ASSESSMENT OF MARKET REACTION ON THE SHARE PERFORMANCE ON THE BASIS OF ITS VISUALIZATION IN 2D SPACE

Ingrida VAICIULYTE¹, Zivile KALSYTE², Leonidas SAKALAUSKAS³, Darius PLIKYNAS⁴

¹Faculty of Business and Technologies, Siauliai State College,
Ausros al. 40, LT–76241 Šiauliai, Lithuania

²Faculty of Electrical and Electronics Engineering, Kaunas University of Technology,
Studentu g. 50, LT–51367, Kaunas, Lithuania

³Faculty of Fundamental Sciences, Vilnius Gediminas Technical University,
Saulėtekio al. 11, LT–10223, Vilnius, Lithuania

⁴Institute of Mathematics and Informatics, Vilnius University,
Akademijos g. 4, LT–08663, Vilnius, Lithuania

E-mails: ¹i.vaiciulyte@svako.lt (corresponding author); ²kalsyte@yahoo.com;

³leonidas.sakalauskas@mii.vu.lt; ⁴dplikynas@gmail.com

Received 22 April 2016; accepted 18 January 2017

Abstract. This paper provides a new methodology for company assessment besides other traditional assessment measures such as share price or forecasts of the analysts. It is suggested to assess the market reaction on change in share price via using graphical approaches. Investors buy shares with the expectation that its price will rise in the future. But sometimes expectations don't coincide with reality and then shares are sold. This work has been taken into account in the asymmetry between expectations of investors and results. In order to identify the position of a company in 2D space, the paper uses classification algorithm of random forests with data on change in share price during the period of the year in the inputs, and the forecasts of analysts, i.e., whether a price will increase or decrease, for the same year in the outputs. Thus, two clusters of companies are seeking to represent: one of the companies whose changes in share price coincide with investors' expectations, and another one – on the contrary. This method can be useful to investors, for whom it is important to identify the market reaction about companies from the whole industry or its branches and analyze its trend.

Keywords: skew t distribution, maximum likelihood method, random forests, forecasting, rating, management, assessment.

JEL Classification: C610, G17.

Introduction

A number of research (see Angelovska 2016; Coussement, Van den Poel 2009; Liang et al. 2015; Lin et al. 2014; Price et al. 2011) have shown that the trend of share markets is not entirely random and can be forecasted, however, various random factors, for ex-

ample, significant price growth or recession periods, change of national currency, expectations of investors, assessments by financial analysts, etc., can influence the financial forecasts. Recommendations of analysts, engaged in assessment of financial market elements, are treated controversially. On one hand, they are commonly based on technical analysis (share price is forecasted by attempting to determine the trends and statistical analogies, following the historical price volatility). On the other hand, recommendations are biased – brokers and brokerage companies tend to highlight positive information and conceal negative one, in order to attract and maintain investment banking business. Ongoing declines of share markets show that indicators of forecasts and recommendations, published by analysts, not always reflect reality (Brown et al. 2015; Peasnell et al. 2016). Barber et al. (2006) researched criteria under which analysts' recommendations should be classified to buy, hold shares expecting future rise or to sell shares at a profit now. They found out that there is a need to discipline the brokers, who commonly tend to recommend buying shares. What is more, the aforementioned authors concluded that the analysts of investment banking companies are more conservative for long-term forecasts than the analysts from independent research companies. Thus, the forecasts of analysts can be used only as additional information, and not as final proposal to buy or sell the shares of specific business.

The paper proposed the method of assessment of market reaction on the share performance enabling to assess the business viability by market participants' view. This problem is relevant to investors, for whom it is important to select the most viable companies from the whole industry without spending too much time. It is explored what companies are assessed by market positively, i.e., as having perspective, and what – negatively, not only on the basis of share prices, analysts' forecasts, but also according to visualization of each business in respect of other companies. The experiments are made with data of companies from Health Care industry. Current share prices and ratings, provided by analysts, are used for describing market assessment. If investors assess business as viable, share price will be high – more investors will want to buy the shares of that company than to sell, thus, share price will increase. Current share price reflects the assessment of company from investor's position. Multivariate skew t distribution was selected for description of investors' opinion about business results (see Sakalauskas et al. 2015). The parameters of probability distribution are assessed by maximum likelihood method. By using random forests classification algorithm, research of market reaction is carried out, and business position in 2D space is identified. During forecasting, proximity matrices are stored. They reflect the similarity of one company with other companies, according to this parameter. The parameter, reflecting market reaction, is selected by using the rating data (see Zavadskas, Turskis 2011). This method can be useful to investors, for whom it is important to identify the market reaction about companies from the whole industry.

1. Probabilistic model

It was found out that some financial data are leptokurtic, characterized by heavy-tailed values, thus, many authors showed that the normal distribution, so often used for analysis of economic indicators, is not suitable (Belovas *et al.* 2006; Fischer *et al.* 2007; Ka-

basinskas *et al.* 2009; Rachev, Mittnik 2000). By the way, this law ignores asymmetry between expectations of investors (probability to get profit) and results. Investors buy shares with the expectation that its price will rise in the future. But sometimes expectations don't coincide with reality and then shares are sold. Therefore, multivariate skew *t* distribution was selected for description of investors' opinion about business results, as data therein are described by normal distribution with random mean vector and random dispersion by taking into account asymmetry and heavy-tailed values (Panagiotelis, Smith 2008; Theodossiou 1998). Sakalauskas *et al.* (2015) proved that the difference between skew *t* distribution mean and data sample mean shows the difference between investors' expectations and business value. If this difference is positive, it means that investors overestimate business shares, if negative – investors' expectations are pessimistic. It is noticed that this interpretation of probabilistic model reflects the real market trend.

2. Assessment of skew t distribution parameters

Let us consider multivariate skew *t* distribution selected for description of investors' opinion about business results. Skew *t* distribution is defined by hierarchical statistical model through multivariate normal distribution, the mean vector of which is also distributed, according to normal distribution, when covariate matrices of both these distributions depend on parameter, distributed, according to gamma law.

Suppose that multivariate random vector $X \in \Re^d$, d = 2 – dimension, is described by probability density:

$$p(x|\mu,\Omega,\Theta,\alpha,\eta) = 2 \cdot \int_{0}^{\infty} \int_{\eta \cdot (z-\mu) \ge 0} f(x|z,t,\Omega) \cdot f(z|\mu,t,\Theta) \cdot \varphi(t|\alpha) dz dt, \qquad (1)$$

 $X \sim N(z,\Omega)$ – multivariate random vector, distributed according to normal distribution

$$z \sim N\left(\mu, \frac{\Theta}{2t}\right)$$
, i.e., with density.

$$f(x|z,t,\Omega) = (t/\pi)^{\frac{d}{2}} \cdot |\Omega|^{-\frac{1}{2}} \cdot e^{-t \cdot (x-z)^T \cdot \Omega^{-1} \cdot (x-z)}, \qquad (2)$$

where μ , Ω , Θ , α , η is a vector of unknown parameters:

 μ – mean vector,

 Ω , Θ – positively defined covariance matrices,

 α – the extinction degree,

 η – scalar, enabling to establish the area of preferences (priorities).

While random variable t is distributed according to gamma distribution with parameter α (see Azzalini, Capitanio 2003), i.e., with density:

$$\varphi(t|\alpha) = \frac{t^{\frac{\alpha}{2}-1}}{\Gamma(\alpha/2)} \cdot e^{-t} . \tag{3}$$

The parameters of the defined skew *t* distribution were estimated by optimization of the likelihood function (see Azzalini, Genton 2008; Sakalauskas, Vaiciulyte 2012). Below, there is a statistical modelling scheme of maximum likelihood:

- suppose that $X^1, X^2, ..., X^K, K = 252$, are independent, uniformly distributed random variables with density function $p(x \mid \mu, \Omega, \Theta, \alpha, \eta)$;
- after calculating the product of density functions, likelihood function is obtained:

$$L(\mu, \Omega, \Theta, \alpha, \eta) = \prod_{i=1}^{K} p(X^{i} | \mu, \Omega, \Theta, \alpha, \eta);$$
(4)

- to facilitate the calculations, the function $L(\mu, \Omega, \Theta, \alpha, \eta)$ is logarithmized;
- maximum likelihood estimates are selected as values of parameters μ , Ω , Θ , α , η at which the function $L(\mu, \Omega, \Theta, \alpha, \eta)$ has the highest value:

$$L(\mu, \Omega, \Theta, \alpha, \eta) \rightarrow \underset{\mu, \Omega, \Theta, \alpha, \eta}{\operatorname{arg max}};$$
 (5)

– maximum of the function $\ln L(\mu, \Omega, \Theta, \alpha, \eta)$ is obtained by equating the respective likelihood function derivatives to zero and solving the resulting system of equations (Sakalauskas, Vaiciulyte 2012), e.g.,

$$\frac{\partial L(\mu, \Omega, \Theta, \alpha, \eta)}{\partial \mu} = \sum_{i=1}^{K} \frac{\partial p(X^{i} | \mu, \Omega, \Theta, \alpha, \eta)}{\partial \mu} \cdot \frac{1}{p(X^{i} | \mu, \Omega, \Theta, \alpha, \eta)} = 0.$$
 (6)

Calculation of maximum likelihood estimates of the parameter, when using numerical methods is an iterative process, where initial estimate values are selected and certain number of iterations is performed until the estimates of adjacent iterations start being characterized by sufficiently low difference. Assessment of multivariate probability model parameters is quite relevant issue in the world, while the results obtained might be widely used in the real life – for financial models, forecasting, cartographic maps, localization of infectious agents, assessing the probabilities of rare events (e.g., deaths, accidents), etc.

3. Random forests

Random forests (RF) classification algorithm (see Breiman 2001) consists of the set of individual decision trees. Each decision tree is prepared by using random two thirds of data from training sample, while the remaining data out-of-bag (OOB) is used for testing. Error for the tested part of data decreases by increasing the number of decision trees (Archer, Kimes 2008; Breiman 2001; Chan, Paelinckx 2008; Genuer *et al.* 2010; Hapfelmeier, Ulm 2013; Wu *et al.* 2014).

Importance of the features is measured by using OOB data. The selected features depend on the number of data, entering OOB. Therefore, while selecting the most important features, 7 different data amounts, entering OOB, are taken. The number of features, falling into OOB, N is selected in the following way: $\sqrt{N} - 2$, $\sqrt{N} - 1$, \sqrt{N} , $\sqrt{N} + 1$, $\sqrt{N} + 2$. The mean of importance of features of each OOB data amount is calculated. By using backward elimination, 5% of the least important features are eliminated until only 1

feature is left. It is observed which set of features was characterized by the lowest OOB error. Then, these features are considered as the most important (Guyon 2008; Kalsyte *et al.* 2013; Verikas *et al.* 2011).

4. Data

The data include the companies (130 in total), characterized by upgrade and downgrade trends of share prices in 2010–2012, from the following sectors of Health Care industry (9 in total): Diagnostic Substances (8), Drug Generic (4), Drug Manufactures Major (5), Drug Manufactures Other (9), Health Care Plans (6), Home Health Care (2), Med. Laboratories and Research (7), Med. Instruments and Supplies (20), Spec. Health Services (3).

Seeking to describe the share prices of a company during the year, primary data processing is used, i.e., sample mean is deducted from the available data and it is divided by standard deviation. In the case considered, this data centering and normalization is required to facilitate the work of the program, otherwise the program would have to optimize under very different scale parameters. After solving the minimization task, it is returned to original scales.

5. Methodology

- 1. The following skew *t* distribution parameters are calculated from closing share prices of 2011;
- 2. The data of clause 1 are used as inputs in RF algorithm. Output data describe whether share prices will rise or fall. Proximity matrices are stored during classification. Apart from a predictor, the RF software also produces a data proximity matrix. To obtain the matrix, for each tree grown, the data are run down the tree. If two observations occupy the same terminal node of the tree, is increased by one. At the end of the run, the proximities are divided by the number of trees. Data proximities possess an important property meaning that only variables used by the forest contribute to the proximity values. Data proximity matrices are used for exploration of data by mapping them in 2D space (see Borg, Groenen 2005);
- 3. The *t*-distributed stochastic neighbor embedding (t-SNE) algorithm was used to perform the mapping in this work (see van der Maaten, Hinton 2008). T-SNE algorithm represents similarities of data points as conditional probabilities. The similarity of x_i to x_i in a high-dimensional space is defined as the conditional probability $p_{j|i}$ that x_j is picked by x_i as its neighbor when neighbors are picked in proportional to their probability density defined by a Gaussian centered on x_i (see van der Maaten, Hinton 2008):

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma^2)},$$
(7)

where σ is a parameter and the values of $p_{i|i}$ are set to zero. Hence, it works best

when the data will be embedded on two or three dimensions (see van der Maaten, Hinton 2008). By using t-SNE dimensional reduction business position is transferred from multidimensional space to 2D and is displayed in a 2D plane. This procedure is repeated for each parameter of skew *t* distribution individually;

4. The obtained graphs are analyzed and it is observed which parameter of skew *t* distribution is the best to reflect the upgrade and downgrade trends of share prices (see Zavadskas, Turskis 2011).

6. Results

Upon assessing the parameters of probability distribution, the obtained coordinates of market reaction enable visualization of business position in respect of other companies. The results show that density integration parameter of skew t distribution has the greatest influence in the model, developed, and seeking to assess priority areas and describe the expectations of investors in more details. As area parameter η is the key indicator of market reaction research, Figure 1 shows the results of examination of this parameter. The companies, which were characterized by downgrade trend in 2010–2012, are depicted as squares, while the companies with upgrade trend — as dots. Coordinates and trend value (downgrade, upgrade) are given by calculating proximity matrices from the RF algorithm by t-SNE algorithm. RF method allows data convert into abstract variables, retaining only the metric (distances) between objects. The method created in this paper allows visualized data of individual economic sectors (companies). Unrated companies are not included in the graph.

Figure 1 shows that in 2010 two clusters are distinguished: the companies with downgrade trend and the companies with upgrade trend and in the next pictures we can see changes of this trend in 2011 and 2012 It is noticed that the ratings of companies from the *Medical Laboratories and Research* sector were characterized by the most significant fall in respect of other sectors, while underestimation of companies is not so significant. By the way, the ratings of companies from *Medical Instruments and Supplies* and *Drug Manufactures Major* sectors, which were assessed positively, are significantly better than the ones of others. Both the downgrade and upgrade trends of the *Diagnostic Substances* sector match and are similar as well as upgrade trend of *Drug Generic* and *Drug Manufactures Other* sectors.

Companies that are further away from intersection boundary of the upgrade and downgrade clusters are more dependent on analysts' opinion, i.e., the expectations of investors and analysts are coincident. However, those companies that are closer to the intersection boundary of the upgrade and downgrade clusters need more data, to forecast their share price, maybe, the financial statements or other, i.e., they are not so heavily dependent on analysts' forecasts. This model is not appropriate for all sectors, because they work in an unstable way. Figure 1 shows, that the downgrade and upgrade trends of the *Diagnostic Substances* sector are at the same point. So this model is not appropriate for the *Diagnostic Substances* sector, because it works in an unstable way. Share price upgrades points of *Drug Generic* and *Drug Manufactures Other* sectors near downgrade points of other sectors share price, so this model can be unstable too.

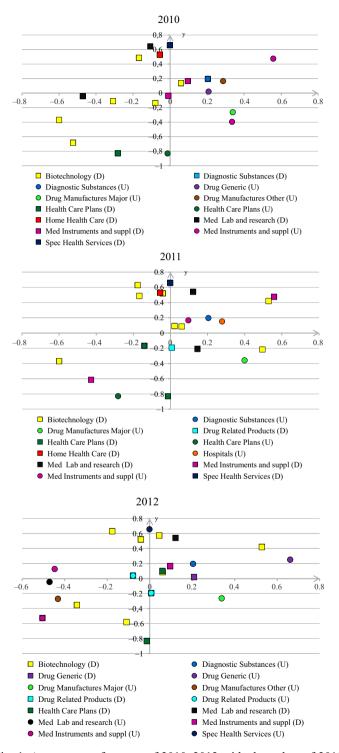


Fig. 1. Assessment of sectors of 2010-2012 with share data of 2011

Conclusions

This paper suggests evaluation of market reaction on the share of firms traded in the share markets via using graphical approaches. It is argued that if share market investors view that the business is viable, they will increase in share price, otherwise its share price will decrease. The method of business visualization for identifying the plane coordinates of the business on the basis of share data and ratings, provided by analysts was created in this paper. The developed techniques were experimentally tested using data concerning companies from different sectors of the health-care industry. Estimates of skew t distribution parameters were used for classifying the companies into upgrade and downgrade clusters. So the share prices and distribution parameters, describing the change in a share price, were used as input data, and in the outputs forecasts of financial analysts were used (where the share price of a company will rise or fall). Proximity matrices were preserved during the classification. Data proximities in the original space were represented by using t-SNE visualization. Then the companies' position was transferred from a multidimensional space to two-dimensional and was displayed in a plane. Companies with the same rating are placed close together in a plane. Also this method allows visualized data of individual economic sectors (companies).

The results of market reaction research show that the figure of parameter of skew t distribution of density integration area is more informative seeking to define the priority area and describe the expectations of investors. Therefore one can conclude that the forecast about the rise or fall of share of the price of a company is forming investor's expectations. Therefore assessments of financial analysts influence the financial situation of business — by raising the ratings they attract investors, and by downgrading — worsen the situation. Thus, financial analysts and investor's opinions are correlated with each other. This method can be useful to investors, for whom it is important to identify the market reaction about companies from the whole industry or its branches and analyze its trend.

The computation time is the primary limitation of this research. The calculation takes too long applying this model in real time to a large number of companies. The research model could be extended by different activities companies and compare the obtained results not only with big corporations, but also with private limited liability companies. To check the model performance in period of crisis and how react to changes of economic cycle is also recommended for further research.

References

Angelovska, J. 2016. Large share movements, reasons and reaction, *Management* 21(1): 1–17. Archer, K. J.; Kimes, R. V. 2008. Empirical characterization of random forest variable importance measures, *Computational Statistics and Data Analysis* 52(4): 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015

Azzalini, A.; Capitanio, A. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65: 367–389. https://doi.org/10.1111/1467-9868.00391

Azzalini, A.; Genton, M. G. 2008. Robust likelihood methods based on the skew-t and related distributions, *International Statistical Review* 76(1): 106–129. https://doi.org/10.1111/j.1751-5823.2007.00016.x

Barber, B. M.; Lehavy, R.; McNichols, M.; Trueman, B. 2006. Buys, holds, and sells: the distribution of investment banks' stock ratings and the implications for the profitability of analysts' recommendation, *Journal of Accounting and Economics* 41(1–2): 87–117. https://doi.org/10.1016/j.jacceco.2005.10.001

Belovas, I.; Kabasinskas, A.; Sakalauskas, L. 2006. A study of stable models of stock markets, *Information Technology and Control* 35(1): 34–56.

Borg, I.; Groenen, P. J. F. 2005. *Modern multidimensional scaling: theory and applications*. New York, NY: Springer-Verlag.

Breiman, L. 2001. Random forests, *Machine Learning* 45: 5–32. https://doi.org/10.1023/A:1010933404324

Brown, L. D.; Call, A. C.; Clement, M. B.; Sharp, N. Y. 2015. Inside the "black box" of sell-side financial analysts, *Journal of Accounting Research* 53(1): 1–47. https://doi.org/10.1111/1475-679X.12067

Chan, J. C. W.; Paelinckx, D. 2008. Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery, *Remote Sensing of Environment* 112(6): 2999–3011. https://doi.org/10.1016/j.rse.2008.02.011

Coussement, K.; Van den Poel, D. 2009. Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers, *Expert Systems with Applications* 36: 6127–6134. https://doi.org/10.1016/j.eswa.2008.07.021

Fischer, M.; Horn, A.; Klein, I. 2007. Tukey-type distributions in the context of financial data, *Communications in Statistics – Theory and Methods* 36(1): 23–35. https://doi.org/10.1080/03610920500476382

Genuer, R.; Poggi, J. M.; Tuleau-Malot, C. 2010. Variable selection using random forests, *Pattern Recognition Letters* 31(14): 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Guyon, I. 2008. Practical feature selection: from correlation to causality, *NATO Science for Peace and Security*, *Series D: Information and Communication Security* 19: 27–43. https://doi.org/10.3233/978-1-58603-898-4-27

Hapfelmeier, A.; Ulm, K. 2013. A new variable selection approach using random forests, *Computational Statistics and Data Analysis* 60: 50–69. https://doi.org/10.1016/j.csda.2012.09.020

Kabasinskas, A.; Rachev, S.; Sakalauskas, L.; Sun, W.; Belovas, I. 2009. Alpha-stable paradigm in financial markets, *Journal of Computational Analysis and Applications* 11(3): 642–688.

Kalsyte, Z.; Verikas, A.; Bacauskiene, M.; Gelzinis, A. 2013. A novel technique to design an adaptive committee of models applied to predicting company's future performance, *Expert Systems with Applications* 40(6): 2051–2057. https://doi.org/10.1115/1.860182_ch11

Liang, D.; Tsai, C. F.; Wu, H. T. 2015. The effect of feature selection on financial distress prediction, *Knowledge-Based Systems* 73: 289–297. https://doi.org/10.1016/j.knosys.2014.10.010

Lin, F.; Liang, D.; Yeh, C. C.; Huang, J. C. 2014. Novel feature selection methods to financial distress prediction, *Expert Systems with Applications* 41(5): 2472–2483. https://doi.org/10.1016/j.eswa.2013.09.047

Panagiotelis, A.; Smith, M. 2008. Bayesian density forecasting of intraday electricity prices using multivariate skew *t* distributions, *International Journal of Forecasting* 24(4): 710–727. https://doi.org/10.1016/j.ijforecast.2008.08.009

Peasnell, K.; Yin, Y.; Lubberink, M. 2016. Analysts' stock recommendations, earnings growth and risk, *Accounting and Finance* (in press). https://doi.org/10.1111/acfi.12202

Price, R. A. III; Sharp, N. Y.; Wood, D. A. 2011. Detecting and predicting accounting irregularities: a comparison of commercial and academic risk measures, *Accounting Horizons* 25(4): 755–780. https://doi.org/10.2139/ssrn.1546675

Rachev, S. T.; Mittnik, S. 2000. Stable Paretian models in finance. New York, NY: Wiley.

Sakalauskas, L.; Kalsyte, Z.; Vaiciulyte, I.; Kupciunas, I. 2015. The relationship between the transparency in provision of financial data and the change in investors' expectations, *Inzinerine Ekonomika – Engineering Economics* 26(1): 33–39. https://doi.org/10.5755/j01.ee.26.1.4061

Sakalauskas, L.; Vaiciulyte, I. 2012. Maximum likelihood estimation of multivariate skew t-distribution, in *Proceedings of the 1st International Conference on Operations Research and Enterprise Systems*, 4–6 February 2012, Algarve, Portugal, 200–203.

Theodossiou, P. 1998. Financial data and the skewed generalized *t* distribution, *Mathematical Science* 44(12): 1650–1660. https://doi.org/10.2139/ssrn.65037

van der Maaten, L. J. P.; Hinton, G. 2008. Visualizing high-dimensional data using t-SNE, *Journal of Machine Learning Research* 9: 2579–2605. [online], [29 March 2015]. Available from Internet: https://lvdmaaten.github.io/publications/papers/JMLR 2008.pdf

Verikas, A.; Gelzinis, A.; Bacauskiene, M. 2011. Mining data with random forests: a survey and results of new tests, *Pattern Recognition* 44(2): 330–349. https://doi.org/10.1016/j.patcog.2010.08.011

Wu, J. L.; Yu, L. C.; Chang, P. C. 2014. An intelligent stock trading system using comprehensive features, *Applied Soft Computing* 23: 39–50. https://doi.org/10.1016/j.asoc.2014.06.010

Zavadskas, E. K.; Turskis, Z. 2011. Multiple criteria decision making (MCDM) methods in economics: an overview, *Technological and Economic Development of Economy* 17(2): 397–427. https://doi.org/10.3846/20294913.2011.593291

Ingrida VAICIULYTE, Assoc. Prof Dr at Siauliai State College, Faculty of Business and Technologies, Lithuania. Her research interests focus on operation research and optimization, statistical modeling and analysis, stochastic programming.

Zivile KALSYTE, PhD student at Kaunas University of Technology, Faculty of Electrical and Electronics Engineering, Lithuania. She is a head of sales in JSC Topkodas. Research interests include financial economics and business continuity management.

Leonidas SAKALAUSKAS, Prof Habil. Dr at Vilnius Gediminas Technical University, Faculty of Fundamental Sciences, Lithuania. He is a president of Lithuianian Operational Research Society. His scientific interests focus on stochastic optimization, queueing theory and other fields of operation research.

Darius PLIKYNAS, Prof Dr at Vilnius University, Institute of Mathematics and Informatics, Lithuania. His research interests include, but are not limited to various computational intelligence, complexity and agent-based simulation approaches in social sciences.