

Санкт-Петербургский государственный университет  
Кафедра математической теории игр и статистических  
решений

**Никитина Анастасия Александровна**

**Магистерская диссертация**

**О методике статистического анализа  
банковских рисков**

Направление 01.04.02

Прикладная математика и информатика

Магистерская программа

Исследование операций и системный анализ

Научный руководитель,  
кандидат физ.-мат. наук,  
старший преподаватель  
Кумачева С. Ш.

Санкт-Петербург

2017

# Содержание

Введение . . . . .	3
Постановка задачи . . . . .	5
Обзор литературы . . . . .	6
Глава 1. Параметры кредитного портфеля . . . . .	8
1.1. Описание данных . . . . .	8
1.2. Обязательные нормативы . . . . .	11
1.3. Оценка параметров кредитного портфеля . . . . .	12
1.4. Оценка величины VaR . . . . .	17
Глава 2. Классификация коммерческих банков . . . . .	20
2.1. Бинарные регрессии . . . . .	20
2.2. Байесовский классификатор . . . . .	25
Глава 3. Использование <i>ARIMA</i> моделей . . . . .	30
3.1. Описание данных . . . . .	30
3.2. Выбор параметров модели <i>ARIMA</i> . . . . .	31
3.3. Исследование остатков модели . . . . .	38
3.4. Примеры построения прогноза с использованием моделей <i>ARIMA</i> . . . . .	40
Выводы . . . . .	44
Заключение . . . . .	45
Список литературы . . . . .	46
Приложение . . . . .	49

## Введение

Банковская система является важной неотъемлемой частью экономики любой страны, так как коммерческие банки имеют связи со всеми секторами экономики. Поэтому мировой финансовый кризис, дестабилизируя экономику в целом, затрагивает и денежно-кредитные институты. В связи с этим проблема обеспечения финансовой устойчивости российских коммерческих банков становится одной из наиболее актуальных задач [1].

В литературе встречается множество толкований понятия „финансовая устойчивость“ кредитной коммерческой организации, однако нет четко сформулированного определения. В настоящей работе под финансово устойчивым будем понимать такой коммерческий банк, который способен в дестабилизирующих условиях сохранить достаточный капитал, превышающий минимальное ограничение, установленное Банком России, а также оказывать банковские услуги и своевременно выполнять обязательства. То есть, это коммерческий банк, способный осуществлять свою деятельность в долгосрочной перспективе.

Традиционно оценка финансовой устойчивости коммерческого банка основывается на таких показателях, как уровень ликвидности, качество активов, достаточность собственного капитала для развития организации, риски, прибыльность деятельности и другие. Однако в условиях кризиса важно не только оценить финансовую устойчивость коммерческого банка, но и поддерживать стабильность функционирования организации. А это уже задача из области управления рисками.

В процессе экономической деятельности банки сталкиваются с разными видами рисков. Такой вид, как внешние риски, к которым относятся страновой, экономический, нормативно-правовой риск, не поддаются управлению организацией. Однако, финансовые риски, включающие в себя

процентный риск, валютный риск, кредитный риск, внебалансовый риск, риск использования заемного капитала, а также риск ликвидности, кроме того, что в наибольшей степени поддаются контролю со стороны банка, также являются ключевыми факторами при управлении активами и пассивами.

Так как для многих коммерческих банков кредитование является одним из основных видов деятельности, доходы от кредитных операций составляют существенную часть в общей структуре доходов коммерческих организаций. Следовательно, эффективность кредитной политики зависит от качества управления организацией кредитным портфелем.

Кредитный портфель представляет собой остатки задолженности на балансовых счетах по действующим кредитным договорам на определенную дату. В данной работе в кредитный портфель включены кредиты, выданные банкам, а также юридическим и физическим лицам.

Неграмотное управление кредитным портфелем может привести к большим финансовым потерям, а, значит, нарушить стабильность функционирования кредитной организации, что чаще всего приводит к ее дефолту. Поэтому качество кредитного портфеля становится одним из основополагающих факторов для выживания и успеха коммерческого банка.

## Постановка задачи

Для эффективного управления кредитным портфелем необходим его анализ по различным характеристикам. Целью данной работы является разработка аналитической модели, позволяющей сформировать „систему раннего оповещения“ об ухудшении качества кредитного портфеля, которое в дальнейшем может привести к отзыву лицензии коммерческого банка.

В классической экономической литературе [2, 3, 4, 5] портфель ценных бумаг описывается такими параметрами, как ожидаемая доходность и риск. Также важными показателями банковской деятельности являются обязательные нормативы Банка России [6]. Все коммерческие банки обязаны предоставлять информацию о соблюдении обязательных нормативов, поэтому по их выполнению легче всего следить за деятельностью коммерческих банков. Описание обязательных нормативов ЦБ, используемых в работе, а также вычисление параметров кредитного портфеля представлено в **Главе 1**.

Помимо анализа параметров, необходимо исследовать надежность функционирования кредитной организации на ближайший год. Для этого, с использованием параметров кредитного портфеля, в **Главе 2** построены модели бинарной регрессии, позволяющие оценить вероятность дефолта коммерческого банка. Также в этой главе описано построение классификатора Байеса, который по вычисленным характеристикам определяет рассматриваемый банк в группу надежных или не очень организаций.

В **Главе 3** проводится построение моделей *ARIMA*, которые позволяют прогнозировать значения параметров портфеля на ближайший год.

## Обзор литературы

Перед проведением данного исследования необходимо было в первую очередь ознакомиться с основными понятиями и методиками финансового риск-менеджмента. Поэтому одними из первых были прочитаны книги [3, 7], собравшие в себе подробную информацию о различных подходах к оценке риска и методах управления ими. Эти книги помогли составить представление о данной области знаний и основном предмете исследования — кредитных рисках.

Параллельно изучались российские методики оценки деятельности коммерческих банков. В первую очередь осуществлялся поиск информации об обязательных нормативах Банка России, основным источником которой стало официальное издание Центрального Банка [6]. В „Вестнике Банка России “ наиболее полно описана структура и суть обязательных нормативов.

Авторы многих статей посвященных финансовому анализу [8] в своих работах ссылаются на методику Кромонава, поэтому она также была изучена. Однако в данном исследовании было решено отказаться от ее использования, поскольку не представляется понятным, как были получены веса для сводного коэффициента.

После прочтения статьи [9], в которой авторы проводят сравнительный анализ обязательных нормативов и показателей из методики Кромонава, стало понятно, что обязательные нормативы и коэффициенты Кромонава имеют большую ошибку оценки надежности коммерческих банков, поэтому было необходимо разработать новые способы оценки финансовой устойчивости.

В поисках подхода к анализу кредитного портфеля с целью ознакомления с классической „портфельной теорией“ была прочитана статья [4].

Оттуда была почерпнута идея оценивать риск портфеля ценных бумаг как дисперсию, хотя основная задача работы заключалась в выборе стратегии формирования портфеля ценных бумаг.

Затем изучалась учебная литература [2, 5, 10], в которой подробно на примерах описано как вычислять доходность и риск портфеля, а также показатель Value-at-Risk, рекомендованный к применению Базельским комитетом по банковскому надзору [11].

Идея применения модели бинарной регрессии к вычисленным показателям появилась после прочтения статьи [12], в которой автор, используя логит-регрессию, анализирует характеристики клиентов банка. Теория по логит и пробит-моделям представлена в книге [13]. Результат применения бинарных регрессий необходимо было сравнить с каким-либо другим способом классификации. Материал о принципах и идеях различных классификаторов изучался по классической книге по эконометрике [14].

Теоретические основы моделирования временных рядов с помощью авторегрессии и модели скользящего среднего, которые применяются в главе 3, изложены в учебном пособии [15]. Также стоит отметить книгу [16] о моделировании временных рядов, в которой на большом количестве примеров и графиков иллюстрируются тонкости практического применения моделей *ARIMA*.

# Глава 1. Параметры кредитного портфеля

## 1.1. Описание данных

Каждый год в Российской Федерации около 100 коммерческих банков лишаются лицензии. Наиболее частыми причинами отзыва лицензии становятся неспособность исполнения обязательств перед кредиторами, несоблюдение требований законодательства в области противодействия легализации доходов, полученных преступным путем, размещение денежных средств в низкокачественные активы, проведение высокорискованной кредитной политики и другие. Так как в данной работе проводится анализ кредитного портфеля, то для исследования были отобраны банки, причиной отзыва лицензии которых стала последняя из перечисленных выше. Выбранные коммерческие банки были поделены на две группы: закрытые в период с ноября 2015 года по февраль 2016 года (17 банков) и те, у которых лицензия была отозвана в период с ноября 2016 года по февраль 2017 года (12 банков). Подобное разделение было проведено для того, чтобы данные банков из первой группы использовать в качестве обучающих при построении классификаторов в Главе 2, а данные по второй группе использовать для контроля и тестирования. Расчеты, описанные в Главах 1 и 3 проводились для всех банков одинаково.

Информация о причинах закрытия банков была взята из раздела „Книга памяти“ сайта Банки.ру [17].

Ниже представлен перечень банков, лицензия которых была отозвана в период с ноября 2015 года по февраль 2016 года:

1. АО «Смартбанк»;
2. ООО «Коммерческий банк „Богородский“»;
3. АО «СтарБанк»;



4. ПАО «Агро-промышленный банк Екатеринбургский»;
5. ООО «Коммерческий банк „Регионально-отраслевой Специализированный Автопромышленный банк“»;
6. ОАО «Акционерно-коммерческий банк „Акция“»;
7. ООО «Коммерческий банк „Милбанк“»;
8. ООО «Коммерческий банк „Расчетный Дом“»;
9. ОАО «Ростовский-на-Дону акционерный коммерческий банк „Капиталбанк“»;
10. ООО «Коммерческий банк „Регнум“»;
11. ООО «Коммерческий Межрегиональный Трастовый Банк»;
12. ООО «Коммерческий банк „Эргобанк“»;
13. ООО «Дил-Банк»;
14. ООО «Коммерческий банк „Ренессанс“»;
15. ОАО «Коммерческий Банк „Максимум“»;
16. АО «Акционерный коммерческий банк „Носта“»;
17. ООО «КБ „Межрегионбанк“»;

Список банков, закрытых в период с конца 2016 года по начало 2017 года, отчеты которых используются в работе:

1. АО «Торговый Городской Банк»;
2. ПАО «Акционерный коммерческий банк „Енисей“»;
3. ПАО «Акционерный коммерческий банк „Новация“»;

4. АО «Акционерный коммерческий банк „Банк на Красных Воротах“»;
5. АО «Вологдабанк»;
6. ООО «Коммерческий банк „Финансовый капитал“»;
7. ООО «Коммерческий банк „Развитие“»;
8. ООО «Центральный коммерческий банк»;
9. ООО «Коммерческий банк „Российский промышленный банк“»;
10. ПАО «Выборг-банк»;
11. АО «Акционерный коммерческий банк „Русский Трастовый Банк“»;
12. АО «Акционерный коммерческий банк „Газстройбанк“».

Основным источником данных по вышеперечисленным банкам стал сайт Куап.ру [18], так как отчеты, представленные там, структурированы более удобным для обработки способом по сравнению с сайтом Центрального банка. Для анализа кредитного портфеля с сайта были выгружены агрегированный баланс, отчет о прибылях и убытках за 2008 — 2016 гг и Форма 135. Последняя содержит информацию по обязательным нормативам. В ячейках агрегированного баланса, так же, как и в оборотно-сальдовой ведомости (Форма 101), записаны остатки по счетам. От Формы 101 он отличается тем, что в нем перегруппированы статьи бухгалтерского баланса, объединяя статьи с одинаковым экономическим содержанием. При решении задачи оценивания кредитного портфеля учитываются выданные на них суммы. В отчете о прибылях и убытках используются только данные о процентных доходах по кредитам банкам, юридическим и физическим лицам.

## 1.2. Обязательные нормативы

В соответствии с политикой, проводимой Банком России [6], в целях регулирования принимаемых банками рисков были приняты следующие нормативы:

- Н1 — норматив достаточности собственных средств (капитала) банка;
- Н2 — норматив мгновенной ликвидности банка;
- Н3 — норматив текущей ликвидности банка;
- Н4 — норматив долгосрочной ликвидности банка;
- Н6 — норматив максимального размера риска на одного заемщика или группу связанных заемщиков;
- Н7 — норматив максимального размера крупных кредитных рисков;
- Н9.1 — норматив максимального размера кредитов, банковских гарантий и поручительств, предоставленных банком своим участникам;
- Н10.1 — норматив совокупной величины риска по инсайдерам банка;
- Н12 — норматив использования собственных средств (капитала) банка для приобретения акций (долей) других юридических лиц.

В таблице (1), представленной ниже, кратко описаны нормативы, наиболее важные с точки зрения анализа данных, необходимого для решения задач текущей работы.

Норматив достаточности собственных средств (Н1) регулирует риск невыполнения банком своих обязательств, а также определяет минимальный размер резервов для покрытия операционного, кредитного и рыночного рисков. Минимально допустимое его значение — 10%.

Норматив мгновенной ликвидности (Н2) регулирует риск банка потерять способность исполнять свои обязательства в течение одного дня.

Таблица 1: Нормативы ЦБ РФ

Показатель	Определение
<b>Н1</b> – Достаточность капитала	$= \frac{\text{Капитал}}{\text{Активы, взвешенные с учетом риска}}$
<b>Н2</b> – Мгновенная ликвидность	$= \frac{\text{Ликвидные активы}}{\text{Обязательства до востребования}}$
<b>Н3</b> – Текущая ликвидность	$= \frac{\text{Ликвидные активы}}{\text{Обязательства до востребования}}$
<b>Н4</b> – Долгосрочная ликвидность	$= \frac{\text{Сумма выданных кредитов}}{\text{Суммарные обязательства}}$

Аналогичен ему норматив текущей ликвидности (Н3), который отличается только тем, что ограничивает риск потери платежеспособности в течение ближайших 30 дней. Минимальные значения для этих нормативов, соответственно, 15% и 50%.

Риск потери ликвидности по причине вложения средств в долгосрочные активы Банк России регулирует с помощью норматива долгосрочной ликвидности (Н4). Максимальное значения для этого норматива — 120%.

Описанные выше 4 норматива будут использованы в Главе 2 при построении правил классификации банков по надежности их функционирования. Однако, строить классификатор только на основе обязательных нормативов некорректно, так как, как правило, банки ведут бухгалтерский учет так, чтобы удовлетворять ограничениям на нормативы [9]. И заметить, что дела у банка складываются неблагополучно, можно только в последний момент.

### 1.3. Оценка параметров кредитного портфеля

Так как в работе стоит задача создания „системы раннего оповещения“ об ухудшении качества кредитного портфеля, было принято решение ввести в рассмотрение такие параметры портфеля, как ожидаемая доходность и риск [2, 4]. Их вычисление основывается на доходности активов,

входящих в портфель. Как уже говорилось ранее, в экономической литературе нет единого стандарта, какие активы следует включать в кредитный портфель, поэтому, исходя из найденных отчетов, будем включать в портфель кредиты, выданные банкам, а также юридическим и физическим лицам.

В учебной экономической литературе [19] оценка доходности актива находится по следующей формуле

$$r_i = \frac{ПП}{KB} \times 100\%,$$

где  $ПП$  — это процентный доход по соответствующему виду кредита за рассматриваемый период, а  $KB$  — средний остаток ссудной задолженности за период. Так как в отчете о прибылях и убытках данные представлены за квартал, то и доходность актива будем вычислять поквартально. Таким образом, была оценена доходность кредитов, выданных банкам, юридическим и физическим лицам.

Ниже, в качестве примера, приведены графики (см. рис. 1, 2) изменения доходности по активам для «Коммерческого банка „Ренессанс“» и «Акционерного коммерческого банка „Енисей“», у которых была отозвана лицензия в конце 2015 года.

Основываясь на оценках доходности активов, можно вычислить ожидаемую доходность кредитного портфеля [2] для каждого квартала

$$E(r_p) = \sum_{i=1}^n r_i \theta_i,$$

где  $\theta_i$  — удельный вес актива в портфеле. Он рассчитывается как отношение стоимости актива к стоимости портфеля.

Изменение значений ожидаемой доходности кредитного портфеля представлены также на примерах «Коммерческого банка „Ренессанс“» и «Акционерного коммерческого банка „Енисей“» (см. рис. 3, 4).

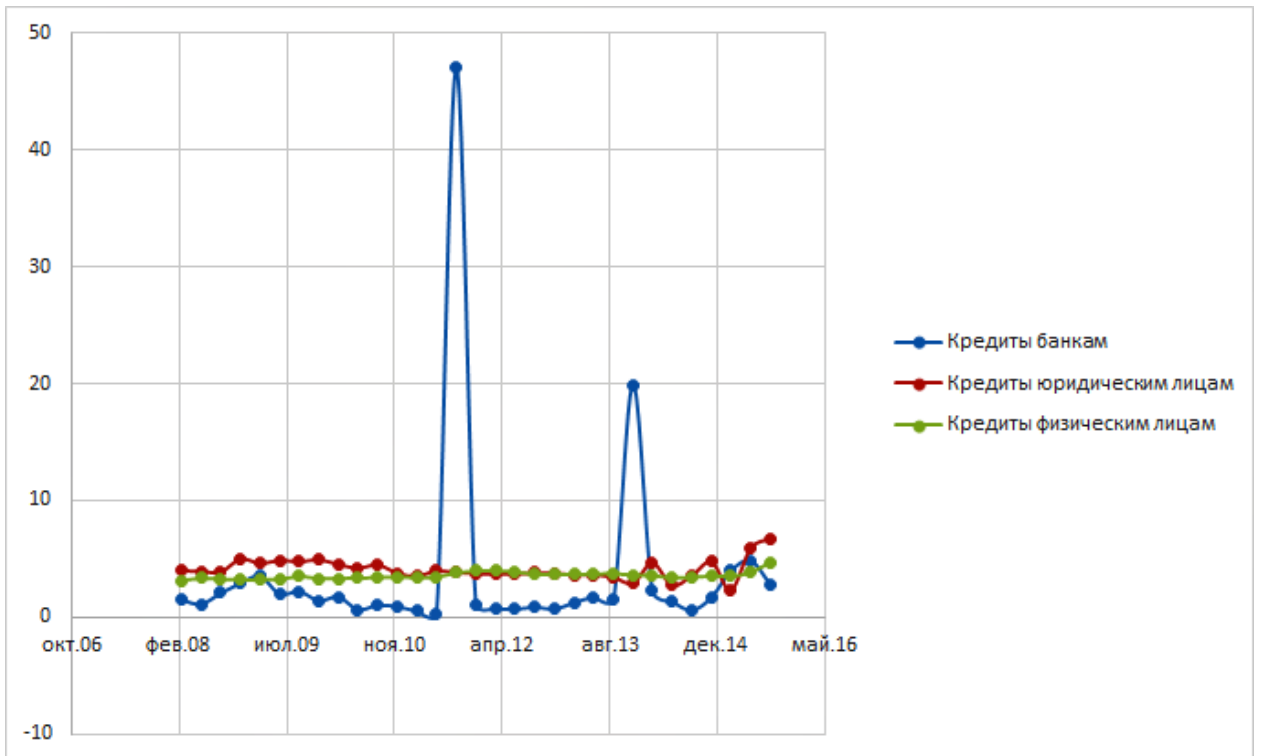


Рис. 1: Доходность по активам «КБ „Ренессанс“»

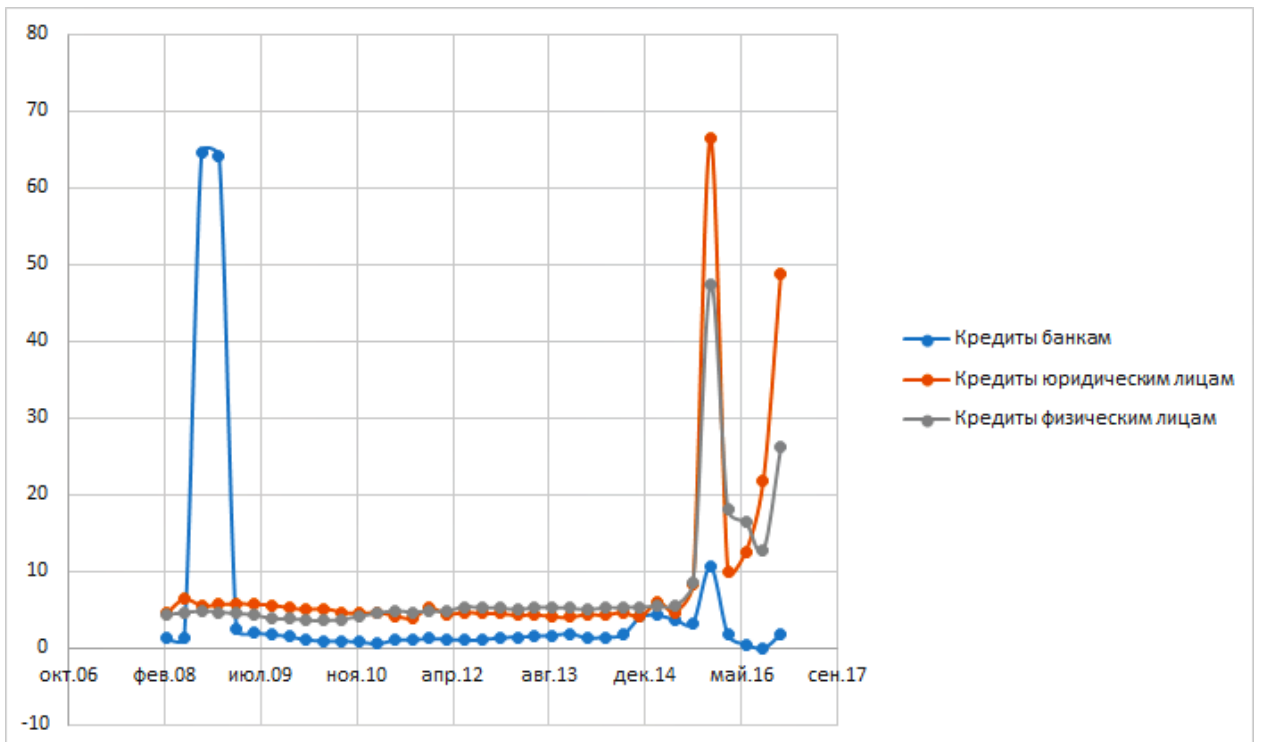


Рис. 2: Доходность по активам «АКБ „Енисей“»

При вычислении риска портфеля используются оценки рисков активов, формирующих портфель. Риск отдельно взятого актива вычисляется

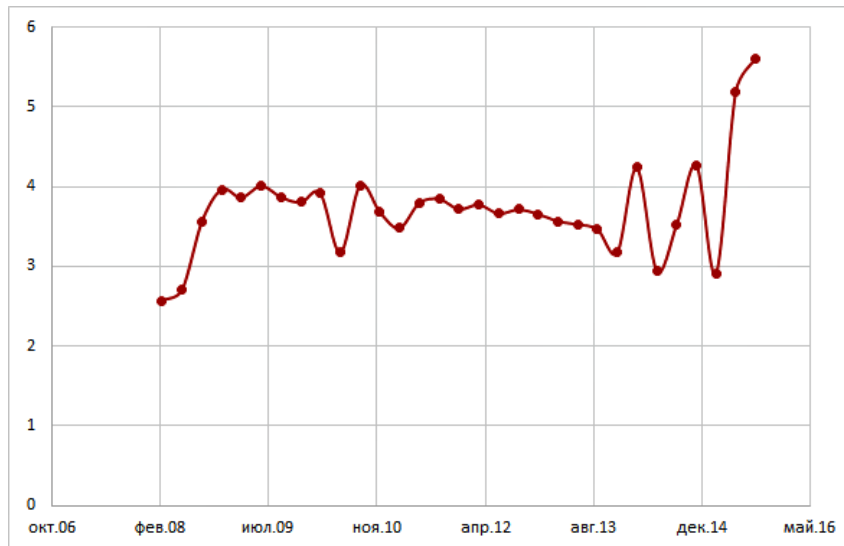


Рис. 3: Доходность кредитного портфеля «КБ „Ренессанс“»

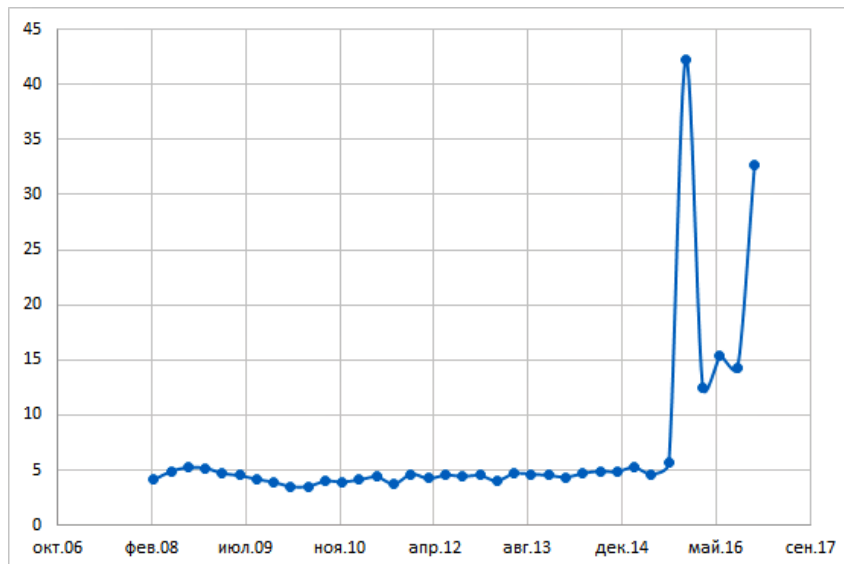


Рис. 4: Доходность кредитного портфеля «АКБ „Енисей“»

как стандартное отклонение доходности актива за определенный период

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{t=1}^k (r_t - \bar{r}_t)^2}$$

В данной работе значение риска актива вычислялось за год.

Зная коэффициенты ковариации между доходностью активов, можно оценить риск всего портфеля [2]:

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n \theta_i \theta_j Cov_{r_i, r_j}, \quad (1)$$

где  $Cov_{r_i, r_j}$  — коэффициент ковариации между доходностью по активам  $i$  и  $j$ , а  $\theta_i, \theta_j$  — как и ранее, доли активов  $i$  и  $j$  в портфеле.

Рассматриваемый портфель состоит из трех активов: кредиты банкам, юридическим и физическим лицам — которые обозначим соответственно  $a, b$  и  $c$ . Поэтому (1) примет вид

$$\sigma_p^2 = \theta_a^2 \sigma_a^2 + \theta_b^2 \sigma_b^2 + \theta_c^2 \sigma_c^2 + 2\theta_a \theta_b Cov_{r_a, r_b} + 2\theta_a \theta_c Cov_{r_a, r_c} + 2\theta_b \theta_c Cov_{r_b, r_c}.$$

Динамика изменения риска кредитного портфеля рассматриваемых банков представлена ниже (см. рис. 5, 6). Между этими и графиками (3, 4) можно проследить связь. С 2013 года ряд ожидаемой доходности кредитного портфеля «Коммерческого банка „Ренессанс“» начинает сильно колебаться, при этом риск с 2013 года растет. Это говорит о возможном ухудшении качества кредитного портфеля. Подобным образом ведут себя и графики ожидаемой доходности и риска «Акционерного коммерческого банка „Енисей“».

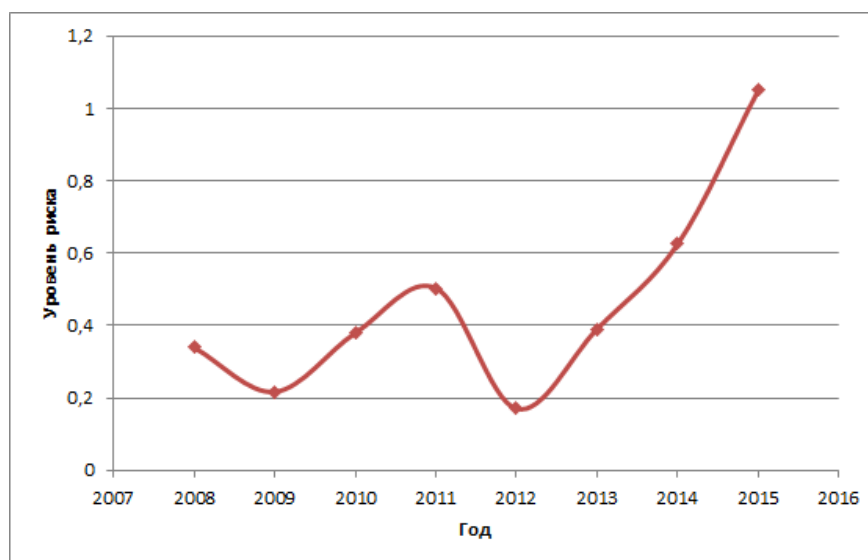


Рис. 5: Риск кредитного портфеля «КБ „Ренессанс“»



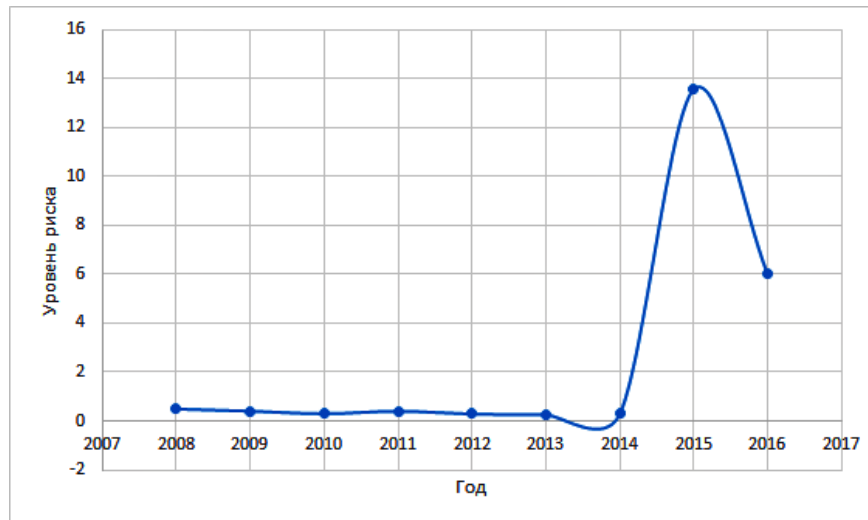


Рис. 6: Риск кредитного портфеля «АКБ „Енисей“»

#### 1.4. Оценка величины VaR

Помимо стандартного отклонения, в финансовом анализе широко используют такой показатель риска, как VaR (Value-at-Risk). «Это выраженная в денежных единицах оценка величины, которую не превысят ожидаемые в течение данного периода времени потери с заданной вероятностью» [3].

По определению из [3] VaR оценивается следующим образом:

$$P(VaR \geq x) = 1 - \alpha,$$

где  $\alpha$  — уровень значимости,  $x$  — величина возможных потерь.

Если предполагать, что стоимость портфеля подчиняется нормальному распределению, то показатель VaR несложно вычислить по следующей формуле [5]

$$VaR = P_t \sigma_p u_\alpha,$$

где  $P_t$  — это объем выданных кредитов на последний отчетный период, а  $u_\alpha$  — квантиль стандартного нормального распределения.

Однако на практике чаще всего закон распределения исследуемой величины неизвестен. Ниже приведена гистограмма (см. рис. 7) объема

выданных кредитов «Коммерческого банка „Ренессанс“» за период с 2008 по 2014 года. Уже по ней видно, что данные не подчиняются нормальному распределению.

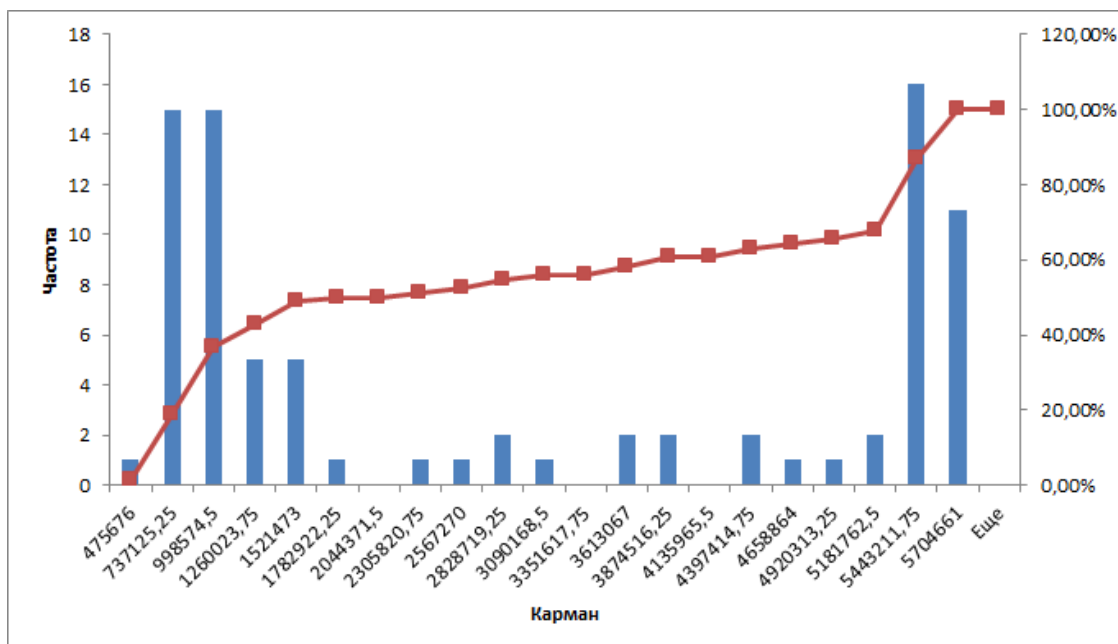


Рис. 7: Гистограмма объема выданных кредитов «КБ „Ренессанс“»

Чтобы убедиться в этом интуитивном заключении, проверим гипотезу о нормальности распределения данных с помощью критериев Шапиро–Уилка [20] и Колмогорова–Смирнова [21], используя пакет R.

В работе был использован Критерий Колмогорова–Смирнова, потому что он является „классическим“ критерием согласия. Однако, на небольших выборках, как в нашем случае, он обладает не очень большой мощностью,  $1 - \beta \approx 0,062$  [22]. Критерий Шапиро–Уилка для небольших выборок обладает более высокой мощностью по сравнению с непараметрическими критериями согласия.

Критерий Шапиро–Уилка показал, что гипотеза о нормальности распределения ряда принимается с вероятностью  $p = 9,64 \cdot 10^{-10}$ . Критерий Колмогорова–Смирнова принимает эту гипотезу с вероятностью  $p = 1,99 \cdot 10^{-4}$ . То есть, приняли альтернативную гипотезу о том, что

ряд не является нормально распределенным.

Так как мы опровергли гипотезу о нормальности распределения ряда ожидаемой доходности, оценим VaR «ручным способом». С помощью пакета R был найден квантиль порядка 0,05 ряда стоимости кредитного портфеля. Он составил 475 676 тыс. рублей. То есть с вероятностью 5% стоимость портфеля через год будет меньше, чем 475 676 тыс. рублей.

В действительности стоимость кредитного портфеля «Коммерческого банка „Ренессанс“» на конец 2015 года составила 7 296 244 тыс. рублей, то есть прогноз VaR подтвердился. Однако, несмотря на такой кредитный портфель, у банка все равно была отозвана лицензия. Это говорит о том, что банк, стараясь выдать как можно больше кредитов, не заботился об их качестве .

## Глава 2. Классификация коммерческих банков

Описанные в Главе 1 параметры могут использоваться не только при анализе кредитного портфеля, но и для прогнозирования вероятности дефолта кредитной организации. Для этого будет использоваться методология бинарной регрессии, а также классификатор Байеса.

### 2.1. Бинарные регрессии

В работах, посвященных управлению рисками [12, 23], авторы часто используют логит и пробит-регрессии при прогнозировании дефолта банка. Особенность этих моделей в том, что зависимая переменная принимает значения 0 или 1. В нашем случае 0 эквивалентен отзыву лицензии у банка, а 1 означает, что банк продолжает функционировать.

Общий вид моделей бинарной регрессии следующий:

$$P\{Y = 1\} = F(z),$$

$$z = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n,$$

где  $\beta_i$  — коэффициенты регрессии,  $x_i$  — факторы модели,  $F(\cdot)$  — некоторая функция распределения.

Логит и пробит-модели отличаются тем, какой закон распределения используются в качестве  $F(\cdot)$ :

- если функцию стандартного нормального распределения

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx,$$

то модель называют пробит-регрессией;

- если же функцию логистического распределения

$$F(z) = \frac{1}{1 + e^{-z}},$$

то модель называют логит-регрессией.

К достоинствам моделей бинарной регрессии можно отнести то, что они не требуют следующих предположений:

- факторы должны подчиняться нормальному закону распределения;
- гомоскедастичность остатков;
- остатки распределены нормально.

Строить бинарные регрессии было решено на основе 17-ти банков, лицензии которых были отозваны в конце 2015 года. Зависимая переменная, соответствующая этим банкам, принимает значение 0. Но для оценки коэффициентов регрессии нужны также банки из другого класса, то есть, стабильно функционирующие. Поэтому в анализ были добавлены еще 6 действующих банков:

1. ПАО «Уральский Банк Реконструкции и Развития»;
2. ПАО «Челябинвестбанк»;
3. ПАО «Совкомбанк»;
4. АО «Тинькофф Банк»;
5. АО «Райффайзенбанк»;
6. ПАО «Банк Уралсиб».

Банки, закрытые в конце 2016 – начале 2017 года, будут использоваться для проверки адекватности построенных моделей.

Так как планируется построение долгосрочного прогноза, то будем использовать значения характеристик, вычисленных на момент за год до отзыва лицензии. Следовательно, для банков, закрытых в конце 2015 года, будем использовать показатели за декабрь 2014 года.

В качестве регрессоров для логит и пробит-моделей были взяты следующие величины:

- уровень риска портфеля (*Risk*) на конец 2014 года;
- норматив достаточности капитала (*H1*) на конец 2014 года;
- норматив текущей ликвидности (*H3*) на конец 2014 года;
- норматив долгосрочной ликвидности (*H4*) на конец 2014 года.

В перечень факторов не был включен показатель ожидаемой доходности кредитного портфеля, потому что в большинстве случаев высокая ожидаемая доходность не являлась гарантом надежности банка. Норматив мгновенной ликвидности (*H2*) также не рассматривается, поскольку данная модель строится для долгосрочного прогноза.

### Результаты логит-регрессии

С помощью пакета R были оценены коэффициенты логит-модели:

$$z = 11,117 - 2,455 \cdot Risk - 0,26 \cdot H1 - 0,052 \cdot H3 - 0,043 \cdot H4.$$

Если при подстановке значений параметров в модель функция распределения  $P = F(z) > 0,5$ , то банк относим к классу стабильно функционирующих. В противном случае кредитная организация признается ненадежной.

Качество построенной модели оценим с помощью коэффициента правдоподобия (Likelihood ratio test). Этот критерий показывает, значимо ли модель, построенная по выбранным факторам, лучше, чем нулевая модель. Статистика критерия вычисляется как разность между отклонениями остатков нулевой и построенной модели. Данный критерий показал (см. Приложение 1), что рассматриваемая модель описывает данные лучше, чем модель без предикторов, с вероятностью  $p = 0,933$ .

Оценим эффективность модели. В таблице (2) представлены результаты логит-модели, примененной к обучающей выборке. К классу *C1* относятся банки, лицензия которых была отозвана, а к классу *C2* — действующие банки.

Таблица 2: Результат логит-регрессии на обучающей выборке

Фактически	Прогнозные		
	<i>C1</i>	<i>C2</i>	Общее число
<i>C1</i>	15	2	17
<i>C2</i>	2	4	6
Общее число	17	6	23

Доля правильно классифицированных банков составляет:

$$\frac{15 + 4}{23} \times 100\% = 82,6\%.$$

Применим теперь модель к контрольной выборке. В нее входят 12 банков, закрытых в конце 2016 года, и те же 6 действующих банков. В модель подставляются значения факторов, вычисленные на декабрь 2015 года. Результаты представлены в таблице (3).

Таблица 3: Результат логит-регрессии на контрольной выборке

Фактически	Прогнозные		
	<i>C1</i>	<i>C2</i>	Общее число
<i>C1</i>	12	0	12
<i>C2</i>	5	1	6
Общее число	17	1	23

По показателям за декабрь 2015 года 5 из 6 стабильных банков были отнесены в категорию ненадежных, что увеличило ошибку классификации. Доля правильно классифицированных банков составила 72,2%.

В таблицах (10, 11) из Приложения 2 представлен перечень анализируемых банков и результаты их классификации.

Можно заметить, что логит-регрессия имеет неплохую точность оценки, однако прогноз такого классификатора излишне пессимистичен, что проиллюстрировано в таблице (3).

### Результаты пробит-регрессии

В пакете R также были оценены коэффициенты пробит-модели:

$$z = 6,721 - 1,885 \cdot Risk - 0,144 \cdot H1 - 0,03 \cdot H3 - 0,027 \cdot H4.$$

Ее качество так же оценивалось на основе коэффициента правдоподобия (Likelihood ratio test). Этот критерий показал (см. Приложение 1), что пробит модель, построенная по выбранным факторам, лучше нулевой модели с вероятностью  $p = 0,932$ .

Эффективность модели будем оценивать так же по обучающей и контрольной выборкам. Результаты классификации представлены ниже в таблицах (4, 5).

Таблица 4: Результат пробит-регрессии на обучающей выборке

	Прогнозные		
Фактически	C1	C2	Общее число
C1	15	2	17
C2	2	4	6
Общее число	17	6	23

Как можно заметить, таблицы (4, 5) совпадают с аналогичными для логит-регрессии, а значит и доля верно классифицированных банков такая же.

В данном случае сложно решить, какая модель лучше. Обе модели



Таблица 5: Результат пробит-регрессии на контрольной выборке

Фактически	Прогнозные		
	<i>C1</i>	<i>C2</i>	Общее число
<i>C1</i>	12	0	12
<i>C2</i>	5	1	6
Общее число	17	1	23

показали одинаковые ошибки классификации. Отличается ненамного только оценка моделей с помощью информационного критерия Акаике (AIC) [24]:

- для логит-модели:  $AIC = 27,634$ ,
- для пробит-модели:  $AIC = 27,642$ .

Данный критерий не только оценивает адекватность модели, но и штрафует за использование лишних параметров. Поэтому, наилучшей считается модель,  $AIC$  которой наименьший.

## 2.2. Байесовский классификатор

Так как обе модели бинарной регрессии показали одни и те же результаты, то было решено для сравнения построить еще один классификатор. В основе многих классификаторов (например, линейного дискриминантного анализа) лежит предположение о нормальности реализации многомерной случайной величины, описывающей объекты исследования. Поэтому данные для обоих классов, описывающие факторы модели (уровень риска и нормативы Н1, Н3, Н4), были проверены с помощью критериев Шапиро–Уилка и Мардиа (Mardia's test) [25] на подчинение многомерному нормальному распределению.

Оба теста отклонили гипотезу  $H_0$  о многомерной нормальности данных, относящихся к закрытым банкам (класс  $C1$ ) (см. Приложение 2). Данные по действующим банкам (класс  $C2$ ) критерий Мардиа признает нормально распределенными, а тест Шапиро – Уилка снова отклоняет нулевую гипотезу. Так как данные для класса  $C1$  нельзя считать нормально распределенными, то и применять можно только классификаторы, не требующие нормальной реализации величин. Поэтому было решено использовать наивный байесовский классификатор [14].

Этот алгоритм классификации основывается только на предположении, что факторы не зависят друг от друга. Несмотря на то, что байесовский классификатор является одним из самых простых в реализации, на малом объеме наблюдений он часто работает лучше других алгоритмов.

При построении классификатора также использовались данные по 17-ти банкам, закрытым в конце 2015 года, и по 6-ти действующим. Все показатели взяты также за декабрь 2014 года.

Оценим точность классификатора с помощью пакета R, используя функцию `train()` из пакета `caret` [26], осуществляющую перекрестную проверку (cross-validation) модели. При перекрестной проверке выборка разбивается на  $k$  частей, из которых  $k - 1$  часть используется для построения модели, а оставшаяся — для тестирования. Процедура повторяется  $k$  раз, то есть каждая из частей по одному разу используется для проверки. В функции `train()` по умолчанию  $k = 10$ .

Точность классификатора на данных из обучающей выборки составила 63,3% (Приложение 1). При проведении перекрестной проверки с использованием контрольной выборки точность составила 86,6%.

Оценим также долю правильно классифицированных наблюдений. В качестве контрольной выборки используются также данные 12-ти банков,

закрытых в конце 2016 года, и 6-ти действующих банков.

Таблица 6: Результат байесовского классификатора на обучающей выборке

Фактически	Прогнозные		
	<i>C1</i>	<i>C2</i>	Общее число
<i>C1</i>	9	8	17
<i>C2</i>	1	5	6
Общее число	10	13	23

Таблица 7: Результат байесовского классификатора на контрольной выборке

Фактически	Прогнозные		
	<i>C1</i>	<i>C2</i>	Общее число
<i>C1</i>	9	3	12
<i>C2</i>	5	1	6
Общее число	14	4	23

В таблицах (6, 7) представлен результат работы классификатора на обучающей и контрольной выборках. Доля правильно классифицированных наблюдений составила:

- для обучающей выборки:

$$\frac{9 + 5}{23} \times 100\% = 60,87\%,$$

- для контрольной выборки:

$$\frac{9 + 1}{18} \times 100\% = 55,55\%.$$

В Приложении 2 размещены результаты классификации коммерческих банков, отобранных для исследования, с помощью всех вышеописанных методов.

Для каждого фактора по отдельности были построены графики плотности распределения (см. рис. 8–11). Красной сплошной линией отмечена

плотность случайных величин из класса  $C1$ , голубая прерывистая соответствует классу  $C2$ .

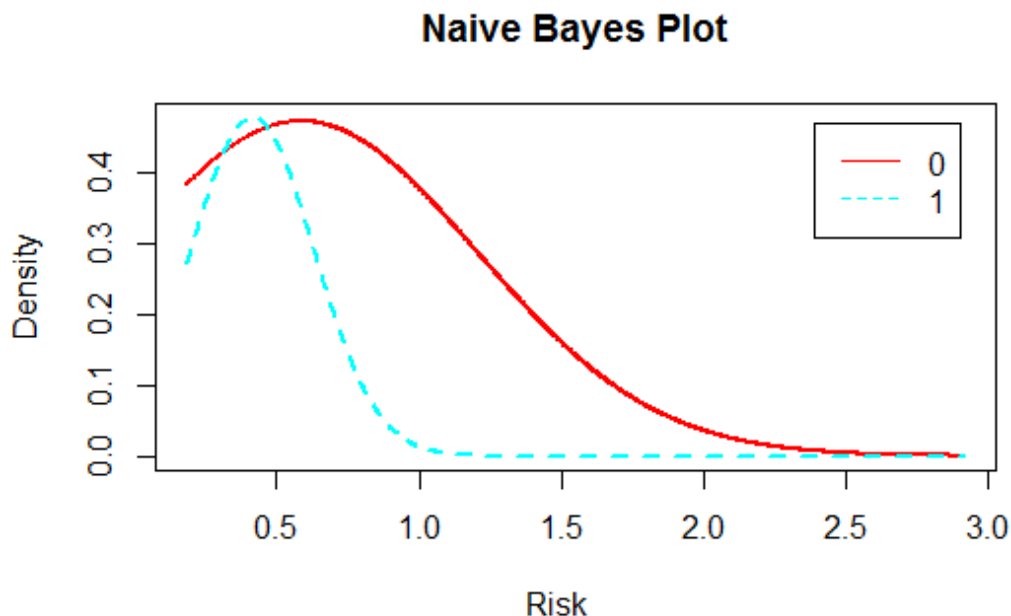


Рис. 8: Плотности случайных величин, описывающих уровень риска

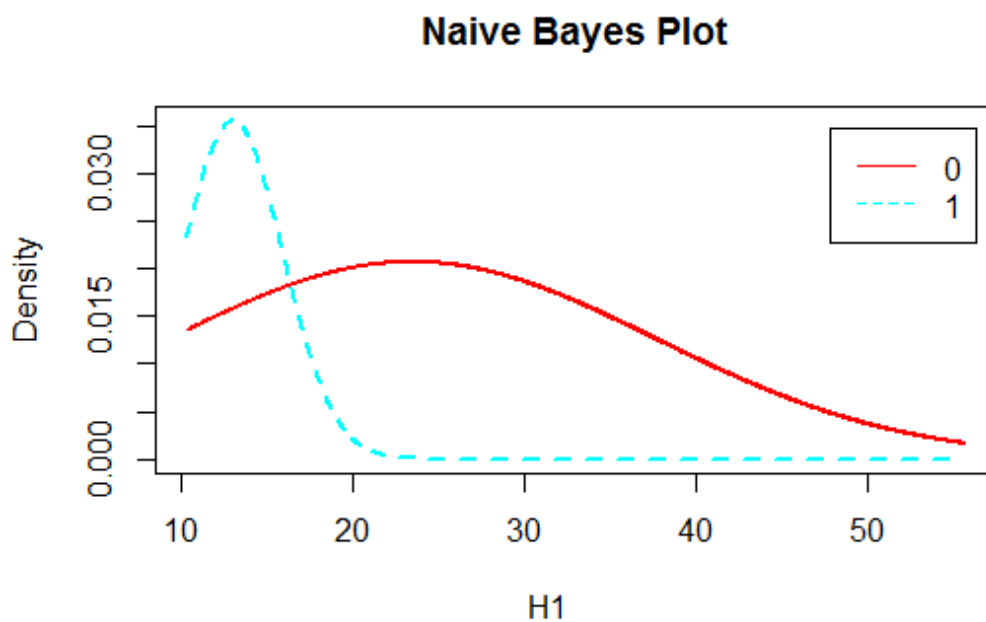


Рис. 9: Плотности случайных величин, описывающих норматив  $H1$

Как можно заметить, сильнее всего классы отличаются друг от друга по таким параметрам, как уровень риска и норматив достаточности капитала. По нормативу долгосрочной ликвидности вообще сложно разделить классы.

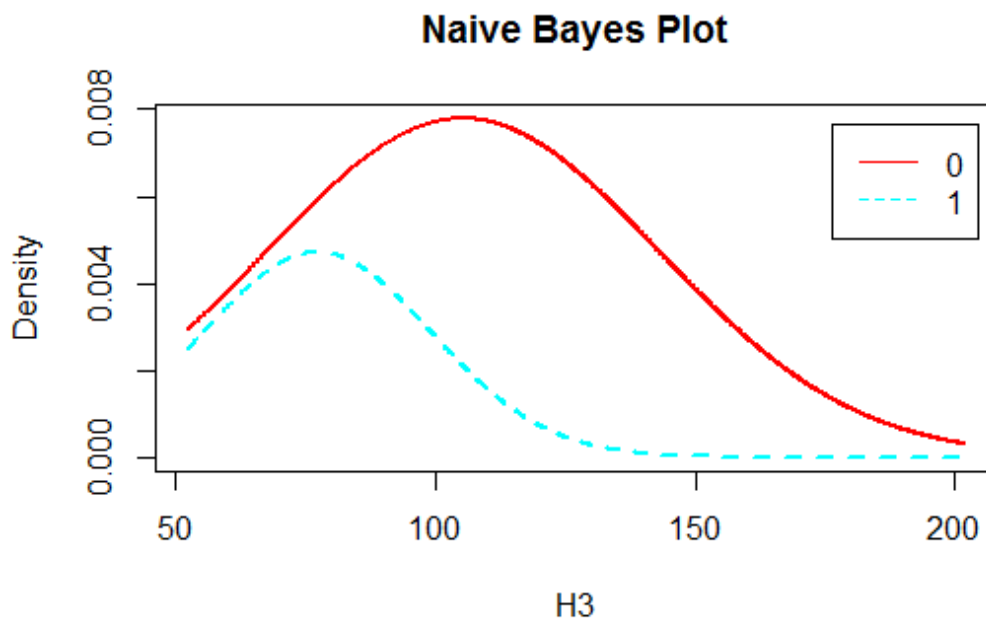


Рис. 10: Плотности случайных величин, описывающих норматив H3

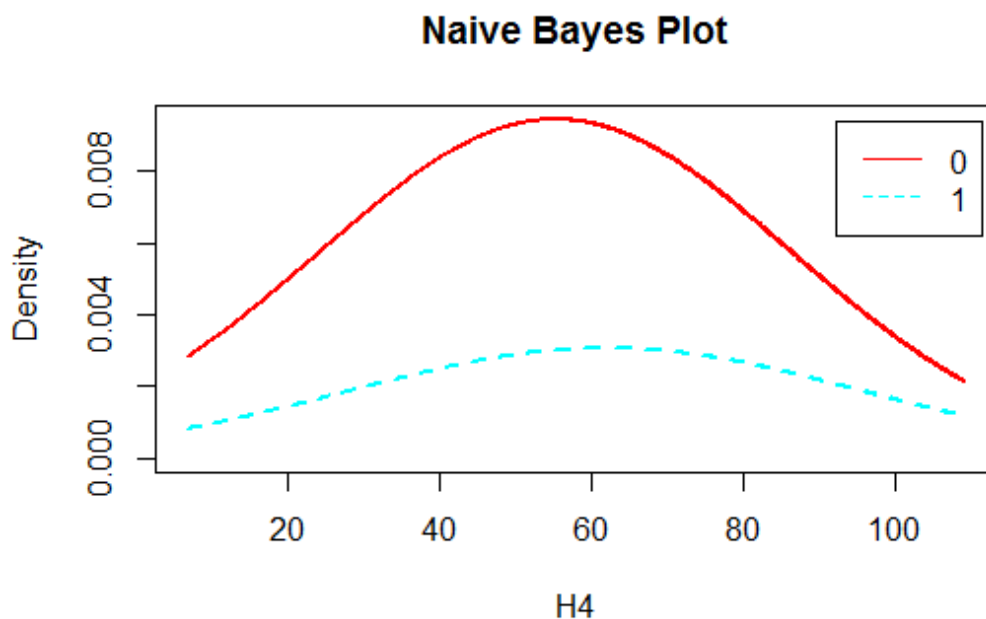


Рис. 11: Плотности случайных величин, описывающих норматив H4

Результаты применения вышеописанных алгоритмов к данным по действующим банкам совпали. Поэтому сравнение классификаторов можно проводить только по их использованию на данных закрытых банков. И модели бинарной регрессии определили больше наблюдений к истинному классу, чем байесовский классификатор.

## Глава 3. Использование *ARIMA* моделей

В этой главе подробнее анализируются оценки доходности активов и кредитного портфеля в целом, вычисление которых было описано в Главе 1.

### 3.1. Описание данных

Оценки доходности для каждого актива, а также кредитного портфеля, вычислены для каждого квартала за период с 2008 по 2015 год (для 12-ти банков по 2016 год), то есть полученный набор оценок можно рассматривать как временной ряд.

В большинстве исследований анализ временных рядов сводится к его декомпозиции [27], то есть разделению ряда на трендовую, сезонную, циклическую и случайную составляющую. Пример аддитивной модели:

$$y_t = u_t + s_t + \nu_t + \epsilon_t,$$

где  $y_t$  — это наблюдаемые значения,  $u_t$  — значение тренда,  $s_t$  — сезонная компонента,  $\nu_t$  — циклическая компонента,  $\epsilon_t$  — случайная компонента.

Подобные модели удобно применять к рядам, отражающим, например, спрос на продукты питания. Это связано с тем, что, в зависимости от времени года, люди больше покупают те или иные продукты, например, летом охотнее покупают мороженое, зимой — мандарины. У таких рядов есть явные сезонные колебания, их тренд также нетрудно проследить.

Однако рассматриваемые в исследовании временные ряды не имеют какого-либо явного тренда, и тем более сезонности. В качестве иллюстрации ниже приведен график (см. рис. 12) оценок доходностей от кредитов юридическим и физическим лицам АО «Торговый Городской Банк». Графики оценок доходностей для «КБ „Ренессанс“» и «АКБ „Енисей“», пред-

ставленные ранее в Главе 1 (см. рис. 1–4), также не имеют тренда и сезонной компоненты. Поэтому, для анализа удобнее использовать *ARIMA* модели [15].

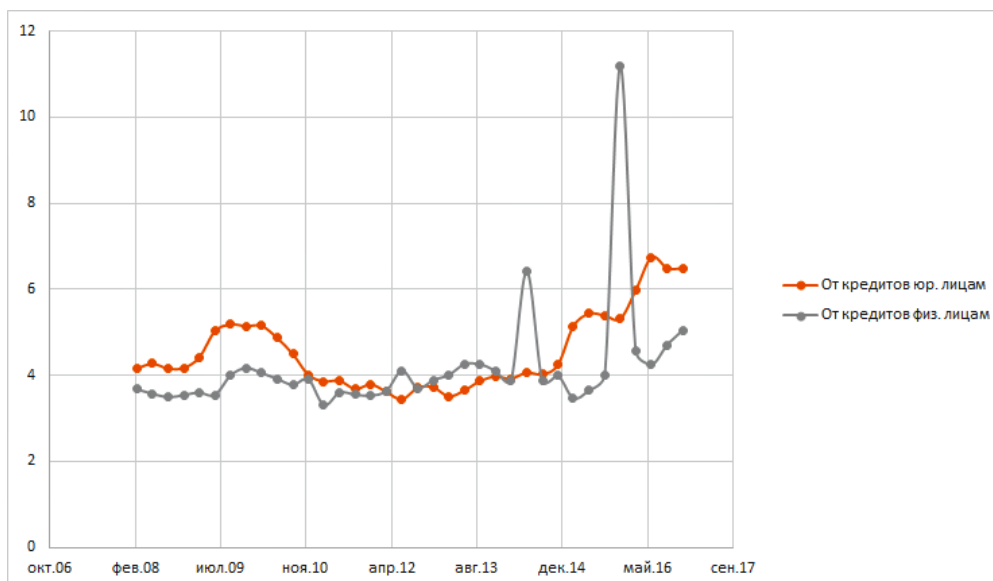


Рис. 12: Доходность по активам АО «Торговый Городской Банк»

### 3.2. Выбор параметров модели *ARIMA*

Модели  $ARIMA(p, d, q)$  являются объединением авторегрессии и модели скользящего среднего [16]. Авторегрессионная модель ( $AR(p)$ ) учитывает влияние предыдущих значений временного ряда на текущее:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \epsilon_t,$$

где  $a_i$  — коэффициенты модели,  $\epsilon_t$  — ошибка модели.

Идея модели скользящего среднего ( $MA(q)$ ) заключается в том, что текущее значение временного ряда может зависеть от случайных ошибок предыдущих наблюдений:

$$y_t = c_1 \epsilon_{t-1} + c_2 \epsilon_{t-2} + \dots + c_q \epsilon_{t-q} + \epsilon_t,$$

где  $c_i$  — коэффициенты модели.

Модель  $ARMA(p, q)$  объединяет в себе модели  $AR(p)$  и  $MA(q)$ :

$$y_t = a_0 + a_1 y_{t-1} + \dots + a_p y_{t-p} + c_1 \epsilon_{t-1} + \dots + c_q \epsilon_{t-q} + \epsilon_t.$$

Авторегрессия и модель скользящего среднего основываются на предположении о стационарности временного ряда. В широком смысле стационарным называется такой ряд, дисперсия и математическое ожидание которого не меняются со временем. Но большинство рядов не являются стационарными. Избавиться от нестационарности временного ряда помогает взятие разности между наблюдениями ряда:

$$\Delta^1 y_t = y_t - y_{t-1}.$$

$ARMA(p, q)$  модель, построенная на основе разностей, обозначается  $ARIMA(p, d, q)$ , где параметр  $d$  отвечает за порядок разности.

В этой главе проводимый анализ будет иллюстрироваться на рядах, описывающих доходность портфеля от кредитов банкам и от кредитов юридическим лицам АО «Торговый Городской Банк». Графики этих рядов приведены ниже (см. рис. 13, 14).

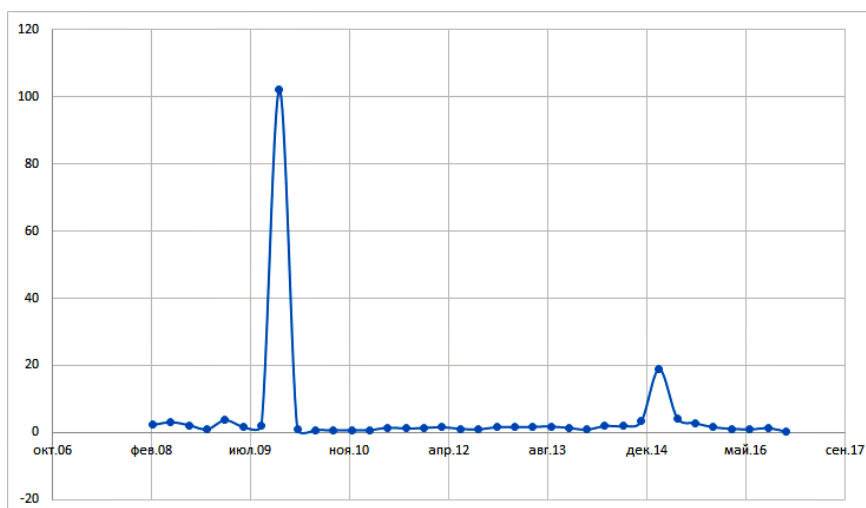


Рис. 13: Доходность от кредитов, выданных банкам

По исходным рядам сложно сказать какой порядок разности будут иметь модели  $ARIMA(p, d, q)$ . Более точные выводы помогают сделать



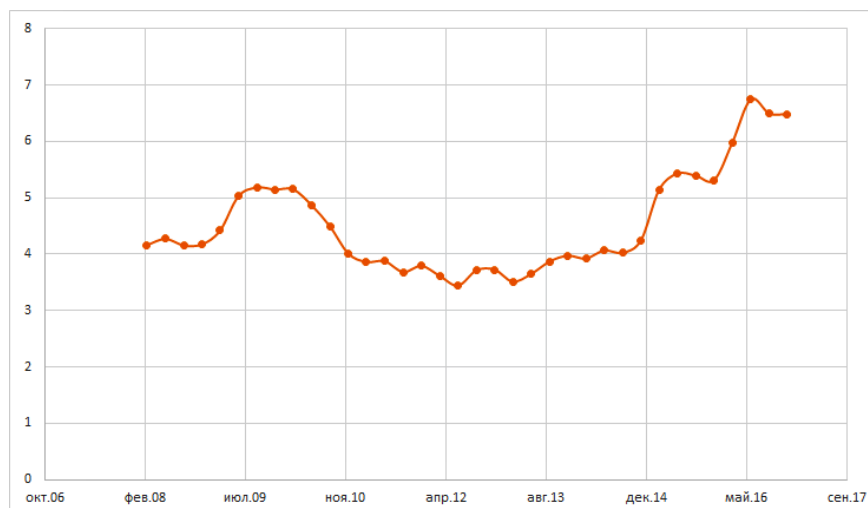


Рис. 14: Доходность кредитов, выданных юридическим лицам

графики автокорреляционной функции (ACF) (см. рис. 15, 16).

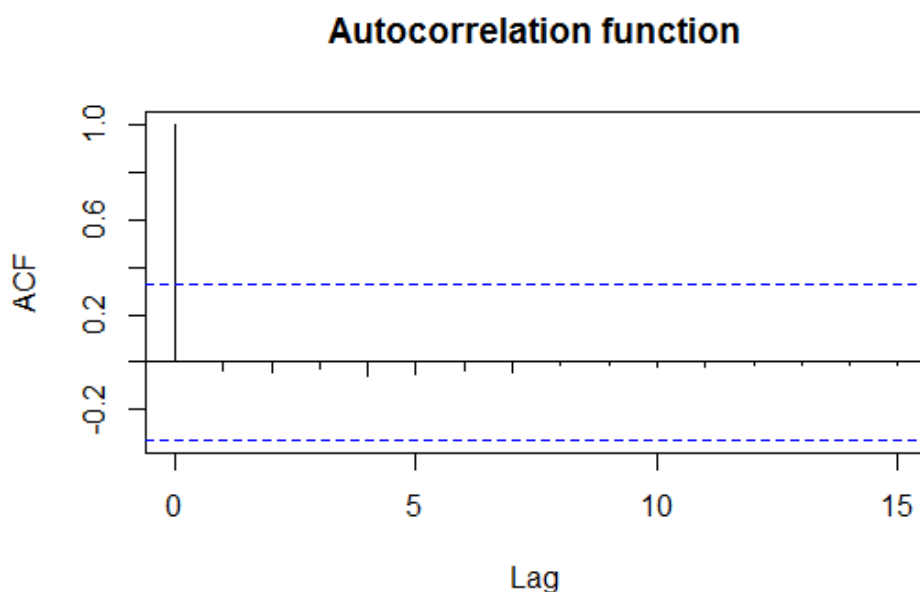


Рис. 15: Коррелограмма ряда доходности от кредитов банкам

Первая коррелограмма (см. рис. 15) не содержит значимых коэффициентов (все значения ниже уровня значимости), следовательно, ряд не имеет тренда и, возможно, является стационарным. График второй автокорреляционной функции (см. рис. 16) имеет значимые коэффициенты на первом, втором и третьем лаге, а значит ряд имеет тенденцию и не является стационарным.

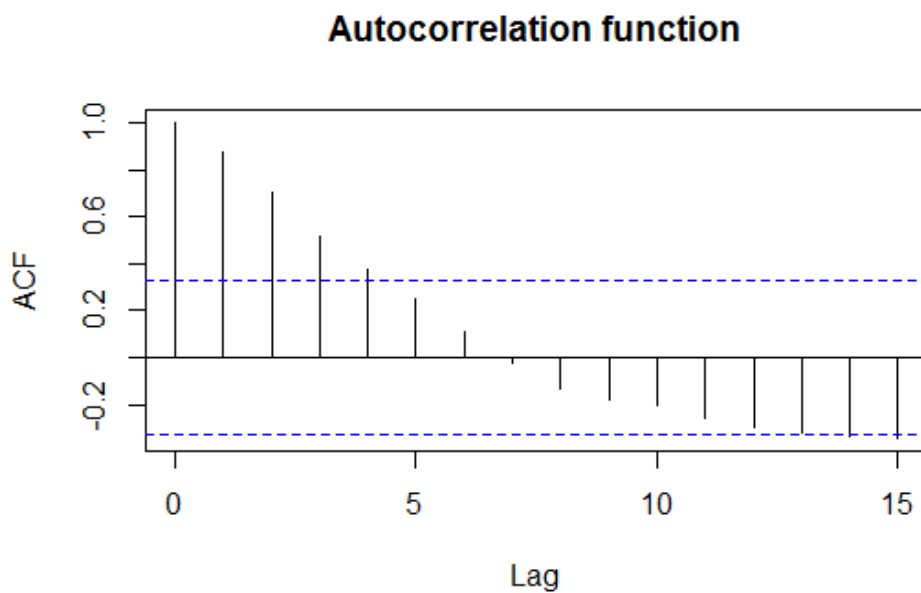


Рис. 16: Коррелограмма ряда доходности от кредитов юридическим лицам

Рассмотрим также графики частной автокорреляционной функции (PACF) (см. рис. 17, 18)

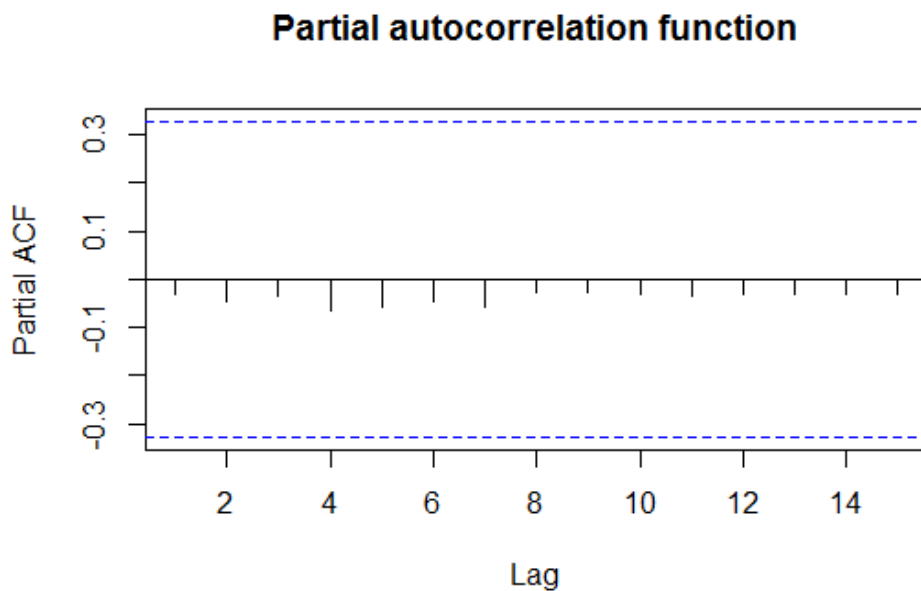


Рис. 17: PACF ряда доходности от кредитов банкам

По значимым коэффициентам частной автокорреляционной функции можно определить период сезонной компоненты. Но в рассматриваемом случае все коэффициенты незначимы, следовательно, ряд не имеет сезон-

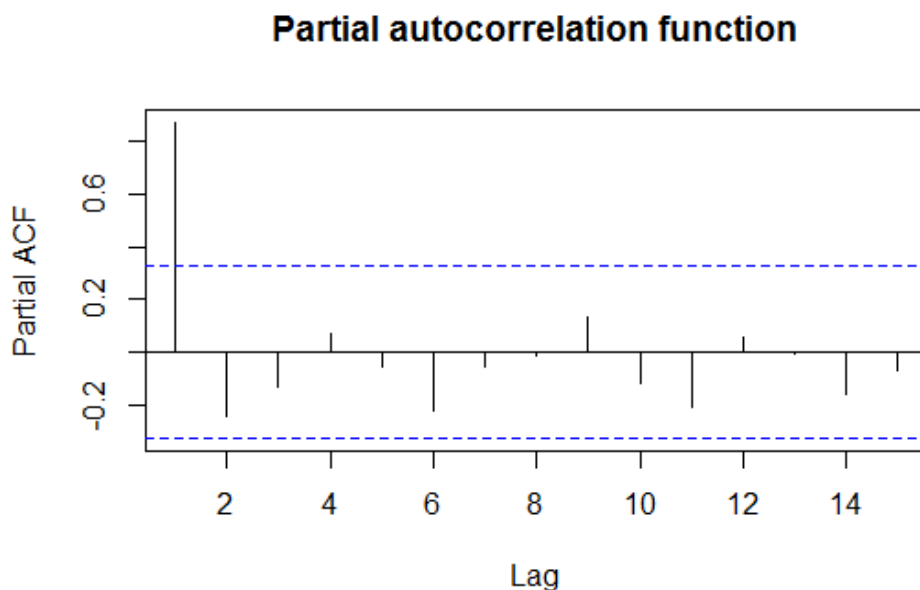


Рис. 18: PACF ряда доходности от кредитов юридическим лицам

ности.

Так как коррелограммы не дают точного ответа, какой порядок разности выбрать для модели, проведем тест Дики–Фуллера на стационарность. Гипотезу о стационарности исходных рядов тест Дики–Фуллера отклонил (см. Приложение 3), следовательно, нужно взять разность между значениями ряда и проверить снова.

Ряд доходности от кредитов банкам оказался стационарным при взятии первой разности, а ряд доходности от кредитов, выданных юридическим лицам, становится стационарным при взятии второй разности. Следовательно, для первого ряда параметр  $d = 1$ , а для второго  $d = 2$ .

Опытные аналитики могут только по графикам ACF и PACF сказать, какие параметры  $p$  и  $q$  следует использовать в модели [16]. Но мы воспользуемся информационным критерием Акаике [24], который уже упоминался в Главе 2. Этот критерий часто применяется для выбора наилучшей модели. Ниже представлена таблица (8) со значениями коэффициента AIC для ряда доходности от кредитов, выданным банкам. Наименьшее значение

Таблица 8: Выбор модели  $ARIMA(p, d, q)$  для ряда доходности от кредитов банкам

Модель $ARIMA(p, d, q)$	Значение $AIC$
$ARIMA(0, 1, 0)$	291,53
$ARIMA(0, 1, 1)$	274,34
$ARIMA(0, 1, 2)$	276,34
$ARIMA(0, 1, 3)$	278,33
$ARIMA(1, 1, 0)$	285,12
$ARIMA(1, 1, 1)$	276,34
$ARIMA(1, 1, 2)$	278,23
$ARIMA(2, 1, 0)$	283,51
$ARIMA(2, 1, 1)$	278,33
$ARIMA(3, 1, 0)$	283,89

Таблица 9: Выбор модели  $ARIMA(p, d, q)$  для ряда доходности от кредитов юридическим лицам

Модель $ARIMA(p, d, q)$	Значение $AIC$
$ARIMA(0, 2, 0)$	12,63
$ARIMA(0, 2, 1)$	11,8
$ARIMA(0, 2, 2)$	5,61
$ARIMA(0, 2, 3)$	7,15
$ARIMA(1, 2, 0)$	13,96
$ARIMA(1, 2, 1)$	9,12
$ARIMA(1, 2, 2)$	7
$ARIMA(2, 2, 0)$	10,55
$ARIMA(2, 2, 1)$	10,36
$ARIMA(3, 2, 0)$	12,5

коэффициента  $AIC$  соответствует модели  $ARIMA(0, 1, 1)$ , то есть она состоит только из модели скользящего среднего порядка 1, построенной на основе первых разностей.

Определим далее параметры для ряда доходности от кредитов юридическим лицам.

По коэффициентам АИС, представленным в таблице (9), для ряда доходности от кредитов юридическим лицам, наилучшей является модель  $ARIMA(0, 2, 2)$ , она представляет собой модель скользящего среднего порядка 2, построенной по ряду вторых разностей.

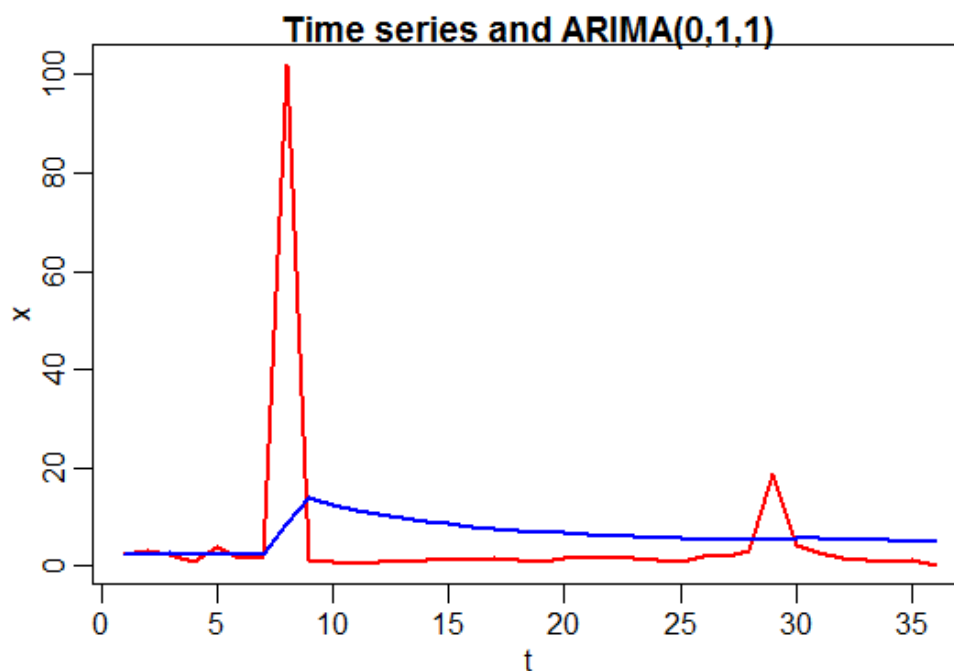


Рис. 19: Исходный ряд доходности от кредитов банкам и модель  $ARIMA$

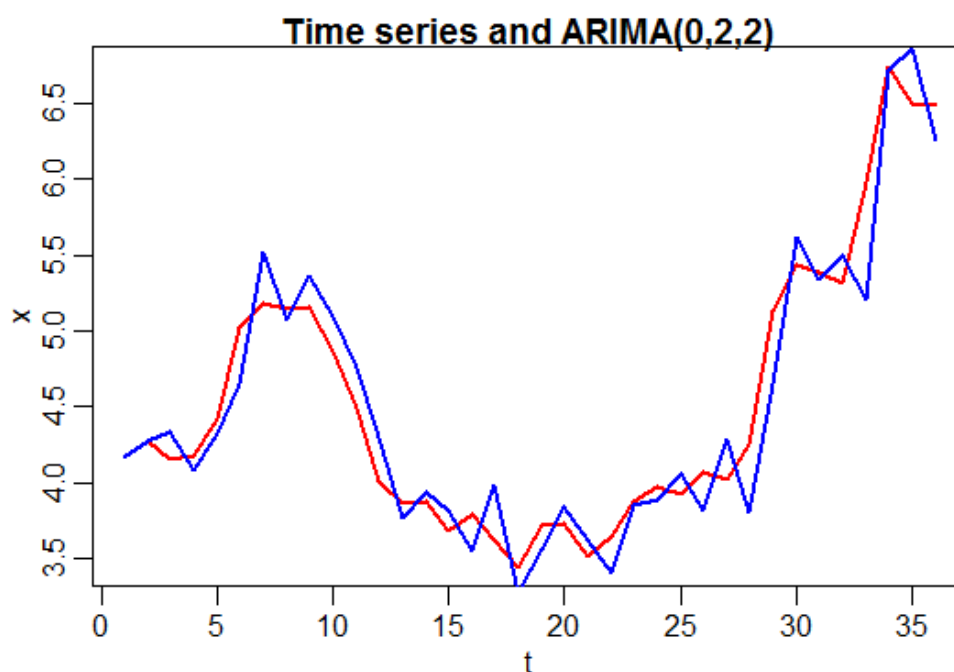


Рис. 20: Исходный ряд доходности от кредитов юридическим лицам и модель  $ARIMA$

На графиках выше (см. рис. 19, 20) красным цветом изображен исходный ряд, а синим — модельный.

### 3.3. Исследование остатков модели

Рассмотрим остатки построенных моделей, которые находятся как разность между наблюдаемыми значениями ряда и модельными. Они должны представлять собой белый шум [16], то есть не коррелировать между собой. Для этого построим коррелограммы остатков.

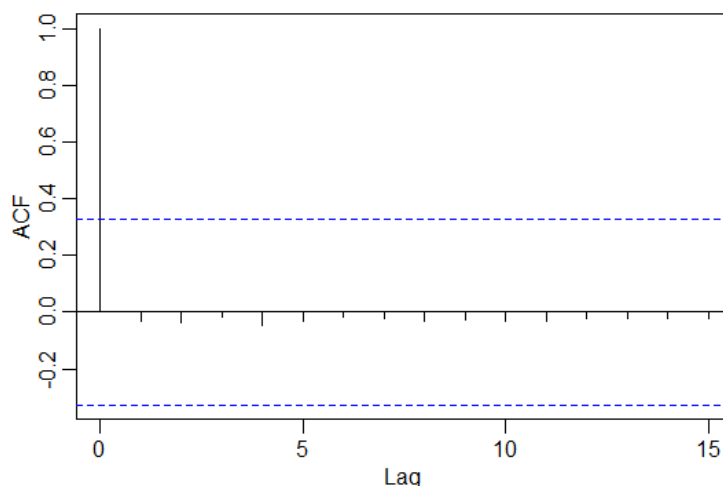


Рис. 21: Коррелограмма остатков модели ряда доходности от кредитов банкам

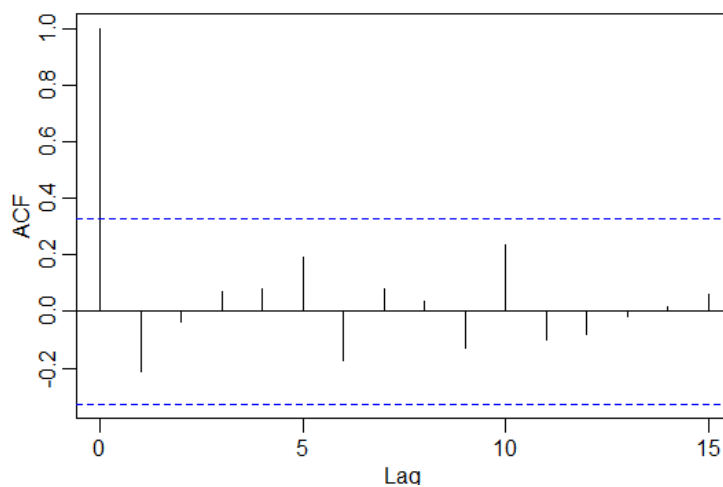


Рис. 22: Коррелограмма остатков модели ряда доходности от кредитов юридическим лицам

Значения автокорреляционных функций на графиках (21, 22) не яв-

ляются значимыми, следовательно, остатки можно считать независимыми. Подтвердим это с помощью критерия Льюнга – Бокса.

Остатки первого ряда тест признает случайными с вероятностью  $p = 0,84$ , а второго с вероятностью  $p = 0,19$  (см. Приложение 3). Так как оба значения вероятности  $p > 0,05$ , то остатки для обоих рядов можно считать независимыми. Это значит, что в построенных моделях  $ARIMA(p, d, q)$  учтены все зависимости.

Для признания моделей адекватными достаточно независимости остатков, которую подтвердил тест Льюнга – Бокса. Но для полноты картины исследуем также остатки на подчинение нормальному распределению.

Для этого изучим гистограммы остатков (см. рис. 23, 24).

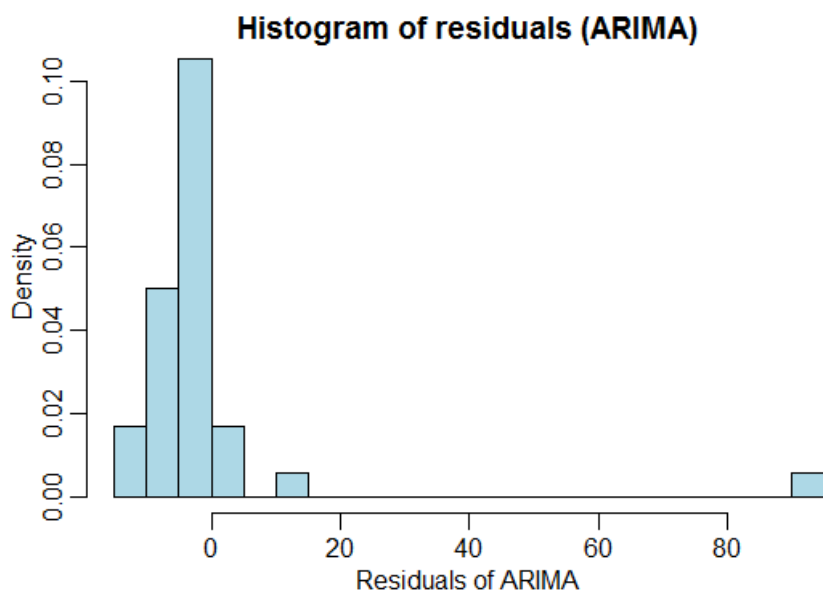


Рис. 23: Гистограмма остатков модели ряда доходности от кредитов банкам

По гистограмме (24) видно, что остатки не подчиняются нормальному распределению. Это подтвердили критерии Шапиро – Уилка и Колмогорова – Смирнова (см. Приложение 3). Гипотезу о нормальности распределения критерий Шапиро – Уилка отклоняет с вероятностью  $p = 0,16$ , а критерий Колмогорова – Смирнова с вероятностью  $p = 0,86$ . Остатки первой модели (см. рис. 23) проверка признала нормально распределенными (см. Приложение 3).

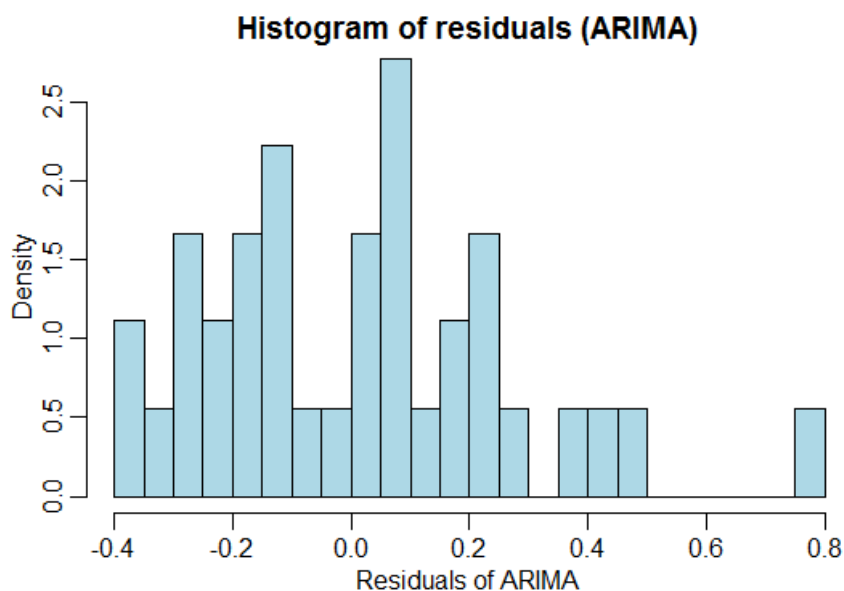


Рис. 24: Гистограмма остатков модели ряда доходности от кредитов юридическим лицам

Несмотря на то, что остатки второй модели не принадлежат нормальному распределению, модель все равно можно считать адекватной и использовать при прогнозировании.

### 3.4. Примеры построения прогноза с использованием моделей *ARIMA*

Построенные модели  $ARIMA(0, 1, 1)$  и  $ARIMA(0, 2, 2)$  далее можно использовать для прогнозирования. Ряды доходности по активам АО «Торговый Городской Банк» содержат по 36 наблюдений за период с 2008 по 2016 год, то есть в год мы имеем 4 наблюдения. Разобьем этот ряд на 2 части. Первая часть состоит из 32-х наблюдений с 2008 по 2015 год. А вторая содержит 4 наблюдения, соответствующие 2016 году. По первой части наблюдений построим прогноз на 2016 год и сравним его с реальными значениями.

В пакете R есть множество функций, позволяющих построить прогноз по имеющейся модели. С точки зрения удобства лучше использовать функцию `sarima.for()`, так как она строит график прогноза и довери-



тельные интервалы с уровнями доверия 0,95 и 0,8.

Графики прогнозов, построенных с помощью функции `sarima.for()` представлены на рисунках (25, 26). Синяя линия соответствует реальным значениям ряда доходности. Красным цветом выделены прогнозные значения. Серая область — это доверительный интервал прогноза.

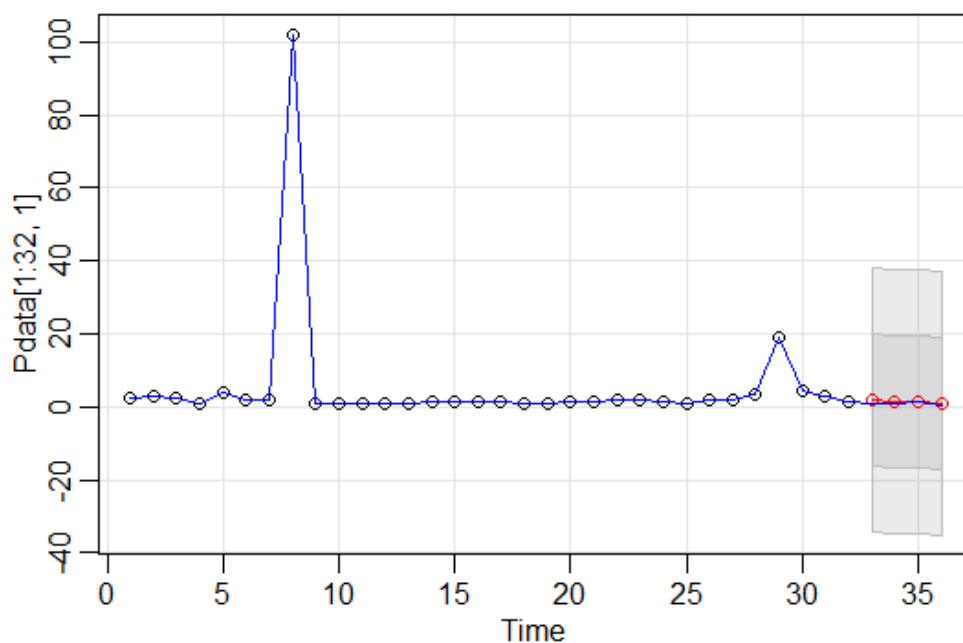


Рис. 25: Прогноз ряда доходности от кредитов банкам

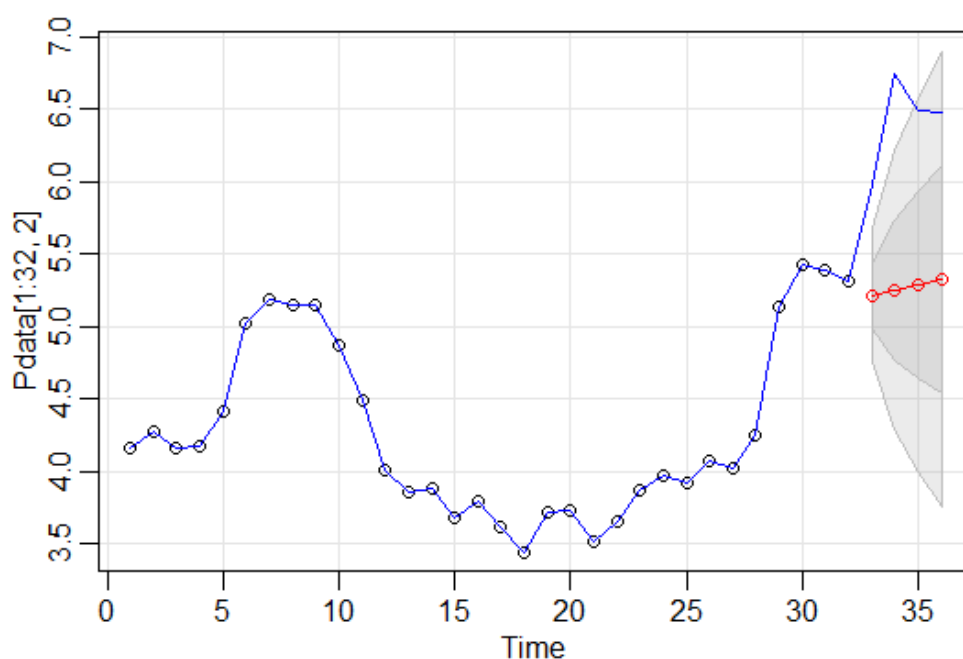


Рис. 26: Прогноз ряда доходности от кредитов юридическим лицам

Оценку прогноза проведем с помощью средней ошибки аппроксимации, вычисляемой по формуле

$$\bar{A} = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t|} \times 100\%,$$

где  $y_t$  соответствует наблюдаемым значениям ряда, а  $\hat{y}_t$  — прогнозным значениям.

Для первого ряда она составила 89,12%, что является довольно большим значением. Для второго ряда ошибка имеет вполне допустимое значение  $\bar{A} = 17,79\%$ .

Больше половины построенных прогнозов имеют ошибку аппроксимации меньше 20%. Поэтому прогнозирование рядов на основе моделей  $ARIMA(p, d, q)$  является довольно точным.

Используя прогноз доходности по активам, коммерческий банк может оценить все параметры кредитного портфеля: ожидаемую доходность портфеля, риск каждого актива и риск портфеля в целом. Но для этого нужен план по выдаче кредитов, чтобы вычислить долю активов в портфеле.

Построим прогноз риска кредитного портфеля на 2017 год на примере банков ПАО «Уральский Банк Реконструкции и Развития» и «КБ „Ренессанс“». Первый банк все алгоритмы классификации отнесли к группе надежных, а второй является показательным примером банка, проводившего высокорискованную кредитную политику. Так как нет информации о том, какие суммы планируется выдавать банкам, юридическим и физическим лицам, то построим также и прогноз доли активов в кредитном портфеле.

На графиках ниже (см. 27, 28) оранжевым цветом отмечен прогноз риска, а синим — вычисленный по имеющимся данным риск.

Для «КБ „Ренессанс“» прогноз риска на 2015 год меньше, чем вы-

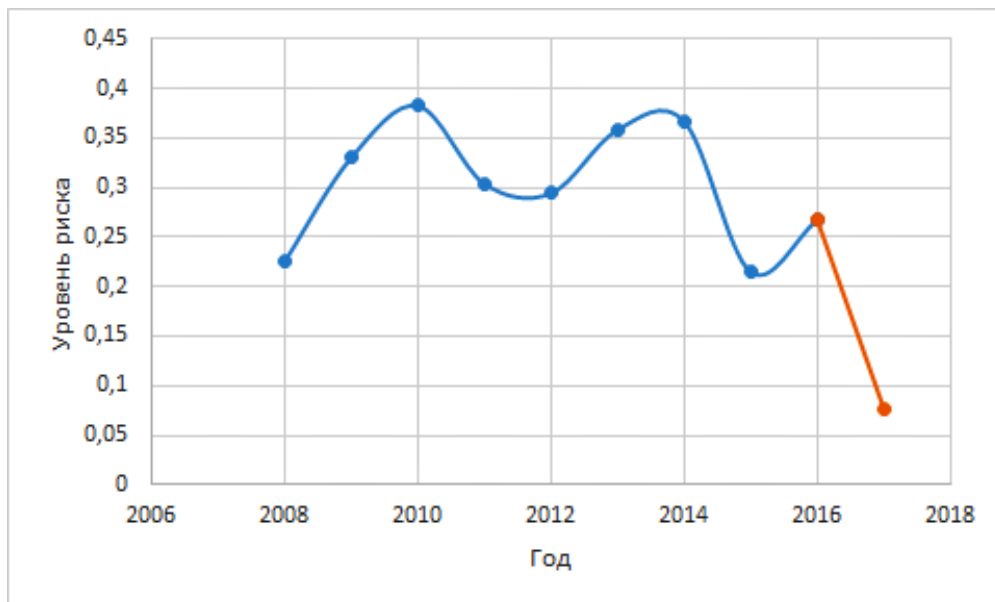


Рис. 27: Прогноз риска кредитного портфеля ПАО «Уральский Банк Реконструкции и Развития»

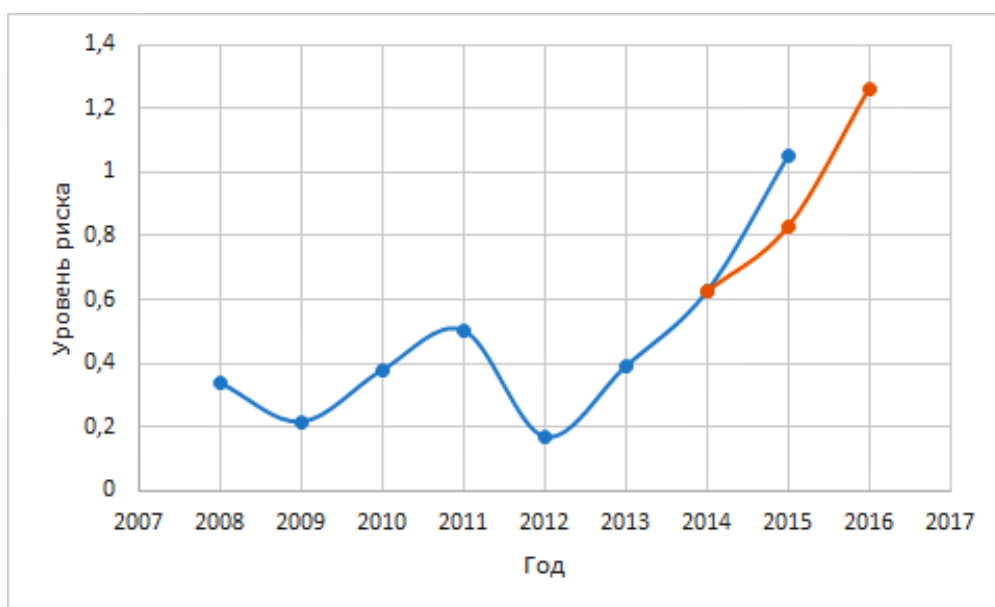


Рис. 28: Прогноз риска кредитного портфеля «КБ „Ренессанс “»

численный, однако, он соответствует тенденции возрастания. И, если бы, у банка не отозвали лицензию в конце 2015 года, то в 2016 году риск вырос бы еще больше.

Прогноз риска для «Уральского Банка Реконструкции и Развития» хороший. Но проверить результаты прогноза можно будет только по итогам 2017 года.

## Выводы

Решение поставленной в исследовании задачи, а именно разработки методики заблаговременной оценки риска, было проведено в 3 этапа.

Первый этап описан в Главе 1. Он также состоял из подзадач, первой из которых был поиск данных, подходящих под тематику исследования и пригодных к обработке. В результате все необходимые данные были найдены на сайте КУАП.ру [18], а именно агрегированный баланс, отчет о прибылях и убытках и Форма 135. Далее на основе найденных данных были вычислены параметры кредитного портфеля такие, как доходность и риск активов, ожидаемая доходность портфеля и риск портфеля.

На втором этапе, в Главе 2, используя значения параметров кредитного портфеля и обязательных нормативов, были построены правила классификации моделей по критерию надежности функционирования. Для этих целей были использованы модели бинарной регрессии и байесовский классификатор. Построенный байесовский классификатор показал низкую точность классификации. Логит и пробит-регрессии показали более высокую долю верно классифицированных банков, но на большем объеме наблюдений результаты были бы лучше, так как, согласно теории, изложенной авторами [28], на каждый фактор в бинарной регрессии должно приходиться 10 наблюдений, что не может быть обеспечено в данной работе.

Последний этап, описанный в Главе 3, заключался в построении модели временных рядов, описывающих доходность активов, формирующих кредитный портфель. Для этого были использованы модели *ARIMA*, объединяющие в себе авторегрессию и модель скользящего среднего. На основе полученных моделей были построены прогнозы рядов доходности активов. Эти прогнозы достаточно точны, а значит, они могут использоваться для прогноза риска портфеля.

## Заключение

В результате проведенного исследования было установлено, что для выявления ухудшения качества кредитного портфеля необходимо оценить такие характеристики, как доходность и риск активов, ожидаемая доходность портфеля и риск портфеля. Непосредственный анализ данных характеристик может выявить общую динамику изменения качества кредитного портфеля. Используя их в совокупности с обязательными нормативами, можно оценить вероятность отзыва лицензии у коммерческого банка через год. Она находится с помощью построенных моделей бинарной регрессии.

Также, построив модель  $ARIMA(p, d, q)$  для рядов доходности активов, можно сделать прогноз на ближайший год. Большинство построенных прогнозов имело среднюю ошибку аппроксимации менее 20%, а значит их можно использовать в дальнейших расчетах. Например, при планировании выдач кредитов, можно рассчитать, какую доходность и какой риск будет иметь такой кредитный портфель. В данной работе доля выданных кредитов также была спрогнозирована, и на основе полученных прогнозов был вычислен риск на ближайший год. Для банков с отозванной лицензией риск вычислялся на последний год работы. И хотя прогнозные значения оказывались меньше, чем вычисленные, они соответствовали общей динамике. Следовательно, кредитная организация может управлять уровнем риска портфеля, планируя показатели выдачи кредитов.

Таким образом, значение вероятности, полученное с помощью логит или пробит-регрессий должно сигнализировать об ухудшении качества кредитного портфеля. А для того, чтобы повлиять на сложившуюся ситуацию в следующем году, коммерческому банку следует при планировании выдачи кредитов построить прогноз по активам и по нему рассчитать будущий риск, воспользовавшись методикой, предложенной в [2, 4, 5]

## Список литературы

1. Печалова М. Ю. Организация риск-менеджмента // Менеджмент в России и зарубежом. 2001. № 1.
2. Буренин А. Н. Рынок ценных бумаг и производных финансовых инструментов: Учебное пособие. М.: 1 Федеративная Книготорговая Компания, 1998. — 352 с.
3. Лобанов А. А., Чугунов А. В. Энциклопедия финансового риск-менеджмента. М.: Альпина Паблишер, 2003. 786 с.
4. Markowitz H. Portfolio Selection // Journal of Finance, Vol. 7. 1952. No 1 P. 77–91.
5. Теория принятия решений. В 2 т. Т. 2: Учебник и практикум для бакалавриата и магистратуры / под ред. В. Г. Халина. М.: Издательство Юрайт, 2016. 431 с.
6. Вестник Банка России / под ред. Г. И. Лунтовский. 2012. № 74. 73 с.
7. Jorion P. Financial Risk Manager Handbook: Second Edition. New Jersey: Wiley, 2003. 736 p.
8. Бобыль В. Методика применения показателей системы риск-менеджмента // Банковский вестник. 2014. № 6.
9. Бухштабер В. М, Оводов И. Г, Шевченко С. Н. Статистический подход к проблеме оценки надежности коммерческих банков // Экономический журнал ВШЭ 1998. № 1.
10. Шоломицкий А. Г. Теория риска. Выбор при неопределенности и моделирование риска. М.: Изд. дом ГУ ВШЭ, 2005. 400 с.

11. Basel III: A global regulatory framework for more resilient banks and banking systems [Электронный ресурс] // Basel Committee on Banking Supervision. 2010. <http://www.bis.org/publ/bcbs189.pdf>
12. Груздев А. В. Метод бинарной логистической регрессии в банковском скоринге // Риск-менеджмент в кредитной организации. 2012. № 1(5). С. 71–88.
13. Greene W. H. Econometric analysis, 5-th edition. New Jersey: Prentice Hall, 2003. 1054 p.
14. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
15. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс: Учеб. — 6-е изд., перераб. и доп. М.: Дело, 2004. 576 с.
16. Светуных И. С., Светуных С. Г. Методы социально-экономического прогнозирования. В 2 т. Т. 2. М.: Издательство Юрайт, 2015. 447 с.
17. Информационный портал Банки.ру. <http://www.banki.ru/banks/memory/>
18. КУАП. Финансовый анализ банков. <http://kuap.ru/>
19. Батракова Л. Г. Экономический анализ деятельности коммерческого банка. Учебник для вузов — 2 изд. М.: Логос, 2007.
20. Shapiro S. S., Wilk M. B. An analysis of variance test for normality // *Biometrika*, Vol. 52. 1965. No 3 P. 591–611.
21. Буре В. М., Парилина Е. М. Теория вероятностей и математическая статистика. СПб.: Изд-во Лань, 2013. 416 с.

22. Лемешко Б. Ю, Лемешко С. Б, Постовалов С. Н. Мощность критериев согласия при близких альтернативах. // Измерительная техника. 2007. № 2. С. 22-27.
23. Мищенко А. В, Чижова А. С. Методология управления кредитным риском и оптимальное формирование кредитного портфеля // Финансовый менеджмент. 2008. № 1. С. 91–105.
24. Akaike H. A new look at the statistical model identification // IEEE Transactions on Automatic Control, Vol. 19. 1974. No 6 P. 716–723.
25. Mardia K. V. Measures of multivariate skewness and kurtosis with applications // Biometrika. 1970. No 3. P. 519–530.
26. Шитиков В. К., Мастицкий С. Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. Электронная книга, адрес доступа: <https://github.com/ranalytics/data-mining>. 2017.
27. Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice. Otexts, 2014. 291 p.
28. Hosmer D. W., Lemeshow S. Applied Logistic Regression. — 2nd ed. Wiley, 2005. 397 p.



## Приложение

### Приложение 1.

В этом приложении представлены результаты тестов из Главы 2, проведенные в пакете R.

#### Likelihood ratio test

Вычисление коэффициента правдоподобия для логит-модели:

```
> p <- with(mylogit, pchisq(null.deviance - deviance,  
  df.null - df.residual, lower.tail = FALSE))  
> 1-p  
[1] 0.9328408
```

Вычисление коэффициента правдоподобия для пробит-регрессии:

```
> p <- with(myprobit, pchisq(null.deviance - deviance,  
  df.null - df.residual, lower.tail = FALSE))  
> 1-p  
[1] 0.9326321
```

#### Проверка многомерной нормальности

Применение критериев Шапиро–Уилка и Мардиа к данным закрытых банков (класс  $C1$ ) из обучающей выборки:

```
> mshapiro.test(t(data[1:17,3:6]))  
Shapiro-Wilk normality test  
data:  Z  
W = 0.46432, p-value = 6.546e-07  
  
> mardiaTest(data[1:17,3:6], cov = TRUE, qqplot = TRUE)
```

## Mardia's Multivariate Normality Test

---

```
data : data[1:17, 3:6]
g1p      : 17.46562
chi.skew : 49.48593
p.value.skew : 0.000262258
g2p      : 29.7776
z.kurtosis : 1.719181
p.value.kurt : 0.08558147
chi.small.skew : 62.37723
p.value.small : 3.033711e-06
Result   : Data are not multivariate normal.
```

---

Проверка данных по действующим банкам (класс *C2*) из обучающей выборки на подчинение многомерному нормальному распределению:

```
> mshapiro.test(t(data[18:23,3:6]))
```

Shapiro-Wilk normality test

data: Z

W = 0.57005, p-value = 0.0001962

```
> mardiaTest(data[18:23,3:6], cov = TRUE, qqplot = TRUE)
```

Mardia's Multivariate Normality Test

---

```
data : data[18:23, 3:6]

g1p      : 13.3526
chi.skew : 13.3526
```

```
p.value.skew      : 0.8617484
g2p               : 18.06483
z.kurtosis        : -1.0492
p.value.kurt      : 0.2940863
chi.small.skew    : 24.1728
p.value.small     : 0.2349253
Result            : Data are multivariate normal.
```

---

### Оценка точности классификатора

Оценим точность алгоритма байесовского классификатора на данных обучающей выборке:

```
> train_control <- trainControl(method = 'cv', number = 10)
> Test <- train(factor(Class) ~ Risk + H1 + H3 + H4, data = data,
  trControl = train_control, method = "nb")
> print(Test)
Naive Bayes
23 samples
 4 predictor
 2 classes: '0', '1'
No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 20, 22, 20, 21, 21, 21, ...
Resampling results across tuning parameters:
```

usekernel	Accuracy	Kappa
FALSE	0.6166667	0.025
TRUE	0.6333333	0.150

Результат перекрестной проверки классификатора Байеса на контрольной выборке:

```
> train_control <- trainControl(method = 'cv', number = 10)
> Test <- train(factor(Class) ~ Risk + H1 + H3 + H4, data = Tdata,
  trControl = train_control, method = "nb")
```

```
> print(Test)
```

Naive Bayes

18 samples

4 predictor

2 classes: '0', '1'

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 16, 15, 17, 16, 17, 17, ...

Resampling results across tuning parameters:

usekernel	Accuracy	Kappa
FALSE	0.6166667	-0.2666667
TRUE	0.8666667	0.5666667

## Приложение 2.

В этом приложении приведен результат классификации коммерческих банков с помощью разных моделей.

Таблица 10: Результаты классификации банков, закрытых в конце 2015 года

Название	Исходный класс	Логит-модель	Пробит-модель	Классификатор Байеса
ООО «Дил-Банк»	C1	C2	C2	C2
ООО «КБ „Ренессанс“»	C1	C1	C1	C2
ОАО «КБ „Максимум“»	C1	C1	C1	C1
АО «АКБ „Носта“»	C1	C1	C1	C1
ООО «КБ „Межрегион-банк“»	C1	C1	C1	C1
ООО «КБ „Регнум“»	C1	C1	C1	C1
ООО «Межтрастбанк»	C1	C2	C2	C2
ООО «КБ „Эргобанк“»	C1	C1	C1	C2
ОАО «Капитал банк»	C1	C1	C1	C1
ООО «КБ „Расчетный Дом“»	C1	C1	C1	C1
ООО «КБ „Милбанк“»	C1	C1	C1	C2
ОАО «АКБ „Акция“»	C1	C1	C1	C1
ООО «КБ „Росавтобанк“»	C1	C1	C1	C2
ПАО «Банк Екатеринбургский»	C1	C1	C1	C2
АО «СтарБанк»	C1	C1	C1	C1
КБ «Богородский»	C1	C1	C1	C2
АО «Смартбанк»	C1	C1	C1	C1
ПАО «УБРиР»	C2	C2	C2	C2
ПАО «Челябинвестбанк»	C2	C1	C1	C1
ПАО «Совкомбанк»	C2	C2	C2	C2

Название	Исходный класс	Логит-модель	Пробит-модель	Классификатор Байеса
АО «Тинькофф Банк»	C2	C2	C2	C2
АО «Райффайзенбанк»	C2	C1	C1	C2
ПАО «Банк Уралсиб»	C2	C2	C2	C2

Таблица 11: Результаты классификации банков, закрытых в конце 2016 года

Название	Исходный класс	Логит-модель	Пробит-модель	Классификатор Байеса
АО «Вологдабанк»	C1	C1	C1	C1
АКБ «Банк на Красных Воротах»	C1	C2	C2	C2
ПАО «АКБ „Новация“»	C1	C1	C1	C2
ПАО «АКБ „Енисей“»	C1	C1	C1	C2
АО «Торговый Городской Банк»	C1	C1	C1	C1
ООО «КБ„Финансовый капитал“»	C1	C1	C1	C1
ООО «КБ „Развитие“»	C1	C1	C1	C1
ООО «Центркомбанк»	C1	C1	C1	C1
ООО «Роспромбанк»	C1	C1	C1	C1
ПАО «Выборг-банк»	C1	C1	C1	C1
АО «Русский Трастовый Банк»	C1	C1	C1	C1
АО «Газстройбанк»	C1	C1	C1	C1
ПАО «УБРиР»	C2	C2	C2	C2
ПАО «Челябинвестбанк»	C2	C1	C1	C1
ПАО «Совкомбанк»	C2	C1	C1	C1
АО «Тинькофф Банк»	C2	C1	C1	C1
АО «Райффайзенбанк»	C2	C1	C1	C1
ПАО «Банк Уралсиб»	C2	C1	C1	C1

### Приложение 3.

В этом приложении представлены результаты тестов из Главы 3, выполненные в пакете R.

#### Тест Дики – Фуллера

Проверим стационарность ряда, описывающего доходность портфеля от кредитов банкам:

```
> adf.test(x,alternative = "stationary")
Augmented Dickey-Fuller Test
data:  x
Dickey-Fuller = -3.2487, Lag order = 3, p-value = 0.09525
alternative hypothesis: stationary

# для ряда первых разностей:
> adf.test(diff(x),alternative = "stationary")
Augmented Dickey-Fuller Test
data:  diff(x)
Dickey-Fuller = -4.3111, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Применим критерий Дики – Фуллера к ряду доходности от кредитов, выданных юридическим лицам:

```
> adf.test(y,alternative = "stationary")
Augmented Dickey-Fuller Test
data:  y
Dickey-Fuller = -0.67117, Lag order = 3, p-value = 0.9632
alternative hypothesis: stationary
```

```

> y1 <- diff(y, differences=1)      # Взятие первых разностей
> adf.test(y1, alternative = "stationary")
Augmented Dickey-Fuller Test
data:  y1
Dickey-Fuller = -2.1774, Lag order = 3, p-value = 0.5042
alternative hypothesis: stationary

```

```

> y2 <- diff(y, differences=2)      # Взятие вторых разностей
> adf.test(y2, alternative = "stationary")
Augmented Dickey-Fuller Test
data:  y2
Dickey-Fuller = -4.8569, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary

```

### **Критерий Льюнга–Бокса**

Проведем проверку остатков первой модели на случайность:

```

> Box.test(residuals(arima_X), type = c("Ljung-Box"))
Box-Ljung test
data:  residuals(arima_X)
X-squared = 0.037563, df = 1, p-value = 0.8463

```

Также проверим остатки второй модели с помощью теста Льюнга–Бокса:

```

> Box.test(residuals(arima_Y), type = c("Ljung-Box"))
Box-Ljung test
data:  residuals(arima_Y)
X-squared = 1.7508, df = 1, p-value = 0.1858

```



## Проверка нормальности остатков

Исследуем нормальность остатков первой модели, используя критерии Шапиро–Уилка и Колмогорова–Смирнова:

```
> shapiro.test(residuals(arima_X))  
Shapiro-Wilk normality test  
data: residuals(arima_X)  
W = 0.37288, p-value = 2.987e-11  
> n <- length(residuals(arima_X))  
> ks.test(residuals(arima_X), "pnorm", mean(residuals(arima_X)),  
          sd(residuals(arima_X))*sqrt((n-1)/n))  
One-sample Kolmogorov-Smirnov test  
data: residuals(arima_X)  
D = 0.37728, p-value = 4.106e-05  
alternative hypothesis: two-sided
```

Проверим, подчиняются ли остатки второй модели нормальному распределению:

```
> shapiro.test(residuals(arima_Y))  
Shapiro-Wilk normality test  
data: residuals(arima_Y)  
W = 0.9561, p-value = 0.1625  
> n <- length(residuals(arima_Y))  
> ks.test(residuals(arima_Y), "pnorm", mean(residuals(arima_Y)),  
          sd(residuals(arima_Y))*sqrt((n-1)/n))  
One-sample Kolmogorov-Smirnov test  
data: residuals(arima_Y)  
D = 0.096252, p-value = 0.861  
alternative hypothesis: two-sided
```