

# Modeling of Modern Business Intelligence Analytical Functions Based on Stochastic Simulation

Vladimir Šimović, Vilko Žiljak, Tomislav Kosić  
University of Zagreb, Croatia

---

A new model of analytical function has been developed for modern Business Intelligence Analysis (BIA). The Croatian model is being used for intelligence investigations of various financial events, markets, subjects or entities, and for financial BIA control methods. The financial BIA functional entropy was predicted utilizing the soft computing process (based on fuzzy logic) and various stochastic simulations. This study explains the simulation-modeling concept of the BIA function. The queuing M/M/s model with priorities was used for solving different stochastic simulations. In practice, the BIA simulation model can be used for measuring analytical capacity.

---

Even in the initial stages of establishing a mathematical model, the analytical approach is often too categorical and inflexible to cope with the intricacy and the complexity of real world financial systems such as that in Croatia. In dealing with such complex systems, we face a lot of uncertainty and imprecision. It is therefore necessary to attempt to exploit the human ability to make rational decisions in an uncertain and imprecise environment, and also to use soft computing tolerance for imprecision and uncertainty, in order to achieve an acceptable low cost model. Soft computing is oriented towards fuzzy logic analysis, artificial neural networks, and probability reasoning, including genetic algorithms, chaos theory, and machine expertise. Fuzzy logic addresses imprecision



and approximate reasoning.

Optimal decision-making is often inadequate in achieving the best design for a queuing system and developing information on the behavior of a queuing system such as the analytical information queuing system of the Business Intelligence Analysis (BIA). The BIA system (with data warehousing) is an aid to technology, which provides governmental and financial organizations with end-to-end solutions for managing, organizing, and exploiting financial and other data throughout the enterprise. This technology provides tools to collect information (concerned financial business) in a single organized data repository based on a common set of financial and other business definitions. After explication of the soft computing of the entropy of the modern financial analytical function (based on fuzzy logic), the second part of this work provides a short explanation of the results of analysis and simulation of modern financial analytical function based on stochastic modeling.



First let us examine several basic concepts and notations. If  $\mathcal{I}$  denotes the universal analytical set, then set  $\mathcal{I}$  contains all the possible analytical elements in each particular analytical context or analytical application from which a set can be formed. The process by which individuals from the universal set  $\mathcal{I}$  are determined to be either members or non-members of a set can be defined by a characteristic function. For a given set  $\mathcal{I}$  where  $\mathcal{I} \subseteq \mathcal{I}$ , this characteristic function assigns a value  $m$  to every  $\mathcal{I} \in \mathcal{I}$  such that  $m = 1$  if  $\mathcal{I} \in \mathcal{I}$  or  $m = 0$  if  $\mathcal{I} \notin \mathcal{I}$ .

This function can be generalized such that the values assigned to the elements of the universal set fall within a specified range and indicate the membership grade of these elements in the set in question. Consequently, larger values denote higher degrees of set membership. Such a function is called a membership function and the set defined by it is a fuzzy set. Here are two different kinds of notations. For example, if  $\mathcal{I} = \{ \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4, \mathcal{I}_5 \}$  and  $\mathcal{I} = \{ \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4, \mathcal{I}_5 \}$ , then a membership function for the elements of the set  $\mathcal{I}$  can be dually denoted:  $\mathcal{I} = \{ \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4, \mathcal{I}_5 \}$ , or  $m_1; m_2; m_3; m_4; m_5$ .

Logically, then, a membership function in fuzzy logic can be represented in a similar but slightly different manner. For exam-



ple, if  $\underline{A} = \{x_1 | 0.9; x_2 | 0.7; x_3 | 0.5; x_4 | 0; x_5 | 0\}$ , and the two fuzzy sets are:  $\underline{A} = \{x_1 | 0.9; x_2 | 0.7; x_3 | 0.5; x_4 | 0; x_5 | 0\}$ , or  $\underline{B} = \{x_1 | 0.9; x_2 | 0.7; x_3 | 0.5; x_4 | 0; x_5 | 0\}$ , then a membership function for the elements of the fuzzy set  $\underline{A}$  and  $\underline{B}$  are denoted thus, for a fuzzy set  $\underline{A} : m_A(x_i)$ , a fuzzy set  $\underline{B} : m_B(x_i)$ .

A fuzzy decision is a special type of fuzzy set. The decision in a fuzzy environment (depending on the context) can be viewed as the intersection of fuzzy constraints and fuzzy objective function(s), where the fuzzy objective function is characterized by its membership function, and represents constraints. In contrast to a no fuzzy environment, the decision in a fuzzy environment is defined as the optimal selection of activities that simultaneously satisfy fuzzy objective function and fuzzy constraints. On this basis, the assumption is that the constraints are not interactive, are logical, and correspond to the intersection. By analogy to crisp (not fuzzy) environments and to crisp decision logic, in fuzzy environments we have slightly different decision logic (usually called *Fuzzy decision logic*). A linguistic variable  $x$  is a variable whose values are words or sentences in natural or artificial language. For example, if intelligence is interpreted as a linguistic variable, then its term set  $T_x$ , as the set of its linguistic values, where each of the terms in  $T_x$  is a fuzzy subset of a universe of discourse, say  $U_x$  or because of practical reasons usually  $I_x$ .

There are two rules associated with a linguistic variable: syntactic rule (which defines the well-formed sentences) and semantic rule (by which the meaning of the terms in  $T_x$  may be determined).

Modern BIA (Data contents and information source evaluation system) (the so-called *4x4x2* evaluation system), is an important concept in the BIA model. It can be viewed as a conceptual tool for reducing the entropy of the modern financial analytical function. The classic or OLD (Data contents and information source evaluation system) has one simple criterion, which deals with three linguistic variables for data contents and information source evaluation purposes (low information, unknown=entropy, high information). Modern BIA or NEW (Data contents and information source evaluation system) (4x4x2 evaluation system) has a highly developed criterion, which deals with a minimum of 32 linguistic variables for data contents and information source evaluation purposes.

Table 1. Simplified Modern BIA Data Contents and Information Source Evaluation System

Intelligence Evaluation System	Transformation of The Linguistic Variables				System
Real 4x2 Evaluation System	X4, X3, C4	X2, B4, X1, A4	C3	C2, B3, C1, A3,	B2, B1, A2, A1
Simplified New Evaluation System	vli = 0.1	li = 0.3	u = 0.5	hi = 0.7	vhi = 0.9

For simplicity's sake, we can say that the modern BIA or NEW data contents and information source evaluation system deals (in the worst case) with only five linguistic variables for data contents and information source evaluation purposes (say, **very low information**, **low information**, **unknown = entropy**, **high information**, **very high information**). Suppose that simplification has been made, and that modern BIA data contents and information source evaluation system (4x2) has a criterion system, and that its term set T<sub>NEW</sub>(intelligence), as the set of its linguistic values, is:

$$T_{NEW}(\text{intelligence}) = \text{vli (very low information)} + \text{li (low information)} + \text{u (unknown/entropy)} + \text{hi (high information)} + \text{vhi (very high information)}.$$

Classical BIA data contents and information source evaluation system has a simple criterion system, which deals with only three linguistic variables for data contents and information source evaluation purposes, and where its term set T<sub>OLD</sub>(intelligence), as the set of its linguistic values, is:

$$T_{OLD}(\text{intelligence}) = \text{li (low information)} + \text{u (unknown/entropy)} + \text{hi (high information)}.$$

Table 2. Classical Data Contents and Information Source Evaluation System

Intelligence Evaluation System	Transformation of The Linguistic Variables System		
Real Classical Evaluation System	Weak (Low) Information	Entropy	Strong (High) Information
Simplified Old Evaluation System	li = 0.3	u = 0.5	hi = 0.7

Simplified, analytical information (financial rumours etc.) is composed of at least two or more elementary parts. Let us suppose that it is composed of only two elementary parts  $m_1(x)$  and  $m_2(x)$ , or  $\text{Analytical}_{\text{information}} = [m_1 \ m_2]$ . For the modern or NEW BIA fuzzy decisions system, it is clear that we have 16 (from 25 possible) compound linguistic variables, with a value greater than 0.5 and which represent real analytical information. The maximal analytical success (good fuzzy decision) of the modern or NEW BIA data contents and information source evaluation system (in the worst case) is 64% (or 16/25), and it is much greater than the maximal analytical success of the classic or OLD BIA evaluation system (which is, in the best case, 55.56%, or 5/9). Also, in the BIA source analytical information set is true ( ) or false ( ), with probability  $p_1$  and  $p_2$ . These relations are:  $p_1 + p_2 = 1$ . During the analytical transformation process, compound analytical information set can have these destination probabilities:  $p_{11}, p_{12}, p_{21}, p_{22}$ , which are represented in the relation:  $p_{11} + p_{12} + p_{21} + p_{22} = 1$ .

The analytically interesting transformation process is represented (from the given and taken set of analytical information) with the equation for analytical information contents  $I(X; Y)$ :  $I(X; Y) = H(Y) - H(Y|X)$ , where destination set probabilities are:  $p_{11}, p_{12}, p_{21}, p_{22}$ . All compound probabilities are:  $p_{11}, p_{12}, p_{21}, p_{22}$ .

$$p(X=0/Y=0) = ((1 - A)\bar{C}) / (\bar{A}B + \bar{A}C)$$

Input for the OLD and NEW BIA system is true ( ) or false ( ) analytical information set, with the same probabilities where:  $p_1, p_2$ . The maximal analytical success of the NEW BIA system (in the worst case) is 64%, but for the OLD BIA evaluation, the system maximal success is 55.56%. When  $p_1$  and  $p_2$ , then input probabilities for the OLD BIA systems are:  $p_1, p_2$ .

and for NEW BIA system are:

From the equation for :

for  
Then the analytical entropy is:  
, where is

analytical noise



Consequently, the analytically interesting transformation process is represented (from the given and taken set of analytical information) with the equation for analytical information contents  $I(X; Y) = HT$  (or analytical transformation). Without comparing the degree of analytical noise, it is clear that in the NEW BIA evaluation system analytically transformed relevant information  $H(Y)$  is only 100 analytic cycles 542.28 bit better than OLD BIA system.



We are using here the specific  $M/M/s$  model, which assumes that all inter-arrival times are independently and identically distributed according to an exponential distribution (our input process is Poisson); that all analytical service times are independently and identically distributed according to another exponential distribution (our analytical service process is Poisson); and that the number of servers is  $s$  (any positive integer); however, in the Croatian BIA practice and related analytical function they vary from minimum 1 to maximum 7. With the equal distribution of analytical supply time, with an expected analytical service time of about  $1/\mu$  ( $\mu_n$  is the mean analytical service rate for the overall system, or expected number of clients (data or information) completing analytical service per unit time, and with exponentially distributed inter-arrival time of analytical information at the expected average rate of  $1/\lambda$  ( $\lambda_n$  is mean arrival rate, or expected number of arrivals per unit time), this represents the most simplified type of Markovian analytical system with an assumed infinite analytical capacity ( $Y = \infty$ ), and with priorities in queue discipline (or without supposed FIFO queue discipline). We are currently researching analytical BIA cases in which there are no possibilities for analytical closeness of a multi-channel model of analytical supply function, or when the utilization factor for the analytical service facility is  $\rho_s < 1 \Leftrightarrow \lambda < s\mu$  (because  $\rho_s = \lambda/s\mu$ ).

In the multi-channel model  $M/M/s$  we have a priority sub-system with  $N$  (where  $N = 1, 2, \dots, k$ ) and relative priority classes where  $W_k$  is steady-state or has a total expected waiting time in the analytical system (including service time, or analytical supply time). The steady state expected number of members of priority class  $k$  in the queuing system (including those being analytically served) is  $L_k$ , and it can be explained in this relation:  $L_k = \lambda_k W_k$ , for  $k = 1, 2, \dots, N$ . The expected waiting time in the queue



(excluding service time) for priority class  $k$  is  $W_q(k)$ , and can be explained in this relation:  $W_q(k) = W_k \cdot \rho / \mu$ . The corresponding expected queue length (tail length) is  $L_q(k)$ , and it can be explained:  $L_q(k) = \lambda_k \cdot W_q(k)$ .



We researched behaviour of developed M/M/s model types in relation to the various intensities of analytical traffic (see Table 3).

Table 3. Exploitation levels ( $\rho_s$ ) of various analytical system (M/M/s) types

$\rho_s = \lambda / (s \cdot \mu)$	$m=4$	$m=5$	$m=6$
$\lambda = 4$ and $s = 1$	-	0.800	0.667
$\lambda = 3$ and $s = 1$	0.750	0.600	0.500
$\lambda = 2$ and $s = 1$	0.500	0.400	0.333
$\lambda = 4$ and $s = 2$	0.500	0.400	0.333
$\lambda = 3$ and $s = 2$	0.375	0.300	0.250
$\lambda = 4$ and $s = 3$	0.333	0.267	0.222
$\lambda = 2$ and $s = 2$	0.250	0.200	0.167
$\lambda = 3$ and $s = 3$	0.250	0.200	0.167
$\lambda = 4$ and $s = 4$	0.250	0.200	0.167
$\lambda = 4$ and $s = 5$	0.200	0.160	0.133
$\lambda = 3$ and $s = 4$	0.188	0.150	0.125
$\lambda = 2$ and $s = 3$	0.167	0.133	0.111
$\lambda = 4$ and $s = 6$	0.167	0.133	0.111
$\lambda = 3$ and $s = 5$	0.150	0.120	0.100
$\lambda = 4$ and $s = 7$	0.143	0.114	0.095
$\lambda = 2$ and $s = 4$	0.125	0.100	0.083
$\lambda = 3$ and $s = 6$	0.125	0.100	0.083
$\lambda = 3$ and $s = 7$	0.107	0.086	0.071
$\lambda = 2$ and $s = 5$	0.100	0.080	0.067
$\lambda = 2$ and $s = 6$	0.083	0.067	0.056
$\lambda = 2$ and $s = 7$	0.071	0.057	0.048

We changed exploitation variables of the analytical system ( $\rho_s = \lambda / s \mu$ ), in all combinations for the values:  $\lambda = \{2, 3, 4\}$ ,  $\mu = \{4, 5, 6\}$ , and for models: M/M/1, M/M/2, M/M/3, M/M/4, M/M/5, M/M/6 and M/M/7 (the number of analytical servers varies from 1 to 7, or  $s = \{1, 2, 3, 4, 5, 6, 7\}$ ). We have completed 62 simulation-modelling experiments with different types of M/M/s multi-channel analytical models. Simulation modelling results were successful(see Table 1). We researched the possibility of rational dimensioning and organization of analytical function,



without remaining in a stationary state, of the developed analytical model as well. Now we will be comparing the potential difference between both BIA M/M/s models (NEW and OLD) by stochastic simulations, (first) in a similar experimental situation, and (second) in a minimally different experimental situation. In both situations, we selected the M/M/5 simulation model for NEW and the M/M/3 simulation model for OLD. For both simulation models, we used the same intensity of analytical traffic. Or concretely,  $\lambda_k = 4$ , and  $\mu = 5$  for both simulation experiments (see Table 1). Also, in both situations the numbers of non pre-emptive classes are minimally different. Consequently, for the NEW simulation model we used only ( $N =$ ) 3 different non pre-emptive classes, and for the OLD simulation model we used only ( $N =$ ) 2 different non pre-emptive classes. The mean exponential distribution of the expected analytical service time (or mean analytical service rate for the overall BIA system,  $\mu_n$ ) was the same ( $1/\mu_n = 0.2$ ) for both experimental situations. In the first case, we used almost the same experimental situation for both simulation models (NEW and OLD). We have  $\lambda_k = 4$ ,  $\mu = 5$ , with  $\lambda_{kNEW} = \lambda_1 + \lambda_2 + \lambda_3 = 1.4 + 1.3 + 1.3 = 4$  for NEW simulation model ( $N = 3$ ), and  $\lambda_{kOLD} = \lambda_1 + \lambda_2 = 2 + 2 = 4$  for OLD simulation model ( $N = 2$ ). Results from simulation-modelling experiments were successful. Due to the problem of stochastic convergence, we made nine different series of stochastic simulations after the simulation modelling experiments with: 100000, 50000, 10000, 5000, 1000, 500, 100, 50, and 10 arrivals of analytical data (or information).

The variables tell us that for the first non pre-emptive priority class we have a significantly lower time ( $W_q(1)$  is lower for 91.94%) for the NEW model than for the OLD model, and that we have a significantly lower number ( $L_q(1)$  is lower for 87.10%) for the NEW model than for the OLD model. In the second case, we used a somewhat different experimental situation for both simulation models (NEW and OLD). We have  $\lambda_k = 4$ ,  $\mu = 5$ , with  $\lambda_{kNEW} = \lambda_1 + \lambda_2 + \lambda_3 = 2 + 1 + 1 = 4$  for the NEW simulation model ( $N = 3$ ), and  $\lambda_{kOLD} = \lambda_1 + \lambda_2 = 3 + 1 = 4$  for the OLD simulation model ( $N = 2$ ). Results from simulation-modelling experiments were successful. The variables tell us that for the first non pre-emptive priority class we have a significantly lower time ( $W_q(1)$  is lower for 92.39%) for the NEW model than for the OLD model, and that we have a significantly lower number ( $L_q(1)$  is lower for 88.52%) for the NEW model than for the OLD model. The NEW model is definitely superior.





From Table 3 and Figure 1, we can easily find from the specific exploitation level (for example it is 30%, or  $p = 0,3$ ), and the specific number of analytical servers (for example  $s = 3$ ) the maximum intensity of analytical traffic (for this example, it ranges from  $4/5$  to  $4/4$ , or  $\lambda/\mu = [4/5, 4/4]$ ). Or in the opposite direction, we can find the specific exploitation level (it can be from 20% up to 25%) from the specified maximum intensity of analytical traffic (say it is in the interval  $\lambda/\mu = [4/5, 4/4]$ ), and the specific number of analytical servers (say it is  $s = 4$ )



In the NEW BIA evaluation system, the analytically transformed relevant information is only 100 analytic cycles 542.28 bit better than in the OLD evaluation system. In the worst case (for the NEW BIA system), the maximal analytical success of the NEW BIA system is 64%, but for the OLD system, it is maximally 55.56%. This conclusion and study provide a solid base for future BIA modeling and simulation process. The benefits of this new analytical model of BIA function consist in the simple method utilized for measuring analytical capacity and capability of analysis



Andrews, P., P., JR., Peterson, M., B. (1990).  
Palmer Enterprises, Loomis, California.

Brandt, S. (1999). Data Analysis: statistical and computational methods for scientists and engineers - 3rd ed., Springer-Verlag, New York Inc., New York.

Han, J. (1999). Characteristic Rules  
Oxford University Press.

Hillier, F., S., Lieberman, G., J., (1995), Introduction to Operations Research 8 International Editions, McGraw-Hill, Inc., Singapore.

Simovic V., Zrinusic Z., Skugor M., (1999), An Application Of Profound Financial Knowledge Discovery Model, Papers and Proceedings - Euro Working Group on Financial Modelling - 25th Meeting, Vienna, session 12.

Zadeh, L., A., (1996), Fuzzy logic, Neural Networks and Soft Computing in Computational Intelligence: Soft Computing



- and Neuro-Fuzzy Integration with Applications, ed. By O. Kaynak, L. A. Zadeh, B. Tuksen, I. Rudas, Springer Verlag, NATO ASI Series, under publication.
- Zadeh, L., A., et all, (1974), Fuzzy sets and their applications to cognitive and decision processes, Academic Press, Inc., Chestnut Hill, MA.
- Ziljak, V., (1982), Simulation with computer (in Croatian language), textbook, Zagreb.