

2. *Ляльков И. М.* Анализ стратегических финансовых рисков компаний топливно-энергетического комплекса // Экономика и предпринимательство. 2017. № 7 (84).
3. *Ляльков И. М.* Международный рынок нефти: риски и вызовы для российских компаний // Новая российская экономика: движущие силы и факторы. Международная научно-практическая конференция молодых ученых, аспирантов и магистрантов (3 декабря 2016 г., Ярославль).
4. *Ляльков И. М.* Методология Value-at-Risk в моделировании ценовых рисков нефти // Современная наука: актуальные проблемы теории и практики. Серия «Экономика и право». 2017. № 11.
5. *Ляльков И. М.* Oil Market: A new paradigm // XXX Международные Плехановские чтения, Москва 2017.
6. *Халл Дж. К.* Опционы, фьючерсы и другие производные инструменты. М. : ИД «Вильямс», 2008.
7. *Хоминич И. П., Ляльков И. М.* Управление финансовыми рисками в компаниях топливно-энергетического комплекса // Финансовые стратегии и модели экономического роста России: проблемы и решения : сб. науч. статей коллектива Финансового факультета науч.-педагог. работников и молодых ученых выпускающих кафедр Образовательно-научного центра «Экономика и финансы».
8. Энциклопедия финансового риск-менеджмента / под ред. А. А. Лобанова, А. В. Чугунова. М. : Альпина Паблишерз, 2009.
9. *Jorion P.* Value at Risk: The New Benchmark for Managing Financial Risk. The McGraw-Hill Companies, Inc. 2007.
10. *Stulz R.* Risk Management and Derivatives. South-Western Mason. 2003.
11. *Tagliafichi R. A.* The estimation of Market VaR using Garch model and a heavy tails distribution. [Electronic resource]. URL: [www.actuaries.org/AFIR/](http://www.actuaries.org/AFIR/) (date of access: 18.07.2018).
12. *Tsay R. S.* Analysis of Financial Time Series. John Wiley and Sons, 2010.

## **КРЕДИТНЫЙ СКОРИНГ. МЕТОДЫ ПОСТРОЕНИЯ СКОРИНГОВЫХ МОДЕЛЕЙ**

**Л. А. Маслова**

*Московский государственный институт международных отношений, Россия*  
E-mail: [l.maslova@inno.mgimo.ru](mailto:l.maslova@inno.mgimo.ru)

В работе проводится сравнение статистической модели кредитного скоринга на основе логистической регрессии с моделью кредитного скоринга, построенной по модели неройной сети - самоорганизующихся карт Кохенена (SOC). Обе модели обладают хорошей классификационной способностью распознавания «плохих» и «хороших» заемщиков на уровне от 75-80%.

## **CREDIT SCORING. METHODS OF CONSTRUCTION OF SCORING MODELS**

**L. A. Maslova**

The paper compares the statistical model of credit scoring on the basis of logistic regression with the model of credit scoring, built according to the model of a non-linear network - self-organizing Kohonen maps. Both models have a good classification ability of recognizing "bad" and

"good" borrowers at a level of 75-80%.

Сегодня финансовыми институтами и банками для предотвращения кризисного состояния признается необходимость ежедневной количественной оценки возможных потерь по различным операциям. Эта оценка необходима также, чтобы выполнить требования со стороны надзорных органов по объему собственного капитала и ликвидности.

В основе процесса управления кредитными рисками клиента в большинстве банков лежит *классификация* потенциальных клиентов и контрагентов по уровню риска на основе системы кредитных рейтингов и оценки вероятности дефолта. Построение профиля риска клиента, позволяет оценить вероятность того, что клиент выплатит/не выплатит долг в краткосрочном периоде 1 год. Оценка осуществляется на основе характеристик клиента, статистики дефолтов, рыночной ситуации. Критерии разрабатывается самим банком.

Согласно требованиям Базель II и III оценка и управление кредитными рисками должна осуществляться на основе внутренних рейтингов. Для этого банки создают свои модели оценки рейтингов – скоринг модели. На основе различных характеристик клиента согласно модели заявителю присваивается определенный рейтинг (scoring), который оценивает степень вероятности его дефолта.

Основные общепринятые модели скоринга делятся на следующие типы: анкетный (заявочный) скоринг; поведенческий скоринг; голлиторский скоринг; антимошеннический скоринг; модели PD, LGD, CCF/EAD в соответствии с Advanced-IRB подходом Базельского комитета.

При решении о выдаче кредита используется анкетный скоринг. Анкетный скоринг подразделяет клиентов на «плохих» и «хороших», присваивает определенные веса с учетом риска (вероятности дефолта). Для разделения клиентов на «плохих» и «хороших» применяются различные математические методы классификации и кластеризации: статистические методы (логистическая регрессия, линейная регрессия, дискриминантный анализ) деревья классификации; нейронные сети; генетический алгоритм; метод ближайших соседей; линейное программирование.

Создание качественных скоринговых моделей является достаточно сложной и недешевой задачей. В работе [1] приводится табл. 1 сравнения точности классификации различными методами оценки кредитных рисков.

Таблица 1

**Сравнение точности классификации различными методами (Thomas, 2000)**

Авторы	Лин. рег	Лог. рег	RPA	LP (лин. прог)	Нейронные сети	Генет. алгоритм
Henley (1995)	43,4	43,3	43,8	-	-	-
Boyle et al. (1992)	77,5	-	75	74,7	-	-
Srinivasa and Kim (1987)	87,5	89,3	93,2	86,1	-	-
Yodas et al. (1997)	68,4	-	62,3	-	62,0	64,5
Desai et.al (1997)	66,5	67,3	67,3	-	6,4	

### Статистическая модель анкетного скоринга.

Рассмотрим применение статистических методов (логистической регрессии [3]) для анкетного кредитного скоринга. Объект исследования – база данных о 3250 клиентах (20 предикторов), предоставленные отделом кредитования российского банка N.

Прежде, чем приступить к построению модели была проведена оценка качества данных на удовлетворение условиям построения модели, как для базы данных, так и для случайной «обучающей» выборки. В частности, соотношение «плохих» и «хороших» заемщиков в базе и выборке должно быть не менее 30%. Переменная «просрочка90» принимает значение 0, если кредит был погашен до 90 дней - заемщик «хороший». Если кредит не был погашен через 90 дней, то такой заемщик считается «плохим». Отношение «плохих» (нет) к «хорошим» (да) кредитам в случайной выборке равно 66,98%.

Для оценки качества базы были проведены также следующие исследования: категоризация интервальных переменных; оценка предикторов по соотношению «плохих» и «хороших» заемщиков; оценка способности предикторов классифицировать заемщиков; кластерный анализ.

*Оценка предикторов по соотношению «плохих» и «хороших» заемщиков.*

Анализ описательных статистик и предсказательной способности предикторов показал, что в исследуемой базе данных практически все предикторы имеют удовлетворительное соотношение «плохих» и «хороших» заемщиков. Поскольку в первоначальной базе категории *возраст* и *стаж работы* была количественной переменной, то была проведена категоризация этих переменных.

### Бинарная логистическая регрессия.

Бинарная логистическая регрессия была построена в пакете SPSS [2] для двух случайных выборок. Результаты логистической регрессии приведены в табл. 2.

Таблица 2

Логистическая регрессия

Предикторы		Стд. ошибка	Вальд	Ст. св.	Знч.	Exp(B)
Нал. свед.в БКИ	1,172	0,139	71,191	1	0	3,23
Пол	0,506	0,113	19,927	1	0	1,658
Статусм.жит	0,374	0,024	242,728	1	0	1,454
Типдоп.уд.личн	-0,368	0,091	16,509	1	0	0,692
Должн.об	0,087	0,041	4,413	1	0,036	1,091
Катвозраст	-0,19	0,081	5,462	1	0,019	0,827
Колвоист.дох	-0,632	0,199	10,123	1	0,001	0,532
Константа	-2,802	0,829	11,42	1	0,001	0,061

Уравнение лог-регрессии имеет вид

$$Z = -2,802 + 1,172 * БКИ + 0,506 * Пол + 0,374 * Статусм - 0,368 * Типудлич + 0,097 * Должноб - 0,19 * катвозраст - 0,632 * колвоистдох$$

Вероятность дефолта рассчитывается по формуле

$$P = \frac{1}{1 + \exp(-z)}$$

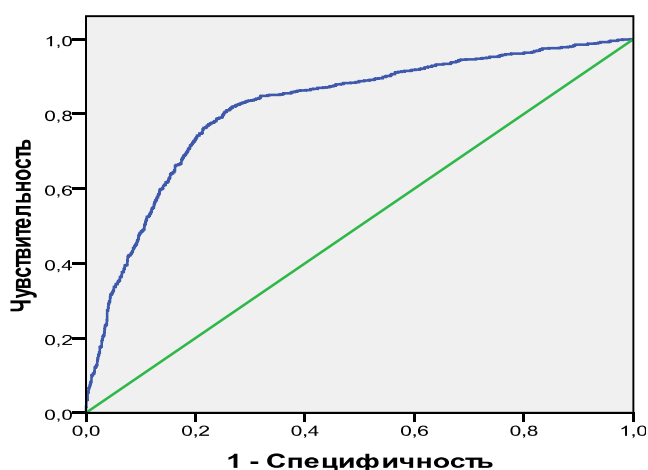
Модель логистической регрессии, должна одинаково хорошо различать как «хороших», так и «плохих» заемщиков. Дискриминирующую способность модели можно оценить по таблице классификации. В нашей модели, правильно распознано 77, 2% из общего числа просроченных кредитов в тестовой выборке.

В модели логистической регрессии коэффициент детерминации не является основной характеристикой точности модели в отличие от модели линейной регрессии. В приведенной выше модели  $R^2=0,389$ . Характеристикой важности предикторов является статистика Вальда. В нашем случае, самым важным предиктором является *Статус места жительства*. Он принимает значения от 1 до 7. Например, 2 – проживает у родственников, 3- другое, 5 - проживает постоянно в непериватизированном жилье, 6 – снимает. Чем больше значение этого предиктора, тем больше вероятность дефолта. Вторым по степени важности является *Наличие сведений в БКИ*.

### **ROC-кривая.**

Другой характеристикой способности классификации модели является ROC-кривая. ROC–кривая показывает зависимость количества верно классифицированных положительных исходов от количества неверно классифицированных отрицательных исходов. Для сравнения двух и более моделей между собой сравниваются площади под ROC-кривыми – этот показатель называется AUC и измеряется от 0,5 до 1. Наша модель имеет показатель AUS равный 0,815, что говорит о высоком качестве модели.

Чем больше значение площади, тем лучше модель. Обычно считают, что значение площади от 0,9 до 1 соответствует отличному качеству модели, от 0.8–0.9 – очень хорошему, 0.7–0.8 – хорошему. По значению площади под ROC–кривой можно вычислить показатель индекс Джинни. Чем круче кривая, тем лучше предсказательные свойства модели.



Диагональные сегменты формируются совпадениями.

ROC-кривая

### **Проверка модели на тестовой выборке.**

Проверка модели была произведена на другой части выборки, содержа-

щей 30% наблюдений. Результаты приведены ниже в табл. 3.

Таблица 3

			Модель	Модель
	Тестовая	Вероятн. деф.	деф>0,75	деф>0,7
дефолтов	397	дефолтов	329	396
процент	0,174352	процент	0,14449	0,173913043
		ошибка распознавания	0,17128	0,002518892

В тестовой выборке было 397 случаев просрочки по оплате кредита, что составило 17,4% от объема выборки равной 973. Если случаем дефолта заемщика принять вероятность большую 0,75, то модель распознает 329 случаев, что дает ошибку 17,4%. Для вероятности 0,7 распознает 396 случаев и ошибка 0,2%.

Таким образом, построенная статистическая модель обладает хорошими прогностическими свойствами. Кроме того позволяет оценивать степень важности предикторов для анкетного скоринга, что позволяет уменьшить число предикторов в анкете.

#### **Модель кредитного скоринга на самоорганизующихся картах Кохонена.**

Сеть Кохонена или самоорганизующиеся карты Кохонена (SOM) являются одним из видов нейросетевых алгоритмов. Карты Кохонена позволяют классифицировать исходные данные с достаточно высоким процентом верных результатов. Кроме того, наглядная визуализация многомерных входных данных дает хорошую информацию о структуре самих данных.

Сеть учится сама понимать структуру данных. Эта нейронная сеть обучается без учителя. В результате обучения входные данные разделяются на кластеры. Многомерные входные данные (нейроны) отображаются на двумерной плоскости выходных нейронов. Графическое представление карт Кохонена дает информацию, по каким параметрам наблюдается сходство объектов. Какое количество наблюдений попадает в данный кластер (мощность кластера). Для этого карты раскрашиваются. Более темные цвета соответствуют большим весам. Близкие веса имеют одинаковый цвет. Карты Кохонена достаточно часто используются для задач классификации, к которым относится анкетный кредитный скоринг.

#### **Результаты обучения.**

Анализ проводился в статистическом пакете Detuctor. Было построено 8 карт с различными входными параметрами: доли обучающего и тестового множества, размер сетки, размер ошибки сети, уровень значимости, радиус обучения.

Как показало исследование при уменьшении размера сетки (в данном случае с 16×12 до 10×6) значительно снижается детализация карты Кохонена, что приводит к отображению всего исходного множества на один кластер. В то же время, изменение остальных параметров не имеет значительной роли в отношении количества получаемых кластеров. Наилучшие результаты на данной

базе данных дают карты с параметрами: карты: размер обучающей выборки 95%, тестовой 5%; размер сетки 16×12 нейронов; уровень ошибки сети – 1; Уровень значимости – 0,1%; Радиус обучения в начале процесса обучения – 4 показал, что

При снижении уровня допустимой ошибки сети снижается количество распознаваемых примеров как на обучающем, так и на тестовом множествах. Уменьшение размеров сетки также приводит к снижению доли распознаваемых примеров. Остальные же рассмотренные параметры почти не оказывают влияния на эту характеристику.

Анализ атрибутов (входных данных) карт показал, что наибольшую итоговую значимость (100%) имеют параметры: «Категория стажа», «Категория возраста», «Количество источников дохода», «Общий трудовой стаж», «Количество иждивенцев», «Тип недвижимости в собственности», «Тип дополнительного удостоверения личности», «Тип транспортного средства», «Пол» и «Статус места жительства». Пять из этих атрибутов (выделено курсивом) совпадают с предикторам лог-регрессии. Атрибуты «Наличие закрытых кредитов в других банках» и «Наличие сведений в БКИ» имеют 99,9% значимость, при этом они попадают в один кластер.

Таблица 4

Таблица результатов построения карт Кохонена

№ Карты Кохонена	Размер карты	Объем обучающей выборки	Число кластеров	Качество классификации					
				Исходное множество все вместе		Обучающее		Тестовое	
				Верно	Неверно	Верно	Неверно	Верно	Неверно
1	16*12	95%	2	77,38%	22,62%	77,2%	22,8%	80,86%	19,14%
2	30*20	95%	16	79,08%	20,92%	79,47%	20,53%	71,6%	28,4%
4	16*12	70%	2	77,51%	22,49%	78,73%	21,27%	74,67%	25,33%
5*	16*12	95%	7	78,4%	21,6%	78,56%	21,44%	75,31%	24,69%

\*Карта строится только по параметрам регрессии

Из табл. 4 видно, что наилучший результат по способности классифицировать заемщиков дает карта под номером 1. Способность верно распознать «плохих» заемщиков на тестовой выборке равна 80,86%.

Другие карты обладают тоже хорошей способностью классифицировать заемщиков. Карта 5 построена по параметрам логистической регрессии. Уменьшение количества параметров не сильно повлияло на результат. Способность верно распознать «плохих» заемщиков равна 75,31%. Получился практически тот же результат, что и в случае лог-регрессии.

### Выводы.

На одной и той же базе данных построены две модели анкетного кредитного скоринга. Обе модели обладают хорошей классификационной способностью распознавания «плохих» и «хороших» заемщиков на уровне от 75-80%. Преимуществом статистической модели является ясная интерпретация атрибу-

тов (предикторов). Для построения обеих моделей необходимо проводить предварительную обработку данных и анализ пригодности данных для анализа. Преимуществом нейросетевой модели является графическая визуализация, которая помогает более глубоко оценить взаимосвязи атрибутов (предикторов), что позволяет выбрать и построить наилучшую модель. Нейросетевая модель требует большего компьютерного времени для построения модели, с требуемыми характеристиками.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Thomas L. C.* A Survey of Credit and Behavioural Scoring // University of Edinburgh. 1999.
2. *Бююль А., Цёфель П.*, SPSS: искусство обработки информации. СПб.ООО «Диал-СофтЮП», 2005.
3. *Сорокин А. С.* Построение скоринговых карт с использованием модели логистической регрессии // Интернет- журнал «Науковедение». 2014. Вып 2. [Электронный ресурс]. URL: <http://publ.naukovedenie.ru/> (дата обращения: 25.08.2018).
4. *Горбаченко В. И.* Сети и карты Кохонена. [Электронный ресурс]. URL: [http://gorbachenko.self-organization.ru/articles/Self-organizing\\_map.p](http://gorbachenko.self-organization.ru/articles/Self-organizing_map.p) (дата обращения: 20.08.2018).

### НЕПАРАМЕТРИЧЕСКАЯ РЕГРЕССИЯ ИНТЕРПОЛЯЦИОННОГО ТИПА

**В. В. Новиков, С. В. Лю**

*Саратовский государственный университет, Россия*  
E-mail: [vvnovikov@yandex.ru](mailto:vvnovikov@yandex.ru), [lyusv@info.sgu.ru](mailto:lyusv@info.sgu.ru)

Рассмотрен класс непараметрических оценок функции регрессии, использующих интерполяционные коэффициенты Фурье-Лагранжа.

### ON AN INTERPOLATION TYPE NONPARAMETRIC REGRESSION

**V. V. Novikov, S. V. Lyu**

We consider a method for constructing nonparametric estimator of a regression function based on the Fourier-Lagrange coefficients.

Рассмотрим, регрессионную модель

$$Y_i = m(X_i) + \varepsilon_i, i = 1, \dots, n,$$

где  $m(x) = E(Y | X = x)$  – неизвестная функция регрессии, подлежащая оцениванию на основе эмпирических данных  $\{(X_i, Y_i)\}_{i=1}^n$ , а  $\{\varepsilon_i\}_{i=1}^n$  – случайные ошибки. Один из непараметрических методов построения оценки  $\hat{m}(x)$  основан на разложении функции  $m(x)$  в ряд Фурье