УДК 336.77

С. Математические и количественные методы
C. Mathematical and Quantitative Methods

ПОСТРОЕНИЕ ГИБРИДНОЙ ЛОГИСТИЧЕСКОЙ МОДЕЛИ ДЛЯ ВЫЯВЛЕНИЯ СКРЫТЫХ ДЕФОЛТОВ ПО ФИНАНСОВОЙ ОТЧЕТНОСТИ КОМПАНИЙ

А. И. ТКАЧЁВ[1)]

[1)]*Национальный банк Республики Беларусь, пр. Независимости, 20, 220008, г. Минск, Беларусь*

Рассмотрены регрессионные подходы к построению моделей оценки кредитного риска. Разработан и описан процесс создания гибридной логистической модели множественного упорядоченного выбора, которая представляет собой систему, состоящую из двух эконометрических моделей (линейной вероятностной модели и логит-модели). Полученные результаты, как инструмент макропруденциального контроля, имеют практическую значимость при проведении анализа реального сектора на микроданных.

Ключевые слова: балансовые показатели; оценка риска; скоринг-модели; микроданные; дефолт.

BUILDING A HYBRID LOGISTICS MODEL TO IDENTIFY HIDDEN DEFAULTS IN THE FINANCIAL STATEMENTS OF COMPANIES

A. I. TKATCHEV[a]

[a]*National Bank of the Republic of Belarus, 20 Niezaliežnasci Avenue, Minsk 220008, Belarus*

This article explores approaches to the construction of credit risk assessment models. The aim of this work is to study quantitative methods for assessing credit risks. In the course of the study, the financial data of companies were processed and systematised, the analysis and synthesis of data was carried out, economic, mathematical and statistical approaches were applied. The process of creating a hybrid logistic model of multiple ordered choice, which is a system of two econometric models (linear probabilistic model and logit model), is described. The results obtained as a tool for macroprudential control are of practical importance when analysing the real sector using microdata.

Keywords: balance sheet indicators; risk assessment; model scoring; microdata; default.

Введение

Существенные изменения в глобальной экономике, которые произошли за последнее десятилетие, привлекли внимание многих ученых, законодателей и практиков, вовлеченных в финансовую сферу. Финансовые рынки характеризуются повышенной волатильностью. В результате отсутствия оценок либо их недостаточности для составления планов по противодействию неявным рискам периодически

Образец цитирования:
Ткачёв АИ. Построение гибридной логистической модели для выявления скрытых дефолтов по финансовой отчетности компаний. *Журнал Белорусского государственного университета. Экономика*. 2021;2:26–38.

For citation:
Tkatchev AI. Building a hybrid logistics model to identify hidden defaults in the financial statements of companies. *Journal of the Belarusian State University. Economics.* 2021;2:26–38. Russian.

Автор:
Артём Ильич Ткачёв – главный специалист управления финансовой стабильности.

Author:
Artem I. Tkatchev, chief specialist, Financial Stability Department.
a.tkachev1992@mail.ru

возникают финансовые кризисы, которые с учетом усложнения экономических процессов и роста взаимозависимостей практически всегда приводят к спаду активности в реальном секторе.

Например, «в России многие неэффективные компании не уходят с рынка, а превращаются в зомби-компании без перспектив роста и расплаты по долговым обязательствам»[1], что было отмечено на организованной Банком России конференции «(Пост)коронавирусная экономика и вызовы для политики центрального банка». «Расшифровывая термин “зомби”, эксперты называют два основных компонента: мертвеца и магию, которая превращает его в некое полуживое существо, – поясняет Кирилл Тремасов, директор Департамента денежно-кредитной политики Банка России. – В роли этого волшебника выступают банки, а главным элементом “волшебства” является реструктуризация кредита, который и поддерживает мертвеца в полуживом состоянии»[2]. В текущих реалиях актуальной задачей для национальных (центральных) и коммерческих банков представляется выявление подобных компаний, поскольку они искажают конкурентную среду, препятствуют эффективному перераспределению труда и капитала между компаниями в целом. Для определения финансовой устойчивости любой компании, помимо моделей оценки вероятности дефолта, основанных на дискриминантном анализе, используются также регрессионные модели.

Отправной точкой в расчете вероятностей дефолта выступает созданная на базе макроэкономического подхода и уже ставшая классической модель Уилсона. Она легла в основу программного продукта *Credit Portfolio View*, разработанного консалтинговой группой *McKinsey & Company* [1]. Названная модель представляет собой инструмент для оценки рисков в отраслях, которые являются наиболее чувствительными к экономическим циклам и поэтому первые реагируют на изменения, происходящие в экономике.

Логистическая регрессия, или логит-регрессия (англ. *logit model*), – это статистическая модель, используемая для предсказания вероятности возникновения некоторого события путем подгонки данных к логистической кривой. Основное преимущество применения такой модели заключается в том, что при интерпретации результатов (они могут принимать значения только в интервале от 0 до 1, также они определяют номинальное значение вероятности наступления несостоятельности предприятия) не возникает проблем. В дискриминантных моделях вероятность банкротства не обозначается номинальным значением. Таким моделям присуще наличие так называемых зон неопределенности, при попадании в которые по значению рассчитанного рейтингового показателя нельзя сделать однозначный вывод о вероятности дефолта. В логит-моделях такие зоны отсутствуют, поскольку если оцененная вероятность больше 0,5, то событие произойдет, а если такая вероятность меньше 0,5 или равна этому значению, то событие не случится.

Одной из логистических регрессионных моделей является модель Олсона [2]. Альтернативным скоринговым подходом выступает модель банкротства Таффлера [3]. Данную линейную регрессионную модель с четырьмя финансовыми коэффициентами ученый разработал для оценки финансового состояния компаний Великобритании. Для этой цели он в период с 1969 по 1975 г. исследовал 46 организаций, постигших дефолт, и столько же финансово устойчивых предприятий.

Среди российских ученых, занимающихся данной проблематикой, выделяется А. М. Карминский [4], среди белорусских – В. И. Малюгин [5], Г. В. Савицкая [6], Н. В. Гринь, А. И. Зубович, П. С. Милевский [7], Е. В. Пытляк [8] и др.

Постановка задачи и описание существующих подходов

В данном исследовании применялись две выборки, содержащие разный набор информации об одном классе объектов (предприятия, люди, дома, деревья и т. д.). В связи с этим регрессионный анализ целесообразно проводить по данным одной из выборок, т. е. использовать ее в качестве обучающей, а полученные данные в ходе этого анализа применять на другой выборке. Как правило, для проведения регрессионного анализа ученые использовали данные, общие для всех выборок. Выборка 1 включает информацию, состоящую из трех форм отчетности, выборка 2 – из двух форм отчетности. Общим источником информации для обеих выборок является одна форма отчетности, которая будет считаться пересечением этих выборок (рис. 1).

Цель настоящего исследования – проанализировать две выборки предприятий и показать, что дополнительная предварительная классификация данных, не являющихся общими для обучающей и тестовой выборок, дает улучшение результатов при проверке на тестовой выборке. Такая ситуация может возникнуть, если в качестве одной выборки будут использоваться данные, подготовленные и предоставленные

[1] *Ведерина Е.* Господдержка спровоцировала рост числа зомби-компаний // Ведомости [Электронный ресурс]. URL: https://www.vedomosti.ru/economics/articles/2020/12/14/850946-gospodderzhka-sprovotsirovala (дата обращения: 28.12.2020).

[2] Там же.

предприятиями в Национальный статистический комитет Республики Беларусь, а в качестве другой выборки – данные отчетности, подготовленной для Министерства по налогам и сборам Республики Беларусь. Поскольку в этих данных имеются как сходства, так и отличия, то использовать сразу всю информацию невозможно. Тогда следует обратиться к более узкому информационному множеству, которое является общим для двух выборок. В итоге может возникнуть ситуация, где обучающая и тестовая выборки отличаются по определенным параметрам. Так, в обучающей выборке можно задействовать параметры, недоступные для тестовой выборки. В данном случае важно раскрыть суть методики и доказать, что если через создание двухшаговой модели привлечь к использованию недостающие при проверке данные, применяемые для предварительной классификации, то результаты, полученные при построении двухшаговой модели, будут лучше, чем при построении одношаговой модели множественного упорядоченного выбора. Данную гипотезу можно проверить на примере предварительной классификации с помощью созданной линейной регрессионной модели.

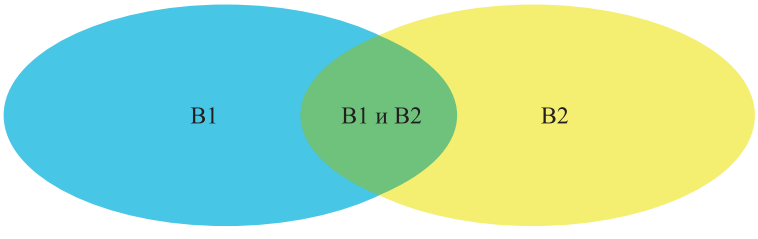

Рис. 1. Схема пересечения выборок:
B1 – выборка 1; B2 – выборка 2

Fig. 1. Scheme of intersection of samples:
B1 – sample 1; B2 – sample 2

В целях проверки этой гипотезы взяты данные реально существующих предприятий. На основе этих данных и построено множество коэффициентов.

Для оценки кредитного риска применяются модели регрессионного анализа. Зависимая переменная y принимает фиксированные значения из некоторого заранее предопределенного набора. В частности, модель с зависимой бинарной переменной имеет два значения (обычно 0 и 1), а также регрессоры x, которые определяют значения зависимой переменной [9].

К моделям регрессионного анализа относятся линейная вероятностная модель, логит- и пробит-модели. Их сравнение представлено на рис. 2.

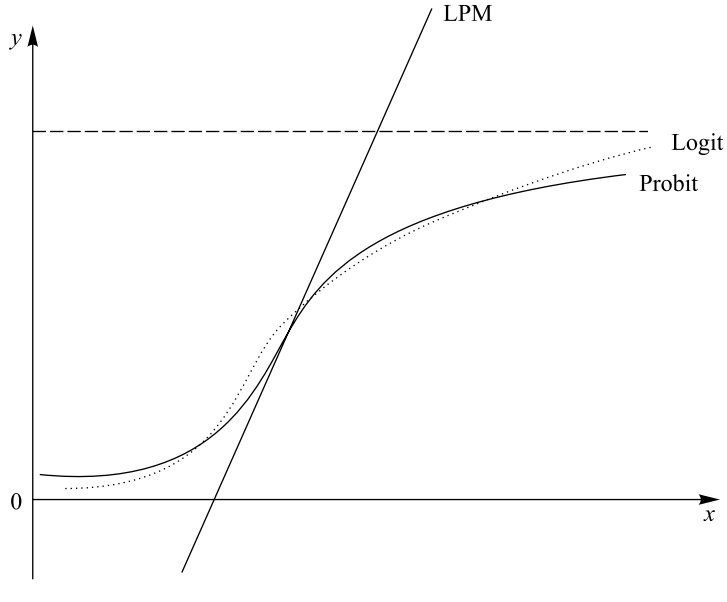

Рис. 2. Диаграмма отказа в выдаче ипотечного кредита
и соотношения платежа и дохода:
LPM – линейная вероятностная модель; logit – логит-модель;
probit – пробит-модель

Fig. 2. Diagram of refusal to issue a mortgage loan
and the ratio of payment to income:
LPM – linear probabilistic model; logit – logit model;
probit – probit model

Необходимость в бинарном моделировании возникает, как правило, если для определения показателя используется порядковая шкала, которая принципиально не может быть преобразована в непрерывную числовую последовательность. Пусть, например, дается оценка полу заявителя, где цифра 0 означает «мужчины», цифра 1 – «женщины». В таком случае построенная обычная линейная регрессия будет предсказывать абсурдные значения y (дробные, отрицательные или больше единицы).

Линейная вероятностная модель – это регрессионная модель, где исходная переменная является бинарной. Одна или несколько объясняющих переменных (от их значения зависит значение исходной переменной) используются для прогнозирования результата, объясняющие переменные могут быть двоичными или непрерывными. Линейная вероятностная модель – модель линейной множественной регрессии, которая применяется к бинарной зависимой переменной. Линейная вероятностная модель находит наилучшую линейную зависимость путем минимизации суммы стандартных отклонений и имеет формулу

$$y_i = \beta_0 + \beta_1 x_{1,\,t} + \beta_2 x_{2,\,t} + \ldots + \beta_n x_{n,\,t} + \varepsilon_{i,\,t},$$

где $y_i,\ i \in \{0,1\}$, – зависимая переменная (платежеспособность клиента); $x_{i,\,t}$, $i = 1, \ldots, n$, – объясняющие переменные (скоринговые характеристики); β_i – вектор неизвестных коэффициентов (параметры модели, скоринговые веса).

В результате построения модели получены значения β – это параметры, которые определяют характер связи между наблюдаемым значением переменной «платежеспособность клиента» и соответствующими скоринговыми характеристиками.

Логит- и пробит-модели – это статистические методы, специально созданные для обработки такого момента, когда объясняемая переменная может принимать только два значения. В нашем случае переменная y_i, которая должна быть объяснена линейной вероятностной моделью, способна принимать значения 0 (отсутствие значения по умолчанию) и 1 (наличие значения по умолчанию). Это привело к некоторым трудностям в формировании модели. В частности, возникающие стандартные ошибки, обычно генерируемые компьютерными программами, скорее всего, будут неправильными. Существуют статистические подходы, специально разработанные для правильной обработки логит- и пробит-моделей [8].

Значения зависимой переменной y_i имеют следующую интерпретацию: если в исследуемом периоде $y_i = 1$, то предприятие признается дефолтным, а если $y_i = 0$, то оно считается нормально функционирующим. Таким образом, вероятность дефолта i-го случая p_i равна вероятности того, что $y_i = 1$. В число компонент вектора факторов $x_i = \left(x_i 0,\ x_i 1, \ldots, x_i k\right)^{\mathrm{T}}$ могут включаться как количественные, так и качественные переменные (финансовые показатели состояния предприятий).

Модель бинарного выбора описывает зависимость вероятности дефолта предприятия p_i от включенных в модель факторов, задаваемых вектором $x_i = \left(x_i 0,\ x_i 1, \ldots, x_i k\right)^{\mathrm{T}}$, и определяется соотношением

$$p_i = P\left(y_i = 1\right) = F\left(x_i^{\mathrm{T}}\beta\right). \tag{1}$$

При этом вероятность того, что предприятие не является проблемным, равна

$$p_i = P\left(y_i = 0\right) = 1 - p_i = 1 - F\left(x_i^{\mathrm{T}}\beta\right).$$

Различают два основных типа модели бинарного выбора [10]:

- пробит-модель (если $F(\cdot)$ – функция стандартного нормального распределения);
- логит-модель (если $F(\cdot)$ – функция логистического распределения вероятностей).

Интерпретация моделей бинарного и множественного выбора основана на использовании так называемой латентной (скрытой, ненаблюдаемой) переменной y_i, которая связана с вектором факторов x_i моделью множественной линейной регрессии:

$$y_i^* = x_i^{\mathrm{T}}\beta + \xi_{i,\,1} = 1,\ 2, \ldots, n, \tag{2}$$

где $\beta = \left(\beta_0,\ \beta_1, \ldots, \beta_k\right)^{\mathrm{T}} - \left(k+1\right)$ – мерный вектор неизвестных параметров; ξ_i – случайная ошибка наблюдения в i-м эксперименте.

Предполагается, что ошибки $\{\varepsilon_i\}$, $i = 1, 2, \ldots, n$, являются независимыми в совокупности и одинаково распределенными случайными величинами с нулевым средним значением и постоянной дисперсией. В логит- и пробит-моделях бинарная переменная y_i связана с ненаблюдаемой переменной y_i^* следующими соотношениями:

$$y_i = 1, \text{ если } y_i^* > c, \tag{3}$$

$$y_i = 0, \text{ если } y_i^* \le c, \tag{4}$$

где c – некоторое пороговое значение. Обычно рассматриваются модели со свободным членом, т. е. полагается, что $x_{i0} \equiv 1$, $i = 1, \ldots, n$. В этом случае β_0 – свободный член, $\beta_1, \ldots, \beta_k$ – коэффициенты регрессии.

При этом модель (2) включает k факторов, а пороговое значение в соотношениях (3) и (4) равно нулю. Если модель (2) включает свободный член, то с учетом симметричности функции распределения $F(\cdot)$ на основании модели (2), соотношений (3) и (4) получаем

$$p_i = P\left(y_i^* > 0\right) = P\left(x_i^{\mathrm{T}}\beta + \xi_i > 0\right) = 1 - P\left(\xi_i \le -x_i^{\mathrm{T}}\beta\right) = F\left(x_i^{\mathrm{T}}\beta\right).$$

Модель (1) является нелинейной по параметрам $\beta = \left(\beta_0, \beta_1, \ldots, \beta_k\right)^{\mathrm{T}}$, и поэтому компоненты вектора β имеют более сложную интерпретацию, чем коэффициенты регрессии в модели типа множественной линейной регрессии [8].

Алгоритм построения гибридной логистической модели

В данном исследовании для решения поставленной задачи – выявления дефолтных компаний – была выдвинута следующая гипотеза: если предварительно проранжировать случаи из выборки 1 с помощью линейной вероятностной модели с применением недостающих в выборке 2 показателей (обучающая выборка), а затем строить по ней на базе общих для выборок 1 и 2 показателей логистическую модель множественного упорядоченного выбора, то получится модель, дающая лучший результат по сравнению с моделью множественного упорядоченного выбора, которая сразу строится на базе общих для обеих выборок показателей (по той же обучающей выборке). Схематичное описание дано на рис. 3.

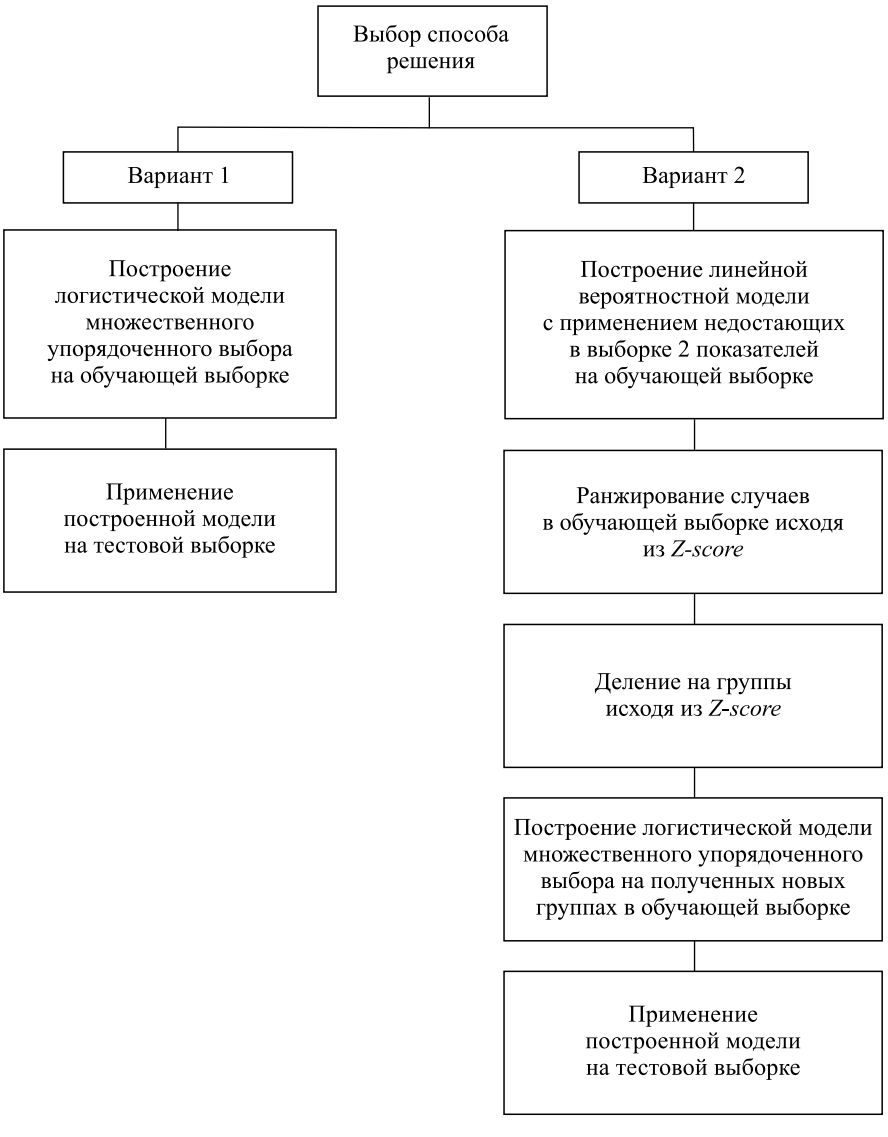

Рис. 3. Алгоритм построения модели
Fig. 3. Algorithm for constructing a model

Обзор результатов

В связи с тем что одно и то же предприятие в разное время имеет различные коэффициенты и может обладать неодинаковым бинарным значением (дефолт или недефолт), было принято решение строить модель на панельной выборке. Дефолтным предприятием в модели принято считать ту организацию, которая на момент оценки имела просрочку по кредитам более 90 дней. Все такие предприятия были обозначены цифрой 1, а нормально функционирующие – цифрой 0. Таким образом, получена выборка, имеющая бинарную классификацию.

В зависимости от вероятности прогнозирования модели наблюдение может принадлежать любому из следующих типов:

- TP – истинно положительный (если наблюдение правильно классифицировано как положительное);
- FP – ложноположительный (если наблюдение ошибочно классифицировано как положительное);
- TN – истинно отрицательный (если наблюдение правильно классифицировано как отрицательное);
- FN – ложноотрицательный (если наблюдение ошибочно классифицируется как отрицательное).

Для моделирования по варианту 1 (см. рис. 3) использовалась обучающая выборка в 600 дефолтных случаях (наблюдений) и 600 нормально функционирующих случаях (наблюдений). Для варианта 2 применялась та же обучающая выборка. Все используемые коэффициенты оказались значимыми, что позволило сопоставить оба варианта построения моделей. На данной выборке была построена линейная вероятностная модель. Статистические значения показателей представлены в табл. 1 и 2.

Таблица 1

Статистическое значение показателей при одношаговом построении по варианту 1

Table 1

Statistical value of indicators for one-step construction according to option 1

Переменные	*Z*	Вероятность (*Z*)	LR	Вероятность (LR)
k_2	–3,06	0,002	–	–
k_{22}	2,616	0,009	–	–
k_{28}	–7,784	0	–	–
Модель в целом	–	–	1571	0

Примечание. *Z* – стандартизированная оценка; LR – отношение правдоподобия; k_2 – коэффициент абсолютной ликвидности; k_{22} – коэффициент оборачиваемости запасов; k_{28} – коэффициент финансовой автономии.

При одношаговом построении (см. табл. 1) применяется только общее пространство классификационных признаков, а при двухшаговом (см. табл. 2) сначала используется пространство, которое есть только в выборке 1, затем на полученных в результате применения модели (шаг 1) новых классах применяется общее пространство классификационных признаков.

Логит-модель, полученная при одношаговом построении по варианту 1, будет иметь вид

$$y = –2{,}55k_2 + 0{,}009k_{22} – 37{,}81k_{28}. \quad (5)$$

Уравнение выражает зависимость кредитоспособности от значений коэффициентов абсолютной ликвидности, оборачиваемости запасов, финансовой автономии.

Названные коэффициенты были выбраны исходя из их значимости при обоих вариантах построения: даже если коэффициент был значим (т. е. имел *p*-значение более 0,05) при построении одной модели, но при этом не был значим при построении другой, то он не использовался.

Коэффициент абсолютной ликвидности имеет *p*-значение, равное 0,002, коэффициент оборачиваемости запасов – 0,009 и коэффициент финансовой автономии – 0 (см. табл. 1), что меньше 0,05. Таким образом, названные коэффициенты являются значимыми. Наиболее естественный критерий качества – вероятность ошибки при оценке прогнозируемых альтернатив. Понятно, что хорошая модель должна давать высокий процент правильных предсказаний.

Коэффициент абсолютной ликвидности рассчитывался как отношение суммы финансовых вложений и денежных средств к краткосрочным обязательствам за вычетом резервов предстоящих расходов. Он показывает, какую часть текущих обязательств компания способна погасить за счет собственных средств и в кратчайшие сроки. Чем большую долю краткосрочных обязательств предприятие может погасить, тем более устойчивым его можно считать. Однако большие остатки денежных средств могут свидетельствовать об их неэффективном использовании, так как денежные средства целесообразно либо реинвестировать, либо инвестировать в другие предприятия, либо использовать для вознаграждения

сотрудников, акционеров. Как правило, рост коэффициента абсолютной ликвидности говорит об улучшении финансового положения предприятия (в данном случае оно становится более финансово устойчивым и платежеспособным), а снижение – об ухудшении финансового состояния предприятия. Иногда снижение коэффициента абсолютной ликвидности может указывать на повышение эффективности использования активов. Чаще всего это происходит тогда, когда значение коэффициента абсолютной ликвидности значительно выше норматива коэффициента.

Таблица 2

Статистическое значение показателей при двухшаговом построении по варианту 2

Table 2

Statistical value of indicators for two-step construction according to option 2

Шаг 1					
Переменные	*t*	Вероятность (*t*)	*F*	Вероятность (*F*)	R_2
с	–7,824	0	–	–	–
k_5	4,959	0	–	–	–
k_6	–7,828	0	–	–	–
k_7	19,32	0	–	–	–
Модель в целом	–	–	2416	0	0,85

Шаг 2				
Переменные	*t*	Вероятность (*t*)	LR	Вероятность (LR)
k_2	–4,588	0	–	–
k_{22}	–4,358	0	–	–
k_{28}	–23,298	0	–	–
Модель в целом	–	–	2114	0

Примечание. *t* – критерий Стьюдента; *F* – критерий Фишера; LR – отношение правдоподобия; R_2 – коэффициент детерминации; k_5 – коэффициент обеспеченности финансовых обязательств активами; k_6 – коэффициент качества дебиторской задолженности; k_7 – коэффициент качества кредиторской задолженности; с – свободный член (пересечение) линии оценки.

Коэффициент оборачиваемости запасов равен отношению себестоимости проданных товаров к среднегодовой величине запасов. Он показывает, сколько раз в среднем продаются запасы предприятия за некоторый период времени. Стоит отметить, что поскольку в модели (5) соответствующий коэффициент равен 0,009, то он не оказывает значительного влияния на результат оценки кредитоспособности.

Коэффициент финансовой автономии характеризует отношение собственного капитала к общей сумме капитала (активов) организации. Он показывает, насколько организация независима от кредитования. Чем выше названный коэффициент, тем более устойчива компания.

При двухшаговом построении многофакторная множественная вероятностная модель, полученная методом наименьших квадратов (шаг 1) по варианту 2, будет иметь вид

$$y = –0,07 + 1,12k_5 + 0,31k_6 + 0,756k_7.$$

Уравнение выражает зависимость кредитоспособности предприятия в момент наблюдения от коэффициента обеспеченности финансовых обязательств активами, коэффициента качества дебиторской задолженности и коэффициента качества кредиторской задолженности.

Коэффициент качества дебиторской задолженности определяется как доля просроченной дебиторской задолженности в дебиторской задолженности всего. Коэффициент качества кредиторской задолженности рассчитывается как доля просроченной кредиторской задолженности в кредиторской задолженности всего. Коэффициент обеспеченности финансовых обязательств активами определяется как отношение долгосрочных и краткосрочных обязательств за вычетом резервов предстоящих платежей к общей величине активов.

Названные коэффициенты можно признать экономически значимыми, поскольку их меньшее значение соответствует более высокой кредитоспособности, что полностью отвечает их экономической интерпретации.

Значимость линейного коэффициента корреляции проверяется с помощью *t*-критерия Стьюдента. При этом выдвигается и проверяется гипотеза о равенстве этого коэффициента нулю. Если гипотеза подтверждается, то *t*-статистика имеет распределение Стьюдента. Если расчетное значение $t_{\text{расч}} > t_{\text{кр}}$, то гипотеза о статистической незначимости коэффициента отвергается, что свидетельствует о значимости линейного коэффициента корреляции, а следовательно, и о статистической значимости зависимости между *y* и объясняющими переменными. Значение $t_{\text{кр}} = 1{,}962$, что по модулю меньше, чем каждое из расчетных значений *t*-статистик. Исходя из вышеизложенного, можно сделать вывод о том, что линейные коэффициенты корреляции значимы.

Была выдвинута нулевая гипотеза об отсутствии значимой взаимосвязи выбранных переменных со случаями наличия кредитоспособности у предприятий. Для проверки данной гипотезы проведено сравнение *p*-значений с уровнем значимости 0,05. У каждой из переменных *p*-значение равно нулю, что меньше выбранного уровня значимости. Таким образом, нулевая гипотеза отклоняется. Это говорит о высоком уровне статистической значимости выбранных параметров.

Для проверки уравнения выдвинем гипотезу H_0 о статистической незначимости коэффициента детерминации и противоположную ей гипотезу H_1 о статистической значимости коэффициента детерминации:

$$H_0 : R^2 = 0,$$
$$H_1 : R^2 \neq 0.$$

Расчетное значение *F*-статистики равно 2416, а табличное – 2,612, что показывает значимость уравнения в целом. (Поскольку *F*-статистика больше $F_{\text{кр}}$, то в итоге принимается гипотеза H_1 о статистической значимости коэффициента детерминации.)

Коэффициент детерминации равен 0,85 (см. табл. 2), что можно интерпретировать как очень высокую зависимость между переменной и объясняющими переменными. Поскольку рассматривается пространственная выборка, то автокорреляции быть не может, она возникает в случае упорядоченных наблюдений, т. е. тогда, когда мы имеем дело с временным рядом, и ошибка текущего наблюдения зависит от ошибки прошлых наблюдений.

Далее на основании модели, применяемой на шаге 1 по варианту 2, в зависимости от *Z-score*[3] дефолтные и недефолтные наблюдения разбивались на классы. В результате среди функционирующих без просрочек недефолтных классов были получены три более устойчивых и менее устойчивых класса и три дефолтных класса, отличающихся глубиной дефолта. В обучающей выборке числовые значения *Z-score* в случаях представленных дефолтных и недефолтных классов предприятий не пересекаются. Графическое изображение линейной вероятностной модели и логит-модели представлено на рис. 4.

Таким образом, был получен своего рода опорный вектор для строительства логистической модели множественного упорядоченного выбора. После разбиения выборки на новые классы на данной выборке строилась логистическая модель множественного упорядоченного выбора, на основании которой выделялись шесть классов (три дефолтные и три недефолтные).

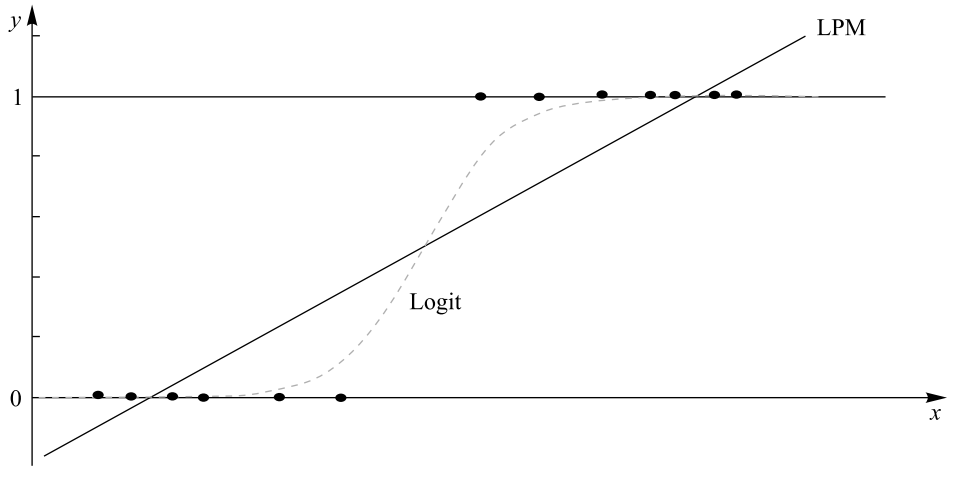

Рис. 4. Сопоставление изображения
линейной вероятностной модели (LPM) и логит-модели (logit)

Fig. 4. Comparison of the image LPM and the logit

[3]Результаты, полученные при построении модели 1 (реализация шага 1 двухшаговой модели) по варианту 2, для каждого случая наблюдения за предприятием.

Из условия максимума логарифмической функции правдоподобия для рассматриваемой модели, которая строится на основании разбиения множества значений переменной y_i на шесть классов, находятся оценки параметров модели, а также пороговых значений.

Логит-модель, полученная при двухшаговом построении (шаг 2 по варианту 2), будет иметь вид

$$y = -0{,}29k_2 - 0{,}02k_{22} - 14{,}35k_{28}.$$

Вероятность (LR-статистика) равна нулю, что показывает значимость уравнения в целом.

Для сопоставления вышеуказанных моделей была построена матрица сопряженности (табл. 3).

После построения каждая из моделей была апробирована на тестовой выборке, в нее вошли 450 случаев наблюдений за предприятиями, относящимися к дефолтным, и 449 случаев наблюдений за предприятиями, относящимися к нормально функционирующим. Некоторые результаты апробации построены по варианту 1 (табл. 4).

Таблица 3

Сопряженность между истинной классификацией и ее оценкой

Table 3

Relationship between true classification and its assessment

Название классов		Истинная классификация		Суммы случаев по классификации, оцененной моделью
		Дефолтные (проблемные)	Нормально функционирующие	
Классификация, оцененная по модели	Дефолтные (проблемные)	TP	FP	TP + FP
	Нормально функционирующие	FN	TN	FN + TN
Суммы случаев по истинной классификации		TP + FN	FP + TN	Проверка равенства суммы случаев по классификациям

Таблица 4

Сопряженность между истинной классификацией и ее оценкой для логит-модели, построенной по варианту 1

Table 4

Relationship between true classification and its assessment for logit model constructed according to option 1

Классы, %		Сумма	Фактический класс*	Проверка**
0	I			
0	100,00	100	1	1
1,52	98,48	100	1	1
0,74	99,26	100	1	1
0	100,00	100	1	1
0	100,00	100	1	1
84,43	15,57	100	0	1
99,88	0,12	100	0	1
56,49	43,51	100	0	1
64,09	35,91	100	0	1
47,41	52,59	100	0	0

Примечание. * – цифра 1 означает случай наблюдения за предприятием, относящимся к дефолтному, цифра 0 означает случай наблюдения за предприятием, относящимся к нормально функционирующему; ** – цифра 1 значит верно, цифра 0 значит ошибочно; зеленым цветом обозначен недефолтный класс, красным – дефолтный.

Исходя из данных табл. 4, можно заключить, что если значение по классу 0 больше значения по классу I, то предприятие функционирует нормально. Если значение по классу I больше значения по классу 0, то предприятие является дефолтным. Сумма вероятностей должна равняться единице. Происходит сверка с истинным значением конкретного случая (дефолт или недефолт) (см. табл. 4, стб. 4). Затем делается вывод о том, верно или ошибочно был классифицирован случай (см. табл. 4, стб. 5).

Результаты логистической модели множественного упорядоченного выбора на базе общих для обеих выборок показателей (вариант 1) представлены в табл. 5.

Таблица 5

Сопряженность между истинной классификацией и ее оценкой на основе логистической модели множественного упорядоченного выбора на базе общих показателей для выборок 1 и 2 (вариант 1)

Table 5

The relationship between the true classification and its assessment based on the logistic model of multiple ordered choice based on indicators common sample for the 1 and 2 (option 1)

Название классов		Истинная классификация		Суммы случаев по классификации, оцененной моделью
		Дефолтные (проблемные)	Нормально функционирующие	
Классификация, оцененная по модели	Дефолтные (проблемные)	447	369	816
	Нормально функционирующие	3	80	83
Суммы случаев по истинной классификации		450	449	899

Чувствительность, или полнота (англ. *sensitivity*), модели рассчитывается как частное от деления общего количества верно определенных дефолтных наблюдений на сумму общего количества верно определенных как дефолтные и общего количества ошибочно определенных как недефолтные наблюдений за предприятиями (TP / (TP + FN)) и равна 99 %.

Специфичность (англ. *specificity*) модели вычисляется как частное от деления суммы верно определенных недефолтных наблюдений на сумму верно определенных недефолтных и ошибочно определенных как дефолтные наблюдений (TN / (TN + FP)) и равна 18 %.

Точность (англ. *precision*) модели показывает, сколько из предсказанных дефолтных наблюдений оказались действительно дефолтными (TP / (TP + FP)), и равна 55 %.

F-мера (F_1) модели 1, построенной по варианту 1, равна 0,706. Она позволяет одновременно оценить точность и чувствительность модели.

Сравнив эти показатели с данными, которые будут получены за счет изменения алгоритма, построенного на тех же показателях, мы узнаем, дает ли новый алгоритм преимущества (табл. 6).

Таблица 6

Сопряженность между истинной классификацией и ее оценкой для логит-модели, построенной по варианту 2

Table 6

Relationship between true classification and its assessment for logit model constructed according to option 2

Интерпретация результатов										Фактический класс*
Классы, %						Сумма, %	Классы, %		Сумма, %	
I	II	III	IV	V	VI		I–III	IV–VI		
0	0	0	0,22	3,32	96,45	100	0	100	100	1
0,51	3,12	29,77	62,41	3,92	0,27	100	33,40	66,60	100	1
0,39	2,38	24,72	67,04	5,12	0,35	100	27,49	72,51	100	1

Окончание табл. 6
Ending table 6

Интерпретация результатов										Фактический класс*
Классы, %						Сумма, %	Классы, %		Сумма, %	
I	II	III	IV	V	VI		I–III	IV–VI		
0	0,02	0,27	11,31	56,66	31,74	100	0,29	99,71	100	1
0	0,02	0,27	11,56	56,94	31,20	100	0,29	99,71	100	1
1,24	7,18	46,62	43,20	1,65	0,11	100	55,04	44,96	100	0
41,73	42,21	14,65	1,39	0,03	0	100	98,58	1,42	100	0
2,38	12,74	55,22	28,74	0,86	0,06	100	70,35	29,65	100	0
2,51	13,31	55,62	27,70	0,82	0,05	100	71,43	28,57	100	0
1,04	6,06	43,31	47,49	1,98	0,13	100	50,40	49,60	100	0

Примечание. * – цифра 1 означает случай наблюдения за предприятием, относящимся к дефолтному, цифра 0 означает случай наблюдения за предприятием, относящимся к нормально функционирующему; зеленым цветом обозначен недефолтный класс, красным – дефолтный.

При применении первого способа интерпретации результатов, полученных при построении модели 2 по варианту 2, случай наблюдения за предприятием следует считать дефолтным либо нормальным исходя из сравнения сумм по классам I–III и IV–VI. Если сумма больше в классах I–III, то такие предприятия являются нормально функционирующими, а если сумма больше в классах IV–VI, то эти предприятия будут относиться к дефолтным (проблемным).

При использовании второго способа интерпретации результатов, полученных при построении модели 2 по варианту 2, отнесение предприятия к нормально функционирующему или дефолтному происходит так же, как описано выше.

Гибридная логистическая модель множественного упорядоченного выбора сначала строится на основе выборки 1 (вариант 2), затем применяются недостающие данные выборки 2 (обучающая выборка). Далее на базе выборок были получены результаты, которые представлены в табл. 7.

Таблица 7

Интерпретация результатов логит-модели, построенной по варианту 2 (способ 1)

Table 7

Interpretation of the results of the logit model constructed according to option 2 (way classification 1)

Название классов		Истинная классификация		Суммы случаев по классификации, оцененной моделью
		Дефолтные (проблемные)	Нормально функционирующие	
Классификация, оцененная по модели	Дефолтные (проблемные)	444	335	779
	Нормально функционирующие	6	114	120
Суммы случаев по истинной классификации		450	449	899

Чувствительность модели составляет 99 %, т. е. она фактически не изменилась, а специфичность модели равна 25 %, что на 7 процентных пунктов выше, чем у модели, построенной согласно варианту 1. Точность (количество предсказанных случаев наблюдения за предприятиями, относящимися к дефолтным) равна 57 %, это на 2 процентных пункта выше по сравнению с моделью, построенной с учетом варианта 1.

F_1 модели 2, построенной по варианту 2 и интерпретированной способом 1, равна 0,722; F_1 модели 1, построенной по варианту 1, равна 0,706; F_1 модели 2, построенной по варианту 2 и интерпретированной способом 1, больше F_1 модели 1, построенной по варианту 1.

Следовательно, у логит-модели, построенной по варианту 2, качество лучше, чем у логит-модели, построенной по варианту 1, так как у нее выше точность и чувствительность.

Результаты, полученные при классификации способом 2, представлены в табл. 8.

Таблица 8

Интерпретации результатов логит-модели, построенной по варианту 2 (способ 2)

Table 8

Interpretation of the results of the logit model constructed according to option 2 (way classification 2)

Название классов		Истинная классификация		Суммы случаев по классификации, оцененной моделью
		Дефолтные (проблемные)	Нормально функционирующие	
Классификация, оцененная по модели	Дефолтные (проблемные)	444	344	788
	Нормально функционирующие	6	105	111
Суммы случаев по истинной классификации		450	449	899

Чувствительность модели составляет 99 %, т. е. она фактически не изменилась, а специфичность модели равна 23 %, что на 5 процентных пунктов выше, чем у модели 1, построенной в соответствии с вариантом 1, но на 2 процентных пункта ниже по сравнению с чувствительностью модели 2, построенной по варианту 2, при классификации способом 1. Точность (количество предсказанных случаев наблюдений за предприятием, относящимся к дефолтным) равна 56 %, что на 1 процентный пункт выше, чем точность модели, построенной по варианту 1.

F_1 модели 2, построенной по варианту 2 и интерпретированной способом 2, равна 0,717; F_1 модели 1, построенной по варианту 1, равна 0,706; F_1 модели 2, построенной по варианту 2 и интерпретированной способом 2, больше F_1 модели 1, построенной по варианту 1.

Итак, у модели 2 качество лучше, так как у нее выше точность и чувствительность.

Заключение

Для наибольшей сопоставимости вариантов построения моделей в настоящей работе использовались именно те коэффициенты, которые будут значимы при обоих вариантах (методах) построения. Таким образом, удалось повысить качество модели, построенной на одних и тех же данных и на одних и тех же показателях. На исходных данных с применением варианта 2 можно построить модель даже с большими прогностическими способностями, однако их нельзя будет сопоставить. Результаты моделирования объясняются тем, что многие государственные предприятия пользуются различного рода государственной поддержкой и не могут самостоятельно и стабильно функционировать, поэтому часть из них, которые попали в выборку как недефолтные, на самом деле являются дефолтными.

Модель, построенная по варианту 2, может использоваться для выявления скрытых дефолтов и способствовать кредитованию только нормально функционирующих предприятий либо осознанному кредитованию проблемных предприятий, у которых есть возможность погашения кредитов с помощью поддержки государства. Определение скрытых дефолтов – это очень важная работа, поскольку многие государственные предприятия пользуются государственной поддержкой и не могут стабильно функционировать без дотаций. В результате часть случаев наблюдений за предприятием попадают в выборку как недефолтные, фактически будучи таковыми. Предложенная модель позволяет выявить такие компании и более точно оценить вероятность их дефолта.

Прогнозируется улучшение полученных результатов путем создания методики, предусматривающей совместное применение гибридной логистической модели множественного упорядоченного выбора и модели, построенной на другой выборке данных, объединенных на основании матричного подхода [11].

Библиографические ссылки

1. Тотьмянина КМ. Обзор моделей вероятности дефолта. *Управление финансовыми рисками*. 2011;1:12–24.
2. Ohlson JA. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*. 1980;18:109–131. DOI: 10.2307/2490395.
3. Kubecová J, Vrchota J. The Taffler's model and strategic management. *The Macrotheme Review* [Internet]. 2014 [cited 2021 March 15];3(2). Available from: http://macrotheme.com/yahoo_site_admin/assets/docs/16MR31Ku.1354035.pdf.
4. Карминский АМ, Моргунов АВ, Богданов ПМ. Оценка вероятности дефолта сделок проектного финансирования. *Журнал Новой экономической ассоциации*. 2015;2(26):99–122.
5. Малюгин ВИ. *Методы анализа многомерных эконометрических моделей с неоднородной структурой*. Минск: БГУ; 2014. 351 с.
6. Савицкая ГВ. *Экономический анализ*. 10-е издание. Москва: Новое знание; 2008. 640 с.
7. Малюгин ВИ, Гринь НВ, Милевский ПС, Зубович АИ, редакторы. *Система статистических кредитных рейтингов предприятий: методика построения, верификации и применения*. Минск: Национальный банк Республики Беларусь; 2013. 75 с. (Банкаўскі веснік. Исследования банка № 5).
8. Малюгин ВИ, Пытляк ЕВ. Оценка устойчивости банков на основе эконометрических моделей. *Банкаўскі веснік*. 2007; 2:30–36.
9. Шитиков ВК, Розенберг ГС, Зинченко ТД. *Количественная гидроэкология: методы системной идентификации*. Тольятти: ИЭВБ РАН; 2003. 463 с.
10. Магнус ЯР, Катышев ПК, Пересецкий АА. *Эконометрика. Начальный курс*. Москва: Дело; 2004. 576 с.
11. Ткачёв АИ, Шипунов АВ. Системы кредитного скоринга. Матричный подход. *Банкаўскі веснік*. 2019;10(674):37–46.

References

1. Totmianina KM. Review of models of default of probability. *Upravlenie finansovymi riskami*. 2011;1:12–24. Russian.
2. Ohlson JA. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*. 1980;18:109–131. DOI: 10.2307/2490395.
3. Kubecová J, Vrchota J. The Taffler's model and strategic management. *The Macrotheme Review* [Internet]. 2014 [cited 2021 March 15];3(2). Available from: http://macrotheme.com/yahoo_site_admin/assets/docs/16MR31Ku.1354035.pdf. Russian.
4. Karminsky AM, Morgunov AV, Bogdanov PM. Assessment of the probability of default of project finance transactions. *Journal of the New Economic Association*. 2015;2(26):99–122. Russian.
5. Malygin VI. *Metody analiza mnogomernykh ekonometricheskikh modelei s neodnorodnoi strukturoi* [Methods for the analysis of multivariate econometric models with a heterogeneous structure]. Minsk: Belarusian State University; 2014. Russian.
6. Savitskaya GV. *Ekonomicheskii analiz* [Economic analysis].10th edition. Moscow: Novoe znanie; 2008. 640 p. Russian.
7. Malygin VI, Grin NV, Milevsky PS, Zubovich AI, editors. *Sistema statisticheskikh kreditnykh reitingov predpriyatii: metodika postroeniya, verifikatsii i primeneniya* [The system of statistical credit ratings of enterprises: a method of construction, verification and application]. Minsk: National Bank of the Republic of Belarus; 2013. 75 p. (Bankawski vesnik. Issledovaniya banka № 5). Russian.
8. Malygin VI, Pytlyak EV. [Assessment of the stability of banks on the basis of econometric models]. *Bankawski vesnik*. 2007; 2:30–36. Russian.
9. Shitikov VK, Rosenberg GS, Zinchenko TD. *Kolichestvennaya gidroekologiya: metody sistemnoi identifikatsii* [Quantitative hydroecology: methods of systemic identification]. Togliatti: Institute of Ecology of Volga Baisin of the Russian Academy of Sciences; 2003. 463 p. Russian.
10. Magnus YaR, Katyshev PK, Peresetskiy AA. *Ekonometrika. Nachal'nyi kurs* [Econometrics. Initial course]. Moscow: Delo; 2004. 576 p. Russian.
11. Tkatchev AI, Shipunov AV. [Credit scoring systems. Matrix approach]. *Bankawski vesnik*. 2019;10(674):37–46. Russian.

Статья поступила в редколлегию 21.03.2021.
Received by editorial board 21.03.2021.