

# ДИНАМИКА ПРОГНОЗНОЙ СИЛЫ МОДЕЛЕЙ БАНКРОТСТВА ДЛЯ СРЕДНИХ И МАЛЫХ РОССИЙСКИХ КОМПАНИЙ ОПТОВОЙ И РОЗНИЧНОЙ ТОРГОВЛИ

Демешев Б.Б.<sup>1</sup>, Тихонова А.С.<sup>2</sup>

Цель данной работы — изучение динамических особенностей линейных и нелинейных методов прогнозирования вероятности банкротства средних и малых российских непубличных компаний отрасли оптовой и розничной торговли. *Моделирование учитывает финансовые и нефинансовые показатели* до и после кризиса 2008—2009 гг. В работе используются два метода прогнозирования: логит-модели и алгоритм случайного леса.

В исходной выборке содержится от 200 до 600 тысяч компаний в зависимости от года. Данные взяты из базы данных Руслана и относятся к периоду 2004—2012 гг.

Понятие банкротства расширено до понятия закрытия из-за серьезных финансовых сложностей, несовместимых с дальнейшим продолжением деятельности компании. Исследуются активные компании и два типа неактивных: компании, ликвидированные в результате легального банкротства, и добровольно ликвидированные компании.

Вместе с финансовыми отношениями, которые рассчитаны с помощью информации из финансовой отчетности, в моделях учитывается неоднородность фирм по возрасту, размеру, территориальному признаку и организационной форме.

Прогнозы строятся вне обучающей модели и обладают довольно высокой прогнозной силой, площадь под ROC-кривой достигает 0,75. Лучшим методом оказался случайный лес, что подтверждает тот факт, что зависимость между финансовыми, нефинансовыми переменными и вероятностью дефолта является нелинейной. Добавление нефинансовых переменных, таких как федеральный округ, возраст компании, улучшает качество прогнозов. Вид организационной формы и размер компании оказывают незначительное влияние на вероятность банкротства. Среди финансовых переменных наиболее важными оказались показатели рентабельности, ликвидности и финансового рычага. Более того, был обнаружен структурный сдвиг, который, вероятно, был спровоцирован кризисом 2008—2009 гг.

Прогнозирование банкротства средних и малых предприятий необходимо банкам и иным кредитным организациям, принимающим решение о предоставлении кредита фирмам.

**Ключевые слова:** прогнозирование банкротства, сравнение моделей, средние и малые предприятия, оптовая и розничная торговля, алгоритм случайного леса, логит-модель.

**JEL:** C14, C45, G30, G33

## Введение

Хотя концепция кредитоспособности сама по себе не нова, интерес к ней сильно повысился в последние годы, после кризиса 2008—2009 гг., когда отсутствие кредитоспособности плательщиков привело к дефолту банков и поставило всю финансовую систему в затруднительную ситуацию. До Базельского соглашения (2007) требования к капиталу крупных и малых компаний были более размыты, однако после этого соглашения различия стали настолько ощутимы, что у банков появились новые стимулы для построения различных моделей для фирм разного размера.

Цель данной работы — изучение динамики прогнозной силы и важности переменных в моделях банкротства средних и малых российских непубличных компаний оптовой и розничной торговли. Задачи исследования: сравнить различные подходы к прогнозированию; описать динамику моделей с 2004 по 2012 г.; проверить гипотезу о влиянии нефинансовых показателей на вероятность банкротства; проверить гипотезу о различиях моделей по правовым формам.

1. Старший преподаватель кафедры математической экономики и эконометрики департамента прикладной экономики НИУ ВШЭ.

2. Студентка первого года магистерской программы «Финансовая экономика» МИЭФ НИУ ВШЭ.

Исследования для компаний среднего и малого бизнеса необходимы, потому что они являются прочной основой для экономики страны и способствуют инновациям и развитию. В то же время построение скоринговых систем для компаний такого размера – одна из основных задач банков, которые составляют свой кредитный портфель. Прогнозирование банкротства средних и малых предприятий необходимо банкам и иным кредитным организациям, принимающим решение о предоставлении кредита фирмам.

В данной работе сравниваются линейные и нелинейные статистические методы на примере логит-модели и алгоритма случайного леса. Сравняются модели до и после кризиса 2008–2009 гг., что не было сделано ранее. В первый раз по российским данным сделана попытка оценить и учесть неоднородность по возрасту и формам организации предприятия. Период наблюдения: 2004–2012 гг. Более того, понятие банкротства расширено до понятия закрытия из-за серьезных финансовых сложностей, несовместимых с дальнейшим продолжением деятельности компании, то есть исследуются действующие на данный момент компании, а также два типа неактивных: ликвидированные в результате банкротства и добровольно ликвидированные компании. Важно отметить, что используется отчетность компаний, адаптированная к международным стандартам финансовой отчетности (МСФО).

### Обзор литературы

Размер предприятия накладывает ограничения на допустимый размер долга. Так, для компаний малого бизнеса гораздо сложнее взять кредит, чем для крупных известных корпораций. В работе Колари, Оу и Шин (Kolari, Ou, Shin, 2006) показано, что маленькие компании гораздо более рискованны, чем крупные компании. Более того, многочисленные исследования – такие, как работа Сирираттанафонкун и Паттаразаммас (Sirirattanaphonkun, Pattarathammas, 2012), показали, что эти типы компаний должны изучаться в отдельности. Обычно все компании делят на две группы: компании крупного бизнеса и компании малого и среднего бизнеса. Компании малого и среднего бизнеса, в свою очередь, делятся на три подкатегории: микро-, малые и средние предприятия.

В данной работе в качестве критериев для разделения на разные категории взяты два параметра: численность персонала и выручка. Эти два критерия выбора категории указаны в Федеральном законе № 209-ФЗ «О развитии малого и среднего предпринимательства в Российской Федерации». В таблице 1 представлены требования к каждому показателю для компаний. Компания каждого размера должна удовлетворять обоим критериям одновременно.

*Таблица 1*

**Критерии определения размера компании**

Размер	Численность персонала (человек)	Выручка
Микро-	1–15	≤ €1,4 млн (60 млн руб.)
Малое	16–100	≤ €9,6 млн (400 млн руб.)
Среднее	101–250	≤ €24 млн (1 млрд руб.)

В современной литературе существует два основных определения дефолта. Часть исследователей (Maleev, Nikolenko, 2010) понимают под дефолтом невозможность выплаты платежей процентов или основного тела долга. Другая часть исследователей, которая более многочисленна, в качестве определения дефолта используют понятие легального банкротства и его различные стадии. Среди сторонников этого определения есть и российские, и иностранные авторы (Hunter, Isachenkova, 2001; Vallini, Ciampi, Gordini, 2009; Khorasgani, 2011). Это определение дефолта кажется более логичным, потому что неспособность платить по своим обязательствам в данный момент не означает невозможности погашения задолженности в предусмотренные законом сроки.

Первые шаги в прогнозировании банкротства компаний были сделаны еще в 1960-х гг. В работе Бивера (Beaver, 1966) предложено использовать анализ относительных показателей (одномерный параметрический метод), в работе Альтман (Altman, 1968) был применен ли-

нейный дискриминантный анализ (ЛДА). Основной недостаток метода Бивера состоял в выборе порога отсечения. В то же время модель Альтмана (Z-score model), которая основана на заранее определенных финансовых показателях, не учитывает некоторые источники дохода различных компаний. Более того, некоторые исследования (Lugovskaya, 2009) показали, что наличие нефинансовых переменных в модели сильно влияет на результат.

Следующим этапом в развитии прогнозирования банкротств стало применение логит- и пробит-моделей (Martin, 1977; Ohlson, 1980). Мартин и Олсон показали, что эти методы зачастую превосходят дискриминантный анализ. К тому же в работе Олсона (Ohlson, 1980) предложено оценивать вероятность дефолта не только на год, но и на два года вперед.

В данной сфере использовался метод опорных векторов (Härdle, Lee, Schäfer, Yeh, 2007), а также нейронные сети (Tam, Kiang, 1992; Altman, Marco, Varetto, 1994). В то же время некоторые методы вызывают сомнения, потому что финансовые отношения, рассчитанные с помощью отчета о прибылях и убытках, зачастую не имеют нормального распределения (Ohlson, 1980; Wilson, Sharda, 1994). В работе Вэй. Ли и Чен (Wei, Li, Chen, 2007) также подчеркивается, что ЛДА может неверно классифицировать исходы, так как ковариационные матрицы дефолтных и активных компаний, скорее всего, неидентичны.

Существует множество финансовых показателей, которые рассчитывает каждая компания. По этой причине перед каждым исследователем стоит вопрос, какие из них выбирать. Однако чаще всего выделяют пять групп показателей: рентабельность, ликвидность, оборачиваемость, финансовый рычаг, обслуживание долга (Altman, Sabato, 2007).

Что касается нефинансовых характеристик, необходимо рассматривать не только размер компании, но и ее возраст и организационную форму. Помпе, Билдербик в своей работе (Pompe, Bilderbeek, 2005) отмечают, что прогнозировать вероятность банкротства молодых фирм сложнее, чем давно существующих компаний. В исследовании Фалькенштейн, Борал и Карти (Falkenstein, Boral, Carty, 2000) отмечается, что связь между финансовыми показателями и риском дефолта различна для публичных и непубличных компаний.

Есть немало работ, в которых рассматривается прогнозирование дефолтов именно компаний среднего и малого бизнеса (Edmister, 1972; Altman, Sabato, 2007; Vallini, Ciampi, Gordini, 2009).

В то же время работ по прогнозированию банкротства российских компаний не так много, как зарубежных. В 1990-е гг. это было связано с коренными переменами в регулировании бизнеса, законах и ведении бухгалтерского учета. Однако в середине 1990-х гг. появилась работа по сравнению банкротств российских и британских фирм (Hunter, Isachenkova, 2001). Позже попытки Зайцевой, Сайффулина и Кадикова адаптировать модель Альтмана (Z-score model) и модель Олсона (O-score model) оказались неудачными.

В работе Луговской (Lugovskaya, 2009) анализируются российские дефолтные компании с помощью ЛДА, в работе Жданова и Афанасьевой (Жданов и Афанасьева, 2011) используются ЛДА и логит-модель, в работе Макеевой и Бакуровой (Макеева и Бакурова, 2006) применяются нейронные сети, а другие авторы (Федорова, Гиленко и Довженко, 2013; Makeeva, Neretina 2013) концентрируются исключительно на логит- и пробит-моделях.

Кроме Луговской, все изучают одну отрасль, например строительство, нефтегазовую или авиастроительную отрасль, и не учитывают размер компаний. К тому же размер выборки во многих случаях крайне мал. Выбор финансовых отношений ставится под вопрос, так как отчетность по российским стандартам отличается от отчетности по международным, поэтому неверно было бы основываться на иностранных источниках для выбора показателей. Более того, к сожалению, в большинстве работ правовые формы никак не обозначены.

## Описание данных

В данной работе исследуются непубличные российские компании среднего и малого бизнеса, относящиеся к оптовой и розничной торговле (по ОКВЭД). Определение размера соответствует критериям, представленным в Федеральном законе Российской Федерации № 209-ФЗ

«О развитии малого и среднего предпринимательства в Российской Федерации». Два типа непубличных компаний – это общества с ограниченной ответственностью (ООО) и закрытые акционерные общества (ЗАО), именно они составляют основную массу микро-, малых и средних компаний.

Понятие легального банкротства расширено до понятия серьезных финансовых сложностей. В анализ включены активные компании и компании, которые из-за финансовых сложностей были ликвидированы, то есть ликвидированные банкроты и ликвидированные добровольно.

Это объединение сделано из-за того, что в соответствии с процедурой банкротства и ликвидации компаний оба типа сталкиваются с невозможностью продолжения работы, отличие лишь в одном: банкроты накопили столько долга, что уже не могут по нему расплатиться, в то время как добровольно ликвидированные компании закрываются, так как бизнес приносит только убытки, а продать его не получается. Однако важно отметить, что если компания хочет добровольно прекратить свою работу, но стоимость имущества не покрывает все обязательства, то компания будет ликвидирована по процедуре банкротства.

Данные по российским фирмам собраны из базы данных российских, украинских и казахских компаний RUSLANA (ruslana.bvdep.com). Период исследования: 2004–2012 гг. Данные годовые.

В таблице 2 представлены нефинансовые переменные, которые были изначально в данных или которые были созданы на основе имеющейся информации.

Таблица 2

**Краткое описание переменных**

Переменные	Описание
Имеющиеся в данных	
Организационная форма	Закрытое акционерное общество Общество с ограниченной ответственностью
Статус	Активное Добровольно ликвидированное Ликвидированное в результате банкротства
Дата статуса	Дата ликвидации (если предприятие было ликвидировано)
Созданные	
Возраст	Возраст компании в годах
Дата создания	Если дата не содержала дня, то ставится середина месяца (гггг-мм-15). Если дата не содержала ни дня, ни месяца, то ставится середина года (гггг-07-01)
Дефолт	Если компания обанкротится в текущем году – 1 Если компания не обанкротится в текущем году – 0
Дефолт в следующем году	Если компания обанкротится в следующем году – 1 Если компания не обанкротится в следующем году – 0
Федеральный округ	Укрупнение административного деления до федеральных округов
Последний доступный размер	Микро-, малое и среднее

Текущий год – каждый год из изучаемого периода от 2004 до 2012 г. в зависимости от того, на какой год строится модель, а текущая дата – последний день текущего года (гггг-12-31).

Опишем создание переменной размера компании. Для каждого года берутся количество персонала компании и величина выручки. С их помощью производится деление в соответствии с законом Российской Федерации, но есть некоторые уточнения:

1. Если данные по обоим критериям попадают в одну группу, то берется размер этой группы.
2. Если данные попадают в разные группы, то численность персонала выбирается основным критерием для выбора типа размера.
3. Если есть данные только по одному из критериев, то размер определяется только относительно этого критерия.

4. Если ни по одному из критериев нет данных, то ставится NA.

Однако и при таком делении пропущенных значений в переменной размера осталось много. Это логично, потому что обычно в год дефолта компания не предоставляла данные. Поэтому была введена переменная «последний доступный размер», то есть если не было данных о размере на текущий год, то брался размер в предыдущем году, если его тоже не было, то брался размер два года назад и так далее.

Распределение дефолтных и активных компаний по годам в зависимости от размера представлено на графиках 1а и 1б. Общее количество компаний плавно растет с 200 тысяч в 2004 г. до 600 тысяч в 2012 г.

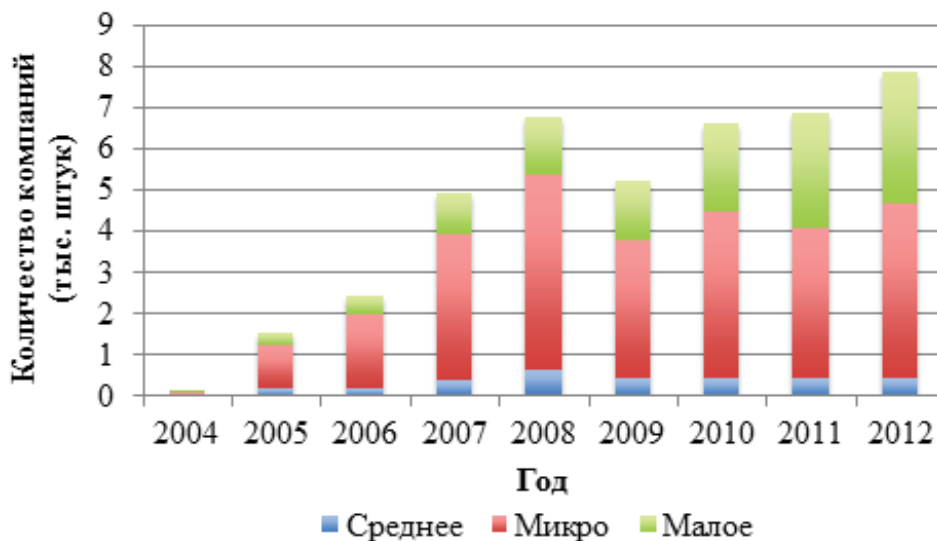


График 1а. Количество дефолтных компаний в зависимости от размера

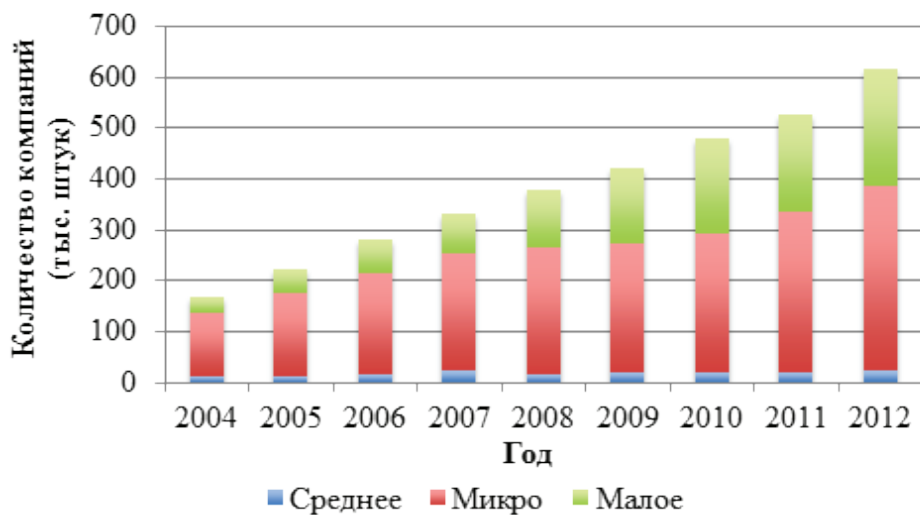


График 1б. Количество активных компаний в зависимости от размера

В выборках доминируют компании микроразмера, на втором месте идут малые предприятия, предприятий среднего размера существенно меньше. Динамика количества банкротств показывает структурные изменения 2008 г.: на год кризиса приходится локальный максимум числа банкротств. Начиная с 2010 г. количество банкротств продолжает расти, что связано с ростом общего количества компаний.

Относительно финансовых переменных: из базы данных были взяты финансовые показатели из баланса и отчета о прибылях и убытках, на основе которых были рассчитаны финансовые отношения. Разные авторы относят показатели в разные группы, но в данной работе классификация содержит пять групп, соответствующих работе Альтмана и Сабато (Altman, Sabato, 2007).

Показатели финансового рычага характеризуют то, как соотносятся заемные и собственные средства компании, отражая финансовую устойчивость фирмы и ее степень риска. Эти отношения отражают эффект увеличения прибыли за счет взятия дополнительного долга из-за недостатка собственного капитала.

Показатели ликвидности характеризуют то, насколько быстро компания может превратить имеющиеся материальные ценности в деньги для покрытия своих финансовых обязательств. Чем больше у фирмы неликвидных активов, тем выше вероятность того, что она не расплатится по долгам.

Показатели рентабельности характеризуют степень эффективности использования средств компании. Они отражают, покрывает ли компания свои затраты доходами, и если да, то получает ли она прибыль.

Показатели обслуживания долга характеризуют кредитоспособность предприятия, отражая то, как быстро оно совершает выплаты по долгу и какая доля денежного потока уходит на выплату процентов или основного тела долга.

Показатели активности характеризуют уровень деловой активности предприятия, то есть оборачиваемость средств компании. Они помогают ей определить необходимый уровень оборотных средств, так как при ускорении оборачиваемости потребность в оборотном капитале падает.

Далее, в таблице 3, мы приводим все финансовые отношения с указанием формулы, также в этой таблице выписаны все нефинансовые переменные. Деление переменных по группам проводилось с помощью прочитанных статей и собственного анализа.

### Алгоритмы оценивания

В данной работе построены отдельные модели для каждого года для прогнозирования дефолта на год вперед, то есть были взяты текущие значения объясняющих переменных и будущее (на следующий год) значение зависимой переменной:

$$P(\text{def}_{i,t+1}) = f_t(x_{i,t}) \quad (1)$$

где  $i$  – идентификационный номер компании,  $t$  – период,  $\text{def}_{i,t+1}$  – индикатор дефолта (равен 1, если компания  $i$  стала банкротом в году  $t$ , и равен 0, если иначе),  $x_{i,t}$  – вектор-строка характеристик компании  $i$  в году  $t$ ,  $f_t$  – функция, описывающая зависимость в году  $t$ .

В статье мы приводим только результаты оценивания по балансированным выборкам. В балансированных выборках содержится равное число банкротов и активных компаний. Для получения балансированной выборки мы берем все компании-банкроты за данный период и к ним добавляем равное количество случайным образом выбранных активных компаний. Результаты по несбалансированным выборкам схожи не только качественно, но и количественно и поэтому не приводятся. Для проверки устойчивости результатов использовались несбалансированные выборки, в которых присутствовали все дефолтные компании, но случайно выбранных активных компаний было 20% от всех активных в тот период.

Стоит подчеркнуть, что в данных очень много пропущенных значений, и перед балансировкой данные были очищены от пропущенных значений по переменным, которые включались в модель. В результате количество наблюдений в моделях составило от 1234 до 3986 компаний.

На примере наиболее популярных представителей сравниваются два принципиально различных класса моделей – обобщенные линейные модели и полностью нелинейные модели. Наиболее известной обобщенной линейной моделью для бинарных данных является логит-модель. Среди нелинейных моделей лидирующее положение по популярности занимает алгоритм случайного леса.

Для удобства мы изложим логит-модель для случая двух объясняющих переменных,  $x$  и  $z$ . В логит-модели предполагается, что существует ненаблюдаемая переменная, склонность к

банкротству,  $def_{i,t+1}^*$ , которая линейно зависит от регрессоров:

$$def_{i,t+1}^* = \beta_1 + \beta_2 x_{i,t} + \beta_3 z_{i,t} + \varepsilon_{i,t} \quad (2)$$

Компания оказывается банкротом, если  $def_{i,t+1}^* > 0$ . Связь между наблюдаемым индикатором банкротства и ненаблюдаемой склонностью к банкротству имеет вид:

$$def_{i,t+1} = \begin{cases} 1, & def_{i,t+1}^* > 0 \\ 0, & def_{i,t+1}^* \leq 0. \end{cases} \quad (3)$$

Логит-модель оценивается с помощью метода максимального правдоподобия, при этом предполагается, что величины  $\varepsilon_{i,t}$  независимы и имеют логистическое распределение.

Алгоритм случайного леса состоит в построении большого количества, 500 в данной работе, классификационных деревьев для каждой модели. Особенности построения деревьев следующие:

1. Каждое дерево строится по случайной выборке с повторениями из исходной. Размер случайной выборки равен размеру исходной.
2. При каждом ветвлении дерева случайным образом предварительно отбирается  $[\sqrt{k}]$  штук из исходных регрессоров. Затем из предварительно отобранных регрессоров выбирается тот, который обеспечивает максимальное падение индекса Джини (индекса «неидеальности» классификации).
3. Каждое дерево строится до тех пор, пока не будет достигнута идеальная классификация.

Суммарное падение индекса Джини по всем узлам, связанным с некоторой переменной, можно интерпретировать как важность переменной для классификации. Падение индекса Джини в узле можно определить как изменение (до и после деления узла на два) вероятности несоответствия типов предприятий, если первое выбирается случайно из всей выборки, а второе выбирается случайно из того же терминального узла дерева, что и первое.

### Выбор объясняющих переменных

Большое количество финансовых отношений затрудняет выбор переменных для построения моделей. Мы остановились на двух вариантах:

1. Выбор финансовых переменных, аналогичных модели Альтмана и Сабато (Altman, Sabato, 2007) для предприятий малого и среднего бизнеса, и добавление нефинансовых переменных.
2. Выбор финансовых переменных с помощью критерия частоты использования в других работах (Bellovary, Giacomin, Akers, 2007) и добавление нефинансовых переменных.

В других исследованиях иногда используется метод главных компонент, однако он обладает двумя недостатками. Во-первых, не факт, что зависимая переменная сильнее всего объясняется именно главными компонентами, вполне возможно, что наибольшую объясняющую силу имеют компоненты с малыми собственными числами. Во-вторых, в данных имеется большое количество пропущенных значений, поэтому реализовать метод главных компонент для большого количества переменных невозможно из-за малого количества наблюдений.

В работе Альтмана и Сабато (Altman, Sabato, 2007) предложена модель, предназначенная именно для прогнозирования банкротств небольших фирм. В силу огромного числа пропущенных значений в базе данных RUSLANA или отсутствия некоторых показателей, мы подобрали наиболее близкие к указанным в исследовании этих авторов переменные.

Финансовая переменная  $sr$  (Total equity / Total assets) не присутствовала в модели Альтмана и Сабато, но мы ее добавили, так как она характеризует степень стабильности предприятия. Также мы рассматриваем два способа пополнения модели Альтмана и Сабато нефинансовыми переменными.

Только возрастом:

$$P(def_{i,t+1} = 1) = f_t(iptd_{i,t}, ebta_{i,t}, stdte_{i,t}, roa_{i,t}, liq_{i,t}, sr_{i,t}, age_{i,t}) \quad (4)$$

Всеми нефинансовыми переменными:

$$P(def_{i,t+1} = 1) = f_t(iptd_{i,t}, ebta_{i,t}, stdte_{i,t}, roa_{i,t}, liq_{i,t}, sr_{i,t}, age_{i,t}, fedreg_{i,t}, lasize_{i,t}, legal_{formi,t}) \quad (5)$$

Здесь для компактности записи для нефинансовых качественных переменных используется выражение вида  $fedreg_{i,t}$ , что означает включение соответствующего количества дамми-переменных.

Второй вариант отбора финансовых показателей – частота упоминаний в исследованиях дефолтов компаний в разных странах. Для этого была взята статья Белловери, Джикомино и Акерс (Bellovary, Giacomino, Akers, 2007), в которой приведено сравнение статей по прогнозированию банкротств компаний с 1930 по 2007 г.

Критерием для выбора стало использование показателя в более чем 27 исследованиях, данный порог составляет половину от максимального количества упоминаний.

Вариант отбора финансовых переменных по популярности мы снова дополняем нефинансовыми переменными двумя способами.

Только возрастом:

$$P(def_{i,t+1} = 1) = f_t(iptd_{i,t}, roa_{i,t}, cr_{i,t}, wcta_{i,t}, ebta_{i,t}, lr_{i,t}, tdta_{i,t}, age_{i,t}) \quad (6)$$

Всеми нефинансовыми переменными:

$$P(def_{i,t+1} = 1) = f_t(iptd_{i,t}, roa_{i,t}, cr_{i,t}, wcta_{i,t}, ebta_{i,t}, lr_{i,t}, tdta_{i,t}, age_{i,t}, fedreg_{i,t}, lasize_{i,t}, legal_{formi,t}) \quad (7)$$

В таблице 3 представлены переменные, входящие во все рассматриваемые нами модели.

Таблица 3

Переменные в моделях

Название в данных	Формула	Альтман		Популярные	
		Возраст	Все	Возраст	Все
ebta	$\frac{EBIT}{Total\ assets}$	+	+	+	+
stdte	$\frac{Short - term\ debt}{Total\ equity}$	+	+		
roa	$\frac{Net\ income}{Total\ assets}$	+	+	+	+
liq	$\frac{Cash}{Total\ assets}$	+	+		
iptd	$\frac{Interest\ paid}{Total\ debt}$	+	+	+	+
sr	$\frac{Total\ equity}{Total\ assets}$	+	+		
cr	$\frac{Current\ assets}{Current\ liabilities}$			+	+



lr	$\frac{\text{Current assets} - \text{stocks}}{\text{Current liabilities}}$			+	+
tdta	$\frac{\text{Total debt}}{\text{Total assets}}$			+	+
wcta	$\frac{\text{Working capital}}{\text{Total assets}}$			+	+
age	Возраст компании	+	+	+	+
fedreg	Федеральный округ			+	+
lasize	Последний доступный размер компании			+	+
legal_form	Организационная форма			+	+

### Эмпирические результаты

Для начала сравним качество прогнозов логит-модели и алгоритма случайного леса для всех четырех вариантов объясняющих переменных. Существует много критериев качества прогноза. К сожалению, самый простой критерий качества, доля верно классифицированных предприятий, не очень пригоден в нашем случае. Связано это с тем, что мы оцениваем качество прогнозов на год вперед, а количество предприятий-банкротов примерно равно 1%. В такой ситуации даже тривиальная модель «все предприятия – активные» будет давать долю верно классифицированных предприятий, равную 0,99. Однако эта тривиальная модель бесполезна для кредитора компании, так как не позволяет определить, станет ли компания банкротом.

Мы используем более сложный показатель качества прогнозов – площадь под ROC-кривой. Этот показатель качества прогнозов лежит от 0 до 1. Он равен 0,5 в случае если вероятность банкротства прогнозируется случайным числом от 0 до 1.

В случае идеального качества прогнозов (будущим банкротам предсказывается вероятность банкротства 1, а будущим активным предприятиям предсказывается вероятность банкротства 0) площадь под ROC-кривой равна 1. Также можно интерпретировать площадь под ROC-кривой как вероятность того, что случайное предприятие-банкрот получит более высокую предсказанную вероятность банкротства, чем случайно выбранное активное предприятие.

Площади под ROC-кривой для сравниваемых алгоритмов и наборов объясняющих переменных представлены на графиках 2а и 2б.

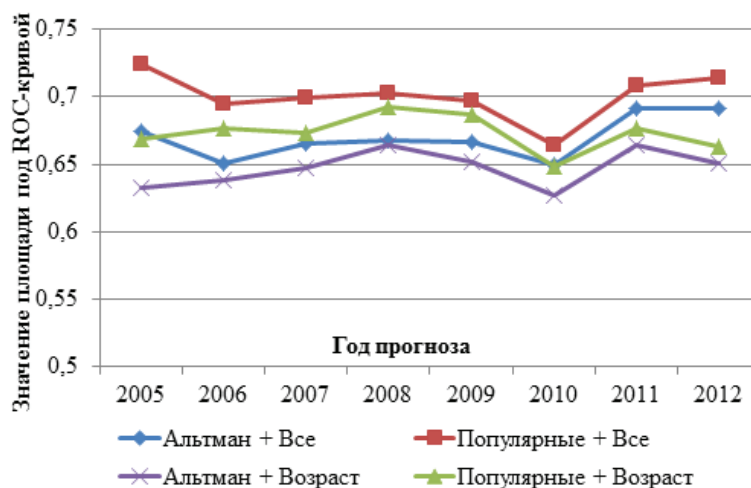


График 2а. Площадь под ROC-кривой (алгоритм случайного леса)

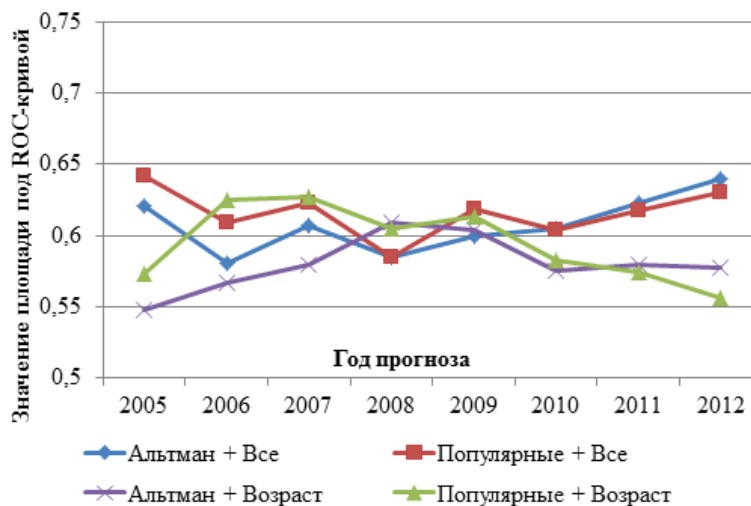


График 2б. Площадь под ROC-кривой (логит-модели)

Из графиков 2а и 2б четко видно несколько интересных фактов.

Во-первых, при любом наборе объясняющих переменных алгоритм случайного леса существенно превосходит логит-модель по качеству прогнозов. Значения площади под ROC-кривой для моделей, оцененных с помощью алгоритма случайного леса, лежат в промежутке от 0,65 до 0,75, а для логит-моделей – в промежутке от 0,55 до 0,65. Из этого следует, что зависимость вероятности банкротства от объясняющих переменных является существенно нелинейной. Кроме нелинейности воздействия объясняющих переменных, вероятно, имеет место неаддитивность эффектов. То есть предельный эффект воздействия одного регрессора на вероятность банкротства нетривиально зависит от значения других регрессоров.

Во-вторых, из нелинейности и неаддитивности зависимости следует, что логит-модель и другие более простые алгоритмы (линейный дискриминантный анализ, квадратичный дискриминантный анализ) малопригодны для прогнозирования банкротств средних и малых российских компаний.

В-третьих, значение площади под ROC-кривой около 0,7 для прогнозов вне выборки, по которой производилось оценивание, считается довольно высоким, что говорит о приемлемости алгоритма случайного леса для прогнозирования банкротств.

В-четвертых, если попарно сравнивать модели с одинаковым набором финансовых показателей и разным набором нефинансовых характеристик, то для алгоритма случайного леса заметно превосходство моделей с добавлением всех нефинансовых показателей. Для логит-моделей прослеживается такая же зависимость, но не столь явно.

В-пятых, наилучшей моделью для алгоритма случайного леса стала модель, включающая в себя финансовые переменные, отобранные по упоминаемости в исследованиях банкротства компаний, и нефинансовые переменные: возраст, организационная форма, федеральный округ и последний доступный размер компании.

В-шестых, на графике 2а виден структурный сдвиг. Модель 2009 г. хуже прогнозирует на 2010 г., что выступает в пользу утверждения о том, что зависимость между вероятностью дефолта и объясняющими переменными изменилась. Скорее всего, это было вызвано кризисом 2008–2009 гг. Важно отметить, что сдвиг проявляется вне зависимости от выбранного набора объясняющих переменных.

Теперь перейдем к анализу влияния отдельных переменных. Для начала рассмотрим предельные эффекты. Предельный эффект в логит-модели можно считать несколькими способами: найти предельный эффект для среднестатистического предприятия и посчитать средний предельный эффект по всем предприятиям. Мы выбрали второй вариант в силу ненормальности распределения объясняющих переменных.

Динамика предельных эффектов переменных в наилучшей логит-модели (все нефинансовые переменные и финансовые согласно популярности использования в других исследованиях) представлена на графике 3.

К сожалению, в силу неаддитивности и нелинейности зависимости вероятности банкротства предельные эффекты в логит-модели довольно нестабильны. Помимо абсолютной величины предельного эффекта сильно колеблется во времени и значимость коэффициентов и предельных эффектов.

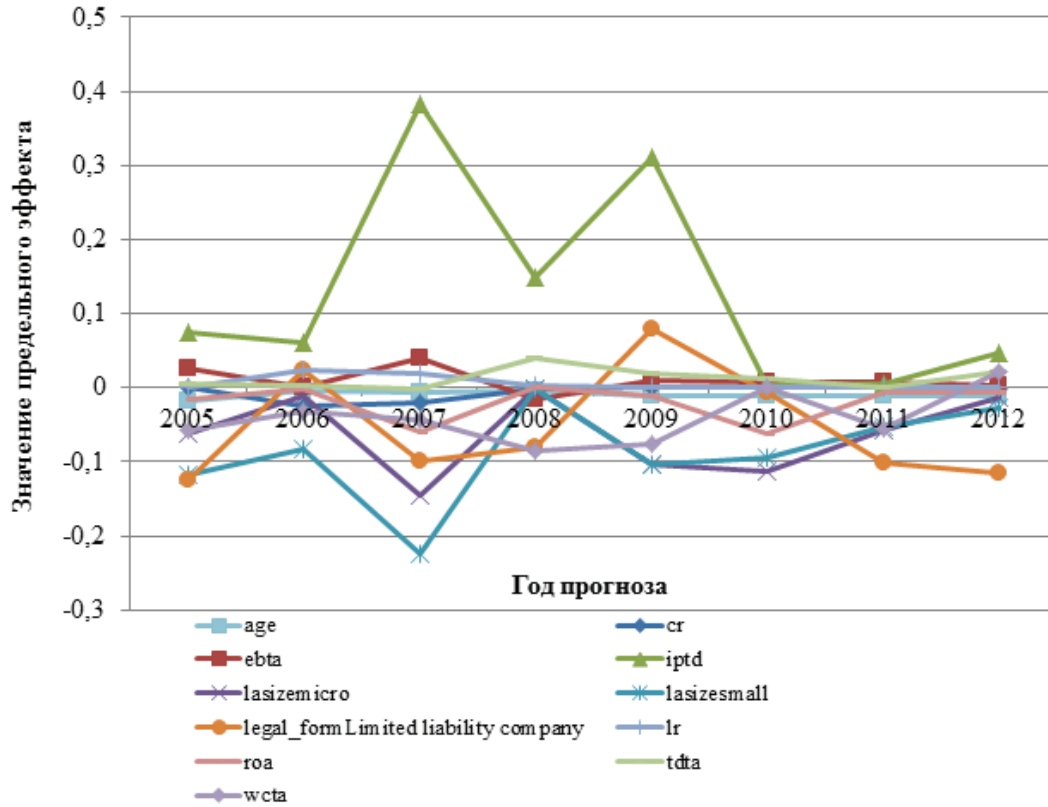


График 3. Предельные эффекты в логит-модели

Остановимся более подробно на переменной, которая была значима во все годы в данной спецификации, – на возрасте компании. На графике 4 изображено влияние предельного эффекта возраста на вероятность банкротства в зависимости от года:



График 4. Предельный эффект возраста в логит-модели

В полном соответствии с ожиданиями значение предельного эффекта оказывается всегда отрицательным, то есть с увеличением возраста компании падает вероятность банкротства. Это логично, потому что компании, функционирующие более длительное время, являются более устойчивыми, чем молодые компании, только начинающие свой бизнес.

Что касается абсолютной величины предельного эффекта, то можно сказать, что после 2008 г. значение эффекта стабилизировалось, и с увеличением возраста компании на один год вероятность дефолта падает примерно на один процентный пункт.

Теперь перейдем к визуализации результатов оценивания алгоритма случайного леса, так как он показал существенное превосходство над логит-моделью. В алгоритме случайного леса аналогом значимости коэффициента выступает его важность.

В данной работе важность измеряется как усредненное по всем деревьям суммарное падение индекса Джини в узлах, относящихся к конкретной переменной. Динамика важности переменных для наилучшей спецификации (популярные финансовые переменные плюс все нефинансовые) представлена на графике 5.

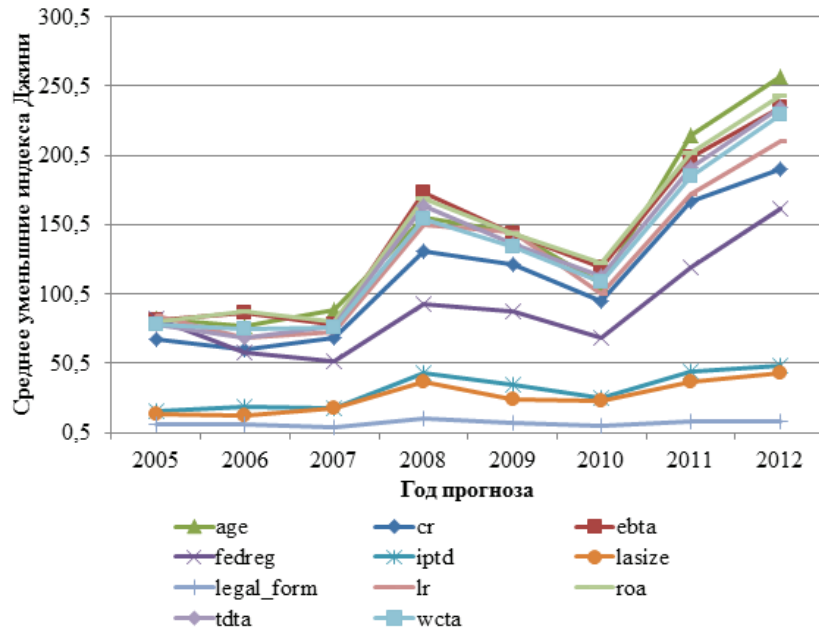


График 5. Важность переменных (алгоритм случайного леса)

Можно выделить две группы переменных: важные и не важные. Начнем с последней группы. В нее входят три переменные: организационная форма (legal\_form), последний доступный размер (lasize), iptd. Так как уменьшение индекса Джини очень мало (до 50,5), можно заключить, что эти переменные не оказывают значительного влияния на вероятность банкротства.

Теперь обратимся к группе важных переменных: среди них есть как финансовые, так и нефинансовые переменные. Возраст компании является одним из лидеров по уменьшению индекса Джини.

Мы снова видим, что нефинансовые характеристики должны быть включены в модели и что они улучшают качество прогнозов. В то же время другая нефинансовая характеристика, федеральный округ, замыкает группу лидеров. Однако важность этой переменной подтверждает, что включение в выборку компаний без учета территориального признака неверно.

Перейдем к важным финансовым показателям. Среди них есть показатели рентабельности (ebta, roa), финансового рычага (tdta), ликвидности (lr, cr, wcta). Проинтерпретируем некоторые из них.

С ростом roa (Net income / Total assets) вероятность стать компанией-банкротом падает, потому что увеличение этого отношения означает, что компания получает больший доход и может в случае необходимости пустить большую часть на выплату долга.

Когда tdt (Total debt / Total assets) увеличивается, с определенного момента компания может оказаться не в состоянии покрывать свой долг. То есть, хотя при увеличении этого соотношения растет финансовый рычаг, также увеличивается финансовый риск.

Показатель ликвидности  $cr$  ( $Current\ assets / Current\ liabilities$ ) характеризует, насколько компания в состоянии покрывать свои краткосрочные обязательства своими краткосрочными активами. Если отношение меньше единицы, то вероятность банкротства растет.

Визуализировать отдельно взятую модель, оцененную алгоритмом случайного леса, крайне сложно, так как случайный лес – это совокупность 500 классификационных деревьев и задача абсолютно точно изобразить случайный лес означала бы изображение всех 500 деревьев. Одно отдельно взятое дерево из этих 500 также плохо пригодно для изображения: во-первых, оно не является репрезентативным, во-вторых, оно строится до идеальной классификации каждого предприятия, то есть является слишком длинным.

Поэтому мы строим для алгоритма случайного леса одно «усредненное» дерево, которое при заданной сложности наилучшим образом делит предприятия на банкротов и активные. На графиках 6а и 6б нарисованы классификационные деревья для моделей со всеми нефинансовыми переменными по 2012 г.

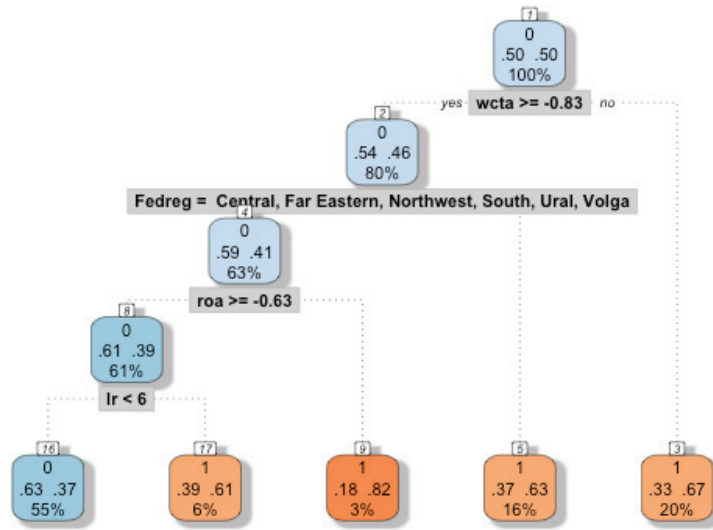


График 6а. Классификационное дерево по 2012 г. по модели со всеми нефинансовыми переменными и популярными финансовыми переменными

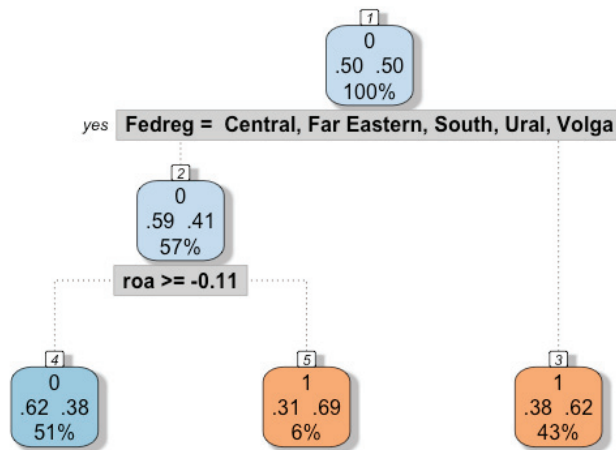


График 6б. Классификационное дерево по 2012 году по модели со всеми нефинансовыми переменными и финансовыми переменными по Альтману

На обоих деревьях видно, что регион играет существенную роль при прогнозировании банкротства. Вероятность банкротства существенно выше в Центральном, Дальневосточном, Южном, Уральском и Волжском федеральных округах. Также при обоих наборах объясняющих финансовых переменных важным является показатель рентабельности  $roa$ . Предприятия с большим значением показателя  $roa$  классифицируются как активные обоими моделями.

Отличие двух деревьев состоит в том, что на дереве, использующем популярные финансовые переменные, на две переменных больше оказались решающими при заданном уровне сложности дерева. Этими переменными стали  $lg$  и  $wcta$ , что еще раз подтверждает, что показатели ликвидности оказывают значительное влияние на вероятность дефолта.

## Заключение

Кратко подводя итоги эмпирического исследования, отметим выводы и перспективы дальнейших исследований.

Из проведенного анализа следует, что нелинейные алгоритмы прогнозирования работают существенно лучше обобщенных линейных. Наиболее известным нелинейным алгоритмом является использованный нами алгоритм случайного леса. Было бы также интересно сравнить прогнозную силу алгоритма случайного леса с другими популярными нелинейными алгоритмами, например с нейронными сетями или методом опорных векторов.

Нефинансовые переменные оказывают существенное влияние на вероятность банкротства, вне зависимости от включаемых в модель финансовых переменных, года построения модели и метода оценивания, и улучшают качество прогнозов.

По важности для прогнозирования переменные достаточно четко делятся на две группы. Не очень существенными оказались форма организации компании, размер и отношения процентных платежей к общей сумме долга. Остальные рассмотренные нами показатели более существенны и сопоставимы по важности.

Исходная модель Альтмана и Сабато уступает по прогнозной силе модели, включающей самые популярные переменные, выбираемые исследователями с 1930 г. по настоящее время. Следовательно, нужно с большой осторожностью подходить к отбору объясняющих переменных.

Среди финансовых переменных наибольшее влияние на вероятность дефолта оказывают показатели рентабельности, ликвидности и финансового рычага.

Обнаружен структурный сдвиг, соответствующий реакции отрасли оптовой и розничной торговли на кризис 2008–2009 гг.

В силу большого количества финансовых отношений необходимо использование методов, предназначенных для большого количества регрессоров. Здесь помимо использованной нами важности переменных можно использовать алгоритм LASSO или байесовские регрессии.

В силу неоднородности компаний по нефинансовым признакам следует ожидать, что отрасль оказывает существенное влияние на вероятность банкротства и структуру зависимости вероятности от объясняющих факторов. Поэтому имеет смысл строить отдельные модели вероятности банкротства для разных отраслей.

В базе данных RUSLANA имеется большое количество пропусков. Возможно следует подумать об агрегировании данных из разных источников. Также может иметь смысл воспользоваться алгоритмами восстановления данных, например, используя меру близости наблюдений алгоритма случайного леса.

## Список литературы

1. Гиленко Е.В., Довженко С.Е., Федорова Е.А. Модели прогнозирования банкротства: особенности российских предприятий. М.: ФГБОУ ВПО «Финансовый Университет при Правительстве Российской Федерации». 2012. С. 85–92.
2. Жданов В.Ю., Афанасьева О.А. Модель диагностики риска банкротства предприятий авиационно-промышленного комплекса // Корпоративные финансы. 2011. № 4. С. 77–89.
3. Макеева Е.Ю., Бакурова А.О. Прогнозирование банкротства компаний нефтегазового сектора с использованием нейросетей // Общественные науки и современность. 2006. № 6. С. 22–30.

4. Altman, E.I. (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The journal of finance*, 4 (23) (1968) 589–609.
5. Altman, E.I., Marco, G., Varetto, F. (1994), Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience), *Journal of Banking & Finance*, 3 (18) (1994) 505–529.
6. Altman, E.I., Sabato, G. (2007), Modelling credit risk for SMEs: Evidence from the US market, *Abacus*, 3 (43) (2007) 332–357.
7. Beaver, W.H. (1966), Financial ratios as predictors of failure, *Journal of accounting research*, 4 (1996) 71–111.
8. Bellovary, J.L., Giacomino, D.E., Akers, M.D. (2007), A review of bankruptcy prediction studies: 1930 to present, *Journal of Financial Education*, 33 (2007) 1–42.
9. Berger, A.N. (2006), Potential competitive effects of Basel II on banks in SME credit markets in the United States, *Journal of Financial Services Research*, 1 (29) (2006), 5–36.
10. Ciampi, F., Vallini, C., Gordini, N. (2009), Using Artificial Neural Networks Analysis for Small Enterprise Default Prediction Modeling: Statistical Evidence from Italian Firms, *Oxford Business & Economics Conference Proceedings, Association for Business and Economics Research (ABER)*, 1 (2009) 1–26.
11. Craig, B.R., Jackson, W.E., Thomson, J.B. (2007), Does government intervention in the small-firm credit market help economic performance?, *Federal Reserve Bank of Cleveland*.
12. Falkenstein, E., Boral, A., Carty, L. (2000), RiskCalc for private companies: Moody's default model, As published in *Global Credit Research*, May.
13. Härdle, W.K., et al. (2007), The default risk of firms examined with smooth support vector machines, *Discussion papers, German Institute for Economic Research*, 757 (2007) 1–30.
14. Hunter, J., Isachenkova, N. (2001), On the Determinants of Industrial Firm Failure in the UK and Russia in the 1990s, *ESRC Centre for Business Research, University of Cambridge*.
15. James, G., et al. (2013), *An introduction to statistical learning*, Springer.
16. Kaplinski, O. (2008), Usefulness and credibility of scoring methods in construction industry, *Journal of Civil Engineering and Management*, 1 (14) 2008 21–28.
17. Khorasgani, A. (2011), Optimal accounting based default prediction model for the UK SMEs, *Proceedings of ASBBS Annual Conference: Las Vegas*, 1 (18) (2011) 149–168.
18. Kolari, J.W., Ou, C., Shin, G.H. (2006), Assessing the profitability and riskiness of small business lenders in the banking industry, *Journal of Entrepreneurial Finance, JEF*, 2 (11) (2006) 1–26.
19. Lugovskaya, L. (2010), Predicting default of Russian SMEs on the basis of financial and non-financial variables, *Journal of Financial Services Marketing*, 4 (14) (2010) 301–313.
20. Makeeva, E., Neretina, E. (2013), The Prediction of Bankruptcy in a Construction Industry of Russian Federation, *Journal of Modern Accounting and Auditing*, 2 (9) (2013) 256–271.
21. Maleev, V., Nikolenko, T. (2010), Predicting Probability of Default of Russian Public Companies on the Basis of Financial and Market Variables. Preprint.
22. Martin, D. (1977), Early warning of bank failure: A logit regression approach, *Journal of Banking & Finance*, 3 (1) (1977) 249–276.
23. Ohlson, J.A. (1980), Financial ratios and the probabilistic prediction of bankruptcy, *Journal of accounting research*, 1 (18) (1980) 109–131.
24. Pompe, P.P., Bilderbeek, J. (2005), The prediction of bankruptcy of small-and medium-sized industrial firms, *Journal of Business Venturing*, 6 (20) (2005) 847– 868.
25. Sirirattanaphonkun, W., Pattarathammas, S. (2012), Default Prediction for Small- Medium Enterprises in Emerging Market: Evidence from Thailand, *Seoul Journal of Business*, 2 (18) (2012) 25–54.

26. Tam, K.Y., Kiang, M.Y. (1992), Managerial applications of neural networks: the case of bank failure predictions, *Management science*, 7 (38) (1992) 926–947.
27. Venables, W.N., Smith, D.M. (2002), *An introduction to R, Network Theory*.
28. Wei, L., Li, J., Chen, Z. (2007), Credit risk evaluation using support vector machine with mixture of kernel, *Lecture Notes in Computational Science and Engineering*, (2007) 431–438.
29. Wilson, R.L., Sharda, R. (1994), Bankruptcy prediction using neural networks, *Decision support systems*, 5 (11) (1994) 545–557.
30. Zeitun, R., Tian, G., Keen, K. (2007), Default probability for the Jordanian companies: A test of cash flow theory, *International Research Journal of Finance and Economics*, 8 (2007) 147–162.