

## Przestrzenne modelowanie zjawisk ekonomicznych w R

dr Katarzyna Kopczewska
Wydział Nauk Ekonomicznych Uniwersytet Warszawski



## Dlaczego metody przestrzenne?

### Pierwsze prawo geografii:

Waldo Tobler (1970) "Everything is related to everything else, but near things are more related than distant things."



#### Mamy przynajmniej 4 motywy:

- Dlaczego nie? ze względu na dostępność danych przestrzennych i rozwiniętego softwaru
- **Dokładność** problem obciążenia i niezgodności estymatorów MNK w przypadku istnienia zależności przestrzennych (niespełnione założenie o losowości błędów)
- Wartość dodana można testować nowe hipotezy
- Aby podążać za naturalnymi wzorcami zjawiska społeczne są najczęściej skorelowana w przestrzeni



Dalej: szczegóły przestrzennego modelowania zjawisk ekonomicznych



## Co mierzymy?

### Przestrzenne zjawiska społeczno-ekonomiczne

#### **Niewidzialne**

- Oddziaływanie polityki
- Wzorce rdzeń-peryferia i ich trwałość
- Dyfuzja za zwyczaj od rdzenia w kierunku peryferii
- Mechanizmy konwergencji i kohezji
- Renta instytucjonalna
- Efekty podziału administracyjnego
- Rola infrastruktury



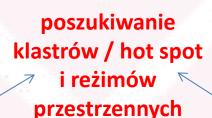
Zazwyczaj chcemy widzieć wpływ tych mechanizmów na widzialne dane/procesy ekonomiczno-społeczne



#### Widzialne

- Migracje
- Handel
- Rynek pracy
- Edukacja
- Rozwój gospodarczy
- PKB i jego wzrost

badanie zjawisk spillover / dyfuzji





Zazwyczaj chcemy widzieć obserwowany efekt i wnioskować o nieobserwowalnej polityce



### Instrumentarium

#### Co przestrzennego możemy wykorzystać w modelowaniu przestrzennym?

### → Macierz wag przestrzennych

- → definiuje strukturę sąsiedztwa, ale też zasięg oddziaływania poszukiwanych efektów przestrzennych
  - → można uwzględnić autokorelację przestrzenną błędu
  - → można uwzględnić opóźnienia przestrzenne i opóźnienia czasowo-przestrzenne zmiennych (objaśniających i objaśnianych)
  - → pozwala policzyć statystyki przestrzenne

### → Odległość pomiędzy lokalizacjami

- → można wykorzystać jako ciągłą zmienną różnicującą poziom innych zmiennych
  - → można użyć jako zmienną objaśniającą
  - → można wykorzystać do budowy macierzy wag odwrotnej odległości



## Dlaczego R?

Co wspólnego ma Poznań i Bergen ?

I co to ma wspólnego z R ???





## Dlaczego R?

Co wspólnego ma Poznań i Bergen ?

I co to ma wspólnego z R ???

### **Prof. Roger Bivand**

Autor środowiska i pakietów przestrzennych w R



Maintainer TaskViews: **Spatial** (Analysis of Spatial Data)

- Reading and writing spatial data
- Point pattern analysis
- Geostatistics
- Disease mapping and areal data analysis
- Spatial regression
- Ecological analysis



## Pakiety przestrzenne w R

**sp:** Classes and methods for spatial data

**spdep:** Spatial dependence: weighting schemes, statistics and models

**splancs:** Spatial and Space-Time Point Pattern Analysis

**classInt:** Choose univariate class intervals

maptools: Tools for reading and handling spatial objects

**splancs:** Spatial and Space-Time Point Pattern Analysis

rgdal: Bindings for the Geospatial Data Abstraction Library

rgeos: Interface to Geometry Engine - Open Source (GEOS)

**spgwr:** Geographically weighted regression

**RColorBrewer:** ColorBrewer palettes

**spatstat:** Spatial Point Pattern analysis, model-fitting, simulation, tests

**sphet:** Estimation of spatial autoregressive models with and without

heteroskedastic innovations

**splm:** Econometric Models for Spatial Panel Data

**spatialprobit:** Spatial Probit Models

**SmarterPoland:** A set of tools developed by the Foundation SmarterPoland.pl



## Pakiety przestrzenne w R w core

**DCluster:** Functions for the detection of spatial clusters of diseases

**deldir:** Delaunay Triangulation and Dirichlet (Voronoi) Tessellation

**geoR:** Analysis of geostatistical data

**gstat:** spatial and spatio-temporal geostatistical modelling, prediction

and simulation

RandomFields: Simulation and Analysis of Random Fields

raster: Geographic data analysis and modeling

**spacetime:** classes and methods for spatio-temporal data

I wiele wiele innych



## Jak zacząć?

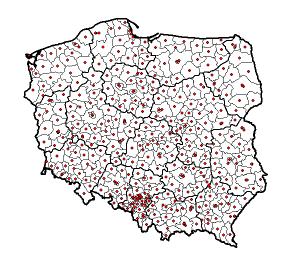
- \* Ekonomiści korzystają głównie z danych obszarowych, wg podziału terytorialnego
- → Mapa shapefile np. z DIVA GIS (open source)
- → Dane w porządku obszarów z mapy, dane z GUS, Eurostat, ECB, IMF, EVS, Diagnoza społeczna, Komenda Główna Policji etc.
- → R pakiety sp, spdep, classInt, maptools, rgeos, RColorBrewer,...
- → Pomysł na badanie: wkład własny

Czyli jak prowadzić badania za 0 zł ©



## Kod R: Starter (1)

```
library(spdep)
library (maptools)
library(sp)
# wczytanie plików shapefile dla powiatów i województw
woj<-readShapePoly("POL adm1.shp",
       proj4string=CRS("+proj=longlat+ellps=clrk80"))
pow<-readShapePoly("POL adm2.shp",</pre>
       proj4string=CRS("+proj=longlat+ellps=cl
# środki powiatów
crds<-coordinates (pow)
# mapa administracyjna
plot(pow)
plot(woj, add=TRUE, lwd=2)
points(crds, pch=21, bg="red", cex=0.8)
```





## Macierz wag przestrzennych

### Sąsiedztwo

$$\tilde{W} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

### Wagi przestrzenne W

$$W = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

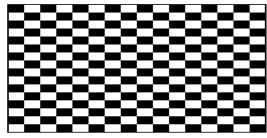
```
# macierz wag przestrzennych W dla powiatów (contiguity matrix)
cont.nb<-poly2nb(as(pow, "SpatialPolygons"))
cont.listw<-nb2listw(cont.nb, style="W", zero.policy=TRUE)</pre>
```

```
# wczytywanie danych z *.csv skonsolidowane i uporządkowane
dane<-read.csv("dane powiaty.csv ", sep=";", dec=",", header=TRUE)
summary(dane)</pre>
```



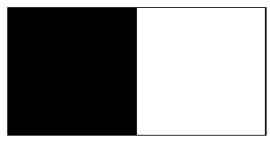
## Autokorelacja przestrzenna

(1)



Negative (extreme)

No correlation



Positive (extreme)

Ujemna autokorelacja przestrzenna – gdy wartości podobne sąsiadują rzadziej niż wg rozkładu losowego

**Brak autokorelacji** – gdy brak jednoznacznego wzorca

Dodatnia autokorelacja przestrzenna - wartości podobne sąsiadują częściej niż wg rozkładu losowego

```
dane12<-dane[dane$rok==2012, ]
wynik<-moran.test(dane12$XA06, cont.listw)
wynik</pre>
```

attributes(wynik)

Moran's I test under randomisation

data: dane12\$XA06
weights: cont.listw

Moran I statistic standard deviate = 13.6153, p-value < 2.2e-16 alternative hypothesis: greater

sample estimates:

Moran I statistic 0.462766094

Expectation -0.002666667

Variance 0.001168579



## Autokorelacja przestrzenna (2)

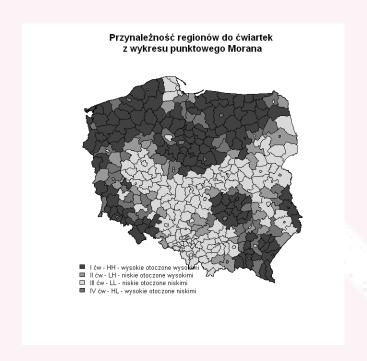
```
# Moran scatterplot
```

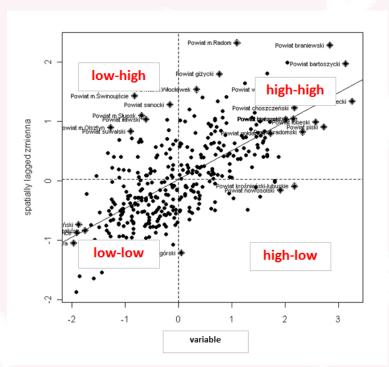
x<-dane12\$XA06

zx<-as.data.frame(scale(x))</pre>

moran.plot(zx\$V1, cont.listw, pch=19,

labels=as.character(dane12\$powiat))







## Badana przykładowa zależność

#### **Y** ~ **X**

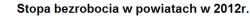
średnie wynagrodzenie ~ stopa bezrobocia

#### **Zmienne kontrolne:**

- udział zatrudnienia w usługach
- odległość powiatu od rdzenia (relatywna lokalizacja)

#### Ważne problemy w modelowaniu:

- Stałość strukturalna w czasie i przestrzeni danej zmiennej
- Zróżnicowanie / bieżność w przestrzeni danej zmiennej
- Zbieżność rozkładów przestrzennych (charakterystyk przestrzennych) badanych zmiennych
  - → jeśli te problemy istnieją, to konieczny taki dobór formy modelu, aby je zniwelować





przedziały<-8

## Rozkład przestrzenny

```
# mapa stopy bezrobocia wg powiatów
zmienna<-dane$XA06[dane$rok==2012]
summary(zmienna)</pre>
```



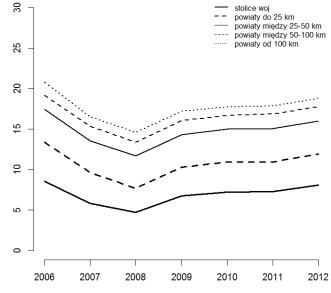
sub1<-dane[dane\$odległość==0,]
sub2<-dane[dane\$odległość<25,]</pre>

## Wykres panelowy

sub3<-dane[dane\$odległość<50 & dane\$odległość>25, ]

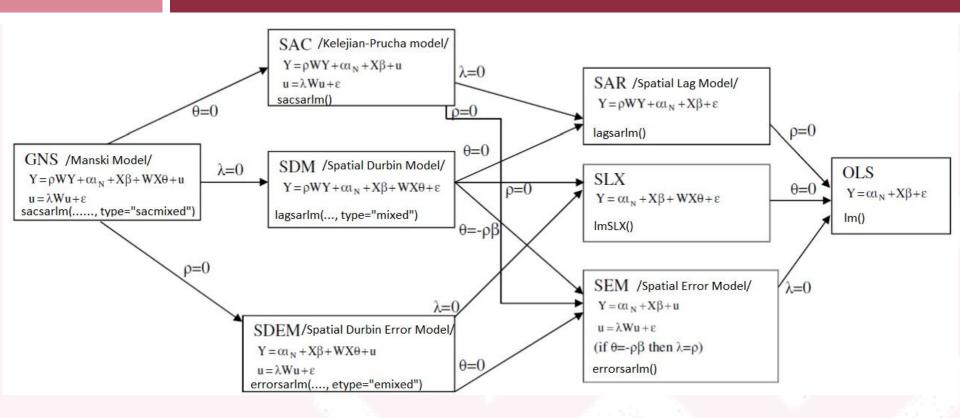
```
sub4<-dane[dane$odległość<100 & dane$odległość>50, ]
sub5<-dane[dane$odległość>100, ]
msub1<-aggregate(sub1$XA06, by=list(sub1$rok), mean,
msub2<-aggregate(sub2$XA06, by=list(sub2$rok), mean,
msub3<-aggregate(sub3$XA06, by=list(sub3$rok), mean,
msub4<-aggregate(sub4$XA06, by=list(sub4$rok), mean, na.rm=TRUE)
msub5<-aggregate(sub5$XA06, by=list(sub5$rok), mean, na.rm=TRUE)</pre>
plot(msub1, type="n", ylim=c(0, 30), axes=FALSE, xlab=" ", ylab=" ")
axis(1); axis(2)
lines (msub1, lwd=3)
lines (msub2, lwd=3, lty=2)
lines (msub3, lty=1, lwd=2)
lines (msub4, lty=2, lwd=2)
lines (msub5, lty=3, lwd=2)
title (main="Rejestrowana stopa bezrobocia
        wg odległości powiatu od miasta wojewódzkiego")
legend ("topright", legend=c ("stolice woj", "powiaty do 25 km", "powiaty między 25-
        50 km", "powiaty między 50-100 km", "powiaty od 100 km"),
        lty=c(1,2,1,2,3), lwd=c(2,2,1,1,1), bty="n", cex=0.8)
```

#### Rejestrowana stopa bezrobocia wg odległości powiatu od miasta wojewódzkiego





## Modele przestrzenne jednookresowe



Elhorst's Flow Chart (Vega & Elhorst, 2013) uzupełniony o kody R



### Zmienne do modelu

```
# model
       przeciętne wynagrodzenie Polska=100% (XA31)
# x1
       stopa bezrobocia
                                                     (XA06)
# x2 udział osób zatrudnionych w usługach
                                                     (XA05/XA20)
# x3
       odległość od miasta woj.
                                                     (odległość)
# estymacja modelu
dane$y<-dane$XA31
dane$x1<-dane$XA06
dane$x2<-dane$XA05/dane$XA20
dane$x3<-dane$odległość
dane12<-dane[dane$rok==2012, ] # dla roku 2012
równanie<-y~x1+x2+x3
```



## Model uogólniony (GNS)

model<-sacsarlm(równanie, data=dane12, cont.listw, type="sacmixed")
summary(model)</pre>

```
Type: sacmixed
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value
                                       Pr(>|z|)
(Intercept) 29.996984 10.034294 2.9894
                                        0.002795
           -0.323845 0.127141 -2.5471
                                        0.010861
\times 1
x2
           50.211856
                      7.146706 7.0259 2.127e-12
                      0.033888 -1.1990 0.230520
         -0.040632
x3
lag.x1 0.311368 0.165833 1.8776 0.060437
                                       0.316751
           17.208147 17.188251 1.0012
laq.x2
            0.033665 0.046503 0.7239
                                       0.469104
laq.x3
```

```
GNS /Manski Model/

Y = \rho WY + \alpha \iota_N + X\beta + WX\theta + u

u = \lambda Wu + \epsilon

sacsarlm(....., type="sacmixed")
```

```
Rho: 0.52892 Asymptotic standard error: 0.131 z-value: 4.033, p-value: 5.5059e-05
Lambda: -0.415 Asymptotic standard error: 0.184 z-value: -2.249, p-value: 0.024515
LR test value: 27.126, p-value: 5.39e-05
Log likelihood: -1393.363 for sacmixed model
ML residual variance (sigma squared): 88.505, (sigma: 9.4077)
Number of observations: 376
Number of parameters estimated: 10
AIC: 2806.7, (AIC for lm: 2823.9)
```



Type: sac

### Model SAC

```
model<-sacsarlm(równanie, data=dane12, cont.listw)
summary(model)</pre>
```

```
Coefficients: (asymptotic standard errors)

Estimate Std. Error z value Pr(>|z|)

(Intercept) 32.854871 7.973674 4.1204 3.782e-05

x1 -0.123572 0.080960 -1.5263 0.1269

x2 55.975653 6.594515 8.4882 < 2.2e-16

x3 -0.015493 0.013934 -1.1119 0.2662
```

```
SAC /Kelejian-Prucha model/

Y = \rho WY + \alpha \iota_N + X\beta + \iota

\iota = \lambda W \iota_N + \epsilon

\iota = \lambda W \iota_N + \epsilon

\iota = \lambda W \iota_N + \epsilon
```

```
Rho: 0.5429 Asymptotic standard error: 0.0847 z-value: 6.4058, p-value: 1.496e-10 Lambda: -0.433 Asymptotic standard error: 0.134 z-value: -3.224, p-value: 0.001262 LR test value: 22.036, p-value: 1.6403e-05 Log likelihood: -1395.908 for sac model ML residual variance (sigma squared): 89.149, (sigma: 9.4418) Number of observations: 376
```

```
Number of parameters estimated: 7 AIC: 2805.8, (AIC for lm: 2823.9)
```



## Przestrzenny model panelowy

### library(splm) dane\$order<-rep(1:376, times=7)</pre> dane<-dane[,c(36,16,1:35)] model.spml<-spml(formula=równanie, data=dane, listw=cont.listw, model="pooling", effect="twoways", lag=TRUE, spatial.error="b") summary(model.spml) Error variance parameters: Estimate Std. Error t-value Pr(>|t|) Spatial autoregressive coefficient: Estimate Std. Error t-value Pr(>|t|) Coefficients: Estimate Std. Error t-value Pr(>|t|) (Intercept) 30.2560632 0.7324545 41.3078 < 2.2e-16 \*\*\* -0.1227367 0.0270357 -4.5398 5.631e-06 \*\*\* x1 x2 63.3827536 2.5849719 24.5197 < 2.2e-16 \*\*\* x3 -0.0180097 0.0049756 -3.6196 0.000295 \*\*\* Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '' 1



## Direct and indirect impacts

```
imp <- impacts (model, tr=trMat, R=2000)</pre>
summary(imp, zstats=TRUE, short=TRUE)
Impact measures (sac, trace):
       Direct
                 Indirect
                                Total
x1 -0.13261940 -0.13773181 -0.2703512
x2 60.07413659 62.38996377 122.4641004
x3 -0.01662756 -0.01726854 -0.0338961
Simulation results (asymptotic variance matrix):
Simulated z-values:
     Direct Indirect
                           Total
x1 -1.571985 -1.3712662 -1.526920
x2 8.569725 2.5810763 4.306607
x3 -1.128902 -0.9835448 -1.082085
```

direct impact: wpływ zmiany
x(i) na zmianę y(i)

indirect impact: wpływ
zmiany x(i) na zmianę y(j)

total impact: łączny wpływ direct + indirect impact



Badanie efektu spillover

#### Simulated p-values:

Direct Indirect Total x1 0.11595 0.1702920 0.12678 x2 < 2e-16 0.0098493 1.6578e-05 x3 0.25894 0.3253394 0.27921



## No właśnie, jak modelować spillover...

### Jest kilka podejść teoretycznych:

- 1) Podstawowe podejście: uwzględnienie opóźnień przestrzennych zmiennej objaśnianej lub błędu w modelu (spatial dependence models). Spillover istnieje, gdy współczynniki przestrzenne są istotne (e.g. Schaltegger & Zemp, 2003; Lundberg, 2001).
- 2) Podobnie jak powyżej, ale opóźnienia przestrzenne dotyczą zmiennych objaśniających (ΘWX) (model Durbina) (Yuandong et al., 2013)
- 3) Przez uwzględnienie innej macierzy wag przestrzennych, np. zależnej od lokalizacji, dochodu per capita czy czynników demograficznych (kompozycja etniczna) (Case & Rosen, 1993)
- 4) Przez uwzględnienie macierzy wag przestrzennych z sąsiadami w promieniu d km, z różnymi punktami odcięcia, co umożliwia ocenę zmian wyników w zależności od kryterium sąsiedztwa (Schaltegger & Zemp, 2003)
- 5) Anselin (2003) proponuje rozróżnienie pomiędzy lokalnymi i globalnymi efektami, przy czym lokalny spillover pojawia się pomiędzy połączonymi lokalizacjami, zaś globalny spillover pomiędzy niepołączonymi lokalizacjami



## No właśnie, jak modelować spillover...c.d.

- 6) LeSage & Pace (2009) proponują wydzielenie efektów bezpośrednich i spillover. Efekty bezpośrednie pojawiają się gdy efekt zmiany zmiennej objaśniającej pojawia się tylko jako oddziaływanie na zmianę zmiennej objaśnianej w tej samej lokalizacji. W przypadku spillover ten wpływ jest widziany w innej, sąsiedzkiej lokalizacji. W sensie operacyjnym, efekty bezpośrednie to elementy diagonalne macierzy pochodnych cząstkowych, zaś efekty spillover są spoza przekątnej (Claeys et al., 2012; Elhorst, 2010).
- 7) LeSage and Pace (2011) proponują jeszcze inny podział, gdzie globalne spillover mają pętlę zwrotną (feedback) tj. zjawisko wraca do regionu emitującego, zaś lokalne spillover nie mają takiej pętli zwrotnej.
- 8) Capello (2007) proponuje spillover jako miarę (indeks) wykorzystywaną w modelach przestrzennej zależności. Dla regionalnego wzrostu byłoby to średnie PKB ważone wagami przestrzennymi zależnymi od odległości pomiędzy regionami.
- 9) Reback (2011) proponuje międzyregionalny model lokalnych spillover, w którym w regionach granicznych obserwuje się spillover. Oddzielenie naturalnych trendów (*local common shocks*) od efektów przestrzennych (spatial spillover) jest możliwe przez model zmiennych instrumentalnych. Prognozy z regionów wewnętrznych są przenoszone do regionów granicznych, a różnica pomiędzy prognozą a wartością obserwowaną jest modelowana jako spillover.



# Przykład: przestrzenny zasięg samorządów regionalnych (1)

## Czy przestrzeń ma znaczenie dla samorządów lokalnych w implementacji regionalnej polityki rozwojowej?

- → Odległość geograficzna jest czynnikiem osłabiającym transmisję polityki rozwojowej
- → Samorządy położone peryferyjnie są słabsze niż te centralne
- → Jak pokazują wyniki, efektywny zasięg centrów regionalnych nie przekracza 25 km (tylko najbliższe sąsiedzko gminy). Pozostałe samorządy można traktować jako peryferyjne.
  - → Badanie na danych gminnych NTS5 z lat 1995-2007

Kopczewska K., 2013, "The spatial range of local governments: does geographical distance affect governance and public service?", The Annals of Regional Science, Volume 51, Issue 3 (2013), Page 793-810, IF=0,901, DOI: 10.1007/s00168-013-0567-z

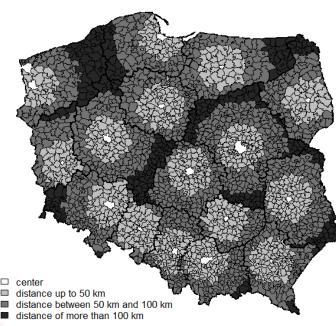


# Przykład: przestrzer zasięg samorządów regionalnych (1)

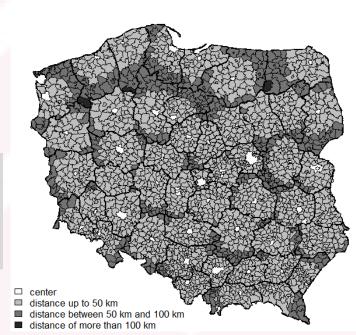
### Odległość między gminą / powiatem a miastem woj.

```
[,1] [,2] [,3] [,4] [,5]
[1,] 0.00000 95.10024 112.82814 155.22502 125.77449
[2,] 95.10024 0.00000 68.46546 78.36275 62.77604
[3,] 112.82814 68.46546 0.00000 52.39279 21.44033
[4,] 155.22502 78.36275 52.39279 0.00000 31.99598
[5,] 125.77449 62.77604 21.44033 31.99598 0.00000
```

#### after the 1999 reform



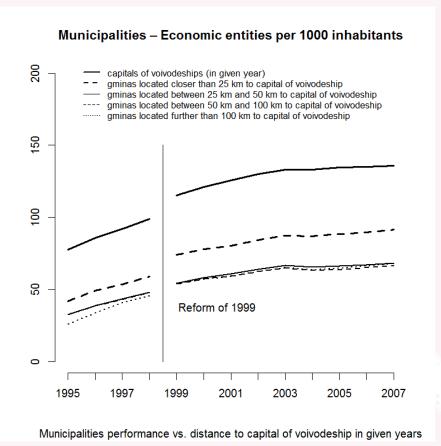
prior to the 1999 reform

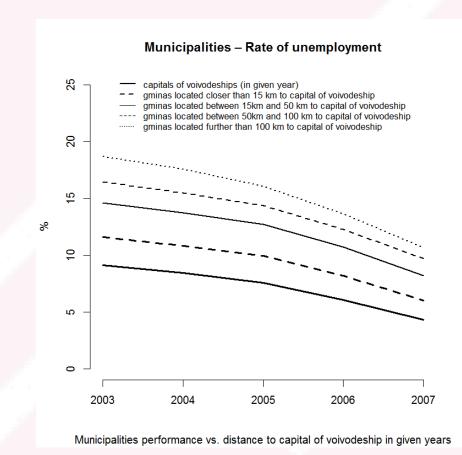




# Przykład: przestrzenny zasięg samorządów regionalnych (3)

Wyniki społeczno-gospodarcze gmin ze względu na odległość do miasta wojewódzkiego







# Przykład: przestrzenny zasięg samorządów regionalnych (4)

$$xx1 = \beta_0 + \beta_1 \cdot xx0 + \beta_2 \cdot city + \gamma_1 \cdot DIST + \gamma_2 \cdot DIST^2 + \gamma_3 \cdot DIST^3 + \gamma_4 \cdot DIST^4 + e$$

$$e = \lambda We + u$$

#### Zmienna objaśniana:

xx1 – relatywne wydatki lub dochody *per capita* w cenach stałych jako Polska = 100 w ostatnim okresie (2007)

### Zmienne objaśniające::

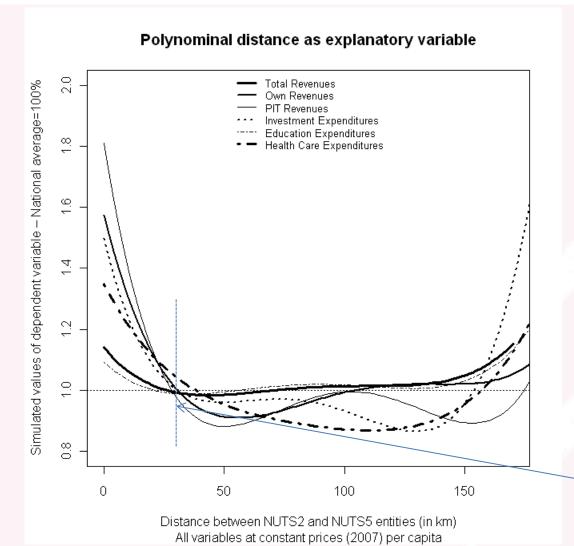
xx0 – jak xx1 ale w pierwszym okresie (1999)

DIST<sup>1</sup>, DIST<sup>2</sup>, DIST<sup>3</sup> and DIST<sup>4</sup>- wielomian odległości Euklidesowej pomiędzy NTS2 i NTS5 CITY – zmienna zero-jedynkowa, =1 dla siedzib władz powiatowych

SRMSE



# Przykład: przestrzenny zasięg samorządów regionalnych (5)



- Dochody ogółem
- Dochody własne
- Dochody z PIT
- Wydatki inwestycyjne
- Wydatki na edukację
- Wydatki na ochronę zdrowia

Model interakcji przestrzennych (spatial interactions model)

Zasięg miast rdzeniowych dyfuzja



### Krótka reklama naukowa



www.ersa.org.pl

## Międzynarodowe stowarzyszenie naukowe regionalistów

- Urbanistyka
- Geografia
- Ekonomia i ekonometria
- Planowanie przestrzenne
- Gospodarka przestrzenna
- Socjologia
- Polityka gospodarcza

- ......

Spotkania i badania naukowe dotyczące regionu, przestrzeni, modelowania, polityki w przestrzeni, specjalizacji etc.



Dziękuję za uwagę!

kkopczewska@wne.uw.edu.pl