

3. Anokhin S., Wincent J., Parida V., Chistyakova N., Oghazi P. Industrial clusters, flagship enterprises and regional innovation. *Entrepreneurship & Regional Development*, 2018, doi: 10.1080/08985626.2018.1537150
4. Информационный ресурс СПАРК. URL: <http://www.spark-interfax.ru/> (дата обращения: 11.07.2018).
5. Антонова И.С., Пчелинцев Е.А. Динамическая оценка эффективности диверсификации экономики моногорода (на примере Кемеровской области) // Регион: экономика и социология, 2018. - №1. – с. 271-287

## ОЦЕНКА ЭФФЕКТИВНОСТИ ПРЕДПРИЯТИЙ ХИМИЧЕСКОЙ ПРОМЫШЛЕННОСТИ РОССИИ В РАЗРЕЗЕ ФОРМ СОБСТВЕННОСТИ: DEA ПОДХОД И ИНДЕКС МАЛМКВИСТА

*Булыкина А.А., Михальчук А.А., Спицын В.В.  
(Томский политехнический университет)  
e-mail: anastasiya.bulykina@mail.ru*

## EVALUATION OF THE EFFICIENCY OF THE RUSSIAN CHEMICAL INDUSTRY ENTERPRISES BY TYPE OF OWNERSHIP: DEA APPROACH AND MALMQUIST INDEX

**Abstract.** This document is devoted to the analysis of the effectiveness of Russian enterprises for the production of chemical products by type of ownership (Russian, Foreign and Joint) for 2012-2016 period. Revenue, Fixed assets and Payroll for 297 enterprises with various type of ownership were chosen as main finance indicators for analysis. Differences between types of ownership were investigated by implementation of statistical criteria, which enable to estimate significance of dissimilarities. To assess effectivity of enterprises' activity we utilized indicators obtained by DEA method, which classifies object as effective if it has the biggest output with the smallest inputs. Hence, Malmquist index was applied to researched indicators for estimation of 2-year progress for each couple of years in row.

**Keywords:** Russian industry, technical efficiency, finance indicators of enterprises, ANOVA, DEA, MPI, crisis, economic sanctions.

В рамках данной работы было проведено исследование деятельности Российских предприятий подразделения DG «Химическое производство» в разрезе форм собственности: РО – российская, ФО – иностранная и JO – совместная. Целью работы является оценка эффективности химических предприятий в разрезе форм собственности за период 2012–2016 гг.

Источниками данных для анализа выступает бухгалтерская отчетность предприятий, полученная из информационной системы СПАРК [1]. Период исследования: 2012-2016гг. В рамках настоящего исследования анализируются следующие важнейшие показатели финансовой отчетности предприятий:

1. выручка (В);
2. основные средства (ОС);
3. оплата труда (ОТ).

В анализируемую базу данных вошли предприятия, которые имеют финансовую отчетность по трем показателям (В, ОС, ОТ) за все отчетные периоды 2012-2016 гг., а также предприятия, финансовые показатели которых не ниже установленных пороговых значений (100 млн. руб. для В, 36 млн. руб. для ОС и 7,5 млн. руб. для ОТ). Таким образом полученная база данных содержит:

1. 226 химических предприятий в РС.
2. 23 химических предприятий в СС.
3. 48 химических предприятий в ИС.

Также для анализируемых показателей была учтена инфляция за исследуемый период. Например, накопленная инфляция в 2014 году в ценах 2012 года составляет примерно 19%, в то время как в 2016 году в ценах 2012 года – примерно 41%.

Сформированная таким образом база данных была использована для исследования различий предприятий в разрезе ФС. С помощью критериев дисперсионного анализа были оценены различия между ФС за каждый год для основных финансовых показателей (В, ОС и ОТ) и показателей эффективности (ТЕ, МРІ), а также рассмотрена динамика каждого показателя в разрезе ФС.

В настоящее время метод DEA (Data Envelopment Analysis) представляет собой развитую методологию сравнительной оценки эффективности функционирования различных производственных объектов по широкому набору входных и выходных показателей их деятельности [2]. Согласно методу DEA, эффективность трактуется как отношение взвешенной суммы выходных параметров (результатов, выгод) к взвешенной сумме входных параметров (ресурсов, затрат), что позволяет классифицировать объекты как эффективные только в том случае, когда они производят наибольшие выходы при наименьших входах и, таким образом, дает возможность определять эффективные объекты и относительную меру неэффективности остальных [3]. DEA располагает все эффективные предприятия на линии фронта эффективности, а неэффективные внутри фронта. Чем ближе к фронту эффективности расположен предприятие, тем выше значение относительной эффективности его управленческой деятельности. Оно будет удовлетворять условию  $\leq 1$ , причем значение 1 указывает точку на границе и, следовательно, выделяет эффективное предприятие. Для оценки относительной технической эффективности (ТЕ) деятельности предприятий с помощью DEA использована модель с переменным эффектом от масштаба VRS, ориентированная как на вход (ТЕвх), так и на выход (ТЕвых).

Помимо статической оценки ежегодного показателя эффективности предприятия, также была рассмотрена краткосрочная 2-летняя динамическая характеристика эффективности - индекс Малмквиста (МРІ) [4]. Значения  $МРІ < 1$ ,  $МРІ = 1$  и  $МРІ > 1$  говорят соответственно о снижении, постоянстве или увеличении эффективности предприятия в течение исследуемого периода. Аналогично [2] была рассмотрена также долгосрочная динамика МРІ, которая оценивалась линейным трендом  $\alpha * t + b$  ( $\alpha > 0$  определяет прогресс, а  $\alpha < 0$  – регресс), что с учетом среднего  $МРІ_{ср}$  на всем периоде позволило провести номинальную кластеризацию компаний в координатах ( $МРІ_{ср}$ ,  $\alpha$ ). При этом,  $МРІ_{ср} > 1$  означает "с приростом технической эффективности в текущем периоде", а  $\alpha > 0$  означает "с тенденцией улучшения МРІ в будущем" или "с тенденцией улучшения прироста технической эффективности в будущем".

Для статистического анализа показателей ТЕ и МРІ использовался дисперсионный и кластерный анализы по следующей методике [5–7]:

1. Тестирование распределения исследуемого показателя на соответствие нормальному закону распределения и определение приоритетных критериев анализа;
2. Дисперсионный анализ исследуемого показателя (проверка статистической значимости различий между разными формами собственности ежегодно и между годами для каждой формы собственности);
3. Кластерный анализ объектов (однородность) по совокупности показателей;
4. Экономическая интерпретация выявленных различий.

Основные финансовые показатели химических предприятий в разрезе ФС представлены в таблице 1.

Таблица 1. Средние и медианы В, ОС и ОТ, 2012–2016 гг., млрд р.

		В			ОС			ОТ		
		2012	2014	2016	2012	2014	2016	2012	2014	2016
Среднее	РС	4,754	4,618	4,802	1,76	1,85	1,938	0,34	0,361	0,42
	ИС	3,925	4,25	4,57	0,892	0,863	1,138	0,239	0,258	0,371
	СС	4,434	4,15	4,929	1,379	4,081	3,589	0,282	0,308	0,394
Медиана	РС	1,231	1,066	1,327	0,26	0,277	0,284	0,134	0,146	0,158
	ИС	1,626	1,836	1,976	0,489	0,409	0,351	0,096	0,105	0,13
	СС	1,846	2,047	1,963	0,47	0,539	0,521	0,093	0,117	0,13

**Анализ эффективности предприятия за период 2012–2016 гг.** Был применен метод DEA, модель с переменным эффектом от масштаба VRS, ориентированная как на вход (ТЕвх), так и на выход (ТЕвых). В качестве входных параметров выбраны показатели ОТ и ОС, в качестве выходного – В.

В ходе проверки рассматриваемых выборок с помощью  $\chi^2$ -критерия Пирсона были выявлены высоко значимые отличия от нормального распределения ( $p < 0,0005$ ), что предполагает при оценке уровней значимости различий предприятий 3-х ФС использовать непараметрические критерии.

Согласно непараметрическому критерию Краскела-Уоллиса для показателя ТЕвх предприятия по всем ФС различаются незначимо для всех лет, за исключением 2014 и 2016 года, где FO статистически значимо превышает RO ( $0,005 < p \approx 0,04 < 0,05$ ). Для показателя ТЕвых различия более выражены, предприятия FO значительно выше предприятий RO и JO за весь исследуемый период: различия между FO и RO высоко значимые для всех лет, а для JO и FO статистически значимые только в 2014гг. ( $0,005 < p \approx 0,02 < 0,05$ ). В обоих случаях предприятия в FO либо эффективнее российских, либо не хуже российских (рис. 1).

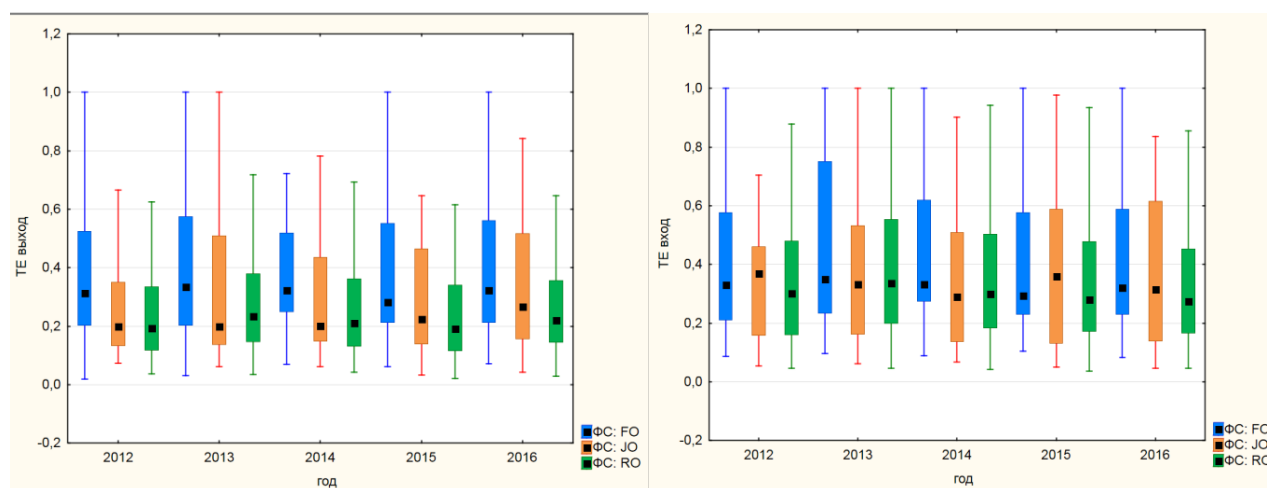


Рисунок 1. Непараметрические характеристики ТЕвых и ТЕвх по годам в разрезе ФС

Результаты, полученные в ходе сравнения эффективности предприятий по химическому производству в разрезе форм собственности показывают неоднородность ТЕвых и Твх. Поэтому для более детального анализа эффективности предприятий был проведен кластерный анализ по совокупности показателей ТЕвых и ТЕвх усредненных за период 2012-2016гг. В результате кластеризации методами k-means и иерархической кластеризации (с помощью метода Варда и расстояния Евклида) было получено разбиение 297 предприятия на 9 кластеров (рис. 2).

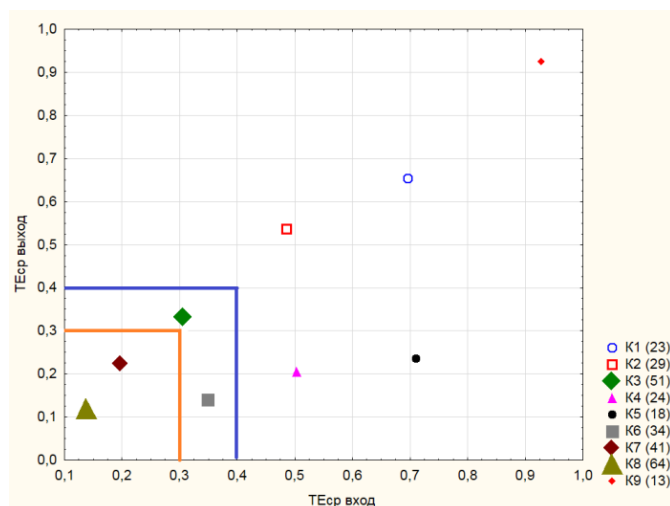


Рисунок 2. Диаграмма рассеяния кластеров по ТЕвх и ТЕвых

В соответствии с критерием Краскела –Уоллиса, кластеры различаются высоко значимо по обоим показателям ТЕср вход и ТЕср выход. Даже кластеры К7, К8 имеют высоко значимые различия по ТЕср вход. Результаты кластеризации показывают низкий уровень эффективности предприятий по химическому производству. Например, кластеры с низкими показателями эффективности, которые отмечены оранжевым прямоугольником на рис. 2 составляют 35% от всех предприятий, а если добавить к ним предприятия К3 и К6, которые показывают низкий уровень эффективности по одному из показателей, то общее количество неэффективных предприятий составит почти 64%.

**Анализ MPI предприятий за период 2012–2016 гг.** Для оценки эффективности предприятий в динамике корректнее всего использовать характеристику MPI, которая оценивает регресс или прогресс эффективности в течение 2-летнего периода. Используя метод DEA с параметрами входа ОС и ОТ и параметром выхода В, были получены выборки MPI для каждой последовательной пары лет (MPI<sub>k</sub>, где k – номер второго года).

Чтобы оценить динамику MPI каждого предприятия, мы использовали линейный тренд  $\alpha \cdot t + b$  ( $\alpha > 0$  определяет прогресс, а  $\alpha < 0$  – регресс) и средним MPI<sub>ср</sub> на периодах 2012-2014гг. до санкций (MPI<sub>24</sub>,  $\alpha_{24}$ ) и 2014-2016гг. после санкций (MPI<sub>46</sub>,  $\alpha_{46}$ ), где в нашем случае  $MPI_{24} = (MPI_3 + MPI_4)/2$ ,  $MPI_{46} = (MPI_5 + MPI_6)/2$ ,  $\alpha_{24} = MPI_4 - MPI_3$ ,  $\alpha_{46} = MPI_6 - MPI_5$ .

Согласно критерию Краскела-Уоллиса предприятия FO статистически значимо превышают предприятия RO в докризисный период по показателю MPI<sub>24</sub> ( $0,005 < p \approx 0,026 < 0,05$ ), тогда как в посткризисный период различия между всеми ФС незначимые ( $p > 0,1$ ). Для показателей  $\alpha_{24}$ ,  $\alpha_{46}$  наблюдаются незначимые различия предприятий всех ФС. Согласно критерию Фридмана, показатель MPI незначимо изменяется для предприятий FO и JO и высоко значимо для предприятий RO ( $p < 0,0005$ ), а показатель  $\alpha$  изменяется незначимо для JO ( $p > 0,1$ ), статистически значимо для FO ( $0,005 < p \approx 0,02 < 0,05$ ) и высоко значимо для RO ( $p < 0,0005$ ) (рис. 3).

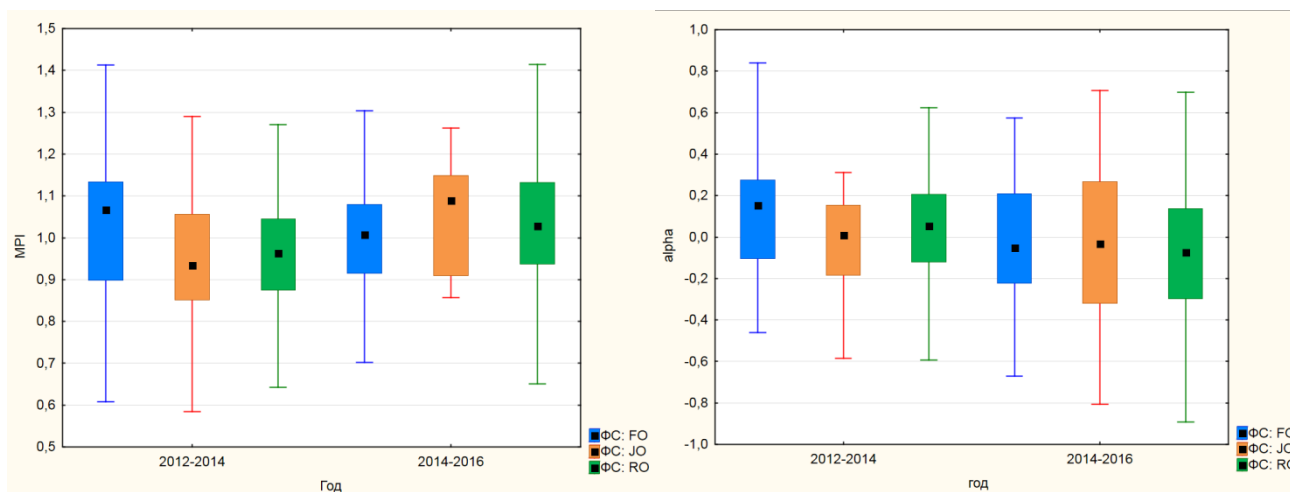


Рисунок 3. Непараметрические характеристики MPI и  $\alpha$

Подобная неоднородность выборок MPI, а также наличие принципиально важных граничных значений характеристик эффективности ( $MPI=1$  и  $\alpha=0$ ) являются поводом для применения кластерного анализа предприятий. Рассмотрим номинальную кластеризацию химических предприятий в разрезе ФС за период до (2012-2014гг.) и после (2014-2016гг) начала введения экономических санкций в координатах ( $MPI_{cr}$ ,  $\alpha$ ). Граничные значения ( $MPI_{cr} = 1$  и  $\alpha = 0$ ) делят все предприятия на 4 основные группы: ( $MPI_{cr}>1$  и  $\alpha > 0$ ) с приростом технической эффективности в текущем периоде и с тенденцией улучшения  $MPI_{cr}$  в будущем, ( $MPI_{cr}< 1$  и  $\alpha > 0$ ) со снижением технической эффективности в текущем периоде и с тенденцией улучшения  $MPI_{cr}$  в будущем, ( $MPI_{cr}> 1$  и  $\alpha < 0$ ) с приростом технической эффективности в текущем периоде, но с тенденцией ухудшения  $MPI_{cr}$  в будущем, ( $MPI_{cr}< 1$  и  $\alpha < 0$ ) со снижением технической эффективности в текущем периоде и с тенденцией ухудшения  $MPI_{cr}$  в будущем.

Численность предприятий в группе с указанием % от числа предприятий ФС в разрезе ФС и периодов приведены в табл. 2.

Таблица 2. Номинальная кластеризация предприятий в координатах ( $MPI_{cr}$ ,  $\alpha$ ).

Период		RO		FO		JO	
		$\alpha > 0$	$\alpha < 0$	$\alpha > 0$	$\alpha < 0$	$\alpha > 0$	$\alpha < 0$
2012-2014гг	$MPI > 1$	51	35	18	11	3	4
		23%	15%	38%	23%	14%	17%
	$MPI < 1$	85	55	13	6	9	7
		38%	24%	27%	12%	39%	30%
2014-2016гг	$MPI > 1$	54	78	7	18	6	9
		24%	35%	15%	38%	25%	39%
	$MPI < 1$	38	56	13	10	4	4
		16%	25%	26%	21%	18%	18%

Согласно табл. 1, в докризисный период 2012-2014гг. для RO и JO доминирующей по численности (в %) была 3-я группа ( $MPI_{cr}<1$  и  $\alpha>0$ ) со снижением технической эффективности в текущем периоде и с тенденцией улучшения  $MPI_{cr}$  в будущем, а для FO доминирующей является группа 1 ( $MPI_{cr}>1$  и  $\alpha > 0$ ) с приростом технической эффективности в текущем периоде и с тенденцией улучшения  $MPI_{cr}$  в будущем. В посткризисный период 2014-2016гг. у RO, JOи FОдоминировать стала 2-я ( $MPI_{cr}> 1$  и  $\alpha < 0$ ) с приростом технической эффективности в текущем периоде, но с тенденцией ухудшения  $MPI_{cr}$  в будущем.

Дополнительную информацию об особенностях динамики эффективности химических предприятий в разных ФС разных периодов дает номинальная кластеризация на уровне групповых средних (рис. 4).

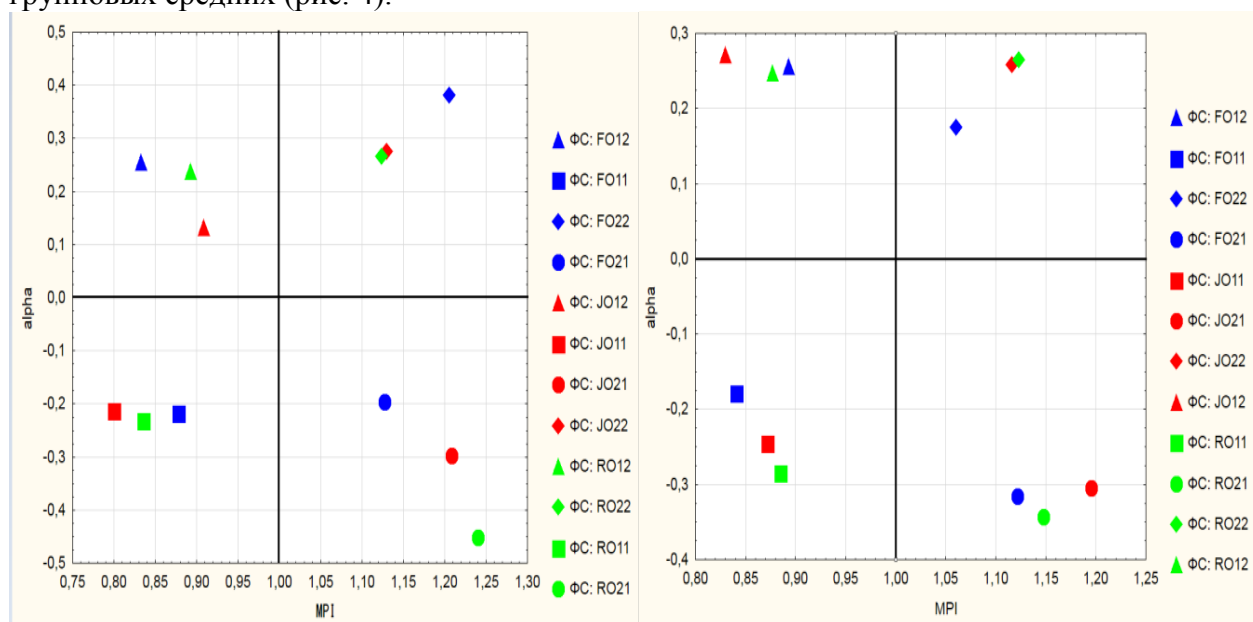


Рисунок 4. Диаграмма рассеяния кластеров по MPI и  $\alpha$ , 2012-2014 (слева), 2014-2016 (справа).

### Выводы.

Проведенное исследование позволило выявить следующие особенности эффективности деятельности химических предприятий в разрезе ФС:

1. Эффективность предприятий ФО по затратам значительно превышает РО только в 2014 и 2016 году. В тоже время по доходам, предприятия ФО значительно выше предприятий РО и JO за весь исследуемый период: различия между ФО и РО высоко значимые для всех лет, а для JO и FO статистически значимые только в 2014гг.
2. Построена высоко качественная 9-кластерная модель 297-ми химических предприятий по совокупности эффективностей затрат и доходов. Результаты кластеризации свидетельствуют о низкой эффективности химических предприятий: 35% предприятий имеют эффективности затрат и доходов меньше 30 %, а  $\approx 64\%$  - меньше 40 %.
3. С помощью индекса производительности Малмквиста MPI оценена долгосрочная динамика эффективности (до и после 2014г.) по линейному тренду ( $\alpha \cdot t + b$ ) MPI в разрезе ФС. Влияние санкций на динамику эффективности по ФС оценена как незначимая для FO и JO, и как высоко значимая для предприятий RO.
4. Проведенная номинальная кластеризация предприятий по химическому производству в разрезе ФС за период до и после 2014г. координатах (MPI<sub>ср</sub>,  $\alpha$ ) установила, что в докризисный период среди предприятий РО и JO доминирующей группой являлась группа предприятий со снижением технической эффективности в текущем периоде и с тенденцией улучшения MPI<sub>ср</sub> в будущем, а для FO доминирующей являлась группа с приростом технической эффективности в текущем периоде и с тенденцией улучшения MPI<sub>ср</sub> в будущем. В посткризисный период у всех ФС доминировать стала группа с приростом технической эффективности в текущем периоде, но с тенденцией ухудшения MPI<sub>ср</sub> в будущем.

Исследование выполнено при финансовой поддержке РФФИ в рамках научно-исследовательского проекта РФФИ «Динамическое моделирование развития российских,

иностранных и совместных промышленных предприятий в России в условиях экономических санкций», проект № 17-06-00584 А.

#### ЛИТЕРАТУРА

5. Информационный ресурс СПАРК [Сетевое издание]. URL: <http://www.spark-interfax.ru/>
6. Malmquist Productivity Index for Multi Time Periods /Jafari Y., et al// International Journal of Data Envelopment Analysis. – 2014. – V.2. – №.1. – P.315-322.
7. Строгонов М.С. Методика факторной оценки регионального инновационного потенциала с применением DEA-технологий//Вестник Забайкальского государственного университета. 2017. Т. 23. № 11. С. 101-108.
8. G. Tohidi, S. Razavyan A circular global profit Malmquist productivity index in data envelopment analysis // Applied Mathematical Modelling. – 2013. – №37. – P. 216–227.
9. Халафян А.А., Боровиков В.П., Калайдина Г.В. Теория вероятностей, математическая статистика и анализ данных: Основы теории и практика на компьютере. STATISTICA. EXCEL.– Москва URSS, 2016. – 317 с.
10. StatSoft, Inc. Electronic Statistics Textbook. – 2013. – StatSoft: Tulsa, OK. URL: <http://www.statsoft.com/textbook/> (дата обращения 12.09.2017)
11. Hill T., Lewicki P. STATISTICS: Methods and Applications.–StatSoft: Tulsa, OK., 2007.–719 с.

#### ПРИМЕНЕНИЕ ИНСТРУМЕНТА DATA MINING ДЛЯ ОЦЕНКИ КРЕДИТОСПОСОБНОСТИ ЗАЕМЩИКА

*А.С. Вершинин\*, Е.И. Губин*

*(г. Томск, Томский государственный университет систем управления и радиоэлектроники)  
e-mail: vershinintmsk@gmail.com*

#### APPLICATION OF THE DATA MINING TOOL FOR CREDIT SCORING

*A.S.Vershinin\*, E.I. Gubin*

*(Tomsk, Tomsk State University of Control Systems and Radioelectronics)  
e-mail: vershinintmsk@gmail.com*

**Abstract.** Due to the increasing need of automatization and improving the quality of the borrower's credit rating and its future behavior, modern scoring models are being improved using data mining and machine learning methods. In this paper, to build a scoring model, we consider such statistical models as logistic regression and decision trees. For these purposes, SAS Enterprise Miner, which contains many methods and tools for finding patterns and predictive modeling, was used.

**Keywords:** credit scoring, data mining, machine learning, predictive modeling, decision trees, regression.

В связи с возрастающей необходимостью в автоматизации и повышении качества оценки кредитоспособности заемщика и его дальнейшего поведения, современные скоринговые модели усовершенствуются благодаря использованию методов интеллектуального анализа данных и машинного обучения. Сегодня эти методы уже не являются чем-то новым, это обязательный пункт банковских бизнес-процессов.

Скоринговая модель – это статистическая модель для прогноза вероятности попадания клиента в категорию «хороший» / «плохой» в течение периода времени после выдачи кредита, в течение которого определяется поведение заемщика по данному кредиту.

К интеллектуальному анализу данных принадлежит большое количество аналитических методов, которые обычно делятся на две большие категории: поиск закономерностей и