A Numerical Approach for finding a Consensus of Individual Credit Assessments

Victor Lapshin, PhD, FRM
(Joint work with Andew Zanochkin)
National Research University Higher School of Economics
vlapshin@hse.ru

Credit Ratings as Expert Information

- Ratings are:
  1. professional opinions;
  2. relative (!!!) rankings.

- 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.

The Data

- We code the rating scales as integers: 0 is the highest rating (e.g. AAA), 1 is the second highest (e.g. AA or AA+), etc. The last one is the default rating.
- Note that this is only an ordinal axis. The numbers themselves don’t mean anything – we only say that rating 2 is better than rating 3 and that 3 is better than 4. But we don’t say that the differences are the same.

The Data

- Consider ***m*** agencies Ag_k, $k=1..m$ issuing to ***n*** companies a_i, $i=1..n$ numerical ratings r^k_i, $k=1..m$; $i=1..n$.
- Note that no single agency has to rate all the companies. We denote `no rating’ as NR.
- Consider the rating table follows:

	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

⇒

	Ag_1	Ag_2		Ag_m
a_1	0	NR		1
a_2	1	1		0
a_3	NR	1	…	2
…	…	…		…
a_n	10	NR		18

The Data

- The problem of aggregating several orderings into a consensus ordering is relatively well-studied. It appears in the problems of aggregating expert opinions, vote results and other operations research problems.
- However, our problem has several distinctive features making it unique.
  - High dimensionality. There are usually many hundreds of companies, each observed at many time moments.
  - Sparsity. Usually each rating agency only rates a relatively small portion of all rated companies. Some companies are rated by two agencies, a few of them by 3, but very rarely more.

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.

Measuring the Discrepancy

- Let the distance between two agencies be the sum of individual discrepancies.

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

- Note that the distance is determined by both the `rated' part and the `unrated' part (proportional to w). Varying w allows us to control the effect of small rating agencies.

Consensus Ranking

- We seek the 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**. It has many nice properties, among which:
  - Satisfies some of the Arrow conditions.
  - Is neutral.
  - Respects the Condorcet majority principle.
  - Respects the Pareto principle.

Nice Properties Explained

- Arrow conditions:
  - Monotonicity: any agency adopting the consensus ranking instead of its own will not change the outcome.
  - Non-imposition: every possible outcome is feasible for some input rankings.
  - Non-dictatorship.
- Condorcet majority principle: if a company wins all pairwise aggregations, it will be at the top of the aggregate ranking.
- Pareto principle: if all agencies actually do rank A higher than B, then so does the consensus.
- Neutrality: renaming / repositioning the alternatives and the agencies does not change anything.

Drawbacks

- Its 2 major drawbacks are:
  - Non-uniqueness.
  - Computational complexity: the problem is NP-complete.
- Non-uniqueness example:

	r_1	r_2
Ag_1	1	2
Ag_2	2	1

	r_1	r_2
Ag_1	1	2
Ag_2	1	1

Another Example

- Independence of irrelevant alternatives: dropping / introducing rated companies will not change the ranking of other companies.

	r_1	r_2	r_3
Ag_1	1	2	-
Ag_2	-	1	1
Ag_3	1	-	1

1 1 1
1
0
0

or

1 2 2
0
0
1

Additional Criterion

- The non-uniqueness problem is very important.
- We may fine-tune some parameters (e.g. penalty for equal ratings).
- We should introduce some convex add-on to the functional to be minimized – J_2.
- Among all medians we seek the one with the minimum value of J_2.

Additional Criterion

- The second functional introduces a metrization. It can be done in many ways, which could result in slightly different results.
- Our choice:
  - Remap all ratings of every agency ***k*** onto [0; 1]. 0 is the best rating, 1 is default. Denote the result as $\boldsymbol{p_i^k}$.
  - Now let $J_2(R) = \sum_{k=1}^{m}\sum_{i=1}^{n}\left(p_i^k - p_i^R\right)^2$
  - This may look very crude, but remember that we only use it to select one median of the many – the best from the best.

Computational Complexity

- Finding all medians is computationally infeasible.
- There are several heuristic algorithms out there based on the branch-and-bound principle.
- We propose an approach based on the Tikhonov regularization and the genetic optimization.

Tikhonov Regularization

- We have a `bad’ functional to minimize:

$$J_1(R) = \sum_{k=1}^{m} d(R, Ag_k) \to \min_R$$

- We add to it a `good’ functional multiplied by a (small) positive number ***λ***.

$$J_2(R) = \sum_{k=1}^{m} \sum_{i=1}^{n} \left(p_i^k - p_i^R \right)^2$$

- We now minimize $J(R) = J_1(R) + \lambda J_2(R)$
- For small λ this is equivalent to our initial problem.

Computational Algorithm

- Most of the existing solutions employ a variant of the branch-and-bound strategy. It is useful in the voting problem when we have relatively few candidates. However, we have thousands of candidates.
- Literature suggests that the most effective class is that of genetic algorithms.
- We adapt a variant by Marti&Reinelt (2011) and introduce the regularization (which helps by the way).

Basic Ideas of the Algorithm

- We maintain a set (population) of solutions.
- At each iteration we throw away the worst ones and introduce new ones by mixing the elements (genes) of the surviving ones.
- The main tuning is in the mixing procedure.
  - Our mixing has two types of mutations: mixing the order in the ranking and introducing/dropping equivalence.

Output of the Algorithm

- The algorithm outputs a ranking.
  - E.g. 1, 1, 2, 3, 4.
- In effect, it groups different rating combinations, each into a new ranking category:
  - AA & a & α -> 1
  - AA & a ------> 2
  - aa & β -------> 3
  - etc.
- Sometimes it merges these groups.

Merging Ranking Groups

- Suppose the general consensus is that $r_1=r_2$.
- One agency rates $r_1>r_2$.
- Do we believe that this is valuable information?

	r_1	r_2
Ag_1	1	1
Ag_2	1	1
Ag_3	1	2

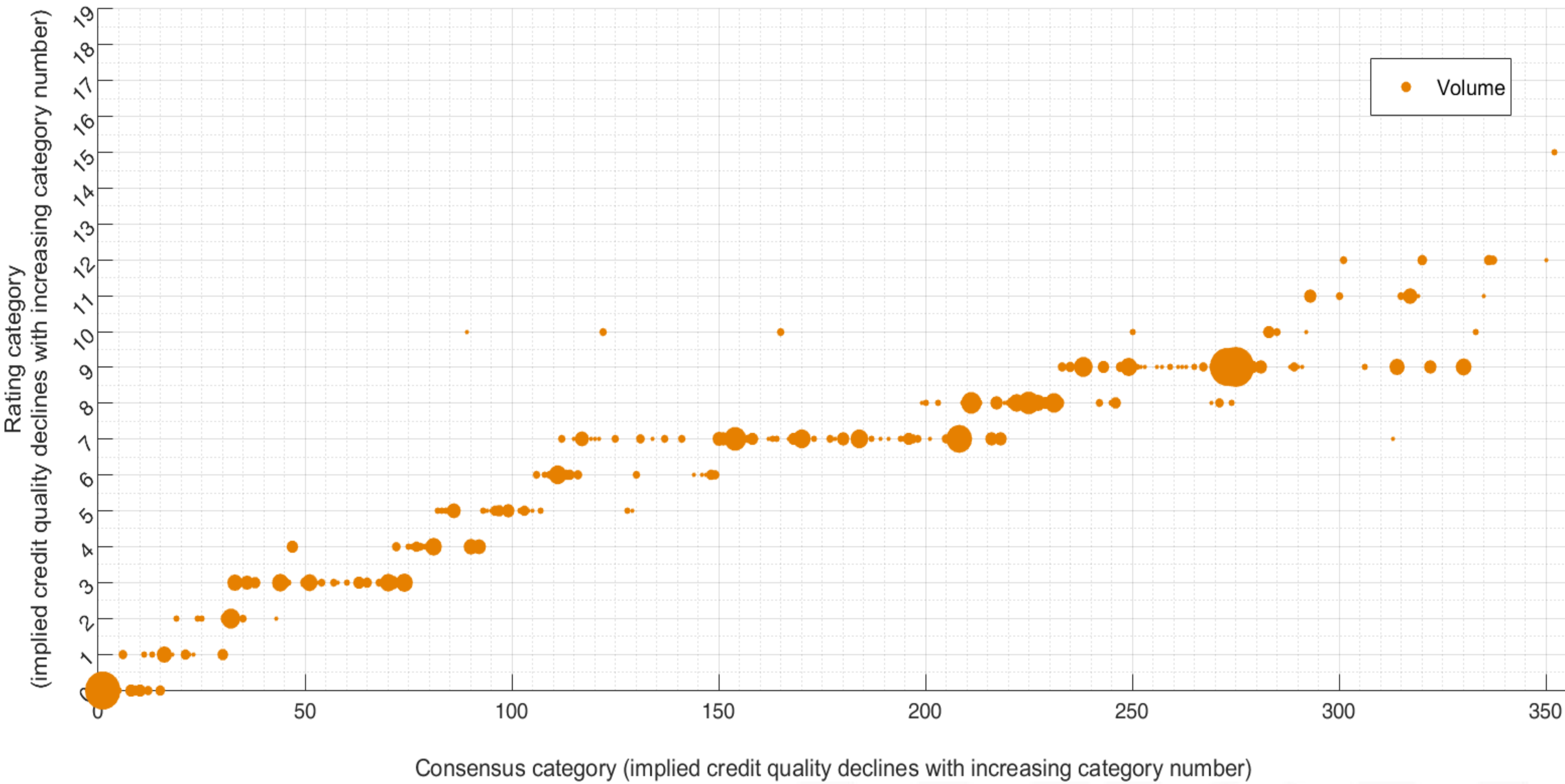
- That really depends on what we are doing and on our beliefs.
- It requires fine-tuning the discrepancy matrix or introducing a penalty for merging groups.

Data Selection

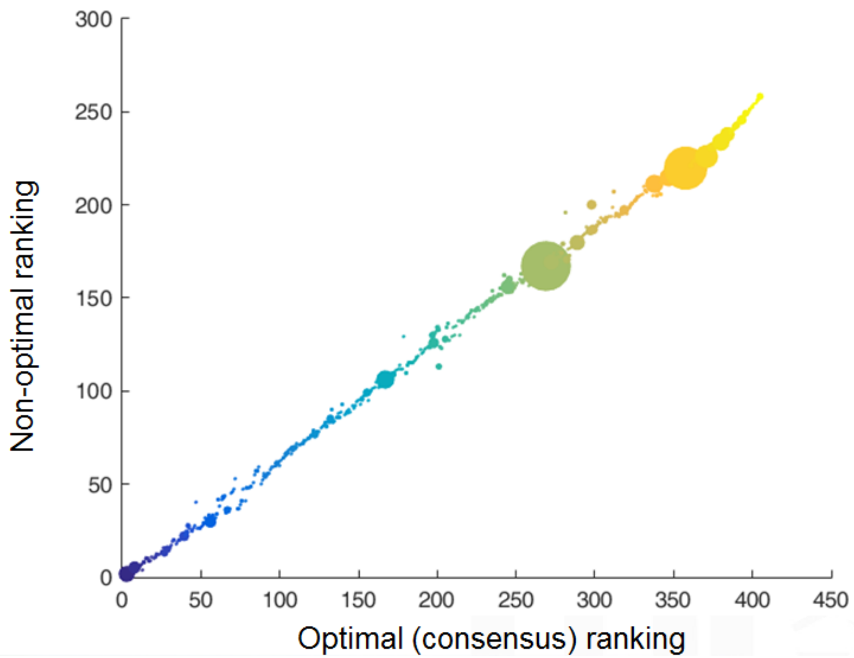
- Period:
  - July 2010 – July 2015.
- Entities:
  - Russian banks.
- Agencies:
  - 3 international agencies + 4 local agencies.
- Scales:
  - National, long-term.
- Timeframe:
  - Quarter.

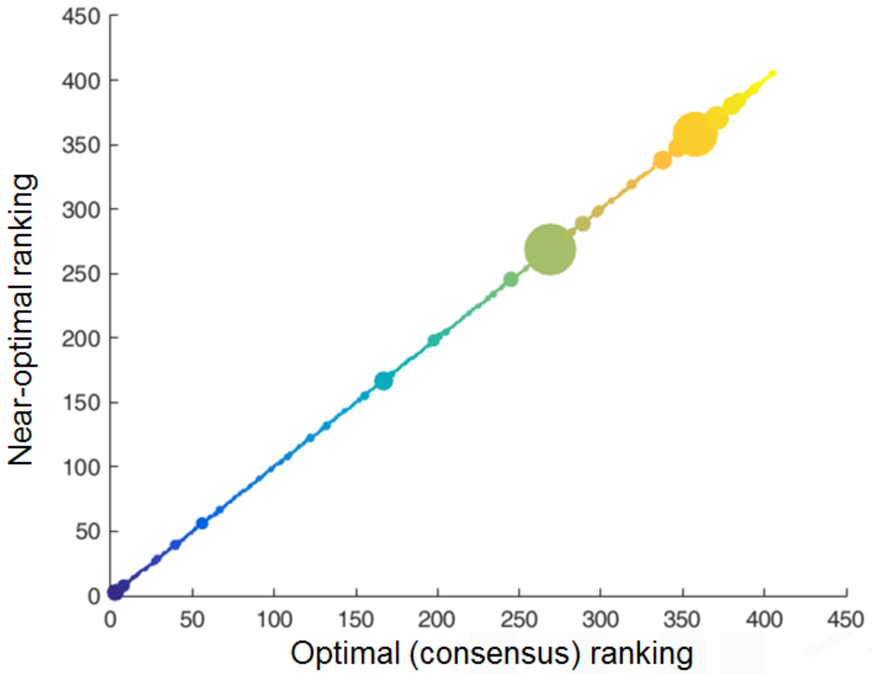
Number of agencies	Observations	% all observations
1	5604	74.5
2	1414	18.8
3	384	5.1
4	100	1.3
5	18	0.2

Consensus Ranking vs Ratings

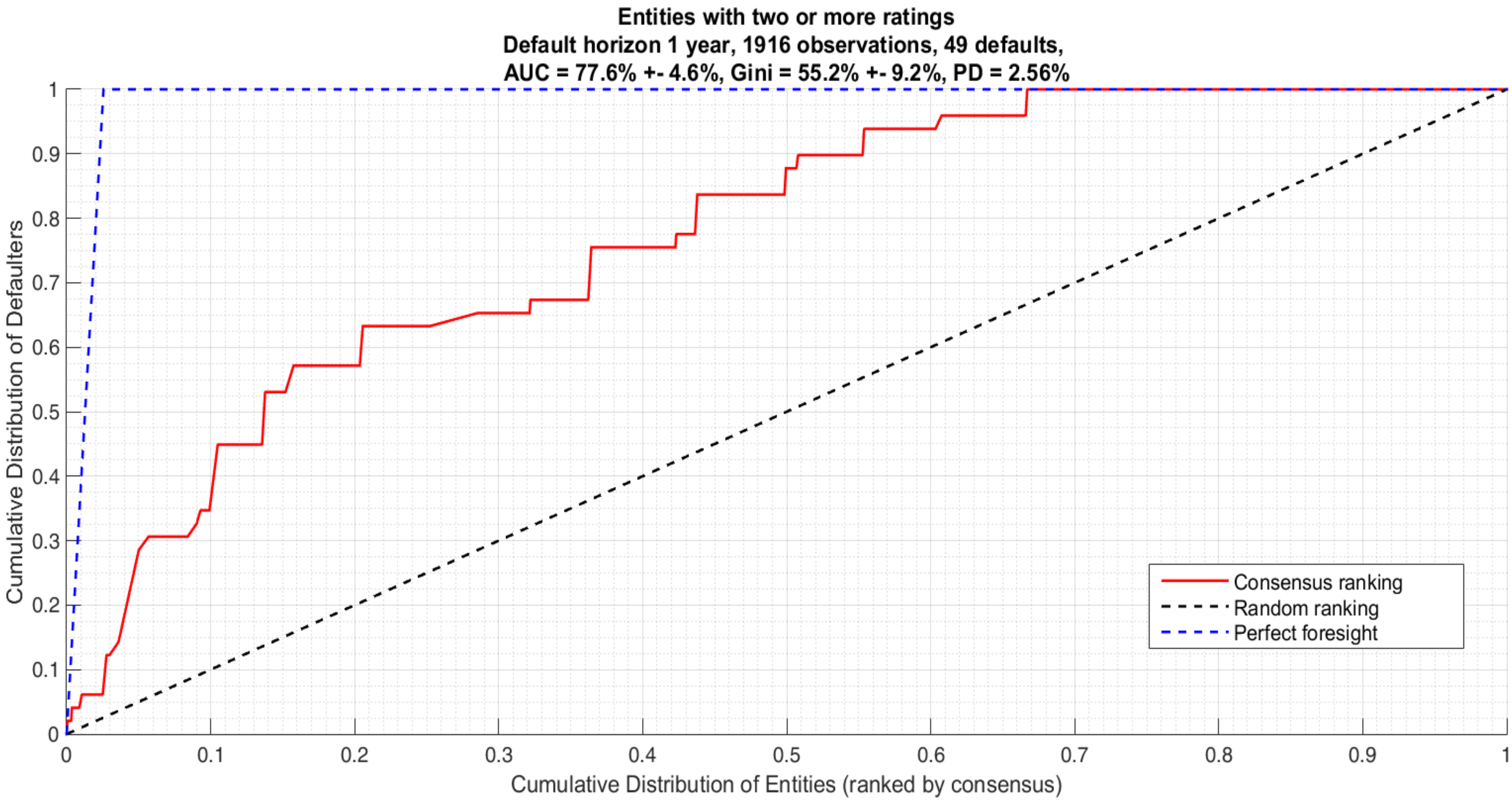


Robustness of Algorithm





Discriminating Power of Consensus Ranking



Further Research

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

Thank you for your attention!

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