

Libraries' pre- and post-Covid-19 life in Poland - spatial approach

(work in progress)

Kateryna Zabarina (p), University of Warsaw
Prof. Codruța Mare, Babeș-Bolyai University



UNIVERSITY
OF WARSAW



FACULTY OF
ECONOMIC SCIENCES

Agenda

1. Introduction (and motivation)
2. Research method
3. Data (source, descriptive statistics)
4. Results and findings

Introduction (and motivation)

- The goal of the paper is to identify main factors which determine number of readers in Polish libraries, both public and specialised (scientific and educational libraries and scientific societies' libraries).
- In this study both types of libraries are analysed, since their activities were affected by COVID-19 in the same way.
- Authors believe that spatial approach is the most appropriate, since it allows to take into account heterogeneity of Polish powiats.

Introduction (and motivation)

- Literature review shows, that the majority of articles concerning libraries is policy-related (what is already done and what has to be done)
- Some studies are about whether people started to use ebooks more
- The most often used method of analysis are surveys and statistics, showing how habits have changed

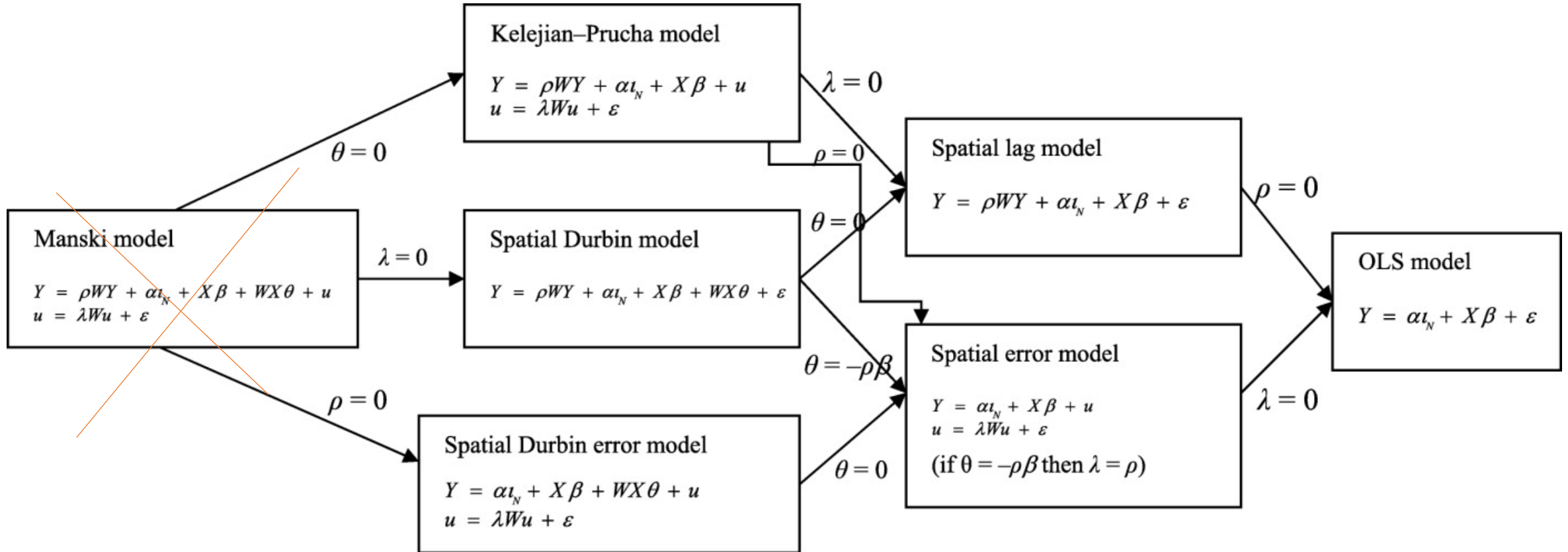
Literature related to habits:

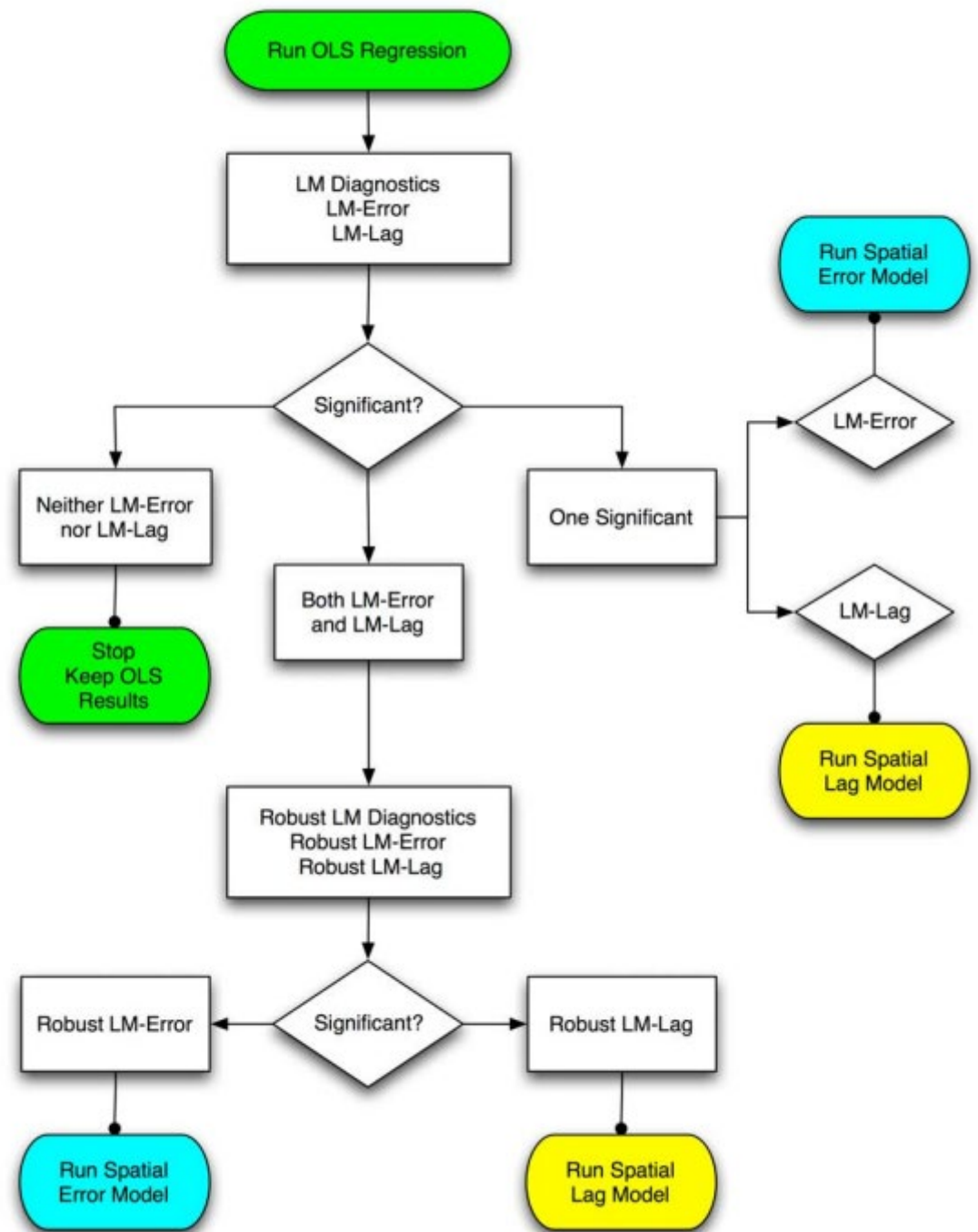
- Surveys (rather mostly)
- Anova analysis of how reading habits had changed (mostly); regression (ordered logit); chi-square test; statistical summary of survey
- Rather smaller groups of people (from university/couple of universities), some teachers and/or students etc.

Research method

- Inside the study we compare performance of six spatial regression models (SAR, SEM, SLX, SDM, SDEM, SARAR (SAC));
- We didn't account for Manski (most general) model – it is rather overidentified;
- We collected powiat's libraries' data (among others, number of readers, computerization of libraries, population, etc.) from 2013 to 2021 and estimated each of abovementioned models for each year in chosen timespan in order to analyse coefficients' dynamics;
- We used both forward and backward options of estimation;
- Later, the best model was determined with likelihood-ratio tests and information criteria.

Research method





Source: Anselin (1988)

Data (source, descriptive statistics)

- Number of readers (annual) in public and other (scientific, educational and scientific societies') libraries (these amounts were summed up into dependent variable);
- Computerisation of libraries: computers used in the library available to readers, computers with Internet access available to readers (in order to calculate percentage of computers with internet and proportion of computers per reader);
- Population in powiat and under 19 (*) population in powiat (in order to calculate percentage of young population);
- Expenditures of powiat (**) budgets (total);
- Revenues of powiat (***) budgets (total);
- Dummy variable indicating whether powiat city is a core city;
- Dummy variable indicating whether given powiat has University (****).

* - Categories in Gus are as follows:

** - These data were collected separately for powiats and cities with powiat status

** - As above

*** - Authors analysed each powiat separately. An object was not considered as university, if it was a branch of another big university. For example, Wrocław University of Science and Technology has several branches, but was counted only as university in Wrocław, and no branch was considered as a university in a given powiat.

Wiek

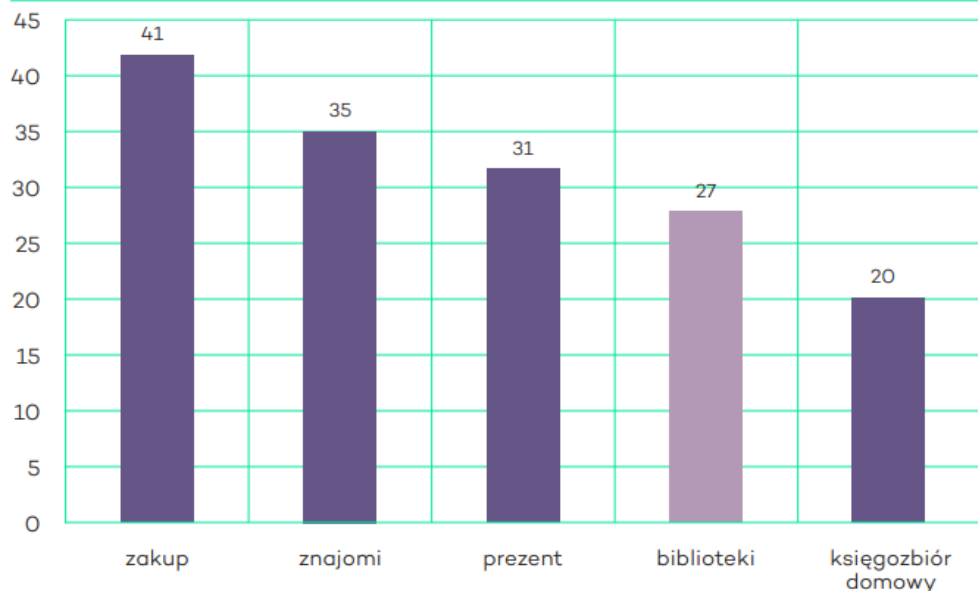
| |
|--------|
| ogółem |
| 0-4 |
| 5-9 |
| 10-14 |
| 15-19 |
| 20-24 |
| 25-29 |

Data (source, descriptive statistics)

Number of readers (total):

| Year | Min | Mean | Max |
|------|---------|----------|-----------|
| 2013 | 1292.00 | 22275.95 | 677000.00 |
| 2014 | 2398.00 | 21394.09 | 672702.00 |
| 2015 | 2424.00 | 20977.78 | 654339.00 |
| 2016 | 2201.00 | 20208.72 | 630457.00 |
| 2017 | 1951.00 | 19782.29 | 629874.00 |
| 2018 | 1710.00 | 19376.04 | 616051.00 |
| 2019 | 1706.00 | 19316.73 | 624145.00 |
| 2020 | 1574.00 | 15482.67 | 469264.00 |
| 2021 | 1405.00 | 14953.89 | 440707.00 |

WYKRES 3. Podstawowe źródła książek w 2019 r. Odsetek czytelników (n=1128) wymienających książki z danego źródła (dane w proc.)

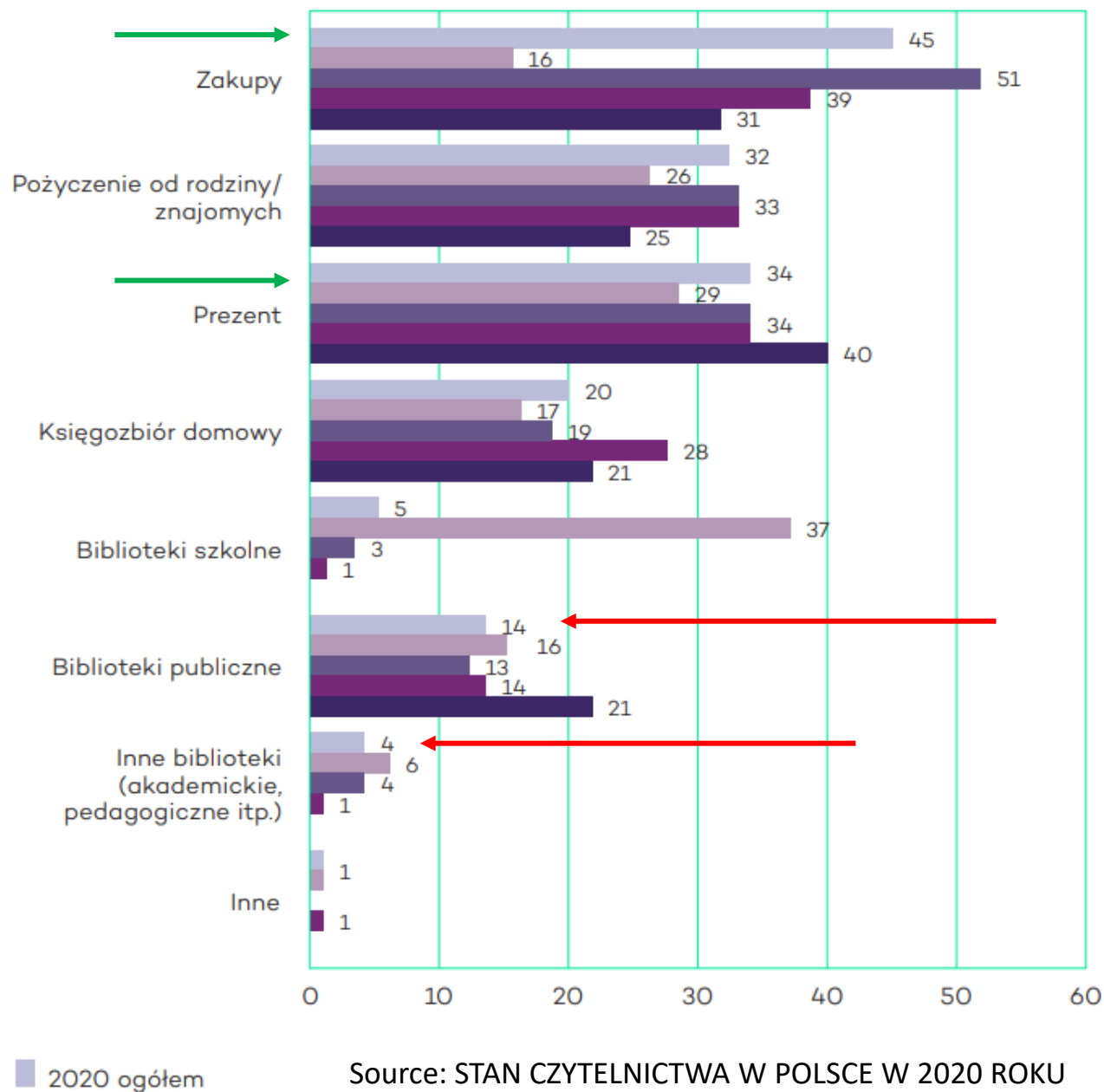


Odsetki nie sumują się do 100%, ponieważ czytelnicy wymieniali książki pochodzące z różnych źródeł.

Źródło: Biblioteka Narodowa

Source: STAN CZYTELNICTWA W POLSCE W 2019 ROKU.
WSTĘPNE WYNIKI

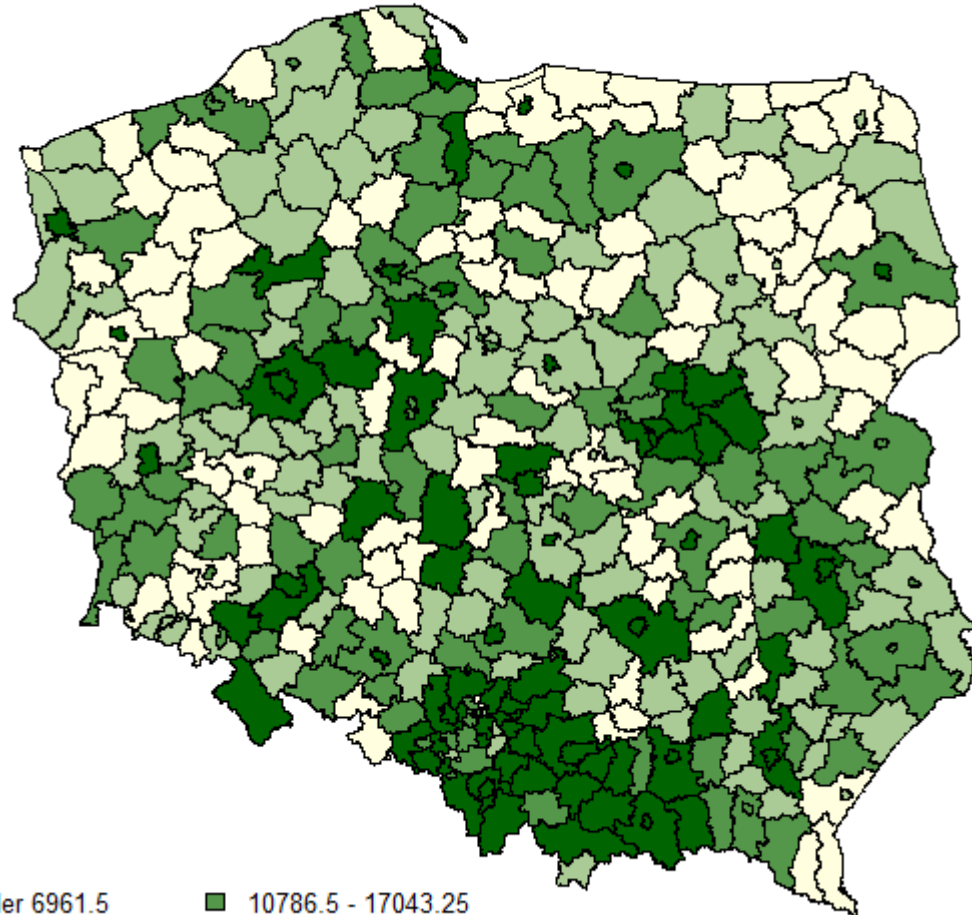
WYKRES 6. Źródła pochodzenia co najmniej jednej przeczytanej książki drukowanej na papierze w 2020 roku oraz według wieku czytelników książek, N=853 (w procentach)



Source: STAN CZYTELNICTWA W POLSCE W 2020 ROKU

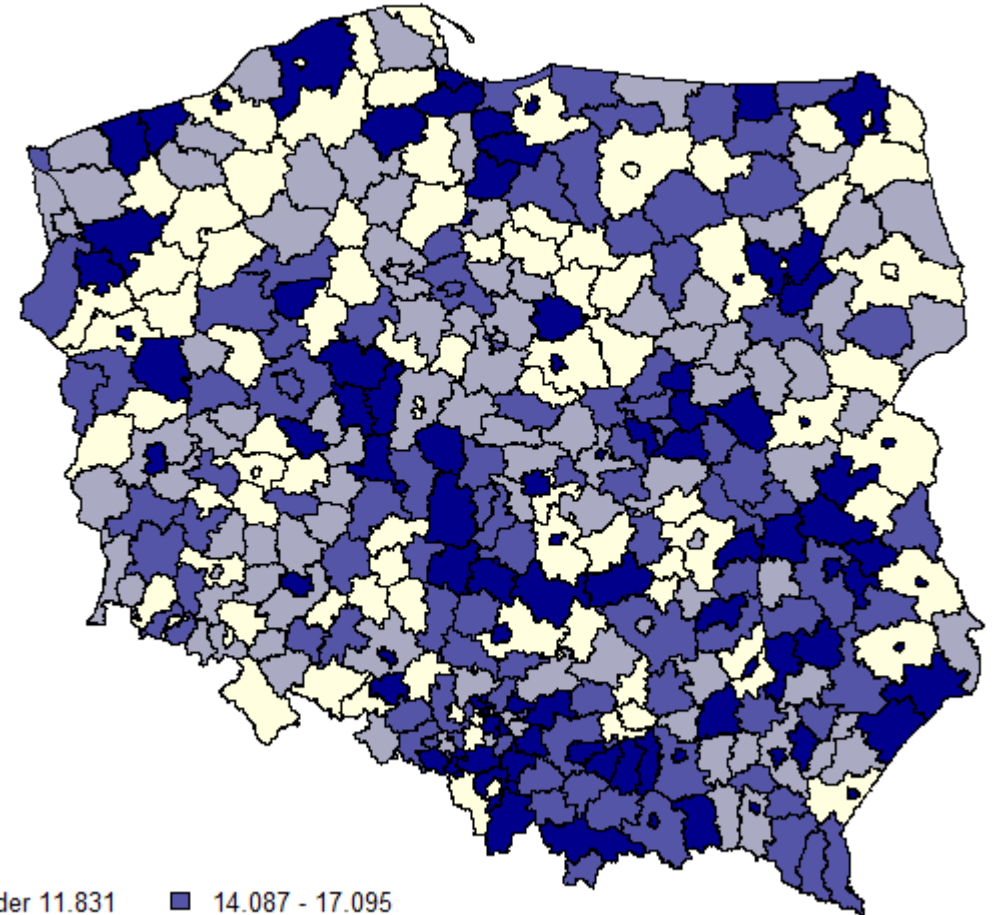
Data (source, descriptive statistics)

Number of readers in libraries, 2019



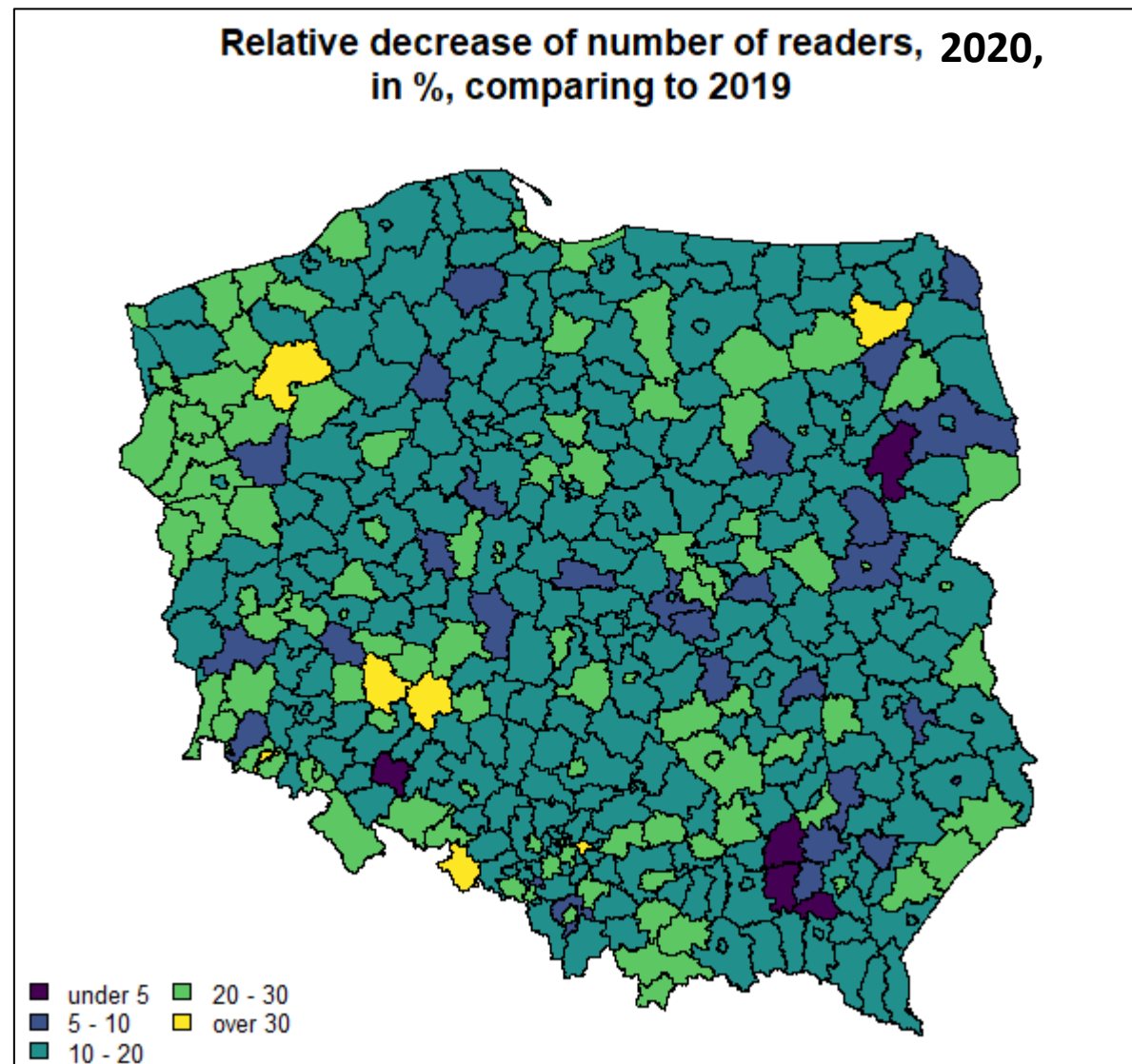
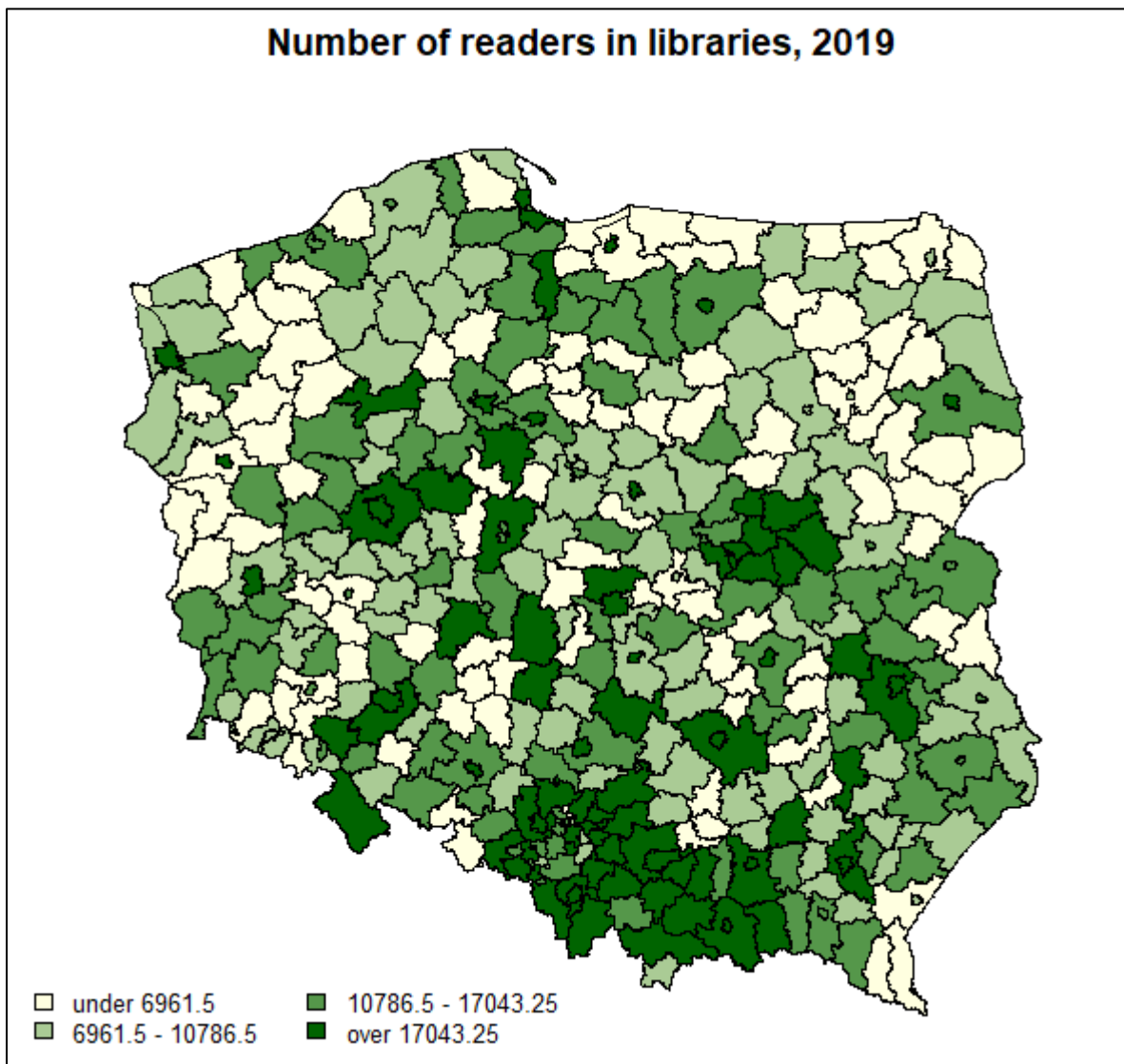
under 6961.5 10786.5 - 17043.25
6961.5 - 10786.5 over 17043.25

Percentage of people using libraries, 2019



under 11.831 14.087 - 17.095
11.831 - 14.087 over 17.095

Data (source, descriptive statistics)



Second map is consistent with GUS report for 2020: „# of readers decreased by 17.9%, # of borrowed books – by 24,5%”

(source: [Ciężki rok polskich bibliotek w pandemii. Według danych GUS spadła liczba czytelników, wypożyczeń i książek w zbiorach : Booklips.pl](https://booklips.pl))

Data (source, descriptive statistics)

Bright side of pandemics:

WYKRES 1. Czytanie książek w latach 2000–2020 (dane w procentach)



Źródło: Biblioteka Narodowa

Results and findings

Preliminary analysis:

| Year | Moran OLS p-v | LMerr p-v | LMLag p-v | RLMerr p-v | RLMLag p-v | SARMA p-v |
|-------------|---------------|-----------|-----------|------------|------------|-----------|
| 2013 | 9,10E-16 | 1,12E-14 | 1,76E-09 | 5,29E-07 | 0,190891 | 4,69E-14 |
| 2014 | 1,11E-15 | 1,34E-14 | 9,86E-09 | 1,75E-07 | 0,358262 | 8,63E-14 |
| 2015 | 3,18E-18 | 0 | 2,41E-09 | 2,88E-09 | 0,506383 | 4,44E-16 |
| 2016 | 4,51E-18 | 1,11E-16 | 5,67E-12 | 5,86E-07 | 0,104238 | 2,22E-16 |
| 2017 | 2,76E-17 | 3,33E-16 | 2,37E-12 | 4,82E-06 | 0,052577 | 6,66E-16 |
| 2018 | 7,10E-18 | 1,11E-16 | 7,40E-13 | 3,99E-06 | 0,048999 | 1,11E-16 |
| 2019 | 1,46E-16 | 2,00E-15 | 6,06E-12 | 1,09E-05 | 0,059981 | 3,33E-15 |
| 2020 | 7,30E-19 | 0 | 2,29E-12 | 3,26E-07 | 0,13878 | 0 |
| 2021 | 1,63E-19 | 0 | 1,79E-12 | 1,05E-07 | 0,19357 | 0 |

➤ Spatial autocorrelation -> should not use OLS

➤ Using Anselin's algorithm:

- Both LMerr and LMLag are significant, so we look further, on RLMerr and RLMLag ->
- RLMerr is significant, so we have to run Spatial Error Model
- RLMLag is significant (on 5% significance level) only for 1 model, hence Spatial Lag Model is not a good choice, according to algorithm

Results and findings

eq <- log(readers) ~ core_city + university + young + internet_perc + log(revenues) + log(expenditures)

| Year | AIC SLX | AIC SEM | AIC SAR | AIC SDEM | AIC SDM | AIC SARAR |
|-------------|---------|---------------|---------|----------|---------|---------------|
| 2013 | 358,16 | 304,24 | 325,85 | 311,67 | 312,12 | 305,99 |
| 2014 | 349,51 | 297,16 | 319,49 | 303,95 | 303,72 | 298,52 |
| 2015 | 350,53 | 287,91 | 316,08 | 296,61 | 296,84 | 288,32 |
| 2016 | 382,15 | 319,30 | 338,61 | 327,28 | 328,32 | 320,99 |
| 2017 | 384,77 | 326,66 | 342,67 | 335,27 | 335,57 | 328,39 |
| 2018 | 401,47 | 342,90 | 359,45 | 346,29 | 348,49 | 344,46 |
| 2019 | 407,99 | 351,76 | 366,64 | 361,26 | 361,41 | 353,31 |
| 2020 | 415,03 | 349,63 | 368,83 | 359,43 | 359,81 | 350,81 |
| 2021 | 399,45 | 341,14 | 363,60 | 345,44 | 345,56 | 340,30 |

| Year | Hausman p-v |
|-------------|-------------|
| 2013 | 0,003581 |
| 2014 | 0,001319 |
| 2015 | 0,00387 |
| 2016 | 0,000116 |
| 2017 | 6,68E-05 |
| 2018 | 2,85E-05 |
| 2019 | 0,000488 |
| 2020 | 0,000513 |
| 2021 | 7,34E-06 |

| Year | BIC SLX | BIC SEM | BIC SAR | BIC SDEM | BIC SDM | BIC SARAR |
|-------------|---------|---------------|---------|----------|---------|---------------|
| 2013 | 413,33 | 339,70 | 361,31 | 370,77 | 371,22 | 345,39 |
| 2014 | 404,68 | 332,63 | 354,95 | 363,06 | 362,83 | 337,92 |
| 2015 | 405,70 | 323,37 | 351,55 | 355,72 | 355,94 | 327,72 |
| 2016 | 437,31 | 354,76 | 374,07 | 386,39 | 387,42 | 360,39 |
| 2017 | 439,93 | 362,12 | 378,13 | 394,37 | 394,67 | 367,79 |
| 2018 | 456,63 | 378,36 | 394,91 | 405,39 | 407,60 | 383,86 |
| 2019 | 463,16 | 387,22 | 402,10 | 420,36 | 420,52 | 392,71 |
| 2020 | 470,20 | 385,10 | 404,29 | 418,54 | 418,91 | 390,21 |
| 2021 | 454,61 | 376,60 | 399,06 | 404,54 | 404,67 | 379,70 |

Green – minimum value, yellow – second minimum value

Hausman test – compares SEM vs OLS. When H0 is rejected, discrepancy between two models is too big, and neither of them should be used (Pace, R. K., & LeSage, J. P., 2008)

Results and findings

| Year | Moran SLX p-v | Moran SEM p-v | Moran SAR p-v | Moran SDEM p-v | Moran SDM p-v | Moran SARAR p-v |
|------|---------------|---------------|---------------|----------------|---------------|-----------------|
| 2013 | 0,00 | 0,00 | 0,76 | 0,75 | 0,73 | 0,77 |
| 2014 | 0,00 | 0,00 | 0,79 | 0,78 | 0,78 | 0,81 |
| 2015 | 0,00 | 0,00 | 0,82 | 0,81 | 0,80 | 0,84 |
| 2016 | 0,00 | 0,01 | 0,78 | 0,77 | 0,75 | 0,79 |
| 2017 | 0,00 | 0,02 | 0,77 | 0,76 | 0,76 | 0,79 |
| 2018 | 0,00 | 0,02 | 0,81 | 0,82 | 0,77 | 0,83 |
| 2019 | 0,00 | 0,03 | 0,79 | 0,79 | 0,79 | 0,81 |
| 2020 | 0,00 | 0,01 | 0,80 | 0,81 | 0,80 | 0,83 |
| 2021 | 0,00 | 0,01 | 0,86 | 0,86 | 0,85 | 0,89 |

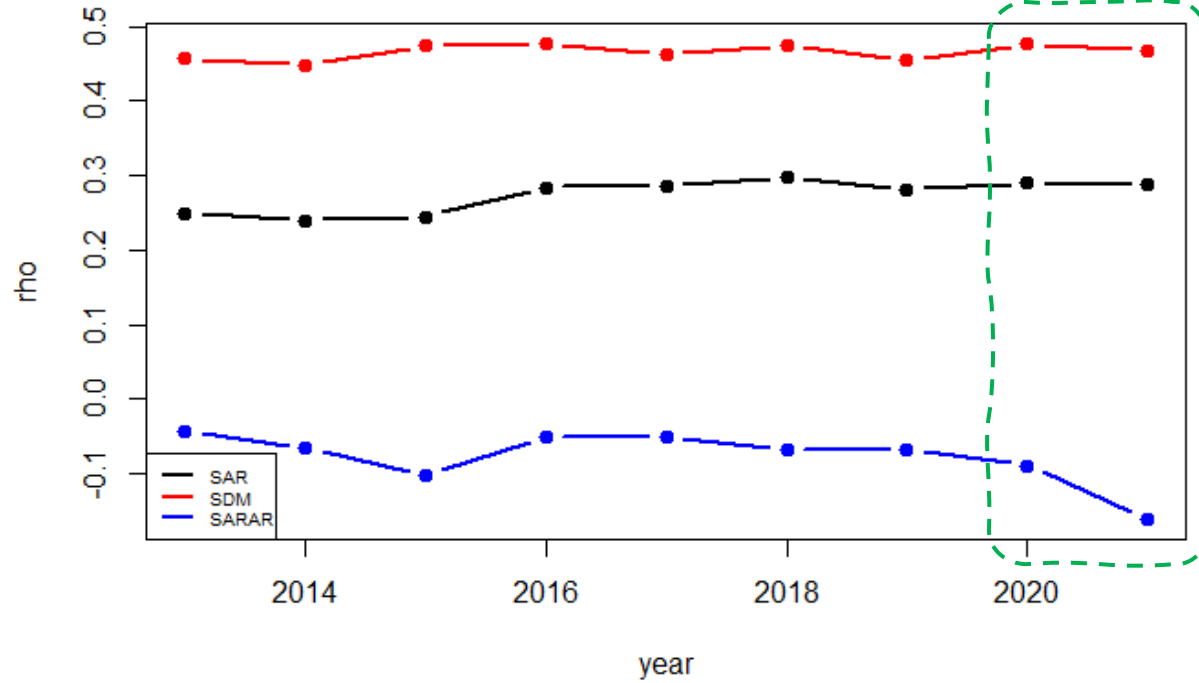
Our conclusions are as follows:

- We see, that in SEM model, spatial autocorrelation is not eliminated, which makes us to choose SARAR (SAC model), according to both AIC and BIC as second best model
- When applying GeTS procedure -> results are surprising.

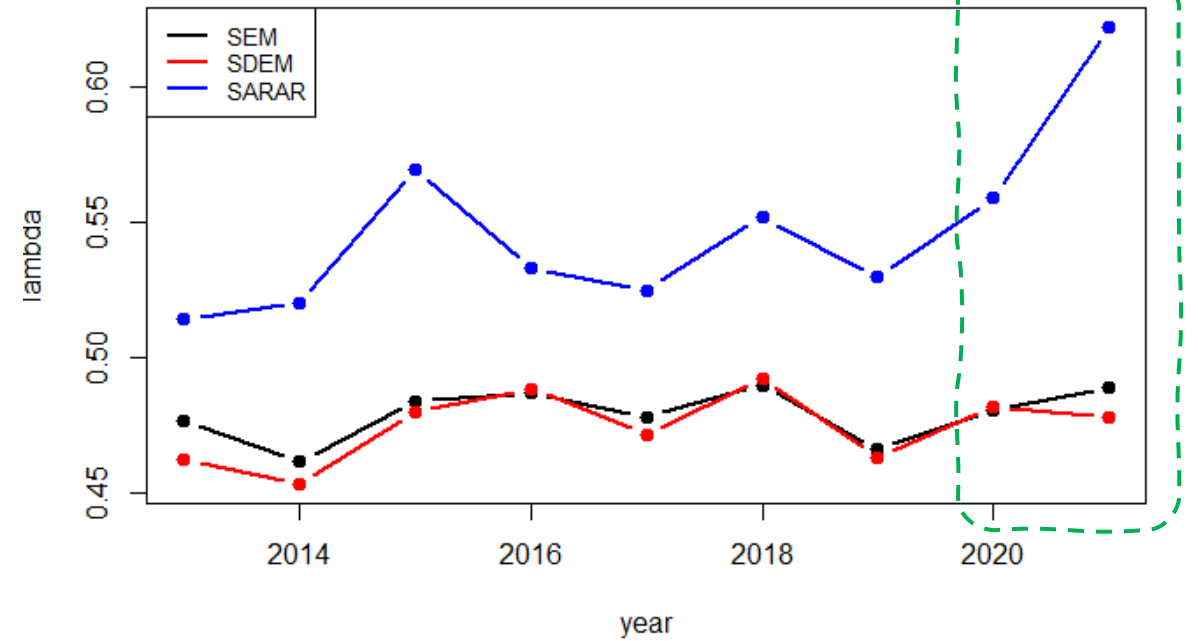
| Year | SARARtoSEMp-v | SARARtoSARp-v |
|------|---------------|---------------|
| 2013 | 0,614 | 0,000 |
| 2014 | 0,421 | 0,000 |
| 2015 | 0,207 | 0,000 |
| 2016 | 0,577 | 0,000 |
| 2017 | 0,608 | 0,000 |
| 2018 | 0,505 | 0,000 |
| 2019 | 0,502 | 0,000 |
| 2020 | 0,363 | 0,000 |
| 2021 | 0,092 | 0,000 |

Results and findings

rho coefficients



lambda coefficients



Results and findings

eq <- log(readers) ~ core_city + university + young + internet_perc + log(revenues) + log(expenditures)

| Year | AIC SLX | AIC SEM | AIC SAR | AIC SDEM | AIC SDM | AIC SARAR |
|------|---------|---------|---------|----------|---------|-----------|
| 2013 | 358,16 | 304,24 | 325,85 | 311,67 | 312,12 | 305,99 |
| 2014 | 349,51 | 297,16 | 319,49 | 303,95 | 303,72 | 298,52 |
| 2015 | 350,53 | 287,91 | 316,08 | 296,61 | 296,84 | 288,32 |
| 2016 | 382,15 | 319,30 | 338,01 | 327,28 | 328,32 | 320,99 |
| 2017 | 384,77 | 326,66 | 342,07 | 335,27 | 335,57 | 328,39 |
| 2018 | 401,47 | 342,90 | 359,45 | 346,29 | 348,49 | 344,46 |
| 2019 | 407,99 | 351,76 | 366,04 | 361,26 | 361,41 | 353,31 |
| 2020 | 415,03 | 349,63 | 368,83 | 359,43 | 359,81 | 350,81 |
| 2021 | 399,45 | 341,14 | 363,00 | 345,44 | 345,56 | 340,30 |

| Year | BIC SLX | BIC SEM | BIC SAR | BIC SDEM | BIC SDM | BIC SARAR |
|------|---------|---------|---------|----------|---------|-----------|
| 2013 | 413,33 | 339,70 | 361,31 | 370,77 | 371,22 | 345,39 |
| 2014 | 404,68 | 332,63 | 354,95 | 363,06 | 362,83 | 337,92 |
| 2015 | 405,70 | 323,37 | 351,55 | 355,72 | 355,94 | 327,72 |
| 2016 | 437,31 | 354,76 | 374,07 | 386,39 | 387,42 | 360,39 |
| 2017 | 439,93 | 362,12 | 378,13 | 394,37 | 394,67 | 367,79 |
| 2018 | 456,63 | 373,36 | 394,91 | 405,39 | 407,60 | 383,86 |
| 2019 | 463,16 | 387,22 | 402,10 | 420,36 | 420,52 | 392,71 |
| 2020 | 470,20 | 385,10 | 404,29 | 418,54 | 418,91 | 390,21 |
| 2021 | 454,61 | 375,60 | 399,06 | 404,54 | 404,67 | 379,70 |

We eliminate SLX, SAR, SEM and SAC

Which one is better now – SDEM or SDM? -> further analysis needed

Results and findings

- Estimation results show, that simple non-spatial regression model cannot be used to determine number of readers in Polish libraries.
- The model which performs best is still a question to solve (interpretation is omitted until the correct specification will be found). Two possible candidates are SDM and SDEM.
- From graphs showing dynamics of rho and lambda parameters, we conclude that they were changing over time, which suggests we may consider to include variables which may explain such dynamics. These graphs show also, that SARAR (despite the fact being chosen by information criteria) seems to be not a good choice.
- Among significant determinants having impact on number of readers – core city dummy, proportion of young population and powiat revenues.

Thank you for attention!
(all questions and suggestions are welcome 😊)