Za predpokladov 1.-5. klasického LRM je odhad parametrov β pomocou metódy najmenších štvorcov

$$\hat{\beta}_{MNS} = (X'X)^{-1}X'y \tag{2}$$

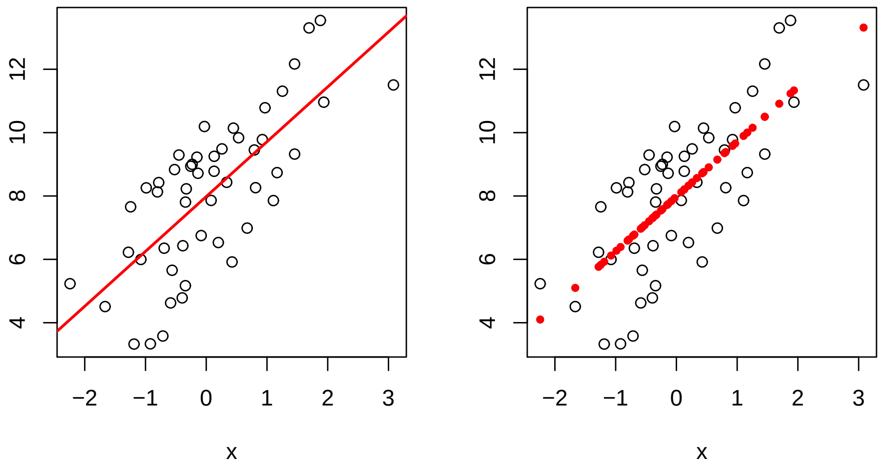
najlepší lineárny nevychýlený odhad, t.j. žiadny iný lineárny nevychýlený odhad nemá menšiu varianciu.

Poznámka: Vo všeobecnosti existujú lepšie odhady pre parameter β, ktoré sú však nelineárne. Ak pridáme ďalší predpoklad:

6. $\epsilon|X \sim N(0, \sigma^2 I)$,

tak odhad parametra β pomocou metódy najmenších štvorcov bude najlepší nevychýlený odhad.

Pre lepšie pochopenie klasickej lineárnej regresia a jej odhadu v dvojrozmernom priestore prikladáme obrázok (1). Na prvej časti obrázku máme znázornené 50 pozorovaných dát a získanú regresnú priamku, pričom vektor parametrov $\hat{\beta}$ sme vypočítali pomocou metódy najmenších štvorcov ako je uvedené vyššie. Na druhej časti obrázku sú znázornené fitované hodnoty, ktoré zodpovedajú regresnej krivke pre každú vysvetľujúcu hodnotu.
Prvý predpoklad pre klasický LRM je nezávislosť pozorovaní pre $i - tu$ a $j - tu$ pozíciu. Avšak pre dáta, ktoré majú v sebe priestorovú závislosť, nie je tento predpoklad o nezávislosti splnený.



Obr. 1: Graficky znázornená získaná regresná priamka a fitované hodnoty

1.2 Priestor a priestorové efekty

Hlavným podnetom vzniku priestorovej ekonometrie bolo zhromažďovanie dát, ktoré mali v sebe zakomponovanú priestorovú závislosť. Sú to dáta, ktoré sú zozbierané s ohľadom na priestor. Pod priestorom si môžeme predstaviť množinu jednotlivých oblastí alebo regiónov, ktoré sú od seba oddelené hranicami. Tieto hranice môžu byť rôzne, napr. hranice štátov Európskej únie, hranice jednotlivých regiónov Slovenska, a i. Pri priestorovej ekonometrii nás zaujíma, či udalosti v jednom regióne ovplyvňujú udalosti v iných regiónoch a vice versa.

Jednoduchým príkladom pre dáta s priestorovou závislosťou sú dáta o populácií, zamestnanosti a vzdelanosti zozbierané na základe administratívnych jednotiek, ktorými môžu byť štáty, provincie či jednotlivé regióny.

1.2.1 Priestorová závislosť

Priestorová závislosť je jednou z vlastností priestorového procesu, v ktorom pozorovania z rôznych oblastiach y_i a y_j, $j \neq i$ sú závislé:

$$E(y_i y_j) \neq E(y_i)E(y_j).$$

Poznámka: Priestorovú závislosť môžeme matematicky zapísať aj nasledovne:

$$y_i = f(y_j); i = 1, \ldots, n; \quad j \neq i.$$

To znamená, že ide o funkčnú závislosť medzi tým, čo sa stane v jednom regióne v priestore a tým, čo sa stane v inom regióne.
Pre jednoduchosť si predstavme pozorovania pre dva rôzne regióny. Potom priestorová závislosť môže nadobúdať tvar:

$$\begin{aligned} y_i &= \alpha_i y_j + X_i \beta + \epsilon_i \\ y_j &= \alpha_j y_i + X_j \beta + \epsilon_j \\ \epsilon_i &\sim N(0, \sigma^2), \quad i = 1 \\ \epsilon_j &\sim N(0, \sigma^2), \quad j = 2 \end{aligned} \tag{3}$$

1.2.2 Priestorová rôznorodosť

Po priestorovej závislosti je priestorová rôznorodosť (z angl. spatial heterogenity) druhým základný priestorový efektom. Každé priestorové pozorovanie, ktoré je závislé od pozície v priestore, nemôže byť homogénneho. Príkladom sú mestské časti, ktoré majú rozdielne populácie alebo príjem, regióny, ktoré majú rozličné stupne technologického rozvoja a i. Tieto aspekty heterogenity môžeme sledovať ako variácie vzťahov v priestore. Preto je potrebné uvažovať regresný vzťah, ktorý má nekonštantnú disperziu pre rezíduá.

1.3 Modely priestorovej ekonometrie

Predstavme si, že v (3) by sme brali do úvahy 3 rôzne regióny a rovnakú funkčnú závislosť. V tomto prípade by sme dostali:

$$
\begin{aligned}
y_i &= \alpha_{i,j} y_j + \alpha_{i,k} y_k + X_i \beta + \epsilon_i \qquad (4)\\
y_j &= \alpha_{j,i} y_i + \alpha_{j,k} y_k + X_j \beta + \epsilon_j\\
y_k &= \alpha_{k,i} y_i + \alpha_{k,j} y_j + X_k \beta + \epsilon_k\\
\epsilon_i &\sim N(0, \sigma^2), \quad i = 1\\
\epsilon_j &\sim N(0, \sigma^2), \quad j = 2\\
\epsilon_k &\sim N(0, \sigma^2), \quad k = 3
\end{aligned}
$$

Dá sa pozorovať, že tento systém rovníc je prakticky nepoužiteľný, lebo obsahuje viacej parametrov ako pozorovaní. Vo všeobecnosti, vzťah závislosti medzi množinou pozorovaní v rámci n regiónov nám môže vyprodukovať až $n^2 - n$ vzťahov.
Na riešenie, ako sa vysporiadať s takto preparametrizovaný problém prišiel *Ord* [11], ktorý sa inšpiroval prácou *Whittla* [13], ktorý pre priestorovú závislosť zaviedol špecifický parametra ρ. Tento parameter popisuje vzťah priestorovej závislosti. Tým pádom môžeme popísať závislosť medzi pozorovaniami premenných y.

1.3.1 Autoregresný proces

Definícia 1.2. *Nech y označuje $n \times 1$ rozmerný vektor a W je $n \times n$ riadkovo normalizovaná váhová matica. Potom priestorový autoregresný proces nadobúda tvar:*

$$y = \rho W y + \epsilon \qquad (5)$$

$$\epsilon \sim N_n(0, \sigma^2 I_n), \qquad (6)$$

kde ρ je 1×1 rozmerný neznámy autoregresný parameter a ϵ je $n \times 1$ rozmerný vektor rezíduí.

Z definície priestorového autoregresného procesu (5) vyplýva, že popisujeme vzťah medzi vektorom y a vektorom Wy, ktorý reprezentuje lineárnu kombináciu susediacich hodnôt každého pozorovania. Ku váhovej matici W a jej konštrukcii sa vrátime neskôr. Pre konkrétne pozorovanie i sa dá prepísať autoregresný proces nasledovne:

$$y_i = \rho \sum_{j=1}^{n} W_{ij} y_j + \epsilon_i \tag{7}$$

$$\epsilon_i \sim N_n(0, \sigma^2). \tag{8}$$

Suma $\sum_{j=1}^{n} W_{ij} y_j$ sa nazýva priestorový posun (z angl. spatial lag). Odhad parametra ρ získame pomocou metódy maximálnej vierohodnosti, ktorý bude podrobnejšie popísaný v kapitole (2) na zložitejších modeloch.

1.3.2 Matica priestorových váh W

Matica priestorových váh W obsahuje údaje o susednosti pozorovaných objektov. Zároveň platí, že matica W je symetrická s nulovými prvkami na diagonále. Je to z dôvodu, že ak región i je susedným regiónom s j, tak aj región j je susedom regiónu i a platí predpoklad, že región i nie je sused sám so sebou. Avšak predpokladáme, že matica W musí byť pozitívne definitná.

Otázkou, ktorou sa budeme teraz zaoberať je ako v matici W určiť susedstvo pozorovaných subjektov a ako zmerať veľkosť tohto susedstva, t.j. vzdialenosť.

Binárna matica priestorových váh W

Najskôr budeme uvažovať prípad, kde vzdialenosť dvoch susedov bude určená jednoduchými binárnymi premennými:

- volíme 1, ak dva pozorované subjekty sú susedné
- volíme 0, ak dva pozorované subjekty nie sú susedné.

Existuje viacero spôsobov ako určiť susedstvo pozorovaných subjektov, my si uvedieme iba niektoré z nich:

Bishop contiguity: $W_{ij} = 1$ vyplníme pre subjekty, ktoré majú spoločný iba jeden bod hranice;

Rook contiguity: $W_{ij} = 1$ vyplníme pre subjekty, ktoré zdieľajú aspoň určitú časť hranice;

Queen contiguity: je kombináciou *Rook* a *Bishop contiguity.* $W_{ij} = 1$ vyplníme pre subjekty, ktoré majú spoločný bod alebo časť hranice. V tomto prípade nezáleží na dĺžke spoločnej hranice.

V odvodení určenia susedstva môžeme pozorovať súvislosť pohybu figúrok v šachu. *Strelec* (z angl. Bishop) sa pohybuje diagonálne, čo predstavuje spoločný rohový bod, *Veža* (z angl. Rook) sa pohybuje horizontálne alebo vertikálne, čo môže predstavovať časť zdieľanej hranice. Nakoniec *dáma* (z angl. Queen) sa pohybuje vertikálne, horizontálne a aj diagonálne, čo v konečnom výsledku predstavuje ich kombináciu na určenie susednosti.

Priestorové vzdialenosti pre maticu W

Ďalší spôsob ako určiť váhové matice je, že nebudeme brať do úvahy len binárnu susednosť, ale i vzdialenosť voči regiónom. Opäť existuje viacero možností určenia vzdialeností pozorovaných subjektov, my si vyberieme len dve na ukážku:

- letecká alebo cestná vzdialenosť dvoch hlavných centier pozorovaných subjektov,
- letecká vzdialenosť centier centroidov pozorovaných subjektov.

Pod pojmom centrum centroidu rozumieme vnútorný bod daného regiónu, tzv. ťažisko. Po získaní vzdialenosti jednotlivých regiónov môžeme vytvoriť maticu W rôznymi spôsobmi. Jeden z príkladov podľa [4] je:

$$w_{ij} = \frac{\beta_{ij}^b}{d_{ij}^a}$$

kde β_{ij} je pomer spoločnej hranice regiónu i a j s celkovou dĺžkou hranice regiónu i, d_{ij} je vzdialenosť medzi dvoma pozorovanými regiónmi i a j a a, b sú kladné konštanty. Z definície vyplýva, že táto matica nemôže byť symetrická. Symetrická by bola len v prípade, ak by mali všetky susedné regióny i a j rovnaké dĺžky hraníc. My nebudeme používať tento prístup pre naše dáta. Je pomerne komplikované získať dĺžku hranice pre jednotlivé regióny. Zvolíme si iný prístup k získaniu váhovej matici, ktorý bude popísaný v praktickej časti.

Váhová matica, ktorá vstupuje do modelu, musí byť riadkovo normalizovaná.

V nasledujúcich častiach si predstavíme tri najviac využívané priestorové regresné modely:

- *Priestorový autoregresný model (SAR)*
- *Model s priestorovými chybami (SEM)*
- *Priestorový Durbinov model (SDM)*

1.3.3 Priestorový autoregresný model

Priestorový autoregresný model získame kombináciou priestorového autoregresného procesu z definície (5) a klasického LRM.

$$y = \rho W y + X\beta + \epsilon \tag{9}$$

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}\epsilon \tag{10}$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

V SAR modely sa odhaduje štandardný regresný parameter β a pridružený skalárny parameter ρ. Nota bene, ak skalárny parameter ρ nadobudne hodnotu 0, tak v prierezových pozorovaniach y nie je priestorová závislosť, a teda sa dostávame k štandardnému regresnému modelu. Podstata modelu SAR oproti klasickej LRM je taká, že do modelu vstupuje priestorová závislosť ako ďalší regresor vo forme priestorového posunu Wy. Inak povedané správanie pozorovaného regiónu závisí od správania sa susedných pozorovaných regiónoch.

1.3.4 Model s priestorovými chybami

Základná motivácia na odvodenie modelu s priestorovými chybami je, že uvažujeme rezíduá, ktoré vykazujú priestorovú závislosť. V tomto modely nám vystupuje priestorová rôznorodosť. Teda prvky variačno-kovariačnej matice mimo diagonály nie sú nulové. Potom model s priestorovými chybami má tvar:

$$y = X\beta + u \tag{11}$$

$$u = \lambda W u + \epsilon$$

$$y = X\beta + (I_n - \lambda W)^{-1}\epsilon \tag{12}$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Parameter λ predstavuje priestorový autoregresný koeficient lineárneho posunu Wu, avšak v tomto prípade týkajúceho sa rezíduí. Podstata modelu spočíva v tom, že v modely nepredpokladáme silnú priestorovú závislosť medzi pozorovanými subjektmi, ale medzi rezíduami.

1.3.5 Priestorový Durbinov model

Posledný model, priestorový Durbinov model vznikne kombináciou SAR a SEM modelu. Pri odvodzovaní tohto modelu sa berie do úvahy, že vektor ϵ sa dá vyjadriť ako lineárna kombinácia matice X a rezíduí v.

$$\epsilon = X\gamma + v$$
$$v \sim N_n(0, \sigma^2 I_n)$$

Substitúciou daného ϵ do modelu SEM a menšou úpravou parametrov dostaneme priestorový Durbinov model:

$$y = \lambda W y + X\beta + WX\theta + v \tag{13}$$
$$y = (I_n - \lambda W)^{-1} X\beta + (I_n - \lambda W)^{-1} WX\theta + (I_n - \lambda W)^{-1} v \tag{14}$$
$$v \sim N(0, \sigma^2 I_n)$$

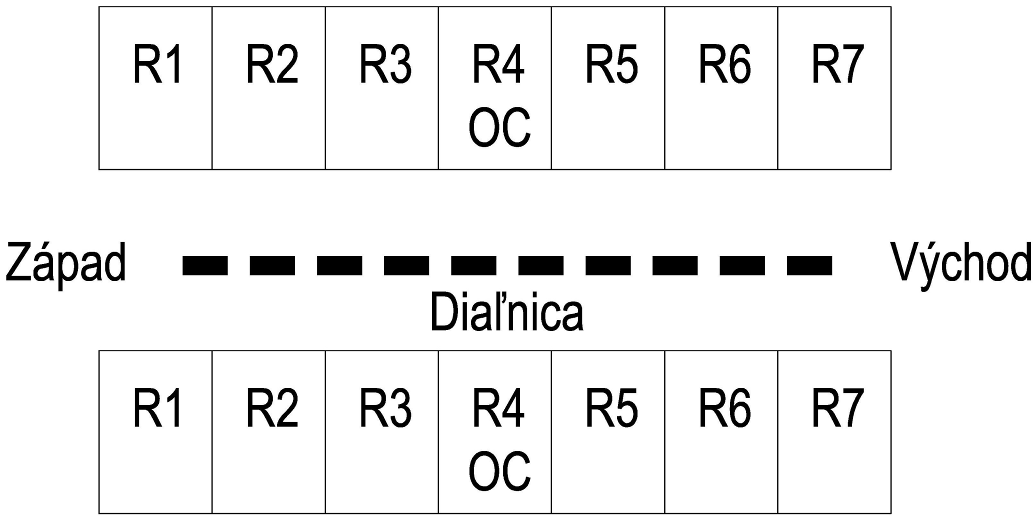
Odhady daných parametrov pre všetky tri modely si odvodíme v druhej kapitole. Teraz si vysvetlíme danú teóriu na ilustračnom príklade.

1.4 Ilustračný príklad

Pre lepšie pochopenie doterajšej teórie uvádzame ilustračný príklad, ktorý je z publikácie [7].

Predstavme si, že máme metropolitnú oblasť, ktorá sa skladá zo siedmych regiónov. V strednom regióne $R4$ sa nachádza hlavné obchodné centrum. Tri regióny sú na východ a tri regióny na západ od tohto centra. Cez stred týchto regiónov ide priamo diaľnica, ktorú všetci obyvatelia využívajú pri ceste do OC a z OC. Na nasledujúcom obrázku (2) je zakreslená štruktúra danej metropolitnej oblasti.

Pozorujme vzorku dát z týchto regiónov, ktorá obsahuje cestovné časy do OC (v



Obr. 2: Metropolitná oblasť s OC

minútach), vzdialenosť regiónov od OC (v kilometroch) a hustotu obyvateľstva jednotlivých regiónoch. Za závislú premennú y zoberme cestovné časy a vysvetľujúce premenné v matici X budú vzdialenosť a hustota obyvateľstva.

$$y = \begin{array}{c} \text{Cestovné časy} \\ \begin{pmatrix} 42 \\ 37 \\ 30 \\ 26 \\ 30 \\ 37 \\ 42 \end{pmatrix} \end{array} \quad X = \begin{array}{cc} \text{Hustota} \quad \text{Vzdialenosť} \\ \begin{pmatrix} 10 & 30 \\ 20 & 20 \\ 30 & 10 \\ 50 & 0 \\ 30 & 10 \\ 20 & 20 \\ 10 & 30 \end{pmatrix} \end{array} \quad \begin{array}{ll} \text{ďaleké predmestie} & \text{R1} \\ \text{okrajové predmestie} & \text{R2} \\ \text{predmestie} & \text{R3} \\ \text{obchodné centrum OC} & \text{R4} \\ \text{predmestie} & \text{R5} \\ \text{okrajové predmestie} & \text{R6} \\ \text{ďaleké predmestie} & \text{R7} \end{array}$$

Môžeme si všimnúť symetrickosť okolo regiónu $R4$. To znamená, že pre jednoduchosť nezávisí, či dochádzame z regiónov, ktoré sú od OC na západ (napr. $R2$ a $R3$) alebo z regiónov, ktoré sú na východ od tohto regiónu (napr. $R5$ a $R6$).

Ako by sme mohli očakávať, pre vzdialenejšie regióny, akú sú $R1$ a $R7$ máme dlhšie cestovné časy oproti bližším regiónom $R3$ a $R5$. To znamená, že čím je väčšia vzdialenosť od OC, tým sa nám zvýši i cestovný čas do OC. Môžeme teda povedať, že máme narušený predpoklad nezávislosti. Zároveň si treba uvedomiť, že cestovný čas v rámci regiónu $R4$ je 26 minút napriek vzdialenosti 0 km, pretože aj v rámci regiónu potrebujeme čas na presun do OC.

Množina pozorovaných cestovných časov predstavuje merania získané v konkrétny deň, t.j. spriemerované cestovné časy do OC za 24 hodín pre jednotlivé regióny. V tomto prípade, niektoré pozorovania môžu byť vysvetlené pomocou efektu kongescie (zahltenia). Preto môžeme očakávať, že dlhší cestovný čas v jednom regióne vedie k dlhším cestovným časom v susediacich regiónoch. Je to zapríčinené predpokladom jednej spoločnej cesty, ktorá je využívaná všetkými obyvateľmi siedmych regiónov. Napríklad kratší cestovná čas v regióne $R5$ môže viesť ku kratším cestovným časom v regiónoch $R6$ a $R7$. Z pohľadu modelovania, sa daný efekt zahltenia nedá vysvetliť premennými *Hustota* a *Vzdialenosť*, pretože cestovný čas v konkrétny deň ovplyvňujú iba cestovné časy susedných regiónov, tzv. efekt spätnej väzby. Naopak vzdialenosť sa nemení zo dňa na deň a takisto aj hustota obyvateľstva sa na dennej báze mení veľmi pomaly. Preto tieto premenné nie sú schopné vysvetliť denný jav zmeny cestovného času. Denné pozorované zmeny v cestovných časoch sa budú lepšie vysvetľovať, ak budeme brať do úvahy cestovné časy zo susedných regiónov. To znamená, že cestovné časy zo susediach pozorovaní y_j berieme ako vysvetľujúcu premennú pre cestovný čas v regióne i, y_i. Rovnako použijeme y_i ako vysvetľujúcu premennú pre cestovný čas v regióne j, y_j pre všetky $i, j \in \{1, 2, ..., 7\}$ a $i \neq j$.

Zadefinujme si jednoduchého suseda prvého druhu k regiónu $R1$. Je ním región $R2$, pretože je to jediný región, s ktorým má spoločnú hranicu. Rovnako, región $R2$ má dvoch susedov prvého druhu, regióny $R1$ a $R3$. Zostrojme $7x7$ rozmernú váhovú maticu V, kde riadky matice sú i-te pozorovanie a stĺpce s indexmi j sú susediace regióny k i-temu regiónu. Zároveň, pre susedov prvého druhu každého regiónu vo váhovej matici

V vyplníme na dané miesto 1, inak 0 ako sme popísali v sekcii (1.3.2)

$$V = \begin{pmatrix} & R1 & R2 & R3 & R4 & R5 & R6 & R7 \\ R1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ R2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ R3 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ R4 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ R5 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ R6 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ R7 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Diagonálne prvky matice sú nulové, teda regióny nie sú považované za suseda samého seba. V ďalšom kroku získanú maticu V riadkovo normalizujeme a dostaneme nezápornú maticu W

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

kde súčet v riadku je rovný 1.

Získanú maticu W priestorových váh vynásobíme hodnotami $7x1$ rozmerného vektora y a získame vektor priestorového posunu pre každé pozorovanie i, $i = 1, \dots, 7$:

$$Wy = \begin{pmatrix} y_2 \\ (y_1 + y_3)/2 \\ (y_2 + y_4)/2 \\ (y_3 + y_5)/2 \\ (y_4 + y_6)/2 \\ (y_5 + y_7)/2 \\ y_6 \end{pmatrix}$$

Na odhad modelu použijeme model SAR.

Zoberme do úvahy dopad zmeny populačnej hustotu v jednom regióne (napr. R2) na cestovné časy do OC pre všetky ostatné regióny. Presnejšie, zdvojnásobme populačnú hustotu regiónu R2 a urobme predikcie dopadu zmeny cestovných časov do OC pre všetky regióny.

$$\tilde{X} = \begin{pmatrix} 10 & 30 \\ \mathbf{40} & 20 \\ 30 & 10 \\ 50 & 0 \\ 30 & 10 \\ 20 & 20 \\ 10 & 30 \end{pmatrix}$$

Použijeme odhady parametrov vypočítaných pomocou metódy maximálnej vierohodnosti $\hat{\beta}' = [0.135, 0.561]$ a $\hat{\rho} = 0.642$. Odhadnutá hodnota parametra ρ udáva pozitívnu priestorovú závislosť v cestovných časoch. Predikcie modelu pre vysvetľované premenné matice X budú nadobúdať tvar:

$$\hat{y^{(1)}} = (I_n - \hat{\rho}W)^{-1}X\hat{\beta}.$$

V nasledujúcej tabuľke porovnajme predikcie modelu $\hat{y}^{(1)}$ s maticou vysvetľovaných premenných X a modelu $\hat{y}^{(2)}$ s maticou vysvetľovaných premenných $\tilde{X}$. Inak povedané ilustrujme ako model vytvára priestorové prelievanie *(spillover effect)* pri zmene hustoty populácie v jednom regióne:

Regióny	$\hat{y}^{(1)}$	$\hat{y}^{(2)}$	$\hat{y}^{(2)} - \hat{y}^{(1)}$
R1:	42.01	44.58	2.57
R2:	37.06	41.06	4.00
R3:	29.94	31.39	1.45
R4: CBD	26.00	26.54	0.53
R5:	29.94	30.14	0.20
R6:	37.06	37.14	0.07
R7:	42.01	42.06	0.05

Porovnaním dvoch predikcií $\hat{y}^{(1)}$, $\hat{y}^{(2)}$ môžeme pozorovať, že zmena hustoty populácie v regióne $R2$ má priamy efekt na tento región, ktorý spôsobí zvýšenie cestovného času pre obyvateľov regiónu $R2$ o 4 minúty. Zároveň táto zmena má aj nepriamy efekt, ktorý ovplyvní zvýšenie cestovných časov aj v ostatných regiónoch. Taktiež sa dá dobre pozorovať aj priestorový efekt susediacich regiónov. Čím sú regióny ďalej od regiónu $R2$, tým aj zmena cestovného času ovplyvnená zmenou hustoty populácie regiónu $R2$ je menšia. Obecne nám model hovorí, že zmena hustoty populácie v regióne $R2$ okamžite vedie k nárastu pozorovaných cestovných časov pre všetky regióny.

Porovnaním priestorového regresného modelu a obyčajného regresného modelu prídeme k záveru, že obyčajný regresný model urobí predikcie, ktorých vplyv zmeny hustoty obyvateľstva regiónu $R2$ ovplyvní cestovné časy iba v regióne $R2$, ako je znázornené v nasledujúcej tabuľke.

Regióny	$\hat{y}^{(1)}$	$\hat{y}^{(2)}$	$\hat{y}^{(2)} - \hat{y}^{(1)}$
R1:	42.98	42.98	0.00
R2:	36.00	47.03	11.02
R3:	29.02	29.02	0.00
R4: CBD	27.56	27.56	0.00
R5:	29.02	29.02	0.00
R6:	36.00	36.00	0.00
R7:	42.98	42.98	0.00

2 Odhady parametrov metódou MLE

V tejto kapitole si odvodíme odhadovanie parametrov modelov popísaných v (1) pomocou metódy maximálnej vierohodnosti (MLE). Odhad modelov budeme robiť na prierezových dátach (neobsahujú časovú štruktúru).

2.1 Odhadovanie parametrov modelov

2.1.1 Priestorový autoregresný model

Prvý model, ktorý budeme odhadovať metódou MLE je priestorový autoregresný model (SAR):

$$
\begin{aligned}
&y = \rho W y + X\beta + \epsilon;\\
&\epsilon \sim N_n(0, \sigma^2 I_n).
\end{aligned}
$$

Na odhad MLE potrebujeme likelihood funkciu závislej premennej y, ktorú získame ako transformáciu náhodnej premennej ϵ pomocou Jacobiho transformácie. Potom je logaritmická vierohodnostná funkcia pre odhad parametrov β, ρ, σ^2 definovaná ako:

$$
\ln L_y(\beta, \rho, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 + \ln \mid I_n - \rho W \mid -\frac{1}{2\sigma^2} e^T e,
$$

kde $e = y - \rho W y - X\beta$. Na odhad parametrov β a σ^2 použijeme podmienky prvého rádu.
Po zderivovaní logaritmickej vierohodnostnej funkcie podľa β a σ^2 dostávame odhady parametrov

$$
\begin{aligned}
\widehat{\beta}_{MLE} &= (X^T X)^{-1} X^T (I_n - \rho W) y\\
\widehat{\sigma}^2_{MLE} &= \frac{1}{n} e^T e,
\end{aligned}
$$

ktoré dosadíme do logaritmickej vierohodnostnej funkcie $\ln L_y(\beta, \rho, \sigma^2)$ a opäť použijeme podmienku prvého rádu na odhad parametra ρ:

$$
\widehat{\rho}_{MLE} = \operatorname{argmax} \frac{d(\ln L_y(\widehat{\beta}_{MLE}, \rho, \widehat{\sigma}^2_{MLE}))}{d(\rho)}
$$

2.1.2 Model s priestorovými chybami

Ďalší v poradí odhadneme model s priestorovými chybami (SEM):

$$y = X\beta + u;$$
$$u = \lambda W u + \epsilon$$
$$\epsilon \sim N_n(0, \sigma^2 I_n).$$

Postupujeme rovnako ako pri predchádzajúcom modely, na odhad MLE potrebujeme likelihood funkciu závislej premennej y, ktorú získame ako transformáciu náhodnej premennej ϵ pomocou Jacobiho transformácie. Logaritmická vierohodnostná funkcia pre SEM model nadobúda rovnaký tvar ako pri odhade SAR modelu, až na e:

$$e = y - \lambda W y - X\beta + \lambda W X\beta.$$

Na odhad opäť použijeme podmienky prvého rádu:

$$\widehat{\beta}_{MLE}(\lambda) = ((X - \lambda W X)^T (X - \lambda W X))^{-1} (X - \lambda W X)^T (y - \lambda W y)$$
$$\widehat{\sigma}^2_{MLE}(\lambda) = \frac{1}{n} z(\lambda)^T z(\lambda),$$

kde $z(\lambda) = y - \lambda W y - (X - \lambda W X)\widehat{\beta}_{MLE}(\lambda)$.
Na odhad parametra λ sa môže použiť iteračná metóda navrhnutá Anselinom, ktorá je popísaná v diplomovej práci [6].

2.1.3 Priestorový Durbinov model

Posledný v poradí odhadneme priestorový Durbinov model:

$$y = \lambda W y + X\beta + W X\theta + v;$$

Postupujeme rovnako ako pri predchádzajúcich modeloch a dostaneme rovnakú logaritmickú vierohodnostnú funkciu ako pre SAR (SEM) model (až na e):

$$e = y - \lambda W y - X\beta - W X\theta.$$

Na logaritmickú vierohodnostnú funkciu opäť použijeme podmienky prvého rádu pre odhad parametrov β, θ a σ^2. Tie potom opätovne dosadíme do logaritmickej vierohodnostnej funkcii a použijeme pre odhad parametra λ. Odhadnuté parametre kvôli

výpočtovej náročnosti nebudeme odhadovať, dostatočne dobre sú opísané v diplomových prácach [12], [6].
Odhady pre modely panelových dát nebudeme uvádzať, pretože sú dostatočne dobre spracované v diplomových prácach [12], [6].

2.2 Moranov korelačný koeficient

Pre overenie priestorovej autokorelácie na dátach použijeme Moranov korelačný koeficient, Moranovo I. Koeficient hľadá odchýlky od náhodných usporiadaní v priestore, t.j. hľadá zhlukovanie dát. Takéto zhlukovanie porušuje náhodnosť, a tým pádom vykazuje priestorovú autokoreláciu. Moranovo I sa používa na prierezové dáta a je popísané ako:

$$I_t = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} (x_{i,t} - \overline{x})(x_{j,t} - \overline{x}) w_{ij}}{m \sum_{i=1}^{n} (x_{i,t} - \overline{x})^2},$$

kde $x_{i,t}$, $x_{j,t}$ je uvažovaná premenná regiónu i, j, v čase t, $\overline{x}$ je priemer premennej x, n je počet regiónov, m je súčet všetkých váh a w_{ij} je váha regiónu i vzhľadom na región j. Hodnota Moranovho I sa pohybuje v rozmedzí $[-1, +1]$. V prípade, že sa Moranovo I nachádza blízko 0, tak hovoríme, že dáta nevykazujú priestorovú autokoreláciu. V opačných prípadoch, ak I_t je blízke 1 hovoríme o pozitívnej priestorovej autokorelácii, inak o negatívnej priestorovej autokorelácii.

3 Popis dát a rôzne konštrukcie váhovej matice

V tejto kapitole budeme aplikovať vyššie popísané metódy priestorovej ekonometrie na reálne dáta, ktoré sú zozbierané na základe ekonomických ročných výsledkov a štatistík pre všetky NUTS 2 regióny štátov Európskej únie.

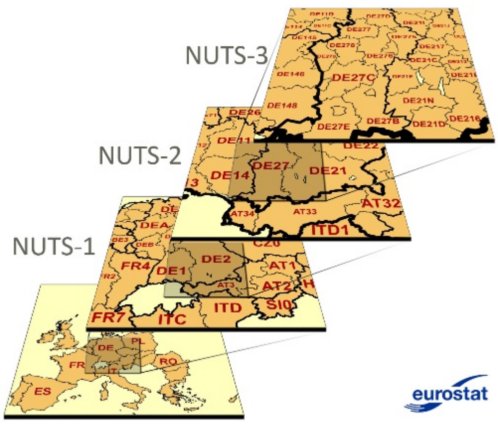
3.1 NUTS 2 regióny

Ako už bolo spomenuté, dáta boli zozbierané vzhľadom na delenie zavedené Štatistickým úradom Európskej komisie - Eurostatom v spolupráci s národnými inštitútmi pre štatistiku, napr. v rámci Slovenska to je Štatistický úrad Slovenskej republiky. Nomenklatúra územných štatistických jednotiek (NUTS), v súčasnosti NUTS 2013 platná od 1. januára 2015 je rozdelenie do vyšších štatistických úrovní, t.j. regiónov a krajov. V súčasnosti existujú tri úrovne delenia NUTS a dve úrovne delenia LAU (Lokálna štatistická územná jednotka, t.j. okresy a obce). Klasifikácia regiónov pomocou NUTS je hierarchický systém, ktorý zhromažďuje dáta viacerých štátov. Tieto dáta sú prehľadnejšie, pretože sa nachádzajú na jednom mieste a dá sa s nimi lepšie pracovať. NUTS sa delia na 3 úrovne:

NUTS 1: hlavné sociálno - ekonomické regióny (98 regiónov)

NUTS 2: základné regióny na uplatňovanie regionálnych politík (276 regiónov)

NUTS 3: malé regióny pre špecifické analýzy (1342 regiónov)



Obr. 3: Rozdelenie NUTS regiónov (Zdroj: Eurostat)

Každý z regiónov má jedinečný kód, ktorý slúži na presnú identifikáciu regiónu a je zostavený nasledovne:

NUTS 1: obsahuje 3 znaky, kde prvé dva sú skratka štátu (napr. AT, DE, SK) a tretí znak je buď pismeno alebo číslica (napr. SK0, UKC)

NUTS 2: obsahuje štyri znaky, kde prvé dva sú opäť skratka štátu, ďalší znak je z predchádzajúceho delenia a posledný je číslica (napr. SK04, UKM3)

NUTS 3: obsahuje päť znakov, ktoré sú rozdelené podľa rovnakej logiky ako pri predchádzajúcom delení (napr. SK042, UKD42)

Naše reálne dáta boli zozbierané, ako sme už spomenuli vyššie, vzhľadom na úroveň NUTS 2, t.j. máme 276 pozorovaných regiónov štátov Európskej únie.

Štáty:	Počet regiónov:
AT - Rakúsko	9
BE -Belgicko	11
BG - Bulharsko	6
CY - Cyprus	1
CZ - Česko	8
DE - Nemecko	38
DK - Dánsko	5
EE - Estónsko	1
EL - Grécko	13
ES - Španielsko	19
FI - Fínko	5
FR - Francúzsko	27
HR - Chorvátsko	2
HU - Maďarsko	7
IE - Írsko	2
IT - Taliansko	21
LT - Litva	1
LU - Luxembursko	1
LV - Lotyšsko	1
MT - Malta	1
NL - Holandsko	12
PL - Poľsko	16
PT -Portugalsko	7
RO - Rumunsko	8
SE - Švédsko	8
SI - Slovinsko	2
SK - Slovensko	4
UK - Veľká Británia	40

Obr. 4: NUTS 2 regióny a ich počet (Zdroj: Eurostat)

Teraz keď už vieme dáta presne identifikovať, tak si zadefinujeme premenné, ktoré budeme používať v modeloch.

3.2 Premenné vstupujúce do modelu

Do nášho modelu nám vstupujú nasledovné premenné, ktorými sú:

HDP - totálny hrubý domáci produkt regiónu, v mil.

Miera nezamestnanosti - celková nezamestnanosť v danom regióne, v %

Miera inflácie - ročná inflácia daného štátu, v %

Hustota obyvateľstva - hustota obyvateľstva daného regiónu na km^2

Časové okno, počas ktorého budeme naše premenné pozorovať si zvolíme medzi rokmi 2005−2014. Zvolili sme si najširšie okno, ktoré nám dovoľujú zozbierané štatistiky pre jednotlivé NUTS 2 regióny. Pre ostatné roky buď boli dáta sčasti neúplné alebo úplne nedostupné. Je to spôsobené aj tým, že zber regionálnych štatistík a celková informatizácia nebola v tom období na takom stupni ako v súčasnosti - napr. len rok pred rokom 2005 vstúpilo do spoločného európskeho priestoru a EÚ 10 štátov.

3.2.1 Hrubý domáci produkt

Prvá pozorovaná ekonomická premenná je hrubý domáci produkt (HDP). Je to základný makroekonomický ukazovateľ, ktorým meriame celkovú peňažnú hodnotu tovarov a služieb vyprodukovaných za dané pozorované obdobie v danej ekonomike. My budeme za obdobie považovať 1 kalendárny rok a zozbierané údaje sú vzhľadom na NUTS 2 regióny. HDP môže byť počítaný viacerými metódami, my si určíme tri základné:

1. **Výrobná metóda:** tvorí ju súhrn všetkých finálnych tovarov a služieb, ktoré boli za dané časové obdobie vyrobené.

2. **Výdavková metóda:** počíta sa ako výdaje jednotlivých sektorov na nákup finálnych tovarov a služieb. Premenné, ktoré vstupujú do výpočtu HDP sú nasledovné:

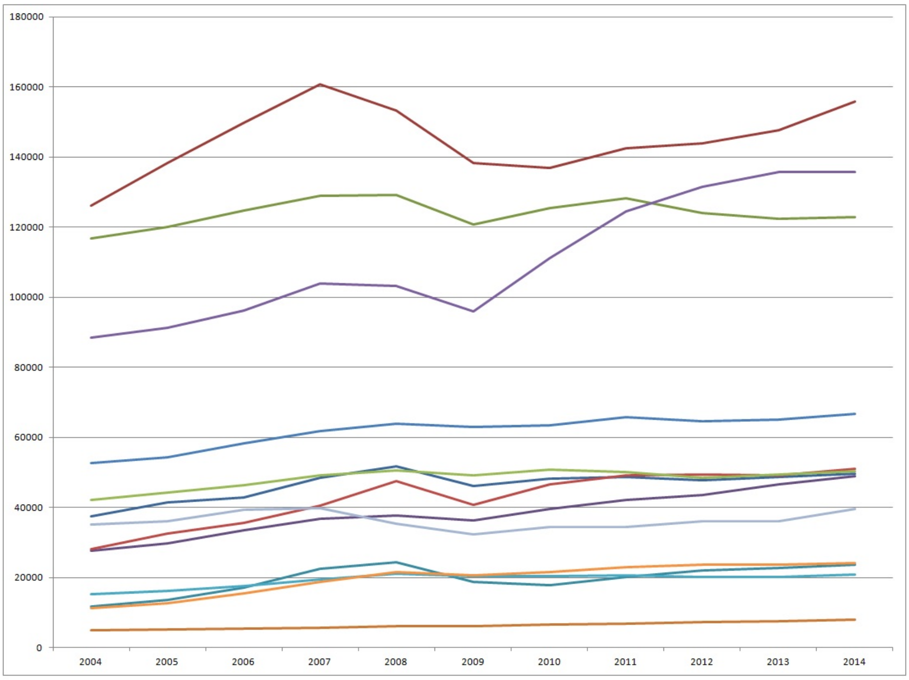
náklady - C, investície - I, vládne výdavky - G, export - E a import - I. Formula na výpočet HDP určená týmito premennými nadobúda tvar:

$$HDP = C + I + G + (E - I)$$

3. Dôchodková metóda: je daná ako súčet hrubého národného dôchodku (HND), amortizácie (A) a nepriamych daní (N). Potom výpočet HDP nadobúda tvar:

$$HDP = HND + A + N$$

V našej práci používame HDP vypočítané podľa 1. metódy, t.j. ako celková hodnota všetkých tovarov a služieb znížená o hodnotu tovarov a služieb používaných na medzispotrebu pri ich výrobe. HDP je merané v miliónoch eur. Na nasledujúcom grafe je zobrazená ukážka vývoja HDP pre pár vybraných regiónov, na ktorom môžeme pozorovať prepad rastu HDP na začiatku roku 2008, kedy bola kríza. Priemerné HDP EÚ sa dostalo na svoju pôvodnú úroveň až o tri roky neskôr a od tohto roku, 2011, je v neustálom raste.



Obr. 5: Vývoj HDP pre pár vybraných regiónov

3.2.2 Miera nezamestnanosti

Ďalšia pozorovaná makroekonomická premenná je miera nezamestnanosti (u). Je to podiel nezamestnaných osôb (U) ku všetkým osobám schopných pracovať, t.j. pracovnej sile (L):

$$u = \frac{U}{L},$$
$$L = U + E,$$

kde E je počet zamestnaných ľudí.

Pracovná sila L je tvorená ľuďmi, ktorí podľa zákona môžu pracovať. Do tejto skupiny nepatria ľudia, ktorí už sú na dôchodku a ani ľudia, ktorí si momentálne nehľadajú pracovnú pozíciu - môže ísť o ľudí, ktorí sú práve na materskej alebo rodičovskej dovolenke, alebo študujú na strednej či vysokej škole. Tým pádom do pracovnej sily patria aj všetky nezamestnané osoby. Za nezamestnanú osobu sa považuje pracovná sila, ktorá spĺňa nasledovné podmienky:

- vek má minimálne 15 rokov;
- v súčasnosti nepracuje a ani neštuduje;
- nepoberá žiadny z dôchodkov ako sú: starobný, výsluhový, invalidný;
- v súčasnosti nepodniká;
- nie je evidovaná v systéme ako dobrovoľne nezamestnaná;
- je práceschopná.

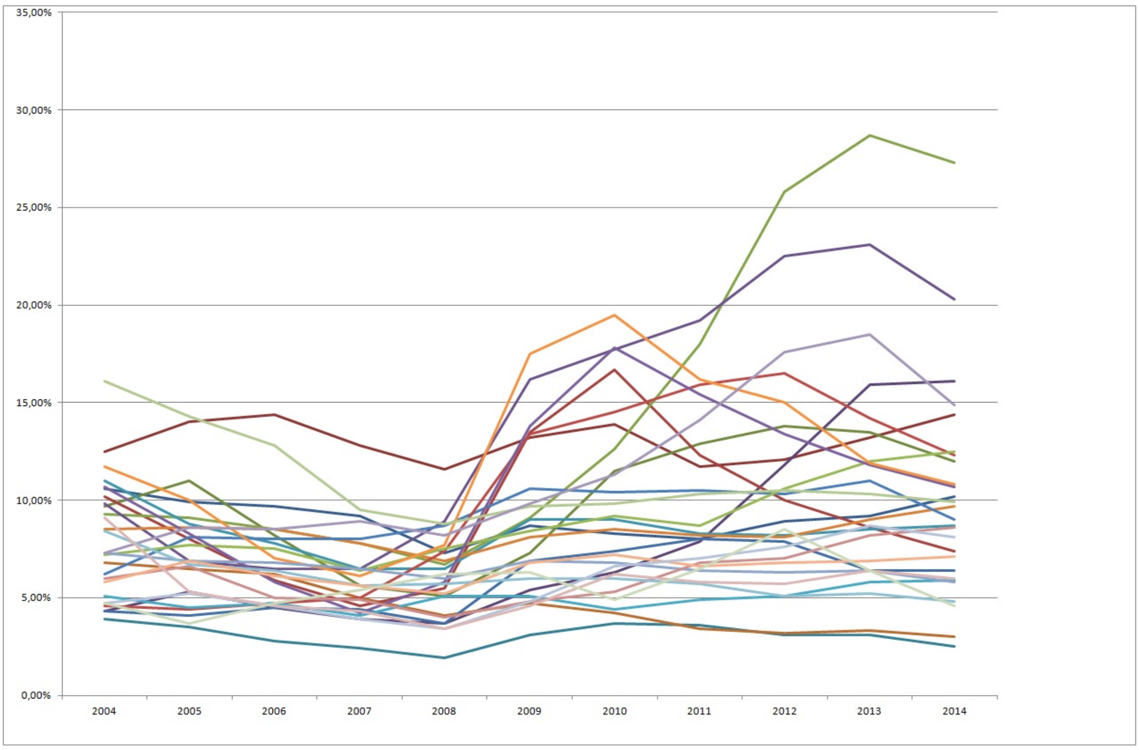
Podľa formy príčiny rozdeľujeme nezamestnanosť na tri hlavné typy:

Cyklický: tento typ nezamestnanosti nastáva v cykloch - v týchto cykloch nadobúda ekonomika mnoho výkyvov. V čase hospodárskej recesie rastie, naopak v čase hospodárskej expanzie klesá.

Frikčný: taktiež trendová nezamestnanosť je krátkodobý typ nezamestnanosti. Vzniká z dôvodu zmeny zamestnania niektorých zamestnancov - takáto zmena potrebuje čas, aby sa potenciálni zamestnanci spojili s novými zamestnávateľmi, alebo hľadania si práce absolventami, či návrat do zamestnania po rodičovskej dovolenke.

Štrukturálny: je to typ nezamestnanosti, kedy chýba dopyt po určitom type zamestnanca. Je to spôsobené obrovským pokrokom v oblasti technológii, ale aj presúvaním zamestnaneckých pozícii do regiónov s nižšími nákladmi na pracovníka.

Miera nezamestnanosti pre naše dáta zozbierané s ohľadom na regióny NUTS 2 je zobrazená na nasledujúcom grafe pre pár vybraných regiónov. Opäť môžeme pozorovať dôsledok finančnej krízy, kedy nám nezamestnanosť v roku 2008 začala významne rásť. Jej rast sa vo veľkej miere zastavil a zastabilizoval na prelome roka 2013, kedy priemerná nezamestnanosť v EÚ klesla z $10,95\%$ na $10,28\%$ a tento trend ďalej pokračoval.



Obr. 6: Vývoj miery nezamestnanosti pre pár vybraných regiónov

3.2.3 Miera inflácie

Predposledná pozorovaná makroekonomická premenná je miera inflácie (π). Miera inflácie nám hovorí o miere nárastu cenovej hladiny a počíta sa nasledovne:

$$\pi_t = \frac{P_t - P_{t-1}}{P_{t-1}},$$

kde

P_t : je buď deflátor HPD alebo index spotrebiteľských cien v čase t

π_t : je miera inflácie v roku t.

Inflácia sa dá vysvetliť aj tak, že 1 peňažná jednotka dnes nie je to isté ako 1 peňažná jednotka včera alebo zajtra. Premenné, ktoré vstupujú do výpočtu miery inflácie sú definované nasledovne:

Deflátor HDP: odráža vývoj cenovej hladiny a mieru znehodnotenia HDP, teda infláciu cien v ekonomike. Vypočíta sa ako podiel nominálneho a reálneho HDP v roku t:

$$P_t = \frac{HDPn}{HDPr}.$$

Rozdiel medzi nominálnym a reálnym HDP spočíva v tom, že nominálny HDP je meraný v bežných, trhových cenách za dané obdobie, pričom reálny HDP je meraný v cenách iného roku (zvyčajne predchádzajúceho).

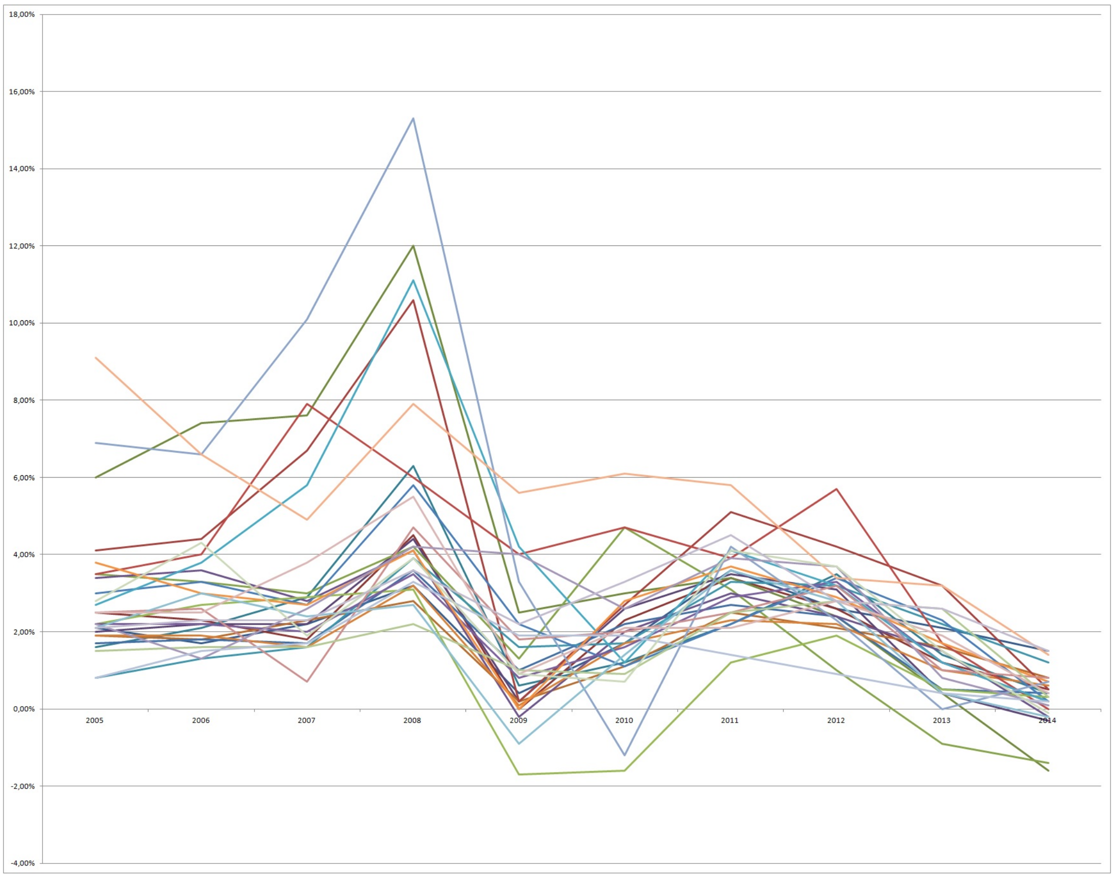
Index spotrebiteľských cien: predstavuje ceny platené spotrebiteľmi (alebo domácnosťami) a počíta sa pomocou tzv. spotrebného koša. Spotrebný kôš v danom roku (základnom roku) zachytáva náklady priemernej domácnosti a porovnáva ich s nákladmi domácnosti v budúcich rokoch.

Inflácia podáva dôležité informácie o vývoji ekonomiky a v rozumnej miere je jej zdravou súčasťou. Podľa rastu cien sa inflácia delí na:

- **Mierna** medziročná zmena cien medzi $1\% - 9\%$
- **Cválajúca** medziročná zmena cien medzi $10\% - 1000\%$

- **Hyperinflácia** medziročná zmena cien vyjadrená viac ako štvorciferným číslom

V prípade poklesu cien, inflácia je záporná, hovoríme o deflácii. Mnohí ekonómovia veria, že deflácia v súčasnej ekonomike predstavuje problém, lebo môže viesť k deflačnej špirále. Oproti inflácii, ktorá v čase znižuje reálnu hodnotu peňazí ju deflácia zvyšuje. Preto sa spotrebiteľ teoreticky môže rozhodnúť odložiť svoju spotrebu na neskôr, čím sa môže ekonomika dostať do špirály, ktorá môže viesť až k jej celému kolapsu.
Naše dáta, v tomto prípade ich máme zozbierané iba vzhľadom na štáty, t.j. 28 štátov, pretože v rámci štátu je miera inflácie vypočítaná z rovnakého spotrebiteľského koša. Eurostat zbiera informácie o miere inflácie ako harmonizované indexy spotrebiteľských cien - tie sa zostavujú pre medzinárodné porovnávanie inflácie spotrebiteľských cien.



Obr. 7: Vývoj miery inflácie pre štáty EÚ

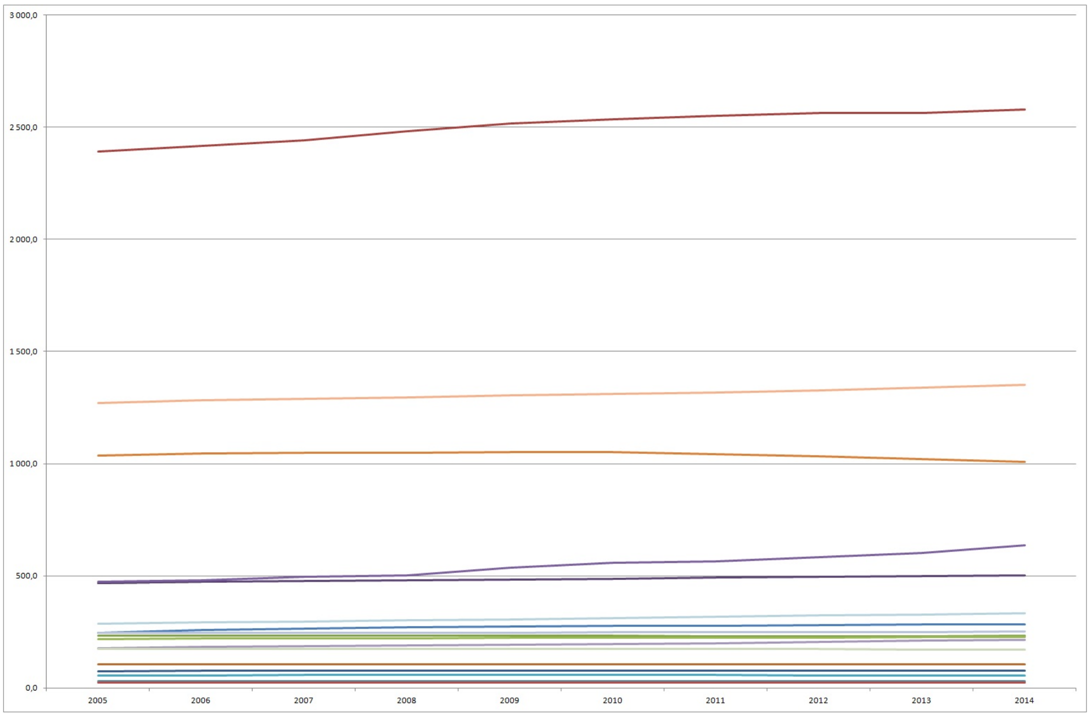
Na grafe (7) máme zobrazené miery inflácie pre všetkých 28 štátov EÚ, na ktorých môžeme pozorovať obrovský nárast v roku začatia finančnej krízy, 2008, a zároveň strmý pád v nasledujúcich obdobiach. Pokračujúci pokles aj v rokoch po kríze sa dá vysvetliť zavedenou monetárnou politikou Európskej Centrálnej Banky (ECB), ktorá sa po recesných rokoch krízy rozhodla opätovne naštartovať rast HDP. Dosiahla to nízkymi úrokovými sadzbami, vďaka ktorým získali finančné inštitúcie lepší prístup k zdrojom peňažného trhu.

3.2.4 Hustota populácie

Posledná pozorovaná premenná je hustota obyvateľstva. je to údaj, ktorý hovorí o podiele počtu obyvateľov a rozlohe územia:

$$H = \frac{P}{A},$$

kde H je hustota, P je populácia daného regiónu a A je jeho rozloha. Na nasledujúcom grafe môžeme pozorovať vývoj hustoty populácie v rozmedzí rokov $2005 - 2014$ pre pár vybraných regiónov.



Obr. 8: Vývoj hustoty populácie pre štáty EÚ

3.3 Rôzne prístupy konštrukcie matice W

V nasledujúcej podkapitole sa budeme venovať rôznym prístupom k odhadom váhovej matice W.

3.3.1 Váhová matica typu Queen contiguity

Prvý spôsob, ktorý si zvolíme na určenie váhovej matice bude najtriviálnejší spôsob typu Queen contiguity. Tento spôsob priradí 1 pod podmienkou, že regióny majú buď spoločnú časť hranice alebo spoločný hraničný bod, inak priradí 0. Túto metódu sme skonštruovali nasledovne:

- zo stránky http://ec.europa.eu/eurostat/web/nuts/history sme si stiahli údaje o NUTS regiónoch, z ktorých sme filtrovali údaje o regiónoch NUTS 2;
- každému regiónu sme priradili jeho susedov;
- podľa predchádzajúceho priradenie sme skonštruovali váhovú maticu W.

Váhová matica W typu Queen contiguity je v našom prípade riedka matica - z celkového počtu 75900 možných susedstiev (už očistených o nulovú diagonálu) máme vyplnených jednotkou 1168 čo tvorí necelých $1,54\%$.
Avšak vyskytujú sa nám tu regióny, ktoré z dôvodu nesplnenia obidvoch kritérií nemajú žiadneho suseda - je ich 21. Sú to nasledujúce regióny:

Kód	Názov regiónu	Kód	Názov regiónu
CY00	Kypros	EL41	Voreio Aigaio
EL42	Notio Aigaio	EL43	Kriti
EL62	Ionia Nisia	ES53	Illes Balears
ES63	Ciudad Autónoma de Ceuta	ES64	Ciudad Autónoma de Melilla
ES70	Canarias	FI20	Åland
FR83	Corse	FRA1	Guadeloupe
FRA2	Martinique	FRA3	Guyane
FRA4	La Réunion	FRA5	Mayotte
ITG1	Sicilia	ITG2	Sardegna
MT00	Malta	PT20	Região Autónoma dos Açores
PT30	Região Autónoma da Madeira		

Ako je možné vidieť, jedná sa o regióny, ktoré sú buď ostrovné európske ($CY00$, $EL43$,...), ostrovné zámorské ($ES70$, $FRA4$,...), alebo autonómne oblasti v afrických štátoch ($PT20$, $PT30$). Tieto regióny sme nútený vylúčiť z váhovej matice W typu queen contiguity, pretože narušujú predpoklad pozitívnej definitnosti - tieto riadky sú lineárne závislé a spôsobujú singulárnosť váhovej matice. Ďalej budeme pokračovať iba s váhovou maticou dimenzie 255, ktorá má vyplnených $1,80\%$ susedstiev.

3.3.2 Váhová matica vzdialeností

Ďalší prístup k tvorbe váhovej matice už nebude taký triviálny ako predchádzajúci typ Queen contiguity. V tomto type priradíme každej dvojici regiónov leteckú vzdialenosť ich úradných centier. Celý tento postup bol rozdelený do nasledujúcich bodov:

1. zhromaždenie údajov o NUTS 2 regiónoch;
2. výpočet vzájomných vzdialeností;
3. preškálovanie váhovej matice.

V nasledujúcich častiach si podrobnejšie popíšeme jednotlivé body tvorby váhovej matice vzdialeností:

1.) **Zhromažďovanie údajov:** Prvým bodom pri tvorbe váhovej matice vzdialeností bude zhromaždenie si údajov o úradných centrách NUTS 2 regiónov. Existuje viacero prístupov ako sa s týmto problémom vysporiadať. My sme si vybrali ten najjednoduchší, a to použiť databázu o regiónoch zo stránky http://www.brrg.de/database.php?language=en&cId=0&dId=47, ktorá žiaľ bola z roku 1995. Preto sme túto databázu overili a údaje chýbajúcich regiónov doplnili.
Druhý spôsob tvorby váhovej matice by bol taký, že každému regiónu určíme jeho ťažisko, tzv. centrum centroidu.

2.) **Výpočet vzdialeností:** Ďalším bodom je výpočet vzdialeností úradných miest NUTS 2 regiónov. Naša databáza obsahuje 276 regiónov, preto je potrebné vypočítať vzdialenosti pre 37950 dvojíc, čo nie je ľahká úloha. Nami zvolené dva postupy nám ju ale výrazne zjednodušia.

2.1.) **Automatické získavanie dát:** Prvotná myšlienka nemanuálneho a nie časovo náročného výpočtu bola naprogramovať kód, ktorý by z vopred vybranej internetovej stránky dokázal pomocou údajov z našej databázy úradných miest regiónov NUTS 2 získať ich vzájomné letecké vzdialenosti, a tieto zapísať do príslušného hárku našej databázy.
Ako prvé sme si vybrali stránku do ktorej budeme zadávať údaje a zároveň budeme údaje aj extrahovať a ukladať do vopred vybraného zošita. Zvolili sme si nasledovnú stránku: http://www.distancefromto.net/, ktorá sa nám svojou jednoduchou štruktúrou zdala adekvátna pre naše požiadavky.
Druhá, komplikovanejšia časť prišla hneď po výbere - naprogramovať daný kód. S týmto problémom som sa rozhodol osloviť kolegyňu Bc. Olíviu Kunertovú z Fakulty matematiky, fyziky a informatiky, UK BA, ktorá navrhla a zostrojila funkčný kód. Celý kód na automatické získavanie dát z internetovej stránky je priložený v prílohe č.1.
Ale výsledok celého nášho úsilia nedosiahol očakávania, pretože pravdepodobne nami vybraná internetová stránka mala ochranu pred robotmi, t.j. po určitom počte opakovaní zadaného úkonu (cca. 50) nás daná stránka odpojila a museli sme program manuálne obnoviť aby pokračoval. V prípade, že by sme zvolili krok zadávania každú sekundu, tak by celkový čas potrebný

na zhromaždenie údajov presiahol 10 hodín. Preto sme sa rozhodli náš kód použiť na iných stránkach, na ktorých však nasledoval rovnaký výsledok. Nakoniec sme sa rozhodli od tohto postupu upustiť a vymyslieť iný, značne jednoduchší.

2.2.) **Haversíniho formula:** Druhý spôsob bola snaha získať, vypočítať vzdialenosti pomocou voľne dostupných informácii. Tieto informácie sme získali z voľne dostupnej databázy, ktorá obsahuje údaje o zemepisných šírkach a zemepisných dĺžkach väčšiny miest na svete. Nachádza sa na stránke http://www.naturalearthdata.com/. Tieto údaje sme prefiltrovali a upravili na našu databázu úradných miest.

Tvorba databázy rozšírenej o zemepisnú šírku a zemepisnú dĺžku bola prvý krok k výpočtu vzdialeností, na ktoré použijeme Haversíniho formulu. Haversíniho formula pre dva rôzne body na sfére slúži na výpočet centrálneho uhla medzi nimi:

$$hav(\frac{d}{R}) = hav(\varphi_2 - \varphi_1) + cos(\varphi_1)cos(\varphi_2)hav(\lambda_2 - \lambda_1),$$

kde

- $hav(\theta) = sin^2(\frac{\theta}{2}) = \frac{1-cos(\theta)}{2}$ je haversíniho funkcia;
- d je vzdialenosť medzi danými dvoma bodmi, sférická vzdialenosť;
- R je polomer sféry (v našom prípade polomer zeme $R = 6378,137km$);
- $\varphi_{1,2}$ je zemepisná šírka (v rad.);
- $\lambda_{1,2}$ je zemepisná dĺžka (v rad.).

Použitím funkcie arcsin na haversíniho funkciu dostaneme vzorec na výpočet vzdialenosti dvoch bodov na sfére:

$$d = 2R \arcsin\left(\sqrt{\sin^2(\frac{\varphi_2 - \varphi_1}{2}) + \cos(\varphi_2)\cos(\varphi_1)\sin^2(\frac{\lambda_2 - \lambda_1}{2})}\right).$$

Vypočítané vzdialenosti sa môžu mierne líšiť od vzdialeností, ktoré sa dajú získať z rôznych verejných zdrojov. Môže to byť spôsobené výpočtovým softvérom (zaokrúhľovanie), alebo detailnejšími údajmi o zemepisných šírkach, zemepisných dĺžkach a polomeru Zeme.

3.) Preškálovanie váhovej matice: Posledný bod tvorby váhovej matice je jej preškálovanie, ktoré vyžadujeme z dôvodu zosilnenia priestorových vplyvov pre bližšie (susednejšie) regióny a zoslabenie pre vzdialenejšie regióny. Táto podmienka zatiaľ nie je splnené a blízke regióny majú malé priestorové vplyvy. Tie zosilníme preškálovaním, ktoré zvolíme nasledovne:

$$\widehat{w_{ij}} = \frac{1}{w_{ij}},$$

kde w_{ij} je vzdialenosť i regiónu od j regiónu. Toto preškálovanie zosilnilo priestorový vplyv bližších regiónov a zároveň zachovalo pomer všetkých ostatných vzdialeností.

V tejto podkapitole sme si uviedli tvorbu symetrickej neznormalizovanej váhovej matice W, ale ako sme už uviedli v (1.3.2) vo výpočtoch používame riadkovo normalizovanú váhovú maticu, ktorá vo väčšine prípadov už nie je symetrická.

4 Analýza dát NUTS 2 regiónov

V poslednej kapitole si popíšeme panelové dáta a modely pre panelové dáta. Zároveň si priblížime model popísaný reálnymi dátami NUTS 2 regiónov krajín EÚ predstavených v (3), vytvorené rôzne váhové matice W, odhady parametrov modelu v štatistickom softvéri `R` a interpretovanie dosiahnutých výsledkov.

4.1 Panelové dáta

V našej práci sa budeme venovať analýzam panelových dát, t.j. dát ktoré pozorujeme viac než jedno časové obdobie. V panelových dátach vystupujú dve hlavné zložky:

- počet pozorovaní n;
- časová perióda t.

Hlavný rozdiel panelových dát oproti časovým radom je veľkosť jednotlivých zložiek. V časových radoch je veľkosť časovej periódy niekoľkokrát vyššia než počet pozorovaní, pričom v panelových dátach to je naopak. V prípade, že sa časová perióda $t = 1$, tak hovoríme o prierezových dátach.
Pri analyzovaní panelových dát sa uvažuje s piatimi možnými modelmi:

1. Model s fixnými efektami;
2. Model s náhodnými efektami;
3. Model s fixnými koeficientami;
4. Model s náhodnými koeficientami;
5. Viacúrovňový model.

V našom prípade budeme pracovať iba s modelmi s fixnými a s náhodnými efektami. V každom modely môžu vystupovať buď priestorovo, alebo časovo špecifické efekty. Tieto efekty v ňom môžu byť zahrnuté obidva, iba jeden alebo žiaden. Podrobnejšie informácie o efektoch v panelových dátach sú spracované v diplomovej práci [6].

4.2 Model makroekonomických dát

Závislou premennou y nášho modelu budeme označovať **Unemployment_Total**, ktorá nám hovorí o miere nezamestnanosti v NUTS 2 regiónoch krajín EÚ a je vyjadrená v %. Do vysvetľujúcej premennej X nám bude vstupovať 5 premenných:

GDP: HDP NUTS 2 regiónov krajín EÚ v mil.;

Population: hustota obyvateľstva na km^2;

Inflation: miera inflácie pre štáty EÚ v %

Unemployment_Male: miera nezamestnanosti mužov NUTS 2 regiónov krajín EÚ;

Unemployment_Female: miera nezamestnanosti žien NUTS 2 regiónov krajín EÚ.

Vzhľadom nato, že nám všetky premenné okrem *GDP* a *Population* vstupujú do modelu v percentách, tak sme sa ich rozhodli zahrnúť do modelu ako log premenné, t.j. log(*GDP*) a log(*Population*). V našom modely budeme pozorovať závislosť nezamestnanosti NUTS 2 regiónov krajín EÚ od hrubého domáceho produktu, hustoty obyvateľstva, inflácie, nezamestnanosti mužov a nezamestnanosti žien. Primárnym cieľom našich bude odhad parametrov a ich signifikantnosti pre rôzne modely konštruovaného s rôznymi váhovými maticami. Skôr ako začneme odhadovať, tak môžeme predpokladať vysokú závislosť zmeny nezamestnanosti od nezamestnanosti mužov a žien.
Odhadovať budeme dva modely, ktoré budú ešte rozdelené podľa zahrnutia fixných (FE), alebo náhodných efektov (RE):

SAR: Priestorový autoregresný model

$$
\begin{aligned}
Unemployment\_Total_{it} = {} & \rho \sum_{j=1}^{n} w_{ij} Unemployment\_total_{jt} + \beta_1 log(GDP)_{it} + \\
& + \beta_2 log(Population)_{it} + \beta_3 Inflation_{it} + \\
& + \beta_4 Unemployment\_Male_{it} + \beta_5 Unemployment\_Female_{it} + \\
& + \mu_i + \delta_t + \epsilon_{it};
\end{aligned}
$$

SEM: Model s priestorovými chybami

$$\begin{aligned} Unemployment\_Total_{it} = &\beta_1 log(GDP)_{it} + \beta_2 log(Population)_{it} + \beta_3 Inflation + \\ &+ \beta_4 Unemployment\_Male_{it} + \beta_5 Unemployment\_Female_{it} + \\ &+ \mu_i + \delta_t + u_{it} \\ u_{it} = &\lambda \sum_{j=1}^{n} w_{ij} u_{it} + \epsilon_{it}. \end{aligned}$$

4.3 Váhové matice

Pri definovaní modelu sme sa rozhodli, že použijeme nasledovné druhy váhových matíc:

1. váhovú maticu typu Queen contiguity očistenú o regióny bez suseda. Táto matica obsahuje váhy pre 255 regiónov.

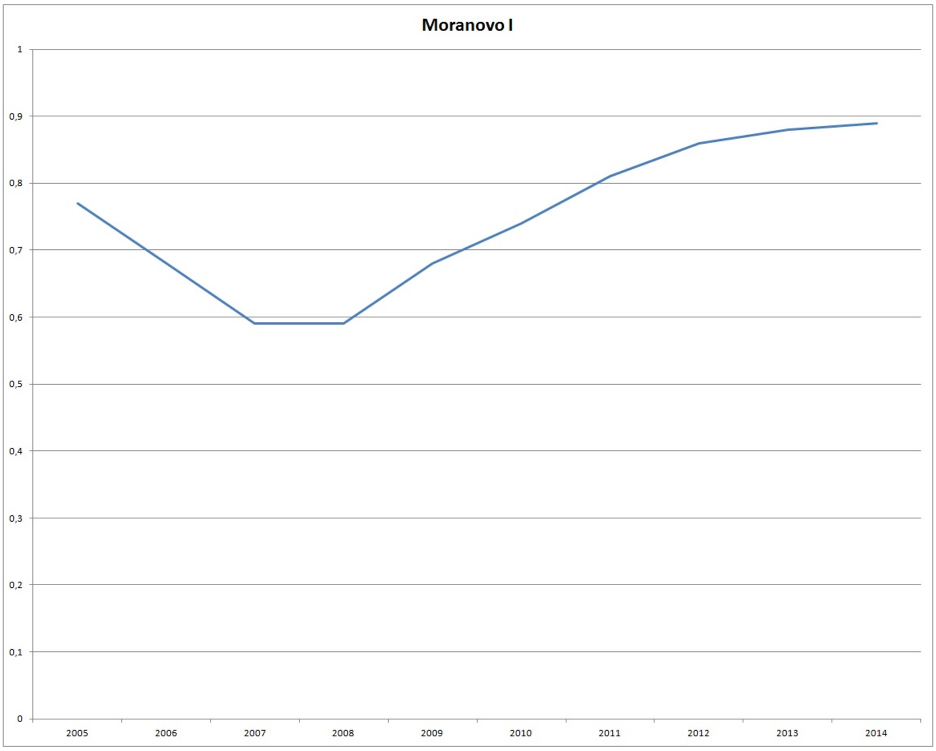
2. váhovú maticu, ktorá každému regiónu so susedom typu Queen contiguity priradí preškálovanú vzdialenosť. Táto matica obsahuje váhy pre 254 regiónov. Je to spôsobené tým, že dva regióny $CZ01$ a $CZ02$ majú nulovú vzdialenosť, kvôli úradnemu centru, ktoré majú spoločné (Praha).

3. váhovú maticu typu Queen contiguity prerobenú tak, aby priradila váhu iba susedovi z rovnakého štátu. Táto matica je oproti typu Queen contiguity zredukovaná a obsahuje váhy pre 250 regiónov.

4. posledný typ váhovej matice priradí preškálované vzdialenosti regiónom rovnakého štátu. Táto matica je medzi práve uvedenými typmi váhových matíc najrozmernejšia, obsahuje váhy pre 270 regiónov.

4.4 Odhad jednotlivých modelov

V nasledujúcej podkapitole si uvedieme odhady pre SAR a SEM modely kombinované s náhodnými a fixnými efektami. Avšak ako prvé si otestujeme prítomnosť priestorovej autokorelácie v dátach pomocou Moranovho I. V softvéri `R` sa táto funkcia nachádza v balíčku `spdep` a je definovaná ako funkcia `moran.test`. Do funkcie vstupujú dva argumenty, premenná a váhova matica, a výsledkom je vypočítana veľkosť Moranovho



Obr. 9: Vývoj Moranovho I

I, ktoré nám hovorí o sile priestorovej autokorelácie. Na danom obrázku môžeme pozorovať zmenu Moranovho I pre premennú *Unemployment_Total* počas 10 rokov.

Z (9) môžeme pozorovať, že Moranovo I bolo počas celých 10 rokov nad hodnotou 0.5, preto môžeme hovoriť o kladnej priestorovej autokorelácii. Odhady Moranovho I pre zvyšné váhové matice a premenné vyšiel veľmi podobne ako je naznačené na obrázku (9). Preto môžeme v našich dátach uvažovať o priestorovej autokorelácii.
Ako ďalšie si odhadneme parametre modelov pre rôzne váhové matice. Nato sa v `R` používa balíček `splm`, v ktorom použijeme funkciu `spml`, do ktorej vstupujú modelované premenné, váhová matica a ďalšie podmienky na špecifikáciu modelu (fixné efekty, ...).

1. typ Ako prvý budeme odhadovať model s váhovou maticou typu Queen contiguity očisteného o regióny bez suseda. Odhady parametrov a ich signifikantnosti [1] sa nachádzajú v nasledujúcej tabuľke:

[1]signifikantnosť na hladine významnosti: "***"0.001, "**"0.01, "*"0.05, "."0.1

Model	Parameter	Odhad	Signifikantnosť
SARFE	ρ	0.0106	***
	log(GDP)	0.0508	.
	log($Population$)	0.2895	*
	Inflation	0.0032	
	Unemployment_Male	0.5573	***
	Unemployment_Female	0.4411	***
SARRE	ρ	0.0101	***
	φ	0.7334	***
	const.	-0.4133	***
	log(GDP)	0.0259	**
	log($Population$)	0.0103	
	Inflation	0.0029	
	Unemployment_Male	0.5639	***
	Unemployment_Female	0.4323	***
SEMFE	λ	0.2308	***
	log(GDP)	0.0343	
	log($Population$)	0.3288	*
	Inflation	0.0016	
	Unemployment_Male	0.5617	***
	Unemployment_Female	0.4445	***
SEMRE	λ	0.2527	***
	φ	0.7190	***
	const.	-0.3300	***
	log(GDP)	0.0229	*
	log($Population$)	0.0065	
	Inflation	0.0013	
	Unemployment_Male	0.5672	***
	Unemployment_Female	0.4355	***

Jednotlivé modely sú označené skratkami, ktoré vznikli spojením skratky modelu a skratky efektu, ktorý do modelu vstupuje, napr. SARFE - priestorový autoregresný model s fixným efektom. Ako môžeme vidieť v tabuľke na rôznych modeloch, tak premenné vstupujúce do modelu sú z väčšej časti signifikantné. Výnimku tvorí premenná *Inflation*, ktorá je v každom modely nesignifikantná. Preto môžeme povedať, že inflácia nemá žiaden významný vplyv na vývoj miery nezamestnanosti. Naopak, ako sme predpokladali miera nezamestnanosti mužov a žien má významný vplyv na mieru nezamestnanosti.

2. typ Ďalší model budeme odhadovať s váhovou maticou typu Queen contiguity s priradenými preškálovanými vzdialenosťami. V tomto prípade tabuľkovú kalkuláciu neuvádzame z dôvodu podobnosti odhadou parametrov s predchádzajúcim odhadom. Avšak na našom príklade môžeme pozorovať, že zmena konštantných váh pre regióny na nekonštantné, ktoré znásobia priestorový efekt, nám dá veľmi podobné odhady.

3. typ V treťom type použijeme váhovú maticu typu Queen contiguity prerobenú tak, aby priradila váhu iba susedovi z rovnakého štátu. Pri výsledkoch môžeme očakávať stratu informácie z vynechania zahraničných susedov, uvidíme ako sa nám to zobrazí na modeloch. V tabuľkovej kalkulácii si zobrazíme iba modely SARFE a SEMFE na ktorých jediných došlo k zásadným zmenám. Ďalšie dva až na miernu zmenu odhadu koeficientov zostali rovnaké:

Model	Parameter	Odhad	Signifikantnosť
SARFE	ρ	0.0241	***
	$\log(GDP)$	0.0674	*
	$\log(Population)$	0.2000	
	Inflation	0.0032	
	Unemployment_Male	0.5517	***
	Unemployment_Female	0.4343	***
SEMFE	ρ	0.1844	***
	$\log(GDP)$	0.0452	
	$\log(Population)$	0.2395	
	Inflation	0.0007	
	Unemployment_Male	0.5642	***
	Unemployment_Female	0.4328	***

Z daného výstupu môžeme sledovať, že okrem inflácie sa stala úplne nesignifikantnou aj premenná *Population*.

4. typ Posledným modelom bude model s váhovou maticou, ktorá priradí preškálované vzdialenosti regiónom rovnakých štátov.

Model	Parameter	Odhad	Signifikantnosť
SARFE	ρ	0.0251	***
	$\log(GDP)$	-0.0156	
	$\log(Population)$	-1.0439	***
	Inflation	0.0094	*
	Unemployment_Male	0.5848	***
	Unemployment_Female	0.4060	***
SARRE	ρ	0.0198	***
	φ	1.1825	***
	const.	-0.3802	*
	$\log(GDP)$	0.0203	
	$\log(Population)$	0.0023	
	Inflation	0.0063	
	Unemployment_Male	0.5814	***
	Unemployment_Female	0.4121	***
SEMFE	λ	0.3336	***
	$\log(GDP)$	-0.1750	*
	$\log(Population)$	-0.7902	**
	Inflation	0.0022	
	Unemployment_Male	0.5966	***
	Unemployment_Female	0.4123	***
SEMRE	ρ	0.3330	***
	φ	1.2134	***
	const.	-0.2810	
	$\log(GDP)$	0.0179	
	$\log(Population)$	-0.0018	
	Inflation	0.0021	
	Unemployment_Male	0.5913	***
	Unemployment_Female	0.4169	***

V tabuľky máme v jednom prípade možnosť prvýkrát pozorovať zápornú odhadnutú hodnotu pre parameter $\log(GDP)$, ktorá nám laicky povedané hovorí, že ak hruby domáci produkt stúpne o 1%, tak nezamestnanosť klesne o 0.18%. Detto máme aj pre premennú *Population*, ktorá hovorí to isté, že ak hustota obyvateľstva stúpne o 1%, tak nezamestnanosť klesne o 0.79%. To sú odhadnuté hodnoty parametrov, ktoré by sme taktiež radi videli aj pri iných modeloch.

4.5 Zhrnutie výsledkov

V kapitole sme sa venovali odhadu parametrov štyroch rôznych modelov konštruovaných štyrmi rôznymi váhovými maticami. V odhadoch parametrov sme mohli pozorovať, že na modelovanú mieru nezamestnanosti mala vždy najväčší vplyv miera nezamestnanosti mužov a žien, pričom miera nezamestnanosti mužov mala vždy vyšší vplyv ako miera nezamestnanosti žien. Inflácia nám vyšla signifikantná len v jednom modely. Ostatné parametre, ktorými sú hustota a HDP nám vyšli rôzne z pohľadu znamienkovej konvencie, t.z. že v jednom prípade nám modelovanú nezamestnanosť ovplyvňovali kladne v inom záporne. Ďalším vhodným krokom by bolo overiť správnosť modelov a navzájom ich porovnať. Touto časťou sa však už nebudeme v našej diplomovej práci zaoberať.

Záver

V diplomovej práci sme sa zaoberali priestorovou ekonometriou. Celú prácu môžeme rozdeliť na dva okruhy. V prvej a druhej kapitole sme sa zaoberali prevažne teoretickou stránkou priestorovej ekonometrie. V tretej a štvrtej kapitole sa venujeme praktickej časti.

V prvej kapitole sme si pripomenuli model klasickej lineárnej regresie, ktorý je v ekonometrii najviac využívaný. Popísali sme si predpoklady, pre ktoré je odhad parametrov daného modelu získaný pomocou metódy najmenších štvorcov najlepší nevychýlený odhad. Pri prechode na dáta, ktoré majú v sebe zakomponovaný priestorový efekt nie sú splnené všetky predpoklady a preto nie je správne túto metódu na odhad parametrov pre takéto dáta využívať. V tejto kapitole sme si zadefinovali priestorové efekty, ktorými sú priestorová závislosť a priestorová rôznorodosť a zadefinovali sme si autoregresný proces. V autoregresnom procese nám vystupuje autoregresný parameter ρ a matica priestorových váh W, ktoré sme si popísali. Na konci kapitoly sme predstavili tri priestorové modely: priestorový autoregresný model, model s priestorovými chybami a Durbinov model, model, ktorý je získaný spojením predchádzajúcich dvoch modelov.

V druhej kapitole sme si podrobne odvodili parametre pre jednotlivé modely pomocou metódy maximálnej vierohodnosti. Odvodenie týchto parametrov nebolo triviálne, nakoľko sa musela využiť transformácia náhodnej premennej ϵ pomocou Jacobiho transformácie. Na konci tejto kapitoly sme si ukázali formulu, ktorou môžeme overiť, či dáta majú v sebe zahrnutú priestorovú autokoreláciu. Táto formula sa nazýva Moranov korelačný koeficient.

V tretej kapitole sme si podrobne popísali dáta, ktorými sú nezamestnanosť, HDP, inflácia a hustota obyvateľstva. Dáta sme brali ohľadom na jednotlivé NUTS 2 regióny. Hlavnú časť tejto kapitoly tvorí popis získanej váhovej matice W. Proces na naplnenie jednotlivý vzdialeností medzi regiónmi sa spočiatku zdal komplikovaný a časovo veľmi náročný. Nakoniec s pomocou študentky informatiky našej fakulty sa tento problém dal jednoducho vyriešiť.

V poslednej kapitole sme si v systéme `R` odhadli parametre pre rôzne modely, ktoré sa líšili rôznymi priestorovými maticami vzdialeností W. Modelovali sme nezamestnanosť vzhľadom na nezamestnanosť mužov a žien, HDP, infláciu a hustotu obyvateľstvo pre jednotlivé NUTS 2 regióny. Spozorovali sme, že najväčší vplyv na mieru nezamestnanosti mala miera nezamestnanosti mužov a miera nezamestnanosti žien. Naopak, inlácia nám okrem jedného modelu vyšla všade nesignifikantná. Ostatné odhadované parametre boli vo väčšine prípadou signifikantné.

Táto práca poskytla základný úvod do metód priestorovej ekonometrie. Medzi jej najväčší prínos by sme zaradili rôzne prístupy k tvorbe váhových matíc, v ktorých by sa ešte určite oplatilo pokračovať. Zároveň táto práca len naznačila možnosti skrývajúce sa v obore priestorová ekonometria, ktoré by sa oplatilo či už preskúmať, alebo aplikovať na reálne dáta.

Zoznam použitej literatúry

[1] Anselin, L., Florax, J. G. M.: *New Directions in Spatial Econometrics: Introduction. In: Anselin, L.; Florax, J.G.M. (eds.) New Directions in Spatial Econometrics.* Springer, Berlin, 1995

[2] Anselin, L.: *Spatial Econometrics: Methods and Models*, Springer Science + Business Media, B.V., Dordrecht, 1988

[3] Baltagi, B. H.: *Econometric analysis of panel data*, 3rd edn. Wiley, Chichester, 2005

[4] Cliff, A. D., Ord, J. K.: *Spatial Autocorrelation*, Pion, London , 1973

[5] Elhorst, J. P.: *Spatial Econometrics: From Cross - Sectional Data To Spatial Panels*, Springer,New York , 2014

[6] Kuricová, K.: *Využitie priestorovej ekonometrie na modelovanie hrubého domáceho produktu SR* [Diplomová práca], Univerzita Komenského, Fakulta matematiky, fyziky a informatiky, Bratislava, 2010

[7] LeSage, J., Pace, R. K.: *Introduction to Spatial Econometrics*, CRC Press Taylor & Francis Group, LLC,New York , 2009

[8] LeSage, J.: *Spatial Econometrics*, Department of Economics, University of Toledo, 1998

[9] Millo, G., Piras, G.: *splm: Spatial Panel Data Models in R*, Journal of Statistical Software, April 2012, Volume 47, Issue 1

[10] Niebuhr, A.: *Spatial dependence of regional unemployment in the European Union* [Odborný článok], Hamburg Institute of International Economics, Hamburg, 2002, dostupné na internete (14.5.2017: https://www.econstor.eu/bitstream/10419/19173/1/186.pdf

[11] Ord, J. K.: *Estimation Methods for Models of Spatial Interaction*, Journal of the American Statistical Association Vol. 70, No. 349 (Mar., 1975), pp. 120-126

[12] Slobodníková, S.: *Metódy priestorovej ekonometrie a modelovanie ekonomického rastu krajín V4* [Diplomová práca], Univerzita Komenského, Fakulta matematiky, fyziky a informatiky, Bratislava, 2011

[13] Whittle, P.: *On stationary processes in the plane*, Biometrika,41, 434-449, 1954

Príloha A

Kód na automatické získavanie dát

```
#!/usr/bin/python

import time
import datetime
import sys
from selenium import webdriver
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
import pandas
from random import random

START_URL = "http://www.distancefromto.net/"

def get_cities(filename):
    cities = pandas.read_csv(filename, sep="\t")
    return cities

def get_city_pairs(cities):
    res = []
    country_year = cities["Country/Year"]
    country_code = cities['Country Code']
    region = cities['Región']
    capitals = cities['Hlavné mesto']
    for i in range(len(capitals)):
        rowA = [country_year[i], country_code[i], region[i], capitals[i]]
        for j in range(i+1, len(capitals)):
            row = rowA[:]
            row.append(country_year[j])
```

```
            row.append(country_code[j])
            row.append(region[j])
            row.append(capitals[j])
            res.append(row)

    return res

def print_city_pairs_to_file(cities, filename):
    country_year = cities["Country/Year"]
    country_code = cities['Country Code']
    region = cities['Región']
    capitals = cities['Hlavné mesto']
    lat = cities['LAT']
    lon = cities['LON']
    res = []
    header = ["Country/Year A", "Country Code A", "Región A",
            "Hlavné mesto A", "LAT A", "LON A", "Country/Year B",
            "Country Code B", "Región B", "Hlavné mesto B", "LAT B", "LON B"]
    res.append(header)
    for i in range(len(capitals)):
        rowA = [country_year[i], country_code[i], region[i],
               capitals[i], lat[i], lon[i]]
        for j in range(i+1, len(capitals)):
            row = rowA[:]
            row.append(country_year[j])
            row.append(country_code[j])
            row.append(region[j])
            row.append(capitals[j])
            row.append(lat[j])
            row.append(lon[j])
            res.append(row)
```

```
    dfKm = pandas.DataFrame(res)
    dfKm.to_csv(filename, index = False, header=False)

def get_measure_button(driver, fromCity, toCity):
    WebDriverWait(driver, 10).until(EC.presence_of_
    _element_located((By.ID, 'distancefrom')))
    fromElement = driver.find_element_by_id('distancefrom')
    fromElement.clear()
    fromElement.send_keys(fromCity)

    toElement = driver.find_element_by_id('distanceto')
    toElement.clear()
    toElement.send_keys(toCity)

    buttonMeasure = driver.find_element_by_id('hae')
    return buttonMeasure

def process_pair(driver, city_pair):
    cityRegionA = city_pair[2]
    cityA = city_pair[3]
    cityRegionB = city_pair[6]
    cityB = city_pair[7]
    fromCity = cityRegionA + " " + cityA
    toCity = cityRegionB + " " + cityB

    buttonMeasure = get_measure_button(driver, fromCity, toCity)
    buttonMeasure.click()
    time.sleep(1)

    tryValues = True
```

```
    while(tryValues):
        distanceKmElement = driver.find_element_by_id('totaldistancekm')
        distanceKm = distanceKmElement.get_attribute('value')
        if (distanceKm != ""):
            distanceKm = distanceKm[:distanceKm.rfind(" ")]
            tryValues = False
            distanceKmElement.clear()
        else:
            driver.get(START_URL)
            buttonMeasure = get_measure_button(driver, fromCity, toCity)
            buttonMeasure.click()
            time.sleep(1)

    return distanceKm

def process_pairs(city_pairs):
    driver = webdriver.Firefox()
    driver.get(START_URL)

    i = 0

    for p in city_pairs:
        distance = process_pair(driver, p)
        p.append(distance)
        print(i, distance)
        i += 1

    driver.quit()
    return city_pairs

def print_results(results, filename):
```

```
    res = []
    header = [["Country/Year A", "Country Code A", "Región A",
             "Hlavné mesto A", "Country/Year B", "Country Code B",
             "Región B", "Hlavné mesto B", "Distance"]]
    res = header + results
    dfKm = pandas.DataFrame(res)
    dfKm.to_csv(filename, index = False, header=False)



if __name__ == "__main__":
    print('Number of arguments: ', len(sys.argv), 'arguments.')
    print('Argument list: ', str(sys.argv))
    infilename = sys.argv[1]
    outfilename = sys.argv[2]
    cities = get_cities(infilename);
    city = "Hlavné mesto"
    city_pairs = get_city_pairs(cities)
    print("City pairs", len(city_pairs))
    #print_city_pairs_to_file(cities, outfilename)
    results = process_pairs(city_pairs)
    print_results(results, outfilename)
```