

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра _____ Математической экономики _____

**Оценивание эконометрических моделей с ограниченными
зависимыми переменными**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента _____ 4 _____ курса _____ 441 _____ группы

направление _____ 09.03.03 — Прикладная информатика _____

_____ механико-математического факультета _____

_____ Мартынова Владимира Александровича _____

Научный руководитель
доцент, к.ф.-м.н. _____

_____ В.В. Новиков _____

Зав. кафедрой
д.ф.-м.н., профессор _____

_____ С.И. Дудов _____

Саратов 2019

Введение. Проблема оценки и управления рисками присутствует во всех секторах экономики, особенно актуальной она является для банковского сектора, играющего значительную роль в развитии страны и обеспечении экономического роста. Задача оценки банковских рисков особенно важна для стран с переходной экономикой, в связи с тем, что и банки и надзорные органы не имеют опыта функционирования в условиях рыночной экономики. Россия страна, неоднократно сталкивавшаяся с банковскими кризисами, необходимость стабильного развития банковской системы подтверждена кризисами 1998, 2004 и 2008 годов. На 01.01.2010 в России было зарегистрировано 1058 банков, такое количество банков не может быть регулярно инспектировано со стороны Центрального Банка или Агентства по Страхованию Вкладов. Данные факты говорят о необходимости использования систем дистанционного анализа состояния банков.

Дистанционный анализ, основанный на анализе банковской деятельности по ежеквартальным и годовым отчетам, позволяет установить проблемные банки, чье финансовое состояние является нестабильным. Дистанционные методы не дают однозначной характеристики надежности того или иного банка, но могут существенно сократить затраты надзорных органов и повысить эффективность их работы. Это связано с тем, что в первую очередь могут проверяться потенциально ненадежные банки, чтобы предупредить их несостоятельность, а затем остальные, что повышает стабильность функционирования банковской системы в целом. Дистанционные методы активно используются регуляторами США и европейских стран, опыт которых подтверждает эффективность и необходимость дистанционного мониторинга. Подобные методы могут применяться и самими банками в качестве системы внутренних рейтингов для оценки надежности контрагента, также дистанционные методы анализа могут использоваться крупными компаниями для мониторинга финансового состояния банка-партнера. Одним из элементов дистанционного анализа может быть автоматизированное прогнозирование дефолта того или иного банка на основе имеющейся финансовой отчетности средствами и методами, применяемыми в современных науках о данных. Этим определяется актуальность темы данной выпускной квалификационной работы. Целью работы является построение регрессионных logit-моделей, на основе которых

можно было бы вычислять вероятность дефолта банков в пределах некоторого горизонта прогнозирования. В работе ширина этого горизонта определена равной двум годам. В рамках достижения данной цели решались следующие задачи:

- сделать обзор методов построения регрессионных моделей с дискретными и ограниченными зависимыми переменными на основе учебной и монографической литературы;

- изучить возможности эконометрической программы Gretl работы с такими моделями;

- по материалам научной периодики ознакомиться со спецификой прогнозирования дефолта банков в условиях Российской Федерации;

- на основе реальной финансовой отчетности российских банков построить регрессионные logit-модели для предсказания возможности дефолта банков в будущем.

Работа состоит из введения, трех разделов, заключения и списка литературы.

В первом разделе "Модели бинарного и множественного выбора" приведены основные теоретические сведения относящиеся к таким разделам как оценка максимального правдоподобия параметров многомерного нормального распределения, оценка максимального правдоподобия, линейная модель вероятности, probit- и logit- модели, модели множественного выбора: номинальные зависимые переменные, модели множественного выбора: порядковые зависимые переменные.

Во втором разделе "Модели с урезанными и цензурированными выборками" приведены основные теоретические сведения относящиеся к таким разделам как урезанные выборки, цензурированные выборки, tobit- модель, модель Хекмана, модели "времени жизни".

В третьем разделе "Оценка вероятности дефолта российских банков" строятся эконометрические модели дефолта российских банков с использованием логит-регрессии. На основе информации Центрального банка Российской Федерации подготовлены данные по основным российским банкам (в выборке более пятисот названий) на основе которых были построены соответствующие logit-модели. Эти модели были протестированы как на самой обучающей выборке, так и вне ее, затем с их помощью была выполнена оценка вероятности дефолта группы российских банков в течение ближайших двух лет.

Probit- и logit-модели. Рассмотрим теоретические предпосылки применения моделей бинарного выбора, одна из которых (logit-модель) используется в практической части работы. Основным недостатком линейной модели вероятности есть следствие предположения о линейной зависимости вероятности $P(y_t = 1)$ от β . Его можно преодолеть, если считать, что

$$P(y_t = 1) = F(x_t' \beta), \quad (1)$$

где $F(\cdot)$ — некоторая функция, область значений которой лежит в отрезке $[0, 1]$. В частности, в качестве $F(\cdot)$ можно взять функцию распределения ней случайной величины. Одна из возможных интерпретаций модели (1) выглядит следующим образом. Предположим, что существует некоторая количественная переменная y_t^* , связанная с независимыми переменными x_t обычным

регрессионным управлением

$$y_t^* = x_t' \beta + \varepsilon_t, \quad (2)$$

где ошибки ε_t независимы и одинаково распределены с нулевым средним и дисперсией σ^2 . Пусть также $F(\cdot)$ — функция распределения нормированной случайной ошибки ε_t/σ . Величина y_t является ненаблюдаемой, а решение, соответствующее значению $y_t = 1$, принимается тогда, когда y_t^* превосходит некоторое пороговое значение. Так, в примере с покупкой автомобиля можно считать, что y_t^* представляет накопления семьи с номеров t . Без ограничения общности, если константа включена в число регрессоров, можно считать это пороговое значение равным нулю. Величину y_t^* можно также интерпретировать как разность полезностей альтернативы 1 и альтернативы 0. Таким образом,

$$\begin{aligned} y_t = 1, & \quad y_t^* \geq 0, \\ y_t = 0, & \quad y_t^* < 0. \end{aligned} \quad (3)$$

Тогда, предполагая, что случайные ошибки ε_t имеют одно и то же симметричное распределение $F(\cdot)$ (т.е. $F(-x) = 1 - F(x)$), получаем:

$$\begin{aligned} P(y_t = 1) &= P(y_t^* \geq 0) = P(x_t' \beta + \varepsilon_t \geq 0) \\ &= P(\varepsilon_t \geq -x_t' \beta) = P(\varepsilon_t \leq x_t' \beta) = F\left(\frac{x_t' \beta}{\sigma}\right), \end{aligned} \quad (4)$$

что с точностью до нормировки совпадает с (см.(1)). Наиболее часто в качестве функции $F = (\cdot)$ используют:

- функцию стандартного нормального распределения:

$$F(u) = \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{z^2}{2}} dz.$$

и соответствующую модель называют *probit*- моделью;

- функцию логистического распределения:

$$F(u) = \Lambda(u) = \frac{e^u}{1 + e^u}.$$

и соответствующую модель называют *logit*-моделью.

В свете рассмотренной выше интерпретации модели (1) использование функции нормального распределения представляется достаточно естественным. Применение функции логического распределения во многом объясняется простой численной реализации процедуры оценивания параметров. Вопрос о том, какую из моделей (*probit* или *logit*) следует использовать в том или ином случае, является достаточно сложным. Можно, например, выбрать ту модель, для которой больше значение соответствующей функции правдоподобия. Отметим также, что для значений u , достаточно близких по модулю к нулю (например при $u \in [-1.2, 1.2]$), функции $\Phi(u)$ и $\Lambda(u)$ ведут себя примерно одинаково, в то же время «хвосты» логистического распределения значительно «тяжелее» «хвостов» нормального распределения. Практический опыт показывают, что для выборок с небольшим разбросом объясняющих переменных и при отсутствии, общественного преобладания одной альтернативы над другой общественные выводы, получаемые с помощью *probit* - и *logit* - моделей будут, как правило, совпадать.

Оценка вероятности дефолта российских банков. Статистическим источником для создания набора данных, по которому строилась модель, является Банк России, с официального сайта которого производился сбор информации. На основе информации представленной в отчете о прибылях и убытках (форма № 102) строится прогноз того, какие банки потерпят дефолт в двухлетний период. В данной работе рассматривается одна пространственная выборка. Горизонт прогнозирования два года выбран из тех соображений, что двухлетний период покрывает среднее время между отзывом лицензии и ликвидацией банка. Основой для прогнозирования является модель бинарного выбора, а именно, *logit*-модель с зависимой переменной *Live*, которой присваивается значение 0, если банк является банкротом (терпит дефолт), и 1, в противном случае. При этом считаем, что банк является банкротом, если лицензия банка была отозвана в двухлетний период от даты отсчета.

Процедура оценивания и прогнозирования проводится по следующей схеме. На основе данных за 2016 год строится *logit*-модель, затем с ее помощью дается ретроспективный прогноз на этот же 2016 год; эти данные можно рассматривать как обучающую выборку. Вне обучающей выборки ретроспектив-

ный прогноз проводится на 2017 и 2018 годы. На основе анализа результатов такого прогнозирования делается заключение об адекватности модели. Затем строится новая модель, для которой в качестве обучающей выборки используются данные за 2018 год. Эта модель снова тестируется, но уже только на обучающей выборке, с помощью ретроспективного прогноза на 2018 год. Наконец, на основе этой модели делается прогноз дефолта российских банков на 2019–2020 годы.

Из выборки был исключен ряд банков: Сбербанк, Банк Москвы, Россельхозбанк, Газпромбанк, МСП БАНК, Национальный Клиринговый Центр, Российский Капитал, Глобэкс, т.к. эти банки заведомо не могут обанкротиться: их поддержит государство.

Все статьи доходов были взвешены по коду 19999 (всего доходов), а расходы по коду 29999 (всего расходов), чтобы сравнение банков было корректным. Состав переменных для модели подбирался так, чтобы значимость коэффициентов была максимальная.

На рисунке (1) приведены исходные показатели, которые были выбраны в соответствии с методикой статьи и среди которых проводился дальнейший отбор показателей, используемых в качестве объясняющих переменных регрессионной модели.

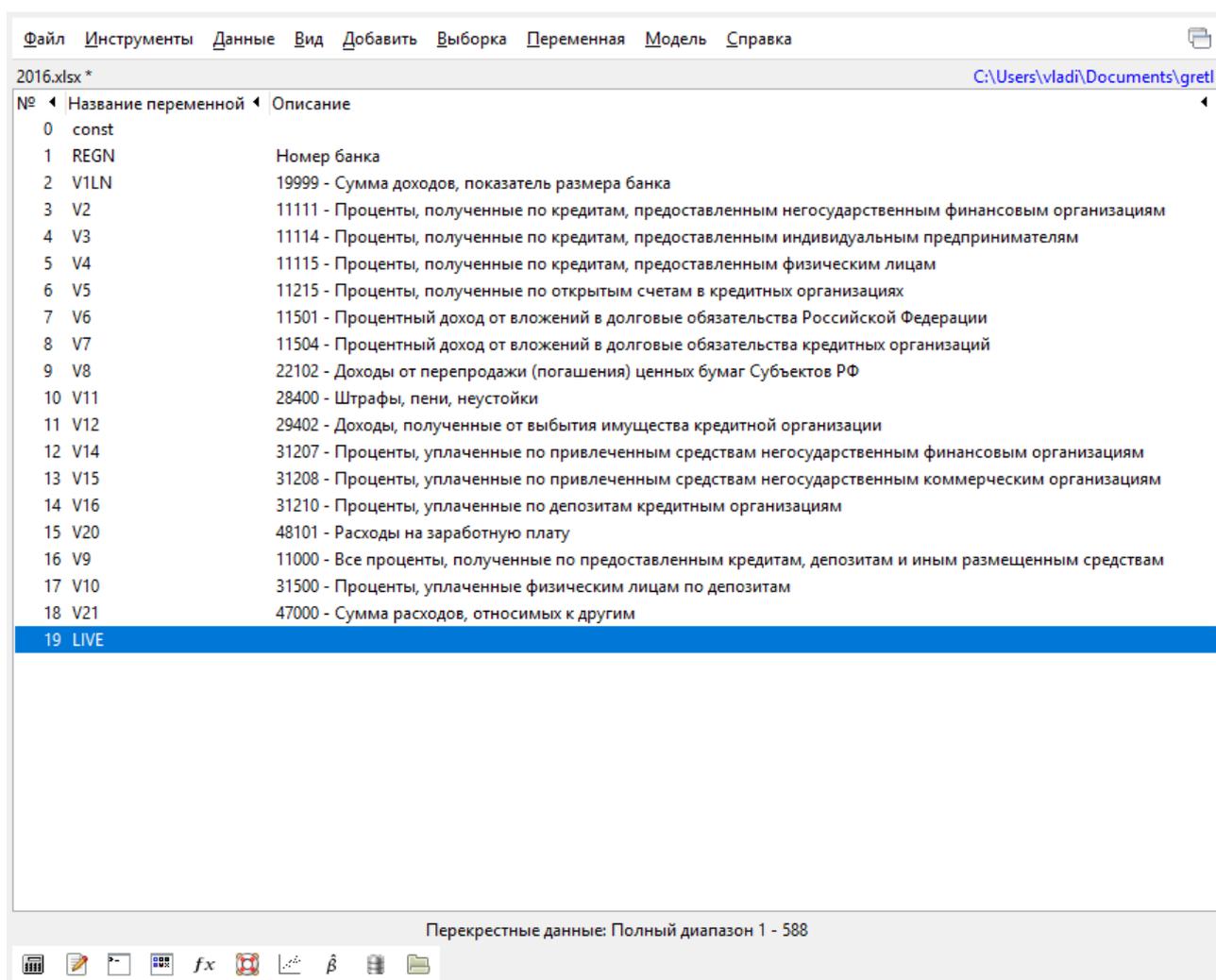


Рисунок 1 — Исходный набор переменных

Были отобраны показатели банков представленные в следующей таблице:

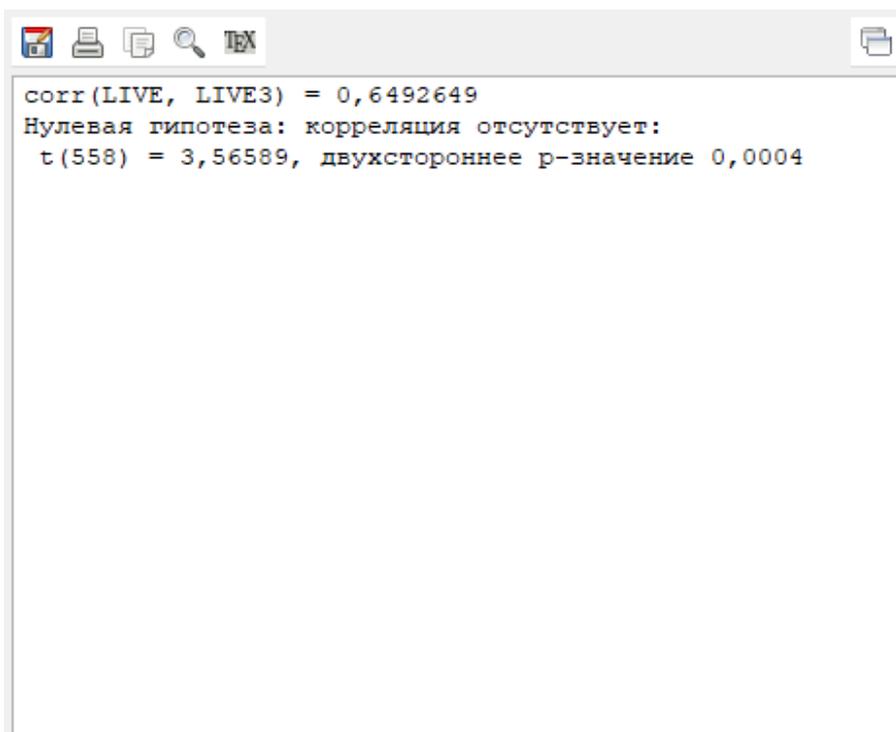
Переменная
$\ln(19999)$
11114/19999
11115/19999
22102/19999
31207/29999
48101/29999

Таблица 1 — Объясняющие переменные

На основе данных за 2018 год аналогичным образом строится и тестируется модель.

$$LIVE3 = 1 / (1 + 2.71828^x - (-1.530 + 0.2301 * V1LN + 2.25745 * V3 + 0.826 * V4 - 263.679 * V8 + 0.513 * V14 + 1.82 * V20)) \quad (5)$$

В данном случае получаем следующее значение коэффициента корреляции:



```
corr(LIVE, LIVE3) = 0,6492649
Нулевая гипотеза: корреляция отсутствует:
t(558) = 3,56589, двухстороннее p-значение 0,0004
```

Рисунок 2 — Коэффициент корреляции между переменными LIVE и LIVE3

Теперь, используя модель (1), выполним прогнозирование дефолта российских банков в течение 2019–2020 годов. В результате получена группа банков с малой вероятностью «выживания» в указанный период. В следующей таблице приведены банки из этой группы и вероятности того, что они не обанкротятся за указанные два года.

Код банка	Вероятность	Код банка	Вероятность	Код банка	Вероятность
3507	0,134216	1115	0,234208	3380	0,292450
596	0,175940	2846	0,248562	3300	0,300168
3328	0,180666	438	0,263141	2807	0,305192
654	0,199377	469	0,265077	236	0,310642
3139	0,204228	2388	0,265824	857	0,320358
2982	0,208672	77	0,269720	3533	0,325258
2048	0,208905	2756	0,272106	3223	0,331709
2605	0,221203	1478	0,276130	2524	0,333947
3470	0,223929	752	0,285036	3255	0,334329
3202	0,227271	2721	0,287973	3420	0,339881
3085	0,340899	969	0,353103	2534	0,359468
842	0,347088	3395	0,362637	2443	0,312174
1948	0,347399	2270	0,353851	2638	0,320230

Таблица 2 — Результат прогнозирования на 2019-2020 годы

3507	"СеверСтройБанк"АО
596	Банк "Уссури"(АО)
3328	ООО "Дойче Банк"
654	АО "Банк Воронеж"
3139	АКБ "Новый Кредитный Союз"(АО)
2982	АО "ГОРБАНК"
2048	ПАО "САРОВБИЗНЕСБАНК"
2605	ООО "РНКО "ВЕСТ"
3470	АО "Тойота Банк"
3202	АО НОКССБАНК
1115	ООО "Костромаселькомбанк"
2846	КБ "СИСТЕМА"ООО
438	ООО "Камкомбанк"
469	ООО "Русский Национальный Банк"
2388	АКБ "АлтайБизнес-Банк"(АО)
77	МОРСКОЙ БАНК (АО)
2756	АО АКБ "Тексбанк"
1478	ООО КБ "Лэнд-Банк"
752	ПАО КБ "МПСБ"
2721	"Северный Народный Банк"(ПАО)
3380	АО "Телекоммерц Банк"
3300	АО "КОШЕЛЕВ-БАНК"
2807	АКБ "Трансстройбанк"(АО)
236	ООО "СПЕЦСТРОЙБАНК"
857	КБ "Долинск"(АО)
3533	ООО "НКО "Глобал Эксчейндж"
3223	АО "Газнефтьбанк"
2524	КБ "МКБ"(ПАО)
3255	ПАО Банк ЗЕНИТ
3420	АО НКО "Частный РКЦ"
3085	АО АКБ "ЭКСПРЕСС-ВОЛГА"
842	ООО КБ "Кетовский"
1948	ООО МИБ "ДАЛЕНА"
969	ПАО "Евразийский банк"
3395	"Банк "МБА-МОСКВА"ООО
2270	ПАО "ОФК Банк"
2534	АО КБ "Пойдём!"
2443	ПАО "МЕТКОМБАНК"
2638	Банк ПТБ (ООО)

Таблица 3 — Таблица соответствия банка коду

Заключение. В рамках данной работы обоснована применимость эконометрических методов для оценки вероятности дефолта российских банков. Основное внимание было уделено практическому применению эконометрических моделей.

В работе получены следующие результаты.

1. Собрана база данных, позволяющая строить вышеуказанные модели;
2. Применена эконометрическая модель бинарного выбора для оценки вероятности дефолта российских банков;
3. Доказана эффективность построенных эконометрических моделей;
4. На основе полученной эконометрической модели вероятности дефолта российских банков дан прогноз де- фолта российских банков на 2019–2020 годы.

Несмотря на относительно низкое качество данных финансовых отчетов банков, эконометрические модели вероятности дефолта, основанные на публично доступной информации, имеют прогнозную силу.

Построенная эконометрическая модель обладает практической применимостью и при некоторой доработке может быть реализована в практике дистанционного анализа банковской деятельности.