



Munich Personal RePEc Archive

Practical Methods for Predicting Customer Retention

Cherkashin, Aleksandr and Sakhadzhi, Vladislav and Guliev,
Ruslan and Bolshunova, Elena

T2 Mobile LLC

22 October 2024

Online at <https://mpra.ub.uni-muenchen.de/123948/>
MPRA Paper No. 123948, posted 16 Mar 2025 20:27 UTC

УДК 338.47:621.395.1:519.246.8:658.8

Черкашин Александр Михайлович

ООО «Т2 Мобайл», Москва, Россия

E-mail: a.cherkashin@t2.ru

Сахаджи Владислав Евгеньевич

ООО «Т2 Мобайл», Москва, Россия

E-mail: vladislav.sakhadzhi@t2.ru

Гулиев Руслан Чингизович

ООО «Т2 Мобайл», Москва, Россия

E-mail: ruslan.guliev@t2.ru

Большунова Елена Васильевна

ООО «Т2 Мобайл», Москва, Россия

E-mail: elena.bolshunova@t2.ru

Практические методы прогнозирования сохранения клиентской базы

Аннотация. В статье рассматриваются методы анализа и прогнозирования сохранения активной абонентской базы в телекоммуникационной отрасли с использованием различных критериев активности абонентов. Результаты исследования показывают, что динамика сохранения активной абонентской базы достаточно точно аппроксимируется убывающей степенной функцией. Это позволяет строить среднесрочные прогнозы по усеченной базе данных об активности абонентов. Вместе с тем нельзя не отметить потенциальные ограничения эффективности применения предложенного подхода для долгосрочного прогнозирования, связанные с изменением динамики оттока абонентов во времени.

Ключевые слова: абонентская база, удержание абонентов, отток абонентов, степенная функция, телекоммуникации, LTV.

Practical Methods for Predicting Customer Retention

Cherkashin Aleksandr

T2 Mobile LLC, Moscow, Russia

E-mail: a.cherkashin@t2.ru

Sakhadzhi Vladislav

T2 Mobile LLC, Moscow, Russia

E-mail: vladislav.sakhadzhi@t2.ru

Guliev Ruslan

T2 Mobile LLC, Moscow, Russia

E-mail: ruslan.guliev@t2.ru

Elena Bolshunova

T2 Mobile LLC, Moscow, Russia

E-mail: elena.bolshunova@t2.ru

Abstract. This study examines methods for analyzing and forecasting the retention of active subscribers in the telecommunications industry using various criteria for subscriber activity. The results demonstrate that the retention dynamics of an active subscriber base can be effectively modeled using a decreasing power function. This allows for medium-term forecasting based on truncated subscriber activity data. However, it is important to note the potential limitations in the effectiveness of the proposed approach for long-term forecasting, associated with changes in subscriber churn dynamics over time.

Keywords: subscriber base, customer retention, customer churn, power law, power function, telecommunications, LTV.

JEL Classification: C53, L96, D12, M31

Введение

Известные исследования в области прогнозирования сохранения и оттока абонентов фокусируются в основном на использовании методов машинного обучения и анализа больших данных (Verhelst et al., 2021; Verbeke et al., 2012). Эти подходы часто требуют значительных вычислительных ресурсов и большого объема данных. При этом они, как правило, направлены на выявление факторов влияющих на склонность абонентов к оттоку (Ribeiro et al., 2023; Jain et al., 2021) или на предсказание поведения конкретных абонентов с целью воздействия на них. Тогда как для задач экономического моделирования нужны точные, но простые в применении методы прогноза поведения абонентской базы в целом или ее крупных сегментов. Например, это необходимо для оценки ожидаемого среднего совокупного дохода от абонента – LTV (Lifetime value) (Gupta & Zeithaml, 2006). Данное исследование сфокусировано на анализе поведения крупных групп абонентов, а также на величине обратной оттоку: сохранении активной абонентской базы или, заимствуя термин из биологии, на выживаемости абонентской базы. Исследование демонстрирует, что сохранение активной абонентской базы достаточно точно моделируется степенными функциями, а их экстраполяция позволяет строить точные прогнозы выживаемости. Но, перед тем как переходить к моделированию сохранения активной абонентской базы необходимо обсудить практические вопросы выбора критериев активности абонентов, так как от этого непосредственно зависит сам показатель и его интерпретация.

1. Признаки активности абонента

Анализ и прогнозирование сохранения активной абонентской базы во многом зависит от выбора критериев, по которым определяется активность абонентов. В этом разделе будут рассмотрены различные подходы к определению активности абонентов и их влияние на оценку выживаемости базы. Предлагается следующая классификация критериев активности:

1. юридические критерии;
2. финансовые критерии;
3. критерии фактического использования услуг.

Каждая из этих категорий имеет свои преимущества и ограничения, которые будут подробно рассмотрены в последующих разделах.

1.1. Юридические критерии активности

Юридические критерии активности основаны на формальном статусе договорных отношений между компанией и абонентом. Основным показателем в этой категории является наличие действующего договора об оказании услуг связи. Но, в сфере массового оказания услуг активность клиента и наличие действующего договора с ним, как правило, мало связаны. Наблюдается значительная задержка между фактическим прекращением использования услуг абонентом и юридическим расторжением договора. Например, многие операторы мобильной связи, включая Т2, предусматривают автоматическое расторжение договора после длительного периода неактивности. Согласно условиям Т2, договор расторгается, если абонент не совер-

шает платных действий в течение 180 дней при нулевом или отрицательном балансе. В некоторых сферах, таких как розничная торговля или разовые услуги, юридические отношения существуют в рамках каждой отдельной транзакции, что делает этот критерий неприменимым для долгосрочного анализа выживаемости. Соответственно, использование юридических критериев для анализа выживаемости абонентской базы имеет существенные ограничения.

Несмотря на ограниченную применимость в телекоммуникационной отрасли, юридические критерии могут быть релевантны для других секторов экономики и правовых систем, где формальный статус договора более тесно связан с фактическим использованием услуг.

1.2. Финансовые критерии активности

Финансовые критерии активности основаны на различных аспектах финансового взаимодействия между абонентом и оператором связи, включая генерацию выручки, платежи абонента, состояние баланса и связанные с обслуживанием затраты. Использование финансовых показателей в качестве критериев активности имеет ряд очевидных преимуществ. Прежде всего, они напрямую отражают экономическую ценность абонента для компании, что особенно важно с точки зрения экономического моделирования.

Однако, несмотря на это преимущество, использование финансовых критериев для оценки выживаемости абонентской базы сопряжено с рядом существенных ограничений. Одним из ключевых факторов, влияющих на достоверность финансовых критериев, является зависимость от учетной политики компании. Моменты начисления выручки могут значительно варьироваться в зависимости от принятых в организации правил. Например, при авансовом платеже за годовое обслуживание выручка может быть начислена единовременно или распределена на весь оплаченный период, что покажет совершенно разную активность абонента. Такая оплата услуг на год вперед может создать, как ложную постоянную активность, даже если абонент фактически прекратил пользоваться услугами через несколько месяцев после оплаты, так и, наоборот, показать полное отсутствие активности за пределами месяца аванса. Ни то, ни другое не даст достоверной информации о реальном поведении абонента, что не позволит использовать эти данные для прогнозирования. Это актуально в том числе в контексте различных маркетинговых акций и специальных предложений, которые могут стимулировать абонентов к долгосрочным предоплатам. Отдельного внимания заслуживает проблема остаточных средств на счетах неактивных абонентов. Если у абонента, переставшего пользоваться услугами, остались средства на счете, их постепенное списание в выручку может показывать ложную активность. Эти явления способны привести к систематическому искажению данных о выживаемости абонентской базы.

Пополнение баланса, как потенциальный критерий активности, с одной стороны, не зависит от учетной политики компании, что является его несомненным преимуществом. Однако оно не обязательно синхронизировано с реальным использованием услуг: абонент может внести крупную сумму и не пользоваться услугами длительное время, что так же приведет к искажению оценки его активности.

Использование затрат в качестве критерия активности сталкивается с теми же проблемами, что и использование выручки, но усугубляется более сложной си-

стемой учета и распределения затрат. Более того, затраты обычно связаны с использованием услуг, что эффективнее измерять напрямую (что будет подробнее описано в следующем разделе, посвященной критериям фактического использования услуг).

Учитывая все вышеизложенное, можно заключить, что несмотря на важность финансовых показателей для экономической оценки абонентской базы, их использование в качестве критериев активности абонента требует крайне осторожного подхода. Эти ограничения могут привести к значительным искажениям в анализе и прогнозировании динамики сохранения активной абонентской базы. Для преодоления описанных ограничений следует тщательно анализировать учетную политику компании перед использованием финансовых критериев. При интерпретации финансовых показателей активности необходимо учитывать специфику бизнес-модели и особенности предоставления услуг. Финансовые критерии, хотя и важны для общей оценки экономической эффективности абонентской базы, не могут считаться универсальным и достаточным инструментом для анализа выживаемости абонентов. Тем не менее, финансовые критерии могут быть эффективны для решения узконаправленных задач (прим.: анализ оплачиваемости абонентской платы в тарифах с фиксированной ежемесячной оплатой).

1.3. Критерии фактического использования услуг

Критерии фактического использования услуг представляют собой наиболее объективный и непосредственный способ оценки активности абонентов в телекоммуникационной отрасли. В отличие от юридических и финансовых критериев, рассмотренных ранее, эти показатели основаны на реальном потреблении услуг связи, что обеспечивает высокую точность и актуальность данных о поведении абонентов. Основное преимущество данных критериев заключается в их независимости от учетной политики компании и других управленческих факторов. Использование услуг связи создает непосредственную нагрузку на сетевую инфраструктуру, отражая реальные физические процессы, которые фиксируются в момент их возникновения. Это позволяет получать актуальные данные о поведении абонентов без временных задержек, характерных для финансовых показателей.

В контексте оператора мобильной связи критерии фактического использования охватывают широкий спектр услуг, включая голосовые вызовы, передачу данных и обмен текстовыми сообщениями. Однако не все виды сетевой активности корректно использовать для оценки выживаемости. Например, технический трафик, генерируемый устройством для поддержания связи с сетью, не может рассматриваться как показатель активности абонента (хотя и свидетельствует об активности устройства). При разработке системы критериев фактического использования услуг необходимо учитывать специфику предоставляемых сервисов и особенности их потребления.

Несмотря на очевидные преимущества, критерии фактического использования услуг тоже имеют определенные ограничения. Основным недостатком является потенциальный разрыв между фактическим использованием услуг и их экономической ценностью для компании. Абонент может активно пользоваться услугами, включенными в пакет, не генерируя дополнительной выручки. С другой стороны, неактивный по этим критериям абонент может регулярно оплачивать услуги, что важно с финансовой точки зрения. Например, абонент может иметь положительный

баланс и регулярно вносить плату, но при этом не пользоваться услугами по различным причинам. Такая ситуация может возникать, например, при наличии у абонента нескольких SIM-карт от разных операторов или при сохранении номера, например, для регистрации на онлайн-платформах. Для преодоления этого ограничения можно оценивать фактическую разницу между абонентами «платящими» и «использующими услуги», и учитывать ее в моделировании. Другой способ – использовать комбинированный подход, сочетающий критерии фактического использования услуг и финансовые критерии.

Выбор конкретного критерия или комбинации критериев должен осуществляться с учетом специфики бизнеса, целей анализа и особенностей исследуемой абонентской базы. Универсального решения, подходящего для всех ситуаций, не существует, и каждой компании необходимо разработать свой подход к определению и измерению выживаемости абонентов. В рамках этой статьи мы будем использовать два специально разработанных критерия активности абонента:

1. Критерий А – учитывает только некоторое фактическое потребление услуг абонентом.
2. Критерий В – комбинированный, включает как фактическое потребление услуг, так и финансовые признаки активности абонента.

2. Соответствие динамики сохранения базы степенным функциям

В контексте телекоммуникационного бизнеса сохранение абонентской базы или выживаемость означает продолжение использования услуг компании абонентом. Сам показатель выживаемости представляет собой долю активных абонентов, в соответствии с выбранным критерием, от изначально подключившихся. Математически выживаемость для каждого периода может быть выражена следующей формулой:

$$R_i = \frac{A_i(T, c)}{G_T} \quad (1)$$

где:

- R_i – выживаемость за i -й период;
- $A_i(T, c)$ – количество активных (в соответствии с критерием активности c) абонентов в периоде i из подключений периода T ;
- G_T – количество подключений в период T ;
- $i \in [1; +\infty]$ – период для расчета выживаемости;
- T – фиксированный начальный период (обычно период подключения);
- c – критерий активности абонента (см. раздел 1).

Расчет и анализ динамики выживаемости, представляют собой когортный анализ (Fader & Hardie, 2009; Zhang & Chang, 2021) и, служат двум основным целям: диагностика текущего состояния и создание основы для построения прогнозных моделей. Для выявления закономерностей в поведении абонентов желательно исследовать динамику выживаемости на значительных временных интервалах. Нами была проанализирована динамика выживаемости абонентов, подключившихся к сети Т2 в ноябре 2015 года, с последующим наблюдением их активности по критерию В на протяжении 105 месяцев (8,5 лет) — до июля 2024 года (Рисунок 1).

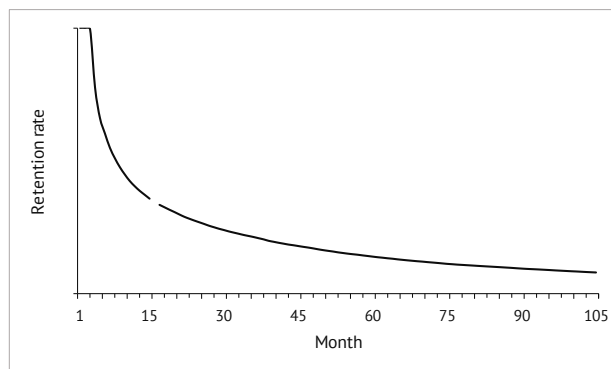


Рисунок 1. В-выживаемость абонентов, подключившихся к Т2 в ноябре 2015 года.

Источник: разработано авторами на основе данных об активности абонентов Т2

Примечание: Рисунок представляет собой графическое изображение временного ряда дискретных значений выживаемости для каждого месяца наблюдений. Здесь и далее дискретный временной ряд изображен в виде непрерывной линии — для наглядности. Вертикальная ось имеет диапазон от 0 до 1 (от 0% до 100%). В рамках статьи для предотвращения раскрытия фактических данных Компании на этом и следующих рисунках убраны подписи значений. Данные по активным абонентам за февраль 2017 года отсутствуют, что выглядит как пропуск на графике.

Визуальный анализ графика демонстрирует монотонный и гладкий характер динамики выживаемости, что позволяет выдвинуть гипотезу о возможности ее аппроксимации относительно простой математической функцией.

Для проверки выдвинутой гипотезы были рассмотрены четыре типа функций:

- степенная: ax^b ;
- экспоненциальная: ae^{bx} ;
- логарифмическая: $a\ln(x) + b$;
- линейная: $ax + b$.

Для простоты расчетов оценка параметров нелинейных функций проводилась после их линеаризации с помощью логарифмирования. Таким образом, параметры функций определялись методом наименьших квадратов (МНК) для соответствующих функций:

- $\ln(y) = \ln(a) + b \ln(x) + \varepsilon$;
- $\ln(y) = \ln(a) + bx + \varepsilon$;
- $y = a\ln(x) + b + \varepsilon$;
- $y = ax + b + \varepsilon$.

Это дает не вполне точную оценку параметров степенного и экспоненциального приближений, так как предполагает мультипликативные остатки для исходных выражений ax^b и ae^{bx} . Тем не менее, мы считаем, их приемлемо использовать для первичной оценки гипотезы и выбора предпочтительной функции.

Результаты аппроксимации представлены на Рисунке 2 и в Таблице 1:

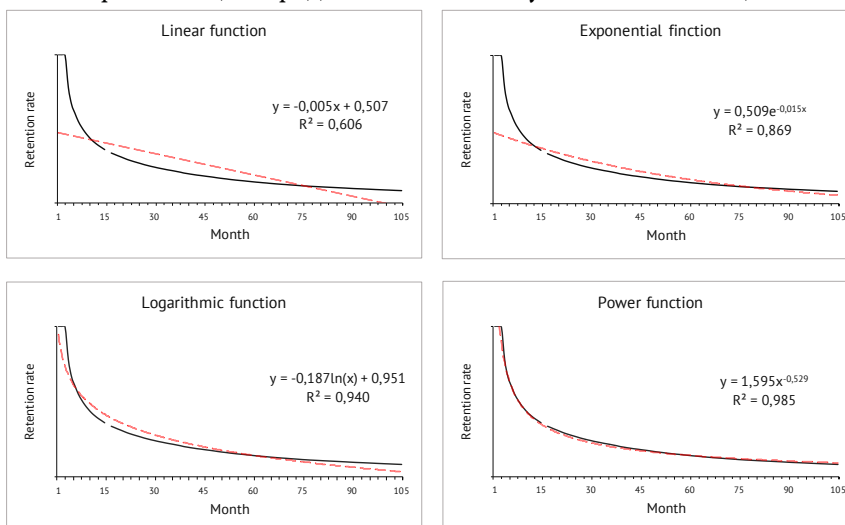


Рисунок 2. Аппроксимация кривой В-выживаемости различными функциями. Источник: разработано авторами на основе данных об активности абонентов Т2.

Анализ результатов показывает, что степенная функция демонстрирует качественную и наилучшую аппроксимацию кривой выживаемости с коэффициентом детерминации (R^2) = 0,985.

Таблица 1

Результаты аппроксимации выживаемости различными функциями

Function type	Function	R^2
Power	$y = 1,595x^{-0.529}$	0,985
Logarithmic	$y = -0,187\ln(x) + 0,951$	0,940
Exponential	$y = 0,509e^{-0.015x}$	0,869
Linear	$y = -0,005x + 0,507$	0,606

Составлено авторами на основе данных об активности абонентов Т2

Для верификации полученных результатов была проведена аппроксимация фактической выживаемости по критерию А для различных групп абонентов, подключившихся к сети компании в разные периоды и через разные каналы продаж. Результаты представлены на Рисунке 3 и в Таблице 2:

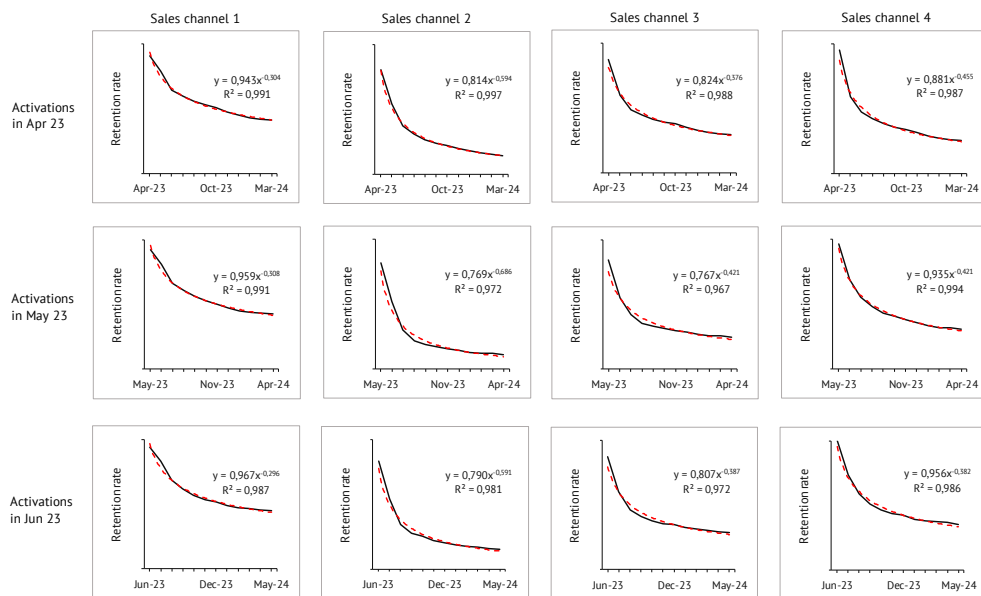


Рисунок 3. Аппроксимация кривых А-выживаемости степенной функцией для различных каналов продаж и периодов подключения. Источник: разработано авторами на основе данных об активности абонентов Т2

Анализ данных Таблицы 2 подтверждает высокую устойчивость результатов аппроксимации степенной функцией. Во всех рассмотренных случаях R² превышает 0,96, что свидетельствует о хорошем соответствии модели эмпирическим данным независимо от канала продаж и периода подключения абонентов.

Таблица 2

Результаты аппроксимации степенной функцией для различных каналов и периодов продаж

Sales channel	Activation month	Function	R ²
1	Apr'23	$y = 0,943 x^{-0,304}$	0,991
	May'23	$y = 0,959 x^{-0,308}$	0,991
	Jun'23	$y = 0,967 x^{-0,296}$	0,987
2	Apr'23	$y = 0,814 x^{-0,594}$	0,997
	May'23	$y = 0,769 x^{-0,686}$	0,972
	Jun'23	$y = 0,790 x^{-0,591}$	0,981
3	Apr'23	$y = 0,824 x^{-0,376}$	0,988
	May'23	$y = 0,767 x^{-0,421}$	0,967
	Jun'23	$y = 0,807 x^{-0,387}$	0,972
4	Apr'23	$y = 0,881 x^{-0,455}$	0,987
	May'23	$y = 0,935 x^{-0,421}$	0,994
	Jun'23	$y = 0,956 x^{-0,382}$	0,986

Составлено авторами на основе данных об активности абонентов Т2

Полученные результаты позволяют сформулировать следующие наблюдения и гипотезы:

1. Универсальность степенной функции: высокая точность аппроксимации выживаемости степенной функцией для различных критериев активности, сегментов абонентской базы и периодов времени может указывать на наличие фундаментальной закономерности процесса сохранения абонентов имеющей вид степенной функции.
2. Зависимость от начальных условий: вариации параметров аппроксимирующей степенной функции для разных групп абонентов могут отражать исключительно влияние начальных условий (канал продаж, период подключения) на долгосрочную динамику выживаемости.
3. Устойчивость к внешним факторам: сохранение формы кривых выживаемости, несмотря на влияние большого количества различных меняющихся факторов (качество услуг, активности по удержанию абонентов, активность конкурентов по привлечению базы). Возможно, это говорит о каком-то динамическом равновесии конкурентного рынка между усилиями компаний по удержанию абонентов и факторами, способствующими их оттоку.
4. Степенные функции широко встречаются при описании закономерностей в самых разных областях окружающего мира: физике, астрономии, биологии, геологии, социологии, лингвистике, психологии, экономике и др. (Andriani & McKelvey, 2007). В исследовании М. Ньюмана (Newman, 2005) рассматриваются степенные законы в таких областях, как физика, биология и экономика, а также их проявления в виде распределения Парето и законе Ципфа. Степенные функции описывают следующие закономерности в экономике и финансах: распределение доходов и богатства, размеры компаний, доходность фондового рынка, объемы торгов, показатели международной торговли и др. (Gabaix, 2009). Преобладание степенной зависимости в описании выживаемости абонентов поднимает ряд фундаментальных вопросов о природе процессов, определяющих поведение абонентов. Например, в исследовании Дж. Стэддона (Staddon, 1978) сделана попытка обосновать возникновение степенных зависимостей в поведении живых организмов при реакции на различные стимулы.

Высокая эффективность аппроксимации выживаемости абонентов степенными функциями открывает возможности для прогнозирования динамики абонентской базы. В следующих разделах будут рассмотрены возможности применения полученных результатов для построения прогнозных моделей и обсуждены ограничения предложенного подхода.

3. Экстраполяция выживаемости и использование ее в качестве прогноза

Выявленная в разделе 2 высокая точность аппроксимации кривых выживаемости степенными функциями открывает перспективы для их использования в прогнозировании. Экстраполяция может позволить оценить выживаемость абонентской базы на длительных временных интервалах без необходимости ожидания фактических данных. Это имеет существенное значение для экономического моделирования и оценки эффективности мероприятий по привлечению абонентов. Прогнозирование выживаемости абонентов играет ключевую роль в расчете таких важных

экономических показателей, как ожидаемая продолжительность жизни абонента (Customer Lifetime) и совокупный доход от абонента (LTV – Customer Lifetime Value) (Kumar, 2014). Эти метрики критически важны для сопоставления с затратами на привлечение абонентов (SAC – Subscriber Acquisition Cost) и, следовательно, для принятия обоснованных решений о запуске, продолжении или прекращении кампаний и механик по наращиванию абонентской базы (Krstevski & Mancheski, 2016).

Для оценки эффективности экстраполяции степенных функций в прогнозировании выживаемости был проведен следующий анализ:

1. использованы данные о выживаемости абонентов по критерию А, подключившихся в апреле 2023 года, за период в 12 месяцев;
2. используя последовательно возрастающее количество фактических значений выживаемости в первые месяцы (от 2 до 11 месяцев) построены приближения степенной функции по методу нелинейной регрессии с аддитивными остатками (см. приложение А);
3. построен прогноз выживаемости как экстраполяция каждой аппроксимирующей функции до 12го месяца; таким образом получены модельные оценки выживаемости по полному периоду в 12 месяцев; при этом каждая модель использует разное количество фактических данных;
4. для каждой модели рассчитан R^2 по данным за 12 месяцев как простая метрика соответствия модельной кривой фактическим данным и, косвенно, как метрика качества прогноза;
5. построен график зависимости значений R^2 каждой модели от количества используемых для ее построения фактических данных.

Результаты анализа представлены на Рисунках 4,5 и в Таблице 3.

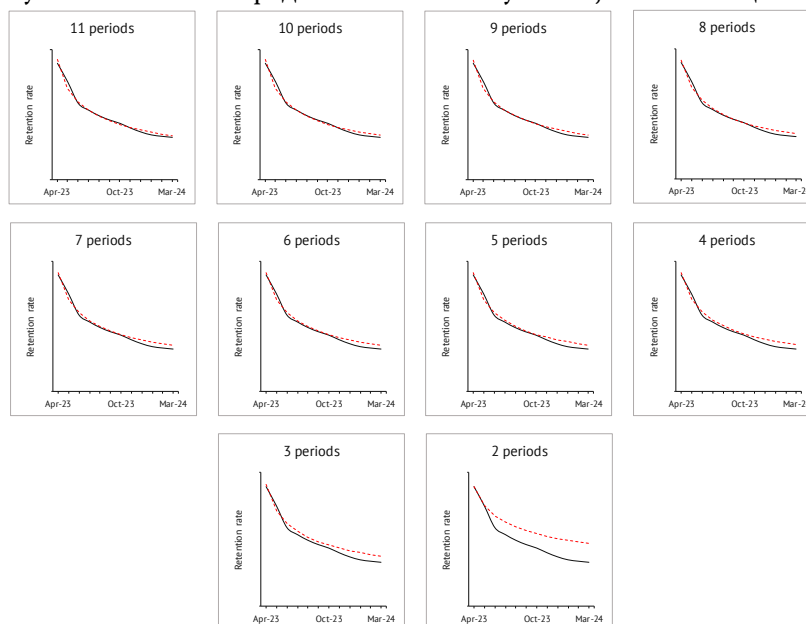


Рисунок 4. Сравнение фактической А-выживаемости с разными прогнозами на основе экстраполяции степенных функций, построенных по разному количеству фактических данных. Источник: разработано авторами на основе данных об активности абонентов T2

Наблюдается резкое повышение качества прогноза при использовании первых трех фактических значений выживаемости, после чего оно показывает лишь незначительное улучшение.

Таблица 3

Параметры аппроксимирующих функций и ошибки прогноза

Actual periods	Forecast periods	Function	R ²
11	1	$y = 0,927x^{-0.370}$	0,988
10	2	$y = 0,925x^{-0.367}$	0,987
9	3	$y = 0,922x^{-0.361}$	0,986
8	4	$y = 0,919x^{-0.356}$	0,984
7	5	$y = 0,917x^{-0.351}$	0,982
6	6	$y = 0,917x^{-0.351}$	0,981
5	7	$y = 0,916x^{-0.348}$	0,980
4	8	$y = 0,915x^{-0.343}$	0,975
3	9	$y = 0,911x^{-0.327}$	0,954
2	10	$y = 0,900x^{-0.241}$	0,575

Составлено авторами на основе данных об активности абонентов Т2

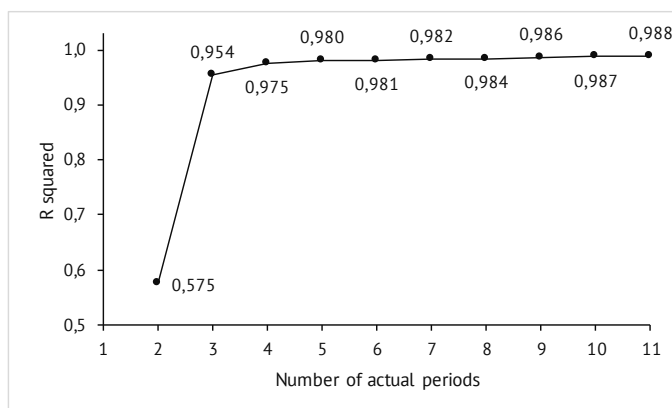


Рисунок 5. Зависимость R² от количества используемых фактических данных в моделях. Источник: разработано авторами на основе данных об активности абонентов Т2

Целью этого анализа было предложить общий подход к оценке качества прогноза, показать применимость прогноза на основе ограниченного набора фактических данных и предложить метод определения необходимого количества фактических данных для его построения. То, что при использовании нескольких фактических значений качество становится достаточно высоким, после чего уже лишь незначительно улучшается, имеет важное практическое значение. Например, можно использовать лишь первые три значения выживаемости для достаточно точной оценки выживаемости в будущем.

Следует отметить, что необязательно брать именно первые несколько фактических значений для экстраполяции. Этот метод дает возможность оценки «гипотетического прошлого» – то есть провести ретроспективный анализ и оценку альтернативных сценариев. Это может быть полезно в случаях, когда требуется проанализировать эффективность уже проведенных акций, направленных на удержание абонентов.

В практических задачах такой прогноз бывает целесообразно строить, используя все доступные значения по выживаемости в конкретный момент. Это могли бы быть значения выживаемости разных групп абонентов, подключившихся в разные месяцы. Например, если мы готовы использовать данные по подключениям всего года, то можно взять выживаемость в первый месяц абонентов, подключившихся в январе, феврале, ..., декабре. Выживаемость на второй месяц абонентов, подключившихся в январе, феврале, ..., ноябре — таким образом, мы получим 12 значений выживаемости в первый месяц, 11 значений на второй и так далее. Это сделает аппроксимацию и прогноз более устойчивым к влиянию сезонности и случайных факторов.

Проведенное исследование демонстрирует высокий потенциал использования экстраполяции степенных функций для прогнозирования выживаемости абонентской базы. Однако, несмотря на выявленную эффективность, данный подход имеет определенные ограничения, которые требуют тщательного рассмотрения. В следующем разделе будет проведен анализ этих ограничений, в частности, будет обсуждаться вопрос долгосрочной применимости степенных функций для моделирования выживаемости.

4. Ограничения применимости

Как было продемонстрировано в разделе 2, степенные функции вида ax^b демонстрируют высокую точность аппроксимации фактической выживаемости абонентской базы. В разделе 3 рассмотрена возможность использования этих функций для построения прогнозов. Однако, несмотря на их эффективность, очевидно, что высокое качество прогнозов, основанных на экстраполяции степенных функций, имеет определенные пределы применимости. С математической точки зрения, при $x \rightarrow +\infty$ функция вида ax^b , при $a > 0$ и $b < 0$, асимптотически стремится к нулю, не пересекая ось X:

$$\lim_{x \rightarrow +\infty} ax^b = 0, \quad \text{при } a > 0, b < 0 \quad (2)$$

Это означает, что в прогнозах, основанных на таких функциях, выживаемость никогда не достигнет нулевого значения. Однако, исходя из общих соображений, очевидно, что такое предположение не может соответствовать реальности. Выживаемость группы абонентов представляет собой агрегированный показатель индивидуальных состояний, каждое из которых в любой момент времени может принимать только два значения: 1 (абонент активен) или 0 (абонент неактивен). Логично предположить, что с течением времени количество активных абонентов будет неуклонно уменьшаться, пока не останется последний активный абонент. После его ухода выживаемость всей группы неизбежно станет равной нулю – событие, которое никогда не будет отражено функцией вида ax^b .

Более того, на выживаемость абонентской базы влияют не только факторы, связанные с желанием клиентов пользоваться услугами компании, но и объективные ограничения, такие как физическая возможность использования услуг связи. Абоненты – это живые люди (за исключением сегмента устройств с SIM-картами, которые также имеют ограниченный срок службы). Неизбежно наступит момент, когда каждый абонент физически не сможет продолжать пользоваться услугами.

Эти рассуждения указывают на потенциальное ограничение в применимости экстраполяции степенных функций для целей долгосрочного прогнозирования.

Для более детального анализа этого ограничения обратимся к графику выживаемости по критерию В и его аппроксимации степенной функцией, представленными на Рисунках 1 и 2. При внимательном рассмотрении области графика, соответствующей более поздним периодам, можно заметить, что, несмотря на высокие показатели общего качества аппроксимации, с увеличением времени наблюдается тенденция к систематической переоценке фактической выживаемости аппроксимирующей кривой.

Для более наглядной демонстрации эффекта переоценки выживаемости построим, во-первых, более точную аппроксимацию (на Рисунке 2 показана аппроксимация подразумевающая мультипликативные остатки), с помощью нелинейной регрессии с аддитивными остатками (см. Приложение А). При этом будем использовать только 3 начальных фактических значения. После чего экстраполируем выживаемость на весь период.

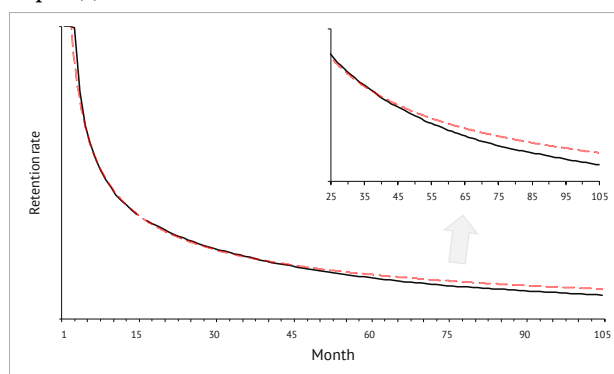


Рисунок 6. Экстраполяция В-выживаемости по начальным фактическим значениям. Источник: разработано авторами на основе данных об активности абонентов Т2

Примечание: ввиду особенностей методологии критерия В первые четыре значения дают искаженное представление о реальной активности абонента, поэтому для экстраполяции были выбраны фактические значения выживаемости на 5, 6 и 7 месяцы. Отклонение в первые несколько месяцев так же объясняются спецификой выбранного критерия активности.

Результаты (Рисунок 6) демонстрируют, что проявляется тенденция к расхождению между прогнозируемой и фактической выживаемостью по мере увеличения возраста этой группы абонентов. Такое наблюдение позволяет выдвинуть гипотезу о возникновении дополнительных факторов, влияющих на выживаемость абонентской базы с течением времени. Как уже говорилось выше, одним из таких факторов может быть физический возраст абонентов, влияние которого на использование услуг связи, предположительно, усиливается с увеличением срока пребывания в абонентской базе.

На основе кривой выживаемости крупных млекопитающих [Deevey, 1947; Pearl & Miner, 1935], в том числе людей (кривая типа I на Рисунке 7), можно сделать предположение об общем виде кривой выживаемости абонента с учетом фактора возраста.

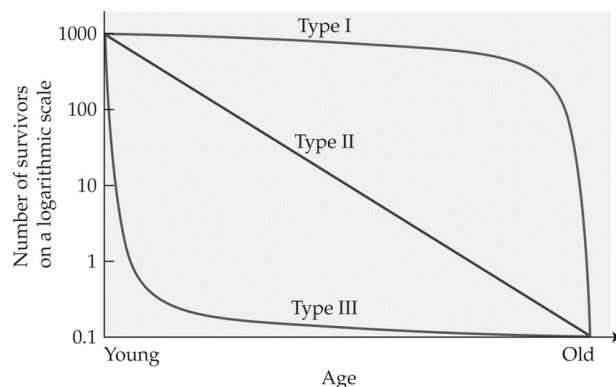


Рисунок 7. Три основных типа кривой выживаемости популяций животных в дикой природе по Э. Диви-мл. Источник: [Hill et al., 2021]

Если рассматривать кривую выживаемости абонента как вероятность сохранения его активности, то фактор физического возраста должен постепенно модифицировать ее в соответствии с кривой типа I (Рисунок 7). Общий вид кривой выживаемости абонента может приобрести S-образную форму, состоящую из двух участков, разделенных промежутком неопределенной длительности (Рисунок 8). Первый участок отражает динамику выживаемости, обусловленную факторами, связанными с качеством услуг и лояльностью. Второй участок демонстрирует влияние иных объективных физических факторов:

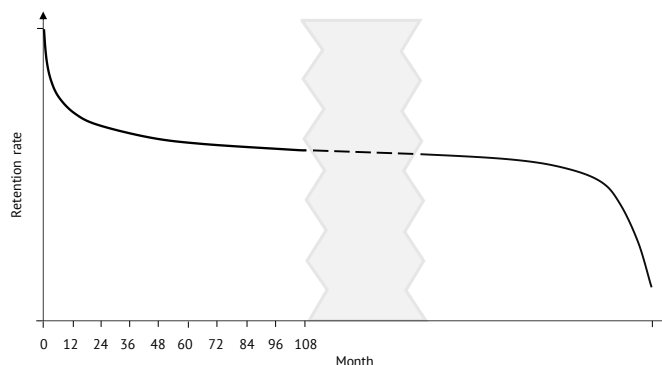


Рисунок 8. Общий вид кривой выживаемости абонента S-образной формы. Источник: разработано авторами

Не исключено, что этот вариант кривой с областями замедления и ускорения мог бы вообще быть общим видом кривой физической выживаемости организмов. Возможно, если подобрать сопоставимые моменты начала наблюдения, то все три типа кривых на Рисунке 7 являются вариантами такой общей кривой, а различия возникают в расположении областей замедления и ускорения, а также в периоде жизни, которому обычно уделяется наибольшее внимание при анализе.

Изучение долгосрочного влияния фактора возраста и других демографических параметров на сохранение абонентской базы представляет собой направление для дальнейших исследований. Такой анализ мог бы способствовать разработке более точных моделей долгосрочного прогнозирования и помочь определить общий вид кривой выживаемости абонента на всем протяжении его жизненного цикла. Тем не менее, как было показано ранее, для задач краткосрочного и среднесрочного прогнозирования этим фактором можно пренебречь.

Заключение

В данном исследовании предложена классификация критериев активности абонентов и обсуждены вопросы их использования на практике. Комбинированные признаки фактического использования услуг, иногда в сочетании с финансовыми признаками, выглядят наиболее универсально для задач моделирования сохранения абонентской базы.

Показано, что динамика сохранения абонентской базы соответствует степенному закону, а именно кривым вида ax^b ($a > 0$ и $b < 0$). Этот факт можно использовать для построения прогнозов сохранения абонентской базы на основе простой экстраполяции, используя, в нашем случае, всего три-четыре фактических значения. Возможность строить простые и точные прогнозы, зная лишь начальные значения выживаемости, находит большую применимость в задачах экономического моделирования и оценки полных совокупных ожидаемых эффектов от подключения абонентов, таких как средний срок жизни абонента и LTV.

Вместе с тем, у долгосрочных прогнозов, основанных на экстраполяции степенных функций есть ограничения, которые предположительно связаны с возрастающим со временем влиянием других факторов, таких как физический возраст абонента. Можно предположить, что общий вид функции выживаемости должен иметь комбинированный вид, в которой начальная часть кривой имеет скорее степенной вид, а конечная – отличается от него.

Данное исследование оставляет открытыми следующие вопросы, каждый из которых требует отдельного изучения:

1. Почему сохранение базы (выживаемость абонентов) так хорошо описывается именно степенными функциями?
2. Почему сохранение базы (и, соответственно, отток) выглядит таким равномерным, и, предположительно, зависящим только от начальных условий? Это проявление динамического равновесия усилий всех участников конкурентного рынка?
3. Какова могла бы быть общая форма долгосрочной кривой выживаемости абонентов?

Список литературы

1. Andriani, P. Beyond Gaussian averages: Redirecting international business and management research toward extreme events and power laws / P. Andriani, B. McKelvey // *Journal of International Business Studies*. – 2007. – №38(7). – P. 1212-1230.
2. Deevey, E. S., Jr. Life tables for natural populations of animals // *The Quarterly Review of Biology*. – 1947. – №22. – P. 283-314.
3. Fader, P. S. Probability models for customer-base analysis / P. S. Fader, B. G. Hardie // *Journal of interactive Marketing*. – 2009. – №23(1). – P. 61-69.
4. Gabaix, X. Power Laws in Economics and Finance // *Annual Review of Economics*. – 2009. – №1. – P. 255-293.
5. Gupta, S. Customer metrics and their impact on financial performance / S. Gupta, V. Zeithaml // *Marketing Science*. – 2006. – №25(6). – P. 718-739.
6. Hill, R. W. Youth in the study of comparative physiology: insights from demography in the wild / R. W. Hill, D. Sleboda, J. J. Millar // *Journal of Comparative Physiology B*. – 2021. – №191. – P. 1-16.
7. Jain, H. Telecom churn prediction and used techniques, datasets and performance measures: a review / H. Jain, A. Khunteta, S. Srivastava // *Telecommunication Systems*. – 2021. – №76(4). – P. 613-630.
8. Krstevski, D. Managerial Accounting: Modeling Customer Lifetime Value – An Application in the Telecommunication Industry / D. Krstevski, G. Manchenski // *European Journal of Business and Social Sciences*. – 2016. – №5(1). – P. 64-77.
9. Kumar, V. Customer lifetime value: The path to profitability. – Boston: Now Publishers, 2008. – 112 p.
10. Newman, M. E. J. Power laws, Pareto distributions and Zipf's law // *Contemporary Physics*. – 2005. – №46(5). – P. 323-351.
11. Pearl, R. Experimental studies on the duration of life. XIV. The comparative mortality of certain lower organisms / R. Pearl, J. R. Miner // *The Quarterly Review of Biology*. – 1935. – №10. – P. 60-79.
12. Ribeiro, H. Determinants of churn in telecommunication services: a systematic literature review / H. Ribeiro, B. Barbosa, A. C. Moreira & R. G. Rodrigues // *Management Review Quarterly*. – 2023. – №74(3). – P. 1327-1364.
13. Staddon, J. Theory of behavioral power functions // *Psychological Review*. – 1978. – №85(4). – P. 305-320.
14. Verhelst, T. Understanding Telecom Customer Churn with Machine Learning: From Prediction to Casual Inference / T. Verhelst, O. Caelen, J. C. Dewitte, B. Lebichot, G. Bontempi // *Artificial Intelligence and Machine Learning*. – 2021. – №1996. – P. 182-200.
15. Verbeke, W. Building comprehensible customer churn prediction models with advanced rule induction techniques / W. Verbeke, D. Martens, C. Mues, B. Baesens // *Expert Systems with Applications*. – 2011. – №38(3). – P. 2354-2364.
16. Zhang J. Z., Chang, C. W. Consumer dynamics: Theories, methods, and emerging directions / J. Z. Zhang, C. W. Chang // *Journal of the Academy of Marketing Science*. – 2021. – №49(1). – P. 166-196.

Приложение А. Оценка параметров нелинейной регрессии степенной функции с аддитивными остатками методом наименьших квадратов

Нелинейная регрессии степенной функции с аддитивными остатками:

$$y_t = ax_t^b + \varepsilon_t, \quad t = 1 \div T$$

Рассмотрим уравнение нелинейной регрессии: $y = ax^b + \varepsilon$

Логарифмируя обе части уравнения получаем: $\ln y = \ln(ax^b + \varepsilon)$,

Обозначим $ax^b = z$, тогда: $\ln y = \ln(z + \varepsilon)$

Используя разложение в ряд Маклорена для функции $f(\varepsilon) = f(0) + f'(0)\varepsilon + O(\varepsilon^2)$, где $f(\varepsilon) = \ln(z + \varepsilon)$, получаем:

$$f(\varepsilon) = \ln(z + \varepsilon) = \ln(z) + \frac{1}{z}\varepsilon + O(\varepsilon^2) \approx \ln(z) + \frac{\varepsilon}{z}$$

Так как $z = ax^b$, то

$$f(\varepsilon) \approx \ln(ax^b) + \frac{\varepsilon}{ax^b} = \ln a + b \ln x + \frac{\varepsilon}{y - \varepsilon} \approx \ln a + b \ln x + \frac{\varepsilon}{y}$$

Таким образом получаем линейную регрессию с гетероскедастичными остатками:

$$\ln y_t = \ln a + b \ln x_t + \frac{\varepsilon_t}{y_t} = \tilde{a} + b \ln x_t + \frac{\varepsilon_t}{y_t}, \quad t = 1 \div T, \tilde{a} = \ln a$$

Регрессионные остатки: $\varepsilon_t = y_t(\ln y_t - \tilde{a} - b \ln x_t)$, $t = 1 \div T$

Сумма квадратов регрессионных остатков:

$$\Sigma(\tilde{a}, b) = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T y_t^2 (\ln y_t - \tilde{a} - b \ln x_t)^2$$

Для нахождения оптимальных параметров необходимо решить систему нормальных уравнений:

$$\begin{aligned} \frac{\partial \Sigma(\tilde{a}, b)}{\partial \tilde{a}} &= - \sum_{t=1}^T 2y_t^2 (\ln y_t - \tilde{a} - b \ln x_t) = 0 \\ \frac{\partial \Sigma(\tilde{a}, b)}{\partial b} &= - \sum_{t=1}^T 2y_t^2 (\ln y_t - \tilde{a} - b \ln x_t) \ln x_t = 0 \end{aligned}$$

Упростив, приходим к системе линейных уравнений:

$$\begin{aligned} \tilde{a} \sum_{t=1}^T y_t^2 + b \sum_{t=1}^T y_t^2 \ln x_t &= \sum_{t=1}^T y_t^2 \ln y_t \\ \tilde{a} \sum_{t=1}^T y_t^2 \ln x_t + b \sum_{t=1}^T y_t^2 (\ln x_t)^2 &= \sum_{t=1}^T y_t^2 \ln y_t \ln x_t \end{aligned}$$

Обозначим суммы:

$$S_1 = \sum_{t=1}^T y_t^2, \quad S_2 = \sum_{t=1}^T y_t^2 \ln x_t, \quad S_3 = \sum_{t=1}^T y_t^2 (\ln x_t)^2, \quad S_4 = \sum_{t=1}^T y_t^2 \ln y_t, \quad S_5 = \sum_{t=1}^T y_t^2 \ln y_t \ln x_t$$

Тогда система уравнений принимает вид:

$$\begin{aligned} \tilde{a}S_1 + bS_2 &= S_4 \\ \tilde{a}S_2 + bS_3 &= S_5 \end{aligned}$$

Решив систему уравнений получим параметры нелинейной регрессии степенной функции с аддитивными остатками:

$$\tilde{a} = \frac{S_4S_3 - S_5S_2}{S_1S_3 - S_2^2}, \quad b = \frac{S_1S_5 - S_4S_2}{S_1S_3 - S_2^2}, \quad a = e^{\tilde{a}}$$