MODELLING DISTRIBUTION DYNAMICS IN R

APPLICATION TO CONVERGENCE ANALYSIS ON A LOCAL LEVEL

PIOTR WÓJCIK UNIVERSITY OF WARSAW

INTRODUCTION

- when modelling socio-economic data one is often interested in **cross-sectional diversity** and **its changes over time**,
- the **simplest approach** uses a **single dispersion measure**, but it tells nothing about the diversity **within the distribution**,
- one can also compare **histograms** or **unidimensional kernel density estimates**, which still tells nothing about the **mobility within the distribution**
- **internal mobility** can be caught by a **transition matrix** or **conditional kernel density estimate** (based on which can also formulate **ergodic distribution**).

OBJECTIVES

- show different approaches of **modelling distribution dynamics** applied in R,
- particular focus on **transition matrices** and **conditional kernel density estimates**,
- **R based application** of recently developed methodology allowing to summarize a two-dimensional conditional kernel density surface with the (univariate) **ergodic distribution** – see Gerolimetto and Magrini (2017),
- present **readable and attractive ways of visualization** of estimation results,
- **practical examples** on simulated and real spatial data.

FURTHER STEPS

- development of own package for modelling regional convergence and within distribution mobility,
- using Rcpp for time efficient calculation of conditional kernel densities on larger samples,

REFERENCES

Gerolimetto, M. and Magrini, S. (2017), "A novel look at long-run convergence dynamics in the united states", *International Regional Science Review* , Vol. 40.

Magrini, S. (2009), "Why should we analyse convergence through the distribution dynamics approach?", *Science Regionali* , Vol. 8, pp. 5–34.

Quah, D. (1996), "Twin peaks: Growth and convergence in models distribution dynamics", *Economic Journal* , Vol. 106, pp. 1045–1055.

Viegas, M. and Antunes, M. (2013), "Convergence at a local level: an exploratory spatial analysis applied to the portuguese municipalities", *Revista Portuguesa de Estudos Regionais* , Vol. 34.

Zambom, A. and Dias, R. (2012), A review of kernel density estimation with applications to econometrics, Discussion Paper arXiv:1212.2812.

CONTACT INFORMATION

web wne.uw.edu.pl/pwojcik
email pwojcik@wne.uw.edu.pl

METHODS & R PACKAGES

Transition matrix (borrowed from Markov chains – see Quah, 1996):

- initial distribution is divided into several intervals (groups),
- matrix M: $d_t = M \times d_{t-1}$ shows probabilities of **mobility** between groups,

Conditional kernel density (A = adaptive): $f^A(y_T|y_0) = \frac{f^A(y_T,y_0)}{f^A_{y_0}(y_0)}$, where

- denominator estimated as: $\hat{f}^A_{y_0}(y_0) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_{y_0} w_i} K\left(\frac{y_0 - y_{0i}}{h_{y_0} w_i}\right)$
- numerator replaced with: $\hat{f}^A(y_T, y_0) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_{y_T} h_{y_0} w_i} K\left(\frac{y_T - y_{Ti}}{h_{y_T} w_i}\right) K\left(\frac{y_0 - y_{0i}}{h_{y_0} w_i}\right)$

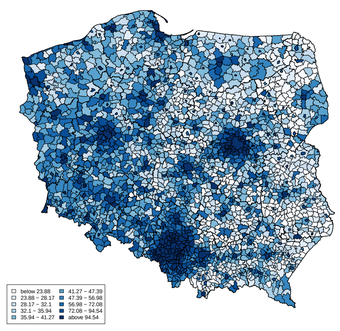
where h_{y_0} and h_{y_T} are optimal bandwidths for initial and final distribution respectively and w_i are observation weights from the two step adaptive estimation procedure.
Calculation of ergodic density based on discretization of conditional kernel density – see Gerolimetto and Magrini (2017).

R packages used in the analysis: markovchain, reshape2, ggplot2, gridExtra

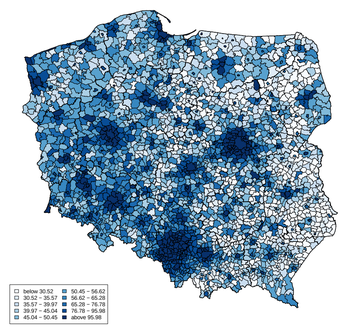
Development of own package for modelling regional convergence and within distribution mobility – in progress.

RESULTS

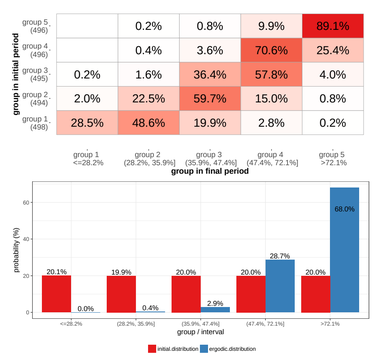
REAL DATA ON RELATIVE INCOME DISTRIBUTION FOR POLISH NUTS5/LAU2 REGIONS



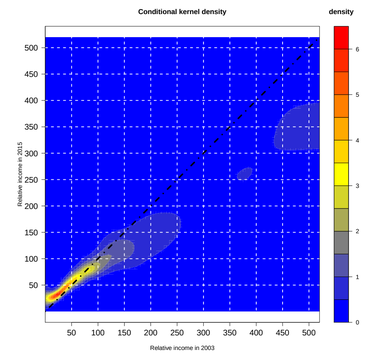
2002



2015

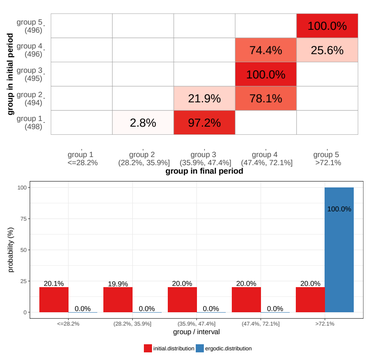


Transition matrix – real data

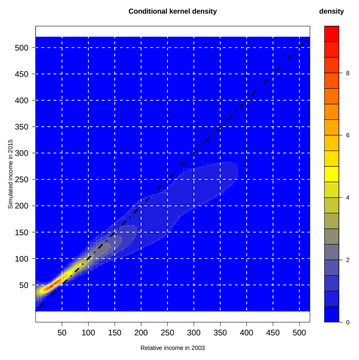


Kernel density – real data

SIMULATED CONVERGENCE

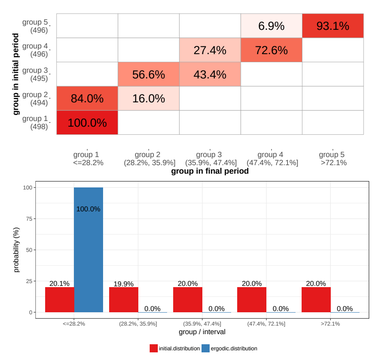


Transition matrix

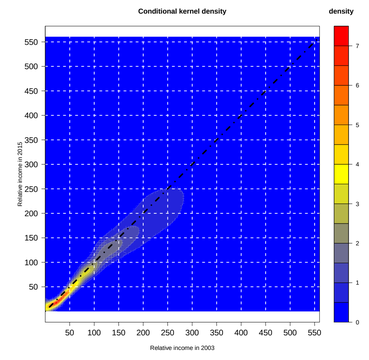


Conditional kernel density

SIMULATED DIVERGENCE

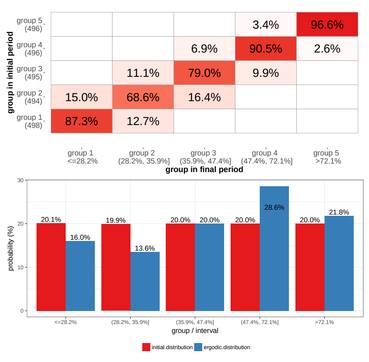


Transition matrix

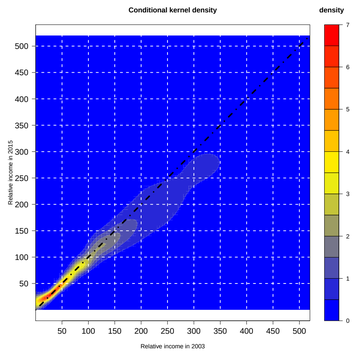


Conditional kernel density

SIMULATED PERSISTENCE

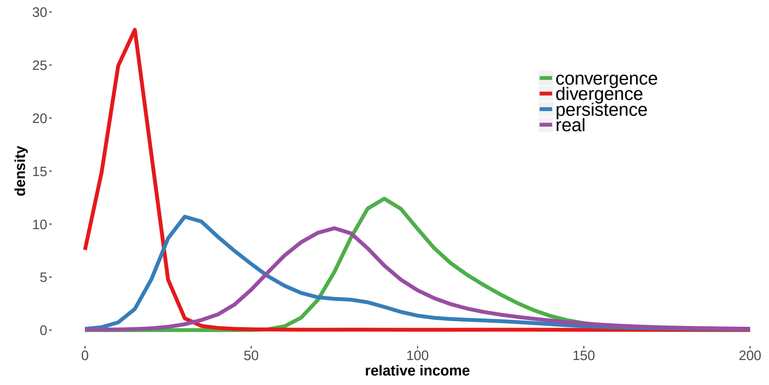


Transition matrix



Conditional kernel density

ERGODIC KERNEL DISTRIBUTIONS



CONCLUSIONS

- distribution of income on a local level is most persistent in highest income groups,
- ergodic pattern for real data is between the simulated convergence and persistence,
- results suggest convergence within several grous (clubs) of regions.