Modelowanie dynamiki rozkładu w R. Zastosowanie do analizy konwergencji na poziomie lokalnym

Piotr Wójcik
pwojcik@wne.uw.edu.pl

Konferencja **WhyR ?**
Warszawa, 27-29 września 2017

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Cel prezentacji

- pokazać różne podejścia do **modelowania dynamiki rozkładu** dostępne w R,
- szczególne skupienie na **macierzach przejścia** i **warunkowych estymatorach gęstości**,
- **zastosowanie w R** względnie nowej metodologii podsumowania dwuwymiarowej warunkowej funkcji gęstości za pomocą (jednowymiarowego) **rozkładu ergodycznego** (Gerolimetto and Magrini 2017),
- zaprezentować **czytelny i atrakcyjny sposób wizualizacji** wyników estymacji,
- **przykłady praktyczne** na rzeczywistych i symulowanych danych przestrzennych.
- wykorzystane **pakiety R**: `markovchain`, `reshape2`, `ggplot2`, `gridExtra`,
- własny pakiet `ModellingConvergence` do modelowania regionalnej konwergencji i dynamiki rozkładu – **w budowie**.

Badania realizowane w ramach grantu Narodowego Centrum Nauki nr 2016/21/B/HS4/00670.

Pojęcie konwergencji

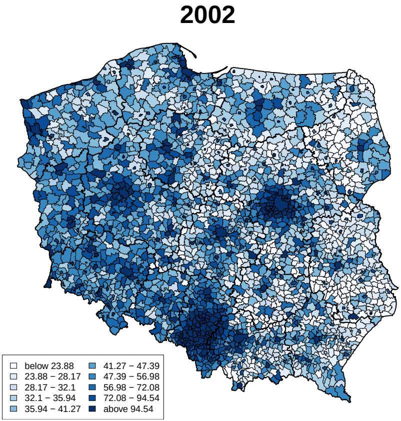
- w naukach ekonomicznych przez **konwergencję** (zbieżność) rozumiane jest upodabnianie się do siebie krajów lub regionów pod względem analizowanego zjawiska,
- oznacza to zmniejszanie się dysproporcji między regionami w czasie,
- zjawisko przeciwne nazywane jest **dywergencją**,
- najczęściej analizowana jest zbieżność pod względem dochodu na 1 mieszkańca lub na 1 zatrudnionego.

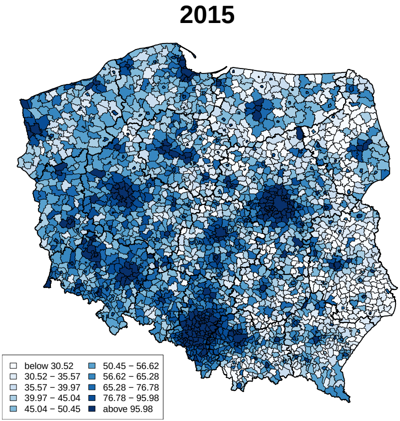
Dane rzeczywiste i symulowane (scenariusze)

- **dane rzeczywiste** – dochody gmin *per capita* z udziału w podatku dochodowym (PIT) w okresie 2002–2015 (Źródło: BDL GUS),
- **symulowana konwergencja** – punkt wyjścia: rzeczywiste dane dla 2002, w roku 2015 zmniejszenie dystansu do średniej krajowej o 25%,
- **symulowana dywergencja** – punkt wyjścia: rzeczywiste dane dla 2002, w roku 2015 zwiększenie dystansu do średniej krajowej o 10%,
- **symulowana stabilność** – punkt wyjścia: rzeczywiste dane dla 2002, w roku 2015 losowa zmiana w zakresie +/-5 p.p. (rozkład jednostajny),

Analiza zjawisk społeczno-gospodarczych

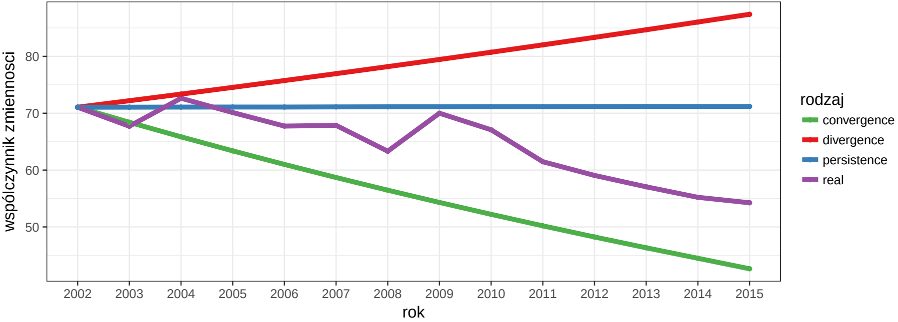
Analizując zjawiska społeczno-ekonomiczne często chcemy zobaczyć ich **przestrzenne zróżnicowanie** i **jego zmiany w czasie**.





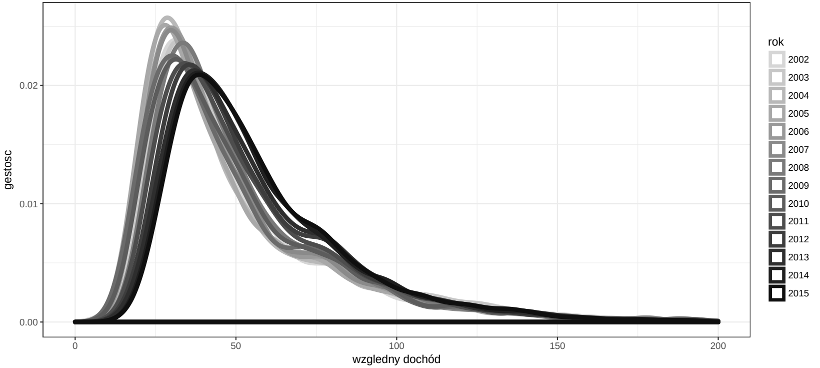
Pojedyncza miara rozkładu

Najprostsze podejście wykorzystuje **pojedynczą miarę zróżnicowania** (np. odchylenie standardowe, współczynnik zmienności), które jednak nie mówi nic o zróznicowaniu **wewnątrz rozkładu**

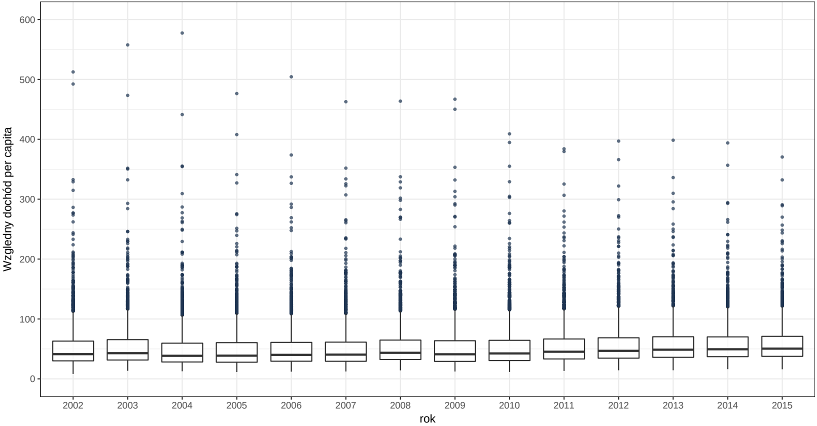


Histogramy i jednowymiarowe estymatory jądrowe

W tym celu można porównać **histogramy** albo **jednowymiarowe estymatory jądrowe**, które jednak wciąż nie informują o **mobilności w ramch rozkładu**



Boxplot


Wzgledny dochód per capita
600
500
400
300
200
100
0
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
rok

Macierz przejścia

- **wewnętrzna mobilność w czasie** może być uchwycona przez zastosowanie **macierzy przejścia** (zapożyczone z procesów Markowa – patrz Quah, 1996):
- początkowy rozkład dzielony na kilka przedziałów (grup) – zwykle równolicznych (grupy kwantylowe),
- macierz M:

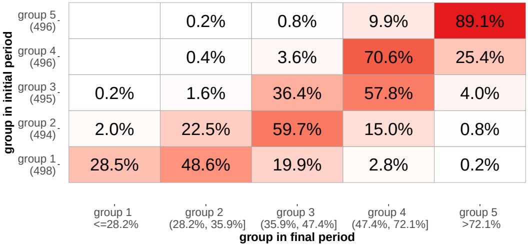
$$d_t = M \times d_{t-1}$$

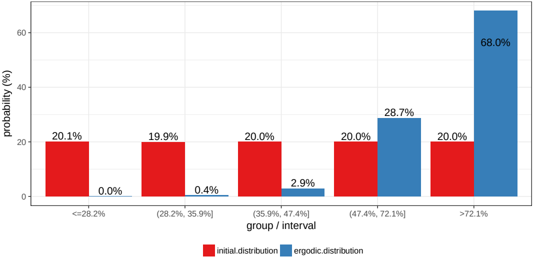
  pokazuje prawdopodobieństwa **mobilności** między grupami,
- macierz podniesiona do dużej potęgi zbiega do **wektora ergodycznego**, który pokazuje długookresowy podział na grupy.

Macierz przejścia – funkcje

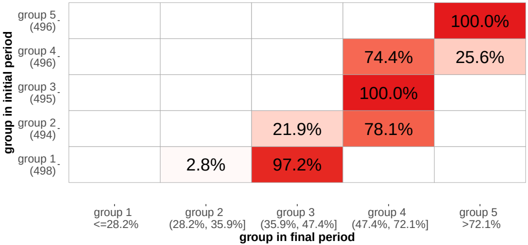
```r
library(ModellingConvergence)

macierz_income <- calculate_trans_matrix(input.data = dane.income,
                                         start = "y_1",
                                         end = "y",
                                         ngroups = 5)

plot_trans_matrix(matrix = macierz_income,
                  gradient.high = "darkred",
                  show.zeros = F)

plot_trans_matrix_erg(matrix = macierz_income)
```

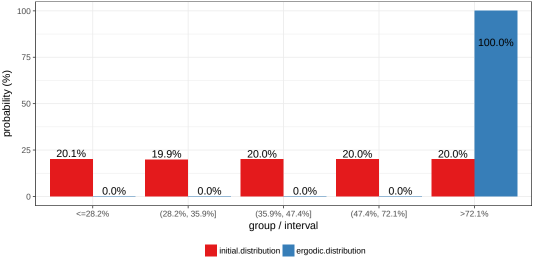
Macierz przejścia – wynik dla rzeczywistych danych



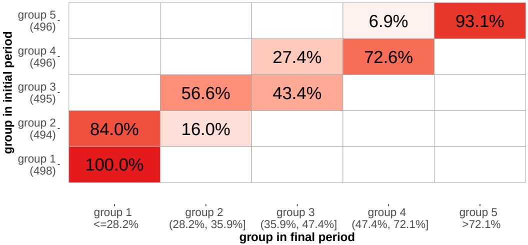


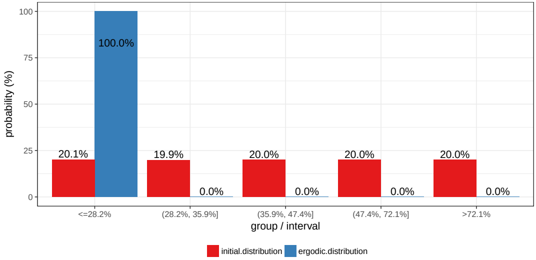
Macierz przejścia – wynik dla symulowanej konwergencji



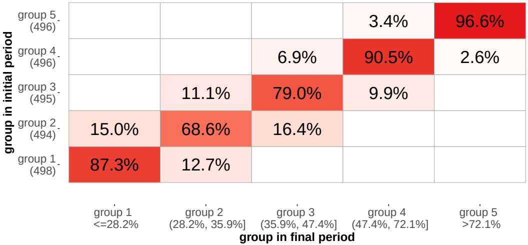


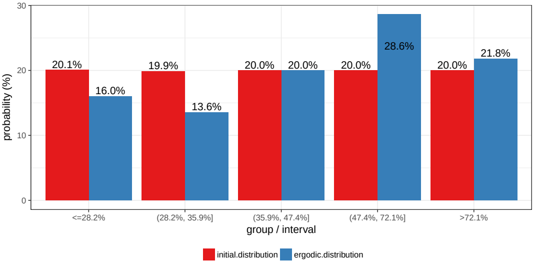
Macierz przejścia – wynik dla symulowanej dywergencji





Macierz przejścia – wynik dla symulowanej stabilności





Warunkowy estymator gęstości

- ciągła alternatywa macierzy przejścia; obliczany zgodnie z formułą:

$$f^A(y_T|y_0) = \frac{f^A(y_T, y_0)}{f^A_{y_0}(y_0)}$$

- mianownik zastępowany jest wyrażeniem:

$$\hat{f}^A_{y_0}(y_0) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_{y_0} w_i} K\left(\frac{y_0 - y_{0i}}{h_{y_0} w_i}\right)$$

- licznik zastąpiony przez:

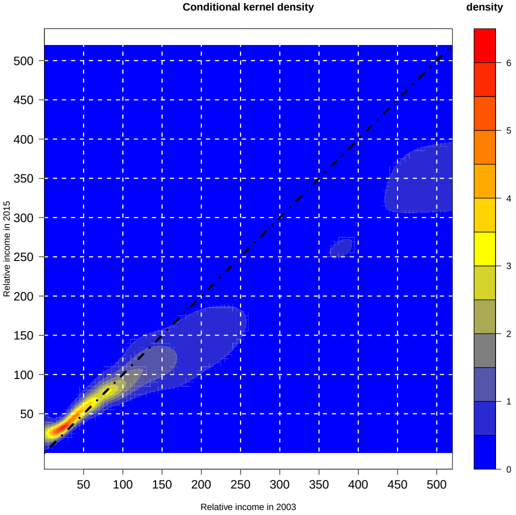
$$\hat{f}^A(y_T, y_0) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_{y_T} h_{y_0} w_i} K\left(\frac{y_T - y_{Ti}}{h_{y_T} w_i}\right) K\left(\frac{y_0 - y_{0i}}{h_{y_0} w_i}\right)$$

- h_{y_0} i h_{y_T} – optymalne pasma estymacji dla rozkładu początkowego i końcowego w_i – wagi obserwacji w metodzie adaptacyjnej.
- rozkład ergodyczny można policzyć przez dyskretyzację warunkowej funkcji gęstości – patrz Gerolimetto and Magrini (2017).

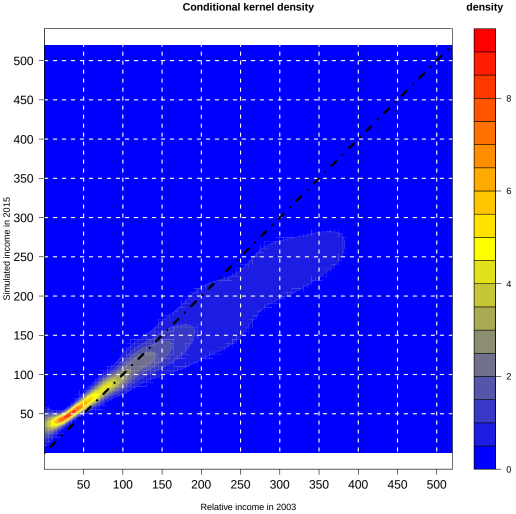
Estymator gęstości – funkcje

```
kernel_income <- calculate_cond_kde(input.data = dane.income,
                                    cond = "y_1", var = "y",
                                    grid_by = 0.05)

kernel_incomeA <-
  calculate_cond_kde_adaptive(input.data = dane.income,
                              cond = "y_1", var = "y",
                              grid_by = 0.05)

plot_kernel(kernel.data = kernel_incomeA,
            gmin = 0, gmax = 550, gby = 50,
            xlab = "Relative income in 2003",
            ylab = "Relative income in 2015",
            BW = F, nlevels = 10)
```

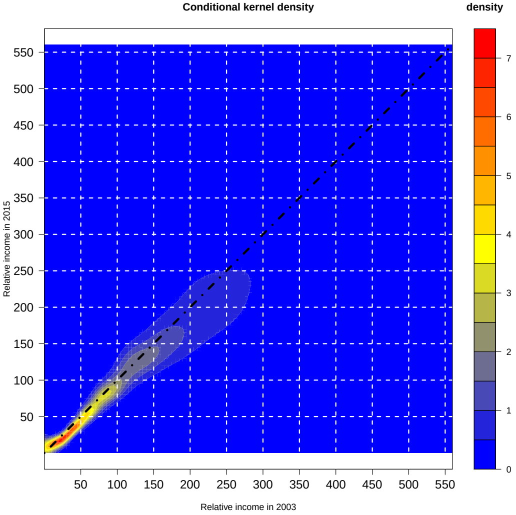
Estymator gęstości – wynik dla rzeczywistych danych



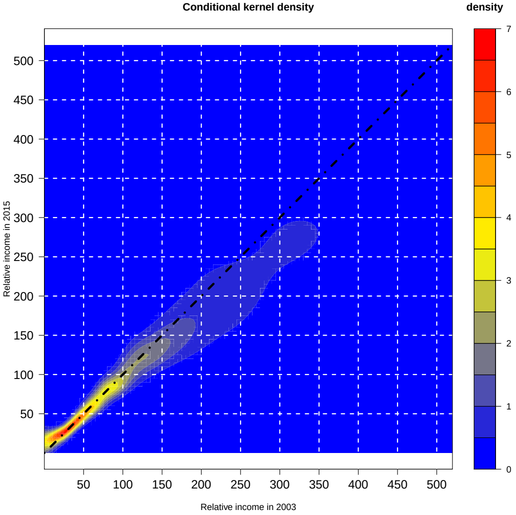
Estymator gęstości – wynik dla symulowanej konwergencji



Estymator gęstości – wynik dla symulowanej dywergencji

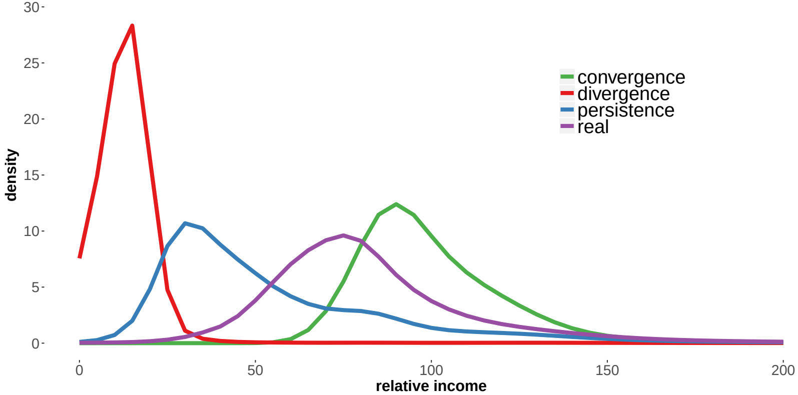


Estymator gęstości – wynik dla symulowanej stabilności



Estymator gęstości – rozkłady ergodyczne

```r
ergodic.real <- ergodicKDE(kernel.data = kernel_incomeA)
```



Wnioski

- prtzedstawione metody pozwalają modelować zmiany w rozkładzie badanego zjawiska (mobilność wewnątrz rozkładu),
- rozkład dochodu na poziomie lokalnym jest najbardziej stabilny w najbogatszych gminach,
- prawdopodobieństwo względnego bogacenia jest wyższe niż ubożenia,
- estymatory jądrowe wskazują na upodabnianie się do siebie podgrup gmin (osobno bogatych, przeciętnych i biednych) – tzw. konwergencję klubów,
- wzorce ergodyczne dla rzeczywistych danych sytuują się między scenariuszem konwergencji i stabilności rozkładu.

Bibliografia

Gerolimetto, Margherita, and Stefano Magrini. 2017. "A Novel Look at Long-Run Convergence Dynamics in the United States." *International Regional Science Review* 40 (3).

Magrini, Stefano. 2009. "Why Should We Analyse Convergence Through the Distribution Dynamics Approach?" *Science Regionali* 8: 5–34.

Quah, Danny. 1996. "Twin Peaks: Growth and Convergence in Models Distribution Dynamics." *Economic Journal* 106: 1045–55.

Viegas, M., and M. Antunes. 2013. "Convergence at a Local Level: An Exploratory Spatial Analysis Applied to the Portuguese Municipalities." *Revista Portuguesa de Estudos Regionais* 34.

Zambom, A.Z., and R. Dias. 2012. "A Review of Kernel Density Estimation with Applications to Econometrics." Discussion Paper arXiv:1212.2812.