

# What are the courts deciding and what will they decide? Topic modelling and time series analysis on the example of the Polish judiciary - by Maciej Światała

comments by Paweł Sakowski  
WNE UW Conference 2022

September 29th, 2022

The general goal of a topic model is to produce interpretable document representations which can be used to discover the topics or structure in a collection of unlabelled documents. An example of such an interpretable document representation is: document X is 20% topic a, 40% topic b and 40% topic c.

The main aim:

- to develop a methodology for analysing variability of **thematic motives** observed in courts' decisions over time:
  - ① **classification** of courts judgements into a number of identified legal problem (*topic modelling*)
  - ② generating time series of relative frequencies of those above and then modelling them with ARIMA (including identification and **forecasting**)

These two topics might be subject of being extended.

Contribution:

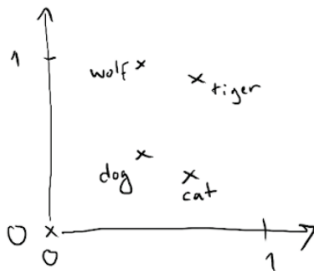
- combining topic modelling with well-established TSA tools

**A very short sample** → only 48 observations for each time series!

- It affects a number of steps in the whole procedure:
  - low power of stationarity steps
  - inaccurate estimates of seasonal variables, ARIMA parameters, ACF & PACF values
  - as a result: optimal orders of ARIMA are affected as well!
  - short out-of-sample period! → extremely hard to compare *ex-post* forecasting errors
- Is it possible to query longer samples?

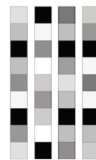
Latent Dirichlet Allocation (LDA) is used to obtain thematical groups of judgements

- other alternatives?
  - word-embeddings?
  - word2vec?
  - **lda2vec?**
- any validations of such classification?



One-hot word vectors:

- Sparse
- High-dimensional
- Hard-coded



Word embeddings:

- Dense
- Lower-dimensional
- Learned from data

# Questions and issues which needs to be addressed I

- ① What is the motivation behind using LDA to assign cases to particular topics? Are there any other interesting alternatives? What about manual assigning?
- ② Unclear description of the way the LDA has been applied to classification process.
- ③ How exactly the perplexity metric was used?
- ④ Include naive forecasts as a benchmark!

# Questions and issues which needs to be addressed II

- ⑤ Calculating averages of probabilities (of topics for a particular document) for a given time interval - lacking description of the process. A scheme perhaps?
- ⑥ Is anywhere in the analysis included the distance in time between:
  - ① beginning of the case in court and final judgement?
  - ② time of all activities involved before the case is presented in the court?
- ⑦ Perhaps there are cases which are ready to be presented in court but have to wait until there are appropriate resources available?
- ⑧ Weekly frequency? Is it worth it?

# Questions and issues which needs to be addressed III

- 9 What is the implication of finding cointegrated series. Why to search for them? Does it provide better forecasts or only useful for classification purposes?
- 10 Scales on vertical axes on time series plots - can those values be interpreted as probabilities? Do they sum up to 100% across all topics in a given month?
- 11 Optimal orders of ARIMA: how? Based on which criterion?
- 12 Why only KPSS results are presented? Why not to present ADF and PP as well?



A simple simulation of AR(1):

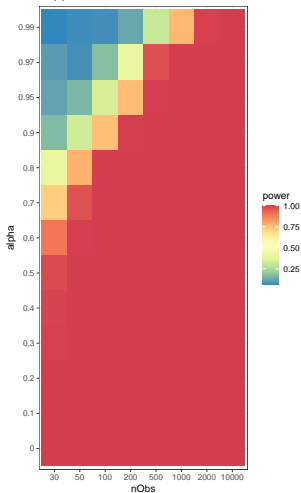
$$y_t = \alpha_1 y_{t-1} + \varepsilon_t$$

where:

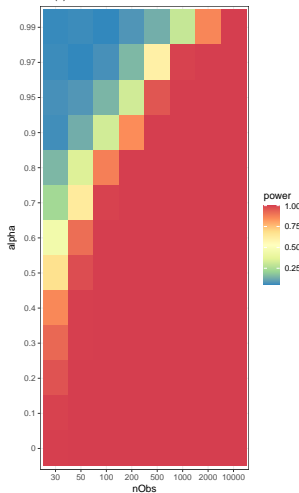
- $t = 1, \dots, T$
- $\alpha_1 \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.97, 0.99\}$
- $T \in \{39, 50, 100, 200, 500, 1000, 2000, 10000\}$
- $\varepsilon_t \text{ IID } \sim N(0, 0.1)$

# DF vs. PP

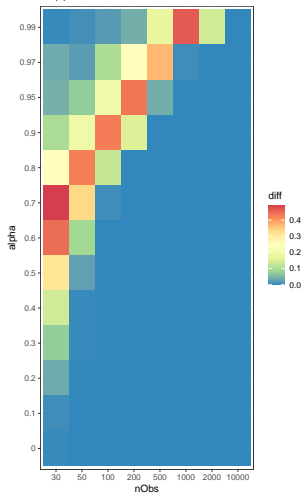
Power of DF test (fraction of correct decisions)  
AR(1) with 500 simulations for each combination



Power of PP test, ie. fraction of correct decisions  
AR(1) with 500 simulations for each combination

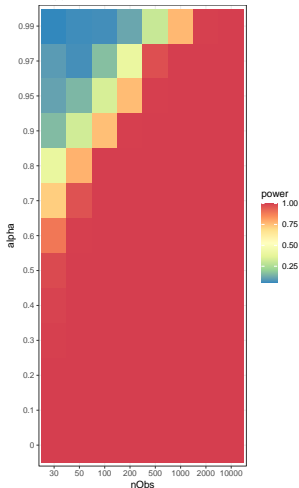


DF vs. PP: differences in power  
AR(1) with 500 simulations for each combination

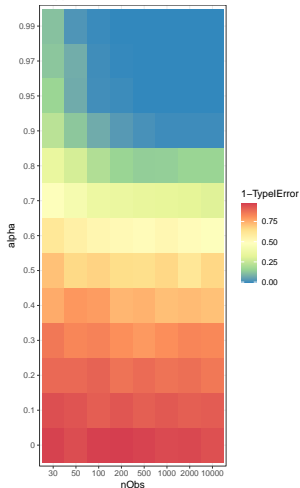


# DF vs. KPSS

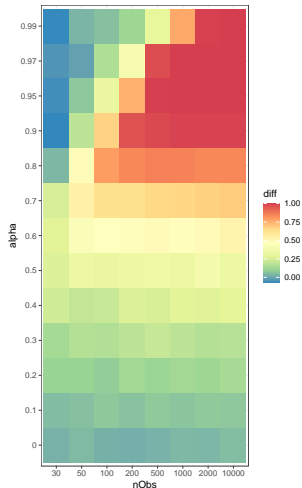
Power of DF test (fraction of correct decisions)  
AR(1) with 500 simulations for each combination



(1 - Type I Error) for KPSS test  
AR(1) with 500 simulations for each combination



DF vs. KPSS: diffs in fractions of correct decisions  
AR(1) with 500 simulations for each combination



# Suggestions

- Why not to concentrate more on classification problem? More about LDA and alternatives, less about ARIMA and tests.
- Query more data if it is possible. With more data you can do:
  - multivariate cointegration with VECM
  - SARIMA models
  - VAR with possibly better forecasts
  - perhaps even LSTM for sequential data!
- Data can be possibly obtained with lower costs using SAOS API.

Documentation:

- <https://www.saos.org.pl/help/index.php/dokumentacja-api/api-pobierania-danych>

Query example:

```
https://www.saos.org.pl/api/dump/judgments?  
    pageSize=100  
    &pageNumber=1&  
    &judgmentStartDate=2016-01-03  
    &judgmentEndDate=2016-01-07  
    &withGenerated=true
```

- longer time series
- other countries (comparison between them)
- seasonal ARIMA models
- multivariate cointegration and VECM
- forecasting with VAR
- forecasting with LSTM