Eurasian Economic Review, 3(2), 2013, 114-135
DOI: 10.14208/eer.2013.03.02.002

ARM'S LENGTH METHOD FOR COMPARING RATING SCALES†

Alexander M. Karminsky*, Richard N. Hainsworth, and Vasily M. Solodkov*****

Abstract: Investors are being encouraged after the global crisis to reduce their dependence on the largest credit rating agencies for risk assessments of companies and securities. Comparing risk assessments from different sources rapidly becomes non-trivial when more than three credit rating agencies are involved. We propose a method for comparing rating scales, and hence constructing correspondence diagrams and tables, thereby treating the rating scales used by different agencies as objects of study. Scales are compared by looking at sets of ratings assigned to similar entities (hereinafter banks) with the assumption that the risk being measured by each credit rating agency is the same for a given rated entity at a given point in time. Studying international bank ratings for a five-year period shows that there are subtle differences for the largest credit rating agencies. A mechanism for constructing mappings between scales could lead to more competition with new credit rating agencies.

Keywords: Rating, Rating Scale, Comparing, Mapping, Regulation
JEL Classification: G21, G24, G32

1. Introduction

Credit ratings have attracted significant attention recently because the ratings assigned to a whole class of securities – collateralized debt obligations – turned out to be almost wholly incorrect. Sovereign ratings, most recently of the USA and Greece, have drawn intense media attention. A great deal of criticism has been leveled at the credit rating agencies, some fair, some unfair. This paper is not about the problems; it is intended as mechanism for alleviating the causes of the problems. We believe that a combination of circumstances have led to a "shared monopoly" of the credit

† This article was written as part of project commissioned by the IFC at the initiative of the Ministry of Finance of the Russian Federation to compare rating scales. The project was carried out by a working group at the Banking Institute of the National Research University Higher School of Economics. Main results were published in the working papers 2010-2011 and at the beginning of 2012 (Hainsworth *et al*. 2012) and (Karminsky and Sosyurko, 2011). Authors are very thankful to all project participants and rating agencies involve in the research. We would also like specially to thank Anatoly Peresetsky, Vladimir Sosyurko, Alexander Vasilyuk, and two anonymous referees for methodological comments, suggestions and discussions.
* Corresponding Author: National Research University Higher School of Economics, International Laboratory of Quantitative Finance (partially supported by Russian Government, Agreement #14.A12.31.0007), Russia. Email: akarminsky@hse.ru
** President RusRating, Russian Rating Agency, CFA, Russia. Email: richard@rusrating.ru
*** National Research University Higher School of Economics, Banking Institute, Russia. Email: vsolodkov@hse.ru

ratings market by three firms, the “Big-3”[1]. It is true that the first two are much larger than Fitch Ratings, which has invested significantly to increase its market share. However, globally, the Big-3 now dominate the biggest credit rating niches.

Adding more rating agencies, and hence their rating scales, increases the complexity of assessing the credit risk of any issuer (or bond etc.). Investors have shied away from this complexity, preferring instead to rely on only two or three credit ratings. The result has contributed to the formation of a monopoly of three rating agencies, which in turn has led to the weaknesses of monopolies in the financial markets, and in part to the global crisis starting in 2007.

Suppose a large investment bank sets up a skilled research team that is able to assess the risk of its various counter-parties. How would this research team express its risk analysis, except as another rating? So how do the ratings of the investment bank's internal research department compare with the ratings of other credit rating agencies? The management of the investment bank, together with its risk managers, dealers and regulators, will want to know how these internal ratings compare with commercially available ratings. In a market with multiple rating agencies, each with their own rating symbols and rating methodologies, the most difficult problem faced by an investor is to compare ratings. The problem of comparing rating scales is therefore of some importance for the whole area of risk management.

We start this analysis by making the assumption that each symbol including plus or minus signs marks a point along a scale of credit risk. When an investor is comparing ratings, they are comparing symbols along a scale.

The approach described here, for clarity we call it the Distance method, relies on the assumption that the credit (financial) risk of a single entity is the same when measured by different rating agencies. When two agencies measure the credit risk of a single entity, they may assign different rating symbols.

Using the analogy of temperature, we have a situation in which the temperatures of a number of different bodies are measured using a variety of thermometers, each of which has different scales. By assuming that the temperature of each body is the same, the difference between the scales is a function of the measurement technology. The mathematical task is to find a relationship between the scales.

Since the ratings assigned to a single entity are themselves opinions and there is some uncertainty as to the accuracy of the measurement, we have to use statistical techniques to average out the uncertainties, as well as to find a mapping between the different scales.

The underlying assumption – that each rating agency is measuring the same thing, namely, credit risk when assigning a rating to an entity – is

[1] We shall refer to the three largest US rating agencies: S&P, Moody's Investor Services, and Fitch Ratings as the Big-3.

not as obvious as it might seem. Even though all the rating agencies speak of their ratings in terms of market parameters such as default risk or recoverability, when the detail of these claims are investigated, considerable differences can arise. What, for example, is a default? Suppose an issuer is delinquent by a small period of time due to technical problems or short-term stress, and then pays in full, or even compensates investors for the interruption? What is the credit rating of a company whose debt has been restructured? What happens when an issuer defaults to one set of creditors (eg. foreign bond holders), whilst fulfilling obligations to domestic creditors? Alternatively, one set of obligations is no longer met (payments by a bank on it bonds), whilst another set of obligations continues to be serviced (servicing customer deposit accounts); is this a complete or partial default?

Some rating agencies interpret their ratings as reflecting probability of default, whilst other rating agencies interpret them as recoverability. More generally, ratings are measures of the financial risk an investor takes on when investing in a security. Indeed as ratings are used more widely to include equities and exotic instruments such as catastrophe bonds, the link with a default on a credit obligation becomes more tenuous. This has led to recommendations that different symbol sets are used to distinguish between the ratings associated with different types of risk.

It could therefore be argued that in fact even when credit rating agencies are assigning a credit rating to a single entity, they are in fact measuring slightly different things and this explains any perceived difference between the ratings they assign to that entity. On some theoretical level, this may be an argument in favor of the idea that credit rating agency scales are fundamentally incompatible. It is not an argument that sits comfortably with practitioners. Investors and regulators are concerned about the risk of economic loss, a loss in the value of their investment, a return below what they expected. The value rating agencies bring is in assessing financial risk for a fairly wide definition of risk, and that risk is perceived by the market to be the “same thing”. So we shall assert that if two rating agencies assign ratings to the same entity, then they are measuring the same thing.

Another issue commonly raised by critics to this study is that ratings are measures of the probability of default. Consequently, harmonization of credit rating scales should occur by defining default probabilities for each gradation and then determine whether the credit rating agencies are accurate in their forecasts. Alternatively, the probability of default should be estimated for each gradation for each agency and then the credit rating scales can be mapped onto each other via the probabilities of default.

Firstly, default probabilities depend on both the company being rated and on the environment it operates in. More companies default in times of economic stress than in times of boom. Secondly, estimating unambiguously a probability of default ex ante for a company is not possible. If it were, then any number of information providers would already be doing so. All of the techniques used in the literature for estimating closeness to default of a company are used by the rating agencies in their methodologies. If some researcher were to determine a new technique for estimating the probability of default and to commercialize the product, he or she would be creating another credit rating agency.

Finally, there is the issue of ex ante credit ratings and ex post default frequencies. We only know about a default once it has taken place. It is then possible to match the default with some credit rating issued earlier. Default frequencies can only therefore be calculated after the defaults – ex post. Investors want information about financial risk when they are buying, perhaps several years before an eventual default. So credit ratings are assigned before any defaults – ex ante.

In order to obtain a statistically significant number of defaults for each gradation takes a considerable period of time, certainly longer than business cycles in the economy. At the same time, rating agencies may change their ratings much more quickly, especially in transition economies. Whilst a mapping exercise based on ex post default frequencies is of interest, it was rejected for our purposes as being impractical.

The research covered in this paper has been driven by the development of credit ratings in the Russian market. There are now seven rating agencies accredited by the Russian Ministry of Finance[2] – four local agencies and the Big-3[3]

Credit ratings have entered into Russian legislation in 2012. When threshold levels are defined for a regulatory purpose, the names of the Big-3 are included in the legislation and the value of the corresponding rating. This practice is, however, considered improper as it confers a legislative preference for the certain proprietary products, such as specific ratings. Consequently, there has been a perceived need to create a mechanism to refer to ratings without specifying either the company name or the companies' specific product name.

More generally, regulators in all countries need to have a single measure for risk, so that they can use the scales for regulatory purposes, and not have a different scale for each company.

Competitively, the dominance of the Big-3 places any newcomer in a difficult position. The single most common question a small rating agency is asked is how does your XYZ rating compare with the MNO rating of one of the Big-3. In other words, investors are seeking a common comparison scale. It is the very same logic that led in primitive economies to the identification of a single commodity as money.

Initiated by domestic rating agencies and the Ministry of Finance Of Russia, a systematic statistical study has been conducted to compare the rating scales of all the rating agencies, which lead to the results below (the Distance method). In comparison to the distance method which used in this paper econometrician approach was used by Aivazian *et al*. (2011) and

[2] Starting from September 2013 regulation is done by Russian Central Bank.

[3] Moody's also owns a domestic agency (Moody's-Interfax), but this is accepted to be operated in coordination with Moody's and so is not considered separately).

alternative non-parametric approach was proposed by Eisl *et al*. (2013).

2. Literature Survey

Rating changes have a significant influence on the purchase and sale of both fixed income and equity risk. The regulatory role of ratings began to grow from the 1970's (Altman and Saunders, 1997; Cantor and Packer, 1995; Partnoy, 2002; Karminsky and Peresetsky, 2009; Aleskerov *et al*. 2010).

The Credit Rating Agencies have given considerable attention to improving their methodologies, especially in the last decade, and have published the rating principles regularly. At the same time, these publications have not included any detailed information, relying instead on descriptions of general principles and the particular features of their approach to ratings. A substantial proportion of the methodologies are based on the expert opinions of analysts.

It is not just the differences between methodologies that are of interest (Altman and Rijken, 2004; Cantor and Packer, 1995), but also the rating process (some details are given below as this is refers to the Russian market) and a comparison of the rating differences with publicly available information (Morgan, 2002; Iannotta, 2006).

Procyclicality has been noted (Altman and Rijken, 2004; Pederzoli and Torricelli, 2005). The through-the-cycle approach should increase the stability of ratings and prevent changes due to short-term fluctuations. At the same time, the approach does not allow agencies to react in a timely manner to significant events. The recent bankruptcy of a series of large companies and banks has raised the question of a review of these methodologies (Servigny and Renault, 2004).

Some credit rating agencies have been adapting their methodologies during the recession. For example, Amato and Furfine (2004) have shown that S&P have not been taking into account business cyclicality in relation to U.S. corporates.

A number of papers have been devoted to the modeling of bank ratings (Caporale *et al.* 2010; Iannotta, 2006; Morgan, 2002; Pagratis and Stringva, 2009; Peresetsky and Karminsky 2008; 2011; Karminsky and Sosyurko, 2010). These papers have defined typical explanatory variables, use models of ordered choice and then examine the process of changes in rating gradation over time for a limited sample of Moody's ratings. At the last paper was demonstrated existence of the statistically dependence difference between Big-3 rating agencies estimations.

The difference between ratings for banks and ratings for corporates has also been studied comparing ratings from different agencies (Iannotta, 2006; Morgan, 2002), and in particular the factors which lead to differences between the largest rating agencies. A similar set of studies focuses on cross-country ratings of banks and corporates (Caporale *et al*. 2010; Ferri *et al*. 2001).

One of the basic problems facing the utilization of credit ratings by regulatory bodies and commercial companies is the comparability of the ratings from different agencies. How can a relationship be established between the positions on a rating scale when there are different levels of defaults and expected loss? Another question is how to account for changes in ratings due to arbitrage when there are systematic differences in ratings, the desire of issuers to obtain the best ratings (rating shopping), and slippage as rating agencies become more accommodating to gain and retain clients. These problems have been cited as reasons underlying the difficulty applying ratings for regulatory use (Cantor and Packer, 1994; Basel, 2010, Karminsky and Solodkov, 2010).

One of the first papers to compare the ratings of many agencies was Beattie and Searle, 1992. A large sample of long-term ratings was gathered from twelve large international credit rating agencies and over 5000 pairwise differences were found between the ratings of the same issuer by different credit rating agencies. The number of rating pairs in which the ratings from two agencies of the same issuer constituted under half the sample, whilst around 20% of the pairs involved differences in excess of two gradations. The differences may be explained as due to differing opinions about the financial stability of the issuers, differing methodologies used by the rating agencies, or differing choices of the qualitative indicators by the rating agencies.

The fundamental question is whether the observed rating distinctions reflect a systematic difference between the scales of the rating agencies (Basel, 2004). The largest numbers of rating pairs in the sample are provided by S&P and Moody's. The average difference between their ratings over 1398 observations is 0.05 of a gradation.

At the same time, when the ratings of eight rating agencies were compared to those of Moody's, the ratings of five agencies were found to be significantly higher (Beattie and Searle, 1992). Ederington (1986) concluded that no systematic difference could be found between the ratings of S&P and Moody's, whilst Morgan (2002) showed that the rating difference was greater, the less transparent the issues, and that this difference was greatest for banks and financial institutions.

Cantor and Packer (1994) studied ratings assigned by four credit ratings agencies in the USA to speculative (junk) bonds in 1989-1993. The authors found that although the ratings of S&P and Moody's were close on average, the ratings of the third or fourth largest rating agencies were more frequently and more strongly differentiated from the Moody's ratings than was shown in Beattie and Searle, 1992. The ratings of third and fourth agencies differed on average from the S&P and Moody's ratings by more than 1.5 graduations.

Cantor and Packer (1994) also compared the Moody's ratings of the international banks with the ratings of nine other rating agencies. Again it was found that the differences were greater on average than the findings in (Beattie and Searle, 1992). For example, the average rating difference (from Moody's) for the three Japanese rating agencies was as much as three gradations.

Guttler and Whrenburg (2007), looked at rating differences and at the adaptation of the rating by one credit rating agency to a change in the rating of the same subject by another credit rating agency, for cases when the issuer is close to default and there is a rating by both S&P and Moody's. The main result is that an increase (decrease) in rating by one rating agency is probable in a short period of time after an increase (decrease) in rating by another credit rating agency. Moreover, the greater the change in rating by the first agency increases the probability of a change in the same direction by another agency.

A number of authors have looked at possible conflicts of interest. One hypothesis is that rating agencies may depress the ratings of companies who have not paid for the ratings in order to pressure them into taking a paid rating (Partnoy, 2002). However, surveys of requested (paid) and unrequested (unpaid) ratings have not uncovered any evidence for this hypothesis. Poon (2003) examined 265 ratings assigned by S&P in 15 countries between 1998 and 2000. He concluded that unrequested ratings were lower on average than requested ratings, although he also found that the effect can be explained as self-selection that is only companies that are certain of their financial condition request to have a rating assigned.

The credit rating agencies also explain the effect in terms of a conservative approach when dealing with an unrequested rating because they do not have as full information about a company with whom they have a rating contract as they would with a company that has entered into a rating agreement. Roy (2006) studied the ratings assigned by Fitch Ratings to Asian banks in 2004 on a requested and unrequested basis. On one hand, the author concluded that the Agency had approximately the same approach to the assignments of ratings because in models of the ratings assigned to the two classes of banks (viz., requesting ratings and not requesting ratings) because the weights associated with the financial indicators of the banks were the same. At the same time, the ratings of banks not requesting a rating were on average 0.9 gradations lower than ratings of banks requesting ratings.

Cantor and Packer (1996) looked for a material selective rating differentiation using the approach developed by Heckman (1979). Cantor and Packer discovered a selective differentiation is improbable and that a large number of non-coincident ratings are due to differences between the rating scales. Prior to 2000, Fitch Ratings only assigned ratings when requested. White (2002) believes that issuers approach Fitch Ratings when the ratings they have received from the two largest agencies do not coincide with the hope of getting a better rating.

It is well-known that an issuer's rating affects its cost of borrowing. When the ratings of Moody's and S&P do not coincide, both ratings affect the yield on the bond and a better prediction of the yield is obtained from the average of the two ratings (Cantor *et al*. 1997). When a rating has been assigned by a third agency, the credit quality of the borrower is perceived by the regulator to be better (Cantor and Packer, 1996). Jewell and

Livingston (1998) studied the behavior of 235 bonds that had ratings assigned by Fitch Ratings, Moody's and S&P during the period from January 1991 to March 1995. The study showed that the cost of borrowing was reduced when the issuer requested a rating from a third rating agency, especially if the rating was higher.

A variety of studies have looked for differences between ratings from different agencies and then analyzing the reasons for these differentiations rather than constructing a mapping between the different scales. Liss and Fons (2006) compared the national rating scales supported by Moody's with its global rating scale. Ratings have been compared in Russia. Matovnikov (2008) looked at the relationship between the gradations of rating scales and the total assets and capital of banks. He demonstrated that such an approach is limited suggested using assessments of creditworthiness (e.g., the methodology of the Central Bank of Russia) given the paucity of data on bank defaults. Hainsworth (2009) compared the rating scales of the Big-3 and the Russian credit rating agencies (RusRating, Expert RA, AK&M and NRA).

Hainsworth used an iterative application of linear regressions to find mappings between the rating scales of all the credit rating agencies. Smirnov and Sholomitsky (2010) attempted to compare the rating scales of the credit rating agencies using estimates of default probability. Even though integral estimates of default probabilities were constructed for each of the four Russian credit rating agencies, it was found that the ratings in each of the gradations could not be statistically distinguished due to the paucity of the statistical data.

A group at the Higher School of Economics in Moscow and the New Economic School has been working on modeling the ratings of the international credit rating agencies in Russia (Peresetsky *et al.* 2004; Karminsky *et al*. 2005, 2006; Karminsky and Peresetsky, 2007; Koshelyuk, 2008). These studies have focused on finding economic and financial factors that affect ratings and on comparing the ratings of different agencies. The difference between them was covered at Karminsky and Sosyurko (2010) and Karminsky (2010).

Although the research was conducted for the Russian market, and some results are given for that market, the comparison of rating scales is actually of wider interest. In our studies (Karminsky and Sosyurko, 2011; Karminsky and Solodkov, 2010; Hainsworth *et al.* 2012) we assumed the rating scales are not the same and that the ratings associated with the different credit rating agencies are correct assessments of credit risk within the context of the credit rating agency making the assignment.

In their study, Eisl *et al*. (2013) used a similar assumption but employed a different technical approach. They used non-parametric methods for the re-mapping of rating scales of Big-3 agencies for the companies from the seven biggest countries (Big-7). The method is based on estimating the relation of rating scales for pairs of raters, based on ordinal co-ratings and correspondence of the structure of all rating classes' relations from a pair of raters. The ideas were proposed and incorporated as rating-class specific re-mapping of one agency's ratings to another's scale.

Methodological difference between the papers is based on slightly different approaches. While Eisl *et al*. (2013) uses classes of ratings, non-parametric approach and direct implementation of PD statistics, our approach is based on more specified ratings derived from grades parametric approximation of maps and using PD statistics only for choosing the type of functions for parameterization.

At the same time in both papers were found the difference in the scales for the three major rating agencies Fitch, Moody's and Standard & Poor's as well as the relations of their rating grades (classes) and discussed the importance of scale correction for benchmarking.

In the years 2012-13 several papers were published covering the closely related problems. For example, Bongaerts *et al*. (2012) discusses the presence of multiple rating agencies. Number of publications (Becker *et al*. 2011; Xia, 2012) touched upon the competition among rating agencies. The role of quantitative analyses in rating process was discussed in Griffin and Tang (2012) as well as credit rating across asset classes was discussed in Cornaggia and Cornaggia (2013).

3. Data Description

In order to avoid problems that might arise because different types of company are being compared, we only used bank ratings data.

The data set used in this study is described in Karminsky and Sosyurko (2010) and contains 3,639 pairwise observations. The data base contains ratings from 290 banks from more than 80 countries, each of which has at least two ratings. For the purposes of this study, however, only the ratings for the years 2006 to 2010 inclusively were used.

The rating symbol is converted into an integer, with the first symbol (lowest risk) assigned to unity, the second to two, and so on. The Big-3 have symbol sets with approximately 25 symbols in the rating scale by grades. The agencies all handle defaults or close defaults in different ways. Moody's, for example, does not issue a default rating, whilst both Fitch and S&P distinguish partial or selective defaults from defaults. These are treated as different ratings. In any case, the number of entities with very low ratings (default or close equivalents) is very small and not statistically significant.

The data is available for a number of cross-sections, such as by year, by country, by type of bank etc. We assume that each agency has confirmed the rating at least once during each year and hence that each rating is a confirmation of the application of the rating agency's methodology. Even though the entity is the same year from year, the assignment each year is assumed to be an independent assignment (Karminsky, 2012).

4. Description of Distance Method

4.1. Overview

Overall, the problem is posed as finding an extremum, i.e., minimizing the distance between the ratings (in our research sum of square of the differences). The data is manipulated to fit into a form of a single equation described below. Each rated entity (bank) in the data set may have one, two or three ratings. Entities with a single rating do not allow for any comparison, and so are eliminated. Each set of entities is split into pairs, thus allowing for a maximum of three pairs for each entity (assuming ratings are assigned by all three agencies). The pairs for each period (year or quarter) are assumed to be independent.

A mapping relating all three rating scales is sought by minimizing the distance between the ratings. If the rating scale of one agency is selected as the base agency, then the minimization problem is easy to specify. At the same time, by choosing one agency (in this paper Moody's, as it has the largest number of observations in Russia) requires the elimination of all pairs not including a Moody's rating. Thus if any entity was rated by S&P and Fitch, but not by Moody's, that entity would be discarded. The robustness of basis scale chosen was analyzed.

The mathematics of finding a mapping between the agencies by minimizing the distance between the ratings also runs into problems when one rating scale is not kept stable. Essentially, a global minimum exists when all the rating scales are transformed so that they are all concentrated into a single point. The problem is akin to the situation with money when it is not linked to a single commodity (such as gold), leading to inflation.

4.2. Detailed Description

Each rating is converted into a numerical scale and then a functional transformation $F_i(\alpha_i)$ is applied to map the rating onto the base scale (Figure 1).

We consider N rating agencies, each of which has a rating scale consisting of an ordered set of symbols. Hence we have N symbolic rating scales, $RS_i, i = 1,\ldots,N$, each of which has its own mapping to a sequence of natural numbers $NS_i, i = 1,\ldots,N$. Once a numerical value is associated with a rating, the numerical values are treated as real numbers, which can be manipulated as part of a continuum. The aim is to find a mapping from the numerical rating scale NS_i to a single base scale *BS*.

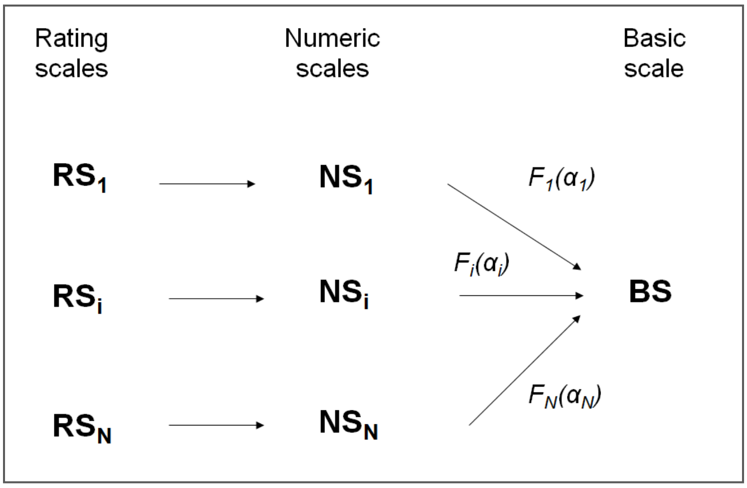

Figure 1. The rating scales mapping scheme

Once we have mappings from the original symbol scales RS_i to the base scale *BS*, the inverse mappings from the numerical base scale to symbols is relatively straight forward. The only difficulty is to transfer numerical characters into regular rating scales to build up correspondence tables.

In addition to the N rating scales, we also have a set of rated entities $A_j = 1,\ldots,K$. Each rating corresponds to a moment in time $t = 1,\ldots,T$, thus leading to a set of assessments of ratings R_{ijt} by a rating agency on scale i of subject j at moment t.

Since the set of ratings is not agreed upon, they can be used as competing assessments. The problem is to find mappings of the rating scales $F_i(RS_i, \alpha_i)$ such that the set of mappings $F_i, i = 1,\ldots,N$ minimizes the integral metric of proximity between pairwise assessments for the same rated entity.

If we denote the proximity (distance) metric between the mappings of two ratings scales i_1 and i_2 using mappings F_{i1} and F_{i2} , respectively, for the same rated entity j at the same point in time t as

$$\mu_{i1i2jt} = \mu(F_{i1}(R_{i1jt}, \alpha_{i1}), F_{i2}(R_{i2jt}, \alpha_{i2})) \quad (1)$$

that is, a proximity metric along a number axis characterizing the base scale *BS*, then the task can be specified as finding mappings F_i and parameters α_i ; *i*=1, …, *N* such that the integral proximity metric is minimized.

Although various distances along the number axis can be used as the proximity of the mappings, we shall use the common root of the sum of the squares all pairwise distances and minimizing the quadratic form. It has computational advantages, generic solution and statistical software is available. Some other variants including information, power and polynomial measures was tested during the research and they give approximately the same robust results. So our decision about Euclidian approach is robust and practically useful and generally accepted.

Thus we have as measure *S* between the mappings' system $\{F_i(\alpha_i), i = 1,..., N\}$

$$S = \sum \mu_{\mathrm{i1i2}jt}{}^2 = \sum (F_{\mathrm{i1}}(R_{\mathrm{i1}jt}, \alpha_{i1}) - F_{\mathrm{i2}}(R_{\mathrm{i2}jt}, \alpha_{i2})))^2. \qquad (2)$$

The summation is over all pairs of ratings (i_1, i_2) for each subject j at time t. We denote the set of such combinations as Q.

The minimization problem is thus

$$\min_{\{\alpha_i, i=1,..,N\}} \sum_Q (F_{i1}(R_{i1jt}, \alpha_{i1}) - F_{i2}(R_{i2jt}, \alpha_{i2}))^2 \qquad (3)$$

and Q is the set { No of quarter t, No of bank j, Rating of the base agency R_{i1jt} for bank j, Rating of another agency R_{i2jt} for bank j}.

We consider some types of approximations F_i, especially linear, power or logarithmic functions such as

$$F_i = a_i * f_i + b_i, \qquad (4)$$

where a_i and b_i are the approximation parameters.

Additional comparing analysis of default statistics for Moody's and S&P and calculation of mappings gives us an opportunity to use a priority logarithmic approximation, which we use in this paper for empirical analysis. When we were restricted the list of rating agencies only Big-3, the linear approximation of mapping function also was useful and give a little better result as will be discussed later.

It also must be mentioned that for previous problem we may use econometrical software packages such as EViews or STATA (we use EViews for calculation) because of using the quadratic criteria. Here is additional argument for using quadratic measure.

5. Results and Commentary

Using rating data for 2006 to 2010 on a quarterly basis data for international banks, which were described, above were defined mapping parameters which are presented at the Table 1. The mapping functions were characterized by the approximation parameters described upper (4). The parameters are presented at the Table for S&P and Fitch using linear and logarithmic functions as approximation of their mapping to Moody's scale as basic and all significant at the 1% level.

These transforms can be depicted graphically using the base

(Moody's) rating scale as the independent axis, and mapping where natural integers (ratings) resulting from the inverse function lie in relation to the independent axis. Figure 2 contains the image for the three rating scales.

Table 1. Scale transformation parameters that minimize proximity metric

	Function Approximation			
Agency	Linear Approximation		Log Approximation	
	a	b	a	b
S&P	0.83	1.017	1.112	0.364
Fitch Ratings	1.024	0.354	1.05	0.154
Observations	10145		10145	
Pseudo R^2	87%		82%	

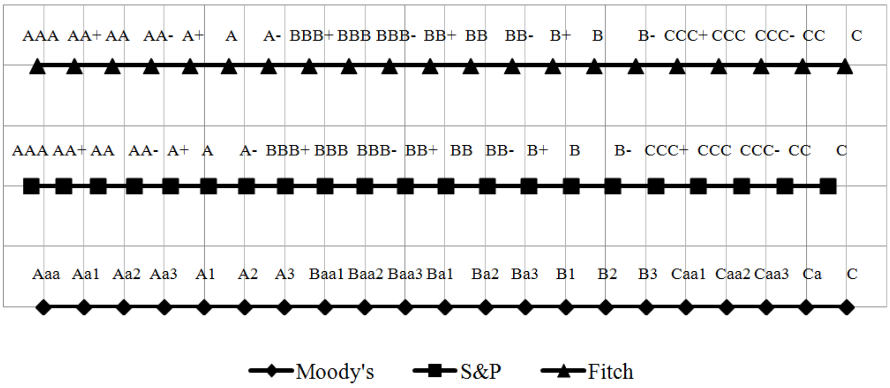

Figure 2. Comparison of Big-3 scales using distance method for logarithmic mapping

This comparison demonstrates quite clearly that the rating scales of the Big-3 credit ratings agencies do not have a simple 1:1 mapping, which would be implied by associating all the first symbols in each scale, then all the second symbols, etc. Thus the AAA rating of S&P is associated with AAA of Fitch and Aaa of Moody's, and then AA+ (S&P) with AA+ (Fitch) and Aa1 (Moody's). The existence of lags between agencies estimations have been shown in Karminsky and Sosyurko (2010).

Suppose, we assume however that the intent of the Big-3 credit rating agencies is in fact to equate their symbols as defined above. Then this statistical analysis shows that S&P is more "conservative" in its ratings than Fitch, which is more "conservative" than Moody's. Indeed for a portion of the risk scale (starting from A+/Aa3), S&P and Moody's have almost identical distances between ratings, but S&P are nearly one notch more conservative.

6. Distance Method Application to Russian Rating Agencies

6.1. Method's Application Study

The study was initiated in order to establish a Figure 3 contains the comparison of the scales for the same period as above (2006-2010) but using quarterly observations for S&P, Fitch, Moody's, RusRating, AK&M, National Rating Agency, and Expert Rating Agency. International rating agencies are used two scales everyone, namely international and local scales.

In this study the limited distance method[4] has also was used. This method deals only with Russian agencies ratings, and using Moody's scale as a base one. The motivation of exploring such method was that according to Basel II the scales' correspondences between scales of Big-3 agencies were defined. Besides it the difference introducing between the Big-3 rating agencies are not yet accepted by international financial authorities, such as the Bank for International Settlements (BIS). We take into consideration that the average difference between those ratings is not more than one grade. So The Big-3 agencies scales have the conventional correspondences, while the relationship between the Big-3 and the domestic Russian agencies (and the domestic agencies assigned by the Big-3) are assessed using the Distance method. Table 2 contains the correspondence table between the scales of seven rating agencies, which was accepted by Russian authorities.

[4] We define it like a method used only in Russia.

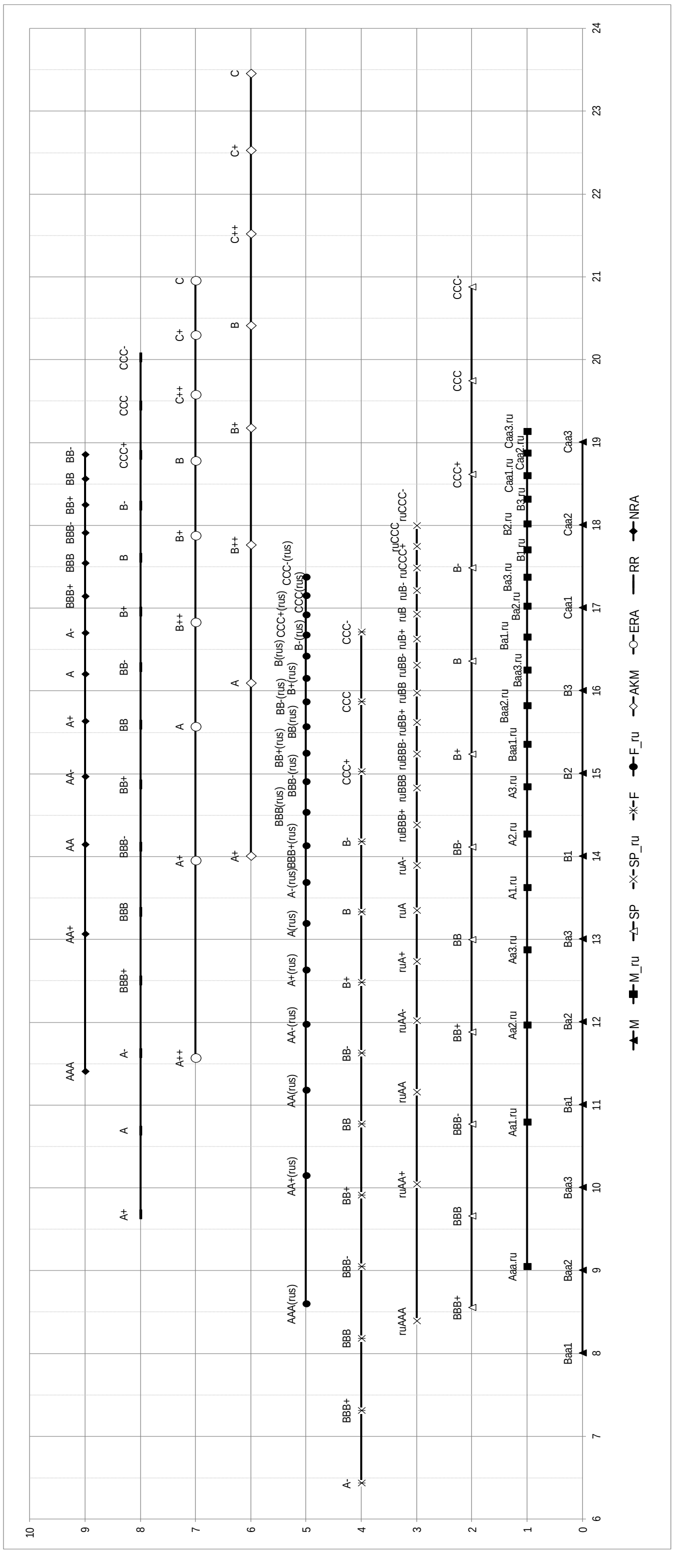

Figure 3. Comparing Russian rating scales

Table 2. Russian rating correspondences (Limited distance method)

Scales as defined for Basel II			**Russian Scales for Rating Agencies registered by the Ministry of Finance of the Russian Federation**				
Moody's	S&P	Fitch	Moody's-Interfax	AK &M	Expert **RA**	National **RA**	RusRating
A3	A-	A-	Aaa.ru Aa1.ru				
Baa1	BBB+	BBB +					
Baa2	BBB	BBB		A++	A++	AAA	
Baa3	BBB-	BBB-					A+
Ba1	BB+	BB+					A
Ba2	BB	BB	Aa2.ru				A- BBB+ BBB
Ba3	BB-	BB-	Aa3.ru	A+	A+	AA+	
B1	B+	B+	A1.ru A2.ru			AA	BBB-
B2	B	B	A3.ru		A	AA-	BB+ BB
			Baa1.ru	A			
B3	B-	B-	Baa2ru			A+	
			Baa3.ru		B++	A	BB-
Caa1	CCC+	CCC	Ba1.ru, Ba2.ru	**B++**		A-	B+
			Ba3.ru			BBB+	
Caa2	CCC		B1.ru B2.ru		B+	BBB BBB-	B
			B3.ru			BB+	B-
Caa3	CCC-		Caa1.ru. Caa2.ru	B+	B	BB BB-	CCC+
			Caa3.ru				CCC

6.2. Regulatory Consequences in Russia

Even Table 3 was considered too complicated for regulatory purposes. Moreover, ratings are used for a variety of goals, some require high levels of creditworthiness, and others, lower levels. The Finance Ministry – in consultation with an Expert Commission consisting of representatives of the rating agencies, both domestic and the Big-3, and ratings users – has adopted a number of rating minima for six different credit worthiness levels (see Table 3).

The relationship between a regulatory requirement for creditworthiness and rating minima is beyond the scope of this paper. Nevertheless, we believe that it is better to have more than one rating threshold than only one. In reality, most regulatory regimes have only one threshold – between investment grade and non-investment grade. Thresholds lead to distortions around the threshold as companies just under the threshold put pressure on the rating agency to increase the rating, while companies under the threshold do not publish the rating if it is possible. Increasing the number of thresholds should reduce the distortion.

Table 3. Rating minima for regulatory purposes

Level	Moody's Investors Service		Standard and Poor's		Fitch Ratings		AK&M	ERA	NRA	Rus Rating
	International	National[1]	International	National	International	National				
	Baa1	Aaa.ru	BBB+	ruAAA	BBB+	AAA(rus)				
	Baa2	Aaa.ru	BBB	ruAAA	BBB	AA+(rus)				
Minimal	Baa3	Aaa.ru	BBB-	ruAA+	BBB-	AA+(rus)	A++	A++	AAA	A-
1	Ba1	Aa1.ru	BB+	ruAA+	BB+	AA(rus)	A+	A+	AA-	BBB-
2	Ba2	Aa2.ru	BB	ruAA	BB	AA-(rus)	A	A	A-	BB-
3	B1	A2.ru	B+	ruA	B+	A-(rus)	B++	B+	BBB-	B-
4	B3	Baa3.ru	B-	ruBBB-	B-	BB-(rus)	B+	B	BB-	CCC
5	Caa2	B3.ru	CCC	ruB-	CCC	B-(rus)				
6	Caa3	Caa3.ru	CCC-	ruCCC-	CCC-	B-(rus)				

[1] Assigned by Moody’s-Interfax

7. Discussion

What we have tried to show in this study is that although it is difficult to compare the ratings scales of credit rating agencies by using only their published ratings, it is possible to compare scales by looking at sets of ratings assigned to similar banks with the assumption that the risk being measured by each credit rating agency is the same for a given rated entity at a given point in time. We are certain that the methodology can be improved. In particular, it would be useful to be able to derive some form of statistic to describe how close a particular credit rating agency's scale is to the average scale, or alternatively, a better statistic to define how close the scales are on average.

The use of one credit rating agency's (Moody's) scale as a base scale was based on a need to simplify the task. As pointed out above, it has the effect of removing from consideration ratings of an entity that has not been rated by Moody's. Further work (nearing a completion) has removed this simplification and we can report that initial results indicate that for the data set described in this article, the final mappings are similar to those presented here.

We believe it is clear from the study that even though the market assumption is wrong; the market assumption is that the scales of S&P, Moody's and Fitch Ratings are the same, namely the first symbol in each sequence (AAA, Aaa, AAA, respectively) correspond to each other, and that each subsequent symbol in each scale also correspond all the way down the scale[5].

We have shown that the rating scales of the Big-3 do not map directly to each other. This then begs several questions:
- How do the rating scales change over time?
- Have the changes in methodology following criticisms of the Big-3 actually affected the mappings between their scales, and if so, in what way? One of

[5] The ratings at the end of each scale are exceptions as the Big-3 do not treat defaults and partial defaults in the same manner.

the Big-3 (Moody's) has been very public about a change in methodology for banks, and there have been across the board changes to bank ratings. However, have the other rating agencies followed suit more quietly? Has Moody's moved away from the other rating agencies?

- Is the mapping we have derived for banks matched by mappings derived for other industries? Our study focused on banks. Considering that the Big-3 tends to have ratings analysts who specialize by industry sector, we think it is very likely that there will be different mappings for each sector (Karminsky, 2010).
- Is there any difference between the Big-3 in specific countries? There is fairly conclusive evidence of bias against countries with rapidly changing (aka developing) economies (see eg. Gultekin-Karakas *et al*. 2011). However, the question raised here is whether this is a bias shared by the Big-3, or whether there are different perceptions amongst the Big-3 Robustness. The models mapping verification partly was done.

The results are sustainable in case of empirical selection of large number of ratings pairs:

- Changing the optimum criteria for which we considered information and power criteria different to quadratic used above;
- Changing the base scale we considered Standard & Poor's and Fitch Ratings like based instead of Moody's scale as discussed above;
- Under the changing of rating to digital scale according Figure 1.

The results sustainability of empirical data set under the change of the time interval was relative stable except the year 2009. With introduction of dummy variables representing years at the regression equation, only the coefficient of 2009 was significant. This could be mainly explained by partial modification of ratings methodology, which was caused by the crisis of 2007-2009.

Eventually the methodology of ratings mapping is expected to change. This was supported on the data bases of 2011 of one of the Russian rating agency. That is why the verification of the method control should be done not less than in one year. Thus, further research on mapping method sustainability is planned in future.

We believe a major issue in the coming years will be rating scale harmonization. Already the European Union is considering the harmonization of ratings (see European Commission Staff Report, 2011). This proposal has been supported by influential industry lobbies (see European Savings Bank Group, 2012).

Currently, each credit rating agency defines its own scale, indeed scales as most agencies have multiple scales to quantify different types of risk. The benefits of a single scale are fairly obvious: comparability and interchangeability for users. On the regulatory side, it will be provide a benchmark to measure the performance of the credit rating agencies.

The credit rating agencies will resist harmonization for almost the same reasons. The Big-3 is likely to resist the most strongly, for their scales are widely accepted and it benefits them to have control over their scales. The existence of different scales is reminiscent of different measures for other characteristics. Although there have been advocates for different scales (the USA still differs from the scientific community and most jurisdictions when measuring temperature, length, weight, etc.), the advantages of a common standard eventually overcome the differences.

The problem historically is to find a mechanism to define the metric, eg., for musical notes it was the mapping of notes to sound frequencies. At present, there is no unique standard for credit risk, for even probability of default is not uniquely defined and Moody's main scale is defined in terms of recoverability. As an interim measure, it is necessary to have some mechanism for comparing the ratings of different credit rating scales, such as the methodology described in this paper.

8. Conclusion

The existence of multiple credit rating scales increases barriers of entry to new credit rating agencies. We have developed a methodology for comparing the rating scales of different credit rating agencies based on the published ratings of entities, where the entities have ratings from more than one credit rating agency.

The results of comparing the ratings of banks around the world for the period of a five-year period show that the jointly accepted mapping from one credit rating agency to another is incorrect. The methodology underpinned a comparison table published by the Ministry of Finance of the Russian Federation for the seven independent credit rating agencies accredited by the Ministry. We believe that in spite of all the problems generated by this comparison table, it will increase transparency of the Russian financial market.

References

Aivazian, S., Golovan, S., Karminsky, A., and Peresetsky, A., 2011. О подходах к сопоставлению рейтинговых шкал [The approaches to the comparison of rating scales], *Applied Econometrics*, 3(23), pp.13-40.

Aleskerov, F., Andrievskaya, I., Penikaz, G., and Solodkov, V., 2010. *Анализ математических моделей базель II* [*Analysis of mathematical models of Basel II*]. Moscow: Fizmatlit, p.288.

Altman, E. and Saunders, A., 1997. Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, 21(11-12), pp.1721-1742.

Altman, E. and Rijken, H., 2004. How rating agencies achieve rating stability. *Journal of Banking and Finance*, 28(11), pp.2679-2714.

Amato, J. and Furfine, C., 2004. Are credit ratings procyclical? *Journal of Banking and Finance*, 28(11), pp.2641-2677.

Basel, 2004. *International Convergence of capital measurement and capital standards: A revised framework.* Basel: Bank for International

Settlements.
Basel, 2010. *Basel III: A global regulatory framework for more resilient banks and banking systems*. Basel: Bank for International Settlements.
Beattie, V. and Searle, S., 1992. Bond ratings and inter-rater agreement. *Journal of International Securities Markets, 6(2),* pp.167-172.
Becker, B. and Milbourn, T., 2011. How did increased competition affect credit ratings? *Journal of Financial Economics*, 101(3), pp.493–514.
Bongaerts, D., Martijn Cremers, K.J., and Goetzmann, W.N., 2012. Tiebreaker: Certification and multiple credit ratings. *Journal of Finance,* 67(1), pp. 113-152.
Cantor, R. and Packer, F., 1994. The credit rating industry. *Federal Reserve Bank of New York Quarterly Review,* 19(2), pp.1-26.
Cantor, R. and Packer, F., 1995. The credit rating industry. *Journal of Fixed Income*, 5(3), pp.10-34.
Cantor, R. and Packer, F., 1996. Discretion in the use of ratings: The case of the NAIC. *Journal of Insurance Regulation*, 15(2), pp.234-255.
Cantor, R., Packer, F. and Cole, K., 1997. Split ratings and the pricing of credit risk. *Journal of Fixed Income*, 7(3), pp.72-82.
Caporale, G.M., Matouse, R., and Stewart, Ch., 2010. EU banks rating assignments: Is there heterogeneity between new and old member countries? *Review of International Economics*, 19(1), pp.189–206.
Cornaggia, J. and Cornaggia, K.J., 2013. Estimating the costs of issuer-paid credit ratings. *Review of Financial Studies,* forthcoming. [online] Available at: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1705843> [Accessed 5 August 2013].
Ederington, L., 1986. Why split ratings occur? *Financial Management*, 15(1) pp.37-47.
Eisl, A., Elendner, H., and Lingo, M., 2013. Re-Mapping credit ratings. In: *INFINITI Conference on International Finance*, Dublin, Ireland, 13-14 June 2011.
European Commission Staff Report, 2011. Executive summary of the impact assessment. *Commission Staff Working Paper,* SEC(2011) 1355/2, p.5. [pdf] Available at: <http://ec.europa.eu/internal_market/securities/docs/agencies/SEC_2011_1355_en.pdf> [Accessed 24 July 24 2013].
European Savings Bank Group, 2012. *ESBG position paper on the proposals for "CRA III".* Brussels: ESBG.
Ferri, G., Liu, L.-G., and Majnoni, G., 2001. The role of rating agency assessments in less developed countries: Impact of the proposed Basel guidelines. *Journal of Banking and Finance*, 25(1), pp.115-148.
Griffin, J. and Tang, D., 2012. Did subjectivity play a role in CDO credit ratings? *Journal of Finance,* 67(4), pp.1293-1328.
Gultekin-Karakas, D., Hisareikhlar, M., and Ozturk, H., 2011. Sovereign

risk ratings: Bias toward developed countries? *Emerging Markets Finance and Trade,* 47(S), pp.69-87.

Guttler, A. and Wahrenburg, M., 2007. The adjustment of credit ratings in advance of defaults. *Journal of Banking & Finance*, 31(3), pp.751-767.

Hainsworth, R., 2009. Сопоставимость уровней кредитных рейтингов, присвоенных разными агентствами [The comparability of the gradations of credit ratings assigned by different agencies]. *Money and Credit*, 12, pp.46-51.

Hainsworth, R., Karminsky, A., and Solodkov, V., 2012. Arm's length method for comparing rating scales. *NRU Higher School of Economics Working Paper*, WP BRP 01/FE/2012.

Heckman, J., 1979. Sample selection bias as a specification error. *Econometrica*, 47(1), pp.153-161.

Iannotta, G., 2006. Testing for opaqueness in the European banking industry: Evidence from bond credit ratings. *Journal of Financial Service Researchs*, 30(3), pp.287-309.

Jewell, J. and Livingston, M., 1998. Split ratings, bond yields, and underwriter spreads. *Journal of Financial Research*, 21(2), pp.185-204.

Karminsky A., Peresetsky, A., and Ryzhov, A., 2006. Модели рейтингов банков для риск-менеджмента [Models of bank ratings in risk management]. *Financial Risk Management*, 4(08), p.362-373.

Karminsky, A. and Solodkov, V., 2010. Единое рейтинговое пространство: миф или реальность? [A unified rating space: Myth or reality?]. *Банковское дело*, 9, pp.56-60.

Karminsky, A. and Peresetsky, A., 2007. Модели рейтингов международных агентств [Models of ratings by international agencies]. *Applied Econometrics*, 1(5), pp.3-19.

Karminsky, A. and Peresetsky, A., 2009. Ratings as measure of financial risk: Evolution, function and usage. *Journal of the New Economic* Association, 1-2, pp.86-104.

Karminsky, A. and Sosyurko, V., 2010. Rating models system: Methods and technology. In: *The First Scientific-Practical International Conference Economics and Management*. Varna, Bulgaria, 16-18 September 2010. Varna: University of Varna.

Karminsky, A., 2010. Rating model opportunities for emerging markets. In: *Challenges for Analysis of the Economy, the Businesses, and Social Progress International Scientific Conference*, Hungary, Szeged, 19-21 November, 2009. Szeged: University Press.

Karminsky, A., 2012. The multiplication of the credit rating agencies effort under IRB approach. *Investment Management and Financial Innovations,* 9(4), pp.78-88.

Karminsky, A., Peresetsky, A., and Petrov, A., 2005. *Рейтинги в экономике: методология и практика* [*Ratings in the economy: Methodology and practice*]. Moscow: Finance and Statistics.

Karminsky, A.M. and Sosyurko, V., 2011. Сопоставление банковских рейтингов различных агентств [Comparison of bank credit ratings for various agencies]. *Journal of the New Economic Association*, 12, pp.123-146.

Koshelyuk, Ju., 2008. *Formation of rating for Russian banks*. Ph.D thesis, Higher School of Economics, Moscow, p.204.

Liss, H. and Fons, J., 2006. Mapping Moody's national scale ratings to global scale ratings. *Moody's Rating Methodology*, December.

Matovnikov, M., 2008. Как уполномочивать рейтинговые агентства для оценки кредитоспособности банков [How to authorize the credit rating agencies to assess the creditworthiness of banks]. *Money and Credit,* 12, pp.26-33.

Morgan, D., 2002. Rating banks: Risk and uncertainty in an opaque industry. *The American Economic Review*, 92(4), pp.874-888.

Pagratis, S. and Stringa, M., 2009. Modeling bank senior unsecured ratings: A reasoned structured approach to bank credit assessment. *International Journal of Central Banking*, 5(2), pp.1-39.

Partnoy, F., 2002. The paradox of credit ratings. In: R. Levich, G. Majononi and C. Reinhart eds. *Ratings, rating agencies and the global financial system*. Boston: Kluwer Academic Publishers, pp.65-84.

Pederzoli, Ch. and Torricelli, C., 2005. Capital requirements and business cycle regimes: Forward-looking modeling of default probabilities. *Journal of Banking and Finance*, 29(12), pp.3121-3140.

Peresetsky, A. and Karminsky, A., 2008. Models for Moody's bank ratings. *BOFIT Discussion Papers*, 17/2008.

Peresetsky, A. and Karminsky, A., 2011. Models for Moody's bank ratings. *Frontiers in Finance and Economics*, 8(1), pp.88-110.

Peresetsky, A., Karminsky, A., and Soest, van A., 2004. Modeling the ratings of Russian banks. *Economics and Mathematical Methods*, 40(4), pp.10-25.

Poon, W.P.H., 2003. Are unsolicited credit ratings biased downward? *Journal of Banking and Finance*, 27(4), pp.593-614.

Roy, P. van, 2006. Is there a difference between solicited and unsolicited bank ratings and, if so, why? *National Bank of Belgium Working Paper Research*, 79.

Servigny, A. and Renault, O., 2004. *Measuring and managing credit risk*. New York: McGraw-Hill, p.467.

Smirnov, S.N. and Sholomitsky, A., 2010. A comparison of the quality of ratings of Russian banks. *Preprint HSE WP16/*2010*/02*.

White, L.J., 2002. An analysis of the credit rating industry. In: M. Ong. ed. *Credit ratings, methodologies, rationale and default risk*. London: Risk Books, pp.17-44.

Xia, H., 2012. Can competition improve the information quality of credit ratings? In: *7th Annual Conference of the Financial Intermediation Research Society Annual Conference*, Minnesota, U.S.A., 1-4 June, 2012. [pdf] Available at: <https://wpweb2.tepper.cmu.edu/wfa/wfasecure/upload/2012_PA_481629_718992_895793.pdf> [Accessed 2 July 2013].