NATIONAL RESEARCH UNIVERSITY

Aggregation of Individual Credit Assessments as a Problem of Consensus in Expert System

Marat Kurbangaleev
Financial Engineering and Risk Management Lab
Higher School of Economics, Moscow
(joint work with Buzdalin A.V. and Smirnov S.N.)

Perm Winter School 2017, Perm
February, 4

Interpretation of Credit Ratings

- Quote from Kiff et al, 2012 (IMF Report):
  - "Standard & Poor's credit ratings are designed primarily to provide relative rankings among issuers and obligations of overall creditworthiness; the ratings are not measures of absolute default probability. Creditworthiness encompasses likelihood of default and also includes payment priority, recovery and credit stability" (S&P, 2009).
  - "Credit ratings express risk in relative rank order, which is to say they are ordinal measures of credit risk and are not predictive of a specific frequency of default or loss. Fitch Ratings' credit ratings do not directly address any risk other than credit risk, ratings do not deal with the risk of a market value loss on a rated security due to changes in interest rates, liquidity and other market considerations" (Fitch Ratings, 2010).
  - Moody's credit ratings are opinions of the credit quality of individual obligations or of an issuer's general creditworthiness... Moody's... ratings are opinions of the relative credit risk of fixed-income obligations.... They address the possibility that a financial obligation will not be honored as promised. Such ratings... reflect both the likelihood of default and any financial loss suffered in the event of default" (Moody's Investors Service, 2009).

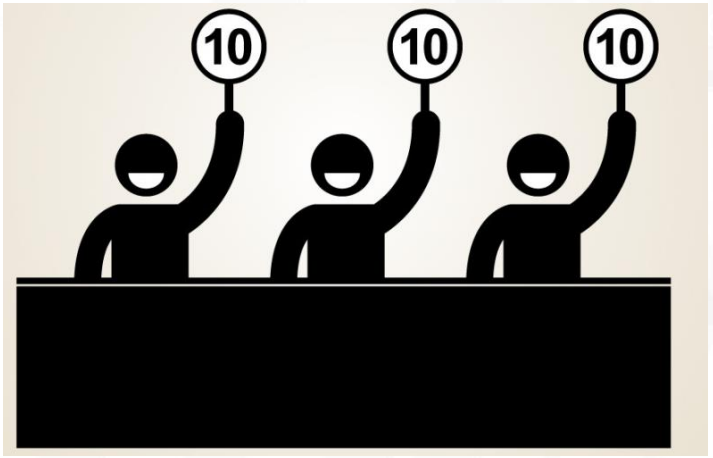
Motivation to Aggregate Ratings

- Several applications are related to described problem.
  - Synthetic general rating.
    - Speaks for itself.
  - Informational content of multiple ratings.
    - Do multiple ratings “beat” credit model based on public information?
  - Mapping of credit rating scales (mainly for using them in regulation).
    - Constructing a “fair” benchmark scale in low-default environment.

Credit Ratings as Expert Information

- Due to these properties, ratings assigned by credit rating agencies can be considered as individual assessments (judgments) in some expert system.

- Example of expert systems:
  - sports judging
  - student rankings
  - etc.



Typical Structure of Consensus Problems

- Expert opinions.
  - typically,unified scales,
  - typically, “all rate all”.
  - typically, opinions are assumed to be independent.
- Function describing discrepancies among expert opinions.
- Criterion.

Features of Rating Data (1st slide)

- Credit risk definition.
  - List of credit events
  - Realized default event vs Event which will happen “almost surely”.
  - Default vs Default + Losses.

- Risk horizon.
  - Short-term vs long-term.
  - Point-in-time vs through-the-cycle.

Features of Rating Data (2nd slide)

- Actualization frequency.
  - Typically, credit agencies update their assessments as a new relevant information arrives.
  - Depending on structure of its assessment process agency may take actions relatively fast or slow.
  - Agencies may also strategically ignore or misrepresent some relevant information to inflate their ratings.

- Number and names of rating agencies servicing particular entity.
  - Depend on
    - entity's purpose;
    - global and local financial regulation;
    - competition among agencies.
  - Can change over time.

Features of Rating Data (3rd slide)

- Scales.
  - Different numbers of categories.
  - Different starting point (credit quality benchmark).
- Types: international, national.
- Generally, national scales are incomparable.
- National and international scales are locally comparable.

Russia Mapping Table

Global-scale long-term local-currency rating	National-scale long-term rating
BBB- and above	ruAAA
BB+	ruAA+
BB	ruAA
BB-	ruAA-
B+	ruA+, ruA
B	ruA, ruA-, ruBBB+
B-	ruBBB, ruBBB-
CCC+	ruBB+, ruBB, ruBB-
CCC	ruB+, ruB, ruB-
CCC-	ruCCC+, ruCCC, ruCCC-
CC	ruCC
C	ruC
R	R
SD	SD
D	D

R--Regulatory supervision. SD--Selective default. D--Default.

Source: www.standardandpoors.com

Selection of Rating Data

- Reasonable selection principles help to deal with such features:
  - select comparable in time rating scales with comparable horizon of risk;
  - pick rating around dates when new relevant information arrives (macroeconomic statistics, financial accounting, etc.)

	Ag_1	Ag_2		Ag_m
a_1	r^1_1	NR		r^m_1
a_2	r^1_2	r^2_2		r^m_2
a_3	NR	r^2_3	...	r^m_3
...	...	...		...
a_n	r^1_n	NR		r^m_n

Measuring the Discrepancy

- For all agencies k,l consider the
- discrepancy matrix $\Delta_{k,l}$ with the
- elements defined by the table.

	$r_i^k < r_j^k$	$r_i^k = r_j^k$	$r_i^k > r_j^k$	r_i^k or r_j^k is NR
$r_i^l < r_j^l$	0	1	2	w
$r_i^l = r_j^l$	1	0	1	w
$r_i^l > r_j^l$	2	1	0	w
r_i^l or r_j^l is NR	w	w	w	0

- Here w is any number. It will not matter.
- Distance between two agencies is the sum of individual discrepancies.

$$d(\mathrm{Ag}_k, \mathrm{Ag}_l) = \sum_{i,j=1}^{m} \delta_{i,j}^{k,l}$$

- Consensus ranking *R* such that it is defined for all companies, and

$$\sum_{k=1}^{n} d(R, Ag_k) \to \min$$

- This is known as the **Kemeny median**.
- Finding all medians is computationally infeasible, therefore we propose an approach based on the Tikhonov regularization and the genetic optimization.

Data Selection

Property	Specification
Period	July 2010 – July 2015
Entities	Russian banks
Agencies	3 international agencies + 4 local agencies
Scales	National, long-term
Timeframe	Quarter

Number of agencies	Observations	% all observations
1	5604	74.5
2	1414	18.8
3	384	5.1
4	100	1.3
5	18	0.2

Data Structure (1st slide)

- Number of pairwise combinations of ratings.

	Ag1	Ag2	Ag3	Ag4	Ag5	Ag6	Ag7
Ag1	x	210	303	321	303	257	54
Ag2	210	x	133	148	176	198	23
Ag3	303	133	x	94	86	195	0
Ag4	321	148	94	x	162	111	93
Ag5	303	176	86	162	x	118	37
Ag6	257	198	195	111	118	x	0
Ag7	54	23	0	93	37	0	x
Multiple ratings	885	480	514	592	554	501	147
All ratings	1174	582	734	3146	1065	674	511
% of multiple ratings	75.38%	82.47%	70.03%	18.82%	52.02%	74.33%	28.77%

Data Structure (2nd slide)

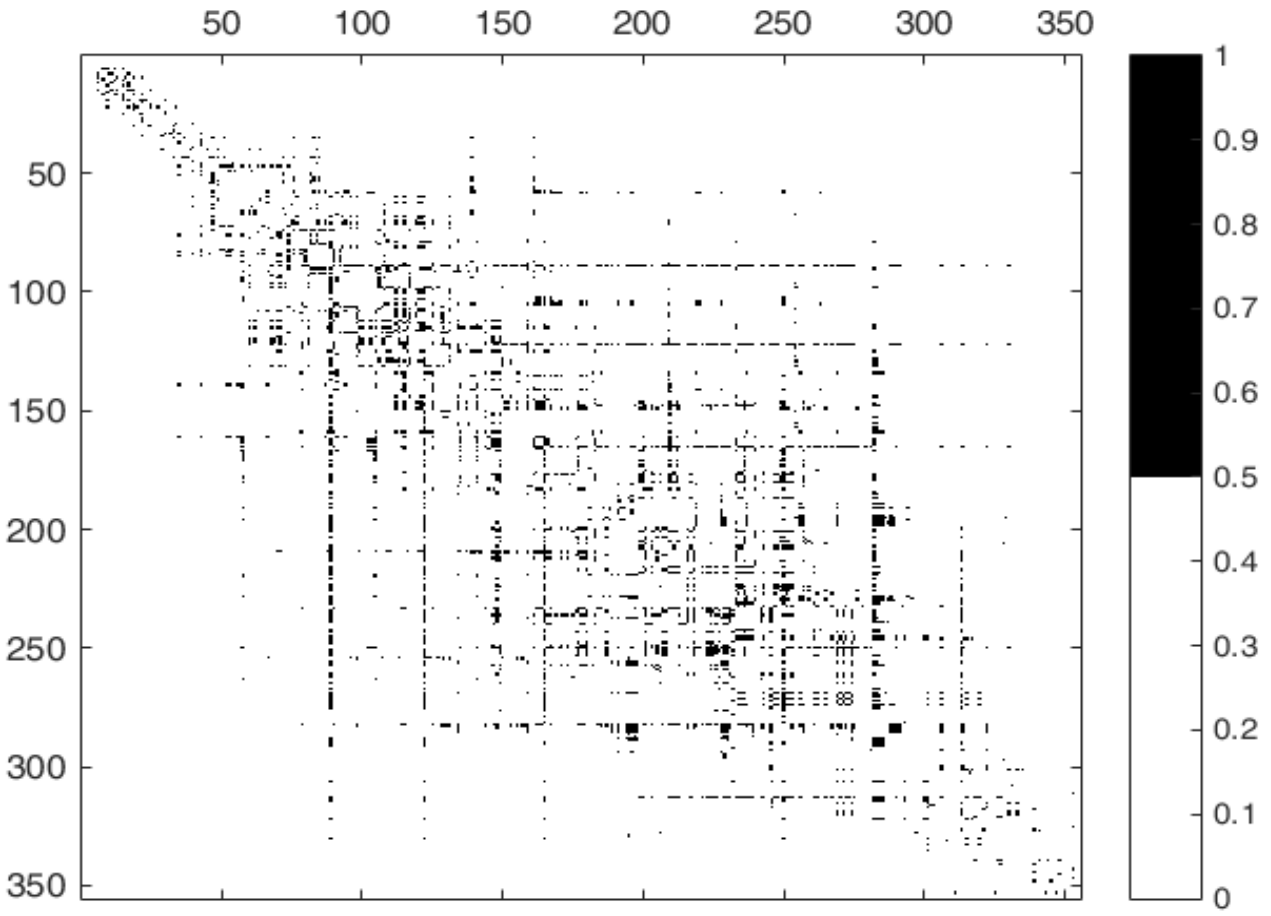
- Kendall τ_x correlation

	Ag1	Ag2	Ag3	Ag4	Ag5	Ag6	Ag7
Ag1	1.00	0.74	0.73	0.46	0.52	0.67	0.21
Ag2	0.74	1.00	0.81	0.55	0.67	0.58	0.52
Ag3	0.73	0.81	1.00	0.37	0.69	0.65	NaN
Ag4	0.46	0.55	0.37	1.00	0.58	0.69	0.54
Ag5	0.52	0.67	0.69	0.58	1.00	0.78	0.75
Ag6	0.67	0.58	0.65	0.69	0.78	1.00	NaN
Ag7	0.21	0.52	NaN	0.54	0.75	NaN	1.00

- See Emond, Edward J., and David W. Mason. "A new rank correlation coefficient with application to the consensus ranking problem." Journal of Multi-Criteria Decision Analysis 11.1 (2002): 17-28.

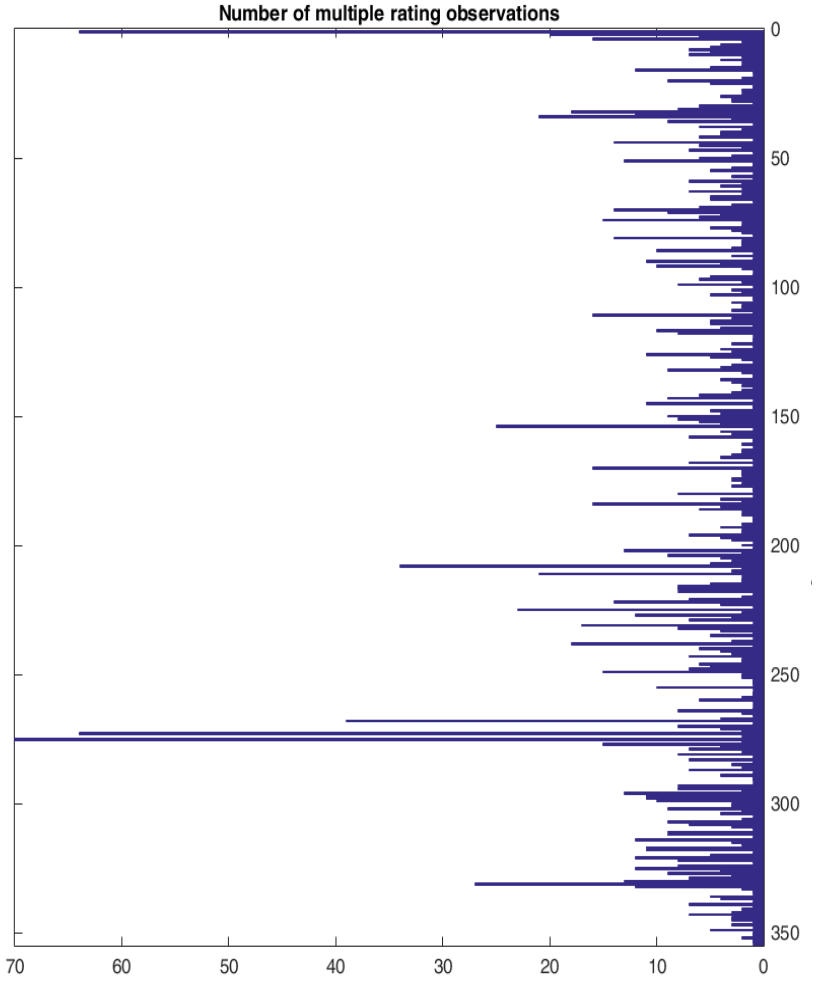
Data Structure (3rd slides)

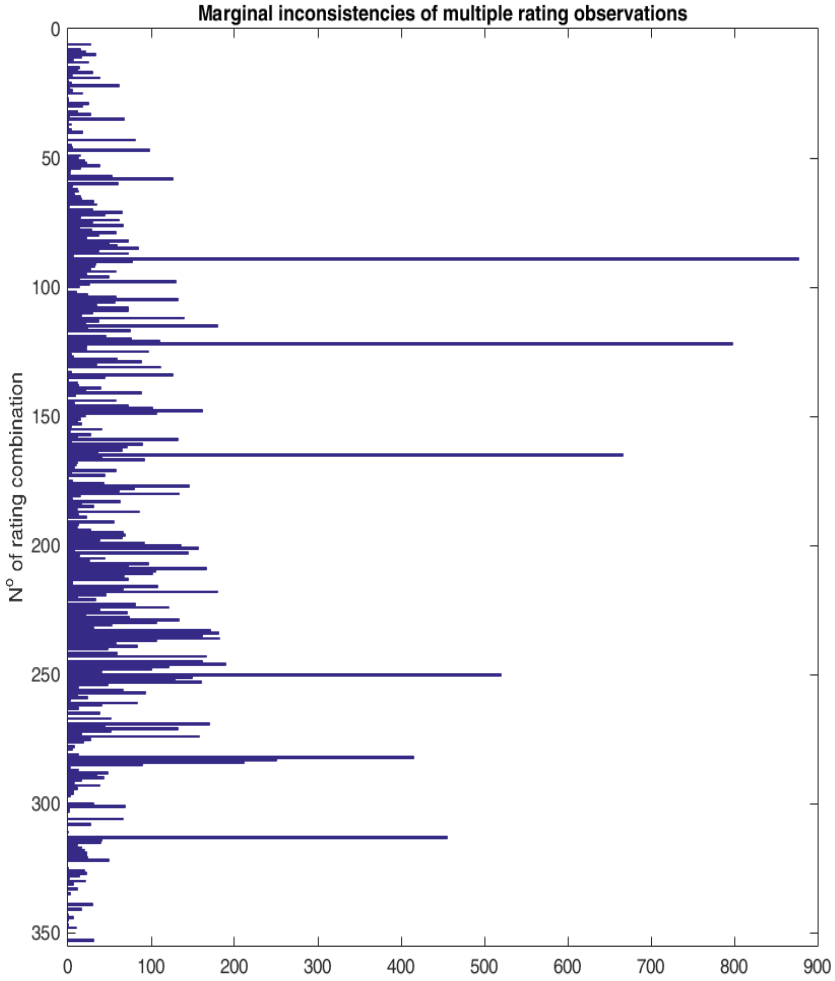
- Map of discrepancies among observed rating combination



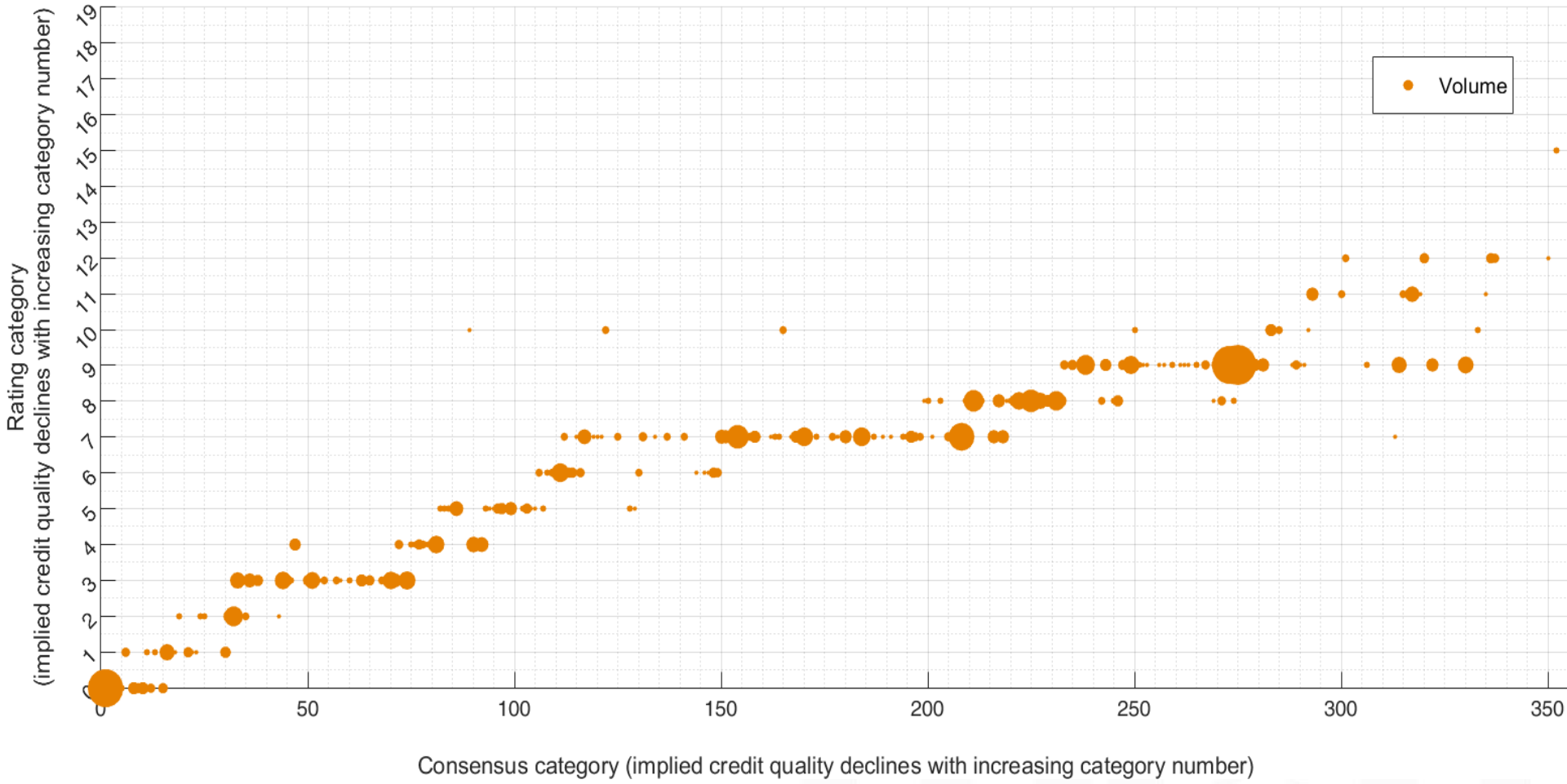
Data Structure (4th slides)

- Contributions to total data discrepancy.

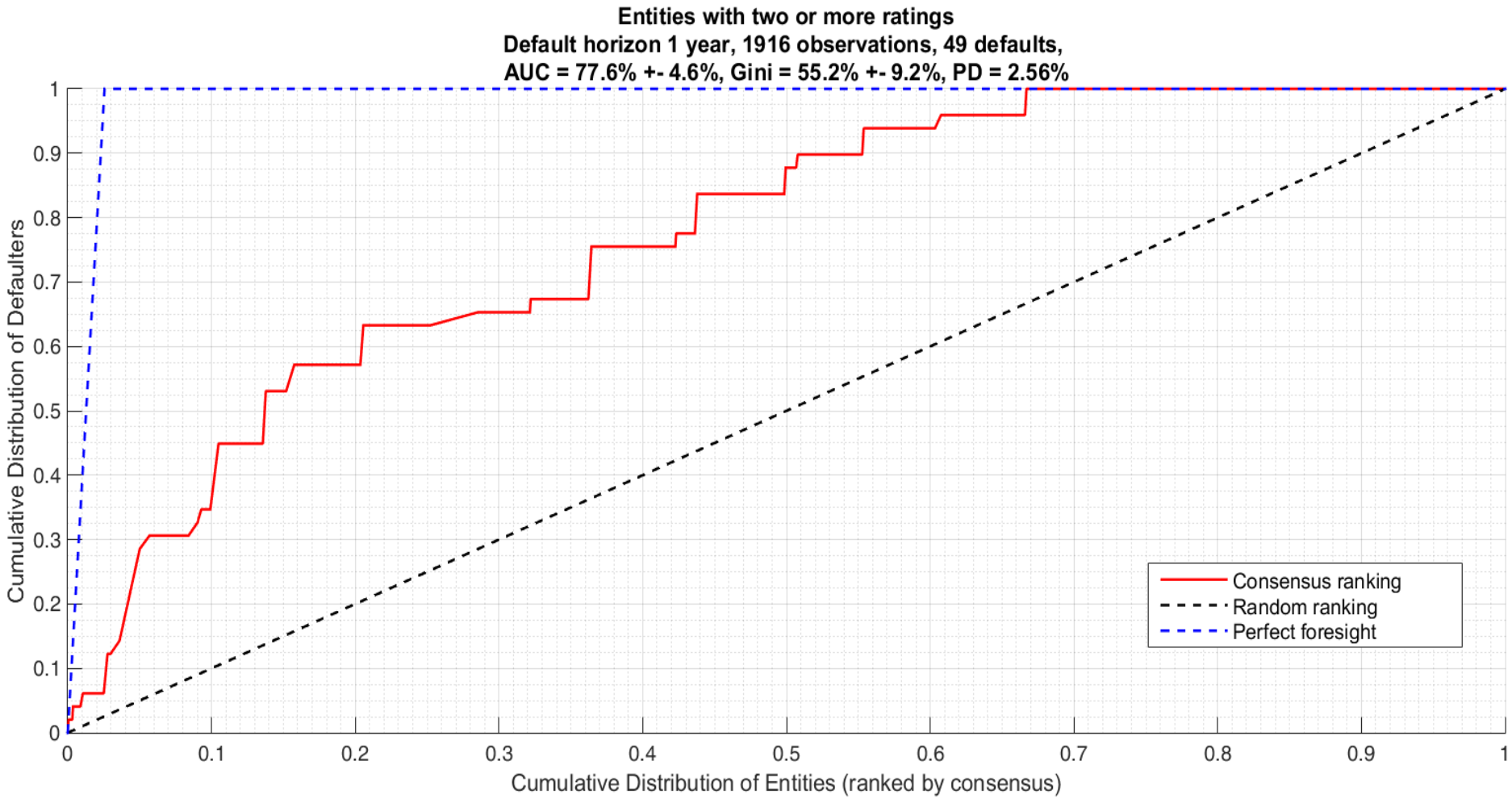




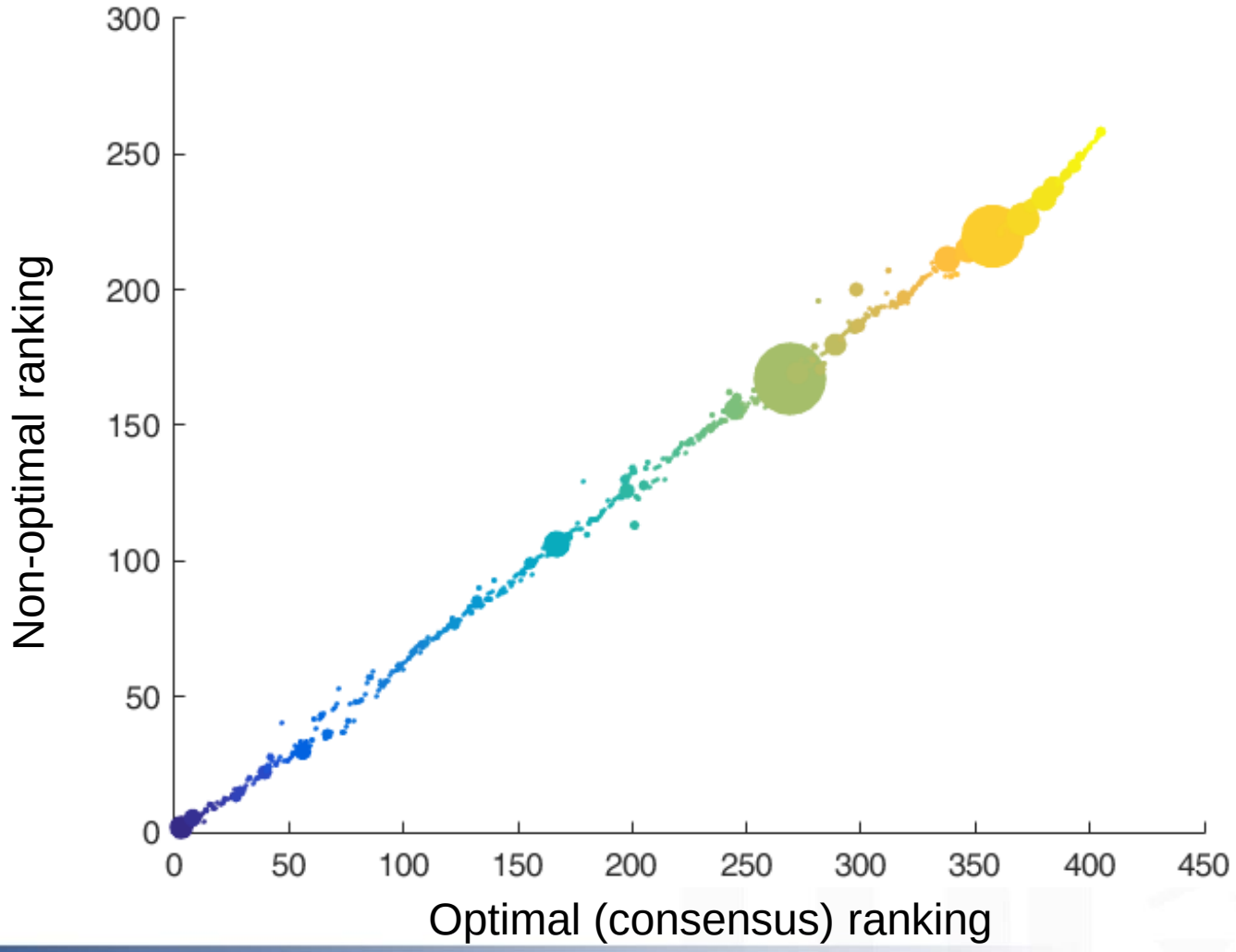
Consensus Ranking vs Ratings



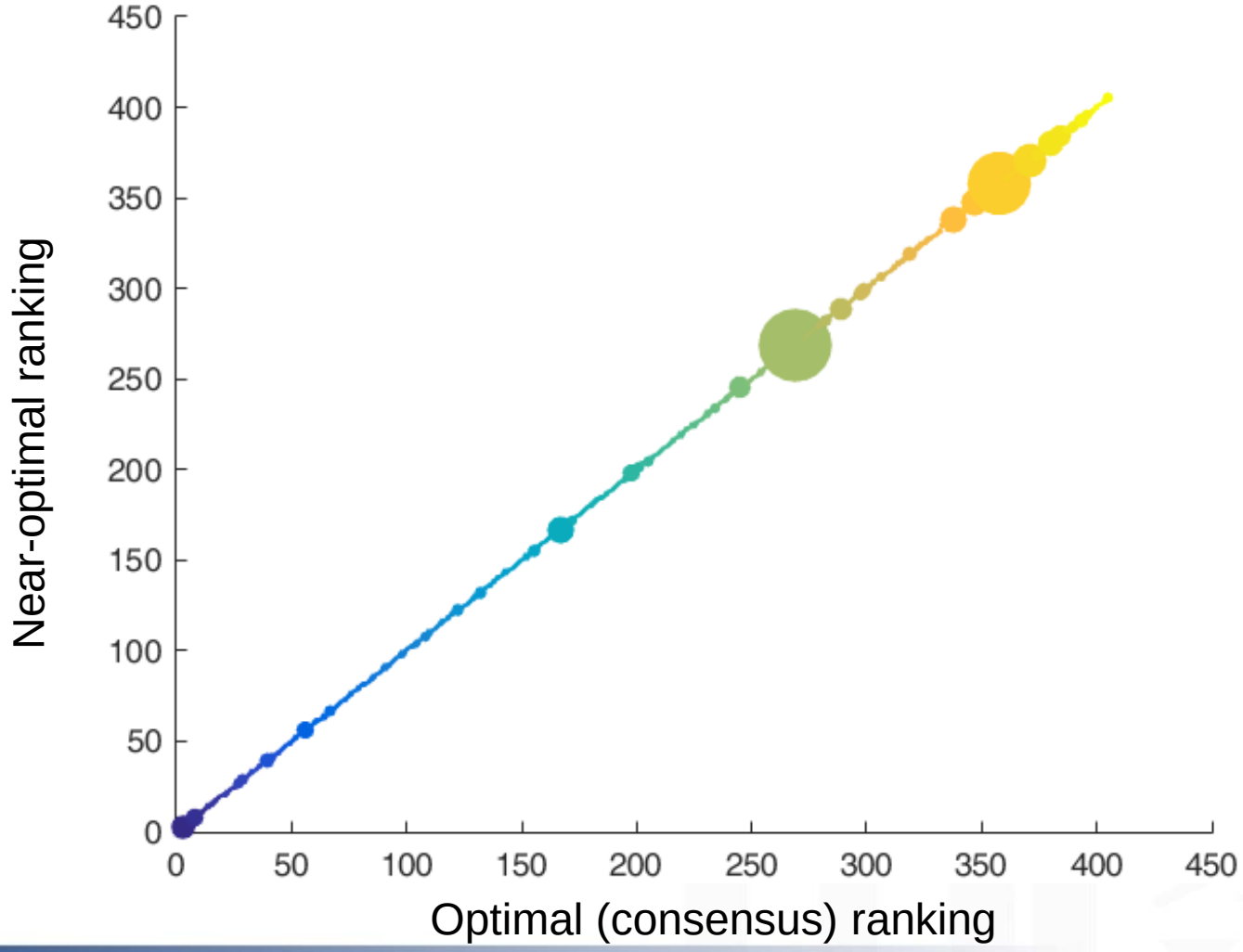
Discriminating Power of Consensus Ranking



Robustness of Algorithm (1st slide)



Robustness of Algorithm (2nd slide)



Further Research

- Incorporation of single ratings.
- Study of informational content of multiple ratings.
- Estimation of probabilities of default.

- Epic speech at PermWinterSchool’18!

NATIONAL RESEARCH UNIVERSITY

Thank you for your attention!

mkurbangalee@hse.ru
20, Myasnitskaya str., Moscow, Russia, 101000
Tel.: +7 (495) 628-8829, Fax: +7 (495) 628-7931
www.hse.ru