Fuzzy Logic Model of Soft Data Analysis for Corporate Client Credit Risk Assessment in Commercial Banking

Brkic, Sabina and Hodzic, Migdat and Dzanic, Enis

American University in Bosnia and Herzegovina, International University Sarajevo, Univerzitet u Bihacu

November 2017

Sabina Brkić, MBA
American University in Bosnia and Herzegovina,
Mije Kerošević Guje 3, 75000 Tuzla
sabc1122@aubih.edu, +387 61 726 701

prof. Migdat Hodžić, PhD
International University Sarajevo,
Hrasnička cesta 15, 71000 Sarajevo
migdathodzic@gmail.com, +387 60 342 02 20

assoc. prof. Enis Džanić, PhD
Univerzitet u Bihaću
Ivana Pavla II b.b., 77000 Bihać
Enis.dzanic@bih.net.ba, +387 61 260 790

FUZZY LOGIC MODEL OF SOFT DATA ANALYSIS FOR CORPORATE CLIENT CREDIT RISK ASSESSMENT IN COMMERCIAL BANKING

MODEL NEIZRAZITE LOGIKE ZA ANALIZU KREDITNOG RIZIKA KOD KORPORATIVNIH KLIJENATA KOMERCIJALNIH BANAKA

ABSTRACT

This paper deals with the use of fuzzy logic as a support tool for evaluation of corporate client credit risk in a commercial banking environment. It defines possibilistic distribution of soft data used for corporate client credit risk assessment by applying fuzzy logic modeling, with a major goal to develop a new expert decision-making fuzzy model for evaluating credit risk of corporate clients in a bank. Currently, predicting a credit risk of companies is inaccurate and ambiguous, as well as affected by many internal and external factors that cannot be precisely defined. Unlike traditional methods for credit risk assessment, fuzzy logic can easily incorporate linguistic terms and expert opinions which makes it more adapted to cases with insufficient and imprecise hard data, as well as for modeling risks that are not fully understood. Fuzzy model of soft data, presented in this paper, is created based on expert experience of corporate lending of a commercial bank in Bosnia and Herzegovina. This market is very small and it behaves irrationally and often erratically and therefore makes the risk assessment and management decision making process very complex and uncertain which requires new methods for risk modeling to be evaluated. Experts were interviewed about the types of soft variables used for credit risk assessment of corporate clients, as well as for providing the inputs for generating membership functions of these soft variables. All identified soft variables can be grouped into following segments: stability, capability and readiness/willingness of the client to repay a loan. The results of this work represent a new approach for soft data usage/assessment with an aim of being incorporated into a new and superior soft-hard data fusion model for client credit risk assessment.

***Key words**: fuzzy logic, credit risk, default risk, commercial banking*

***JEL**: C53, G21, G32*

SAŽETAK

Ovaj rad koristi neizrazitu logiku kao alat za procjenu kreditnog rizika korporativnih klijenata komercijalnih banaka. Rad definira mogućnošćnu distribuciju mekih podataka koji se koriste za procjenu kreditnog rizika korporativnih klijenata koristeći neizrazitu logiku, sa ciljem razvoja novog ekspertnog sistema za procjenu

gore pomenutog rizika. Trenutno, procjena kreditnog rizika kompanija je nejasna i neprecizna, te je pod utjecajem brojnih internih i eksternih faktora koje nije moguće jasno definirati. Za razliku od tradicionalnih metoda za procjenu kreditnog rizika, neizrazita logika može inkorporirati lingvističke izraze te mišljenja eksperata čime je prikladniji za slučajeve gdje imamo neprecizne ili nepotpune podatke, kao i za modeliranje rizika koji nisu potpuno shvaćeni. Model neizrazite logike koji se koristi u ovom radu je baziran na iskustvima korporativnog kreditiranja komercijalne banke u Bosni i Hercegovini. Ovo tržište je malo te se ponaša neracionalno i često nepravilno što čini procjenu rizika te odluke menadžmenta kompleksnim i nesigurnim što kreira potrebu za novim modelom za procjenu rizika. Eksperti su intervjuirani za vrste mekanih varijabli koje se koriste za procjenu kreditnog rizika korporativnih klijenata, kao i za inpute potrebne za generisanje funkcija članstva tih mekanih varijabli. Sve identificirane mekane varijable mogu se grupisati u slijedeće segmente: stabilnost, sposobnost i spremnost klijenta na vraćanje kredita. Rezultati ovog rada predstavljaju novi pristup upotrebe mekih podataka, sa ciljem inkorporacije istih u novi i superiorni model fuzije mekih i tvrdih podataka za procjenu kreditnog rizika klijenata.

***Ključne riječi**: neizrazita logika, kreditni rizik, rizik neplaćanja, komercijalno bankarstvo*

***JEL**: C53, G21, G32*

1. INTRODUCTION

Credit risk is one of the largest risks faced by commercial banks and it is assuming increased importance in a changing regulatory regime and quite volatile market conditions. Commercial banks generally use different methods and procedures for assessing credit risk but it is the primary job of a loan officer to evaluate an applicant's financial position in order to determine whether the applicant will be able to repay the bank loan it requests/applies for. Risk analysis techniques are powerful tools that help professionals manage uncertainty and can provide valuable support for decision making. Risk analysis helps us to take both certain and uncertain elements and include them in a calculation of specific scenarios of the future events. These techniques can be either qualitative or quantitative depending on the information available and the level of detail that is required (Bennett and Bohoris, 1996, pp. 467-475). Quantitative techniques rely heavily on statistical approaches while qualitative techniques rely more on judgment than on statistical calculations.

The complex and uncertain nature of loan processing has enforced banks, and other financial institutions, make loan decisions by utilizing experienced lending officers to perform the essential tasks and evaluations. A loan officer has to fully understand the level of risk a loan would entail. Only when all risk factors are satisfactory the loan is approved. To make such risk assessment of a corporate client, a loan officer has to understand and assess the following: the financial position, repayment ability and strength of the company, whether the company has a sound record of credit worthiness, work history, what is applicants experience and management skills, does the company have a sound business plan which demonstrates his/her understanding of the business and his/her commitment to the success of the business, is company's cash-flow solid and stable, willingness to repay debt and many more. Such analysis incorporates not only the economic data but also the qualitative information concerning the borrower. Data which is subject to this analysis can be classified as hard and soft data. Hard data is usually objective, they express a measure and thus are measurable, quantitative and crisp, while soft data is linguistic, qualitative, subjective and non-measurable.

In addition to experienced loan officers, many banks usually use various types of scoring models to assess credit risk of a borrower before disbursing a loan. Scoring models were initially introduced to standardize the decision making process and to increase the transparency of a bank's business. They are usually estimated with historical data and statistical methods. Scoring models generally do not

follow the Basel II regulatory capital framework definitions since their primary aim is not to fulfill the supervisory requirements but to provide internal decision support. In addition to traditional scoring models, complexity and repetitiveness of decisions in the finance area have made financial services sector an area where various kinds of expert systems have found their many applications. An expert system that evaluates loan applications reduces the time and improves the quality of the evaluation (Gupta & Celtek, 2001, p.18). Expert systems are computer programs which summarize knowledge of human experts and are used for decision-making and/or problem-solving.

While bank risk professionals strive for a better understanding of risk and employ complex models for risk assessment, many risks are still unknown or not well understood. Traditional risk models are based on probability and classical set theory which are widely used for assessing market, credit, insurance and trading risk. However, many risks still cannot be analyzed sufficiently by applying classical probability models because of lack of sufficient experience data, lack of knowledge and vagueness, as well as complex cause-and-effect relationships that are inherent in certain risk types. Many authors believe that the best way to solve obstacles in facing with any type of uncertainties is by utilizing fuzzy logic and theory of possibility. It provides a mathematical advantage to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning. "Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth, truth values between completely true and completely false" (Gupta & Celtek, 2001, p.20). By applying fuzzy logic most variables of a model are described in linguistic terms which makes fuzzy logic models more intuitively similar to the human reasoning. For risks that do not have a proper quantitative probability model, a fuzzy logic system can help model the cause-and-effect relationships, assess the degree of risk exposure and rank the key risks in a consistent way, considering both the available data and experts opinions (Shang & Hossen, 2013, p.3).

Fuzzy logic has been utilized in various industry areas such as, in artificial intelligence, computer science, control engineering, decision theory, expert systems, logic, management science, operations research, pattern recognition and robotics (Zimmermann, 2001, pp. 158-241; 369-404). Considering risk assessment, many studies of fuzzy logic have appeared in different business areas such as information security, software development, ground water nitrate risk management, system failure, civil hazardous materials, natural hazards, bank, etc. (Zirakja & Samizadeh, 2011, pp. 99-100).

Fuzzy set theory and fuzzy logic were introduced by Lotfi A. Zadeh in 1965 who was almost single-handedly responsible for the early development in this field. Fuzzy Set Theory is a mathematical theory for describing impreciseness, vagueness and uncertainty. A fuzzy set is a collection of objects with graded membership. As an extension of his theory of fuzzy sets and fuzzy logic (Zadeh, 1965, pp. 338-353), the "Theory of Possibility" was developed by Zadeh in 1978 in which he explained that possibility distributions were meant to provide a graded semantics to natural language statements by interpretation of membership functions of fuzzy sets as possibility distributions. He introduced the concept of possibilistic fuzzy distributions, contrary to random and probabilistic distributions, and noticed that what is probable must initially be possible, but not vice versa. Zadeh (1978) wrote that in dealing with soft data, encountered in various fields, the standard practice was to rely almost completely on probability theory and statistics and he stressed out that those techniques could not cope effectively with those problems in which the softness of data is non-statistical in nature. Soft data encounter predominance of fuzziness. Author's rationale for using fuzzy logic for soft data analysis "rests on the premise that the denotations of imprecise terms which occur in soft database are for the most part fuzzy sets rather than probability distributions" (Zadeh, 1981, pp. 515-541). The difference between probability and possibility is that the concept of possibility is an abstraction of our intuitive perception while concept of probability depends on likelihood, frequency, proportion or strength of belief.

The purpose of this study is to design and develop fuzzy distribution of soft data/variables used for

corporate client credit risk assessment by applying fuzzy logic. Terms fuzzy and possibilistic distributions are used interchangeably. Expert sample is created ad hoc with a commercial bank in Bosnia and Herzegovina that was willing to take part in this project at this initial phase. We are now in a process of adding data from other local banks for the purpose of expanding the relevant soft database. Top senior credit risk assessment experts from this bank were interviewed and they have provided all information about the process, data processing and inputs used for credit risk assessment. The next step was to identify and generate a list of the most significant soft variables and their descriptions. For this purpose, mentioned experts and relevant literature were consulted in the definition of membership functions. Experts were interviewed for the types of soft variables used for conducting credit risk assessment of corporate clients in a local commercial bank in Bosnia and Herzegovina, as well as for describing the membership functions of these soft variables. Experts have provided inputs for generating universe of discourse and the number and description of membership functions related to each soft variable. Data processing is done by listing all identified soft variables and by mapping their membership values into membership functions based on inputs from interviewed experts. The final step was in generating graphical illustration of possibility distribution of each soft variable. The results of this work represent a new approach for soft data usage/assessment with an aim of being incorporated into a new and superior soft-hard data fusion model (Hodzic, 2016a, pp. 58-66, Hodzic, 2016b, pp. 17-32) for credit risk assessment that will assist bank managers in identifying credit risk factors and improve evaluation of the corresponding default risks of their loan applicants. Design and development of possibilistic distribution of soft data/variables used for corporate client credit risk assessment serves as the first step in this process.

In this paper, we first present a general overview of credit risk assessment in commercial banking, followed by a brief explanation of fuzzy logic and Theory of possibility. Next section provides an overview of the results of this study based on which possibility distribution model of identified soft variables is developed, used by the bank for assessing the credit risk of a corporate loan applicant. Finally, we make conclusions and give directions for future research.

2. SOFT DATA COLLECTION AND FUZZY DISTRIBUTIONS BASED ON EXPERT INTERVIEWS

Considering Zadeh's work, in which he explained that soft database are for the most part fuzzy sets rather than probability distributions, in this section we show fuzzy distributions of soft variables used in a commercial bank for credit risk assessment of corporate clients. All identified soft variables can be grouped in following segments: stability, capability and readiness/willingness of the client to repay a loan. Each of these segments have a variety of impact on the assessments going from low impact to medium and high. The following soft variables are identified and used by the targeted bank for the purpose of credit risk assessment of a corporate client:

- Stability/Capability of the loan applicant based on company size considering its total assets,
- Stability/Capability of the loan applicant based on company size considering its total income,
- Stability/Capability of the loan applicant based on company size considering its total number of employees,
- Stability of the loan applicant considering number of years the company is doing business,
- Stability of the loan applicant considering number of years it operates profitably (considering operating income),
- Stability of the loan applicant considering number of days of blocked bank accounts in the last year,

- Stability of the loan applicant considering company's repayment history in the bank (if already a client),
- Stability of the loan applicant considering company's worst regulatory classification found in the Central Credit Registry,
- Stability of the loan applicant considering future development of the company,
- Stability of the loan applicant considering company's competition,
- Readiness/Willingness/Character of the management of the company to repay the loan,
- Capability/Quality of the company's management.

For linguistic semantic values Kickert (1979) demonstrated that there is no significant computational difference between selection of a wide versus a narrow universe of discourse, so no limits in this regard were used in this study. Upon identification of soft variables and their universe of discourse, it was necessary to develop their membership functions. The identification of the number, size and shape of the membership function of a soft variable is often subjective and unpredictable/arbitrary/unmotivated, since its purpose is to incorporate the expert knowledge into a fuzzy system. Although many shapes can be used, there are authors that recommend the use of simple membership function shapes, such as triangles and trapezoids in order to improve maintainability and execution results (Viot, 1993, pp. 26-33; Kosko, 1993, pp. 76-81). For the purpose of this study membership functions are generated completely unmotivated and are expressing the interviewed experts own opinion and experience.

The results of this study are provided in Figures 1-9. Moreover, due to confidentiality we do not disclose estimation results that have been given by the bank experts but we are instead showing graphical illustration of the possibility distribution results.

Stability and capability of a company is assessed considering company's size based on its total assets, total income, as well as total number of employees. There are five outputs that can be estimated such as extremely small, small, medium, large and extremely large company. Fuzzy distributions of these variables, as shown in Figure 1-3, demonstrate trapezoid membership functions. The output for extremely small shows a left trapezoid membership function, while output for extremely large shows right trapezoid membership function. The distributions for small, medium and large demonstrate trapezoid membership function that are not represented in smooth lines, but can be considered as extended trapezoid membership functions. Such distributions are result of experts' estimations which is based on grouping the values from the universe of discourse in certain ranges and assigning the truth level per defined ranges.

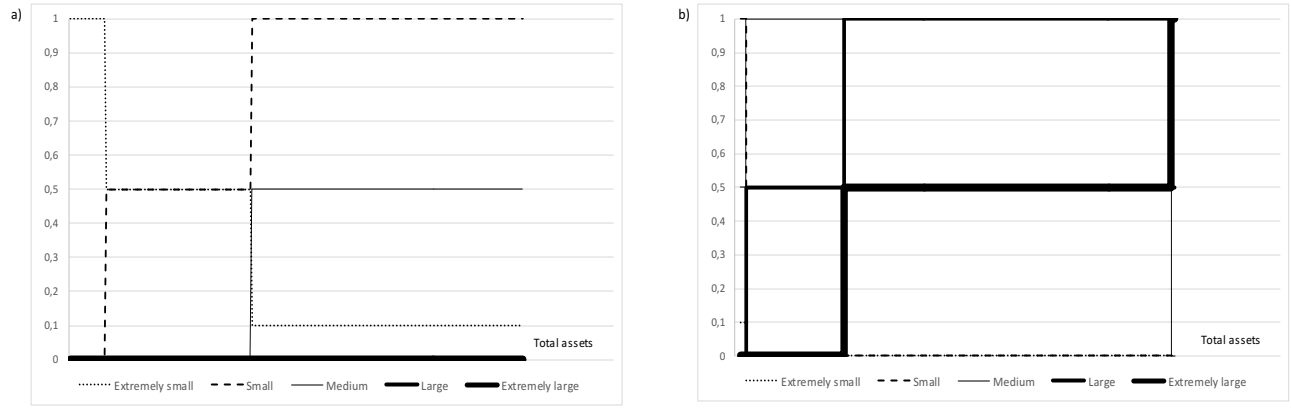

Figure 1 – Stability/Capability of the loan applicant based on company size considering its total assets: a) represents the first 1% of total universe of discourse, b) represents the remaining 99% of total universe of discourse

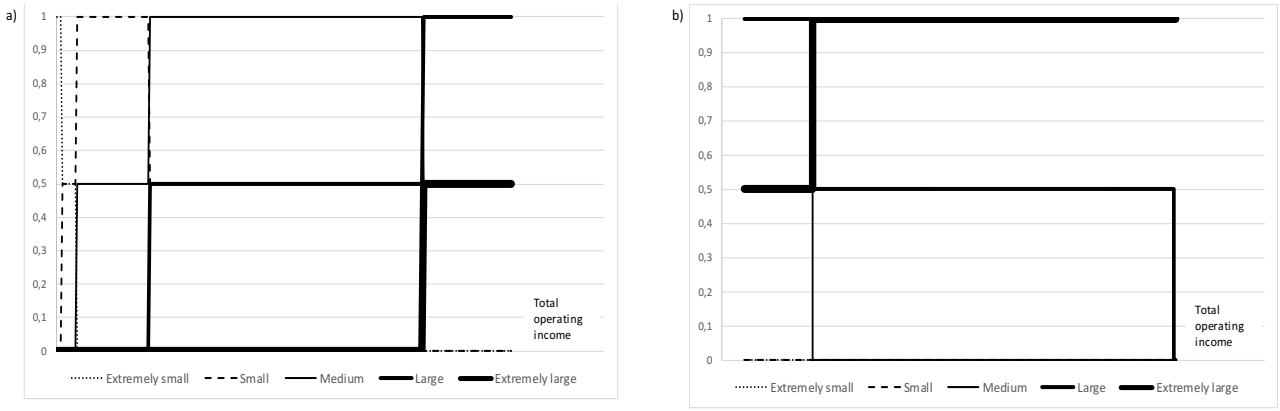

Figure 2 – Stability/Capability of the loan applicant based on company size considering its total income: a) represents the first 5% of total universe of discourse, b) represents the remaining 95% of total universe of discourse

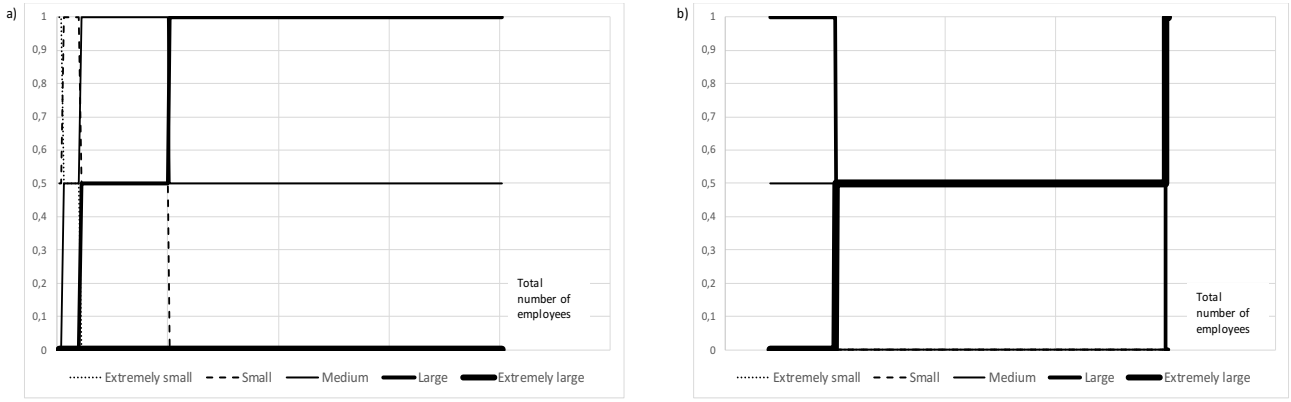

Figure 3 – Stability/Capability of the loan applicant based on company size considering its total number of employees: a) represents the first 10% of total universe of discourse, b) represents the remaining 90% of total universe of discourse

Stability of the company is also assessed based on the number of years a company is doing business, as well as number of years it operates profitably (considering operating income). Figures 4 and 5 show fuzzy distributions for unstable, less stable, stable, very stable and extremely stable company.

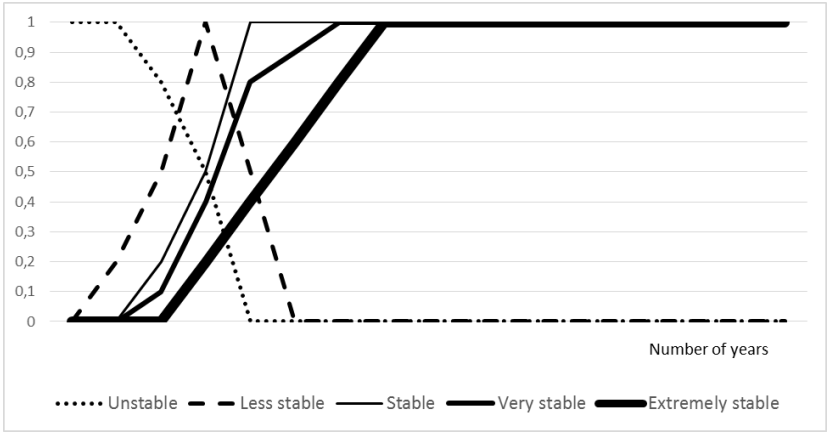

Figure 4 – Stability of the loan applicant considering number of years the company is doing business

Fuzzy distributions of stability of the company based on number of years it is in the business show left trapezoid membership function for unstable company, right trapezoid for stable, very stable and extremely stable company, while less stable is a triangular membership function. Same output of fuzzy distributions is shown in case of stability of the company based on number of years it operates profitably (considering operating income).

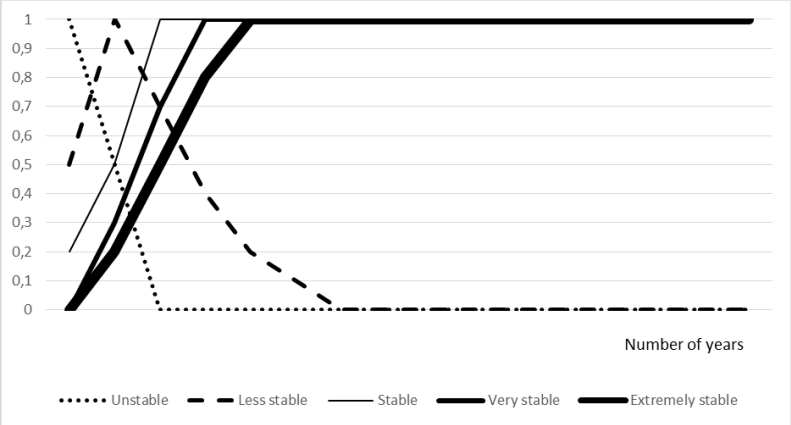

Figure 5 – Stability of the loan applicant considering number of years it operates profitably (considering operating income)

Number of days of blocked bank accounts a company had in the last year is also an indicator of the stability (and capability) of a company and based on this variable a bank is assessing whether a company is unstable, less stable, stable, very stable and extremely stable company. Fuzzy distributions of these variables, as shown in Figure 6, demonstrate trapezoid membership functions. The output for the unstable one shows a right trapezoid membership function, while output for stable, very stable and extremely stable shows left trapezoid membership function. The distribution for less stable company demonstrate left and right trapezoid membership function. All shown distributions are not represented in smooth lines, but can be considered as extended trapezoid membership functions. Such distributions are result of experts' estimations which is based on grouping the values from the universe of discourse in certain ranges and assigning the truth level per defined ranges.

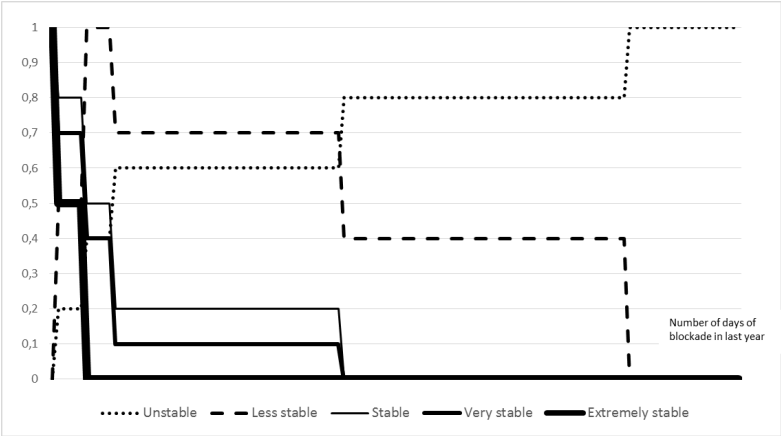

Figure 6 – Stability of the loan applicant considering number of days of blocked bank accounts in the last year

Same results, in terms of types of distributions, is shown in case of assessing stability (and capability) of a company's repayment history in the bank (if already a client), as shown in Figure 7, and company's worst regulatory classification from the Central Credit Registry. Compared to

distributions shown in Figure 6 and Figure 8, interviewed experts have designated more granularity for ranges defined for the universe of discourse in case of distributions shown in Figure 7.

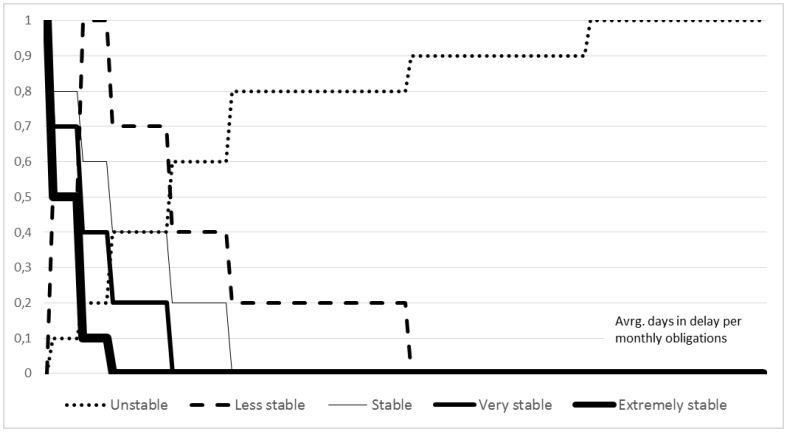

Figure 7 – Stability of the loan applicant considering company's repayment history in the bank (if already a client)

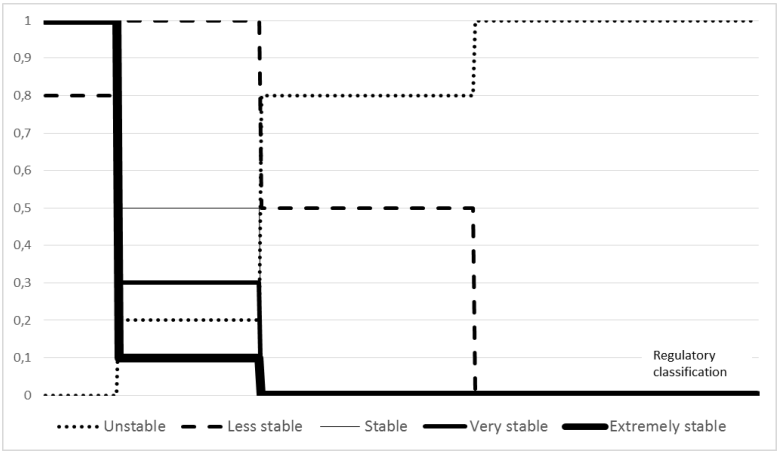

Figure 8 – Stability of the loan applicant considering company's worst regulatory classification found in the Central Credit Registry

The last four soft variables, that interviewed experts use for assessing the credit risk of a loan application for a corporate client, show identical membership functions in terms of types of distributions, as demonstrated in Figure 9. The membership functions of future development of the company, readiness/willingness/character of the management of the company to repay the loan and capability/quality of the company's management are assessed using the same values of universe of discourse i.e., negligible, weak, average, good and excellent. Company's competition impact/influence is assessed by using values critical, strong, moderate, low and negligible impact as the universe of discourse.

The output for the unstable ones shows a S-membership function, while output for stable and extremely stable show a Z-membership function. The distribution for less stable company demonstrate left and right triangular membership function, while very stable show a right triangular membership function. All shown distributions are not represented in smooth lines.

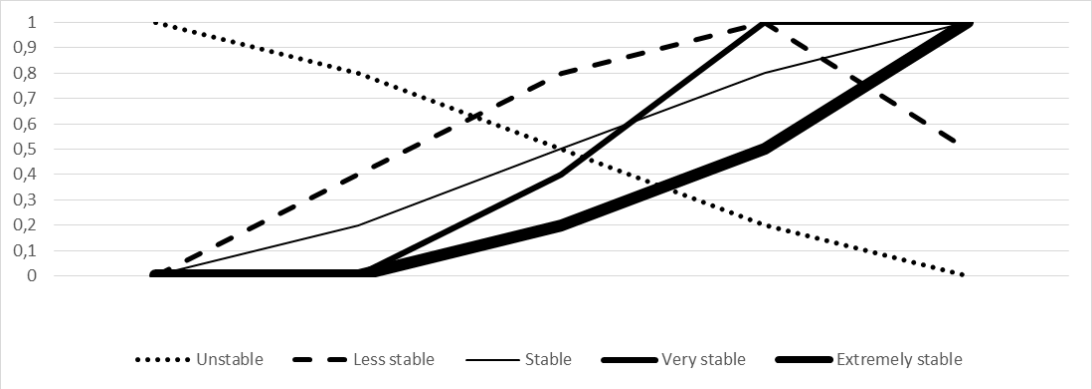

Figure 9 – Possibilistic distribution for: Stability of the loan applicant considering future development of the company, Stability of the loan applicant considering company's competition, Readiness/Willingness/Character of the management of the company to repay the loan, Capability/Quality of the company's management.

In general, results show that fuzzy modeling of soft data, used for corporate client credit risk assessment in targeted bank, is presented as left and/or right triangular and trapezoid fuzzy distributions, as well as various combinations of them.

3. CONCLUSION

In this paper, a possibility distribution model of key/main soft variables, used for corporate client credit risk assessment in commercial banking, has been developed based on specific input data given by the experts. We have presented a new methodology for transforming linguistic and intuitive (soft) information about bank credit risk data into a series of mathematical fuzzy (possibilistic) distributions which can be handled quantitatively and combined (fused) with related probabilistic data.

We conducted a series of interviews with risk specialists from a local bank in order to collect a variety of soft data related to credit risk assessment of local corporate clients. Similar interviews will be conducted in the future with other local banks, as well as extended to incorporate soft data used for private individual credit risk assessment.

Our final aim is to be able to improve bank credit risk assessments and other relevant and highly needed information, where various assessment and modeling approach is suggested and/or anticipated, by using exact and more precise mathematical methodology.

4. FUTURE RESEARCH

The results of this work represent a new approach for corporate client soft data usage/assessment in commercial banking with an aim of finally being incorporated into a new and superior soft-hard data fusion model (Hodzic, 2016a, pp. 58-66, Hodzic, 2016b, pp. 17-32) for client credit risk assessment and other similar assessments. Future research will be focused on analysis of hard and soft data of a pool of targeted corporate clients that did not default, classified as Good choice of the bank, and those that did default, classified as Bad choice of the bank. Hard data will be analyzed by using probabilistic mathematical methodology (random data models), while soft data by applying the possibility mathematical methodology (fuzzy data models) and results of this study. The results of such analysis will be used as a testing sample for the new and superior soft-hard data fusion model for client credit risk assessment. Further tests are necessary to conclude whether this approach can be validated such as interviews with more experts from different banks to validate the possibility distributions, as well as main/key soft variables and their membership functions. Further data can be obtained from other

banks, as well as performing stress testing on the outputs of the model, in order to test the final system and compare its predictions against actual bank decisions. The research will be extended to incorporate a study on private individual customers are well.

5. REFERENCE LIST

Bank for International Settlements. (June 2006). International Convergence of Capital Measurement of Capital Measurement. *Basel Committee on Banking Supervision, A Revised Framework Comprehensive Version*.

Bennett, J.C., Bohoris, G.A., Aspinwall, E.M., & Hall, R.C., (1996). Risk analysis techniques and their application to software development. *European Journal of Operational Research, vol. 95, (3)*, pp. 467-475.

Gupta, Vipul K . & Celtek, S., (2001). A fuzzy expert system for small business loan processing, *Journal of International Information Management: Vol. 10*, Article 2. Available at: http://scholarworks.lib.csusb.edu/jiim/vol10/iss1/2

Hodzic, M., (2016a). Fuzzy to Random Uncertainty Alignment, *Southeast Europe Journal of Soft Computing, 5*, pp. 58-66.

Hodzic, M., (2016b). Uncertainty Balance Principle, *IUS Periodicals of engineering and natural sciences, 4, No.2*, pp. 17-32.

Kickert, W. J. M., (1979). An example of linguistic modeling, the case of Mulder's theory of power, in: M.M. Gupta, R.K. Ragade & R.R. Yager (eds.): *Advances in Fuzzy Set Theory and Applications*, North-Holland, Amsterdam.

Kosko, B., & Isaka, S., (1993). Fuzzy logic, *Scientific American 269*, pp.76-81.

Shang, K., Hossen, Z., (2013). *Applying Fuzzy Logic to Risk Assessment and Decision-Making*, Casualty Actuarial Society, Canadian Institute of Actuaries, Society of Actuaries.

Viot, G., (1993). Fuzzy logic: Concepts to constructs, *AI Expert, 8*, pp. 26-33.

Zadeh, L. A., (1965). Fuzzy sets, *Information and Control, vol. 8*, pp. 338-353.

Zadeh, L. A., (1978). Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems 1*, pp. 3-28.

Zadeh, L. A., (1981). *Possibility theory and soft data analysis*, Selected papers by Lotfi Zadeh, pp. 515-541.

Zirakja, M.H., Samizadeh, R., (2011). Risk Analysis in E-commerce via Fuzzy Logic. *Int. J. Manag. Bus. Res., 1 (3),* pp. 99-112.