

ПЕРМСКИЙ ФИЛИАЛ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО  
АВТОНОМНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»  
Факультет экономики, менеджмента и бизнес-информатики

Ширяева Марина Андреевна

**ПРОГНОЗИРОВАНИЕ ОБЪЕМОВ ПРОСРОЧЕННОЙ ИПОТЕЧНОЙ  
ЗАДОЛЖЕННОСТИ С ИСПОЛЬЗОВАНИЕМ ИНДЕКСА РИСКА  
ИПОТЕЧНОГО ДЕФОЛТА**

Выпускная квалификационная работа - БАКАЛАВРСКАЯ РАБОТА  
студента образовательной программы бакалавриата «Экономика»  
по направлению подготовки 38.03.01 Экономика

Руководитель

к.э.н.

Д.Б. Потапов

Консультант

к.э.н.

А.М. Лозинская

Пермь, 2020 год

## Оглавление

Аннотация .....	3
Введение .....	5
Обзор литературы.....	10
Постановка исследовательского вопроса.....	27
Методология исследования .....	30
Данные и их предварительный анализ .....	38
Эмпирические результаты .....	45
Заключение.....	53
Список использованной литературы .....	57
Приложение 1. Динамика задолженности по ИЖК .....	62
Приложение 2. Графики динамики контрольных переменных и гистограммы их распределения.....	63
Приложение 3. Корреляционная матрица парных коэффициентов корреляции Пирсона.....	66
Приложение 4. Результаты ADF теста в процедуре Доладо-Дженкинсона-Сосвилла- Ривьеро.....	67
Приложение 5. Результаты оценивания модели ARIMAX без включения индекса риска ипотечного дефолта для объемов просроченной ипотечной задолженности .....	68
Приложение 6. Прогнозные значения первой разности объемов просроченной ипотечной задолженности и границы 95% доверительного интервала .....	69

## Аннотация

Показатель просроченной задолженности является основным показателем качества ипотечного портфеля российских банков. Для поддержания качества ипотечного портфеля на высоком уровне банкам необходимо прогнозировать данный показатель. В настоящей работе предпринята попытка прогнозировать объемы просроченной ипотечной задолженности по срокам задержки платежей до 30, до 90, до 180 и свыше 180 дней, с использованием индекса риска ипотечного дефолта (ИРИД). Идея расчета индекса была впервые предложена в работе (Chauvet et. al. 2016). В данной работе ИРИД строится в режиме реального времени на основе статистики о частоте поисковых запросов Google, исходящих от российских пользователей, и с учетом особенностей того, как они формулируют поисковые запросы. На агрегированных месячных данных о российских макроэкономических показателях за период 2009-2019 гг. было показано, что авторегрессионные модели интегрированного скользящего среднего с включением ИРИД имеют большую прогнозную силу, чем аналогичные модели без включения ИРИД. В качестве метрики прогнозного качества использовалась среднеквадратичная ошибка прогноза, значение которой сокращалось в среднем на 0,12% при включении ИРИД в модель.

## Abstract

The overdue debt indicator is the main indicator of the mortgage loan portfolio's quality of Russian banks. Banks need to forecast this indicator to maintain the mortgage loan portfolio's quality at a high level. This study is contributed to predict the volume of overdue mortgage debt with 30, 90, 180 and over 180-day delay in payments using the mortgage default risk index (MDRI). The idea of the MDRI calculation was first proposed by Chauvet et. al. (2016). In this paper, the MDRI was built in real time on the basis of statistics available on the frequency of Google search queries. When selecting search queries for MDRI we took into account the characteristic property of search queries formulation by Russian. The analysis is based on the monthly aggregated data on Russian macroeconomic indicators in 2009-2019. It shown that autoregressive models of the integrated moving average with the MDRI have a greater predictive power than similar models without the MDRI. The root mean squared error (RMSE) of the forecast was used as a metric for predictive quality. The RMSE decreased by an average of 0.12% when MDRI was included in the model.

## Введение

Рынок ипотечного жилищного кредитования (ИЖК) в России достаточно молод, ему всего 15 лет, и он сильно отличается от западных аналогов. Во-первых, в странах запада ипотека оказывает большее влияние на экономику страны, например, доля ипотеки в ВВП США составляет 65%, в то время как в России – только 6%. Во-вторых, на западе широко распространена высокорисковая (subprime) ипотека, в России выдача такой ипотеки минимальна. В-третьих, в странах запада осуществляется секьюритизация высокорисковых ипотечных кредитов, тогда как в России секьюритизация запрещена на законодательном уровне. Наконец, более половины всех кредитов на западе выдается с переменной процентной ставкой, в России, по большей степени, ставка фиксирована. Тем не менее, рынок ИЖК России укрепился и достиг значительного прогресса. Создана вся ключевая инфраструктура, выросли объемы кредитования, увеличилось число заемщиков. На сегодняшний день на рынке ИЖК также произошел важный переход от ориентации кредитных организаций на увеличение количества выданных ипотечных ссуд к борьбе за их качество.

Одним из основных показателей, отражающих качество ипотечного портфеля кредитной организации, является показатель просроченной задолженности, так как ипотечные ссудные операции сопровождаются кредитным риском, то есть риском невозврата долга или неисполнения обязательств заемщиком (Черникова, Щербаков Евстефеева, 2016). Чем ниже доля просроченной ипотечной задолженности в портфеле кредитной организации, тем выше его качество, из чего следует, что кредитная организация становится более устойчивой к воздействию внешних негативных экономических факторов. Проблема просроченной ипотечной задолженности также возрастает в связи с переходом и необходимостью исполнения банками положений Базельского соглашения III о новых

требованиях к оценке достаточности капитала и новых обязательных нормативов ликвидности банка (Зобова, Самойлова, 2012).

Согласно данным Центрального Банка Российской Федерации (ЦБ РФ) по состоянию на 1 января 2019 г. размер просроченных задолженностей в кредитных банковских портфелях находится в пределах от 0,35 до 3,53%. За истекший год наблюдается рост совокупной задолженности по ИЖК. Так, на 1 января 2019 г. величина задолженности по ИЖК в рублях по сравнению с 1 января 2018 г. увеличилась на 23,9%, составив 6 376,8 млрд рублей (Приложение 1). Очевидно, что столь существенный рост просроченной задолженности за 2019 г. обусловлен, прежде всего, текущей экономической ситуацией, в частности, снижением цен на нефть на рынке сырья и обвалом рубля, что непосредственно оказывает влияние на исполнение заемщиками своих обязательств. Кроме того, условия ипотечного кредитования остаются достаточно обременительными для большинства заемщиков, поскольку наблюдаются:

- 1) высокий первоначальный взнос по ипотечному кредиту (минимум 15-20% от стоимости недвижимости);
- 2) значительный размер обязательного платежа по отношению к доходам заемщика (в среднем около 40-50 ежемесячных окладов заемщика);
- 3) регулярность платежей (уплату процентов и погашение основного долга требуется делать регулярно, так как отсрочки, приводящие к капитализации в силу значительности базиса, делают возрастание долга стремительным) (Языков, 2011).

В связи с этим и в силу определённых обстоятельств, заемщики перестают выполнять свои обязательства вовремя и в полном объеме: задерживают платеж на один или несколько месяцев или вовсе оказываются не в состоянии совершить выплаты по ипотечным ссудам и объявляют дефолт.

Таким образом, для банковских организаций остается актуальным вопрос прогнозирования объемов просроченной задолженности по ИЖК.

Данный прогноз позволит кредиторам определить минимально допустимый уровень резервов на возможные потери по ссудам и сформировать комплекс мер по минимизации реальных потерь банков, возникающих при росте просроченной задолженности. Повышение точности прогнозов и увеличение предупреждающих интервалов для принятия решений важны для успеха политики управления кредитными рисками банка.

Изучение деталей возникновения убытков при задолженностях и/или дефолтах заемщиков по ИЖК вызывает определенные затруднения, поскольку официальная статистическая информация по данному вопросу, сбор, обработка и ее публикация сопряжены со значительными временными лагами. Преодолеть эти сложности можно, используя набирающую популярность и активно развивающуюся методологию построения опережающих индикаторов отдельных видов рисков, в частности, кредитных, ликвидности, валютных и процентных рисков. Назначение данных индикаторов – раннее оповещение, которое позволит заблаговременно подготовиться к возникновению этих рисков и смягчить их последствия (Солнцев и др., 2011). Одним из примеров такого индикатора является индекс риска ипотечного дефолта (ИРИД или Mortgage Default Risk Index, MDRI), методология построения которого была предложена в работе М. Шове и др. (2016). Данный индекс представляет собой прокси-переменную для измерения риска наступления ипотечного дефолта заемщика. Методология построения ИРИД основана на анализе динамики специфических запросов интернет-пользователей в поисковой системе Google, что делает его уникальным среди других показателей рынка ИЖК. ИРИД неплохо объясняет динамику индексов цен на жилье и потери права выкупа заложенного имущества, а также обладает довольно высокой прогнозной силой, то есть улучшает прогнозы объемов просроченной задолженности по ипотечному кредиту (Chauvet et. al., 2016). Тем не менее, попытки исчислить аналогичный индекс для российской экономики не предпринимались.

Таким образом, цель данной работы – прогнозирование объемов просроченной ипотечной задолженности с использованием индекса риска ипотечного дефолта. Стоит отметить, что ИРИД строился по аналогии с работой (Chauvet et. al., 2016), но с использованием российских данных и был частично модифицирован в связи с тем, что российские пользователи иначе формулируют свои поисковые запросы.

Для достижения поставленной цели были сформулированы следующие задачи исследования:

- изучение литературы, относящейся к моделированию прогнозных моделей объемов задолженности по ипотеке, а также к использованию запросов в системе Google как инструмента текущей оценки и прогноза рыночной ситуации;
- обоснование используемой методологии, в частности апробация методологии построения ИРИД, освещенной в работе М. Шове и др. (2016), для российского рынка ИЖК, сбор, анализ и преобразование к стационарному виду макроэкономических показателей российского рынка ИЖК;
- построение авторегрессионной модели прогноза объемов задолженности по ИЖК без включения ИРИД в качестве независимой переменной и с включением, и сопоставление прогнозного качества соответствующих моделей;
- представление выводов, рассмотрение ограничений и направлений дальнейшей работы.

Практическая значимость работы заключается в том, что индекс риска ипотечного дефолта может быть использован как механизм раннего выявления возникновения просроченной или проблемной задолженности по ипотечному кредитованию в будущем. Более того, прогностическая модель, представленная в данном исследовании, может помочь кредитным учреждениям, банковским аудиторским компаниям, регулирующим органам и инвестиционным фирмам в прогнозировании объема будущих



просроченных платежей по ипотеке на всем рынке или в рамках конкретных портфелей. Кроме того, компетентные инвесторы в недвижимость, которые работают в близлежащих районах, могут получить выгодную информацию о тенденциях таких просроченных платежей. Таким образом, данное исследование может получить должное внимание со стороны всех участников рынка ИЖК.

Структура выпускной квалификационной работы. В первой части исследования представлен обзор литературы и рассмотрен исследовательский вопрос, его теоретическое обоснование. Во второй части обоснована методология исследования, подробно охарактеризованы используемые данные и проведен их предварительный анализ. В следующей части детально описаны эмпирические результаты. В последней части сформулированы основные результаты и выводы приведенного исследования, ограничения и направления дальнейшей работы.

Объем работы: 70 страниц. Количество использованных источников литературы – 65.

## Обзор литературы

Просроченная задолженность представляет собой задолженность по кредитам с истекшим, согласно условиям договора, сроком погашения (Статистический сборник «Сведения о рынке жилищного (ипотечного жилищного) кредитования в России №5», 2018). Однако в банковской практике понятия «просроченная задолженность» и «проблемная задолженность» тождественны друг с другом и чаще всего используются как взаимозаменяемые. Следует отметить, что между просроченной и проблемной задолженностью существуют различия. Понятие проблемных ссуд можно найти в Положении Банка России от 28 июня 2017 г. № 590-П «О порядке формирования кредитными организациями резервов на возможные потери по ссудам, ссудной и приравненной к ней задолженности» (далее – Положение № 590-П). Согласно п. 1.7 Положения № 590-П в целях определения размера расчетного резерва в связи с действием факторов кредитного риска ссуды классифицируются на основании профессионального суждения в одну из пяти категорий качества:

- I (высшая) категория качества (стандартные ссуды): отсутствие кредитного риска (вероятность финансовых потерь вследствие неисполнения либо ненадлежащего исполнения заемщиком обязательств по ссуде равна нулю);
- II категория качества (нестандартные ссуды): умеренный кредитный риск (вероятность финансовых потерь вследствие неисполнения либо ненадлежащего исполнения заемщиком обязательств по ссуде обуславливает ее обесценение в размере от 1 процента до 20 процентов);
- III категория качества (сомнительные ссуды): значительный кредитный риск (вероятность финансовых потерь вследствие неисполнения либо ненадлежащего исполнения заемщиком

обязательств по ссуде обуславливает ее обесценение в размере от 21 процента до 50 процентов);

- IV категория качества (проблемные ссуды): высокий кредитный риск (вероятность финансовых потерь вследствие неисполнения либо ненадлежащего исполнения заемщиком обязательств по ссуде обуславливает ее обесценение в размере от 51 процента до 100 процентов);
- V (низшая) категория качества (безнадежные ссуды): отсутствует вероятность возврата ссуды в силу неспособности или отказа заемщика выполнять обязательства по ссуде, что обуславливает полное (в размере 100 процентов) обесценение ссуды.

Как следует из смысла Положения № 590-П просроченными могут быть ссуды II-IV категорий качества. Тогда как проблемными могут называться ссуды с высоким кредитным риском, то есть ссуды IV категории качества. Таким образом, понятие «проблемная задолженность» в строгом смысле является более узким, чем понятие «просроченная задолженность».

Связанное с просроченной ипотечной задолженностью понятие – это ипотечный дефолт. Согласно п. 13.3. Положения Банка России от 6 августа 2015 г. N 483-П «О порядке расчета величины кредитного риска на основе внутренних рейтингов» дефолт заемщика – это ситуация невозможности погашения заемщиком своих обязательств. Отсюда следует, что просроченная ипотечная задолженность возникает в случае ипотечного дефолта заемщика.

Литература, посвященная проблемам возникновения просроченной задолженности в кредитных портфелях банков, широка и состоит из теоретических и эмпирических исследований. С теоретической точки зрения можно выделить две основные конкурирующие теории о причинах возникновения просроченной задолженности по ипотечным кредитам (Aristei, Gallo, 2016). Первая, называемая «теорией справедливости», предполагает, что стратегическое решение заемщика

о неплатежеспособности основывается на рациональной оценке затрат и выгод от продолжения (или прекращения) погашения ипотеки. Согласно этой теории, заемщики перестают выполнять свои обязательства в тот момент, когда стоимость приобретаемого жилья меньше стоимости полученного под него ипотечного кредита. Вторая теория называется «теория платежеспособности» и предполагает, что заемщики будут воздерживаться от допущения долгов, если они имеют достаточные доходы для выполнения периодических платежей без чрезмерного финансового бремени. В рамках этой теории, решение заемщика не является стратегическим, а определяется неблагоприятными событиями, такими как потеря работы, проблемы со здоровьем или негативные шоки дохода, которые вызывают трудности с погашением задолженности (Aristei, Gallo, 2016). Однако существуют теоретические работы, которые сфокусированы на теории «двойного триггера», или другими словами, на теории «двойного воздействия» на возникновение ипотечной задолженности и дефолта (Foote et al., 2008). Согласно данному исследованию, теорий «справедливости» и «платежеспособности» недостаточно для полного объяснения дефолта по ипотечным кредитам, а сама теория «двойного воздействия» является результатом сочетания отрицательной разности между стоимостью обеспечения и ипотечной кредитной задолженностью и неблагоприятных экономических потрясений (Bhutta et al., 2010).

В дополнение к научной литературе также изучаются факторы, которые влияют на возникновение ситуации неплатежей, которые ведут к возникновению просроченной задолженности/вероятности возникновения ипотечного дефолта при взятии ипотечного кредита. Чаще всего выделяют три основные группы факторов:

- 1) социально-демографические характеристики заемщика;
- 2) параметры ипотечного займа;
- 3) макроэкономические факторы.

В первую очередь, рассмотрим социально-демографические характеристики заемщика. Авторы в своих работах концентрируются на возрасте, социальном статусе, месте работы и уровне доходов заемщика. Так, например, в работе У. Лин и др. (2015) поясняется, что из-за различий в социальном и экономическом положении семьи иммигрантов могут недостаточно интегрироваться в принимающее общество и, следовательно, с большей вероятностью будут являться должниками по ипотеке, чем идентичные жители местного населения. Д. Аристей и М. Галло (2016), в свою очередь, заключили, что домохозяйства, возглавляемые молодыми, безработными или иммигрантами, имеют более высокую вероятность возникновения задолженности и в большей степени страдают от неблагоприятных экономических условий, связанных с кризисом. Л. Диас-Серрано (2005) указывает, что волатильность доходов заемщика увеличивает риск неплатежеспособности ипотеки, в то время как Р. Гомес-Сальвадор и др. (2011) обнаружили, что домохозяйства в группах с более низким доходом, глава которых безработный или иммигрант, с большей вероятностью попадут в задолженность.

Условия кредитных договоров также могут быть фактором, способствующим ипотечным дефолтам. Эти условия, как правило, включают соотношение кредита к стоимости, сроки кредитования, суммы кредита, место расположения и сроки эксплуатации приобретаемого жилья. Опять обратимся к работе Д. Аристей и М. Галло (2016), в которой обнаружена положительная и статистически значимая связь между изменениями параметров кредита, особенно теми, которые связаны с изменением процентной ставки, и последующей вероятностью возникновения дополнительных трудностей с погашением ипотечной ссуды. Более того, А. Фустер, П. Виллен (2017) утверждают, что размер платежа оказывает экономически значимое влияние на поведение заемщика при погашении, например, снижение процентной ставки на 3%, которое согласно выборке их исследования примерно соответствует сокращению ежемесячного платежа в

два раза и по оценкам, уменьшает ежемесячную вероятность возникновения просроченной задолженности по ипотеке примерно на 55%.

Третий важный тип факторов – это макроэкономические условия. Очевидно, что банки могут не предотвратить увеличение вероятности возникновения просроченной задолженности из-за изменений в общей экономической ситуации. В частности, экономический спад приводит к безработице среди заемщиков, снижению цен на обеспечение кредита и повышению процентных ставок по кредитам, что приводит к неспособности заемщиков выплачивать проценты (Mondria et. al., 2017). Некоторые исследования в области ипотечного жилищного кредитования сосредоточены на влиянии такого макроэкономического фактора, как безработица. К. Герарди и др. (2013) количественно оценили влияние, которое потеря занятости оказывает на ипотечный дефолт, и пришли к выводу, что в среднем и при прочих равных безработица увеличивает вероятность дефолта на 5-13 процентных пунктов по сравнению с выборочной средней нормой дефолта 3,9%. Вышеуказанный вывод подтвердили в своей работе С. Мосетти, Е. Вивиано (2016), исследовав влияние резкого сокращения уровня дохода заемщика, вызванных потерей работы, на просроченные платежи по ипотеке на финансовом рынке Италии. Авторы обнаружили, что безработица более чем в два раза увеличивает вероятность возникновения у заемщиков задержек и невыплат банку ипотечного займа. Результаты исследования также свидетельствуют о том, что инструменты, предназначенные для смягчения последствий шоковых потрясений доходов, такие как, например, пособие по безработице, снижение налогов и в целом любая политика, которая ослабляет финансовые ограничения домашних хозяйств, могут оказывать положительное влияние не только на благосостояние домашних хозяйств, но и на финансовую стабильность, поскольку они снижают вероятность дефолта.

Не менее важным является международно-правовой статус страны, в которой осуществляется ипотечная сделка. Б. Дуйган-Бамп и С. Грант

(2009) показывают, что задолженность имеет важные долгосрочные последствия для домохозяйств и что различия в финансовых и судебных учреждениях в разных странах, отражающие различия в способах наказания за дефолт по ссуде, существенно влияют на поведение по погашению задолженности. Схожим образом Т. Джаппелли и др. (2013) показывают, что задолженность домохозяйств более чувствительна к неблагоприятным потрясениям, когда домохозяйства имеют большую задолженность, и подчеркивают, как институциональные факторы, такие как обеспечение соблюдения договорных прав, эффективность судебной системы и банкротство, оказывают существенное влияние на финансовую неустойчивость домохозяйств. Заметим, что все обозначенные выше авторы использовали множественную регрессию при проведении исследований.

Помимо установления причин и факторов, влияющих на вероятность возникновения просроченной задолженности по ИЖК, кредиторам также необходимо иметь возможность прогнозировать эту вероятность. Однако для этого необходимы индивидуальные детализированные данные, которые часто оказываются изменчивыми и конфиденциальными, что делает процесс сбора соответствующей информации сложней. Тем не менее, в открытом доступе находятся агрегированные данные, доступные в реальном режиме времени, которые, в свою очередь, можно использовать для прогнозирования самих объемов просроченной ипотечной задолженности. Это также актуально и важно для кредиторов, поскольку соответствующие прогнозы помогут заблаговременно подготовиться к наступлению кредитного риска и предпринять меры по сокращению общих потерь от этого риска, в частности, от роста объемов просроченной задолженности. Данной проблеме уделено должное внимание в предшествующих эмпирических работах. Более того, группы факторов, влияющие на объемы просроченной задолженности по ИЖК аналогичны тем, которые влияют на вероятность ее возникновения. Рассмотрим подробнее эти работы и факторы.

В работе (Crook, Banasik, 2012) авторы построили прогноз объемов просроченной ипотечной задолженности на рынке США на 4 квартала вперед, с использованием регрессионных моделей временных рядов. Для этого авторы использовали агрегированные квартальные данные об объемах просроченной задолженности по ипотеке заемщиков США за двадцатилетний период вплоть до 2008 г. и далее сравнивали точность прогнозов, полученных на основе краткосрочной динамической модели исправления ошибок (Error Correction Model, ECM) с данными, полученными с помощью интегрированной модели авторегрессии скользящего среднего (Autoregressive Integrated Moving Average Model, ARIMA). В качестве метрики прогнозного качества они использовали среднеквадратичную ошибку модели (Root Mean Square Error, RMSE) и получили следующие результаты: 5244 для модели ECM и 5137 для модели ARIMA. Таким образом, авторами было обнаружено, что точность прогноза обеих моделей схожа друг с другом.

В свою очередь в работе (Lin, Yang, 2003) разрабатывается уникальная модель прогнозирования с использованием нестационарной цепи Маркова, которая способна предсказать вероятность возникновения просроченных платежей по ипотеке. Данные по ипотечному жилищному кредитованию, полученные крупным тайваньским финансовым учреждением с 1 января 1996 г. по 30 июня 1998 г. на месячной частоте, используются для изучения эффективности прогноза предложенной авторами модели с использованием цепи Маркова и модели ARIMA, путем сравнения величины среднего абсолютного отклонения (Mean Absolute Deviation, MAD). В моделях в качестве зависимой переменной выступали объемы просроченных задолженностей с разными сроками задержки платежей: до 30, до 60, до 90 и свыше 90 дней. В качестве контрольных переменных выступали факторы, относящиеся к социально-демографическим характеристикам заемщика (доход, образование, пол, стаж работы) и к параметрам ипотечного займа (соотношение кредита к стоимости, сумму и источник кредита). В результате



получилось, что значение MAD для моделей с цепями Маркова меньше, чем для моделей ARIMA, что говорит о лучшей точности и качестве прогноза разработанной авторами модели (0,71% против 0,88% для прогноза объемов просроченной 30-дневной задолженности, 0,54% против 0,51% для прогноза объемов просроченной 60-дневной задолженности, 0,19% против 0,20% для прогноза объемов просроченной 90-дневной задолженности и, наконец, 0,14% против 0,16% для прогноза объемов просроченной задолженности по срокам задержки платежей свыше 90 дней).

Тем не менее, не всегда модели прогнозирования временных рядов обладают высокой точностью. Так, например, П. Миранда и П. МакГрат (2011) построили векторную авторегрессионную модель (Vector Autoregression model, VAR) для прогнозирования изменения уровня просроченной задолженности по ипотечным кредитам (по срокам задержки платежей от 30 до 59 дней, от 60 до 89 дней и т.д.) в США за период с 1987 по 2010 гг. (на ежеквартальной частоте). В модели VAR в качестве предикторов изменения просроченной задолженности по ипотечным кредитам использовались агрегированные показатели, относящиеся к макроэкономическим факторам (индекс цен на жилье Кейса-Шиллера, уровень безработицы США, доходность индекса S&P500), к социально-демографическим характеристикам заемщика (процентное изменение личных сбережений и доходов населения) и к параметрам ипотечного кредита (процентные ставки по 30-летним ипотекам). Все вышеуказанные переменные были преобразованы в стационарный вид путем взятия первых разностей. Авторы построили соответствующие прогнозы на трех периодах: 2005Q1-2008Q2, 2008Q1-2009Q4, 2001Q1-2003Q4. Чтобы проверить надежность результатов прогнозирования с помощью модели VAR, рассчитывалась сумма квадратов ошибок (Sum of Squared Errors, SSE) как сумма квадратов разности между прогнозируемым и фактическим изменением уровня просроченной задолженности по ипотечным кредитам. По результатам исследования SSE для первого периода (2005Q1 – 2008Q2),

который совпадает с так называемым ипотечным кризисом США, равняется 0,74, что приблизительно равно SSE для периода стабильности на рынке ИЖК в США (2001Q1-2003Q4) – 0,71. Однако, SSE для оставшегося посткризисного периода равняется 3,01. Таким образом, получившиеся результаты показывают, что период времени с первого квартала 2008 г. по четвертый квартал 2009 г. действительно является экстраординарным периодом – статистическим выбросом. Под экстраординарным здесь авторы подразумевают, что переменные, которые достаточно хорошо предсказывают фактические сдвиги в уровнях просроченной задолженности по ипотечным кредитам за период с 1987 по 2006 гг., имеют тенденцию к занижению с 2007 г. Таким образом, данная модель не может предвидеть всплеск задолженности по ИЖК среди заемщиков.

Поэтому все чаще исследователи прибегают к поиску альтернативных методов прогнозирования, одним из которых является метод машинного обучения. А. Хандани и др. (2010) использовали метод машинного обучения для построения нелинейных непараметрических моделей прогнозирования объемов просроченных ипотечных платежей. В работе были использованы индивидуальные месячные данные о банковских транзакциях, кредитной истории и балансе счета отдельных заемщиков, полученные из клиентской базы крупного коммерческого банка США за период с января 2005 г. по апрель 2009 г. Авторы утверждают, что их прогнозы на вневыборочном периоде сильно коррелируют с реализованными просрочками по ипотечным кредитам, причем коэффициент детерминации линейной модели составляет 85% для ежемесячных прогнозов на 6- и 12-месячных горизонтах. Кроме того, практическая ценность данного прогноза была оценена авторами путем суммирования сэкономленных средств от сокращения выдачи ипотечных кредитов высокорисковым заемщикам: с учетом всех допущений потенциальная чистая выгода от данных прогнозов составляет 6-25% от общих потерь. Учитывая масштаб общепромышленных транзакций

в течение 2008 г., даже 6% экономии затрат составили бы сотни миллионов долларов.

Схожим образом в работе (Parnes, 2018) был применен Байесовский метод машинного обучения для прогнозирования уровня просроченной задолженности по различным срокам просрочки платежей до 30, до 60 и до 90 дней на рынке США. В исследовании были использованы абсолютные квартальные показатели о совокупных объемах просроченных ипотечных платежей в США, полученные от Ассоциации ипотечных банкиров (Mortgage Bankers Association's, MBA), в период с 2004 г. по 2010 г. Д. Парнес (2017) использовал все имеющиеся у него исторические данные для построения соответствующего прогноза, а для проверки устойчивости и надежности полученных результатов построил этот же прогноз с использованием метода «скользящего окна» (moving window) размером в четыре последовательных квартала и получил сопоставимые результаты. В общем и целом в исследовании тестировалась гипотеза о том, что аномальные показатели просроченной задолженности по ипотечным кредитам могут быть выявлены еще до начала полномасштабного жилищного кризиса (например, в 2009-2010 гг.), данная гипотеза подтвердилась в ходе исследования. Таким образом, авторы резюмируют, что аномальные показатели просроченной задолженности могут служить ранними признаками жилищного кризиса.

С одной стороны, применение методов машинного обучения при прогнозировании объемов просроченной ипотечной задолженности сопряжено с рядом трудностей: отсутствие прозрачности, жесткие требования к обучающей выборке, сложность интерпретации результатов. С другой стороны, модели прогнозирования временных рядов просты при моделировании, обладают единообразием подходов и неплохой прогнозной силой, что делает эти модели более предпочтительными в использовании.

В работе (Chauvet et. al., 2016) также были использованы модели временных рядов для прогнозирования объемов просроченной ипотечной

задолженности. Более того, в работе был выдел еще один фактор, влияющий на объемы просроченной ипотечной задолженности – это частота поисковых интернет-запросов пользователей в Google. Количественно оценить объемы и частоту поисковых запросов помогает сервис статистики Google Trends<sup>1</sup>. Данный сервис визуализирует динамику запросов, путем построения линейного графика временного ряда со временем по горизонтальной оси и частотой поиска по вертикальной оси. Данные временных рядов предлагаются с дневной, недельной и месячной периодичностью, начиная с 2004 г., и могут быть загружены в формате csv.

М. Шове и др. (2016) разработали индекс риска ипотечного дефолта (ИРИД), который представляет собой прокси-переменную риска ипотечного дефолта заемщика и характеризует изменение во времени общего настроения заемщиков на рынке ипотечного кредитования. Строится ИРИД на основе анализа динамики таких запросов американских интернет-пользователей, как «государственная помощь по ипотеке», «лишение права выкупа заложенного имущества» и др., с использованием Google Trends на ежедневной, еженедельной и ежемесячной частоте с 2004 г. по 2015 г. Авторами было обнаружено, что ИРИД предсказывает просроченную задолженность в срок до 60 дней, что является распространенным показателем дефолта в жилищной отрасли США. Для этого были построены модели прогностической регрессии с зависимыми переменными в виде индексов просроченной ипотечной задолженности в срок до 30, 60 и до 90 дней и индекс потери права выкупа заложенного имущества (Mortgage Foreclosure Index) от Bloomberg по всем основным и субстандартным (высокорисковым) ипотечным кредитам на национальном уровне в течении месячной периодичности. ИРИД тесно коррелирует со всеми видами просроченной задолженности, но наиболее тесно связана с просроченной задолженностью в срок до 60 дней. Действительно, корреляция между ИРИД и показателями 60-дневной просрочки колеблется от 0,72 до 0,86. Эти результаты

---

<sup>1</sup> <https://trends.google.com/trends/>

согласуются с ожиданиями авторов, поскольку ипотечные кредиторы, как правило, стремятся решить проблему просроченной ипотечной задолженности или инициировать процедуру обращения взыскания после 60 дней. Кроме того, просрочка свыше 60 дней считается дефолтом в большей части жилищной литературы США (An et al., 2016). В подтверждение вышесказанного, также стоит отметить, что увеличение ИРИД на одно стандартное отклонение приводит к увеличению стандартного отклонения на 0,21 для 60-дневных просрочек по обычному ипотечному кредиту и увеличению стандартного отклонения на 0,24 для 60-дневных просрочек для субстандартной (высокорисковой) ипотеки.

М. Шове и др. (2016) проверили способность ИРИД улучшать качество прогнозов ранее упомянутых показателей рынка ИЖК. Для этого сначала строились авторегрессионные (AR) модели для каждой переменной, а затем в эти модели добавляли авторегрессионные лаги и лаги ИРИД, и сопоставляли качество моделей с учетом ИРИД и без учета ИРИД относительно среднеквадратичной процентной ошибки (Mean Squared Percentage Error, MSPE). Значимое отличие между значениями MSPE наблюдалось для моделей прогнозирования 30-дневной просрочки (0,772 для модели с ИРИД против 1,579 для модели без ИРИД) и 60-дневной просрочки (0,526 для модели с ИРИД против 0,904 для модели без ИРИД). Таким образом, результаты свидетельствуют о том, что данный регрессор довольно хорошо предсказывает просрочку платежей по ипотеке в срок до 30 и 60 дней.

Далее авторами ИРИД используется в качестве опережающего индикатора кредитного риска для раннего обнаружения кризисных явлений на рынке ИЖК. Другими словами, на ежедневной и еженедельной частотах строятся модели прогностической регрессии, где в качестве зависимой переменной используются индексы цен на жилье, включая индекс

Кейса-Шиллера<sup>2</sup> (Case-Shiller index), индекс цен на жилье от Федерального агентства жилищного финансирования<sup>3</sup> (Federal Housing Finance Agency, FHFA) и еженедельный индекс цен на жилье от Е. Аненберг и С. Лауфер<sup>4</sup> (2014), а ИРИД выступает в качестве независимой переменной. Результаты оценивания данных моделей указывают на то, что увеличение ИРИД приводит к снижению доходности цен на жилье в соответствии с негативными внешними факторами, обусловленными просрочкой платежей или обращением взыскания долга по закладной.

Кроме того, в работе было обнаружено, что ИРИД является ведущим индикатором настроений на рынке жилья. Для этого была построена еще одна модель прогностической регрессии, но уже на ежемесячной частоте. Влияние ИРИД оценивалось на индекс настроений потребителей Мичиганского университета, который строится на анализе социальных опросов домохозяйств США о том, хорошее ли сейчас время для покупки жилья или плохое. В результате за увеличением значения показателя ИРИД следует увеличение числа ответов на заданный вопрос, в которых говорится, что сейчас неподходящее время для покупки жилья, потому что будущее неопределенно или потому, что семья не может позволить себе купить жилье. Таким образом, волатильность и изменения ИРИД выявляют ипотечный кризис до того, как аналогичная информация будет получена с помощью традиционных обследований и опросов.

Н. Аскитас и К. Циммерман (2014) схожим образом использовали данные из сервиса Google Trends о частоте поисковых запросов для выявления объемов просроченной ипотечной задолженности. Авторы оценили интенсивность поисков Google для поискового запроса «письмо о трудностях» («hardship letter»). Письмо о трудностях – это не что иное, как

---

<sup>2</sup> Индекс Кейса-Шиллера – это индекс цен на жилье по 20-ти крупнейшим городам США, ежеквартальная оценка изменения цен на отдельные односемейные дома в США.

<sup>3</sup> Индекс цен на жилье от FHFA – это средневзвешенный индекс повторных продаж, то есть он измеряет средние изменения цен при повторных продажах или рефинансировании на одни и те же объекты недвижимости, на основе данных, охватывающих все 50 штатов и более 400 американских городов.

<sup>4</sup> Индекс цен на жилье от Е. Аненберг и С. Лауфер – это индекс цен на дома с повторной продажей, который описывает стоимость домов на дату заключения контракта, когда определяется цена, а не на дату закрытия сделки, когда имущество передается.

письменный запрос на реструктуризацию или рефинансирование ипотечного кредита, который заемщик направляет в свой банк, как это общепринято в США. Поэтому, как утверждается в исследовании, спрос на консультацию о том, как писать данное письмо, должен быть ведущим показателем распространенности просроченной ипотечной задолженности. Поиск «письмо о трудностях» – это, пожалуй, самый ранний момент времени, когда заемщик сигнализирует о своем ухудшающемся финансовом состоянии, и Google Trends является первым инструментом для обнаружения этой тенденции. Представленная в работе прогнозная авторегрессионная модель (Hardship Letter Model, HL) в качестве зависимой переменной содержит совокупный объем просроченной задолженности на рынке ипотеки США, а в качестве регрессора – относительную интенсивность поиска запроса «письмо о трудностях» в системе Google (квартальные данные за период 2004-2013 гг.) В результате утверждается, что данный регрессор довольно хорошо предсказывает просрочку платежей по ипотеке.

Использование частоты поисковых запросов из сервиса Google Trends в прогностических моделях улучшает результаты прогноза не только объемов просроченной ипотечной задолженности, но и других макроэкономических переменных. Одними из первых это подтвердили Х. Чой и Х. Вэриан (2012). Авторы показали, что модель авторегрессии первого порядка с временными рядами Google Trends о запросах в отношении определенной марки автомобилей обладает лучшими прогностическими свойствами по сравнению с обычной авторегрессионной моделью первого порядка. Аналогично доказано, что информация, полученная с помощью статистики Google Trends, улучшает прогноз заявок по безработице, потребительской уверенности, туристической статистики (Choi, Varian, 2012). Со времени данных публикаций и по настоящее время научную литературу обогатили множество схожих работ, одни из последних представлены в табл. 1.

Использование статистики пользовательских запросов Google для  
прогнозирования экономических переменных

Автор (год публикации)	Анализируемая страна	Прогнозируемая переменная
Chauvet, Gabriel and Lutz, 2016	США	Объем просроченной ипотечной задолженности
Askitas, Zimmermann, 2014	Германия	Объем просроченной ипотечной задолженности
Veldhuizen, Vogt and Voogt, 2016	Дания	Количество сделок и транзакция на рынке жилья
Saxa, 2015	Чехия	Темпы роста выдачи ипотечных кредитов
Huang, Hui-Kuang Yu and Rodriguez-Garcia, 2019	Тайвань	Темпы роста выдачи ипотечных кредитов
Gotz and Knetsch (2019)	Германия	ВВП
Лазарян, Герман (2018)	Россия	ВВП
Smith (2016)	Великобритания	Уровень безработицы
D'Amuri and Marcucci (2017)	США	Уровень безработицы
Suhoj (2009)	Израиль	Уровень безработицы
Anvik, Gjelstad (2010)	Норвегия	Уровень безработицы
Борочкин (2013)	Россия	Уровень безработицы Чистый ввоз/вывоз капитала частным сектором
Li, Shang, Wang, and Ma (2015)	Китай	Индекс потребительских цен
Li, Ma, Wang, and Zhang (2015)	США	Цены на нефть
Li, Pan, Law, and Huang (2017)	Китай	Количество международных туристических прибытий в страну
Bangwayo-Skeete and Skeete (2015)	Страны Карибского бассейна	Количество международных туристических прибытий в страну
Siliverstovs and Wochner (2018)	Швейцария	Количество международных туристических прибытий в страну
Adachi, Masuda, and Takeda (2017)	Япония	Курс акций, биржевая прибыль
Tang and Zhu (2017)	США	Курс акций, биржевая прибыль
Yung and Nafar (2017)	США	Доходы от инвестиций в недвижимость (REITs)
Dimpfl and Jank (2016)	США	Волатильность фондового рынка
Moussa, Delhoumi, and Ouda (2017)	Франция	Волатильность фондового рынка
Peltomaki, Graham, Hasselgren (2018)	Страны с развивающейся рыночной экономикой	Волатильность обменного курса
Afkhami, Cormack, and Ghoddusi (2017)	США	Волатильность цен на энергоносители
Campos, Cortazar, Reyes (2017)	США	Волатильность цен на энергоносители



Таблица 1 (Продолжение)

Использование статистики пользовательских запросов Google для  
прогнозирования экономических переменных

Автор (год публикации)	Анализируемая страна	Прогнозируемая переменная
Столбов (2011)	Россия	Динамика депозитов физических лиц
Carriere-Swallow and Labbe (2013)	Чили	Продажа автомобилей
Barreira, Godinho, and Melo (2013)	Франция, Италия, Португалия, Испания	Продажа автомобилей

В рамках данной исследовательской работы наибольший интерес представляют работы, посвященные рынку ипотечного жилищного кредитования. Вклад в изучение взаимосвязи между поведением поиска в интернете и реальным поведением домохозяйств внесла работа С. Вельдхейзена и др. (2016). Поисковые данные Google об объемах веб-запросов слова «ипотека» используются авторами как прокси-переменная рыночной активности, а именно, как прокси-переменная количества сделок и транзакций на голландском рынке жилья. Основные выводы, к которым приходят авторы:

- поиск в Google слова «ипотека» в текущем и предшествующем месяце в значительной степени положительно связан с жилищными сделками;
- этот же поиск шесть и девять месяцев назад значительно положительно связан с транзакцией на рынке жилья текущего месяца;
- добавление данных поиска в Google в качестве прокси-переменной в модель прогнозирования объемов сделок на голландском рынке жилья увеличивает объяснительную силу этой модели на 4 процентных пункта.

Аналогичные результаты получились и в исследовании чешского и тайваньского жилищного рынка (Saha, 2015; Huarng et al., 2019, соответственно). В работах обнаружена тесная взаимосвязь темпов роста числа веб-поисков, связанных с ипотечным кредитованием, непосредственно, с темпом роста выданных ипотечных кредитов. Более того, выборочные краткосрочные прогнозы показывают, что объем поисков в Google улучшает краткосрочные прогнозы относительно ипотечного кредитования.

Таким образом, во всех вышеописанных исследованиях статистика частоты поисковых запросов из Google Trends используется в качестве инструмента прогнозирования различных показателей в различных областях экономики, начиная от рынка ИЖК и финансового рынка и заканчивая рынком туристических услуг. Однако, Э. Бурдо и Э. Кинцлер (2017) отмечают ограничения и проблемы использования статистики Google Trends на практике. Во-первых, в большинстве исследований не описано то, как выбираются «лучшие» показатели Google Trends. Во-вторых, не очевидно, что один или два термина могут быть лучшими прогностическими факторами для временного ряда, и даже если это так, то может оказаться, что эти термины будут отсутствовать в будущем, так как привычки веб-пользователей изменятся.

## Постановка исследовательского вопроса

На сегодняшний день проблема исполнения заемщиками своих обязательств по ИЖК, решается каждым банком самостоятельно, исходя из особенностей его устройства, величины активов, характеристик кредитного портфеля, количества проблемных кредитов и т.д. (Черникова, Щербаков Евстефеева, 2016). Традиционно банковские механизмы снижения риска возникновения просроченной или проблемной задолженности связаны с процедурами оценки платежеспособности потенциального клиента на этапе выдачи кредита, однако даже самый надежный андеррайтинг не может спрогнозировать будущее (Брызгалов и др., 2015). Обзор предшествующей литературы показал, что анализ запросов Google чрезвычайно актуален для исследований прогнозирования макроэкономических переменных (например, Smith, 2016; Yung, Nafar, 2017; Campos et. al., 2017; Gotz, Knetsch, 2019, и др.), в частности объемов просроченной задолженности по ипотечному кредитованию (например, Askitas, Zimmermann, 2014; Chauvet et.al., 2016).

Google является самой популярной поисковой системой в мире, через которую осуществляется около 70% всех интернет-запросов (Столбов, 2011). Согласно некоторым оценкам, количество поисковых запросов в Google достигло 2 трл. в 2016 г. (Jun et. al., 2018). Данная компания запустила общедоступный инструментарий анализа статистики пользовательских запросов – Google Trends, который показывает, как часто выполняется поиск определенного ключевого слова по отношению к общему количеству запросов в разных регионах мира и на разных языках. Информация о поведении интернет-пользователей, которые могут искать определенные поисковые элементы несколько раз в течение дня, позволяет судить о намерениях людей совершить те или иные действия, заранее выявить и спрогнозировать возможные значительные события

(Борочкин, 2013; Vuono et al., 2017). Соответственно, информация, полученная с ресурса Google Trends, используется преимущественно с целью краткосрочного прогнозирования различных показателей и для понимания текущей экономической активности и настроений.

Из вышесказанного следует, что перспективным представляется использование статистики запросов в Google Trends для отслеживания настроений заемщиков на рынке ИЖК. Однако, сегодня тема прогнозирования объемов просроченной ипотечной задолженности с использованием поисковых интернет-запросов, не освещена в российских исследованиях. Существующие работы по данной тематике опираются лишь на теоретические аспекты возникновения просроченной задолженности по ИЖК (Языков, 2011; Брызгалов и др., 2015, Черникова, Щербаков Евстефеева, 2016). В зарубежной литературе этот вопрос поднимается уже давно, в частности разработан индекс риска ипотечного дефолта на основе анализа динамики поисковых запросов пользователей, который используется в качестве инструмента прогнозирования ипотечной задолженности (Chauvet et. al., 2016).

Как утверждают сами авторы работы (Chauvet et. al., 2016), использование статистики пользовательских запросов в Google открывает новые возможности на рынке ипотечного жилищного кредитования и имеет ряд неоспоримых преимуществ. Во-первых, они являются чрезвычайно своевременными и охватывают потенциально обширную выборку респондентов, нежели результаты социологических опросов (например, индекс кредитного оптимизма (доверия) и т.п.). Во-вторых, продолжая сравнение с социологическим опросом, это также означает, что информация постоянно собирается по более широкому кругу вопросов, а не только по нескольким заранее определенным вопросам (McLaren, 2011). В-третьих, в отличие от официальной статистики они не пересматриваются, а значит, им можно больше доверять при построении прогнозов. Наконец, простота

сбора и предобработки этих данных позволяет без труда встроить их в используемые прогнозные модели.

Таким образом, в данной работе был поставлен следующий исследовательский вопрос: позволяет ли включение индекса риска ипотечного дефолта повысить точность прогнозирования объемов просроченной задолженности по ИЖК в реалиях российского рынка?

Для этого в работе рассчитывается ИРИД посредством апробации методологии, предложенной в статье М. Шове и др. (2016), но с использованием российских данных о частоте поисковых интернет-запросов, связанных с ипотекой. ИРИД также был модифицирован: в связи с тем, что российские пользователи иначе формулируют свои поисковые запросы в основу ИРИД легли только часто встречающиеся поисковые запросы российских пользователей, в частности, запросы сформулированные в виде вопроса, чего не было в работе (Chauvet et. al., 2016). Исходные данные разбиваются на обучающую (in-sample) и тестовую (out-of-sample) выборки в соотношении 90:10. Далее на обучающей выборке строятся авторегрессионные модели скользящего среднего (Autoregression Moving Average Model, ARMA) для объемов просроченной задолженности по ипотечному кредиту без включения ИРИД в качестве экзогенной переменной и аналогичная модель с включением ИРИД. После построения прогнозов объемов просроченной ипотечной задолженности на тестовой выборке, сопоставляется качество прогнозов для модели с учетом ИРИД и для модели без учета ИРИД по критерию квадратного корня из среднеквадратичной ошибки (Root Mean Square Error, RMSE).

## Методология исследования

На первом этапе исследования, как уже было сказано ранее, апробируется предложенная в статье М. Шове и др. (2016) методика построения индекса риска ипотечного дефолта с использованием данных о поисковых интернет-запросах от Google Trends, исходящих от российских пользователей сети Интернет. Google Trends отображает не абсолютные данные об объеме поиска, а относительную популярность поискового запроса в виде индекса объема поиска (Search Volume Index – SVI), а также корректирует данные об объеме поиска, чтобы в дальнейшем упростить сравнение между запросами. Каждая точка данных делится на общее количество поисковых запросов по географии и временному диапазону, которые она представляет, для сравнения относительной популярности. Другими словами, относительная популярность – это отношение объема поиска запроса к сумме объемов поиска всех возможных запросов. Полученные числа затем масштабируются в диапазоне от 0 до 100, где 100 баллов означают наивысший уровень популярности запроса, 50 – уровень популярности запроса, вдвое меньший по сравнению с первым случаем. Таким образом, Google Trends устраняет повторные поиски одного и того же пользователя в течение короткого периода времени, чтобы дать лучшую картину динамики популярности запроса.

На рис. 1 для демонстрации показан пример SVI для поискового запроса «ипотека», запрашиваемый российскими пользователями в сервисе Google в период с 2004 по 2020 гг. В этом случае, точка В представляет собой максимальное из таких соотношений с течением времени, и поэтому SVI за сентябрь 2010 г. равен 100, а SVI за февраль 2020 г. равен 68, что рассчитывается как отношение А к В, умноженное на 100. Ниже представлены формулы расчета значения SVI в точке А и в точке В (1) и (2):

$$A = \frac{SV_{t_1,l}(\text{"ипотека"})}{\sum_q SV_{t_1,l}(q)} \quad (1)$$

$$B = \frac{SV_{t_2,l}(\text{"ипотека"})}{\sum_q SV_{t_2,l}(q)} \quad (2)$$

где:

$A$  – значение индекса SVI (популярность) для поискового запроса «ипотека» в сентябре 2010 г.;

$B$  – значение индекса SVI (популярность) для поискового запроса «ипотека» в феврале 2020 г.;

$SV_{t_1,l}(q)$  – общее количество поисковых запросов в Google по запросу  $q$ , исходящих из местоположения  $l$  в момент времени  $t_1$ ;

$SV_{t_2,l}(q)$  – общее количество поисковых запросов в Google по запросу  $q$ , исходящих из местоположения  $l$  в момент времени  $t_2$ .

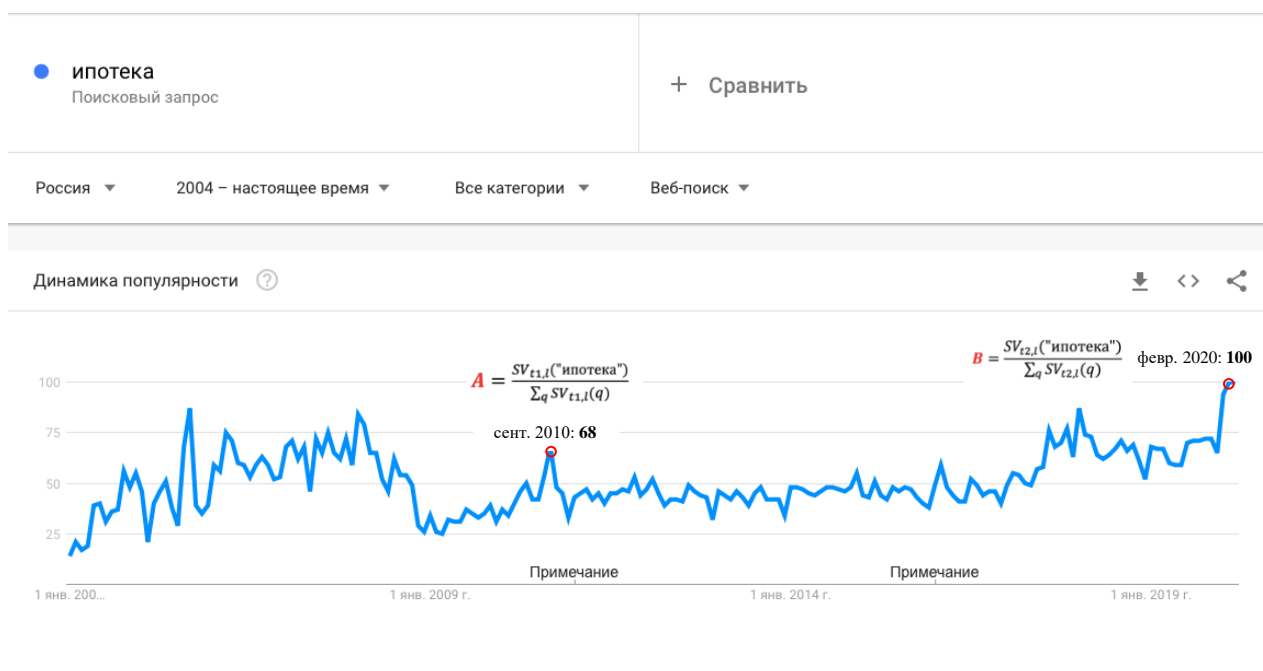


Рис. 1. Пример динамики популярности для поискового запроса «ипотека», полученной из сервиса Google Trends\*

\*Источник: веб-сайт статистики Google Trends (<https://trends.google.com/trends/>).

Чтобы построить индекс риска ипотечного дефолта, начать необходимо с рассмотрения ключевых слов, связанных с ипотекой, в сочетании со словом «помощь». При вводе в Google Trends слов «помощь в ипотеке» создается отчет, содержащий аналогичные запросы. Далее появляется возможность

составить список поисковых терминов из тех, которые выделены Google Trends, при условии, что они содержат ключевое слово «ипотека» и индикатор или сигнал бедствия. Этот процесс приводит к 12 ключевым поисковым запросам, представленным в табл. 2.

Таблица 2

Поисковые запросы Google Trends	
№	Поисковый интернет-запрос
1	помощь в ипотеке
2	помощь в погашении ипотеки
3	помощь в ипотеке от государства
4	помощь государства в погашении ипотеки
5	ипотека помощь
6	помощь по ипотеке
7	помощь государства в выплате ипотеки
8	долг по ипотеке
9	долги по ипотеке
10	нечем платить ипотеку
11	нечем платить ипотеку что делать
12	что делать если нечем платить ипотеку

Далее для каждого запроса из табл. 2 посредством Google Trends собираются значения SVI за доступный временной промежуток времени, то есть с января 2004 г. по март 2020 г. После находится среднее значение индекса SVI для каждого месяца, а затем нормализуется по следующей формуле (3):

$$X(i) = \frac{X(i)}{\max(X(i))} \times 100 \quad (3)$$

где:  $X(i)$  –  $i$ -ый элемент в векторе  $X$ .

Таким образом, значение  $X(i)$  будет представлять собой значение ИРИД в конкретном месяце. Согласно М. Шове и др. (2016), чем выше ИРИД, тем выше уровень обеспокоенности заемщика о своей избыточной долговой нагрузке. Следовательно, ИРИД способен сигнализировать о наступлении неблагоприятной ситуации на рынке ипотечного кредитования.



На втором этапе исследования исходная выборка данных была поделена на обучающую и тестовую выборки. Статистические тесты эффективности прогноза модели обычно проводятся путем разбиения заданного набора данных на период внутри выборки, используемый для начальной оценки параметров и выбора модели, и период вне выборки, используемый для оценки эффективности прогнозирования. Эмпирические данные, основанные на результатах прогнозов вне выборки, обычно считаются более надежными, чем данные, основанные на результатах внутри выборки, которые могут быть более чувствительными к выбросам. Более того, как правило, оценка качества модели, сделанная по результатам вне выборки, оказывается несмещенной (Hansen, Timmermann, 2012). Однако, в эмпирической литературе не существует общепринятых и универсальных правил о соразмерности обучающей и тестовой выборок. В большей части литературы исследователи определяют размер тестовой выборки как 10-15% от всего количества наблюдений. Таким образом, в данной работе внутривыборочный период представляет собой период с марта 2010 г. по апрель 2018 г. (98 месяцев, 90,7% от общего числа наблюдений), а вневыборочный – с мая 2018 г. по февраль 2019 г. (10 месяцев, 9,3% от общего числа наблюдений).

На третьем этапе исследования, чтобы изучить позволяет ли включение ИРИД в модель повысить прогнозную силу модели, строятся модели авторегрессии интегрированного скользящего среднего (Autoregression Integrated Moving Average eXtended, ARIMAX), которые объединяют в себе модель ARIMA с экзогенными переменными, с зависимой переменной отвечающей за объемы просроченных ипотечных платежей по срокам задержки платежей до 30, 90, 180 и свыше 180 дней. В настоящее время наибольшее применение в задачах прогнозирования экономических показателей получили методы авторегрессионного анализа временных рядов и методы машинного обучения. Однако поскольку метод машинного обучения обладает жесткими требованиями к обучающей

выборке, сложностью алгоритмов обучения сети и отсутствием прозрачности, то его применение для прогнозирования объемов просроченной задолженности по ИЖК становится затруднительным. В связи с этим наибольший интерес для прогнозирования объемов просроченной ипотечной задолженности представляет модель ARIMAX, которая в общем случае имеет вид (4):

$$\Delta y_t^s = \alpha + \sum_{i=1}^q \beta_i * \Delta y_{t-i}^s + \sum_{j=1}^m \delta_j * \varepsilon_{t-j} + \varepsilon_t + \sum_{l=1}^r \gamma_l * \Delta MDRI_{t-l} + \sum_{f=0}^k \omega_f * \Delta Controls_{t-f} \quad (4)$$

где:

$\Delta$  – оператор первой разности исследуемого временного ряда;

$\Delta y_t^s$  обозначает прогнозируемую зависимую переменную, т.е. задолженность по ипотечному жилищному кредиту (в млн руб.) по срокам задержки платежей в  $s$  дней – от 1 до 30 дней / от 31 до 90 дней / от 91 до 180 дней / свыше 180 дней – в момент времени  $t$ ;

$\Delta y_{t-i}^s$  – задолженность по ИЖК по срокам задержки платежей в  $s$  дней с лагом  $i$  месяцев;

$\Delta MDRI_{t-l}$  – индекс риска ипотечного дефолта с лагом от  $l$  месяцев до  $r$  месяцев;

$\Delta Controls_{t-f}$  – контрольные (экзогенные) переменные с лагом от  $f$  месяцев до  $k$  месяцев. В качестве контрольных переменных выступают средневзвешенная ставка по ипотечным кредитам ( $INTEREST_t$ ), уровень безработицы населения РФ ( $UNEMPL_t$ ) и среднедушевые доходы населения РФ ( $INCOME_t$ ). Согласно исследуемой литературе, данные показатели являются одними из наиболее популярных для проведения подобного исследования (Lin, Yang, 2003; Miranda, McGrath, 2011; Gerardi et. al., 2013; Mocetti, Viviano, 2016; Mondria et. al., 2017). Более того, выбор независимых переменных основывался на корреляционном анализе, в качестве контрольных выбирались показатели, обладающие наиболее высоким

коэффициентом корреляции с зависимой переменной, что представлено в разделе описания данных данной работы.

$\beta_i$  – неизвестные коэффициенты авторегрессии;

$\delta_j$  – неизвестные коэффициенты скользящего среднего;

$q, m$  – коэффициенты, задающие порядок модели авторегрессии и скользящего среднего соответственно;

$\alpha, \gamma_l, \omega_f$  – неизвестные коэффициенты для экзогенных переменных;

$\varepsilon_{t-j}, \varepsilon_t$  – белый шум (i.i.d.), показывает случайные колебания процесса.

Подбор спецификации модели проводился на этом же этапе исследования. Первоочередной задачей при моделировании и прогнозировании временных рядов с помощью модели ARIMA является решение проблемы нестационарности. Стационарность временных рядов означает постоянство среднего значения и вариации (Канторович, 2002). С инструментальной точки зрения для выявления нестационарных процессов используются тесты на наличие единичного корня. Наиболее популярный из них – расширенный тест Дики-Фуллера (Augmented Dickey-Fuller test, ADF). Он позволяет уверенно (с вероятностью выше 99%) отвергнуть гипотезу о наличии единичного корня с учетом константы и временного тренда. (Канторович, 2002). В качестве методики проведения данного тестирования предлагается процедура Доладо-Дженкинсона-Сосвилла-Ривьеро. Таким образом, используя тест ADF, для каждого ряда на обучающем интервале выборки был определен порядок интегрирования  $d$ , другими словами, минимальное количество периодов, необходимое для вычисления разности временных рядов для получения стационарного ряда. В итоге к каждому исходному ряду применили взятие первой разности, что описано подробнее в разделе данные работы.

Далее следующим образом выбиралась оптимальная величина лагов независимых переменных, включенных в модели (1)-(4). Во-первых, выбор лага происходил на основе  $t$ -статистик и  $F$ -статистик. Методом перебора необходимо было найти ту длину лага, когда  $t$ -статистика и  $F$ -статистика

значимы. Во-вторых, выбор лага происходил на основе информационного критерия Акаике (AIC). Длина лага с минимальным информационным критерием предпочтительнее. В-третьих, величина лага выбиралась таким образом, чтобы устранить автокорреляцию остатков.

Наконец, для каждого ряда с порядком интегрирования  $d$  подбирались параметры  $p$  и  $q$  модели  $ARIMA(p,d,q)$ , отвечающие за количество слагаемых авторегрессионной части и за количество слагаемых части скользящего среднего соответственно. Выбор наилучшей спецификации производился путем перебора моделей  $ARIMA(p,d,q)$  с заданными параметрами  $q$  и  $p$ , где  $p = 1, \dots, 6$  и  $q = 1, \dots, 6$ , с последующей сортировкой моделей по информационному критерию Акаике (AIC). Также в моделях, где это было необходимо, была учтена и включена в модель компонента сезонности.

Следом строилась аналогичная модель  $ARIMAX$ , но без включения ИРИД (обозначения сохраняются прежние):

$$\Delta y_t^s = \alpha + \sum_{i=1}^q \beta_i * \Delta y_{t-i}^s + \sum_{j=1}^m \delta_j * \varepsilon_{t-j} + \varepsilon_t + \sum_{f=0}^k \omega_f * \Delta \text{Controls}_{t-f} \quad (5)$$

Оценка параметров моделей  $ARIMAX$ , представленных уравнениями (4) и (5), проводилась с помощью метода максимального правдоподобия (ММП).

На четвертом этапе исследования строится прогноз объемов просроченной ипотечной задолженности по моделям  $ARIMAX$  на тестовой выборке на десять месяцев вперед (с мая 2018 г. по февраль 2019 г.). На завершающем пятом шаге сопоставляется качество моделей (4) и (5) по такому критерию прогнозной силы моделей, как квадратного корня из среднеквадратичной ошибки (RMSE), который может быть рассчитан по следующей формуле:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\Delta y_{i, \text{прогноз}} - \Delta y_{i, \text{факт}})^2} \quad (6)$$

где:

$n$  – количество наблюдений для тестовой выборки;

$\Delta u_{i,\widehat{\text{прогноз}}}$  – прогнозные значения первой разности объемов просроченной ипотечной задолженности, полученные с помощью модели;

$\Delta u_{i,\text{факт}}$  – фактические значения первой разности объемов просроченной ипотечной задолженности.

Поскольку включение нерелевантных переменных в модель может увеличить RMSE, то наилучшей моделью прогнозирования можно признать ту, которая минимизирует RMSE. С одной стороны, квадратный корень из среднеквадратичной ошибки довольно часто используется в качестве стандартной статистической метрики для измерения точности прогнозирования объемов просроченной ипотечной задолженности (например, Crook, Banasik, 2012; Chauvet et. al., 2016). С другой стороны, в аналогичных эмпирических работах также используются такие статистические метрики, как среднее абсолютное отклонения (Lin, Yang, 2003) или сумма квадратов ошибок (Miranda, McGrath, 2011). В целом, в литературе не существует никакого общего мнения о наиболее подходящей метрике для ошибок модели. Тем не менее, Т. Чай и Р. Дракслер (2014) утверждают, что RMSE целесообразно использовать, чем другие метрики, например, среднюю абсолютную ошибку (Mean Absolute Error, MAE), когда ошибки модели соответствуют нормальному распределению. Авторы выделяют еще одно явное преимущество RMSE: эта метрика легко интерпретируется, поскольку имеет те же единицы, что и исходные значения.

## Данные и их предварительный анализ

Для эмпирического исследования используются два набора данных. Первый набор данных содержит информацию из сервиса статистики ключевых слов Google Trends о 12 поисковых запросах, представленных ранее в табл. 2. Выборка по соответствующей статистической информации используется за максимально доступный продолжительный период времени, на месячной частоте и глубина выборки составляет 195 месяцев (январь 2004 г. – март 2020 гг.).

Второй набор данных включает в себя временные ряды, полученные из официальной статистики, опубликованной на сайтах Росстата и Центрального Банка России, и представляет собой агрегированные данные о таких макроэкономических показателях, как средневзвешенная ставка по ипотечным кредитам, уровень безработицы населения РФ и среднедушевые доходы населения РФ. В работе используются данные с интервалом в один месяц с февраля 2010 г. по февраль 2019 г. (всего 108 наблюдений для каждого показателя). В таб. 3 приведен полный список переменных и их описательные статистики.

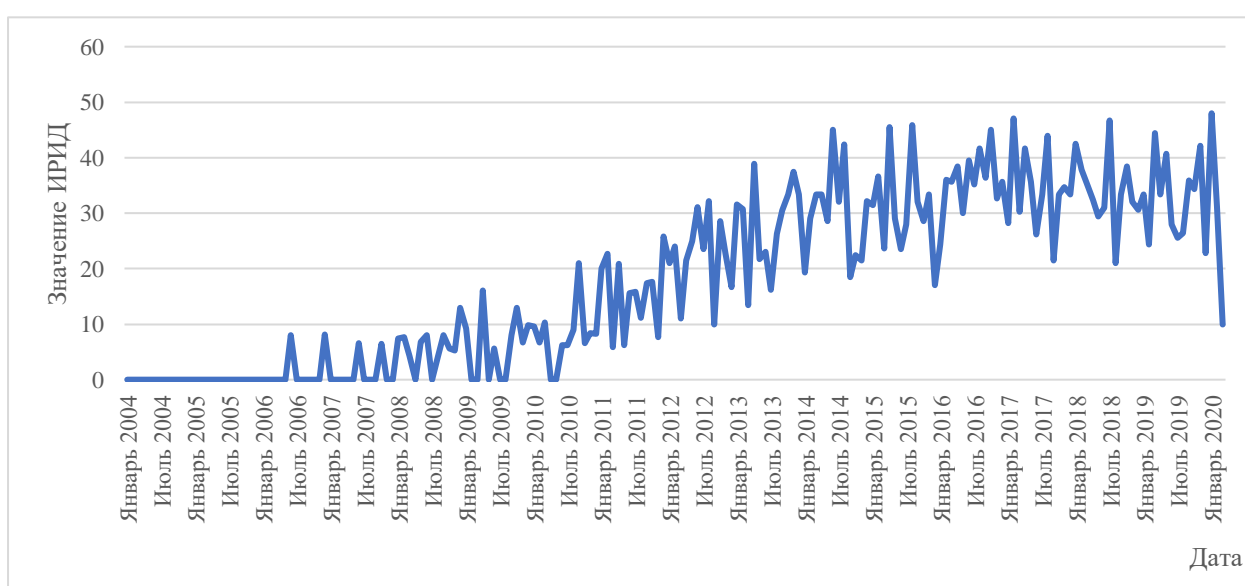


Рис. 2. Динамика индекса риска ипотечного дефолта  
(январь 2004 г. – январь 2020 г.)

Таблица 3

## Исследуемые переменные и их описательные статистики

Название переменной и ед. изм.	Обозначение	Среднее значение	Медиана	Стандартное отклонение	Минимум	Максимум	Коэф. эксцесса	Коэф. асимметрии
Индекс риска ипотечного дефолта, усл. ед.	MDRI	26.65	28.80	11.34	0.00	47.06	2.51	-0.40
Общая сумма задолженности по ИЖК с просроченными платежами от 1 до 30 дней, в млн. руб.	DELINQ30	64 035	56 767	27 303.04	19 079	129 814	2.13	0.44
Общая сумма задолженности по ИЖК с просроченными платежами от 31 до 90 дней, в млн. руб.	DELINQ90	20 200	21 269	8 026.89	7 811	38 146	2.18	0.15
Общая сумма задолженности по ИЖК с просроченными платежами от 91 до 180 дней, в млн. руб.	DELINQ180	12 918	13 426	5 645.69	5 019	26 185	2.14	0.26
Общая сумма задолженности по ИЖК с просроченными платежами свыше 180 дней, в млн. руб.	DELINQ_MORE180	72 011	57 778	24 247.53	45 137	111 570	1.35	0.39
Средневзвешенная ставка по ипотечным кредитам, выданным в течение месяца	INTEREST	12.26	12.30	1.17	9.55	14.53	3.20	-0.69
Уровень безработицы населения РФ в возрасте 15 - 72 лет, %	UNEMPL	5.65	5.40	0.79	4.50	8.70	5.92	1.57
Среднедушевые доходы населения РФ, в руб.	INCOME	26 892	26 853	5 480.13	18 958	38 848	2.12	0.25

Как видно из табл. 3 среднее значение ИРИД в период с января 2004 г. по январь 2020г. равняется 26,65, с минимальным значением 0,00 и максимальным – 47,06. Согласно динамике ИРИД (рис. 2) минимальное значение достигалось в период с января 2004 г. по июнь 2006 г., что может быть связано с тем, что рынок ИЖК РФ начал активно развиваться только с 2005-2007 гг.: снижение уровня инфляции и ставки по ипотечному кредитованию и рефинансированию, которые постепенно создавали подходящие условия для развития рынка ипотеки. Более того популярность поискового сервиса Google на территории России пришлась примерно на эту же пору. Таким образом, причин для возникновения просроченных платежей и беспокойства заемщиков в более ранние периоды не было. Максимальное значение ИРИД, равное 47,06, говорит о том, что на рынке ИЖК наблюдается паническое настроение среди заемщиков, по аналогии с работой (Chauvet et. al., 2016).

На динамику индекса значительно влияют экономическая ситуация в России, что может объяснять высокую частоту колебаний ряда от месяца к месяцу. Как видно из рис. 2 в начале исследуемого периода индекс риска ипотечного дефолта находился на относительно невысоком уровне вплоть до августа 2009 г. Затем индекс риска ипотечного дефолта начал расти скачкообразно.

Из рис. 2 также видно, что дальнейшая динамика индекса характеризуется заметными подъемами в июне 2014 г., апреле 2015 г., августе 2015 г. и в феврале 2017 г. Это, в свою очередь, можно объяснить следующими произошедшими событиями на российском рынке ипотечного жилищного кредитования. Во-первых, в июне 2014 г. президентом РФ был подписан Федеральный закон «О внесении изменений в статьи 31 и 61 Федерального закона «Об ипотеке (залоге недвижимости)», согласно которым он дополняется нормами, касающимися установления механизмов страхования ответственности лица, являющегося должником по обеспеченному ипотекой обязательству, а также финансового риска



кредитора от возникновения убытков, вызванных недостаточной стоимостью заложенного имущества. Во-вторых, в апреле 2015 г. Правительством РФ была утверждена программа помощи ипотечным заемщикам, согласно которой из резервного фонда Правительства РФ будут выделены дополнительные денежные средства в размере 2 млрд рублей для оказания поддержки отдельным заемщикам, у которых сохраняются сложности с погашением ипотечных кредитов. Соответственно, в августе 2015 г. эта программа начала работать. В-третьих, в феврале 2017 г. объем выдачи ипотеки в России сократился относительно аналогичного периода прошлого года на 22%, была отменена программа выдачи ипотечного кредита с государственной поддержкой. Это привело к тому, что ставки по ипотечным кредитам на новостройки выросли примерно на полтора процентных пункта.

Наибольшее среднее значение объемов просроченной задолженности по ИЖК в период с февраля 2010 г. по февраль 2019 г. можно наблюдать у категории задолженности по срокам задержки платежей свыше 180 дней – 72 011 млн. руб., далее по убыванию идет категория задолженности по срокам задержки платежей до 30 дней, среднее значение объемов просроченной задолженности которой равно 64 035 млн. руб. (табл. 3). На рис. 3 представлена динамика задолженности по ипотечным жилищным кредитам по срокам задержки платежей в срок до 30, 90, 180 и свыше 180 дней. Данные ряды также характеризуется высокой и изменяющейся во времени волатильностью. Предварительный визуальный анализ данных временных рядов показывает, что в данных имеется положительно направленный тренд вплоть до марта 2016 г.

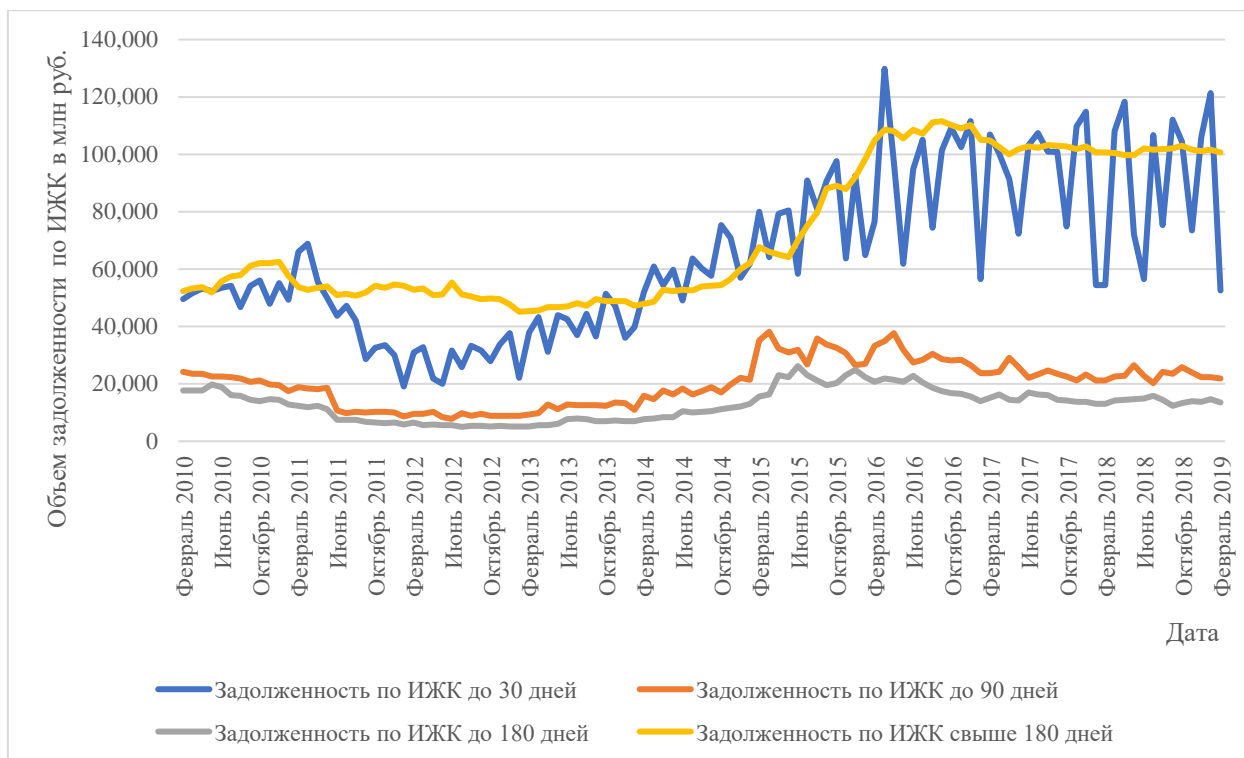


Рис. 3. Динамика просроченной задолженности по ИЖК по срокам задержки платежей до 30, 90, 180 и свыше 180 дней (февраль 2010 г. – февраль 2019 г.) \*

\*Сост. по источнику: Статистический сборник «Сведения о рынке жилищного (ипотечного) кредитования в России №5. Москва, 2018. С. 16-23

Говоря об экзогенных переменных, следует подчеркнуть, что средняя средневзвешенная ставка процента по ипотечным кредитам равна 12,26%, при максимальном значении, равном 14,53%, и минимальном – 9,55% (табл. 3). Динамика экзогенных переменных представлена в Приложении 2. Временные ряды характеризуются наличием тренда, для ряда среднедушевых доходов населения РФ – положительный возрастающий тренд, а для рядов средневзвешенной ставки процента по ипотечным кредитам и уровнем безработицы – ниспадающий. Также можно наблюдать периодически повторяющиеся сезонные эффекты для рядов доходов и уровня безработицы. Кроме того, все исследуемые ряды имеют распределение отличное от нормального с наличием пиков и тяжелыми хвостами: коэффициент эксцесса для каждой экзогенной переменной существенно превышает ноль. Отметим также, что уровень безработицы

имеет наибольший коэффициент эксцесса и коэффициент асимметрии, следовательно, этот временной ряд наиболее волатилен среди рассматриваемых в работе. Таким образом, вышеописанные характеристики рядов подтверждают их нестационарность.

В Приложении 3 представлена матрица выборочных парных коэффициентов корреляции Пирсона. Анализ корреляционной матрицы показал, что индекс риска ипотечного дефолта обладает статистически значимой корреляцией со всеми исследуемыми макроэкономическими показателями. Стоит отметить, что ИРИД коррелирует со всеми изучаемыми переменными, в том числе и со всеми случаями просроченных ипотечных платежей, что согласуется с результатами, полученными в работе М. Шове и др. (2016). Однако, в данном исследовании наибольший коэффициент корреляции Пирсона наблюдается между ИРИД и задолженностью по ИЖК на срок свыше 180 дней – 0,52, при этом в работе М. Шове и др. (2016) наиболее тесная связь наблюдается между ИРИД и показателями 60-дневной просрочки – коэффициент корреляции Пирсона равен 0,86. Данное различие может быть связано с тем, что банки США обычно направляют уведомление через суд заемщикам о наличии у них просроченной задолженности по ипотеке спустя 60 дней, тем самым побуждая заемщиков после получения данного уведомления искать в Интернете какую-либо информацию о случаях просроченных ипотечных платежей (Chauvet et al., 2016), в то время как в России период направления кредитором уведомления о просроченной задолженности и обращения в суд чуть больше и обычно на практике достигает 180 дней.

Согласно процедуре Доладо-Дженкинсона-Сосвилла-Ривьеро, результаты которой представлены в Приложении 4, гипотеза о наличии единичных корней в изучаемых временных рядах принимается на всех трех уровнях значимости для всех трех вариантов (тест с константой и трендом, тест с константой, тест без константы и тренда). Об этом свидетельствуют р-значения, равные больше, чем 0,1 для всех временных рядов, что, в свою

очередь, больше всех трех уровней значимости (0,01/0,05/0,1). Исключением является временные ряды ИРИД и уровень безработицы населения РФ. Для ряда ИРИД в тесте с константой и трендом гипотеза о наличии единичного корня отклоняется на 10% уровне значимости, но, тем не менее, эта же гипотеза в двух оставшихся тестах принимается на 5% и 10% уровнях значимости, о чем свидетельствуют р-значения, равные 0,07 и 0,82. Для ряда уровня безработицы гипотеза о наличии единичного корня отклоняется в тестах с константой и трендом и с константой, о чем свидетельствуют р-значения, равные 0,004 и 0,001. Соответственно для теста без константы и тренда эта же гипотеза для ряда уровня безработицы принимается на 1% уровне значимости. Таким образом, для того, чтобы приблизить временные ряды к стационарному виду и повысить устойчивость условного среднего значения было решено ко всем исходным показателям применить взятие их первых разностей. Проверив первые разности этим же тестом, можно сделать вывод о том, что они являются стационарными рядами (р-значения для каждого временного ряда меньше, чем 0,1, что, в свою очередь, меньше всех трех уровней значимости) (Приложение 4). Следовательно, все ряды являются интегрированными порядка 1, т. е. рядами I (1).

## Эмпирические результаты

В ходе исследования было построено четыре модели ARIMAX, с помощью которых были спрогнозированы объемы просроченной задолженности по ипотечному кредиту. В Таблице 4 представлены результаты оценивания данных моделей.

Таблица 4

Результаты оценивания модели ARIMAX с включением индекса риска ипотечного дефолта для объемов просроченной

	Зависимая переменная			
	$\Delta DELINQ30_t$	$\Delta DELINQ90_t$	$\Delta DELINQ180_t$	$\Delta DELINQ\_MORE180_t$
	(1)	(2)	(3)	(4)
<i>AR</i> (1)	-0.88 *** (0.11)			0.882 *** (0.08)
<i>AR</i> (2)	-0.99 *** (0.14)			
<i>AR</i> (3)	-0.63 *** (0.14)			
<i>AR</i> (4)	-0.37 *** (0.11)			
<i>MA</i> (1)		-0.31 *** (0.15)		0.06 *** (0.16)
<i>SMA</i> (1)		0.58 *** (0.22)		
	Индекс риска ипотечного дефолта			
$\Delta MDRI_t$	-51.88 (169.95)	-19.9 (28.42)	3.29 (17.02)	2.24 (28.18)
$\Delta MDRI_{t-1}$	268.99 (166.95)	11.16 (28.14)	-9.59 (23.31)	-15.09 (42.12)
$\Delta MDRI_{t-2}$	34.28 (165.21)	-42.27 (30.72)	19.23 (25.12)	-60.99 (46.51)
$\Delta MDRI_{t-3}$	-151.14 (169.62)	-94.92 *** (27.71)	39.78 (25.82)	-104.25 ** (48.35)
$\Delta MDRI_{t-4}$	537.50 *** (186.52)	-104.97 *** (37.39)	68.25 ** (26.84)	-78.14 (49.12)
$\Delta MDRI_{t-5}$	318.28 * (172.82×10 <sup>2</sup> )	-87.62 *** (31.72)	19.24 (24.31)	-13.66 (43.76)
$\Delta MDRI_{t-6}$	-57.64 (186.65)	-24.65 (25.39)	-13.66 (17.58)	21.28 (29.05)

Таблица 4 (Продолжение)

Результаты оценивания модели ARIMAX с включением индекса риска  
ипотечного дефолта для объемов просроченной

	Зависимая переменная			
	$\Delta DELINQ30_t$	$\Delta DELINQ90_t$	$\Delta DELINQ180_t$	$\Delta DELINQ\_MORE180_t$
	(1)	(2)	(3)	(4)
Среднедушевые доходы населения РФ				
$\Delta INCOME_t$	1.75 *** (0.64)	-0.04 (0.16)	-0.01 (0.06)	0.05 (0.11)
$\Delta INCOME_{t-1}$	-2.03 *** (0.56)	-0.41 *** (0.28)	0.005 (0.06)	0.19 ** (0.08)
$\Delta INCOME_{t-3}$	0.066 *** (0.56)	-0.21 (0.20)	0.03 (0.05)	0.07 (0.14)
$\Delta INCOME_{t-6}$	-0.69 (0.47)	0.09 (0.17)	0.13 * (0.07)	0.09 (0.13)
Уровень безработицы населения РФ				
$\Delta UNEMPL_t$	$1.21 \times 10^4$ *** ( $4.16 \times 10^3$ )	-1593.70 (1180.30)	324.84 (626.84)	-2136.40 ** (996.36)
$\Delta UNEMPL_{t-6}$	$6.75 \times 10^3$ ** ( $3.38 \times 10^3$ )	-1356.40 (1007.30)	925.13 * (536.57)	-605.85 (880.43)
Средневзвешенная ставка по ипотечным кредитам				
$\Delta INTEREST_t$	$1.10 \times 10^4$ ** ( $4.81 \times 10^3$ )	-604.33 (801.52)	1170.90 ** (558.93)	963.82 (974.42)
$\Delta INTEREST_{t-1}$	$-9.29 \times 10^3$ * ( $4.89 \times 10^3$ )	2501.00 *** (791.04)	529.57 (530.76)	826.49 (929.39)
Константа	810.67 * (416.06)			
Количество наблюдений	98	98	98	98
Логарифм функции правдоподобия	-988.93	-832.47	-782.08	-831.25
Информационный критерий Акаике	2016.46	1700.95	1596.17	1695.5

*Примечание.* В скобках приведены робастные стандартные ошибки. Уровни значимости: \*, \*\*, \*\*\* – 10, 5 и 1 % соответственно.

Полученные результаты носят ограниченный характер и являются неустойчивыми, поскольку находятся спецификации, в которых есть значимое влияние первой разности ИРИД вплоть до пятого лага, но направление влияния у данных регрессоров разное. Лаговые значения ИРИД оказывают положительное влияние на 30-дневную и 180-дневную ипотечную

просрочку. Так, например, в среднем и при прочих равных при увеличении первой разности ИРИД четыре месяца назад и пять месяцев назад на 1 ед. первая разность объемов просроченной задолженности по ИЖК в срок до 30 дней увеличится на 537,5 и 318,28 млн. руб., соответственно. Аналогично, с ростом первой разности ИРИД четыре месяца назад первая разность 180-дневной просрочки вырастет на 68,25 млн. руб. в среднем и при прочих равных. Обратное влияние оказывает первая разность ИРИД на первую разность объемов просроченной ипотечной задолженности в срок до 90 дней и свыше 180 дней. В среднем и при прочих равных при увеличении первой разности ИРИД три, четыре и пять месяцев назад на 1 ед. первая разность объемов просроченной ипотечной задолженности по ИЖК в срок до 90 дней сократится на 94,92, 104,97 и 87,62 млн руб. соответственно. В данном случае знаки коэффициентов не соответствуют ожиданиям, но, тем не менее, результаты частично согласуются с результатами, полученными в работе М. Шове и др. (2016): на объем просроченной ипотечной задолженности в срок до 90 дней статистически значимое отрицательное влияние оказывает ИРИД с лагом в четыре месяца. Схожим образом сокращается первая разность ипотечной просрочки со сроком задержки платежей свыше 180 дней на 104,25 млн. руб. при увеличении первой разности ИРИД с лагом в 4 месяца.

Значение коэффициентов для моделей ARIMAX без включения ИРИД в качестве экзогенной переменной представлены в Приложении 5. Согласно показателям качества модели, выраженным в виде информационного критерия Акаике (AIC), модели с включением ИРИД являются лучшими по сравнению с моделями без учета ИРИД. Значение AIC для каждой спецификации в среднем сокращается на 3,5 ед. при включении в модель ИРИД. Модель (3) с учетом ИРИД является наилучшей с минимальным значением AIC 1596,17, что превышает значения AIC для аналогичной модели без учета ИРИД (1598,60). Если сравнивать модели с включением ИРИД и без включения ИРИД по значению логарифма функции правдоподобия, то мы получим такие же результаты (значение логарифма

функции правдоподобия в среднем увеличивается на 6,43 ед. при включении в модель ИРИД).

Отметим, что распределение остатков, полученных для всех четырех моделей, является нормальным (проведена проверка на нормальность при помощи теста Q-теста Льюнга-Бокса).

Из табл. 4 можно также заключить, что первые разности исследуемых контрольных переменных тоже оказывают статистически значимое влияние на первые разности объемов просроченной ипотечной задолженности. В большой степени воздействует на первую разность объемов 30-дневной просроченной задолженности: из 8 контрольных переменных 7 оказались значимыми, исключением является первая разность среднедушевых доходов населения РФ с лагом в 6 месяцев. Поскольку предшествующие работы изучали влияние групп факторов на возникновение/объем просроченной ипотечной задолженности только в текущий момент времени, то мы можем сопоставить наши выводы данным результатам лишь частично (только для нелагированных контрольных переменных). Первая разность уровня безработицы в текущий момент времени оказывает статистически значимое на 1% уровне значимости отрицательное влияние на первую разность объемов просроченной задолженности по ИЖК по срокам свыше 180 дней, что соотносится с результатами, полученными в предшествующей литературе (Miranda, McGrath, 2011; Gerardi et. al., 2013; Mocetti, Viviano, 2016). Более того, мы также частично соглашаемся с выводами работ (Aristei, Gallo, 2016; Fuster, Willen, 2017; Miranda, McGrath, 2011), о том, что с ростом процентной ставки по ипотеке объем просроченных платежей увеличивается (обнаружено положительное статистически значимое на 5% уровне значимости влияние первой разности средневзвешенной ставки по ипотечным кредитам в текущий момент времени на первую разность объемов 30-дневной и 180 дневной просроченной задолженности). Что касается, доходов населения, то наши выводы о том, что первая разность среднедушевых доходов населения в текущий момент оказывает



положительное статистически значимое на 1% уровне значимости влияние на первую разность объемов 30-дневной просроченной задолженности, противоречит результатам в работах (Lin, Yang, 2003; Diaz-Serrano, 2005; Gomez-Salvador et. al., 2011). Но этот же регрессор в предыдущий месяц оказывает отрицательное статистически значимое на 1% уровне значимости влияние на первую разность объемов 30-дневной и 90-дневной просроченной задолженности по ИЖК.

Результаты прогнозирования объемов просроченной задолженности по ИЖК с помощью использования ИРИД на 10 месяцев вперед, то есть на период с мая 2018 г. по февраль 2019 г., представлены на рис. 5. Синей линией на рис. 5 отображены прогнозные значения первой разности объемов просроченной ипотечной задолженности по срокам задержки платежей до 30, до 90, до 180 и свыше 180 дней, серая область, верхняя и нижняя линии которой образуют коридор с уровнем доверия 95% (Приложение 6). В ближайшие десять месяцев первая разность объемов просроченной ипотечной задолженности с уровнем доверия 95% будет лежать в этом коридоре. Согласно рис. 5 для каждой категории задолженности ожидается прирост объемов на февраль 2019 г.

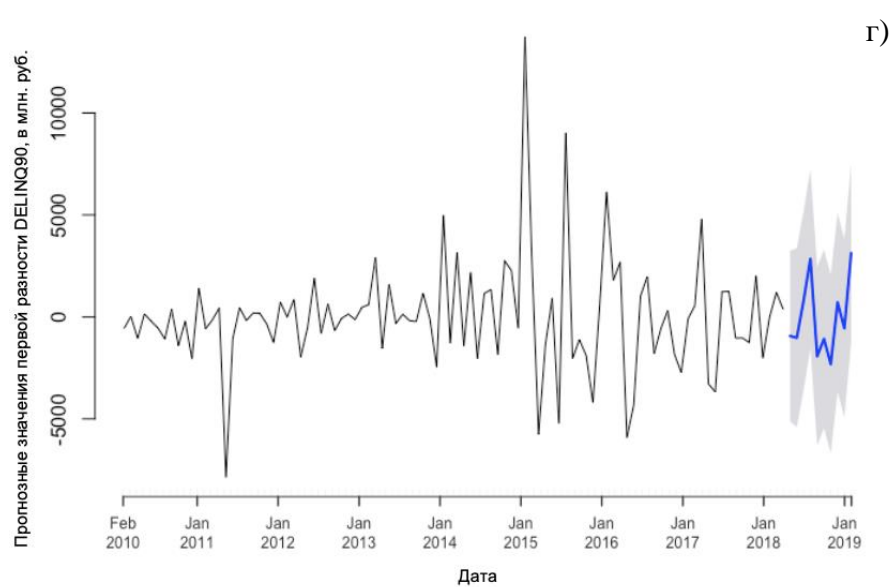
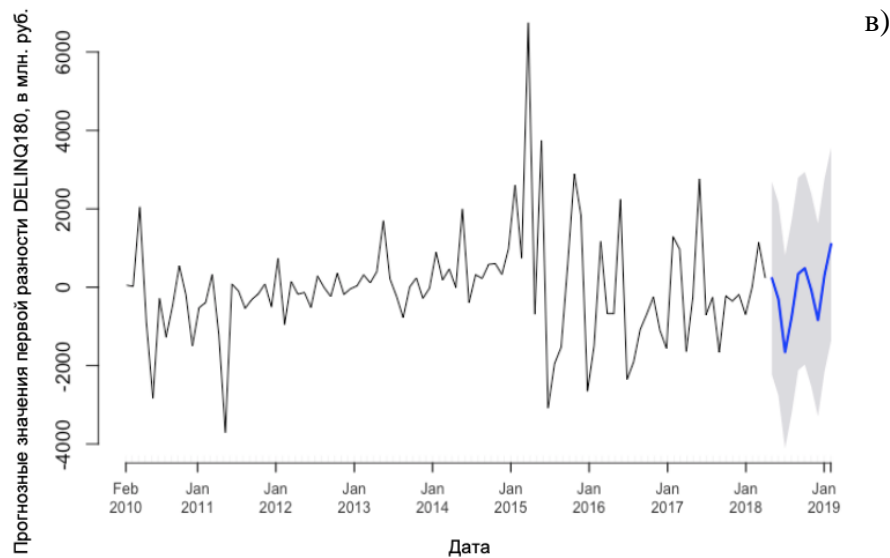
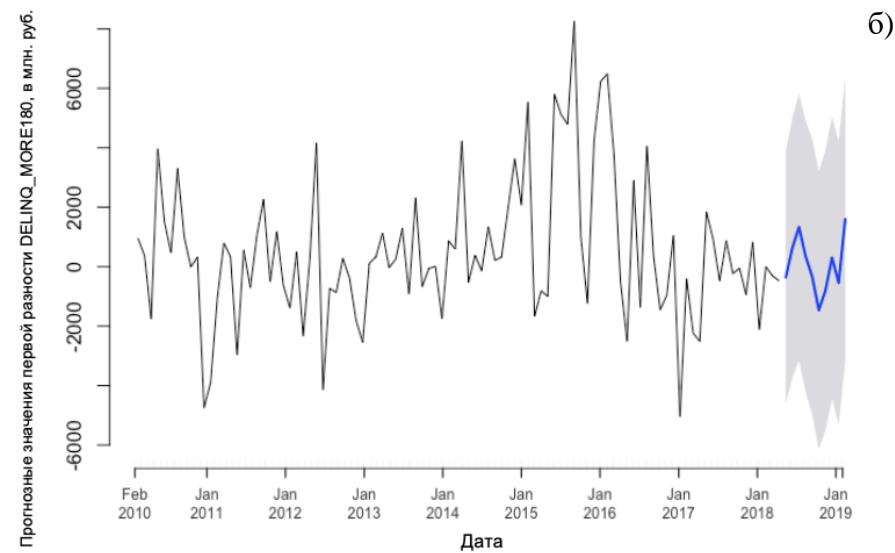
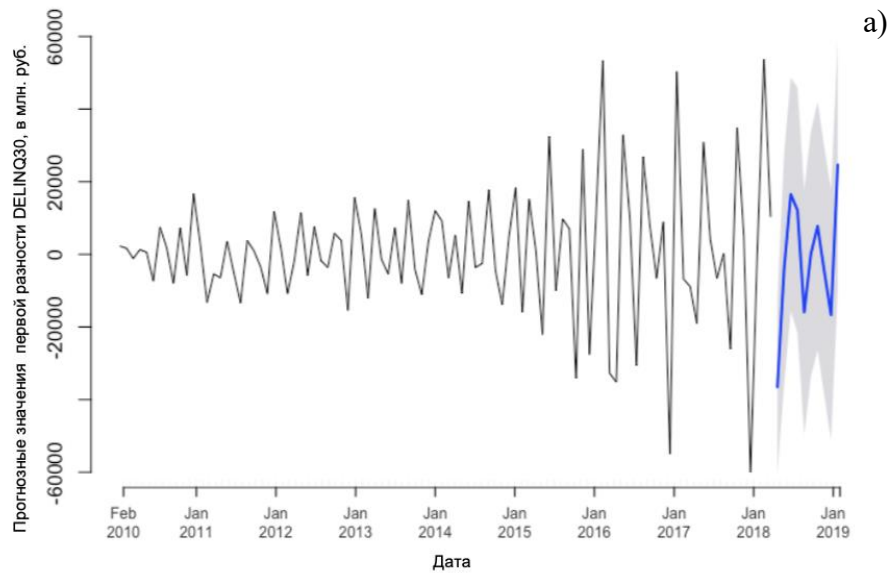


Рис. 5. Реальные и прогнозные значения первой разности объемов просроченной задолженности по ИЖК в срок а) до 30 дней; б) до 90 дней; в) до 180 дней; д) свыше 180 дней в соответствии с моделями (1)-(4) соответственно

Для ответа на вопрос о том, позволяет ли включение ИРИД повысить точность прогнозирования объемов просроченной ипотечной задолженности по ИЖК, рассчитана такая статистика качества прогноза как квадратный корень из средней квадратичной ошибки прогнозирования, представленная в табл. 5.

Таблица 5

Сравнение качества прогноза посредством RMSE для моделей с и без ИРИД

	Зависимая переменная			
	$\Delta DELINQ_{30_t}$	$\Delta DELINQ_{90_t}$	$\Delta DELINQ_{180_t}$	$\Delta DELINQ\_MORE_{180_t}$
	(1)	(2)	(3)	(4)
	Значения RMSE			
Контрольные переменные	43 258.34	2 894.68	1 401.57	1 730.55
$\Delta MDRI$ и Контрольные переменные	37 255.58	2 687.35	1 247.07	1 469.69
Разница, %	-0.13	-0.07	-0.11	-0.15

В целом можно отметить, что изначальное предположение о том, что включение ИРИД в модель улучшает качество прогноза, подтвердилось, о чем говорит общее снижение показателя – среднеквадратичной ошибки прогноза. Сравним точность прогноза для моделей (1)-(4). Можно сказать, что прогнозное качество модели (1) значительно выросло после включения ИРИД. Снижение средней квадратичной ошибки прогноза (RMSE снизился с 43 258,32 до 37 255,58) говорит о том, что в абсолютных величинах отклонение прогноза от реального значения в модели с включением ИРИД меньше, чем то же отклонение у модели без него. Наибольшая разница в значениях RMSE произошла при прогнозировании первой разности объемов просроченной ипотечной задолженности в срок свыше 180 дней с помощью модели (2) (RMSE снизился с 1730,55 до 1469,69). Для моделей (2) и (3)

прогнозирования первой разности объемов 90-дневной и 180-дневной просроченной ипотечной задолженности RMSE сократился на 7% и 13% соответственно. Несмотря на то, что в работе (Chauvet et. al., 2016) использовалась другая метрика прогноза качества (среднеквадратичная процентная ошибка, MPSE), полученные результаты о том, что ИРИД обладает довольно высокой прогнозной силой, то есть улучшает прогнозы объемов просроченной задолженности по ипотечному кредиту, совпадают с нашими выводами.

Согласно полученным прогнозным значениям, наибольший объем просроченной ипотечной задолженности нужно было ожидать в августе 2018 г., причем для категории задолженности с 30-дневными просроченными платежами. Объем задолженности составил бы 106 428,5 млн. руб., однако прогнозное значение показателя выше фактического (75 425 млн. руб.) в 1,4 раза. Наименьший объем просроченной ипотечной задолженности нужно было ожидать в декабре 2018 г., причем для категории задолженности с 180-дневными просроченными платежами. В данном случае объем задолженности составил бы 11 758,93 млн. руб., что меньше фактического значения показателя (13 648 млн. руб.) в 0,8 раз. Если предположить, что кредитные организации установили бы сумму резервов на возможные потери на период с мая 2018 г. по февраль 2019 г. в размере суммы полученных прогнозных значений объемов просроченной задолженности по срокам задержки платежей до 30 дней, то на начало марта 2019 г. на балансе кредитных организаций осталось бы 31 736,68 млн. руб. Однако, что касается категорий задолженности с другими сроками задержки платежей, установленного резерва бы не хватило, чтобы покрыть фактические потери, то есть прогнозные значения оказались завышенными по сравнению с фактическими.

## Заключение

В работе прогнозируются объемы просроченной задолженности по ипотечным кредитам. Данная тема остается актуальной для кредитных организаций, поскольку, зная прогнозные значения объемов просроченной ипотечной задолженности, кредитные организации смогут определить минимально допустимый уровень резервов на возможные потери по ссудам и сформировать комплекс мер по минимизации реальных потерь банков.

Более того, проведенное исследование демонстрирует возможности использования статистики интернет-запросов пользователей Интернета в Google для оценки настроений заемщиков на рынке ипотечного жилищного кредитования. Показана перспективность использования этих данных для более эффективного моделирования прогнозов объемов просроченной задолженности по ИЖК. Для этого на основе информации о частоте поисковых запросов, таких как «ипотека помощь» или «долг по ипотеке», и путем их компиляции был построен индекс риска ипотечного дефолта, согласно методологии предложенной в работе Шове и др. (2016), на российских данных.

Цель данного исследования состояла в том, чтобы оценить прогностические возможности ИРИД. Поделив исходные данные на обучающую (март 2010 г. – апрель 2018 г.) и тестовую (май 2018 г. – февраль 2019 г.) выборки, в работе было оценено четыре модели авторегрессионного скользящего среднего с экзогенными переменными, где в качестве зависимой переменной выступали объемы просроченных задолженностей по ипотечному кредиту по срокам задержки платежей до 30, 90, 180 и свыше 180 дней соответственно. В качестве контрольных переменных выступали ИРИД, средневзвешенная ставка по ипотечным кредитам, уровень безработицы населения РФ и среднедушевые доходы населения РФ. На тестовой выборке были рассчитаны месячные прогнозы с использованием имеющихся данных с мая 2018 г. по февраль 2019 г. Аналогичный процесс

был реализован для построения этих же авторегрессионных моделей скользящего среднего, но без включения ИРИД в качестве контрольной переменной. На основании результатов полученных прогнозов оценивалась прогностическая сила индекса с использованием квадратного корня из среднеквадратичной ошибки. В результате, все модели с включением ИРИД и ИРИД с лагами от одного до шести месяцев превосходили авторегрессии без включения ИРИД по указанному критерию, значение RMSE сокращалось примерно на 0,12%. Данный вывод частично согласуется с результатами, полученными в работе (Chauvet et. al., 2016) в части того, что включение ИРИД в модель прогноза объемов просроченной ипотечной задолженности улучшает точность прогноза. Однако точность данного прогноза в работе (Chauvet et. al., 2016) выше, поскольку метрика качества прогноза (MPSE) улучшается в среднем на 0,45%. Более того, результаты исследования свидетельствуют о том, что не только текущие значения ИРИД оказывают статистически значимое влияние на объем просроченной задолженности, но и его запаздывающие значения, вплоть до 5 лага. Согласно построенному прогнозу, наибольший объем просроченной ипотечной задолженности стоило ожидать в августе 2018 г. в размере 106 428,5 млн. руб., а наименьший – в декабре 2018 г. в размере 11 758,93 млн. руб. Если предположить, что кредитные организации установили бы сумму резервов на возможные потери на период с мая 2018 г. по февраль 2019 г. в размере суммы полученных прогнозных значений объемов просроченной задолженности, то на начало марта 2019 г. на балансе кредитных организаций осталось бы 31 736,68 млн. руб., однако данный вывод касается только категории 30-дневной просроченной задолженности, для остальных категорий задолженности соответствующего резерва не было бы достаточно, чтобы покрыть фактические потери. Стоит отметить, результаты исследования свидетельствуют о том, что на объемы просроченной ипотечной задолженности статистически значимое влияние оказывают средневзвешенная ставка по ипотечному кредиту, уровень безработицы

населения РФ, среднедушевые доходы населения РФ и их лаговые значения в зависимости от количества дней просрочки платежа.

В то время как ИРИД действительно представляется для кредитных организаций, банковских аудиторских компаний и инвестиционных фирм как универсальный инструмент прогнозирования объемов просроченных задолженностей и новая мера риска дефолта по ипотечным кредитам, существуют некоторые ограничения для его использования. Во-первых, ИРИД фиксирует риск дефолта по ипотечным кредитам только на уровне заемщиков, и, следовательно, не дает никакой информации о состоянии кредиторов. Таким образом, высокий уровень ИРИД может не обязательно сигнализировать о наступлении финансового кризиса или экономического спада на рынке ИЖК, если кредиторы достаточно хорошо подготовлены для того, чтобы справиться с соответствующим повышенным уровнем просроченных задолженностей. Во-вторых, ИРИД может быть уязвим к изменениям в поведении поиска в Интернете. Например, если поисковая система Google перестанет пользоваться популярностью у пользователей или появятся новые продукты (например, голосовой поиск), то это приведет к тому, что привычки веб-пользователей изменятся и часть поисковых запросов будет отсутствовать в будущем, тогда полезность ИРИД при прогнозировании объемов просроченной задолженности по ипотеке будет минимальная или в целом отрицательно скажется на качестве прогнозов. На данный момент нет оснований полагать, что последние технологические изменения изменят поведение пользователей в Интернете, но риск остается. В-третьих, ИРИД является индикатором раннего предупреждения о наступлении риска дефолта только на рынке ипотечного кредитования. Тогда, если отрицательный шок возникает на другом рынке и в другом секторе экономики, а затем впоследствии передается на рынок ИЖК, то ИРИД может выступать в качестве запаздывающего индикатора во время соответствующего спада.

Стоит также отметить, что существуют некоторые специфические характеристики данных онлайн-поиска, которые должны быть приняты во внимание исследователем:

- использовать только четкие и интуитивно понятные ключевые слова и элементы поиска (Ross, 2013);
- учитывать сезонность и цикличность данных (Buono, 2017)
- учитывать экстремальные события, такие как физические явления или политические события, которые могут привести к выбросам (хотя иногда могут быть полезны, поскольку они иллюстрируют поведение экономических агентов) (Vosen, Schmidt, 2011). Вышеописанные характеристики могут использоваться в качестве перспективных идей и направлений для дальнейшей работы. Кроме того, будущим направлением также могут быть использование статистики запросов других поисковых систем (например, Яндекс), а также в социальных сетях (например, Facebook) (Столбов, 2011). Их использование позволит учитывать настроения большего количества потенциальных заемщиков. Кроме того, в дальнейшем возможно построение ИРИД в региональном разрезе. В перспективе исследования возможно также применение более продвинутых эконометрических моделей прогнозирования (например, модели, основанные на искусственном интеллекте или на имитационном моделировании).



## Список использованной литературы

### Нормативно-правовые акты

1. Федеральный закон от 16.07.1998 N 102-ФЗ «Об ипотеке (залоге недвижимости)».
2. Положение Банка России от 6 августа 2015 г. N 483-П «О порядке расчета величины кредитного риска на основе внутренних рейтингов».
3. Положение Банка России от 28 июня 2017 г. № 590-П «О порядке формирования кредитными организациями резервов на возможные потери по ссудам, ссудной и приравненной к ней задолженности».

### Специальная литература

4. Борочкин А. А. Использование статистики поисковых запросов в сети Интернет для краткосрочного прогнозирования макроэкономических переменных // Проблемы и Суждения. 2013. №8. С. 27–32.
5. Зобова, Е. В., Самойлова, С. С. Управление кредитным риском в коммерческих банках // Социально-экономические явления и процессы. 2012. №12 (046), С. 74–81.
6. Канторович Г.Г. Анализ временных рядов. Лекция 7 // Экономический журнал ВШЭ. 2002. Т. 6, № 2. С. 268–273.
7. Лазарян С. С., Герман Н. Е. Прогнозирование текущей динамики ВВП на основе данных поисковых запросов // Финансовый журнал: Макроэкономическое моделирование. 2018. №6. С. 83–94.
8. Солнцев, О. Г., Мамонов, М. Е., Магомедова, З. М. Опыт разработки системы раннего оповещения о финансовых кризисах и прогноз развития банковского сектора. 2011. С. 41–76.
9. Статистический сборник (интернет-версия) «Сведения о рынке жилищного (ипотечного жилищного) кредитования в России №5». 2018. [Электронный ресурс] URL: [https://www.cbr.ru/Collection/Collection/File/15723/Stat\\_digest\\_mortgage\\_05.pdf](https://www.cbr.ru/Collection/Collection/File/15723/Stat_digest_mortgage_05.pdf) (дата обращения 10.05.2020).

10. Столбов М. Статистика поиска в Google как индикатор финансовой конъюнктуры // Вопросы экономики. 2011. №11. С. 79–93.
11. Черникова, Л. И., Евстефеева, С. А. (2016). Просроченная задолженность как индикатор состояния банков // Деньги и кредит. 2016. №5. С. 53–56.
12. Языков, А. Д. Стратегия банка по минимизации потерь в области ипотечного жилищного кредитования // Банковское дело. 2011. №44 (476). С. 37–46.
13. Adachi Y., Masuda M., Takeda F. Google search intensity and its relationship to the returns and liquidity of Japanese startup stocks // Pacific-Basin Finance Journal. 2017. Vol. 46. P. 243-257.
14. Afkhami M., Cormack L., Ghoddusia H. Google search keywords that best predict energy price volatility // Energy Economics. 2017. Vol. 67. P. 17–27.
15. An, X., Deng, Y., Gabriel, S.A. Default Option Exercise over the Financial Crisis and Beyond // Working Paper. 2016. P. 1–54.
16. Anenberg, E., Kung, E. Estimates of the size and source of price declines due to nearby foreclosures // American Economic Review. 2014. Vol. 104 (8), P. 2527–2551.
17. Anvik C., Gjelstad K. Just Google it. Forecasting Norwegian unemployment figures with web queries // Center for Research in Economics and Management (CREAM). 2010. Vol. 11. P. 1–76.
18. Aristei, D., & Gallo, M. The determinants of households' repayment difficulties on mortgage loans: evidence from Italian microdata // International Journal of Consumer Studies. 2016. Vol. 40 (4), P. 453–465.
19. Askitas, N., Zimmermann, K. F. Detecting Mortgage Delinquencies with Google Trends // Institute of Labor Economics (IZA) Discussion Paper. 2011. Vol. 5895. P. 2–17.
20. Bangwayo-Skeete P. F., Skeete R. W. Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach // Tourism Management. 2015. Vol. 46 (C). P. 454-464.
21. Barreira N., Godinho P., Melo P. Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends // Netnomics. 2013. Vol. 14. P.129–165.
22. Bhutta, N., Dokko, J., Shan, H. The Depth of Negative Equity and Mortgage Default Decisions // Finance and Economics Discussion Series Working Paper. 2010. Vol. 35. P. 1–45.
23. Buono D., Mazzi G. L., Kapetanios G., Marcellino M., Papailias F. Big data types for macroeconomic nowcasting // Eurostat Review on National Accounts and Macroeconomic Indicators. 2017. Vol. 1. P. 93–145.

24. Burdeau E., Kintzler E. Assessing the use of Google Trends to predict credit developments // 61st World Statistics Congress of the International Statistical Institute. 2017. P. 1–6.
25. Campos, I., Cortazar G., Reyes T. Modeling and predicting oil VIX: Internet search volume versus traditional variables // *Energy Economics*. 2017. Vol. 66. P. 194–204.
26. Carriere-Swallow Y., Labbé F. Nowcasting with Google Trends in an Emerging Market // *Journal of Forecasting*. 2013. Vol. 32 (4). P. 289–298.
27. Chauvet, M., Gabriel, S., Lutz, C. Mortgage default risk: New evidence from internet search queries // *Journal of Urban Economics*. 2016. Vol. 96, P. 91–111.
28. Chai, T., Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature // *Geoscientific Model Development*. 2014. Vol. 7 (3). P. 1247–1250.
29. Choi, H., Varian, H. Predicting the present with google trends // *Economic Record*. 2012. Vol. 88. P. 2–9.
30. Crook, J., Banasik, J. Forecasting and explaining aggregate consumer credit delinquency behaviour // *International Journal of Forecasting*. 2012. Vol. 28(1), P. 145–160.
31. D'Amuri F., Marcucci J. The predictive power of Google searches in forecasting US unemployment // *International Journal of Forecasting*. 2017. Vol. 33. P. 801–816.
32. Diaz-Serrano, L. Income volatility and residential mortgage delinquency across the EU // *Journal of Housing Economics*. 2005. Vol. 14(3), P. 153–177.
33. Dimpfl T., Jank S. Can Internet search queries help to predict stock market volatility? // *European Financial Management*. 2016. Vol. 22 (2). P. 171–192.
34. Duygan-Bump, B., Grant, C. Household debt repayment behaviour: what role do institutions play? // *Economic Policy*. 2009. Vol. 24, P. 107–140.
35. Foote, C., Gerardi, K., Willen, P. Negative equity and foreclosure: theory and evidence // *Journal of Urban Economic*. 2008. Vol. 64, P. 234–245.
36. Fuster, A., Willen, P. S. Payment Size, Negative Equity, and Mortgage Default // *American Economic Journal: Economic Policy*. 2017. Vol. 9(4), P. 167–191.
37. Gerardi K., Herkenhoff, K.F., Ohanian, L.E., Willen, P.S. Unemployment, Negative Equity, and Strategic Default // *Social Science Research Network (SSRN)*. 2013. Vol. 4. P. 1–50.
38. Gomez-Salvador, R., Lojschova, A., Westermann, T. Household sector borrowing in the euro area – a micro data perspective. // *European Central Bank of Frankfurt Occasional Paper*. 2011. Vol. 125, P. 1–40.

39. Götz B., Knetsch A. Google data in bridge equation models for German GDP // *International Journal of Forecasting*. 2019. Vol. 35. P. 45–66.
40. Hansen, P. R., Timmermann, A. Choice of Sample Split in Out-of-Sample Forecast Evaluation // *Working Paper*. 2012. P. 1–42.
41. Huarng K. H., Yu H. K., Rodriguez-Garcia, M. Qualitative analysis of housing demand using Google trends data // *Economic Research-Ekonomska Istrazivanja*. 2019. Vol. 1. P.409–419.
42. Jappelli, T., Pagano, M., Di Maggio, M. Households' indebtedness and financial fragility // *Journal of Financial Management, Markets and Institutions*. 2013. Vol. 1, P. 23–46.
43. Jun, S. P., Yoo, H. S., Choi, S. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications // *Technological Forecasting and Social Change*. 2018. Vol. 130. P. 69–87.
44. Khandani, A. E., Kim, A. J., Lo, A. W. Consumer credit-risk models via machine-learning algorithms // *Journal of Banking and Finance*. 2010. Vol. 34 (11). P.2767–2787.
45. Li X., Shang W., Wang S., Ma J. A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data // *Electronic Commerce Research and Applications*. 2015. Vol. 14 (2). P. 112–125.
46. Li X., Ma J., Wang S., Zhang X. How does Google search affect trader positions and crude oil prices? // *Economic Modelling*. 2015. Vol. 49. P. 162–171.
47. Li X., Pan B., Law R., Huang X. Forecasting tourism demand with composite search index // *Tourism Management*. 2017. Vol. 59. P. 57–66.
48. Lin, C., Yang, S.-Y. A Forecasting model for the likelihood of delinquency, default or prepayment: the case of Taiwan // *International Journal of Business*. 2003. Vol. 8(2), P. 203–211.
49. Lin W-S., Tou J-C., Lin S-Y., Yeh M-Y. Effects of socioeconomic factors on regional housing prices in the USA // *International Journal of Housing Markets and Analysis*. 2014. Vol. 7. P. 30–41.
50. McLaren N. Using internet search data as economic indicators // *Quarterly Bulletin*. – 2011. Vol. 2. P. 134–140.
51. Miranda, P., McGrath, P. Forecasting aggregate US mortgage delinquencies // *Journal of Applied Financial Research*. 2011. Vol. 1. P. 5–83.
52. Mocetti, S., Viviano, E. Looking behind mortgage delinquencies // *Journal of Banking and Finance*. 2017. Vol. 75, P. 53–63.
53. Mondria, J., Wu, T., Zhangz, Y. The Determinants of International Investment and Attention Allocation // *Journal of International Economics*. 2010. Vol. 82(1), P. 85–95.

54. Moussa F., Delhoumi E., Ouda O. B. Stock return and volatility reactions to information demand and supply // *Research in International Business and Finance*. 2017. Vol. 39. P. 54–67.
55. Parnes, D. Abnormal mortgage delinquencies as housing crisis early symptoms // *International Journal of Housing Markets and Analysis*. 2018. Vol. 11(2), P. 412–432.
56. Peltomaki J., Graham M., Hasselgren A. Investor attention to market categories and market volatility: The case of emerging markets // *Research in International Business and Finance*. 2018. Vol. 44. P. 532–546.
57. Ross, A. Nowcasting with Google Trends: a keyword selection method // *Fraser of Allander Economic Commentary*. 2013. Vol. 37(2). P.54–64.
58. Saxa B. Forecasting Mortgages: Internet Search Data as a Proxy for Mortgage Credit Demand // 4th Research Conference «Addressing Structural Rigidities in View of Monetary Policy Transmission Effectiveness». 2015. P. 107–123.
59. Siliverstovs B., Wochner D. S. Google Trends and reality: Do the proportions match? Appraising the informational value of online search behavior: Evidence from Swiss tourism regions // *Journal of Economic Behavior and Organization*. 2018. Vol. 145 (C). P. 1–23.
60. Smith P. Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data // *Journal of Forecasting*. 2016. Vol. 35 (3). P. 263–284.
61. Suhoy T. Query indices and a 2008 downturn: Israeli data // *Bank of Israel Discussion Paper*. 2009. P. 1–32.
62. Tang W., Zhu L. How security prices respond to a surge in investor attention: Evidence from Google Search of ADRs // *Global Finance Journal*. 2017. Vol. 33. P. 38-50.
63. Veldhuizen, S., Vogt, B., Voogt, B. Internet searches and transactions on the Dutch housing market // *Applied Economics Letters*. 2016. Vol. 23 (18). P. 1321–1324.
64. Vosen, S., Shmidt T. Forecasting private consumption: survey-based indicators vs. Google Trends // *Journal of Forecasting*. 2011. Vol. 30, P. 565-578.
65. Yung K., Nafar N. Investor attention and the expected returns of REITs // *International Review of Economics and Finance*. 2017. Vol. 48. P. 423-439.

Приложение 1. Динамика задолженности по ИЖК (01.01.2013-01.01.2019 гг.)

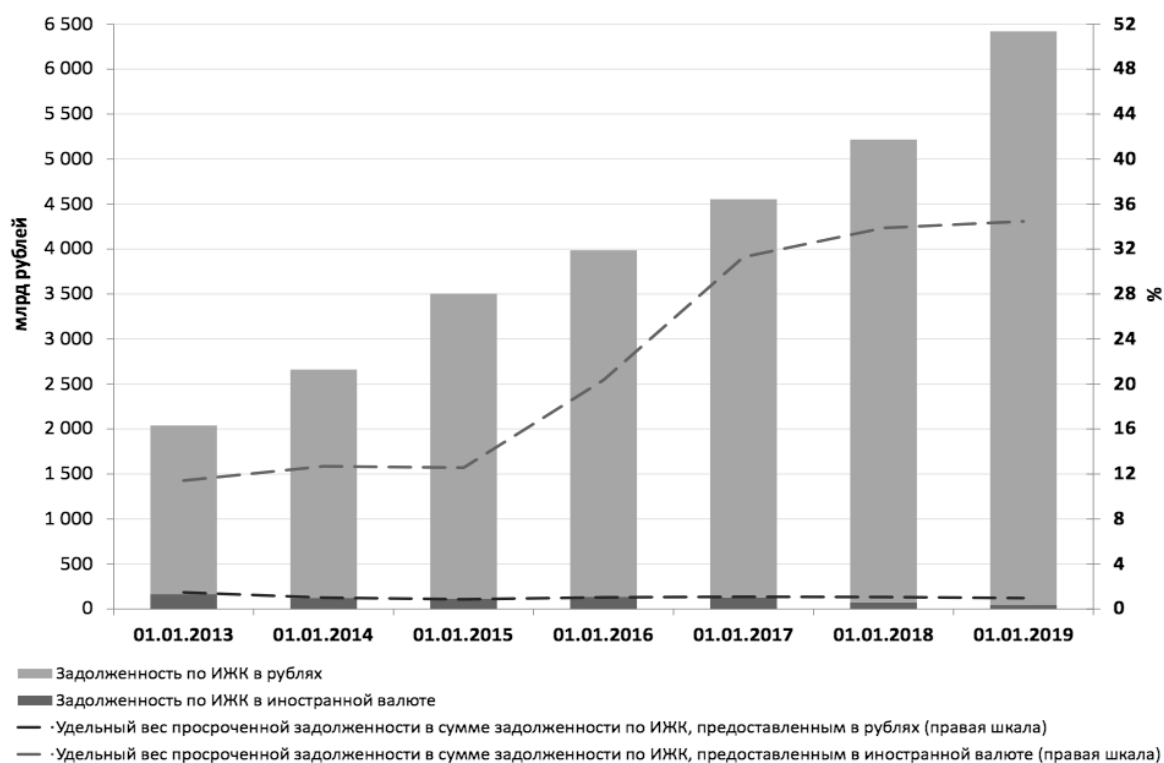


Рис. 1П. Динамика задолженности по ИЖК (01.01.2013-01.01.2019 гг.)

Источник: Аналитические материалы ЦБ РФ «О состоянии рынка ипотечного жилищного кредитования в 2018 г.»

## Приложение 2. Графики динамики контрольных переменных и гистограммы их распределения

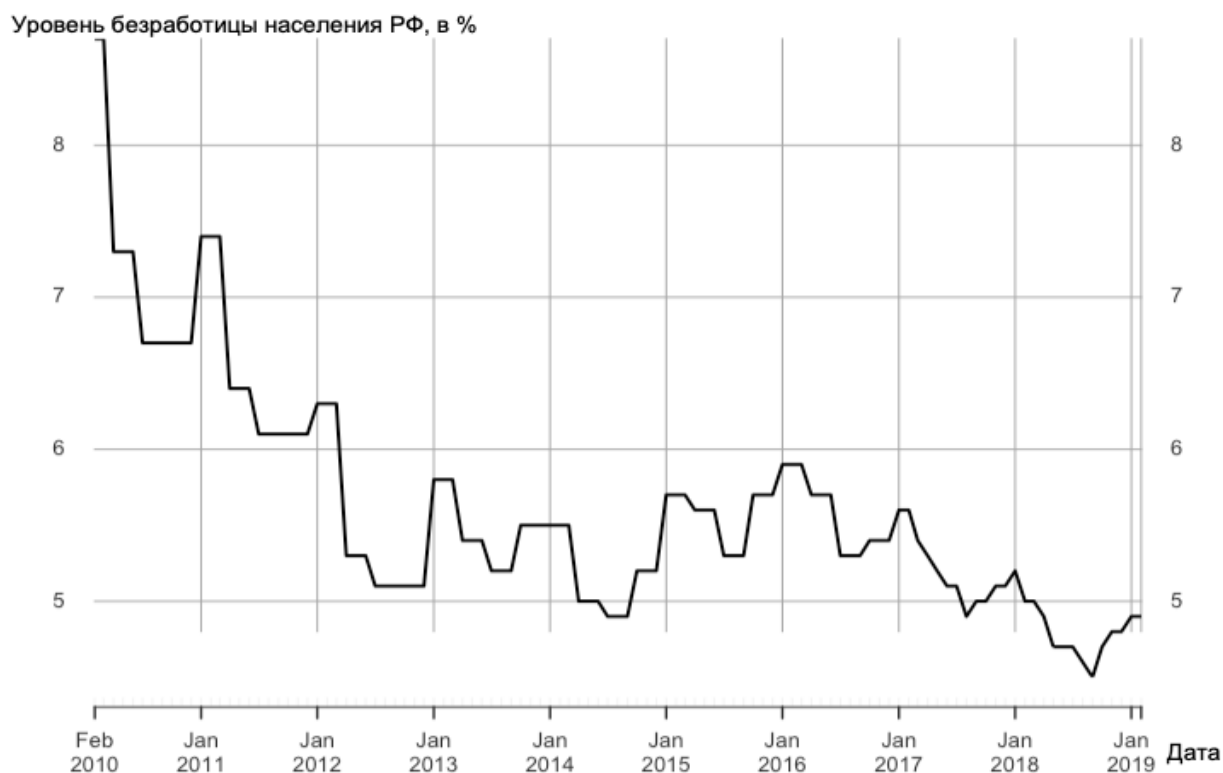


Рис. 2П. Динамика уровня безработицы населения РФ в возрасте 15 - 72 лет, в % (февраль 2010 – февраль 2019)

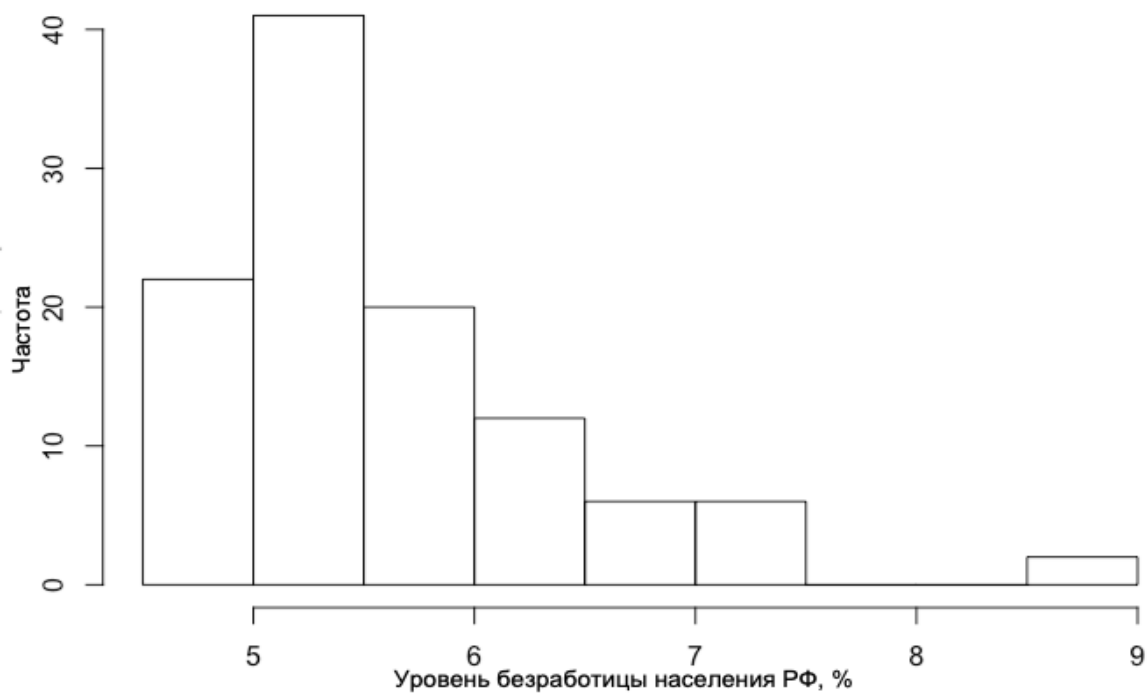


Рис. 3П. Гистограмма распределения уровня безработицы населения РФ

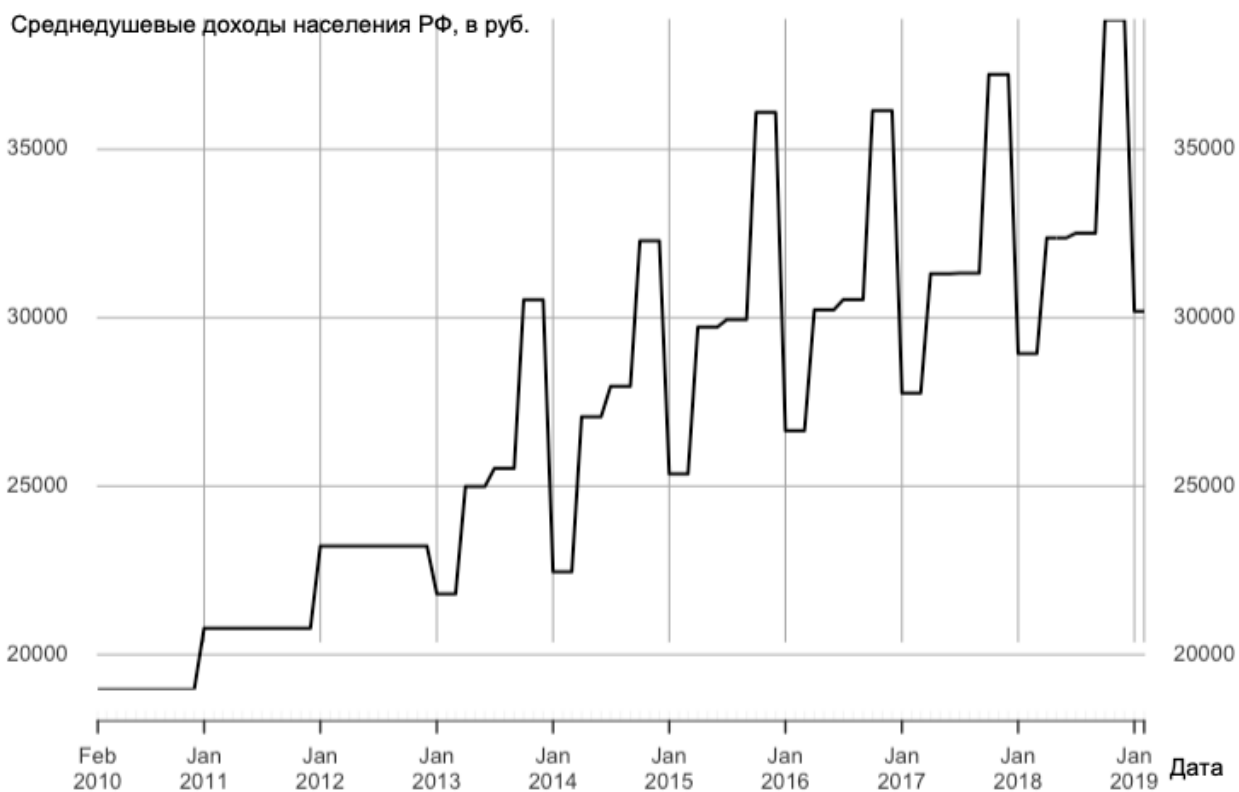


Рис. 4П. Динамика среднедушевых доходов населения РФ, в руб.  
(февраль 2010 – февраль 2019)

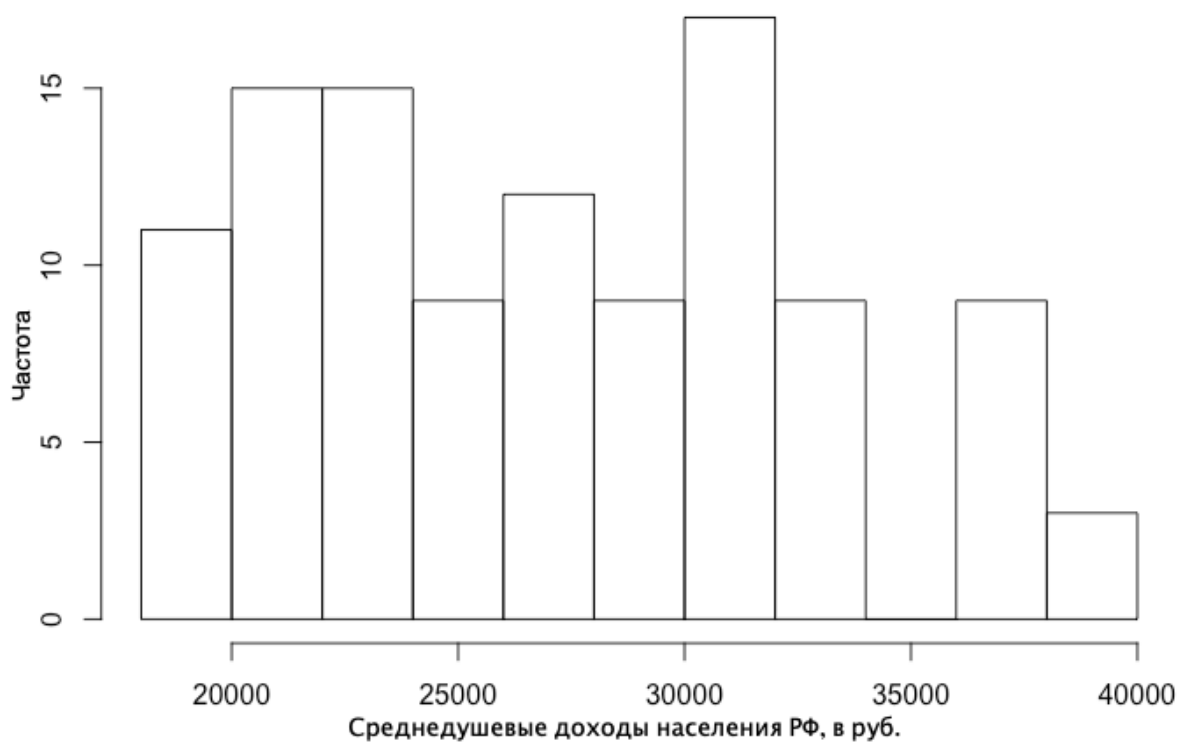


Рис. 5П. Гистограмма распределения среднедушевых доходов населения РФ



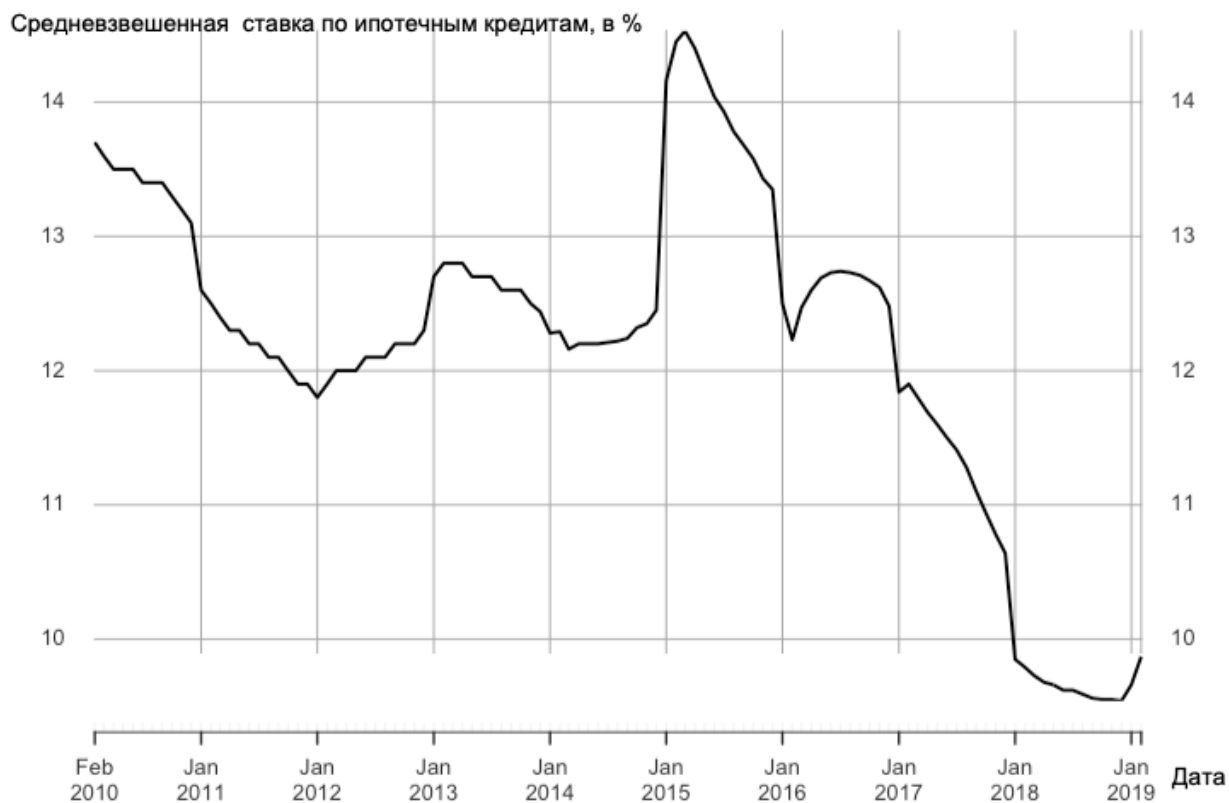


Рис. 6П. Динамика средневзвешенной ставки по ипотечным кредитам, в %  
(февраль 2010 – февраль 2019)

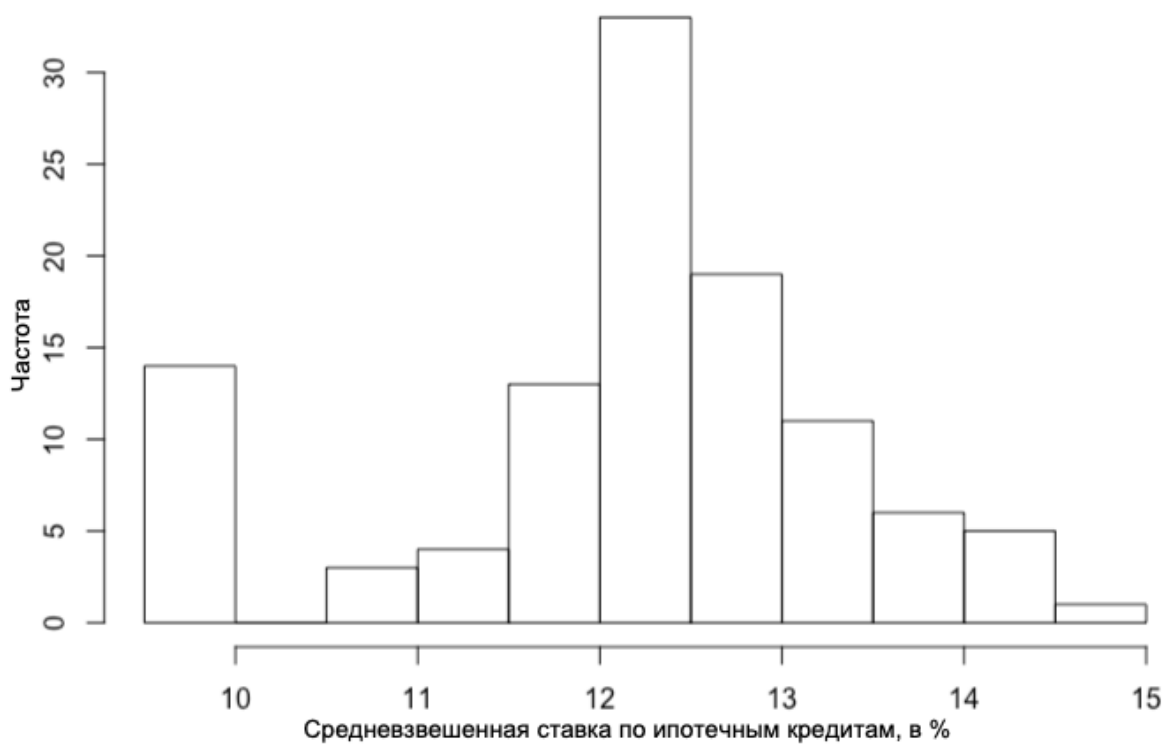


Рис. 7П. Гистограмма распределения средневзвешенной ставки по ипотечным кредитам

Приложение 3. Корреляционная матрица парных коэффициентов корреляции Пирсона

	DELINQ30	DELINQ90	DELINQ180	DELINQ_MORE180	INTEREST	UNEMPL	INCOME	MDRI
DELINQ30	1.00*							
DELINQ90	0.72*	1.00*						
DELINQ180	0.66*	0.91*	1.00*					
DELINQ_MORE180	0.81*	0.69*	0.63*	1.00*				
INTEREST	-0.26*	0.15	0.22**	-0.45*	1.00*			
UNEMPL	-0.29*	-0.03	0.13	-0.38*	0.52*	1.00*		
INCOME	0.68*	0.49*	0.39*	0.72*	-0.42*	-0.67*	1.00*	
MDRI	0.47*	0.35*	0.18***	0.52*	-0.27*	-0.64*	0.65*	1.00*

*Примечание.* Уровни значимости: \*, \*\*, \*\*\* - 10, 5 и 1 % соответственно.

Приложение 4. Результаты ADF теста в процедуре Доладо-Дженкинсона-  
Сосвилла- Ривьеро

Таблица 1

Р-значения ADF теста в процедуре Доладо, Дженкинса,  
Сосвилла-Ривьеро для исходных данных

Переменная	Константа и тренд	Константа	Без константы и тренда
DELINQ30	0.644 *	0.876*	0.858*
DELINQ90	0.396 *	0.371*	0.397 *
DELINQ180	0.685 *	0.571 *	0.354 *
DELINQ_MORE180	0.845 *	0.898 *	0.938 *
MDRI	9.99×10 <sup>-12</sup>	0.07 **	0.82 *
INTEREST	0.835 *	0.762 *	0.214 *
UNEMPL	0.004	0.001	0.04 ***
INCOME	0.950 *	0.457 *	0.999 *

*Примечание:* \*, \*\*, \*\*\* Нулевая гипотеза не отклоняется на 10%, 5%, 1% уровне значимости соответственно.

Таблица 2

Р-значения ADF теста в процедуре Доладо, Дженкинса,  
Сосвилла-Ривьеро для первых разностей

Переменная	Константа и тренд	Константа	Без константы и тренда
Δ DELINQ30	2.45×10 <sup>-9</sup>	2.13×10 <sup>-9</sup>	1.41×10 <sup>-15</sup>
Δ DELINQ90	7.55×10 <sup>-15</sup>	5.10×10 <sup>-16</sup>	6.38×10 <sup>-58</sup>
Δ DELINQ180	1.49×10 <sup>-12</sup>	1.15×10 <sup>-13</sup>	2.27×10 <sup>-9</sup>
Δ DELINQ_MORE180	9.48×10 <sup>-9</sup>	7.51×10 <sup>-10</sup>	3.02×10 <sup>-16</sup>
Δ MDRI	4.09×10 <sup>-12</sup>	9.13×10 <sup>-13</sup>	7.92×10 <sup>-34</sup>
Δ INTEREST	7.42×10 <sup>-9</sup>	5.89×10 <sup>-10</sup>	3.16×10 <sup>-17</sup>
Δ UNEMPL	3.26×10 <sup>-10</sup>	2.46×10 <sup>-11</sup>	5.36×10 <sup>-24</sup>
Δ INCOME	1.59×10 <sup>-7</sup>	1.28×10 <sup>-8</sup>	6.38×10 <sup>-58</sup>

*Примечание:* \*, \*\*, \*\*\* Нулевая гипотеза не отклоняется на 10%, 5%, 1% уровне значимости соответственно.

Приложение 5. Результаты оценивания модели ARIMAX без включения индекса риска ипотечного дефолта для объемов просроченной ипотечной задолженности

Зависимая переменная:				
	$\Delta DELINQ_{30_t}$	$\Delta DELINQ_{90_t}$	$\Delta DELINQ_{180_t}$	$\Delta DELINQ\_MORE_{180_t}$
	(1)	(2)	(3)	(4)
<i>AR</i> (1)	-0.86 *** (0.10)			0.68 *** (0.24)
<i>AR</i> (2)	-1.02 *** (0.14)			
<i>AR</i> (3)	-0.52 *** (0.14)			
<i>AR</i> (4)	-0.28 ** (0.11)			
<i>MA</i> (1)		-0.26 * (0.15)		-0.33 (0.33)
<i>SMA</i> (1)		0.38 *** (0.13)		
Среднедушевые доходы населения РФ				
$\Delta INCOME_t$	1.77 *** (0.55)	-0.05 (0.15)	-0.02 (0.06)	-0.12 (0.11)
$\Delta INCOME_{t-1}$	-1.98 *** (0.47)	-0.45 *** (0.12)	-0.01 (0.05)	-0.29 *** (0.08)
$\Delta INCOME_{t-3}$	0.008 (0.55)	-0.22 (0.21)	-0.03 (0.08)	-0.15 (0.14)
$\Delta INCOME_{t-6}$	-1.08 ** (0.46)	0.03 (0.17)	0.11 (0.07)	-0.09 (0.12)
Уровень безработицы населения РФ				
$\Delta UNEMPL_t$	0.008 ** (3.98×10 <sup>3</sup> )	-452.32 (1321.36)	-220.46 (660.31)	-1198.40 (902.20)
$\Delta UNEMPL_{t-6}$	3.34×10 <sup>3</sup> (3.31×10 <sup>3</sup> )	-176.84 (1036.54)	358.67 (543.83)	-49.22 (902.20)
Средневзвешенная ставка по ипотечным кредитам				
$\Delta INTEREST_t$	6.90×10 <sup>3</sup> (4.50×10 <sup>3</sup> )	453.23 (936.30)	930.42 (570.38)	881.67 (937.74)
$\Delta INTEREST_{t-1}$	-6.04×10 <sup>3</sup> (4.53×10 <sup>3</sup> )	2667.41 *** (907.79)	773.99 (559.21)	679.22 (989.22)
<i>Constant</i>	934.63 ** (418.59)			
Количество наблюдений	98	98	98	98
Логариф функции правдоподобия	-994.23	-841.18	-789.30	-835.75
Информационный критерий Акаике	2019.87	1706.37	1598.60	1698.50

Примечание. В скобках приведены робастные стандартные ошибки. Уровни значимости: \*, \*\*, \*\*\* – 10, 5 и 1 % соответственно.

Приложение 6. Прогнозные значения первой разности объемов просроченной ипотечной задолженности и границы 95% доверительного интервала

Дата	Прогнозное значение первой разности, в млн. руб.	Нижняя граница 95% дов. интервала, в млн. руб.	Верхняя граница 95% дов. интервала, в млн. руб.	Прогнозное значение, млн. руб.	Фактическое значение, млн. руб.
<i>DELINQ30<sub>t</sub></i> , спецификация (1), ARIMA (4,0,0)					
Май 2018	-36 535.89	-60 304.39	-12 767.39	81 899.11	72 085
Июнь 2018	-4 147.21	-35 881.81	27 587.39	77 751.90	56 518
Июль 2018	16 546.92	-15 582.99	48 676.83	94 298.82	106 624
Август 2018	12 129.68	-21 621.28	45 880.64	106 428.50	75 425
Сентябрь 2018	-15 928.14	-49 680.48	17 824.19	90 500.36	112 021
Октябрь 2018	141.63	-33 613.61	33 896.86	90 641.99	104 283
Ноябрь 2018	7 769.66	-26 407.69	41 947.00	98 411.65	73 445
Декабрь 2018	-4 756.34	-38 934.72	29 422.04	93 655.31	106 079
Январь 2019	-16 688.613	-51 187.91	17 810.68	76 966.70	121 416
Февраль 2019	24 676.65	-9 843.52	59 196.82	101 643.35	52 565
<i>DELINQ90<sub>t</sub></i> , спецификация (2), SARIMA (0,0,1) (0,0,1)					
Май 2018	-934.26	-5 116.76	3 248.23	21 902.74	26 558
Июнь 2018	-1 024.45	-5 402.77	3 353.88	20 878.29	22 718
Июль 2018	796.43	-3 581.89	5 174.76	21 674.72	20 309
Август 2018	2 849.25	-1 529.06	7 227.56	24 523.97	24 182
Сентябрь 2018	-1 935.44	-6 313.59	2 442.71	22 588.53	23 460
Октябрь 2018	-1 075.17	-5 453.32	3 302.99	21 513.36	25 771
Ноябрь 2018	-2 322.02	-6 700.17	2 056.13	19 191.34	23 870
Декабрь 2018	715.68	-3 662.47	5 093.84	19 907.02	22 332
Январь 2019	-560.62	-4 938.77	3 817.53	19 346.40	22 249
Февраль 2019	3 133.98	-1 244.18	7 512.13	22 480.38	21 795
<i>DELINQ180<sub>t</sub></i> , спецификация (3), ARIMA (0,0,0)					
Май 2018	230.034	-2 226.19	2 686.27	14 570.04	14 634,00
Июнь 2018	-305.43	-2 761.67	2 150.79	14 264.61	14 961,00
Июль 2018	-1653.42	-4 109.65	802.82	12 611.20	15 796,00

Приложение 6. Прогнозные значения первой разности объемов просроченной ипотечной задолженности и границы 95% доверительного интервала  
(Продолжение)

Дата	Прогнозное значение первой разности, в млн. руб.	Нижняя граница 95% дов. интервала, в млн. руб.	Верхняя граница 95% дов. интервала, в млн. руб.	Прогнозное значение, млн. руб.	Фактическое значение, млн. руб.
<i>DELINQ180<sub>t</sub></i> , спецификация (3), ARIMA (0,0,0)					
Август 2018	-763.03	-3 219.26	1 693.20	11 848,17	14 394
Сентябрь 2018	340.51	-2 115.72	2 796.75	12 188,68	12 316
Октябрь 2018	484.06	-1 972.17	2 940.29	12 672,74	13 185
Ноябрь 2018	-77.82	-2 534.05	2 378.41	12 594,92	14 077
Декабрь 2018	-835.99	-3 292.22	1 620.25	11 758.93	13 648
Январь 2019	313.49	-2 142.75	2 769.72	12 072.42	14 621
Февраль 2019	1 099.48	-1 356.75	3 555.72	13 171.90	13 462
<i>DELINQ_MORE180<sub>t</sub></i> , спецификация (4), ARIMA (1,0,1)					
Май 2018	-360.86	-4 594.40	3 872.69	99 486.14	99 727
Июнь 2018	629.74	-3 742.97	5 002.45	100 115.88	101 955
Июль 2018	1 334.12	-3 144.06	5 812.29	101 450.00	101 546
Август 2018	350.21	-4 208.49	4 908.90	101 800.21	101 911
Сентябрь 2018	-335.74	-4 956.21	4 284.74	101 464.47	102 051
Октябрь 2018	-1 467.22	-6 135.28	3 200.85	99 997.25	103 084
Ноябрь 2018	-804.18	-5 509.01	3 900.65	99 193.07	101 599
Декабрь 2018	301.25	-4 432.03	5 034.53	99 494.32	101 116
Январь 2019	-545.42	-5 300.76	4 209.92	98 948.90	101 710
Февраль 2019	1 596.59	-3 175.87	6 369.05	100 545.49	100 643