

**Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный институт культуры»**

На правах рукописи

ДЕЛИЦЫН

Леонид Леонидович

**НАУЧНЫЕ ОСНОВЫ РАЗРАБОТКИ И ПРИМЕНЕНИЯ
КОЛИЧЕСТВЕННЫХ МОДЕЛЕЙ РАСПРОСТРАНЕНИЯ
НОВЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

Специальность 05.25.05 – Информационные системы и процессы

ДИССЕРТАЦИЯ

на соискание ученой степени

доктора технических наук

Москва – 2015

Оглавление

ВВЕДЕНИЕ.....	4
ГЛАВА 1. КОЛИЧЕСТВЕННЫЕ МОДЕЛИ РАСПРОСТРАНЕНИЯ СЕТИ ИНТЕРНЕТ	21
1.1 Моделирование и прогнозирование распространения сети Интернет с учетом воспроизводства населения.....	23
1.1.1 Разработка количественной модели распространения нововведения: система уравнений, алгоритм численного решения и процедура идентификации параметров	24
1.1.2 Прогнозирование распространения сети Интернет на основе разработанной количественной модели	32
1.2 Некоторые аналитические модели распространения новых информационных технологий и их приложения	44
1.2.1 Аналитическая модель распространения нововведения с учетом обучения	45
1.2.2 Модель динамики цифрового разрыва.....	51
1.3 Многостадийные модели распространения и интенсивности использования новых информационных технологий.....	57
1.3.1 Совместное моделирование количества информированных, намеренных воспользоваться и пользователей новой технологии	58
1.3.2 Совместное моделирование количества пользователей и абонентов	75
1.3.3 Многостадийная модель охватов аудитории сети Интернет	80
ГЛАВА 2. МОДЕЛИ ИНТЕНСИВНОСТИ ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И РЕСУРСОВ	85
2.1 Количественная модель интенсивности использования сети Интернет.....	87
2.1.1 Разработка модели интенсивности (частоты) использования сети Интернет.....	90
2.1.2 Прогнозирование динамики охватов аудитории сети Интернет на основе разработанной количественной модели.....	114
2.2 Моделирование онлайн-панели при помощи смеси бета-биномиальных распределений... ..	125
2.3 Модель охвата аудитории информационных ресурсов WWW.....	136
2.3.1 Модель поведения пользователей информационного ресурса WWW	139
2.3.2 Способ расчёта охватов аудитории информационных ресурсов WWW по данным счетчиков.....	153
ГЛАВА 3. СОВМЕСТНОЕ МОДЕЛИРОВАНИЕ ПАРКА И РЫНКА ТЕХНИЧЕСКИХ СИСТЕМ	160
3.1 Модели распространения мобильной связи среди индивидов и в домохозяйствах	163
3.1.1 Модели с постоянным потенциалом распространения	163
3.1.2 Модели с линейно растущим потенциалом распространения.....	193
3.2 Совместное моделирование компьютерного рынка и парка компьютеров.....	205
3.2.1 Модель распределения срока службы компьютера	206
3.2.2 Прогнозирование парка и рынка ПК, планшетных компьютеров и сотовых телефонов.....	217

ГЛАВА 4. МОДЕЛЬ ВЫБОРА НОВЫХ ИКТ ДЛЯ ИНВЕСТИРОВАНИЯ ВЕНЧУРНЫМ ФОНДОМ.....	223
4.1 Модель принятия решения об инвестировании в информационные и телекоммуникационные технологии венчурным фондом	224
4.2 Проблемы венчурного инвестирования в новый способ предоставления программного обеспечения	251
ЗАКЛЮЧЕНИЕ	260
СПИСОК ЛИТЕРАТУРЫ.....	264
ПРИЛОЖЕНИЕ А	294
ПРИЛОЖЕНИЕ Б.....	295

ВВЕДЕНИЕ

Актуальность темы исследования

Информационный «взрыв» XXI века сопровождается ускоренной разработкой и быстрым распространением новых технологий, их эволюцией, конкуренцией и замещением. У каждого из таких устройств, как персональный компьютер, мобильный телефон, цифровая камера и пластиковая карточка, сегодня насчитывается более двух миллиардов пользователей. Не менее популярны новые услуги, среди которых – доступ к сети Интернет и информационным ресурсам WWW, электронная коммерция и реклама, видеоконференции, дистанционное обучение, программное обеспечение в виде услуги (SaaS). В то же время завершают жизненный цикл и выходят из употребления широко распространенные ранее технологии: телеграф, факс, компакт-диски, а в ряде стран – и проводная телефонная связь.

Стремительное развитие и быстрая смена технологий¹, изменяя поведение пользователей и потребителей, вынуждают руководителей и специалистов принимать решения в условиях высокой неопределённости, а значительный объём разнородных и противоречащих друг другу статистических показателей усложняет выбор альтернатив. В итоге проблемы оценки современного состояния и прогнозирования распространения новых информационных и коммуникационных технологий (ИКТ), ранее интересовавшие преимущественно производителей компьютерной техники, связистов, финансистов и маркетологов, попадают в поле внимания руководителей и специалистов издательской, библиотечной и образовательной отраслей [293-296, 212].

Основу моделирования и прогнозирования распространения новых информационных технологий по сей день образуют классические модели

¹ М.Я.Дворкина оценивает протяженность периода, в течение которого инновация в современной библиотеке (например, предоставление услуг через сеть Интернет) сохраняет статус новшества, лишь в один-два года по окончании внедренческих работ [59].

диффузии инноваций, созданные около полувека назад и обладающие весьма ограниченными объяснительными и прогностическими возможностями. Разработка и апробация более пригодных для решения практических задач моделей становится возможной благодаря, во-первых, доступности статистических и социологических данных об использовании новых технологий, и, во-вторых, – возросшей вычислительной мощности компьютерной техники и наличию необходимого программного обеспечения. Данная работа посвящена разработке и внедрению количественных моделей распространения ИКТ, которые позволяют решать задачи прогнозирования, анализа сценариев и оценки воздействия регулирования, что существенно расширяет инструментарий поддержки принятия решений руководителями учреждений, организаций и компаний перечисленных выше отраслей.

Объект и предмет исследования

Объектом исследования является процесс распространения новых информационных технологий, который рассматривается как коммуникационный процесс распространения информации о нововведении. Традиционно основными считались два канала распространения сведений – средства массовой информации и межличностное общение, в то время как обучение в образовательном учреждении и в семье ранее оставались за рамками рассмотрения.

Предметом исследования являются количественные модели динамики (изменения во времени) статистических показателей распространения новых информационных технологий. В частности, в работе предложены модели динамики аудитории Интернета, охвата аудитории информационных ресурсов WWW, парка компьютеров, мобильной связи, предоставления программного обеспечения в качестве сервиса (SaaS), обосновано применение таких моделей в качестве инструмента поддержки принятия решений.

Цель и основные задачи исследования

Целью исследования является разработка и апробация количественных моделей распространения новых информационных и телекоммуникационных технологий как инструмента поддержки принятия решений.

В соответствии с поставленной целью в диссертации поставлены и решены следующие **задачи**:

1. Разработать количественную модель распространения нововведения в обществе, сегментированном по полу и типу населенного пункта, учитывающую процессы воспроизводства населения, а также изменение доступности использования нового продукта или услуги во времени. Применить разработанную модель для построения прогноза распространения сети Интернет в России и в различных типах населенных пунктов (города, села). Вывести и использовать аналитические решения уравнений распространения нововведения в частном случае стабильной популяции, для изучения влияния обучения использованию нововведения в семье и в школе на распространение информационных технологий.
2. Разработать и апробировать на данных массовых социологических опросов многостадийную модель роста интенсивности использования нововведения. Построить среднесрочные прогнозы суточного, недельного и месячного охватов аудитории российского Интернета.
3. Разработать и апробировать на данных социологических опросов модель кумулятивных кривых охвата аудитории Интернета, на данных онлайн-панелей и счетчиков – модель кумулятивных кривых охвата аудитории информационных ресурсов WWW.
4. Разработать и апробировать на данных социологических опросов об использовании мобильной связи индивидами и в домохозяйствах количественные модели распространения нововведений, в которых единицей, принимающей решения, является не индивид, а домохозяйство.

5. Разработать модель и прогноз распространения компьютерного парка и рынка России с привлечением социологических данных и онлайн-опросов.
6. Разработать метод выбора новых проектов в сфере ИКТ для венчурного инвестирования. Проанализировать новые направления в предоставлении ИТ-услуг, такие как предоставление программного обеспечения в качестве услуги (SaaS) при помощи разработанного метода и выработать рекомендации для венчурных инвесторов и ИТ-предпринимателей.

Научная новизна исследования состоит в обосновании и последующей разработке комплекса средств анализа сценариев и построения прогнозов распространения новых информационных технологий, в состав которого входят:

- количественная модель распространения новой информационной технологии, учитывающая процессы воспроизводства населения, изменение доступности нововведения с течением времени, а также обучение использованию новой технологии в учебных заведениях;
- аналитические решения уравнений указанной модели, позволившие описать историческую динамику и построить прогнозы распространения Интернета в ряде стран, где статистические данные противоречат предсказаниям классических моделей распространения нововведений;
- многостадийная модель интенсивности использования информационного ресурса, на основе которой построены прогнозы суточного, недельного и месячного охватов аудитории российского Интернета;
- модели интенсивности использования информационных ресурсов в форме вероятностных смесей, позволяющие полностью описать кумулятивные кривые охватов аудитории онлайн-ресурсов, построенные по данным онлайн-панелей и счетчиков;
- модель распространения нововведения, рассматривающая в качестве единицы, принимающей решения, домохозяйство, и при этом описывающая три вида данных: о личном владении сотовыми телефонами,

об использовании сотовой связи индивидом и о наличии сотового телефона в домохозяйстве;

- оценка размера эксплуатационного парка технических систем, использующая функцию распределения времени отказа устройства, полученную с привлечением данных ежегодных онлайн-опросов;
- инструментарий отбора проектов в сфере новых информационных технологий венчурным фондом на основе метода машинного обучения, включающий способ визуализации агрегированных экспертных оценок в виде двумерной карты.

Личное участие автора в получении результатов, изложенных в диссертации:

1. Автор построил систему интегро-дифференциальных уравнений распространения новых информационных технологий с учетом процессов воспроизводства населения и динамики стоимости доступа, разработал программное обеспечение для её численного решения, нашёл некоторые аналитические решения этой системы при помощи элементарных и специальных функций. Автор рассчитал среднесрочные прогнозы полной (полугодовой) аудитории Интернета в России и ряде развитых и развивающихся стран, а также в российских городах и сёлах.
2. Автором построены среднесрочные прогнозы суточного, недельного и месячного охватов аудитории российского Интернета.
3. Автор предложил модель поведения онлайн-панели в виде смеси бета-биномиальных распределений для описания наборов кумулятивных кривых охвата аудитории информационных ресурсов WWW.
4. Автором разработаны модели принятия решения об использовании новой информационной технологии домохозяйством, в которой первым владельцем нового продукта становится глава домохозяйства, а остальные получают продукт в личное владение с определенной задержкой. Автор провел расчеты

по этой модели, и построил прогнозы динамики количества пользователей мобильной связи в России.

5. С целью определить функцию распределения срока службы (т.н. функцию надежности) компьютера автор разработал и с 2004-го по 2008-й год проводил онлайн-опросы о сроке службы компьютера на Интернет-портале Rambler.
6. Автор разработал диаграмму “неопределенность – привлекательность”, моделирующую принятие венчурным фондом решений по предложенным инновационным проектам.

На защиту выносятся следующие **положения**:

- Разработанная количественная модель распространения новой информационной технологии, учитывающая изменение её доступности во времени и процессы воспроизводства населения, может быть применена для описания и прогнозирования динамики количества пользователей сети Интернет среди городского и сельского населения.
- Предлагаемая многостадийная модель интенсивности использования новой информационной технологии позволяет построить прогнозы охватов аудитории сети Интернет за сутки, неделю и месяц.
- Представленные в работе вероятностные модели интенсивности использования информационных ресурсов позволяют описать накопленные охваты аудитории WWW-сайтов (в частности, поисковых систем и порталов), измеренные как с применением онлайн-панелей, так и при помощи счётчиков.
- Моделирование динамики эксплуатационного парка технических систем, объединяющее данные онлайн-опросов о возрастном составе устройств со статистикой поставок, повышает обоснованность оценки величины российского парка устройств доступа к сети Интернет, в частности, персональных компьютеров и мобильных телефонов.

- Инструментарий поддержки принятия решений на основе методов машинного обучения позволяет венчурному фонду повысить релевантность выбираемых аналитиками инвестиционных предложений и ускорить обработку заявок.

Достоверность научных положений, рекомендаций и выводов подтверждается корректным применением методов системного анализа, математического моделирования, статистических методов, принятия решений, репрезентативностью данных массовых социологических опросов и объективностью их анализа.

Методологическая база исследования

Методологической основой моделирования и прогнозирования распространения новых информационных технологий в данной работе служит теория распространения нововведений. Следуя Э.Роджерсу [449], мы определяем нововведение (инновацию) как любую идею, объект, практику, которая воспринимается как новая индивидом или иной единицей (домохозяйством, организацией), принимающей решение об ее использовании, а распространение нововведений (“диффузию инноваций”) как информационный процесс, в ходе которого инновация с течением времени распространяется по определенным каналам между членами социальной системы.

Основными методами, которые применяются в данной работе, являются методы информационного моделирования, системного анализа, прикладной статистики, теории случайных процессов, численные методы решения дифференциальных и интегральных уравнений, методы решения обратных задач и идентификации систем.

Эмпирическая база исследования

Исследование использует данные статистики Международного союза электросвязи (МСЭ), публикации Росстата, Министерства Информационных технологий и связи Российской Федерации, массовых репрезентативных опросов Фонда “Общественное мнение”, ВЦИОМ, Левада-центра, Ромир, GfK, КОМКОН, World Internet Institute, ARD/ZDF, данные маркетинговых исследований IDC, Gartner, IT-Research, а также данные онлайн-панели TNS и интернет-счётчика Rambler’s Top 100 об аудитории российских Интернет-проектов. Кроме того, использованы данные разработанных и проведенных автором в 2004-2008 гг. на Интернет-портале Rambler онлайн-опросов, а также массив экспертных оценок показателей венчурных проектов в сфере ИКТ, выставленных аналитиками инвестиционного фонда ФИНАМ ИТ и студентами МГУКИ в 2005-2011 гг.

Степень разработанности проблемы

В качестве фундамента для построения количественных моделей распространения информационных технологий в нашей работе используется **теория распространения нововведений (диффузии инноваций)**, в частности, работы Ц.Гриликеса [370,371], С.К.Додда [349], Э.Мэнсфилда [413], Дж.Коулмана, И.Катца и Г.Менцеля [337], Э.Роджерса [449], С.Ожги [438], Э.Песемье [443], Р.Фишера и Р.Прая [364], Ф.Басса и соавторов [315,316,317,318,315], П.Леквалла и К.Валбина [401], Р.Диксона [348], В.Махайана и соавторов [406,407,408,410,411,409,412], Дж.Додсона и Э.Мюллера [350], К.Изингвуда и др. [354,356], Ш.Калиша [390], Д.Мидгли [428], В.Камакуры и Ш.Баласубраманьяна [392,392], К.Лоуренса и У.Лоутона [399], Р.Бьюли и Д.Фибига [324], У.Путсиса и С.Шринивасана [518], Дж.Робертса и Дж.Латтина [448], Б.Ратчфорда и др. [447], П.Стоунмана [463], А.Беммаора [322], П.Стеффенса и Д.Мурти [461], М.Хана и др.[512], Д.Джейна и Р.Рао [387], А.Жоланда [388], Х.Скиадаса [455,456], М.Н.Шарифа и соавторов [457,458,459], К.Ван ден Бульте и соавторов [470,471], П.Мейера [425], М.Коэна и др.[336],

Р.Перес и соавторов [442], Дж.Гольденберга и др. [368], С.Молдован и Дж. Гольденберга [430].

В нашей стране уже в 1959 году логистические модели роста были применены Г.Э.Влэдуцем, В.В.Налимовым и Н.И.Стяжкиным [41] при изучении развития науки; с этим их применением отечественный читатель знакомился и в учебниках информатики А.И.Михайлова, А.И.Черного и Р.С.Гиляревского [222] и К.В. Тараканова [269]. Отдельные модели распространения нововведений рассматривались в восьмидесятые годы XX века, а среди более новых публикаций следует отметить работы О.С.Багатуровой и А.Г.Мамиконова [17], К.А.Багриновского и др. [18], Р.Л.Болдырева и др. [26], В.Б.Булгака и др. [28], Л.Е.Варакина [31,32], Л.М.Гельмана и М.И.Левина [49], О.Г.Голиченко [51], В.М.Московкина [226], Р.М.Нижегородцева и соавторов [230,231,232], В.М.Полтеровича и Г.М.Хенкина [244], В.Д. Секерина и др. [259], И.Н.Щепиной [300], Ю.А.Кобленц-Мишке [186,187] и др. В работах А.А. Самарского и А.П.Михайлова [257], В.А.Шведовского [290], А.П.Михайлова и соавторов [176,223,224] а также ряде других публикаций рассмотрены модели распространения информации в обществе. Сущность информации и методы её количественного изучения обсуждается в монографиях Ю.Н.Столярова [266], К.К.Колина [196], А.И.Михайлова и др. [222], К.В.Тараканова [269], Р.С.Гиляревского [50] и ряде других. Исследования распространения нововведений в социальных сетях получили отражение в монографии Д.А.Губанова, Д.А.Новикова и А.Д.Чхартишвили [56].

Пионером в области применения статистики к изучению развития и распространения изобретений стал французский социолог Жан-Габриель Тард, книга которого “Законы подражания” [271], впервые изданная в 1890 году, уже в 1892 была переведена на русский язык. Важный этап развития моделирования распространения инноваций пришёлся на шестидесятые годы прошлого века. Во-первых, опыт изучения распространения инноваций был обобщён Э.Роджерсом в монографии «Диффузия инноваций» [449], которая по сей день является одной из

наиболее цитируемой книг в области изучения коммуникаций. Во-вторых, Ф.Басс модифицировал известное с XIX в. уравнение П.Ф.Ферхюльста [476,477], предложил интерпретацию параметров этого уравнения с позиций теории коммуникаций, и применил его на для прогнозирования новых товаров [315]. Успех прогнозов Ф.Басса и многочисленные публикации его учеников в течение четырех десятилетий привлекли к изучению распространения инноваций значительное число исследователей.

Отметим классическую монографию В.Махайана и Р.Петерсона [408], посвященную количественным моделям распространения нововведений, и изданный в виде монографии детальный обзор достижений последних десятилетий Р.Перес, Э.Мюллера и В.Махайана [442], коллективные монографии под ред. В.Махайана, Э.Мюллера и Й.Уинда [412], и А.Вишваната и Дж.Барнетта [478]. Методам прогнозирования распространения нововведений посвящена серия обзоров Н.Мида и Т.Ислама [421,422], статья П.Паркера [440]. С позиций экономической теории распространение нововведений рассмотрено в монографиях К.Тиртла и В.Руттана [467], П.Стоунмана [463]. Сетевым моделям распространения нововведений посвящена монография Т.Валенте [469]. Монография Р.Бэнкса [314], посвященная явлениям роста и диффузии является, по-видимому, наиболее полной энциклопедией логистических моделей и примеров их применения. Ряд важных работ, посвященных распространению новых технологий, был опубликован исследователями Международного института прикладного системного анализа (IIASA) в Вене (Н.Накиченович [435,434,308], Ч.Марчетти [414,417,415,416], А. Грюблер [375,376,377], В.Петерка и Ф.Флек [444]).

Лишь в последние десять лет появились количественные модели распространения нововведений, учитывающие воспроизводство населения. В частности, в статьях Ф.Центроне, А.Гойя и Е.Салинелли [329,328] найдены точные решения уравнений распространения нововведений для частного случая стабильной популяции. В работе Х.-К.Кортеса, И.Ломбаны и Р.Виллануэвы [338]

модель распространения электронной коммерции в Испании построена на основе подхода Г.Хезкота [381], при котором система интегро-дифференциальных уравнений в частных производных аппроксимируется системой обыкновенных дифференциальных уравнений. За исключением этих работ, в исследованиях распространения нововведений процессы воспроизводства населения игнорируются. Однако эти процессы детально рассмотрены в демографии, экологии и математической теории распространения эпидемий, которая традиционно служит основным источником заимствований коммуникативных моделей распространения нововведений (поскольку, на наш взгляд, является математической теорией межличностных коммуникаций). В этой связи остаются актуальными основополагающие труды А.Лотки [404], В.Вольтерра [42], Р.Пирла и Л.Рида [441], А.Мак Кендрика [477], В.Кермака и А.Мак Кендрика [393], а также работы Н.Бейли [313], Н.Кейфица и Х.Касвелла [394], Ф.Хоппенстедта [382], Р.Гуртина и Р.Мак-Кеми [378,379], Р.Элдеркина [361], О.Дикмана и Дж.Хестербика [346], М.Ианелли, М.Марчевой и Ф.Милнера [384], Ф.Брауэра и К.Кастилло-Чавеса [326], Х.Тиеме [466], Х. фон Фёрстера [479,480], Г.Хезкота [381], Э.де Руса [343], А.Цулариса [468], О.В.Барояна, Л.А.Рвачева и Ю.Г.Иванникова [21], Г.И.Марчука [218], Н.В.Перцева [239,240]. Важными источниками моделей социальных процессов с учетом воспроизводства населения являются монографии Д.Бартоломью [22], В.Вайдлиха [30] и Д.Хелбинга [380]. Среди российских работ следует выделить учебник математической демографии О.В.Староверова [265], монографию Ю.И.Хмелевского [285], посвященную теории самовоспроизводящихся систем, монографии И.Д. Колесина [189,190], Ю.М.Плотинского [242], А.К. Гуц и соавторов [58]. В работах Ю.Н.Гаврильца и соавторов [см., например, 44] предложена модель динамики установки индивидов (предрасположенности к определённому виду поведения), которую мы используем при исследовании изменения интенсивности использования Интернета.

В 1995 г. Я.Нильсен [436] предложил применить модель Ф.Басса [315] и ее расширение, модель NSRL [355], для исследования **распространения Интернета** и WWW (Всемирной Паутины). Среди исследований, использующих модели диффузии инноваций для изучения распространения Интернета в европейских странах, можно отметить работу М.Форнерино [365], в которой проведено сравнение моделей Ф.Басса [315] и NUI [356], причем сделан выбор в пользу последней. Р.Роуз [450], изучая распространение Интернета в России, выделил важное подмножество индивидов, которые не пользуются Сетью, но знакомы с пользователями и подготовлены для восприятия этого нововведения. Глобальные прогнозы распространения Интернета были построены Т.Модисом [429] и С.Арбесманом [307]. Прогнозированию распространения Интернета, в первую очередь – широкополосного, посвящена коллективная монография под ред. Д.Лумиса и Л.Тэйлора [403], прогнозированию инфокоммуникационных технологий – брошюра Л.К. Уэнстона и Дж.Х.Уэнстона [472]. Отметим также работу С.Кииски и М.Похьйола [395], где была получена оценка эластичности потенциала распространения Интернета по отношению к стоимости доступа.

Исследованию распространения Интернета в России посвящены монографии Ю.Ю. Перфильева [238], Г.Л. Смоляна, В.Н. Цыгичко и Д.Д. Хан-Магомедова [262] (см. также монографию В.Н.Цыгичко [286]), исследование коллектива ВНИИПВТИ под руководством В.А. Конявского [199]. Наиболее полное описание источников российских социологических данных об использовании Интернета содержится в монографиях А.В. Чугунова [287,288]. К.К.Колин [196,195] выделил информационное неравенство в качестве важной проблемы XXI века; Е.Л.Вартанова и Я.Н.Засурский [34,33] указывают на сохранение информационного неравенства в течение многих лет даже между европейскими государствами. Преодолению “цифрового разрыва” в обществе посвящена диссертация А.В.Короткова [202]. Ц.Хуанг и Х.Чен [383] предложили способ оценки динамики цифрового разрыва между странами при помощи модели

диффузии нововведений и исследовали различия в использовании Сети домохозяйствами в Великобритании и России.

Распространение Интернета в России рассматривается как процесс диффузии инноваций в книге Ю.Ю.Перфильева [238], при этом автор указал на географическую неоднородность распространения Сети, ее сравнительно высокую доступность в мегаполисах и дороговизну и техническую недоступность в сельских регионах. Диффузионный подход был распространен Ю.В.Деарт на сети связи [60], этим автором и ее коллегами были построены логистические модели и прогнозы распространения сотовой связи и Интернета в России [62,61], важную роль в которых играют межстрановые сопоставления. Классические трех- и четырехпараметрические модели распространения нововведений, помимо уже отмеченных работ, использовались для моделирования и прогнозирования динамики Интернета, мобильной связи и других информационных и телекоммуникационных технологий в работах А.В.Абилова [1], К.Е.Афанасьевой и В.И.Ширяева [12,11], В.Б.Булгака и др. [28], Л.Е.Варакина [32,31], Т.А.Кузовковой и Л.С.Тимошенко [206], С.Ю.Казанцева и И.Э.Фролова [178,178], М.И. Левина и др. [211] и других авторов. Актуальность для библиотечно-информационной отрасли прогнозов распространения таких нововведений, как смартфоны, планшетные компьютеры, читающие устройства, мобильные библиотеки, акцентируется в обобщающих ежегодных докладах Я.Л.Шрайберга [293,296,294,295].

При прогнозировании суточного, недельного и месячного **охвата аудитории информационных ресурсов** мы опирались на модели медиаметрии и поведения потребителей, рассмотренные в работах Р.Мезерингэма [424], Ж.-Л.Шандона [330], П.Данахера и соавторов [340, 341], Й.Парка и П.Фейдера [439], Дж.Ли и Л.Кербак [400], в монографиях Р.Руста [452], Дж.Росситера и П.Данахера [451], Э.Эренберга [360], П.Лифлэнга [514], Г.А.Шматова [291,292], а также в монографии под ред. Ю.Г.Рязанова и Г.А.Шматова [256].

Модель Ф. Басса [315] была применена М.Райтом и др. [481] для описания динамики числа абонентов **мобильной связи** в Новой Зеландии. Х.Груббер и Ф.Вербовен [374, см. также 373] использовали логистическую модель для исследования распространения мобильной связи в странах Европейского союза. Согласно их выводам, наиболее сильно на процесс диффузии повлияли переход от аналоговых стандартов к цифровым и конкуренция операторов связи. Л.Франк [366] использовал логистическую функцию с переменным во времени параметром внутренних воздействий для моделирования распространения мобильной связи в Финляндии. А.Ботельо и Л.Пинто [325] применили логистическую модель для объяснения динамики распространения мобильной связи в Португалии, а Х.Михалакелис с соавторами [426] построили модели динамики сотовой телефонии в Греции при помощи большого количества S-образных кривых. Ф.Центроне и др. [329] применили упоминавшуюся выше модель распространения нововведений с учетом демографических процессов для изучения распространения мобильной связи в ряде стран Европы. Изучая эффект “наверстывания”, Дживанис и Скиадас [367] предложили первую многостадийную модель распространения мобильной связи. Отметим, что в большинстве работ диффузионные модели применялись к данным операторов мобильной связи о количестве абонентов (SIM-карточек). Важный вопрос о связи числа пользователей и абонентов сотовой связи был поставлен лишь в статье И. С. Березина [24].

Распространению технических систем, и в частности – компьютеров, посвящен значительный объем литературы, в частности, работы С.Чоу [333], Ц.Грилихеса и соавторов [372,323], Дж.Ольсона и С.Чои [437], В.Камакуры и Ш.Баласубраманьяна [392], П.Стеффенса [462], П.И.Басс и др. [320,319,321], А.А.Ананьева и И.М. Ильенкова [6], В.В.Васильева и Т.Ю.Салютиной [36], В.Б.Соколова [264], Л.Е.Варшавского [35], Т.А.Кузовковой и Л.С.Тимошенко [206]. Наиболее общие модели обновления технического парка технических систем были разработаны Ю.Н.Григоровым, П.В.Даниловым и Ю.Н.Павловским

[54], см. также [53,236], ряд стохастических моделей технических и социально-экономических систем рассмотрен в монографии К.В.Тараканова и др. [270].

При разработке моделей **отбора инновационных проектов** в сфере информационных технологий мы основывались на работах А.Бретта [27], Дж.Мартино [419], С.Я.Бабаскина [15,16], А.И.Каширина и А.С.Семенова [182], при этом полезными оказались сопоставления с моделями распространения новых информационных технологий в организациях, разработанными Ф.Дэвисом [342], Р.Багоцци и соавторами [311,312], В.Венкатешем и др. [475], Дж. Тенга и др. [465].

Теоретическая значимость работы заключается в

1. Разработке количественной модели распространения новых информационных технологий в неоднородном обществе с учетом демографических процессов и изменения доступности нововведения во времени. Объяснении на основании построенной модели динамики распространения Интернета в ряде случаев, которые не поддается описанию при помощи классических логистических моделей.
2. Создании единого подхода к моделированию охвата аудитории и интенсивности использования сети Интернет и информационных ресурсов WWW при помощи вероятностной смеси непрерывных распределений интенсивности простейших потоков событий.
3. Разработке многостадийной модели роста интенсивности использования информационных ресурсов, в которой на каждой стадии пользователя характеризует непрерывное распределение частоты использования.
4. Обосновании сложного и протяженного характера процесса принятия решений потенциальным потребителем нововведения в сфере ИКТ, в ходе которого первым пользователем нововведения становится домохозяйство (или его глава), а последующие члены семьи получают устройство (или услугу) с задержкой.

Практическая значимость работы

Модели апробированы автором на данных социологических опросов, и были внедрены следующими организациями:

1. ЗАО Артикул Продакшн при разработке прогноза аудитории официального проекта Сочи-2014.
2. ЗАО ФИНАМ при разработке исследований компаний ИТ-сектора в 2006-2012 гг., а также при отборе инновационных проектов в сфере ИКТ в 2006-2012 гг.
3. ООО ЮДУ для прогнозирования количества запросов к сервису.

Соответствующие акты о внедрении приведены в приложении **Б** к диссертации.

Результаты исследования применяются в учебном процессе в Институте информационных коммуникаций и библиотек МГУКИ в читаемых автором курсах “Поисковые системы Интернета”, “Количественные методы исследования социокультурной среды”, “Социология виртуального пространства”, и факультете журналистики МГУ им. М.В.Ломоносова в курсах “Организация, хранение и поиск данных в Интернете” и “Методы компьютерного анализа данных в гуманитарных науках”.

Материалы диссертационного исследования, научные выводы и рекомендации могут быть полезны для дальнейших научных разработок и практической деятельности организаций, компаний, работающих в области в производства и внедрения новых информационных технологий.

Апробация работы

По теме диссертационного исследования автор сделал тридцать два доклада на международных и всероссийских конференциях, кроме того, основные результаты исследования были представлены автором на одиннадцати семинарах в г. Москве. Общее число печатных публикаций автора по теме исследования составляет 121 работу.

Автором опубликованы 23 статьи в изданиях, включенных в перечень ВАК (шесть из них – в соавторстве), 11 работ в журналах, не включенных в перечень (четыре из них – в соавторстве), одна монография.

Кроме того, 65 работ автора (26 статей и 39 тезисов докладов) опубликованы в сборниках трудов международных конференций, 11 работ (4 статьи и 7 тезисов докладов) – в сборниках трудов всероссийских конференций, 13 (8 статей и пять тезисов докладов) – в материалах региональных и институтских конференций и коллективных сборниках. Восемнадцать из этих работ опубликованы автором в соавторстве, остальные – самостоятельно.

Структура и объем диссертации

Диссертационная работа состоит из введения, четырёх глав, заключения, списка литературы, включающего 482 отечественных и зарубежных источника, двух приложений. Объём текста работы составляет 299 страниц, включая 80 рисунков и 12 таблиц.

ГЛАВА 1. КОЛИЧЕСТВЕННЫЕ МОДЕЛИ РАСПРОСТРАНЕНИЯ СЕТИ ИНТЕРНЕТ

Актуальность моделей распространения новых информационных технологий в неоднородной социально-экономической системе вызвана значительными отклонениями данных социологических опросов от предсказаний классических логистических моделей. Опубликованные ранее прогнозы распространения Интернета в России и других странах, построенные на основе классических трехпараметрических логистических моделей [48,147,211], оказались существенно занижены. Простая экстраполяция линейного тренда количества пользователей российского Интернета на период с 2005 по 2009 год [177,178] оказалась точнее логистических моделей. Однако данные о распространении Интернета в развитых странах и в российской столице (см. рис.1.1 и рис.1.4) свидетельствуют о том, что приток пользователей постепенно замедляется, и линейный тренд нельзя успешно эксплуатировать долго.

Кроме того, предсказания логистических моделей противоречат данным социологических и статистических служб. В частности, данные опросов Фонда “Общественное мнение” (ФОМ) об использовании россиянами Интернета с 2002 по 2007 год точнее всего описывает трехпараметрическая логистическая модель с весьма низким потенциалом распространения среди взрослых россиян — 34% [48]. Однако уже в 2006 году более половины россиян в возрасте от 18 до 24 лет пользовалась Интернетом. В таком случае, предсказанной моделью однородного общества потенциал распространения Интернета в 34% подразумевает, что значительная часть россиян, по мере взросления, откажется от использования Всемирной Сети. Такой вывод не подтверждается наблюдениями.

Отметим, что хотя тренд показателей проникновения Интернета во многих странах является S-образным, он не обязательно является логистическим. В этом убеждает рис.1.1. Изображенные в нижней части этого рисунка цепные приросты показателей проникновения, рассчитанные с шагом один год, не образуют

симметричный “колокол”, характерный для логистической функции. Они асимметричны, с резким началом и сравнительно длинным и пологим “хвостом” справа от максимального значения цепного прироста, которое в Швеции и в США было достигнуто в 1997 году. Если бы мы поставили задачу описания данных при помощи одной из S-образных кривых с малым числом параметров, то хорошее приближение на достаточно длительном промежутке времени можно было бы получить при помощи модели Гомперца [348,408,472,473] или NUI [356,365,408].

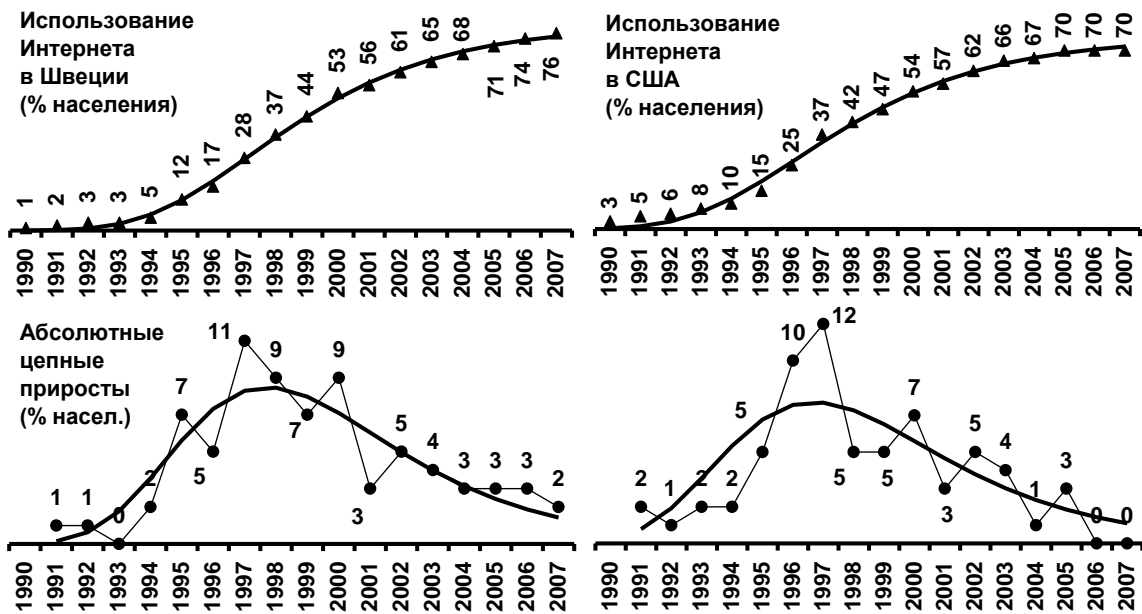


Рис. 1.1. Использование Интернета в Швеции и в США

К сожалению, показатели распространения Интернета в 2008-2013 гг. в Швеции и США модель Гомперца предсказывает столь же неверно, как и логистическая модель. Более того, описанные модели не помогут нам объяснить данные о распространении Интернета в разных возрастных группах или среди лиц с разным уровнем дохода. Существенно больший объем данных можно описать и объяснить при помощи разработанной нами общей модели распространения нововведений, учитывающей процессы воспроизводства населения и изменение доступности нововведения с течением времени [128,301].

1.1 Моделирование и прогнозирование распространения сети Интернет с учетом воспроизводства населения

Мы используем разработанную нами модель распространения Интернета в неоднородном обществе и идентифицируем ее параметры при помощи данных социологических опросов ФОМ [281,282] и ВЦИОМ об использовании Интернета в России в Москве. Разрешение на использование данных опросов ФОМ приведено в **Приложении А** к настоящей работе. Данные ВЦИОМ брались из пресс-выпусков организации.

Одну из проблем, возникающих при практическом применении классических моделей распространения нововведений, легко проиллюстрировать на простом примере. По данным опросов ВЦИОМ, проведенных в сентябре и ноябре 2008 г., доля пользователей Интернета среди взрослых россиян (в возрасте 18 лет и старше) составила от 31% до 35%. По данным ФОМ, собранным осенью 2008 г., полная взрослая аудитория также достигла уровня 30% [234]. В то же время в возрастной группе от 18 до 24 лет проникновение Интернета по данным как ВЦИОМ, так и ФОМ, составило 63%. В сентябре 2009 г. ВЦИОМ сообщил о стабилизации проникновения Интернета среди взрослых россиян на уровне 32%, а ФОМ по итогам летних опросов получил близкую величину, 35%. *Поскольку классические модели распространения нововведений игнорируют воспроизводство населения, полученная при их помощи оценка будущего уровня распространения Сети в России лишь немного превышает 35%. Здравый смысл, однако, подсказывает, что молодые поколения постепенно придут на смену взрослым и пожилым, а следовательно, итоговое проникновение Интернета (или каких-либо иных глобальных компьютерных сетей) окажется по крайней мере не меньше 63%.*

Важным фактором будущего роста распространения Интернета стало подключение в 2006-2008 гг. всех российских школ к Сети в рамках национального проекта «Образование». В случае продолжения действия этой программы, в ближайшие годы российские выпускники будут вступать во

взрослую жизнь, обладая навыками использования Интернета. В благоприятных экономических условиях от использования глобальных компьютерных сетей молодые россияне уже вряд ли откажутся.

1.1.1 Разработка количественной модели распространения нововведения: система уравнений, алгоритм численного решения и процедура идентификации параметров

Система уравнения процесса распространения нововведений с учётом воспроизводства населения

Уравнения процесса распространения нововведений с учетом взросления, рождаемости и смертности, которые можно рассматривать как частный случай общих уравнений социальной динамики, получены нами в совместной работе с Э.А.Юриной [301] и использованы нами в [75,79,68,78,69, 80,128]. Детальный вывод уравнений проведён в [301], здесь мы ограничимся обсуждением переменных и соотношений между ними.

Рассмотрим так называемое атомизованное или атомарное общество, состоящее из индивидов, каждый из которых является лицом, принимающим решения (ЛПР), каждый индивид способен влиять на решение другого, и способность повлиять на чужое решение одинакова для всех. Такая модель является весьма упрощенной, ведь в реальном обществе индивид является членом таких ячеек как домохозяйство, коллектив сотрудников, учебная группа, совокупность жильцов микрорайона и т.п.. В модели атомизованного общества принадлежность к таким группам считается несущественной и не рассматривается. Модели, учитывающие принадлежность индивида к семье и к домохозяйству рассмотрены в четвертой главе данной работы.

Будем предполагать, что общество сегментировано по типу населенного пункта и полу², а в остальном – “атомизовано”, то есть состоит из индивидов, самостоятельно принимающих решения. Каждый индивид может находиться в

² Гендерные аспекты информатизации [5] привлекают значительное внимание исследований, в связи с чем нами были построены прогнозы распространения Интернета отдельно среди мужчин и женщин [66].

одном из трех состояний – невосприимчивости к нововведению, восприимчивости и, наконец, использования нововведения.

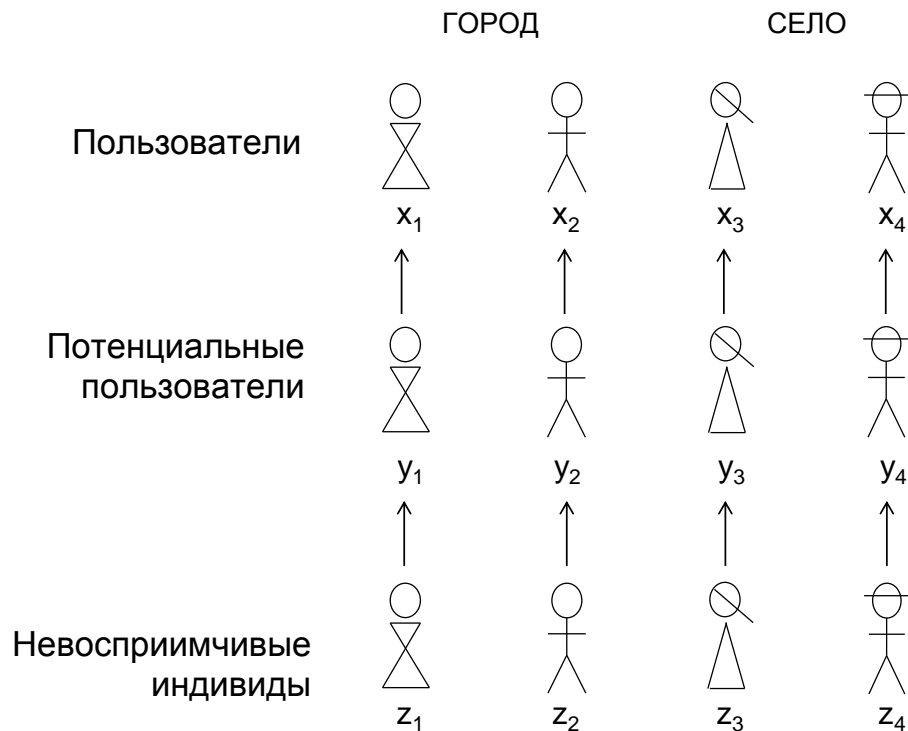


Рис. 1.2. Состояния индивидов и допустимые переходы между ними

Смена состояний происходит однонаправленно (рис.1.2). Индивиды, находящиеся в первом состоянии (невосприимчивые), не имеют возможности использовать нововведение в силу технических, финансовых или иных причин. Индивиды во втором состоянии, которых принято называть потенциальными пользователями, могут использовать нововведение, однако не пользуются им, потому что ещё не слышали о нём, или не приняли положительного решения о его использовании. Переходы невосприимчивых непосредственно в стадию использования нововведения, минуя состояние восприимчивости, не допускаются.

Пусть непрерывная переменная π соответствует моменту рождения индивида. В каждый момент времени t каждый i -й сегмент поколения π разбивается на три непересекающихся класса (подмножества), в зависимости от

состояния, в котором находится индивид. Плотности на единицу возраста количества индивидов, принадлежащих этим классам, удовлетворяют уравнению

$$x_i(t, \pi) + y_i(t, \pi) + z_i(t, \pi) = k_i(t, \pi),$$

где $x_i(t, \pi)$, $x_i(t, \pi)$, $z_i(t, \pi)$, $k_i(t, \pi)$ – плотности на единицу возраста количества пользователей, потенциальных пользователей, представителей «невосприимчивого» класса и населения в целом.

Определим повозрастные потенциалы распространения Интернета как суммы удельных численностей пользователей и потенциальных пользователей нововведения

$$m_i(t, \pi) = x_i(t, \pi) + y_i(t, \pi).$$

Для описания распространения нововведения с учетом взросления, рождаемости и смертности нами была предложена [75,69,128,301] следующая система интегро-дифференциальных уравнений

$$\begin{aligned} \frac{\partial x_i}{\partial t} &= h_i(t, \pi) y_i - \mu_i(t, \pi) x_i \\ \frac{\partial y_i}{\partial t} &= -h_i(t, \pi) y_i + f_i(t, \pi, u(t)) - \mu_i(t, \pi) y_i \\ \frac{\partial z_i}{\partial t} &= -f_i(t, \pi, u(t)) - \mu_i(t, \pi) z_i \end{aligned} \quad (1)$$

с граничными условиями (в общем случае – нелокальными)

$$y_i(t, t) = \sum_{j=1}^J c_{ij} \int_{-\infty}^t B_j(t, \pi) k_j(t, \pi) d\pi, \quad (2)$$

$$x_i(t, t) = 0, \quad z_i(t, t) = 0$$

и ограничениями на параметры $\sum_{i,j=1}^J c_{ij} = 1, \quad c_{ij} \geq 0,$

где $B_j(t, \pi)$ – возрастные коэффициенты рождаемости для поколения π сегмента j в момент времени t ,

$h_i(t, \pi)$ – функция риска ($h_i(t, \pi) \Delta t$ есть вероятность того, что потенциальный пользователь станет пользователем за время Δt , при условии, что еще не был им в момент t),

$\mu_i(t, \pi)$ – носит название силы смертности,

$f_i(t, \pi, u(t))$ – число переходов из «невосприимчивого» класса в класс потенциальных пользователей в единицу времени,

$u(t)$ – «траектория» управляющих параметров (например, цены нововведения) во времени,

c_{ij} – определяют доли сегментов (в частности, мальчиков и девочек) среди новорожденных,

J – количество выделенных сегментов (в данной работе $J = 4$).

Далее за начальный момент времени мы принимаем начало 2001-го года, этот момент не совпадает с началом распространения нововведения. Значения $x_i(t_0, \pi)$, $y_i(t_0, \pi)$, $z_i(t_0, \pi)$, мы не считаем заранее известными.

Мы предполагаем, что влияние коммуникантов на реципиентов при межличностном общении аддитивно, поэтому функция риска имеет вид

$$h_i(t, \pi) = p_i(\pi) + \frac{1}{K(t)} \sum_{j=1}^J \int_{-\infty}^t q_{ij}(\pi, \theta, t) \cdot x_j(t, \theta) d\theta, \quad (3)$$

где $K(t)$ – полная численность населения, π – момент рождения реципиента, θ – момент рождения коммуниканта, $p_i(\pi)$ – коэффициенты внешних воздействий, $q_{ij}(\pi, \theta, t)$ – коэффициенты внутренних воздействий.

Такая форма функции риска распространяет известную модель Ф.Басса [315] на случай зависимости влияния пользователей на потенциальных пользователей от возрастов участников коммуникации.

В этой работе мы полагаем равным нулю коэффициент внешних воздействий $p_i(\pi) \equiv 0$. Мы предполагаем также, что коэффициент внутренних воздействий увеличивается с возрастом реципиента $a = t - \pi$ до достижения совершеннолетия

$$q(\pi, \theta, t) = \begin{cases} q & \text{для } t - \pi \geq 18(\text{лет}) \\ q(t - \pi) & \text{для } 7 \leq t - \pi \leq 18(\text{лет}) \\ 0 & \text{для } t - \pi \leq 7(\text{лет}) \end{cases} \quad (4)$$

Более простая модель с параметром q , не зависящим от возрастов коммуниканта

и реципиента, не позволила нам удовлетворительно описать данные опросов, и была отвергнута.

Алгоритм численного решения уравнений модели

Используемые разностные уравнения для системы (1)-(2), имеют вид [301]:

$$x_j(t + \Delta t) = \left\{ x_j(t) + \left(p_j(t) + q_j(t) \frac{\bar{N}_j(t)}{K(t)} \right) y_j(t) \Delta t \right\} [1 - \mu_j(t) \Delta t]$$

$$y_j(t + \Delta t) = \left\{ y_j(t) - \left(p_j(t) + q_j(t) \frac{\bar{N}_j(t)}{K(t)} \right) y_j(t) \Delta t + f_j(t) \Delta t \right\} [1 - \mu_j(t) \Delta t]$$

$$z_j(t + \Delta t) = \{ z_j(t) - f_j(t) \Delta t \} [1 - \mu_j(t) \Delta t],$$

где сила смертности $\mu_j(t)$ вводится отдельно для каждого поколения

$$\mu_j(t) \approx \frac{1}{\Delta t} \left[1 - \frac{l_j(t + \Delta t)}{l_j(t)} \right],$$

а $l_j(t)$ – функция дожития (вероятность, что к моменту времени t индивид жив).

Функция влияния пользователей на представителя j -го поколения линейна:

$$\bar{N}_j(t) = \sum_i \beta_{ij} x_i(t).$$

Для учета рождаемости при переходе к каждому следующему моменту времени требуется добавлять когорту новорожденных

$$y_{j(t)}(t + \Delta t) = \sum_{i=i \min}^{i \max} B_i(t) K_i(t),$$

где фертильности $B_i(t)$ учитывают рождаемость в различных поколениях.

Последняя формула использует три упрощающих предположения. Во-первых, мы предполагаем, что дети любого из сегментов рождаются потенциальными пользователями и не попадают в недоступное подмножество. Разумеется, это верно не для любого общества, однако мы надеемся, что для российского общества такое предположение допустимо. Во-вторых, мы предполагаем, что рождаемость не зависит от того, к какому подмножеству – x , y или z – принадлежат родители. Наконец, и особенно важно – предполагается, что в атомизованном обществе, принятие решений об использовании нововведений

детьми не зависит от того, являются ли пользователями их родители. В следующих разделах данной главы последнее предположение будет обсуждаться более детально.

Общее число принимаемых во внимание когорт $J(t)$ на каждом шаге вычислений следует увеличивать на единицу:

$$J(t + \Delta t) = J(t) + 1.$$

Выделение большого числа сегментов пользователей приводит к необходимости решать интегро-дифференциальные уравнения баланса (1-4) относительно большого числа переменных. Для идентификации параметров системы (1) необходимо минимизировать функционал ошибки, что, зачастую требует решить саму систему несколько сотен или даже тысяч раз. Наконец, для оценки доверительных интервалов при помощи метода Монте-Карло [3,344] (такой способ был применён в нашей кандидатской диссертации [163]) или при помощи метода бутстреп (bootstrap) [358], требуется несколько тысяч раз повторить процедуру минимизации.

При условии векторизации уравнений (2) возможно осуществить достаточно быстрое решение столь большой системы при помощи популярных программных пакетов. Для этого необходимо заменить обращение к отдельным элементам матриц операциями с матрицами в целом или, если это невозможно, с подматрицами. Известно, что иногда при этом достигается уменьшение времени счета задачи в сотни раз.

Ниже приведен пример фрагмент кода на языке свободно распространяемого пакета SCILAB³, содержащий векторизованную разностную версию уравнений (1). При этом используется матричное умножение и поэлементное умножения “.*”. В данной работе мы использовали простейшую разностную схему Эйлера; в дальнейшем мы предполагаем использовать метод Р.Е. Миккенса [516].

³ См. <http://www.scilab.org/>

```

for it = 1:nt
    // Расчет функции риска
    popfct = 1/sum(P) ;
    HMAT = QAge . * ( PMAT + popfct * CommX * X * QMAT );

    // Блок диффузии нововведения
    DY = Y .* HMAT;                DZ = Z .* LMAT ;
    Y = Y - DY + DZ;      X = X + DY;      Z = Z - DZ;

    // Учет интернетизации школ
    if it == ktschool then
        Y(ks1:ks2,:) = 0;                Z(ks1:ks2,:) = 0;
        X(ks1:ks2,:) = P(ks1:ks2,:);
    elseif it > ktschool then
        Y(ks1,:) = 0; Z(ks1,:) = 0; X(ks1,:) = P(ks1,:);
    end;

    // Расчет рождаемости
    BVECT = OneV *( FERTILITY .* P );
    BVECT = mbabies * [ BVECT(2:$), 0 ] + ( 1 - mbabies ) * BVECT;
    P(2:$,:) = P(1:$-1,:) .* R;  P(1,:) = BVECT;

    // Учет рождаемости и смертности
    X = [ ZeroV; X(1:$-1,:) .* R ];
    Y = [ BVECT; Y(1:$-1,:) .* R ];
    Z = [ ZeroV; Z(1:$-1,:) .* R ];

    // Сохранить суммы значений X(t) для вывода
    XsumVect( : , : , it + 1 ) = SMAT * X;
end

```

При расчетах мы использовали ПК с процессором Intel Core™ 2 Duo и 3,25 ГБ ОЗУ под управлением Windows XP Professional версии 2002 SP 3. Время работы содержащей приведенный код программы при указанном выше числе переменных и 120 шагах времени составило в среднем 1,8 секунды [151]. Скромное время решения системы (2) позволяет в течение получаса подобрать неизвестные параметры модели при помощи процедур нелинейной оптимизации,

когда число вызовов минимизируемой функции составляет две-три тысячи. В дальнейшем при расчётах мы использовали не систему SCILAB в ОС Windows, а интерпретируемый язык Python с расширением NumPy [см., например, 29] и операционную систему Linux, что связано с большей распространённостью Python, что повышает шансы развивать наше программное обеспечение. Отметим, что в наших расчётах использовалась сегментация по полу и по типу населённого пункта (т.е. рассматривались мужчины и женщины, проживающие в городах и сёлах).

Идентификация параметров модели

Три вида параметров нашей модели являются свободными:

- 1) начальные значения потенциалов $m_i(t_0, \pi)$ для рожденных до t_0 ;
- 2) начальные показатели распространения нововведения $x_i(t_0, \pi)$;
- 3) неотрицательный параметр внутренних воздействий q .

Мы предполагаем, что все россияне, рождённые после 2000 года, потенциально способны стать пользователями Интернета, поэтому постулируем $m_i(\pi, \pi) = k_i(\pi, \pi)$ для всех $\pi \geq t_0$.

Зависимости $m_i(\pi, \pi)$ и $x_i(0, \pi)$ были аппроксимированы кусочно-линейными функциями с семью узлами. В итоге общее число неизвестных параметров модели \mathbf{m} составило 15 для каждого сегмента (сочетания пола и региона), в то время как совокупное число измерений ФОМ составило 166 для каждого сегмента.

Для сравнения с данными ФОМ, которые относятся к возрастным отрезкам от 18 до 24 лет, от 25 до 34 лет и т.д. (см.рис.1.3 и рис.1.4), расчетные значения $x_i(t, \pi)$ суммируются внутри соответствующих отрезков

$$\Delta x_i^{(n)}(t_j; \mathbf{m}) = \int_{\pi_j}^{\pi_{j+1}} x(t_j, \pi) d\pi. \quad (5)$$

Определение неизвестных параметров проведено методом наименьших квадратов при помощи минимизации суммы квадратов расхождений между данными наблюдений $d_{i,l}^n$ и расчетными значениями

$$\Phi(\mathbf{m}) = \sum_{n,l} w_{n,l} |d_{i,l}^n - \Delta x_i^{(n)}(t_l; \mathbf{m})|^2. \quad (6)$$

После идентификации параметров модели появляется возможность построения прогнозов величин показателей $\Delta x_i^{(n)}(t_l; \mathbf{m})$. В классических моделях диффузии нововведений часто предполагается, что нововведение доступно всем индивидам с самого начала, поэтому редко ставится вспомогательная задача прогнозирования динамики невосприимчивого класса $z_i(t, \pi)$ или, эквивалентно, динамики потенциалов $m_i(t, \pi)$. Однако в случае распространения Интернета такое предположение является чрезмерно грубым. Среди определяющих динамику $m_i(t, \pi)$ факторов можно выделить доступность услуги доступа к Сети, её стоимость и уровень доходов населения.

1.1.2 Прогнозирование распространения Интернета на основе разработанной количественной модели

Учёт подключения учебных заведений к Интернету

Первый из опубликованных нами прогнозов (рис.1.3) учитывал интернетизацию школ, но предполагал консервацию текущей стоимости доступа к Интернету для частных лиц [301]. Для упрощения расчетов предполагалось, что с осени 2007 года *все* подростки в возрасте от 12 до 18 лет начали использовать Интернет и, более того, затягивать в Сеть взрослых. Это предположение оказалось слишком смелым. По данным опросов, проведенных ФОМ летом 2008 года, лишь 75,5% подростков в возрасте от 14 до 17 лет используют Интернет хотя бы раз в полгода, дома или в школе. Этот показатель возрастал в Москве до 97,4%, и убывал с размерами населенного пункта, снижаясь до 59,9% в селах. Около трети (32%) учащихся старше 12 лет заявили, что не имеют возможности пользоваться Интернетом в учебном заведении, и лишь 30% сообщили, что могут выйти в Интернет не только на специальных занятиях, но и по необходимости [234]. Тем не менее, проникновение Интернета среди подростков уже

существенно превосходило уровень проникновения среди взрослых (равный 32-35%), и мы сочли возможность пренебречь описанными выше расхождениями.

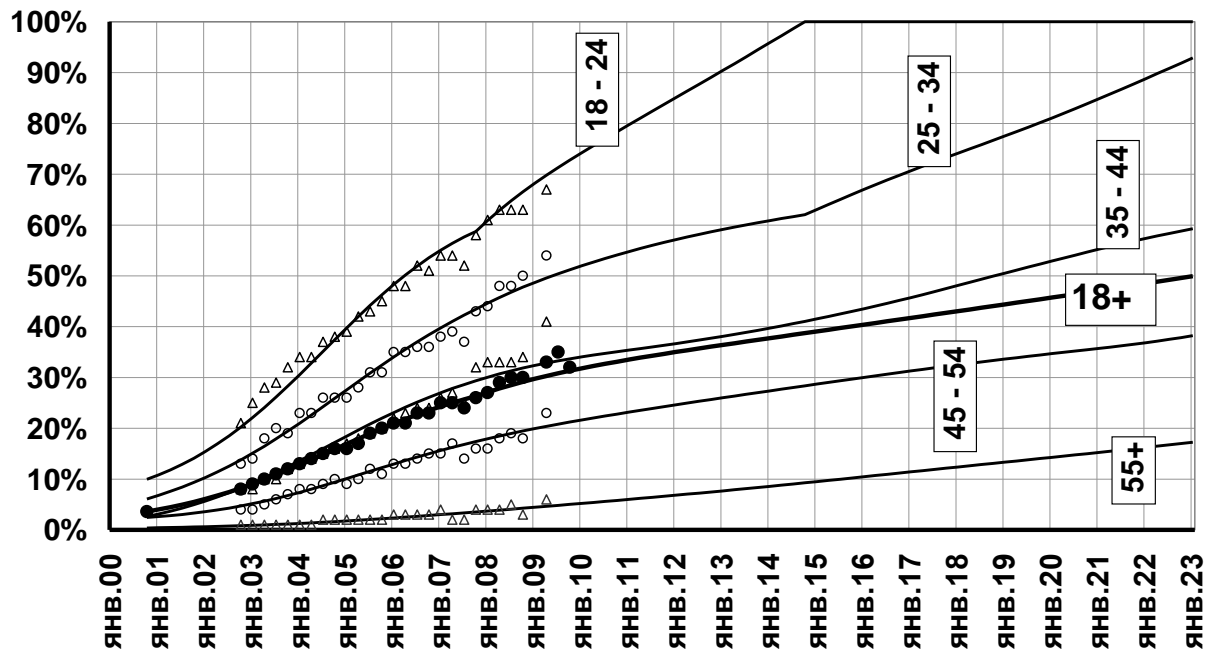


Рис.1.3. Результаты моделирования распространения Интернета в возрастных группах с учетом интернетизации школ

На рис.1.3 кружками и треугольниками изображены данные ФОМ и ВЦИОМ о проникновении Интернета в отдельных возрастных группах. Закрашены только те кружки и треугольники, которые соответствуют агрегированным данным, относящимся к населению России в возрасте 18 лет и старше, а символы, соответствующие отдельным возрастным группам, не закрашены. Результаты численного моделирования изображены сплошными линиями. Возрастные группы, к которым относятся кривые, указаны при помощи меток. Так, метка “18+” означает “население России в возрасте 18 лет и старше”.

При условии полной интернетизации российских школ, уровень распространения Интернета в возрастной категории 18–24 года достигнет 100% уже в 2014 году⁴. Однако на средний уровень распространения Интернета среди

⁴ Интернетизация школ существенно увеличила некоторые технические показатели доступности Сети, в частности, по оценке С.В.Пехтерева в 2007 г. “за один год количество VSAT (малогабаритных спутниковых терминалов), в первую очередь благодаря национальному проекту “Образование” (школьный Интернет) и госпрограмме по Универсальным услугам связи, практически утроилось” (См.: <http://www.telemultimedia.ru/art.php?id=367>)

взрослой части населения это почти не влияет. Таким образом, школьный Интернет – необходимое, но не достаточное условие движения страны к информационному обществу.

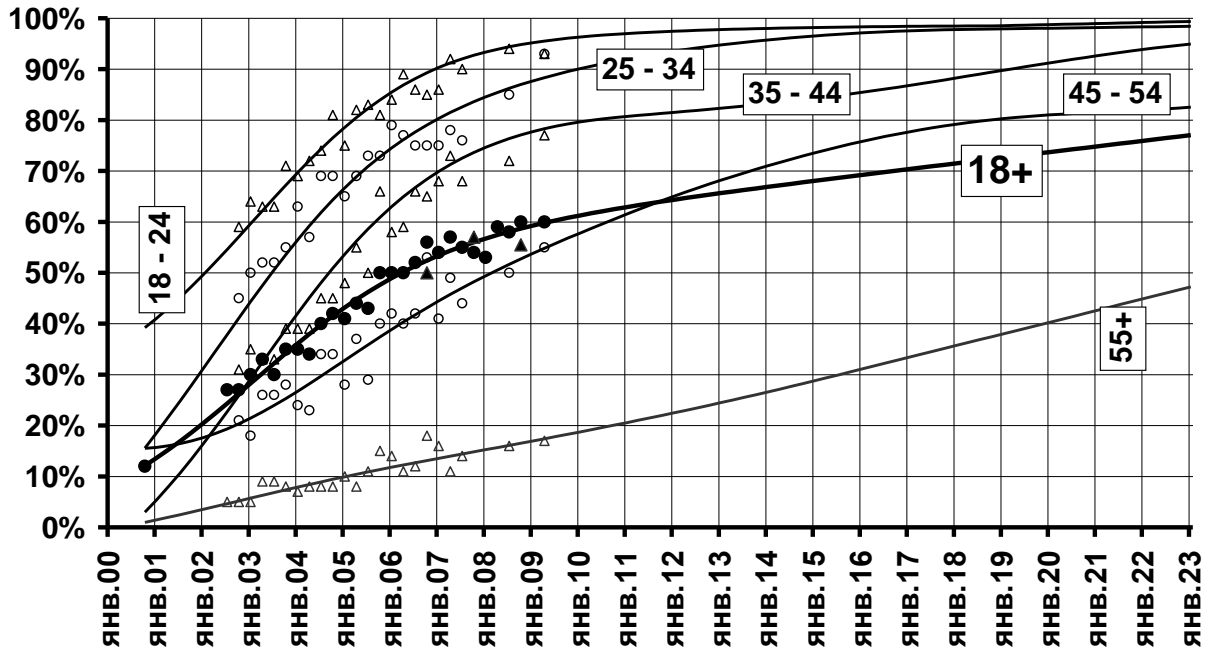


Рис.1.4. Результаты моделирования распространения Интернета в Москве

Прогнозирование распространения Интернета в г. Москве

На рис.1.4 изображены данные ФОМ и ВЦИОМ о распространении Интернета среди взрослых москвичей, сплошными линиями показаны результаты моделирования [120]. В Москве, так же, как в Швеции и США (см. рис.1.1), период роста доли пользователей Интернета среди взрослого населения с 10-15% до 60% составил 8 лет. В соответствии с предсказанием модели, в последующие 7-8 лет в Москве этот показатель должен был вырасти до 70%. В США и Швеции аналогичный переход занял всего лишь три года, вследствие быстрого подключения к Сети старших возрастных групп (см.рис.1.13-1.15). Так, в Швеции в 2008 году Интернетом пользовались 73% жителей в возрасте от 55 до 64 лет и 38% – старше 65 лет [363]. Но как показывает наша модель, “естественным путем”, без проведения специальных мероприятий по обучению пожилых людей использованию Интернета, в Москве уровень проникновения Интернета в группе

лиц старше 55 лет достигнет 40% только к 2020 году. Внутренний “цифровой разрыв” между поколениями сказывается даже в российской столице.

Уточним, что при моделировании распространения Интернета среди взрослых москвичей мы не рассматривали отдельно распространение Сети среди школьников. Уровень проникновения Интернета в возрастной группе 18-24 года уже летом 2007 года составил 90%, таким образом, эффект перехода обученных использованию Сети школьников в эту возрастную группу не может быть велик.

Прогнозирование распространения Интернета в Российской Федерации Сценарий быстрого роста

Несомненно, рост благосостояния в последнее десятилетие сделал Интернет доступнее для россиян. При постоянном или снижающемся уровне благосостояния существенно повысить уровень проникновения Интернета в России могло бы относительное снижение цен на доступ в Интернет. Альтернативой снижению цен может быть существенный рост полезности Интернета для жителей российских регионов. Как показывает современная история распространения технологий, в некоторых случаях при снижении темпов прироста финансовых или натуральных показателей компании ищут пути удешевления товаров или услуг для конечных пользователей, с целью увеличить потенциал рынка. Так, в 1999 году, когда мобильной связью в России пользовались в основном обеспеченные слои населения, крупные операторы сотовой связи, ощутив замедление роста своих абонентских баз, приняли решение о выводе услуги мобильной связи на массовый рынок и радикально снизили стоимость минуты разговора. В результате, рост числа пользователей возобновился, их доля в 2000 году достигла 2%, а сейчас личными сотовыми телефонами владеют более 70% взрослых россиян. То же произошло с широкополосным доступом (ШПД) в Интернет в мегаполисах: волна снижения цен на ШПД прошла в Москве, а, начиная с 2007 года, целый ряд провайдеров выступал с такими же инициативами и в регионах.

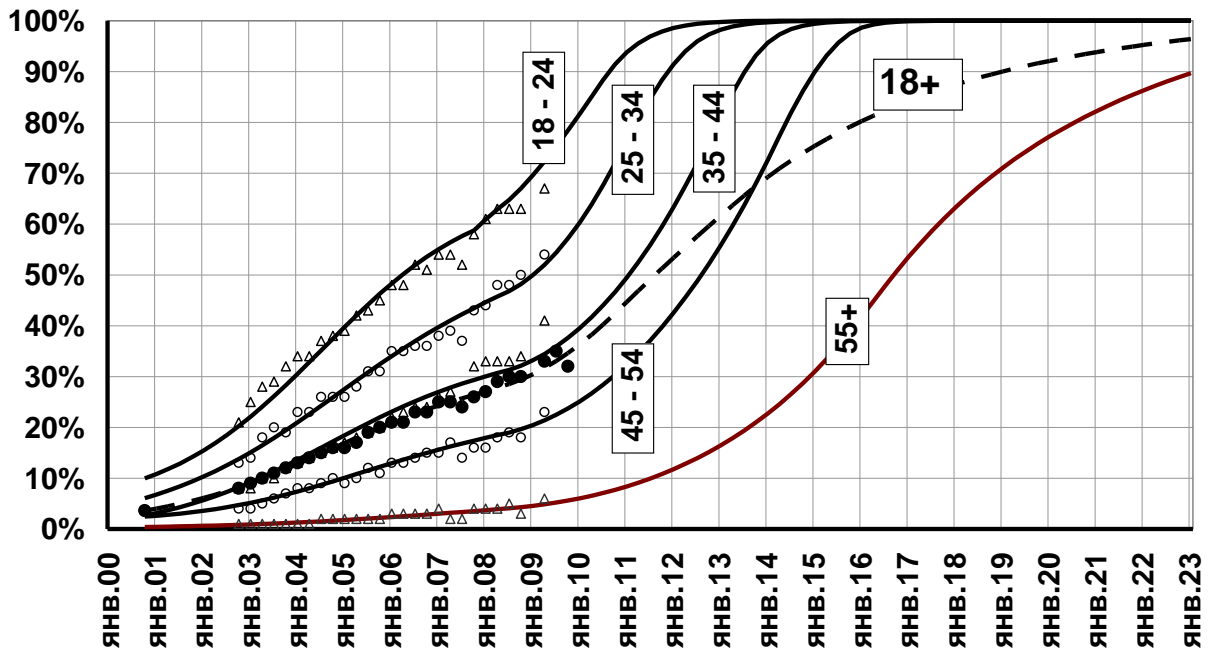


Рис.1.5. Результаты моделирования распространения Интернета в возрастных группах с учетом интернетизации школ и снижения стоимости доступа

При снижении стоимости доступа в Интернет рост численности его пользователей мог бы заметно ускориться (рис.1.5). По нашим расчетам [85], если бы, начиная с весны 2008 года, доступ дешевел на 5% в квартал, то половина взрослых россиян начала бы пользоваться Интернетом уже к концу 2011 г. При этом требуется снижение стоимости доступа не только в городах-“миллионниках” (где проживает только 27 миллионов россиян), а повсеместно. Отметим, однако, что апробированных моделей ценовой эластичности потенциала рынка для интернет-индустрии нет, за исключением работы С.Киискии и М.Похьола [395], где эластичность была получена в ходе совместного моделирования падения цен и роста количества подключенных к Сети хост-компьютеров в странах ОЭСР, поэтому полученную оценку нельзя считать вполне надежной.

Сценарий отложенного роста

Как свидетельствовали опубликованные прессой заявления компаний, в 2009 году повсеместное снижение стоимости доступа к Интернету для частных лиц не планировалось [233]. Кроме того, ряд крупных российских провайдеров

уже тогда объявил о замораживании инвестиций в строительство широкополосного доступа в Интернет в регионах, в том числе и в крупнейших городах-миллионниках. Те компании, которые не объявили об этом прямо, сообщили о “пересмотре временных рамок”. В то время как московские провайдеры выражали осторожную надежду, что жители столицы не откажутся от Интернета, многие аналитики обосновывали сворачивание строительства ШПД в регионах неизбежным падением спроса. Тогда более вероятной представлялась динамика развития российского Интернета, изображенная на рис.1.6.

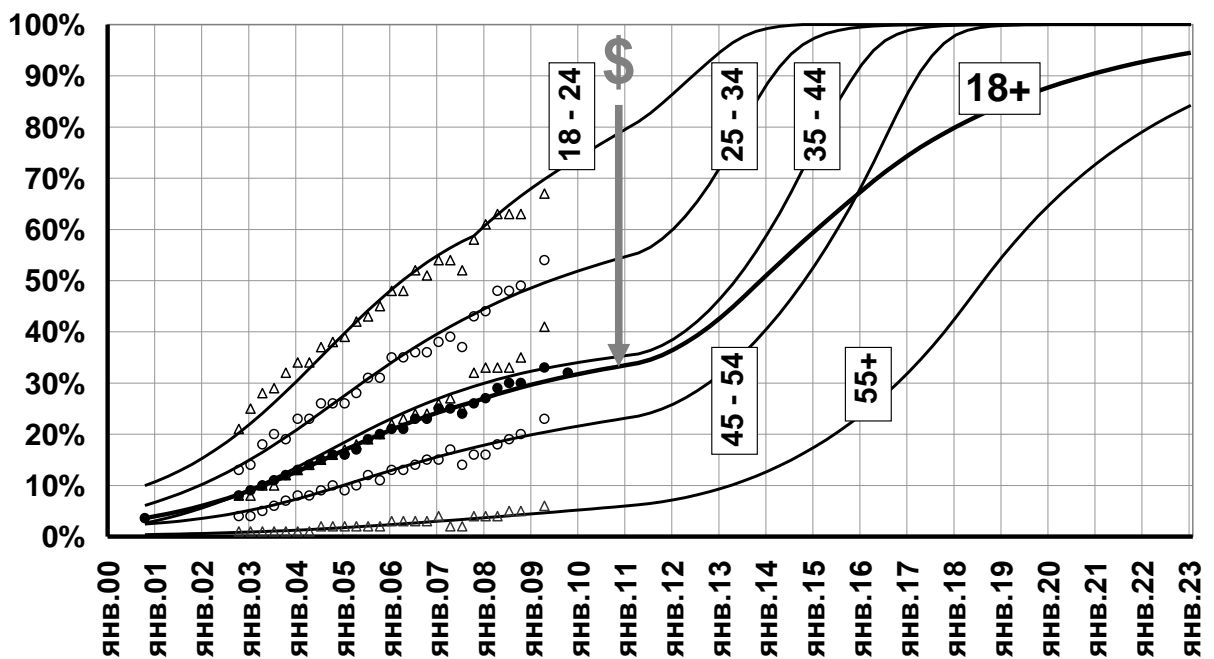


Рис.1.6. Результаты моделирования распространения Интернета в различных возрастных группах в условиях замораживания регионального развития

Более того, в условиях рецессии растет вероятность стагнации российского Интернета [301], которая неизбежна без обучения использования Интернета школьников. Такой сценарий еще год назад казался невозможным, однако сейчас представляется вполне вероятным, поскольку региональные власти могут и не найти миллиард рублей для инвестиций в отдаленное будущее.

Сопоставление с опубликованными прогнозами других авторов

На рис.1.7 сплошной линией показан опубликованный Л.С.Тимошенко прогноз Мининформсвязи распространения Интернета в Москве [272]. Этот прогноз носил нормативный характер, так что потенциал распространения Сети был установлен в качестве желаемой (но необязательно – достижимой) цели. Закрашенная область на рис.1.7 ограничена сверху оптимистическим прогнозом Минэкономразвития для России [246] снизу – консервативным. Дополнительно показаны два наших сценария – инерционный, помеченный цифрой 1, и сценарий быстрого роста, помеченный цифрой 2.



Рис.1.7. Сравнение прогнозов Мининформсвязи и Минэкономразвития с данными массовых опросов

Мы видим, что прогноз использования Интернета москвичами, разработанный Мининформсвязи, заметно превышает данные опросов. Заметим, что Мининформсвязи прогнозирует аудиторию с учетом детей, в то время как ФОМ и ВЦИОМ опрашивают только взрослых, однако, как показали наши расчеты, если бы опрашивали и детей, то это сказалось бы только на абсолютных показателях, но не на относительных (доле пользователей среди населения).

Прогнозируемый Минэкономразвития тренд распространения Интернета в России повторяет прогноз Мининформсвязи для Москвы, но отстает от него на пять лет. Наша траектория быстрого роста (кривая '2' на рис.1.7), отстает от прогноза Минэкономразвития еще на один год, была рассчитана в предположении ежегодного снижения стоимости доступа не менее, чем на 20%. Однако в прогнозах Минэкономразвития не содержится ни описания каких-либо мер, стимулирующих такое снижение, ни каких-либо иных мер, упрощающих россиянам использование Сети.

Обсуждение результатов прогнозирования распространения Интернета в Российской Федерации и в г.Москве

Рис.1.8. иллюстрирует результаты прогнозирования для России в целом, проведенного нами в 2009 году в соответствии с описанным выше сценарием быстрого роста (рис.1.5), при этом закрашенные треугольники и кружки соответствуют опубликованным на тот момент данным ФОМ и ВЦИОМ, сплошные линии – результатам расчётов, а незакрашенные кружки и треугольники – новым данным ФОМ за 2009-2013 гг.

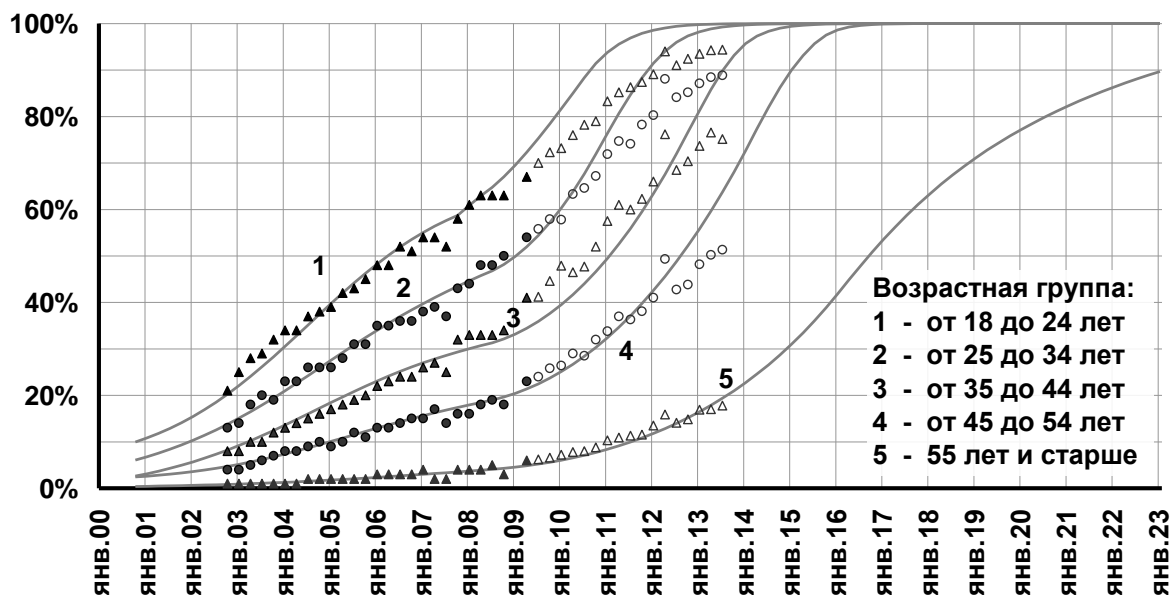


Рис.1.8. Результаты прогнозирования доли пользователей Интернета в России по возрастным группам

Для России в целом средняя квадратическая ошибка прогноза доли пользователей в возрасте 18 лет и старше за четыре года (с лета 2009 по весну 2013 гг.) составила 3 процентных пункта, максимальная ошибка – 6 п.п. Максимальная ошибка в отдельной возрастной группе составила 12 п.п. – для возрастной группы от 18 до 24 лет.

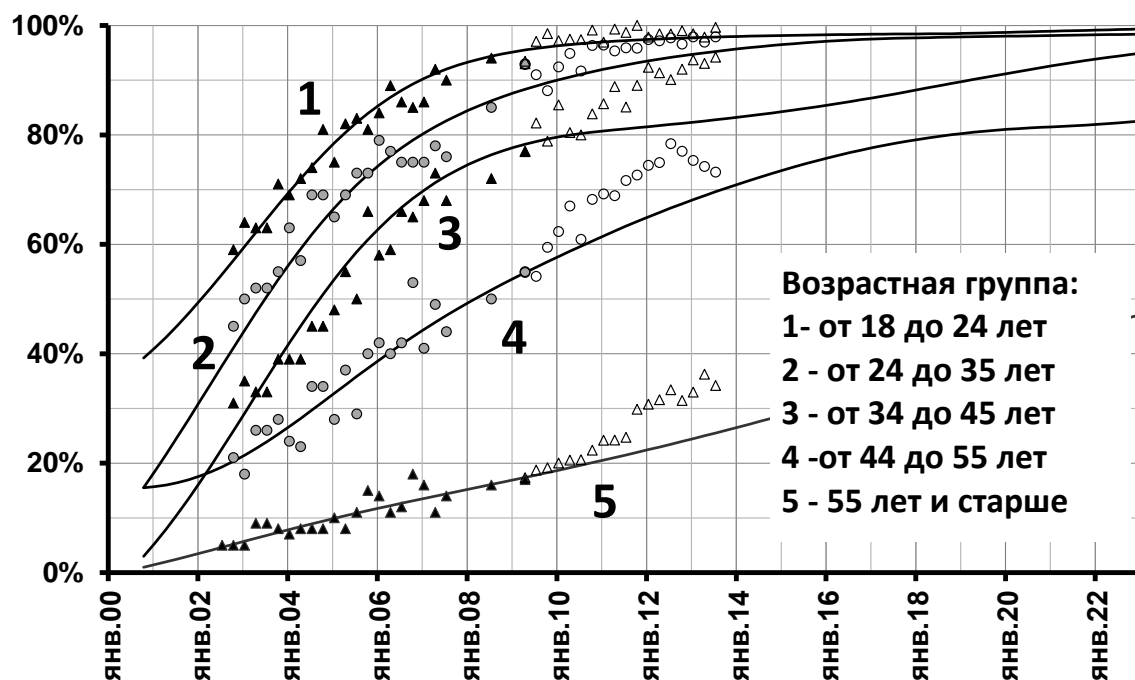


Рис.1.9. Результаты прогнозирования доли пользователей Интернета в Москве по возрастным группам

В то время как в России прогнозные значения к концу пятилетнего периода оказались, как правило, ниже реально наблюдаемых, в случае Москвы наблюдается обратная картина: в столице свежие итоги опросов ФОМ превысили наши прогнозы во всех возрастных группах (рис.1.9). Можно указать две возможные причины такого характера отклонений. Во-первых, в то время как наш общероссийский прогноз предполагал рост потенциала проникновения Интернета за счёт снижения стоимости доступа к Сети, прогноз для Москвы был сделан в предположении постоянства потенциала, поскольку мы полагали, что в столице услуга уже доступна каждому. Однако рост потенциала распространения Интернета мог иметь место и в Москве, поскольку благодаря распространению

мобильного Интернета возможность использовать Сеть получили лица среднего возраста, работающие не в офисе (охранники, продавцы, рабочие, водители). Второе возможное объяснение заключается в возможном усилении проигнорированного нами коэффициента “внешних воздействий” $p_i(\pi)$, поскольку в этот период в Москве конкурирующие между собой провайдеры доступа (преимущественно – операторы мобильной связи) размещали значительное количество рекламы новых тарифов доступа к Сети.

Мы заключаем, что успех прогнозирования показателей проникновения Интернета $x_i(t, \pi)$ существенно зависит от прогнозов значений потенциалов проникновения Сети $m_i(t, \pi)$. Однако прогноз динамики потенциалов является не менее сложной задачей, чем прогнозирование показателей распространения Сети, поскольку требуется предвидеть решения провайдеров доступа. В свою очередь, при принятии решений провайдеры доступа и регулирующие органы привлекают прогнозы динамики количества пользователей, и эти прогнозы могут быть построены на основе иных теорий и методов (таких как теория критической массы [46,47,357,474], “закон Джиппа” [32,206,64], полиномиальная и логистическая экстраполяция [206,405], эконометрические зависимости [167,203,10] и т.п.). С нашей точки зрения прогресс в этой области может быть достигнут путём детального изучения процесса управления информатизацией, включая описание кейсов, описывающих принятие решений в этой сфере [216,215]. Кроме того, необходима разработка методов измерения потенциалов $m_i(t, \pi)$, то есть доли населения, которая имеет возможность использовать новые информационные технологии, хотя и не делает этого в настоящее время.

Прогнозирование распространения Интернета в городских и сельских населённых пунктах Российской Федерации

При прогнозировании распространения Интернета в российских городах и сёлах [144] мы ограничились простейшими предположениями о поведении

потенциала рынка, и предлагаем два гипотетических итоговых сценария – оптимистический и консервативный. Мы полагаем

$$f_i(t, \pi, u(t)) = \gamma \cdot z_i(t, \pi) \quad (7)$$

для поколений, рожденных до 2001 года, что позволяет прогнозировать соответствующие потенциалы проникновения. Как уже упоминалось, мы предполагаем, что в более молодых поколениях невосприимчивые индивиды отсутствуют. Коэффициент γ равен $0,12/\text{год}$ для оптимистического сценария и нулю – для консервативного сценария.

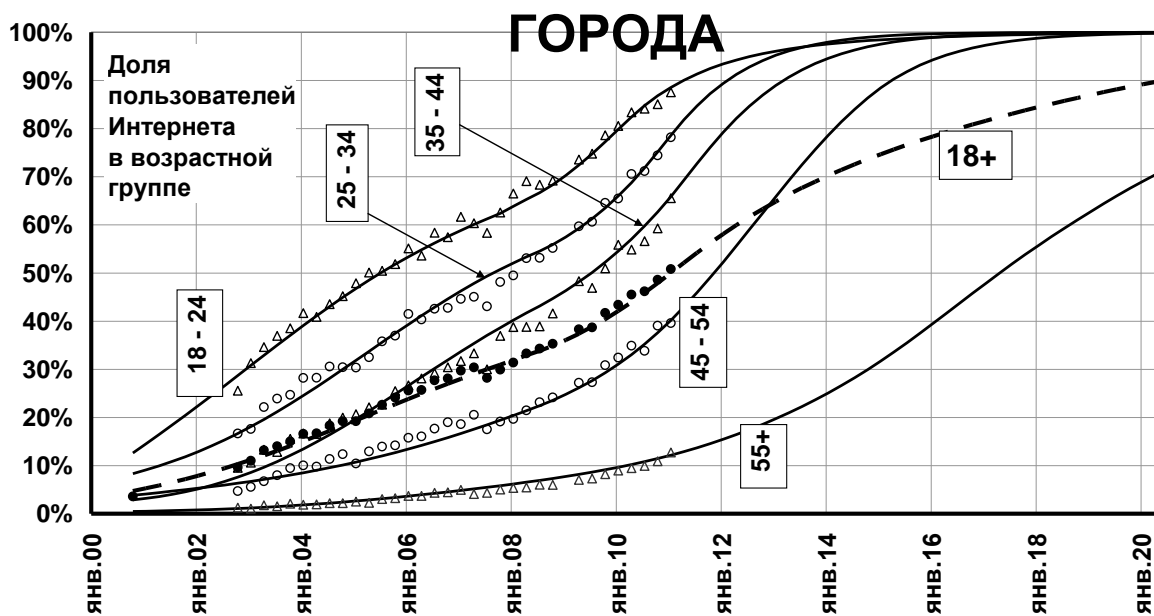


Рис.1.10. Результаты моделирования распространения Интернета в российских городах в ряде возрастных групп (оптимистический сценарий). Кружки и треугольники – данные опросов ФОМ, сплошные линии – результаты расчетов.

На рис.1.10 представлены результаты моделирования и прогнозирования распространения Интернета в российских городах в различных возрастных группах с учетом интернетизации школ и снижения стоимости доступа. Показан только оптимистический прогноз, предполагающий, что совокупная численность пользователей и восприимчивых индивидов, начиная с 2012 года, будет

возрастать на 12% ежегодно, благодаря технологическому прогрессу, снижению стоимости доступа к Сети и упрощению оконечных устройств доступа.

По оптимистическому сценарию к концу 2018 года проникновение Сети в городах достигнет 86%, в том числе в старшей возрастной группе (55 лет и старше) составит 61%. Нами был рассмотрен также консервативный сценарий, при котором рост проникновения Сети продолжится только благодаря обучению использованию Интернета в школах и смене поколений. В этом случае через шесть лет пользователями станут 78% российских горожан, в том числе 40% пожилых людей.

При моделировании структуры коммуникаций в городах предполагалось, что на решения горожан влияют только другие горожане (но не жители сёл). Кроме того, мы учли, что сверстники влияют друг на друга сильнее, чем остальные индивиды, однако это не дало сколько-либо заметного эффекта по сравнению с простой моделью, представленной уравнением (4).

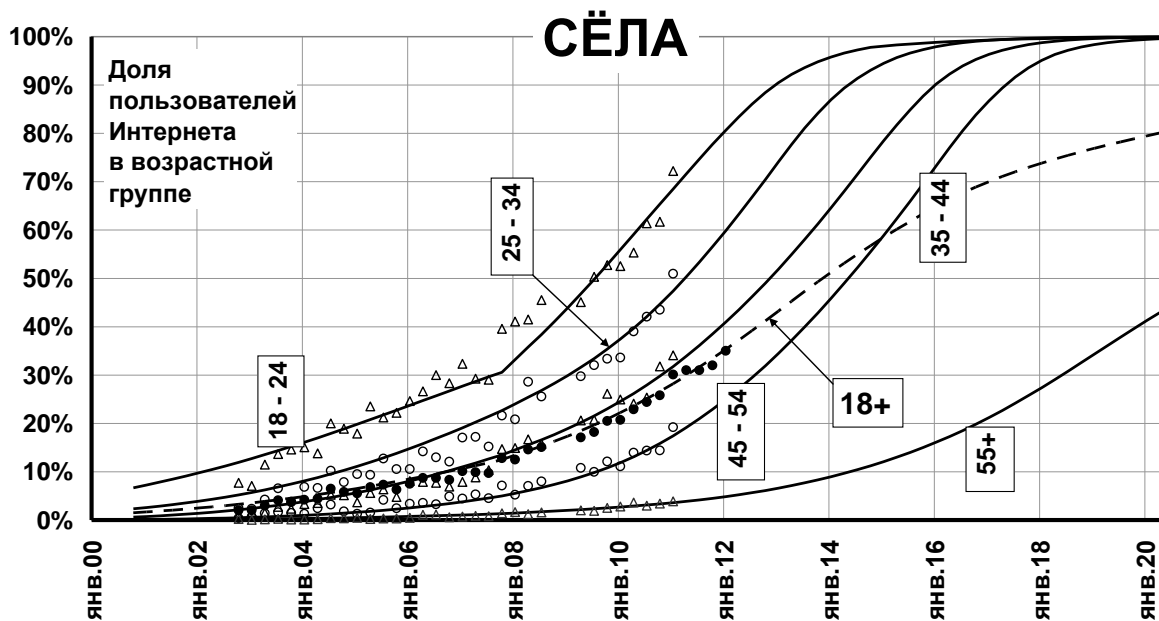


Рис.1.11. Результаты моделирования распространения Интернета в российских сёлах в ряде возрастных групп.

При моделировании распространения Интернета в сёлах мы допустили возможность влияния горожан на принятие сельским жителем решения о

подключении к Сети. Как отмечают Перес и др. [442], помимо непосредственных контактов между индивидами, регионы могут воздействовать друг на друга и другими способами, в нашем случае речь может идти, например, об экспансии городских провайдеров доступа к Интернету в сельские населенные пункты. Результаты моделирования представлены на рис.1.11.

В соответствии с оптимистическим сценарием (рис.1.11), к концу 2018 года 76% сельчан будут использовать Сеть. В старшей возрастной группе Интернетом воспользуются 33% граждан. Если же реализуется консервативный сценарий, то средний уровень проникновения в селах составит 55%, а в старшей возрастной группе – 17%. В этом случае и в более молодых возрастных группах проникновение Сети останется неполным.

Расчитав взвешенное среднее из оптимистических прогнозов для городского и сельского населения, мы получили прогноз распространения Интернета в России в целом. Отличия от опубликованного нами ранее прогноза, в котором не учитывались различия между сельскими и городскими жителями [111], составили лишь полтора процентных пункта. Эти отличия оказались значительно меньше, чем мы ожидали, основываясь на анализе прогнозов глобального распространения Интернета [429,307], проведённом автором совместно с Н.С. Матвеевой [159]. Причиной столь малых различий, по-видимому, является относительно небольшая доля населения, приходящаяся на “запаздывающий” регион, которая в России составляет четверть населения, в то время как глобально – шесть седьмых.

1.2 Некоторые аналитические модели распространения новых информационных технологий и их приложения

Классические модели распространения нововведений, такие как логистическая модель и модель Басса [315] не учитывают воспроизводство населения, что ограничивает их применение инновациями со сравнительно короткими характерными временами. Разработанные нами в предыдущих разделах модели распространения нововведений в неоднородном обществе с

учетом демографических процессов решают указанную проблему, однако представляют собой систему интегро-дифференциальных уравнений в частных производных, которые требуют разработки специальных методов численного решения. Оценка точности численных решений представляет самостоятельную сложную проблему. Поэтому желательно провести качественный анализ решений уравнений распространения нововведений с учетом демографических процессов хотя бы в небольшом числе случаев, допускающих аналитическое решение. Благодаря недавним работам Ф.Центроне, А.Гойя и Э.Салинелли [329,328] обнаружилось, что такие случаи существуют, более того, возможны относительно простые аналитические модели некоторых сценариев, для которых полные системы уравнений баланса пока написать не удастся. Соответствующие модели проанализированы в данном разделе, который основан на ряде результатов, полученные в 2011 г. совместно с Т.А.Подлесной [154] и С.А.Белоцерковским [152,155,156,157]. Для ряда практически важных частных случаев получены точные решения для уравнений распространения нововведений в стабильной (экспоненциально растущей или убывающей) популяции с возрастной структурой. Решения для модели Центроне, Гойя и Салинелли [329] с возрастной структурой получены для случая постоянной силы смертности. Для моделей с всеобщим и обязательным обучением детей использованию нововведения или с обучением только в семьях пользователей получены точные решения и для изменяющейся с возрастом силы смертности. Для общего случая граничных условий задача сведена к линейному интегро-дифференциальному уравнению, которое можно решить численно при помощи преобразования Лапласа.

1.2.1 Аналитическая модель распространения нововведения с учетом обучения

Ф.Центроне, А.Гойя и Э.Салинелли [329,328] построили и проанализировали модель распространения нововведения с учетом рождаемости и смертности для случая, когда

- 1) смертность не зависит от возраста индивида $\mu(t, \pi) \equiv \mu$;

- 2) коэффициенты рождаемости не зависят от возраста $b(t, \pi) \equiv b$;
- 3) популяция стабильна (экспоненциально изменяется с сохранением возрастной структуры).
- 4) все новорожденные пополняют класс потенциальных пользователей;
- 5) нет таких индивидов, которым нововведение недоступно по финансовым или техническим причинам;
- 6) функция риска стать пользователем $h(t, \eta(t))$ одинакова для любого индивида и не зависит от его возраста.

Преимуществом этой модели, которая в [329] названа моделью CGS, по сравнению с классической логистической моделью, является возможность анализа влияния естественного роста населения на распространение инноваций, что актуально для многих регионов мира. Модель CGS можно описать при помощи системы обыкновенных дифференциальных уравнений

$$\frac{dX}{dt} = \left(p + q \frac{X}{K} \right) \cdot Y - \mu X, \quad \frac{dY}{dt} = - \left(p + q \frac{X}{K} \right) - \mu Y + bK, \quad \frac{dK}{dt} = (b - \mu)K.$$

Здесь $X(t)$ – количество пользователей нововведения; $Y(t)$ – количество потенциальных пользователей, еще не использующих нововведение; $K(t)$ – численность населения. Используются также постоянные неотрицательные параметры: p – параметр внешнего воздействия, q – параметр внутреннего воздействия; b – коэффициент рождаемости; μ – коэффициент смертности.

Для уравнений модели CGS её авторами получено явное решение [329]

$$K(t) = K(0)e^{(b-\mu)t}, \quad X(t) = K(t) \cdot \xi(t), \quad Y(t) = K(t) \cdot (1 - \xi(t)),$$

где используются вспомогательная (логистическая) функция

$$\xi(t) = \frac{1 - \tilde{p} \tilde{\tau}_0 \exp(-(\tilde{p} + \tilde{q})(t - t_0))}{1 + \tilde{q} \tilde{\tau}_0 \exp(-(\tilde{p} + \tilde{q})(t - t_0))} \tilde{M}$$

и семь комбинаций параметров и начальных условий исходной задачи

$$\tilde{\tau}_0 = \frac{\tilde{M} - \xi_0}{\tilde{p}\tilde{M} + \tilde{q}\xi_0}, \quad \xi_0 = \frac{X(t_0)}{K(t_0)}, \quad \eta = \sqrt{\beta^2 + 4pq}, \quad \beta = q - p - b.$$

$$\tilde{p} = \frac{\eta - \beta}{2}, \quad \tilde{q} = \frac{\eta + \beta}{2}, \quad \tilde{M} = \frac{\eta + \beta}{2q} = \frac{2p}{\eta - \beta} = \frac{\tilde{q}}{q} = \frac{p}{\tilde{p}}.$$

Модель CGS справедлива лишь для постоянной силы смертности $\mu(t, \pi) \equiv \mu$ (где t – текущий момент времени, π – момент рождения индивида), не характерной для человеческого общества.

Обобщённая модель CGS

В предлагаемом нами обобщении модели CGS сила смертности зависит от момента рождения индивида (или, эквивалентно, от его возраста), т.е. снимается ограничение (1). Ослабляя ограничение (4), допускаем немедленное зачисление в пользователи нововведения детей, рождённых в семьях пользователей (аналогично наследственным болезням)⁵.

Пусть нововведение начинает распространяться в момент $t_0 = 0$. Мы предполагаем, что плотности (на единицу возраста) количества пользователей $x(t, \pi)$ и потенциальных пользователей $y(t, \pi)$ нововведения, рожденных в момент π , при $t \geq \pi$ удовлетворяют уравнениям баланса, рассмотренным в предыдущих разделах

$$\frac{\partial x}{\partial t} = h(t, \eta(t))y(t, \pi) - \mu(t, \pi)x(t, \pi), \quad \frac{\partial y}{\partial t} = -h(t, \eta(t))y(t, \pi) - \mu(t, \pi)y(t, \pi) \quad (1)$$

с нелокальными граничными условиями для поколений (при $t \geq 0$)

$$x(t, t) = b_x X(t) + b_y Y(t), \quad y(t, t) = (b - b_x)X(t) + (b - b_y)Y(t), \quad (2)$$

где численности пользователей $X(t)$ и потенциальных пользователей $Y(t)$ вычисляются при помощи интегрирования по моменту рождения π

$$X(t) = \int_{-\infty}^t x(t, \pi) d\pi, \quad Y(t) = \int_{-\infty}^t y(t, \pi) d\pi, \quad K(t) = \int_{-\infty}^t k(t, \pi) d\pi, \quad (3)$$

а их доли в населении суть $\xi(t) = X(t)/K(t), \quad \eta(t) = Y(t)/K(t).$ (4)

⁵ От допущения (2) тоже можно отказаться, при этом удастся получить систему, состоящую не из двух, а из трех интегральных уравнений. Предприняв попытку отказаться также и от допущения (6) и позволить функция риска зависеть от π , мы не смогли получить каких-либо простых решений.

Здесь $K(t)$ – описывает полную численность населения, $k(t, \pi)$ – плотность популяции на единицу возраста. Функции $k(t, \pi)$ считаются известными и непрерывными при $t \geq \pi$

$$k(t, \pi) = \begin{cases} bK(\pi)l(t, \pi) & \text{при } t \geq \pi > 0 \\ k(0, \pi) \frac{l(t, \pi)}{l(0, \pi)} & \text{при } t > 0 \geq \pi \\ 0 & \text{при } t < \pi \end{cases}$$

Заданы начальные условия

$$y(t_0, \pi) = \varphi(\pi), \quad x(t_0, \pi) = k(t_0, \pi) - \varphi(\pi). \quad (5)$$

Условия (2) содержат два новых параметра по сравнению с CGS:

b_x – определяет количество детей, которые родились в семьях пользователей, и которые обучены (дома или в школе) использованию нововведения;

b_y – определяет количество детей, которые обучены в обязательном порядке (в школе), хотя их родители к моменту t и не являются пользователями.

Определим функцию дожития

$$l(t, \pi) = \exp\left(-\int_{\pi}^t \mu(\theta, \pi) d\theta\right) \text{ для } t \geq \pi \quad \text{и} \quad l(t, t) = 1 \text{ для } t \leq \pi, \quad (6)$$

и плотность распределения популяции по моментам рождения

$$\rho(t, \pi) = b \cdot l(t, \pi) K(\pi) / K(t). \quad (7)$$

Как и в CGS, функция риска стать пользователем $h(t, \eta(t))$ не зависит от π .

Определим также две вспомогательные функции

$$u(t) = \begin{cases} \exp\left(\int_0^t h(\theta; \eta(\theta)) d\theta\right) & \text{для } t \geq 0 \\ 1 & \text{для } t < 0 \end{cases} \quad \text{и} \quad v(t) = \begin{cases} \eta(t)u(t) & \text{для } t \geq 0 \\ \eta(0) & \text{для } t < 0 \end{cases} \quad (8)$$

Утверждение 1 (лемма эквивалентности).

Решения интегральных уравнений

$$y(t, \pi) = \begin{cases} y(\pi, \pi)l(t, \pi) \frac{u(\pi)}{u(t)} & \text{для } t \geq \pi > 0 \\ y(0, \pi) \frac{l(t, \pi)}{l(0, \pi)} \frac{1}{u(t)} & \text{для } t \geq 0 \geq \pi \\ 0 & \text{для } t < \pi \end{cases} \quad (9)$$

являются также решениями интегро-дифференциальных уравнений (1) с условиями (2) и (5), и наоборот, решения (1),(2),(5) удовлетворяют (9).

Утверждение 2 (аналог уравнения восстановления)

Функции $v(t)$ и $u(t)$ связаны уравнением

$$v(t) = \eta_0(t) + \delta_x \int_0^t \rho(t, \theta) u(\theta) d\theta + \delta_{xy} \int_0^t \rho(t, \theta) v(\theta) d\theta, \quad (10)$$

где $\eta_0(t) = \frac{1}{K(t)} \int_{-\infty}^0 \varphi(\pi) \frac{l(t, \pi)}{l(0, \pi)} d\pi$ – доля потенциальных пользователей, рожденных не позже $t_0 = 0$ и доживших до t .

Функция $\eta_0(t)$ известна, поскольку вычисляется по известным функциям.

Здесь введены обозначения: $\delta_x = \frac{b-b_x}{b}$ и $\delta_{xy} = \frac{b_x-b_y}{b}$.

Доказательства обоих утверждений приведены в нашей статье [164]. В частности, формулу (10) можно получить при помощи интегрирования (9) по моментам рождения и последующего умножения обеих сторон уравнения на $u(t)$.

Утверждение 3 (следствие выбора формы функции риска).

Примем предположение Ф.Басса [315] о том, что функция риска воспользоваться нововведением линейно зависит от $\xi(t)$, т.е.

$$h(t, \eta(t)) = p + q\xi(t) = p + q - q\eta(t). \quad (11)$$

Тогда функции $u(t)$ и $v(t)$ связаны вторым интегральным уравнением

$$u(t) = e^{(p+q)t} - q \int_0^t e^{(p+q)(t-\theta)} v(\theta) d\theta. \quad (12)$$

Действительно, из (8) следует $\frac{du}{dt} = u \cdot h(t; \eta) = (p+q)u - q\eta u = (p+q)u - qv$.

Решение этого уравнения относительно $u(t)$ имеет вид (12).

Уравнения (10) и (12) образуют систему линейных интегральных уравнений Вольтерра. В векторных обозначениях

$$\mathbf{f}(t) = \mathbf{g}(t) + \int_0^t \mathbf{G}(t, \theta) \mathbf{f}(\theta) d\theta, \quad (13)$$

$$\text{где } \mathbf{f}(t) = \begin{bmatrix} u(t) \\ v(t) \end{bmatrix}, \quad \mathbf{g}(t) = \begin{bmatrix} e^{(p+q)t} \\ \eta_0(t) \end{bmatrix} \quad \text{и} \quad \mathbf{G}(t, \theta) = \begin{pmatrix} 0 & -qe^{(p+q)(t-\theta)} \\ \delta_x \rho(t, \theta) & \delta_{xy} \rho(t, \theta) \end{pmatrix}. \quad (14)$$

Теорема (теорема 3.11 монографии П.Линца [402, стр.46]).

Если $\mathbf{g}(t)$ и $\mathbf{G}(t, \theta)$ непрерывны в треугольнике $0 \leq \theta \leq t \leq T$ (в смысле непрерывности всех их компонент), то система (13) имеет единственное непрерывное решение на отрезке $0 \leq t \leq T$.

Отсюда вытекает существование решения системы интегро-дифференциальных уравнений (1) для непрерывных функций $l(t, \pi)$.

Утверждение 4

Обобщенную модель CGS описывает система интегральных уравнений

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{g}}(t) + \int_0^t \hat{\mathbf{G}}(t, \theta) \hat{\mathbf{f}}(\theta) d\theta, \quad (15)$$

$$\text{где } \hat{\mathbf{f}}(t) = \begin{bmatrix} u(t)e^{-(p+q)t} \\ v(t)e^{-(p+q)t} \end{bmatrix}, \quad \hat{\mathbf{g}}(t) = \begin{bmatrix} 1 \\ \eta_0(t)e^{-(p+q)t} \end{bmatrix} \quad \text{и} \quad \hat{\mathbf{G}}(t, \theta) = \begin{pmatrix} 0 & -q \\ \delta_x \hat{\rho}(t, \theta) & \delta_{xy} \hat{\rho}(t, \theta) \end{pmatrix}. \quad (16)$$

Здесь введено обозначение $\hat{\rho}(t, \theta) = \rho(t, \theta)e^{-(p+q)(t-\theta)}$ и рассмотрены только $t \geq 0$.

Следствие

Пусть

- 1) функция дожития $l(t, \pi) = l(t - \pi)$ зависит только от разности аргументов;
- 2) популяция стабильна при $t \geq 0$: $K(t) = K(0)e^{\lambda t}$;
- 3) популяция и ранее была стабильной, т.е. начальные условия имеют вид

$$k(0, \pi) = K(0)e^{b\pi}.$$

Параметр λ определяется из уравнения Лотки $1 = b \int_0^{\infty} e^{-\lambda a} l(a) da$.

Тогда $\rho(t, \pi)$, $\hat{\rho}(t, \pi)$ и $\mathbf{G}(t, \theta)$ зависят только от разности аргументов, и уравнения (15) становятся уравнениями свёртки.

Замечание

Практическую проблему при применении дискретного преобразования Фурье для решения уравнений свёртки (15) создаёт разрывность в нуле функции дожития и разрыв в нуле производной $u(t)$. Поэтому мы применяем

преобразования Фурье не для расчета $u(t)$, а для расчета $\tilde{u}(t) = u(t) - u_{CGS}(t)$, где $u_{CGS}(t)$ подобрана так, чтобы $\tilde{u}(t)$ была непрерывна в нуле вместе с первой производной. Поскольку $u_{CGS}(t)$ вычисляется аналитически, находим $u(t) = u_{CGS}(t) + \tilde{u}(t)$.

Наши дальнейшие направления поиска аналитических решений обобщённой модели CGS [145] связаны с другими аппроксимациями функции $l(a)$, которые используются в биологии продолжительности жизни [45].

1.2.2 Модель динамики цифрового разрыва

В данном разделе решена задача о распространении нововведения в обществе, где количество индивидов, не имеющих возможности воспользоваться нововведением в сфере ИКТ, убывает со временем. Мы разделяем общество на три поколения (см.рис.1.12): “доцифровое” поколение, “цифровые мигранты” (взрослые) и “цифровые аборигены” в терминологии Дж. Пэлфри и У. Гассера [247]. Рассмотрим распространение новшества в стационарной популяции с функцией дожития в виде ступеньки длиной T лет. Такую функцию дожития называют функцией дожития II типа [7].

Пусть все индивиды, которым в момент t_0 исполнилось более L лет (“доцифровое” поколение), отвергают нововведение. Остальные индивиды принимают решения принять или отвергнуть новинку, при этом информация к ним поступает как через СМИ, так и в ходе межличностного общения. Кроме того, дети (“цифровые аборигены”) могут научиться использовать нововведение в семье или в школе. На отрезке $t_0 \leq t \leq t_0 + T - L$ умирают только представители “доцифрового” поколения, и модель можно описать дифференциальными уравнениями

$$dX / dt = h(t, X)Y + (\beta_x - \gamma)X + \beta_y(Y + Z),$$

$$dY / dt = -h(t, X)Y + (\beta - \beta_x + \gamma)X + (\beta - \beta_y)(Y + Z),$$

$$dZ / dt = -\beta Z,$$

с начальными условиями $X(t_0) = 0$, $Y(t_0) = K \cdot L/T$, $Z(t_0) = K(1 - L/T)$.

Здесь $X(t)$ и $Y(t)$ – количество соответственно пользователей и потенциальных пользователей нововведения; $Z(t) = K(1 - L/T - \beta(t - t_0))$ – количество индивидов старше L лет; K – численность популяции; $h(t, X) = p + qX(t)/K$ – т.н. функция риска; p и q – коэффициенты внешнего и внутреннего воздействия; $\gamma \geq 0$ – интенсивность прекращения использования; $\beta = 1/T$ – коэффициент рождаемости; неотрицательные параметры $\beta_x \leq \beta$ и $\beta_y \leq \beta$ характеризуют обучение детей нововведению в семье и школе.

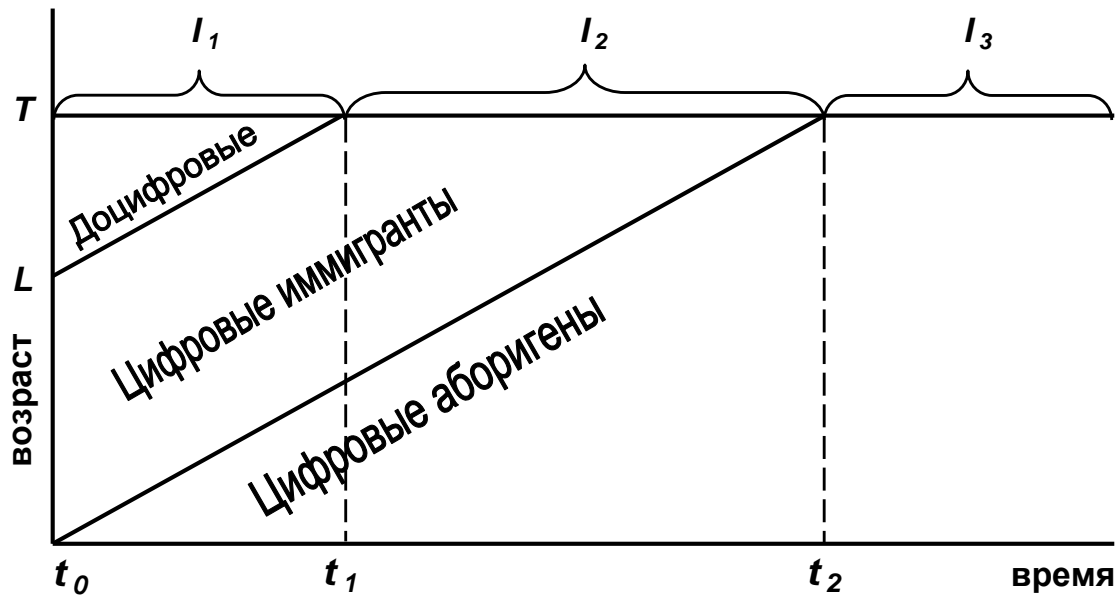


Рис.1.12. Три поколения, различно воспринимающие нововведение, на сетке Лексиса.

Проведя подстановку $h(t, X) = u'(t)/u(t)$ и замену переменной $\tau = \sqrt{q\beta/2} \cdot t$, а также добавив условие $u(0) = 1$, получим задачу Коши для уравнения Эрмита

$$\frac{d^2u}{d\tau^2} = 2(\tau + \sigma)\frac{du}{d\tau} - 2\nu u$$

с параметрами $\nu = \frac{(\beta_x - \gamma)p - \beta_y(p + q)}{q\beta}$ и $\sigma = \frac{p + q + \beta_x - \gamma - \beta_y - qZ(t_0)/K}{\sqrt{2q\beta}}$.

Решением этой задачи является линейная комбинация функций Эрмита первого и второго рода (при $\beta_x = \beta_y = \beta$, $\gamma = 0$ решение выражается через

интегральную функцию Доусона). Вычислив $u(t)$, рассчитываем $h(t, X) = u'(t)/u(t)$, и, наконец, $X(t) = (h(t, X) - p)K/q$.

На последующих отрезках времени: $[t_0 + T - L, t_0 + T]$, $[t_0 + T, t_0 + 2T - L]$ и т.д. задача сводится к решению системы линейных дифференциальных уравнений.

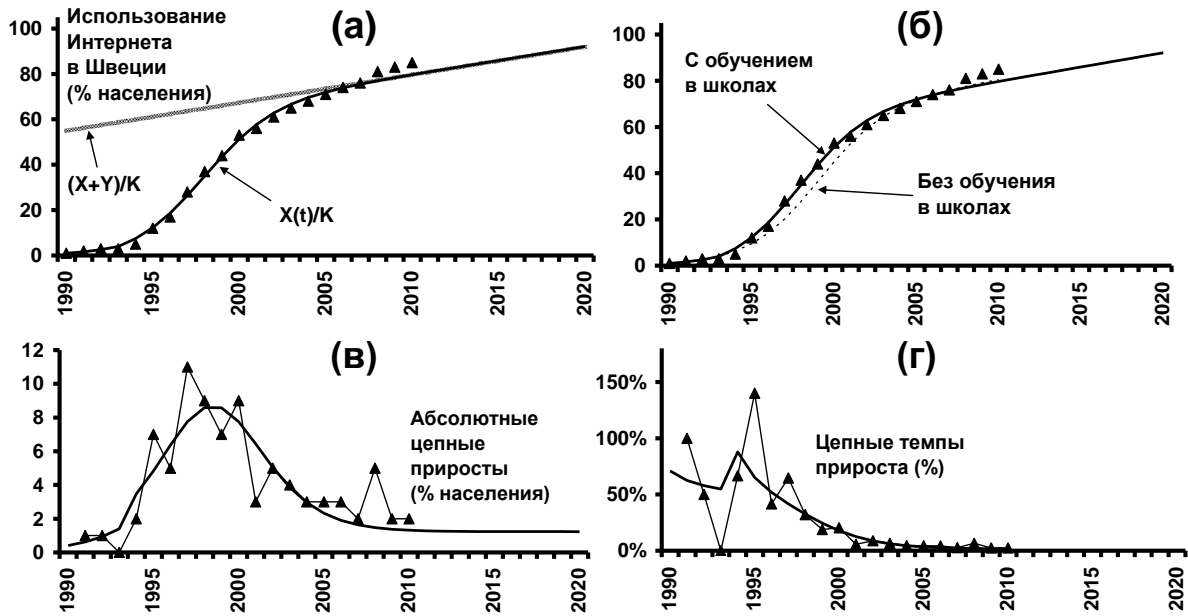


Рис. 1.13 Распространения Интернета в Швеции: данные World Internet Project (треугольники) и результаты расчетов при помощи модели (жирные линии), параметры которой получены методом наименьших квадратов.

Используя полученные решения, мы предлагаем новую интерпретацию данных опросов World Internet Project о проникновении Интернета в ряде стран в 1990-2010 гг. На рис.1.13 показаны данные о распространении Интернета в Швеции, где с 1994 года началась программа всеобщего и обязательного обучения школьников использованию Интернета. Эффект этой программы проявляется на рис.1.13 (г) в резком подъеме цепных темпов прироста доли пользователей в населении (треугольники), который отчасти удалось отразить в нашей модели (сплошная линия).

Важный частный случай при $\beta_x = \beta_y = \beta$ сводится к рассмотренному в следующем разделе решению, выражающемуся через интеграл Дюсона, и исследованному нами в работах [131,115,116,112,122].

Случай линейно растущего потенциала распространения нововведения

Как отмечено выше, хотя тренд показателей проникновения Интернета во многих странах является S-образным, он не обязательно является логистическим. В этом убеждают, например, рис. 1.1 и 1.13.

В случае, если распространение нововведения в отдельных возрастных группах нас не интересует, а требуются лишь агрегированные данные по населению в целом, можно использовать упрощенный подход, позволяющий избежать численного интегрирования дифференциальных уравнений. Такой подход менее трудоемок, и при его помощи были получены все модели и прогнозы распространения Интернета в зарубежных странах, которые использованы в наших работах [129,140, 345]. В этом случае модель имеет вид обыкновенного дифференциального уравнения [408]

$$\frac{dX}{dt} = \left(p + q \frac{X}{K} \right) (M(t) - X) \quad (1)$$

с начальным условием $X(0) = X_0$, где $M(t)$ – потенциал распространения нововведения, который может изменяться во времени.

Известно решение уравнения (1) в квадратурах, удовлетворяющее заданному начальному условию [408]. Это решение имеет громоздкий вид и в общем случае требует использование численного интегрирования

$$X(t) = -\frac{p}{q}K + \frac{\exp \left[pt + \frac{q}{K} \int_0^t M(\theta) d\theta \right]}{\frac{1}{X(0) + \frac{p}{q}K} + \frac{q}{K} \int_0^t \exp \left[p\theta + \frac{q}{K} \int_0^\theta M(\tau) d\tau \right] d\theta} \quad (2)$$

Нами показано, что если потенциал распространения нововведения линейно растет со временем,

$$M(t) = a + bt. \quad (3)$$

то можно выразить решение (4) через т.н. “интеграл Доусона” специальную функцию, для вычисления которой предназначено большое количество алгоритмов и программ.

При вычислениях оказывается практически целесообразным разделять случаи $b > 0$ и $b = 0$.

Решение для случая $b = 0$ и постоянного $M = K$ хорошо известно [см., например, 408], мы перепишем его в модифицированном виде

$$X(t) = \frac{1 - p\theta \exp[-(p+q) \cdot t]}{1 + q\theta \exp[-(p+q) \cdot t]} \cdot M, \quad (4)$$

где

$$\theta = (M - X(0)) / (pM + qX(0)). \quad (5)$$

Для $b > 0$ нам удалось выразить решение уравнения (1) с условием (2) для потенциала вида (2) при помощи специальной функции, которая носит название интеграла или интегральной функции Доусона [305]. Выражение имеет вид

$$X(t) = \left[-p + \frac{1}{c \cdot e^{z^2(0) - z^2(t)} + \sigma \cdot \text{daw}(z(t))} \right] \frac{K}{q}, \quad (6)$$

где

$$\text{daw}(x) = e^{-x^2} \int_0^x e^{u^2} du \quad - \text{интегральная функция Доусона,}$$

$$z(t) = \frac{t + \tau}{\sigma} \quad - \text{нормированное время,}$$

$$c = \frac{1}{p + q \frac{X(0)}{K}} - \sigma \cdot \text{daw}(z(0)), \quad \tau = \frac{pK + qa}{qb}, \quad \sigma = \sqrt{\frac{2K}{qb}} \quad - \text{параметры, не зависящие от}$$

времени (постоянные величины).

Область применения решения (6) ограничена областью $0 \leq t \leq (K - a)/b$ (поскольку требуется $M(t) \leq K$), а вне этой области следует использовать классическое решение (4).

При использовании интеграла Доусона, который доступные библиотечные функции вычисляют с точностью 18-20 знаков после запятой, не требуется численно интегрировать дифференциальное уравнение распространения

инновации (1) или вычислять интеграл (2) с помощью численных методов, соответственно не нужно выбирать шаг интегрирования и оценивать сходимость алгоритма. В данной работе при расчетах мы использовали функцию DawsonIntegral() библиотеки ALGLIB, реализованную С.Бочкановым на основе библиотеки Cephes [432].

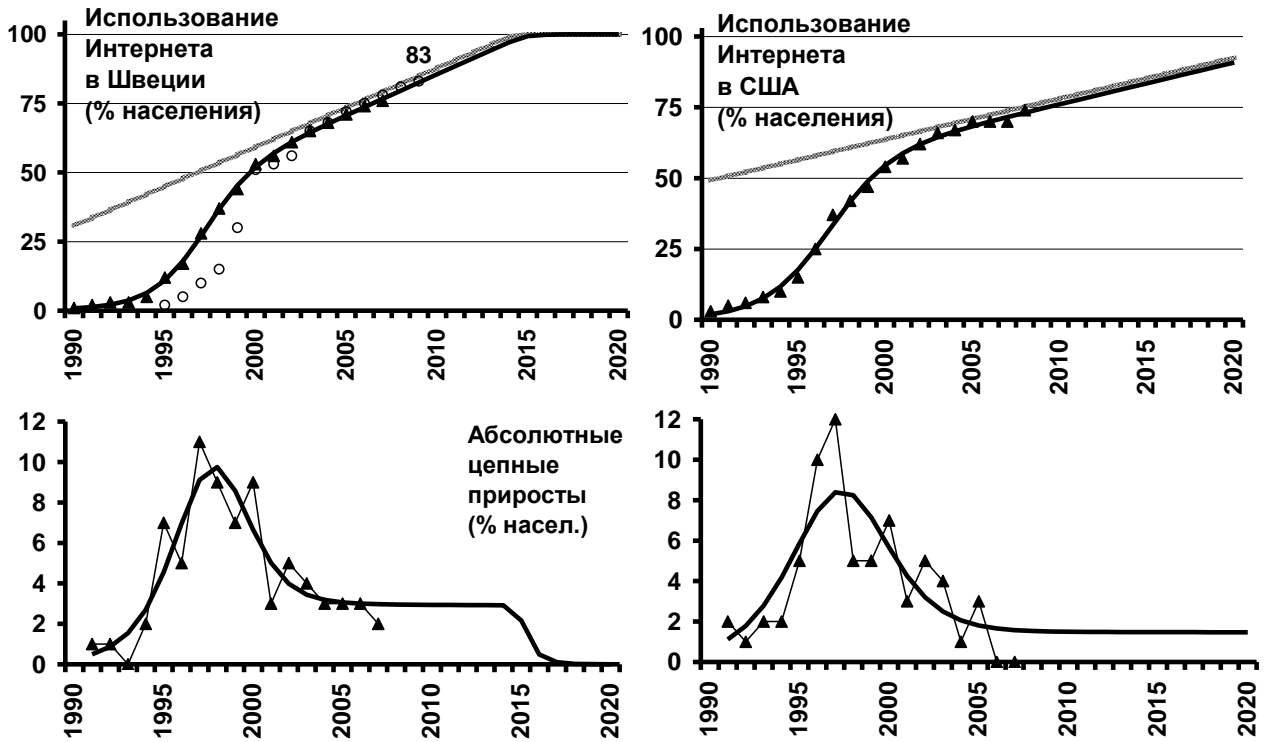


Рис. 1.14. Использование Интернета в Швеции и в США

При помощи модели (6) выполнены прогнозы распространения Интернета в США (рис.1.14), Швеции, Германии и Португалии на основе данных [363].

На рис.1.15 кружками и треугольниками показаны данные опросов ФОМ и ВЦИОМ о распространении Интернета среди москвичей в возрасте 18 лет и старше. Ромбы изображают аналогичные данные World Internet Institute об использовании Сети шведами. Квадратиками изображены данные ARD-ZDF-Online-Studies об использовании Интернета в Германии лицами старше 14 лет [43]. Пунктирными линиями показаны прогнозы, построенные нами при помощи модели распространения нововведения с растущим потенциалом [128]. Рис.1.15 демонстрирует, что тренды распространения Интернета в различных регионах

могут быть очень похожи, в первом приближении отличаясь лишь временным сдвигом. Так, тренд доли пользователей Интернета в Москве приблизительно повторяет тренды двух других стран, отставая на пять лет от Швеции и на два года от Германии. Москва “ближе” к Швеции и Германии по уровню доходов, чем Россия в целом, поэтому различия в распространении Интернета в общих чертах сводятся к простому отставанию во времени.

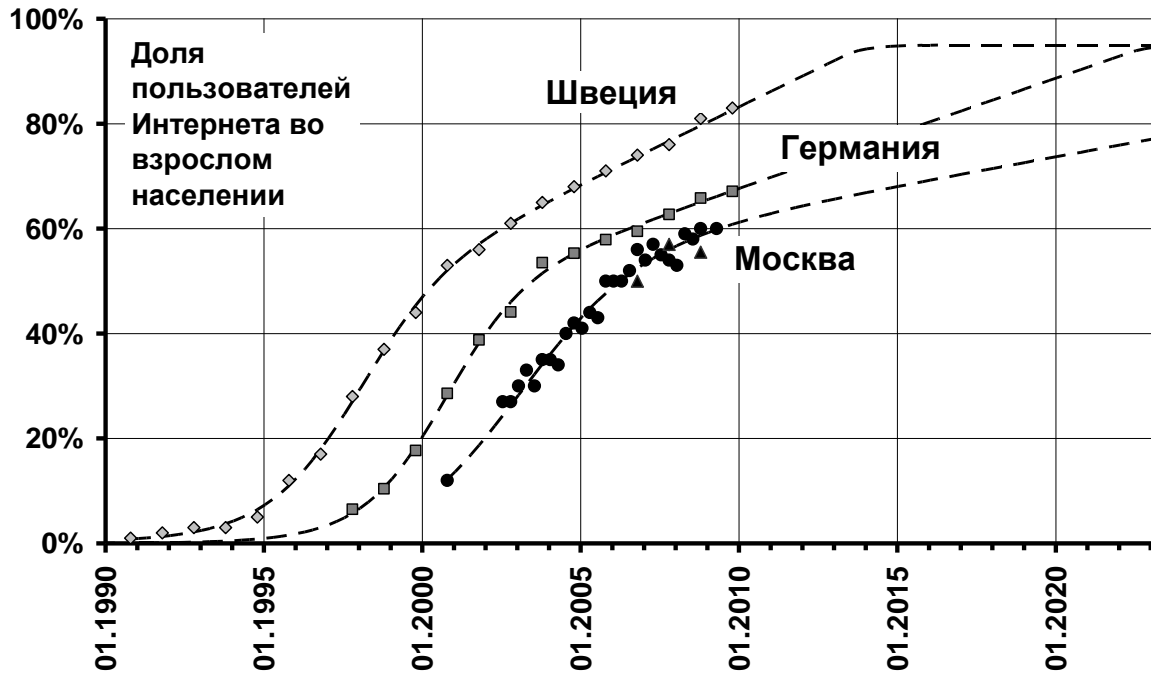


Рис.1.15. Распространение Интернета в Швеции, Германии и в Москве.

В наших работах [129,140, 345] приведены результаты аналогичных расчетов для Португалии.

1.3 Многостадийные модели распространения и интенсивности использования новых информационных технологий

В этом разделе рассмотрены многостадийные, то есть включающие не менее трёх стадий, модели распространения новых информационных и телекоммуникационных технологий (ИКТ), способные включить информацию о знаниях и намерениях лиц или групп, принимающих решения. Простая многостадийная модель, которая учитывает стадии осведомленности и

использования нововведения, применена к данным массовых репрезентативных опросов о распространении Интернета в России и в Москве. Полученная в итоге чрезвычайно высокая величина задержки между осведомленностью и формированием желания воспользоваться Сетью, составляющая 15 лет, может означать, что процесс распространения Интернета в России определяется не столько распространением информации в обществе, сколько технической и финансовой доступности услуги. Основные результаты этого раздела опубликованы в работах автора [94,100,107,93,104,92,105,108,110,111].

Классические одностадийные модели диффузии инноваций [315,408], как правило, успешно аппроксимируют исторические данные о динамике процесса распространения нововведения, однако обладают рядом важных недостатков как с точки зрения задачи объяснения всех доступных данных измерений, так и в задачах прогнозирования. Простые модели с тремя или четырьмя параметрами не позволяют задействовать доступную информацию о намерениях потенциальных пользователей, которая способна помочь определить потенциал распространения нового товара или услуги [308]. Зачастую классические модели оказываются непригодными для прогнозирования статистических показателей распространения ИКТ более, чем на два или три года вперед, что побуждает искать способы привлечения дополнительной информации, помогающие заглянуть в более отдаленное будущее.

1.3.1 Совместное моделирование количества информированных, намеренных воспользоваться и пользователей новой технологии

В этом разделе мы рассматриваем некоторые многостадийные модели распространения нововведений, способные включить информацию о знаниях и намерениях лиц или групп, принимающих решения.

Первая многостадийная модель принятия решения потребителем (модель AID), была предложена Э. Льюисом еще в 1898 году для применения страховыми агентами. В настоящее время при выделении стадий в распространении нововведений наиболее часто используется модель, разработанная Э. Роджерсом

и включающая пять стадий [449]. Дж. Додсон и Э. Мюллер [350], опираясь на модель воздействия рекламного сообщения, предложенную в 1961 г. Р.Лавиджем и Г.Стейнером, построили количественную многостадийную модель распространения нововведений, содержащую стадии неведения, осведомленности и использования, образующие иерархию. В современных моделях, помимо стадий, учитываются также неоднородность общества, например, по уровню доходов или по типу населенного пункта, однако в большинстве известных нам моделей в обществе не выделены ячейки, такие, как сообщество, домохозяйство или семья.

Далее мы рассматриваем примеры применения многостадийных моделей распространения инноваций для изучения распространения информационных и телекоммуникационных технологий в России и российской столице. Предлагаемые нами модели развивают подход Додсона и Мюллера [350] и применены к комплексным статистическим данным о распространении Интернета в России, собранным различными организациями при помощи различающихся способов, а также в Москве и ряде стран мира. В нашей работе [100] две альтернативные многостадийные модели применены к комплексу данных о распространении Интернета в Швеции, при этом обнаружилось, что данные могут быть с равной точностью описаны существенно различающимися моделями.

Разработка многостадийной модели

Главная особенность рассматриваемой в этом разделе модели, развивающей подход Дж. Додсона и Э. Мюллера [350], состоит в учете доступности нововведения: отдельно рассмотрены те индивиды, которым нововведение доступно, и те, которым оно недоступно по техническим, финансовым или иным причинам. Поскольку мы используем данные опросов россиян об индивидуальном использовании Интернета, то единицей, принимающей решение, в нашей модели является индивид.

Разделим общество, в котором распространяется нововведение, на две части по полу (женщины и мужчины), а каждую часть – по географии проживания и

иным признакам на однородные непересекающиеся сегменты, пусть i обозначает номер сегмента. Специфика многостадийных моделей распространения нововведений, выделяющая их среди общих моделей социальной динамики, проявляется в том, что в каждый момент времени t каждый i -й сегмент каждого поколения π разбивается на несколько непересекающихся подмножеств, соответствующих стадиям использования нововведения. Дополнительно мы классифицируем индивидов на тех, кому нововведение доступно или не доступно (по техническим или финансовым причинам). Мы исключаем из рассмотрения состояние, в котором индивид использует недоступное ему нововведение, предполагая такое явление редким.

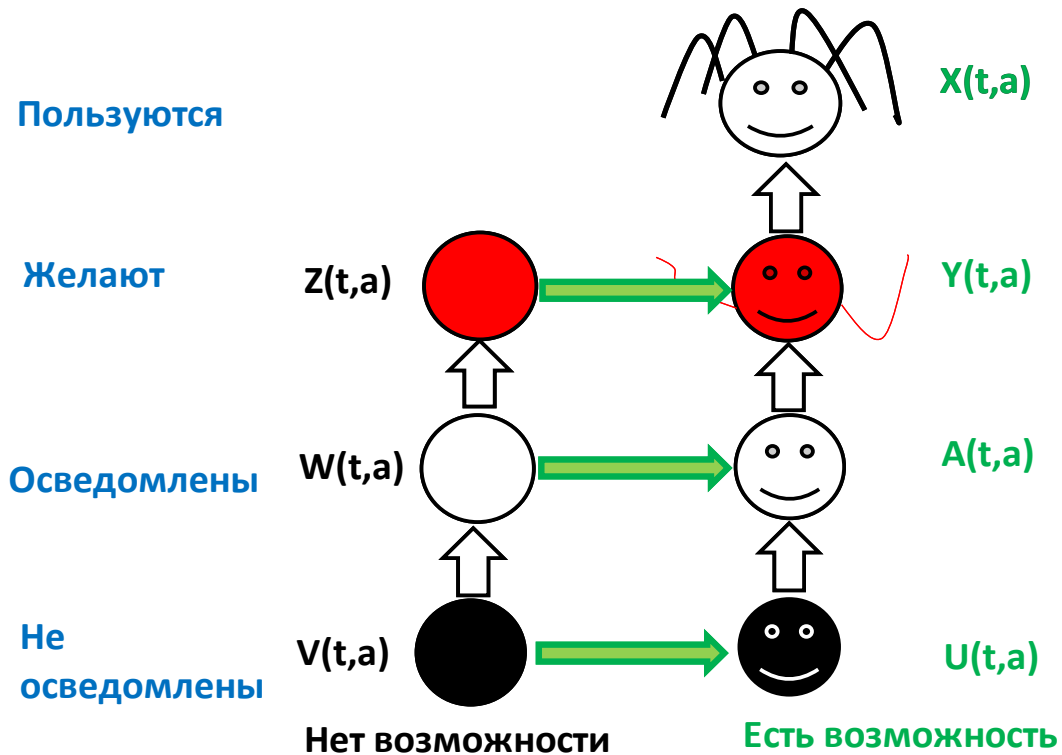


Рис. 1.16. Четыре стадии распространения нововведения

Данные опросов о распространении Интернета в России [209,235] (см. также раздел “СМИ и Интернет” WWW-сайта www.fom.ru), позволяют выделить семь состояний и, соответственно, семь подмножеств индивидов, численности которых (точнее, плотности на единицу возраста) для каждого поколения π в каждый момент t удовлетворяют уравнению баланса:

$$K_i(t, \pi) = X_i(t, \pi) + Y_i(t, \pi) + Z_i(t, \pi) + A_i(t, \pi) + W_i(t, \pi) + U_i(t, \pi) + V_i(t, \pi), \quad (1)$$

где $X_i(t, \pi)$ — количество⁶ пользователей нововведения,

$Y_i(t, \pi)$ — количество желающих приобрести нововведение и имеющих такую возможность (техническую и финансовую),

$Z_i(t, \pi)$ — число желающих приобрести нововведение, но не имеющих возможности это сделать (в частности, осенью 2008 года ФОМ установил, что среди желающих пользоваться Интернетом, составляющих 20% городского населения России, 40% не могут пользоваться Сетью по финансовым причинам и 12% — по техническим),

$A_i(t, \pi)$ — число информированных (осведомлённых) о нововведении и имеющих возможность, но не имеющих желания его приобрести,

$W_i(t, \pi)$ — число информированных (осведомлённых) о нововведении, не имеющих ни желания, ни возможности его приобрести,

$U_i(t, \pi)$ — число не осведомлённых о нововведении, потенциально имеющих возможность его приобрести,

$V_i(t, \pi)$ — число не осведомлённых о нововведении, не имеющих возможности его приобрести.

На рис.1.16 эти семь подмножеств изображены в виде кружков, образующих иерархию стадий. Улыбающиеся лица соответствуют наличию возможности приобрести нововведение. Вертикальные стрелки соответствуют переходам индивидов из предыдущей стадии в последующую, горизонтальные - возникновению технической и финансовой возможности использовать нововведение. В иерархической модели, изображённой на рис.1.16, индивид не способен “перескочить” через стадию. Отметим, что на рис.1.16 не отражены процессы рождаемости и смертности, которые учтены в предлагаемой ниже формуле (2) и которые затрагивают каждое из состояний.

⁶ Здесь и далее следовало бы использовать выражение “плотность на единицу возраста” или “концентрация”, но мы для краткости называем этот показатель просто “количеством” или “числом”.

Если всем стрелкам на рис.1.16 отвечают ненулевые интенсивности переходов (потоков событий), то лишь одно состояние – использование нововведения – является поглощающим [37]. Это означает, что в перспективе все рассматриваемые нами индивиды будут использовать нововведение (кроме умерших).

На основе сделанных предположений мы предложили представить модель системой уравнений:

$$\frac{\partial}{\partial t} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ A_i \\ W_i \\ U_i \\ V_i \end{bmatrix} = \begin{bmatrix} 0 & h_i & 0 & 0 & 0 & 0 & 0 \\ 0 & -h_i & f_i^{ZY} & a_i & 0 & 0 & 0 \\ 0 & 0 & -f_i^{ZY} & 0 & w_i & 0 & 0 \\ 0 & 0 & 0 & -a_i & f_i^{WA} & g_i & 0 \\ 0 & 0 & 0 & 0 & -f_i^{WA} - w_i & 0 & g_i \\ 0 & 0 & 0 & 0 & 0 & -g_i & f_i^{VU} \\ 0 & 0 & 0 & 0 & 0 & 0 & -f_i^{VU} - g_i \end{bmatrix} \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ A_i \\ W_i \\ U_i \\ V_i \end{bmatrix} - \mu_i \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ A_i \\ W_i \\ U_i \\ V_i \end{bmatrix}. \quad (2)$$

Граничные условия, как и в предыдущих разделах, имеют вид⁷

$$U_i(t, t) = \sum_{j=1}^J c_{ij} \int_{-\infty}^t B_j(t, \pi) K_j(t, \pi) d\pi, \quad (3)$$

$$\text{При этом мы требуем } \sum_{i=1}^J c_{ij} = 1, \quad c_{ij} \geq 0, \quad (4)$$

где $B_j(t, \pi)$ - рождаемость в поколении π сегмента j в момент времени t , $\mu_i(t, \pi)$ – интенсивность смертности. Для простоты подразумевается, что у всех новорожденных есть потенциальная возможность приобрести нововведение.

Остановимся на содержании уравнения (4). Подразумевается, что при рождении индивид попадает в подмножество $U_i(t, \pi)$ — он или она, безусловно, еще не знают о нововведении, но потенциально имеют возможность его использовать, поскольку государство уже обеспечивает всеобщее и обязательное обучение детей в школах использованию компьютеров, программного обеспечения и Интернета. В случае более дорогостоящих нововведений (например, космического туризма) некоторым сегодняшним младенцам

⁷ К сожалению, в наших публикациях [2.3,2.4,2.5,2.6,2.7,2.8,2.9,2.10,2.11] в уравнении (3) были неверно указаны пределы интегрирования и аргументы U .

инновация и через двадцать лет может не стать доступнее по причине дороговизны, в таком случае рационально перераспределить новорожденных между сегментами $U_i(t, \pi)$ и $V_i(t, \pi)$. Количество младенцев в формуле (4) получается путем суммирования рожденных в момент t от матерей всех поколений детородного возраста, эта величина вычисляется при помощи произведения $B_j(t, \pi)K_j(t, \pi)$ и интегрируется от $\pi = -\infty$ до $\pi = t$. Поскольку в момент t возникает поколение, которые мы характеризуем также датой t , в левой части (3) имеет вид $U_i(t, t)$.

Функции $h_i(t, \pi)$, $a_i(t, \pi)$, $w_i(t, \pi)$, $g_i(t, \pi)$ описывают условные вероятности перехода пользователя из предыдущей стадии в последующую (например, $g_i(t, \pi)$ – от неосведомленности к осведомленности), а функции $f_i(t, \pi, u(t))$ – переходы из «недоступных» подмножеств (численностью V, W, Z) в подмножества потенциальных пользователей (численностью соответственно U, A, Y), $u(t)$ – «траектория» управляющих параметров (в частности, цен) во времени.

Модель без изменения доступности Интернета

Первая многостадийная модель распространения Интернета в России и в Москве была построена нами в 2009 году [100,107,93,104,92,105,108,110,111] и не учитывала изменения доступности Сети. Мы запретили все горизонтальные переходы на рис.1.16, положив все $f_i(t, \pi, u(t)) \equiv 0$, благодаря чему состояния образовали два изолированных подмножества – имеющих и не имеющих доступ к Сети. Кроме того, положив $Z_i(t, \pi) \equiv 0$ и $w_i \equiv 0$, мы исключили состояние, в котором осведомлённый индивид хотел бы использовать Интернет, несмотря на отсутствие в данный момент доступа, несмотря на цитированные выше свидетельства ФОМ о наличии таких индивидов⁸.

⁸ В работе [442] рассмотрено формирование интереса к предстоящим кинопремьерам, при этом, благодаря рекламе в СМИ желание посмотреть фильм формируется, несмотря на отсутствие возможности. Признавая возможное наличие такого состояния, мы не учитываем его в моделях по причине отсутствия у нас данных, которые могли бы подтвердить или опровергнуть результаты моделирования.

Рассмотрим однородное общество, состоящее лишь из одного сегмента индивидов. Для простоты, следуя подходу [350], положим все $g_i(t, \pi)$ одинаковыми и равными

$$g_i(t, \pi) \equiv g(t) = p + q \frac{X(t) + Y(t) + A(t) + W(t)}{K}, \quad (5)$$

где $X(t)$ – полное количество пользователей нововведения. Разрабатывая эту модель, мы предполагали, что и в ближайшее десятилетие не все россияне будут осведомлены о существовании Интернета, поэтому положили величину количества потенциально осведомлённых равной некоторой, изначально неизвестной, величине K , которая не должна превосходить численность взрослого населения региона.

Все функции риска $h_i(t, \pi)$ и $a_i(t, \pi)$, как и в работе [350], были положены постоянными ($h_i(t, \pi) = \bar{h}$ и $a_i(t, \pi) = \bar{a}$), что эквивалентно экспоненциально распределенным задержкам между состояниями осведомленности, желания и использования. Выбор таких функций риска игнорирует межличностные коммуникации на всех стадиях, кроме перехода от неосведомлённости к осведомлённости.

Пренебрегая процессами воспроизводства населения, получаем систему обыкновенных дифференциальных уравнений

$$\frac{d}{dt} \begin{bmatrix} X \\ Y \\ Z \\ A \\ W \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0 & \bar{h} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\bar{h} & 0 & \bar{a} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\bar{a} & 0 & g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g \\ 0 & 0 & 0 & 0 & 0 & -g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -g \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ A \\ W \\ U \\ V \end{bmatrix} \quad (6)$$

Здесь $X(t)$ и другие зависимые переменные обозначают уже не плотности, а просто численности описанных выше множеств индивидов.

При расчётах мы добавили ещё одну промежуточную стадию “знания”, чтобы отразить различия между поверхностной осведомлённостью и уверенным знанием. Ввести эту стадию нас вынудили не только теоретические

представления [449], но и социологические данные, поскольку в 2000-м году около 80% российских респондентов “что-то слышали” об Интернете, но лишь 40% твёрдо знали, о чём идёт речь (см. рис.1.18). Модифицированная система уравнений модели имеет вид

$$\frac{d}{dt} \begin{bmatrix} X \\ Y \\ Z \\ A \\ W \\ U_2 \\ V_2 \\ U_1 \\ V_1 \end{bmatrix} = \begin{bmatrix} 0 & \bar{h} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\bar{h} & 0 & \bar{a} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\bar{a} & 0 & \bar{b} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \bar{b} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\bar{b} & 0 & g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\bar{b} & 0 & g \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ A \\ W \\ U_2 \\ V_2 \\ U_1 \\ V_1 \end{bmatrix} \quad (7)$$

Для процесса передачи информации используем функцию риска⁹

$$g_i(t, \pi) \equiv g(t) = p + q \frac{X(t) + Y(t) + A(t) + U_2(t) + W(t) + V_2(t)}{K}. \quad (8)$$

Все начальные условия были нулевыми, кроме исходного количества неосведомлённых $U_1(0) = U_0$, $V_1(0) = K - U_0$. Величину U_0 мы включили в состав неизвестных параметров модели. Другими неизвестными параметрами модели были \bar{h} , \bar{a} , \bar{b} , q и K . Предполагалось, что на стадии распространения информации (осведомлённости) межличностные коммуникации доминируют, т.е. $p = 0$. Расчёты величин, входящих в уравнения (6) были проведены способом конечных разностей. Значения неизвестных параметров были определены путём минимизации функционала невязки между расчётными и фактическими количествами пользователей $X(t)$, пользующихся или желающих пользоваться $M(t) = X(t) + Y(t)$, знающих $I(t) = X(t) + Y(t) + A(t) + W(t)$ и осведомлённых $N(t) = I(t) + U_2(t) + V_2(t)$.

Для Москвы значения параметров оказались равны $\bar{h} = 6,94$, $\bar{a} = 1,75$, $\bar{b} = 4,76$, $q = 8,83$ (размерностью всех параметров является доля индивидов в процентах в

⁹ В наших предыдущих работах было опубликовано неверное выражение для $g_i(t, \pi)$, но расчёты проводились по формуле (7), поэтому результаты расчётов, отображённые на рис.1.21 и 1.22 остаются в силе.

месяц), $K = 0,94$, $U_0 = 0,85$ (размерность обеих величин – доля населения столицы). Иными словами, 94% москвичей способны воспринять информацию о существовании Сети, при этом 80% москвичей составляют потенциальную аудиторию Интернета. Коэффициент внутренних воздействий q и интенсивность перехода в состояние использования \bar{h} оказались сравнительно высокими, а интенсивность перехода из состояния знания в состояние наличия желания \bar{a} – относительно низкой. Соответствующая средняя задержка между этими двумя состояниями $1/\bar{a}$ составляет 4,75 года, а среднюю задержку между осведомлённостью москвича о существовании Интернета и началом использования Сети при помощи рис.1.17 мы оценивали в 7 лет [100,107].

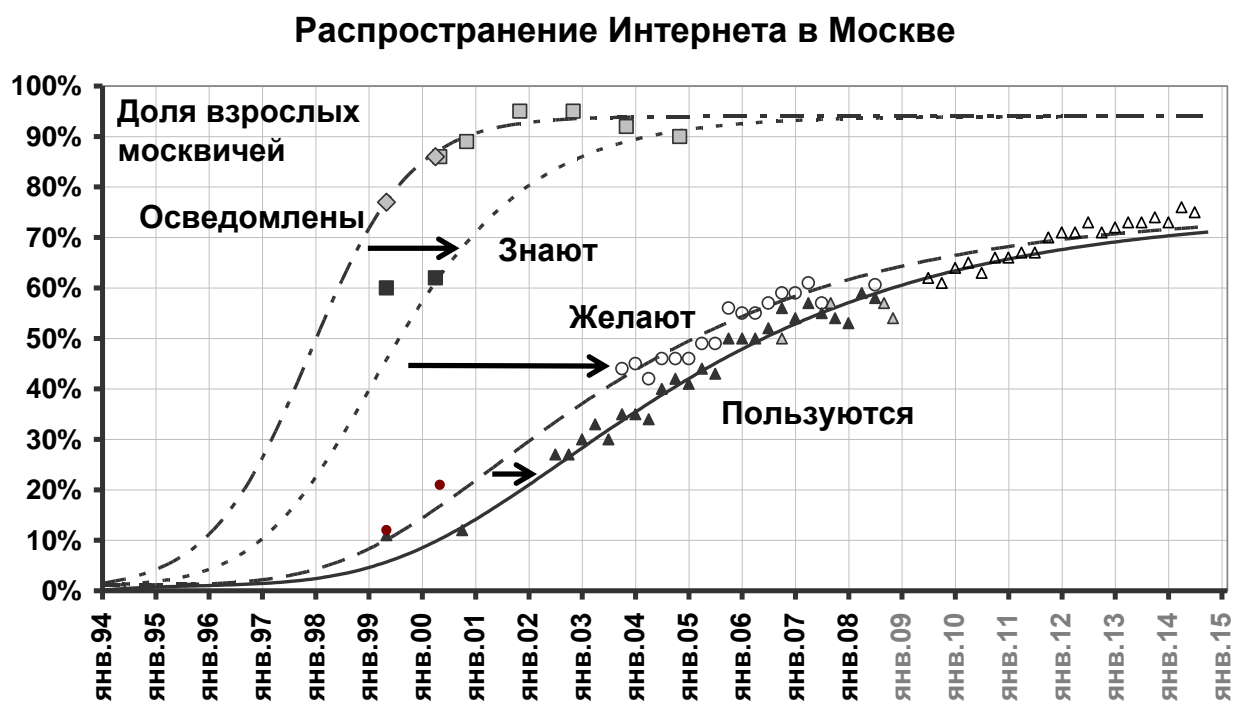


Рис. 1.17. Динамика распространения Интернета в Москве по данным ФОМ, Левада-центра и ВЦИОМ.

Результаты моделирования, опубликованные нами на их основе прогнозы на 2009-2014 гг. [100,107], а также данные массовых социологических опросов изображены на рис.1.17. При построении рис.1.17 и 1.18 использованы следующие обозначения:

- серые прямоугольники – данные Левада-центра о доле осведомлённых;

- серые ромбы – аналогичные данные ФОМ;
- чёрные прямоугольники – данные ФОМ о знании пользователями Интернета ($I(t)/K$);
- белые кружки – данные ФОМ о желании использовать Интернет ($M(t)/K$);
- чёрные треугольники и кружки – данные ФОМ об использовании Интернета, опубликованные до 2009 года ($X(t)/K$);
- серые треугольники – данные ВЦИОМ об использовании Интернета;
- белые треугольники – данные ФОМ об использовании Интернета, опубликованные после публикации нашего прогноза.

Максимальное превышение фактическими данными о количестве пользователей $X(t)$ нашего прогноза составило 5,5% взрослого населения Москвы, а максимальное относительное отклонение фактических данных от прогнозных показателей за шесть лет составило 8%. К сожалению, несмотря на то, что ФОМ опрашивал россиян о намерении (желании) использовать Интернет по крайней мере до 2010 года, открытая публикация данных была прекращена, поэтому сравнить фактические и прогнозные значения количества желающих использовать Сеть мы не можем.

Как показывает рис.1.17, многостадийная модель “справляется” с задачей описания S-образного не логистического роста. Таким образом, отклонения от логистического роста могут быть объяснены не только при помощи процессов воспроизводства населения и неоднородности общества, обсуждавшихся в предыдущих разделах. Оказывается, что даже столь упрощенная многостадийная модель удовлетворительно описывает четыре набора данных, полученные тремя различными социологическими организациями (ФОМ, Левада-центром и ВЦИОМ) на протяжении 15 лет между 1999 и 2014 гг. Мы, однако, предполагаем, что учёт воспроизводства населения необходим и в многостадийных моделях, и без этого усложнения данные о распространении Интернета в Москве в следующие пять лет весьма значительно отклонятся от нашего рис.1.17.

В целом по России величина задержки между осведомлённостью и решением воспользоваться Интернетом – существенно выше, чем в Москве (что очевидно из сравнения рис.1.17 и рис.1.18). В наших работах [100,107,92,108,111] эта задержка была оценена нами в 15 лет. Мы предположили, что столь высокая величина задержки между осведомленностью и желанием воспользоваться может означать, что, несмотря на формальное соответствие данных за 1999-2008 годы нашей модели (в целом лишь незначительно модифицирующей модель Додсона и Мюллера [350]), процесс распространения Интернета в России определяется не распространением информации о Сети в российском обществе, а, во-первых, взаимным влиянием и убеждением (или “подражанием” в духе Г. Тарда [464,271]), и, во-вторых, доступностью услуги. В течение многих лет недоступность Всемирной Сети для россиян (в первую очередь по причине дороговизны), не позволяет сформировать желание стать ее пользователем.

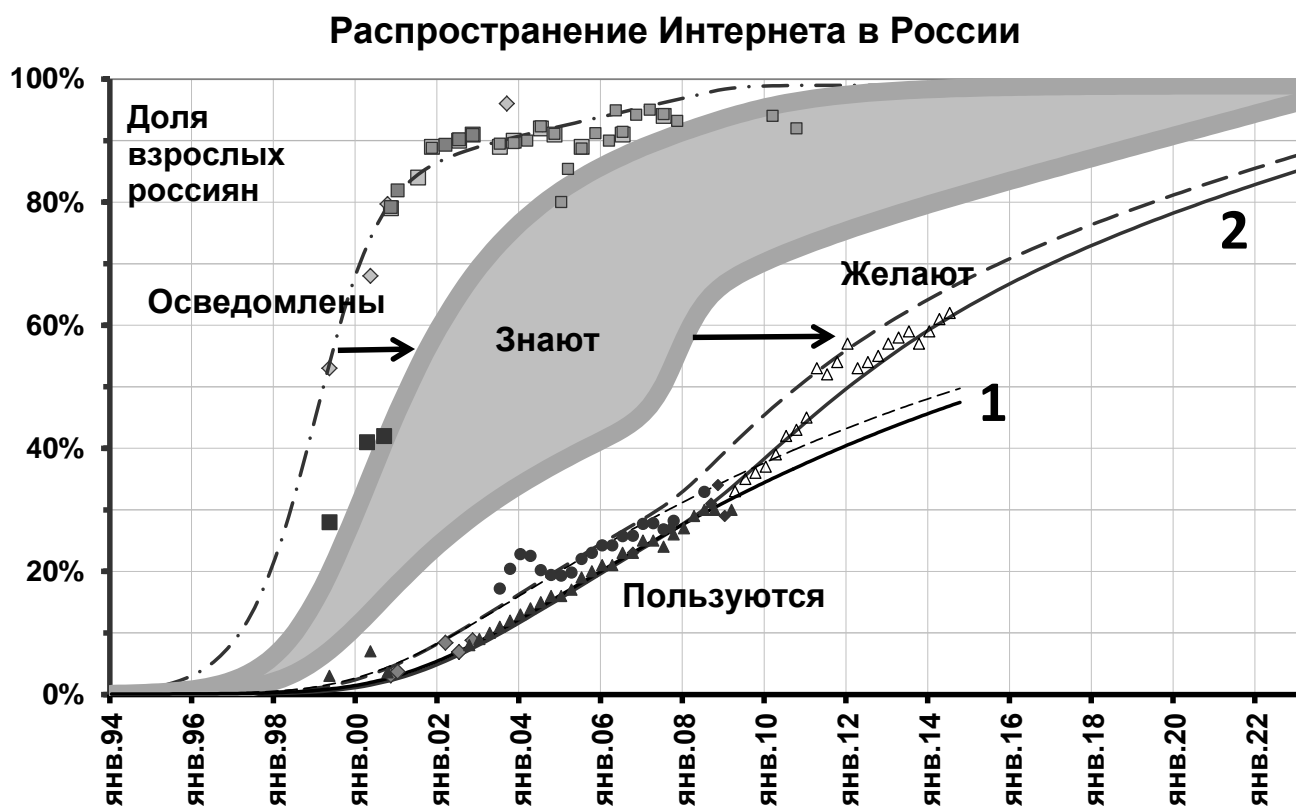


Рис. 1.18. Динамика распространения Интернета в России по данным ФОМ, Левада-центра и ВЦИОМ. Цифрами отмечены наши прогнозы, построенные: (1) – в 2009 и (2) – в 2014 гг.

Как свидетельствует рис.1.18, опубликованный нами прогноз (помеченный на рисунке символом “1”) оказался существенно занижен. Рис.1.18 показывает, что отклонения данных от нашего прогноза составили 12-20 процентов населения России. Эти расхождения вынуждают нас модифицировать изложенную выше модель, включив в неё изменение доступности Интернета.

Наиболее важной причиной расхождений мы считаем изменение доступности Интернета, поэтому ниже представлен способ учёта этого фактора.

Модель, учитывающая рост доступности Интернета

Сравнивая рис.1.17 и рис.1.15, мы заключаем, что показатель распространения Интернета $X(t)$ в Москве лишь незначительно отстаёт от аналогов в развитых странах [129,140]. Тренд доли пользователей Интернета в Москве приблизительно повторяет тренды двух других стран, отставая на пять лет от Швеции и на два года от Германии. Можно предположить, что параметры модели, отвечающие за межличностные коммуникации и принятие решений (\bar{h} , \bar{a} , \bar{b} , q) должны быть приблизительно одинаковы в Москве, в России, и в других странах, а основной причиной отличия показателей являются различия в доступности нововведения. В новой модели мы распространяем приведённое выше “московское” значение параметра $\bar{h} = 6,94$ на всю Россию и фиксируем его.

Основываясь на результатах предыдущих разделов, мы выделяем три сегмента индивидов: “доцифровой” сегмент, “цифровых мигрантов” и “цифровых аборигенов” (см. рис.1.14). Представители “доцифрового” сегмента в нашей модели не только отказываются от использования Интернета, но и отвечают, что ничего о нём не слышали. Для этих поколений мы полагаем $\bar{a} = 0$ и $\bar{h} = 0$, а также $X(t) \equiv 0$, $Y(t) \equiv 0$. Численность доцифровых поколений в нашей модели составляет 23% взрослого населения России в 1994 году, после чего убывает до нуля со скоростью 1,5 процента взрослого населения в год. Два других поколения восприимчивы к нововведению, но поначалу оно не всем доступно.

Разрешив “горизонтальные” переходы между состояниями на рис.1.16, т.е. отменив условие $f_i \equiv 0$, получаем систему уравнений модели, учитывающих динамику доступности Интернета:

$$\frac{d}{dt} \begin{bmatrix} X \\ Y \\ Z \\ A \\ W \\ U_2 \\ V_2 \\ U_1 \\ V_1 \end{bmatrix} = \begin{bmatrix} 0 & \bar{h} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\bar{h} & 0 & \bar{a} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\bar{a} & f & \bar{b} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -f & 0 & \bar{b} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\bar{b} & f & g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\bar{b} - f & 0 & g \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g & f \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g - f \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ A \\ W \\ U_2 \\ V_2 \\ U_1 \\ V_1 \end{bmatrix} \quad (9)$$

По-прежнему, все начальные условия положим нулевыми, кроме исходного количества неосведомлённых $U_1(0) = U_0$, $V_1(0) = K - U_0$ ¹⁰.

Функция $f(t)$, отвечающая за изменение доступности Интернета, заранее неизвестна, требуется построить её модель и прогноз. Мы задаём $f(t)$ в виде интенсивности переходов (в демографических терминах – “силы смертности”), соответствующей функции дожития

$$\Phi(t) = \begin{cases} \tilde{a} + \tilde{b}t + \frac{\tilde{c}}{1 + \exp(-\tilde{d}(t - \tau))} & \text{при } t \leq t^* \\ \Phi(t^*) \leq 1 & \text{при } t \geq t^* \end{cases} \quad (10a)$$

Таким образом

$$f(t) = -\frac{1}{\Phi} \frac{d\Phi}{dt} \quad (10b)$$

Параметры a, b, c, d включаются в число неизвестных параметров модели и будут определены методом наименьших квадратов. Ниже будет показано, что непосредственно $f(t)$ нам не потребуется, поскольку достаточно знать $\Phi(t)$.

¹⁰ Мы обосновываем нулевые начальные условия как содержательными аргументами (малым количеством пользователей и желающих пользоваться Интернетом в начале 90-х гг. XX), так и небольшим количеством доступных нам данных, которое не позволяет нам определить значения слишком большого количества неизвестных. Таким путём мы избегаем проблему моделирования переменных (характеристик процесса), не являющихся *наблюдаемыми* [236].

Определим также функцию $\Psi(t)$, описывающую вероятность, что индивид остался неосведомлённым к моменту t (такую же, как использованная в предыдущих разделах)

$$\Psi(t) = \Psi_0 \exp\left(-\int_0^t g(\tau) d\tau\right), \quad \text{тогда} \quad g(t) = \frac{1}{\Psi} \frac{d\Psi}{dt}.$$

Предположим, что 8-10% неосведомлённых о существовании Интернета, которых выявили опросы Левада-центра и ФОМ – это представители самых старших (“доцифровых”) поколений. Используя результаты разделов 1.1.2.1 и 1.1.2.2. положим

$$\Psi(t) = \begin{cases} \frac{1}{q} - \left[-p + \frac{1}{c \cdot e^{z^2(0)-z^2(t)} + \sigma \cdot \text{daw}(z(t))} \right] \frac{1}{q} & \text{при } t_0 \leq t \leq T \\ 0 & \text{при } t > T \end{cases}. \quad (11)$$

Мы полагаем $T = 20$, а в качестве начала отсчёта выбрали 1.01.1994 года.

После того, как $\Phi(t)$ и $\Psi(t)$ определены, для расчёта $X(t)$ вместо системы (10) достаточно решить три линейных обыкновенных дифференциальных уравнения (основные расчётные уравнения)

$$\frac{dX}{dt} = \bar{h}(M - X) \quad \frac{dM}{dt} = \bar{a}(I(t)(1 - \Phi(t)) - M) \quad \frac{dI}{dt} = \bar{b}(K(1 - \Psi(t)) - I(t)) \quad (12)$$

где $M(t) = X(t) + Y(t)$ – сумма количества пользователей и желающих воспользоваться;

$I(t) = X(t) + Y(t) + A(t) + W(t)$ – общее число знающих о нововведении.

Доказательство.

Рассмотрим сперва часть системы (9), отвечающую за осведомлённость. Введём две вспомогательные переменные:

$R(t) = U_2(t) + A(t) + Y(t) + X(t)$ – количество осведомлённых, имеющих доступ;

$E(t) = V_2(t) + W(t)$ – количество осведомлённых без доступа.

Запишем уравнения для количества осведомлённых и неосведомлённых

$$\begin{aligned} \frac{dE}{dt} &= g(t)V_1 - f(t)E & \frac{dR}{dt} &= g(t)U_1 + f(t)E(t) \\ \frac{dV_1}{dt} &= -(g(t) + f(t))V_1 & \frac{dU_1}{dt} &= -g(t)U_1 + f(t)V_1 \end{aligned}. \quad (13)$$

Этим уравнениям удовлетворяют функции (проверяется дифференцированием)

$$\begin{aligned} E(t) &= K \cdot \Phi(t)(1 - \Psi(t)) & R(t) &= K \cdot (1 - \Phi(t))(1 - \Psi(t)) \\ V_1(t) &= K \cdot \Phi(t)\Psi(t) & U_1(t) &= K \cdot (1 - \Phi(t))\Psi(t) \end{aligned} \quad (14)$$

Определим теперь $N(t) = E(t) + R(t)$ – общее число осведомлённых;

$F(t) = U_1(t) + R(t)$ – общее число имеющих доступ.

Подставляя в эти определения решения (14), получаем

$$N(t) = E(t) + R(t) = K(1 - \Psi(t)) \quad (15)$$

$$F(t) = U_1(t) + R(t) = K(1 - \Phi(t)). \quad (16)$$

Способ расчёта $N(t)$ при помощи интеграла Дюсона дан в предыдущих разделах.

Тогда мы можем выразить решения в интуитивно понятном виде

$$\begin{aligned} E(t) &= (K - F(t))N(t)/K & R(t) &= F(t)N(t)/K \\ V_1(t) &= (K - F(t))(K - N(t))/K & U_1(t) &= F(t)(K - N(t))/K \end{aligned} \quad (17)$$

Приступим теперь к задаче вычисления количества знающих.

Определим $S(t) = I(t) - W(t)$ как общее число знающих, имеющих доступ. Для $S(t)$

и $I(t)$ из системы (9) следуют уравнения

$$\frac{dI}{dt} = \bar{b}(N(t) - I(t)) \quad (18)$$

$$\frac{dS}{dt} = \bar{b}U_2 + f(t)W \quad (19)$$

Первое уравнение является линейным и решается в квадратурах, поскольку $N(t)$ нам известна (формула 15).

Покажем, что выражение

$$S(t) = \frac{F(t)I(t)}{K} \quad (20)$$

является решением уравнения (19).

Дифференцируя (20), и учитывая $\frac{dF}{dt} = -K \frac{d\Phi}{dt} = f(t)(K - F)$ (см. 16 и 10b), получаем

$$\frac{dS}{dt} = \frac{F}{K} \frac{dI}{dt} + \frac{I}{K} \frac{dF}{dt} = \frac{F\bar{b}(N - I)}{K} + \frac{If(t)(K - F)}{K} = b \left(\frac{FN}{K} - S \right) + f(t)(I - S) = b(R - S) + f(t)(I - S).$$

Здесь мы воспользовались полученной выше формулой (17): $R(t) = F(t)N(t)/K$.

Но $R - S = U_2$, а $I - S = W$. Таким образом, доказано, что выражение (20) является решением уравнения (19).

С учётом введённых определений из (9) получим систему уравнений:

$$\frac{dX}{dt} = \bar{h}(M - X) \quad \frac{dM}{dt} = \bar{a}(S - M) \quad \frac{dI}{dt} = \bar{b}(N - I) . \quad (21)$$

Чтобы превратить систему (21) в (12), достаточно в (21) вместо $S(t)$ подставить выражение (20), а вместо $N(t)$ – выражение (15).

Утверждение доказано.

Последнее упрощение, которое мы здесь предпринимаем, состоит в том, что мы распространяем систему (12) на произвольный момент времени, хотя, как отмечено в предыдущих разделах, она остаётся в силе лишь в течение T лет, пока живы представители “доцифрового” поколения. Это допущение упрощает вычисления, но является одним из наиболее уязвимых мест данной модели.

Построенный прогноз получен при помощи расчёта по формулам (10)-(12), и незначительно отличается от полученных ранее в разделе 1.1.2. Отличие моделей состоит в том, что в модели раздела 1.1.2 доминирующую роль играли межличностные коммуникации, в то время как в построенной здесь модели – внешние воздействия и переходы между стадиями. Тот факт, что обе модели породили приблизительно одинаковые прогнозы, означает, что мы не можем с уверенностью утверждать, что определяющую роль при распространении Интернета играют именно межличностные, а не массовые коммуникации.

Что касается прогнозирования, то, как и в разделе 1.1.2, мы предполагаем, что наибольшее важным для построения точного прогноза является определение потенциала распространения нововведения (в данном разделе роль потенциала распространения выполняет $S(t)$ в уравнениях (21)). Относительное влияние каналов коммуникаций (межличностные коммуникации или воздействие СМИ) при этом трудно определить, используя одни лишь данные массовых социологических опросов.

Вообще говоря, возможность аппроксимации одних и тех же статистических показателей при помощи существенно различных моделей – это

общее свойство социально-экономических систем, и процесс диффузии нововведений не составляет исключения. Более того, различные управленческие задачи могут требовать построения различных моделей одних и тех же явлений [25]. В силу сказанного мы рекомендуем российским социологическим организациям включить в регулярные опросы россиян об использовании Интернета вопросы о том, имеет ли кто-либо из членов семьи респондента доступ к сети Интернет, и о том, намерены ли респонденты сами начать пользоваться Сетью.

Наобходимо отметить, что Р. Роуз [450], изучая распространение Интернета в России, выделил дополнительно подмножество индивидов, которые не пользуются Сетью, но знакомы с пользователями и, следовательно, лучше подготовлены для восприятия этого нововведения. Такое подмножество отсутствует в нашей модели, поскольку оно отчасти нарушает иерархию стадий, и поскольку Р.Роуз опубликовал данные только одного опроса.

Следует также отметить ограниченность нашей попытки представить многостадийную модель как марковский процесс с дискретными стадиями и непрерывным временем. В наших моделях пользователь переходит в следующую стадию независимо или под действием контакта с каким-либо индивидом, находящимся на более высокой стадии овладения нововведением. При этом, однако, игнорируется длительность нахождения индивида на той или иной стадии. В случае, когда переход на новую стадию неправомерно описывать процессами без последствия, на наш взгляд, интересные возможности предоставляют модели, разработанные О.В.Барояном и Л.А.Рвачевым [21] и их коллегами для моделирования эпидемий гриппа, где вводится дополнительная переменная, описывающая время нахождения индивида на определенной стадии. Однако, внедрение такой модели требует более объемных вычислений, чем модели, рассматриваемые в нашей работе.

По итогам применения описанных здесь многостадийных моделей, мы приходим к двум выводам. Во-первых, как показывают рис.1.17 и рис.1.18,

различия между состояниями осведомлённости и желания использовать Сеть, не сводятся к простой задержке во времени. Во-вторых, для моделирования перехода от состояния желания использовать Сеть к фактическому использованию, достаточно простого временного сдвига¹¹.

В то время как область применения количественных многостадийных моделей распространения инноваций в качестве инструмента анализа систем весьма широка, успешное применение этого инструмента при моделировании каждой новой предметной области потребует постановки и решения специфических вспомогательных задач, подобных задачам компенсации искажающего влияния прибора. Необходимость решения таких задач вызвана спецификой вида доступных исследователю данных о распространении нововведений, который в каждом случае имеют свои особенности. Следующие разделы данной главы, а также вторая глава диссертационной работы посвящены решению подобных вспомогательных задач.

1.3.2 Совместное моделирование количества пользователей и абонентов

В настоящее время оценки количества пользователей передаются российским Министерством связи и массовых коммуникаций в Международный союз электросвязи (МСЭ), после чего публикуются совместно с соответствующими показателями других стран. Кроме того, МСЭ публикует данные о количестве абонентов доступа к Интернету, предоставляемые государствами-членами этой организации. Такие данные представляются нам ценными, поскольку данные о пользователях, в силу различия методик измерения, часто оказываются несопоставимы, и автору часто приходилось слышать предположение, что статистика количества абонентов должна дать более устойчивую картину распространения Интернета в развивающихся странах. Действительно, интуитивно такой показатель кажется проще и понятнее и не требует изучения

¹¹ К аналогичному выводу можно прийти, изучая данные опросов Левада-центра о распространении банковских карточек.

выборочного метода и особенностей его реализации. Как показывает изучение данных МСЭ, в западноевропейских странах количество абонентов сети Интернет является более стабильным показателем, чем количество пользователей. Однако это оказывается не всегда так в развивающихся странах, где статистические органы иногда существенно изменяют, а то и вообще прекращают публикацию количества абонентов.

Поскольку домохозяйства на два порядка многочисленнее предприятий, можно предположить, что количество абонентов характеризует в первую очередь домашнее использование Интернета. В таком случае количество абонентов отражает более интенсивное использование Интернета индивидами, чем то, которое изучалось в первом разделе этой главы.

На рис. 1.19 изображены результаты моделирования данных о численности пользователей и абонентов сети Интернет в Швеции, опубликованных Международным союзом электросвязи (МСЭ) на веб-сайте (icteye.ict.int), а также веб-сайтом “Показатели достижения целей развития, сформулированных в Декларации Тысячелетия” ООН (mdgs.un.org).

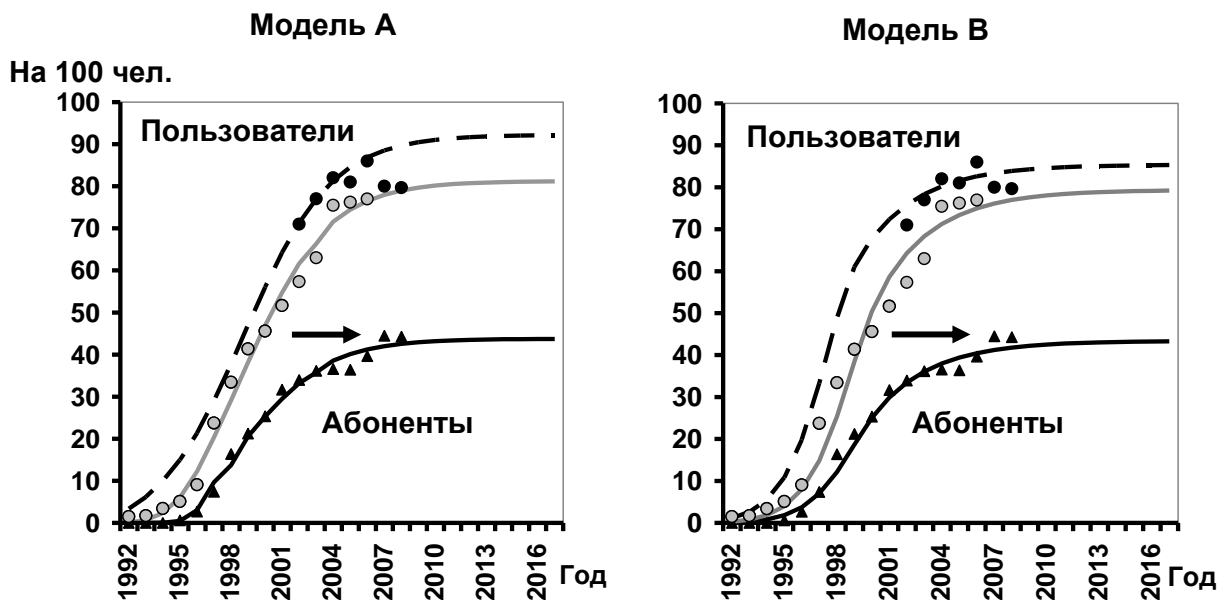


Рис. 1.19. Сравнение двух моделей численности пользователей и количества абонентов сети Интернет в Швеции

Для применения многостадийной модели к данным о пользователях и абонентах выбрана Швеция, а не Российская Федерация, по следующим причинам. Во-первых, распространение Интернета в Швеции, в основном, завершилось (см. рис. 1.1,1.19), что позволяет изучать процесс диффузии этой услуги на протяжении значительной части ее жизненного цикла. Во-вторых, данные МСЭ о России, на наш взгляд, чрезвычайно неточны. МСЭ сообщает о 1,3 миллионах абонентов сети Интернет в России в 2006 году и 21,5 миллионах абонентов – в 2007 году. При этом, количество пользователей (не абонентов!) в 2007 году было оценено в 21,14 миллиона человек. По-видимому, российское телекоммуникационное ведомство до 2007 года не сообщало в МСЭ актуальный показатель количества абонентов, а позже сочло целесообразным считать его эквивалентным количеству пользователей¹². Мы полагаем, что российская методика подсчета количества пользователей и абонентов постепенно будет согласована с европейской, но сейчас нам приходится ограничиться европейскими данными.

На рис. 1.19 стрелками показано направление перехода индивидов от стадии “использования” к стадии подписки на услугу. На стадии использования индивид может и не быть абонентом Интернета, а использовать Сеть в офисе, в учебном заведении, в интернет-кафе или у друзей. Черными кружками на рис. 1.19 изображена т.н. полная аудитория, т.е. доля лиц в возрасте от 16 до 74 лет, использующих Интернет хотя бы раз в течение 12 месяцев, а серыми кружками – “активные” пользователи, использовавшие Интернет в течение последнего квартала. Черные треугольники показывают количество абонентов. Сплошными линиями изображены результаты моделирования (синтетические данные).

В левой части рис. 1.19 представлены результаты, полученные при помощи модели А, представляющей собой описанную в разделе 1.3.1 модификацию модели Додсона и Мюллера [350]. Здесь предполагается, что:

¹² О.Н.Вершинская и О.А.Алексеева [38], анализируя причины снижения позиции России в рассчитанном ООН и МСЭ Индексе готовности стран к электронному правительству, отмечают, что позиция России по подындексу развития телекоммуникаций определена по статистическим данным за 2005 год, т.е. с трёхлетним запаздыванием.

а) некоторая часть полной аудитории постепенно начинает использовать Интернет активно (хотя бы раз в квартал);

б) некоторая часть активной аудитории постепенно подписывается на услугу доступа к Интернету, увеличивая количество абонентов.

В обоих случаях под “постепенностью” подразумевается экспоненциально распределенная во времени задержка. Не все пользователи становятся абонентами, поскольку для использования Интернета достаточно одного подписчика услуг на домохозяйство. В модели **A** общество образует т.н. клику, в которой каждый индивид оказывает влияние на каждого. Общество “атомизовано” в том смысле, что на процесс не влияет наличие каких-либо ячеек общества, например, семей или домохозяйств.

В правой части рис. 1.19 представлены результаты вычислений при помощи модели **B**, которая аналогична модели **B**, подробно описанной далее в третьей главе настоящей работы и в нашей работе [100] и использует следующие допущения:

1. Условный “глава семейства” первоначально обучается использованию Интернета на работе, в учебном заведении, в интернет-кафе или у друзей, так он становится пользователем.

2. Домохозяйство, глава которого научился использовать Интернет, спустя некоторое время приобретает доступ к Интернету, задержка распределена экспоненциально (в общем случае, некоторые главы семейств могут и не подключить свои квартиры, но в случае Швеции отвечающий за такую задержку дополнительный параметр модели оказался нулевым).

3. Период времени между приобретением домохозяйством доступа к Интернету и началом его использования несколькими взрослыми членами (период обучения) в этой работе мы считаем пренебрежимо малым (иными словами, одновременно с абонентом появляется и некоторое количество новых пользователей – для Швеции мы получили величину 0,97).

4. “Главы семейств” становятся пользователями как под действием внешних факторов (СМИ, рекламы, наличия Интернета в офисе), так и под действием других пользователей (вне зависимости от того, являются ли они абонентами).

Таким образом, модель **В** можно представить системой уравнений

$$\begin{aligned} \frac{du_1}{dt} &= \left(p + q \frac{u}{K} \right) (K_1 - u_1) \\ \frac{ds}{dt} &= \frac{a}{\tau} u_1 - \frac{1}{\tau} s \\ u_2 &= b \cdot s \\ u &= u_1 + u_2 \\ K &= K_1 + K_2 \end{aligned} \tag{22}$$

с нулевыми начальными условиями и ограничением $K \leq 100\%$. Параметры $p, q/K, K_1, a, b, \tau$ могут быть найдены путем подгонки синтетических и наблюдаемых данных о количестве пользователей $u(t)$ и абонентов $s(t)$ методом наименьших квадратов. Для расчета активной аудитории используется отдельная простейшая модель с экспоненциальной задержкой, аналогичная модели **А**.

Как демонстрирует рис. 1.19, обе модели (**А** и **В**) с приблизительно одинаковой погрешностью описывают все три вида доступных данных – о количестве пользователей, активных пользователей и абонентов. Подобные результаты получены нами для ряда других стран Северной и Западной Европы. На рис. 1.19 прогнозные значения на 2010-2016 гг. практически совпадают благодаря тому, что процесс распространения Интернета в Швеции практически закончен. Однако в случае стран, где этот процесс находится на более ранней стадии, не следует надеяться, что различные модели приведут к одинаковым прогнозам. Для обоснованного выбора модели (**А** или **В**) данных, публикуемых в настоящее время МСЭ и ООН, недостаточно, что приводит нас к необходимости использовать дополнительные данные опросов о распространении Интернета в домохозяйствах. Такие данные собраны в странах Евросоюза и систематически публикуются проектом Eurostat. За пределами Евросоюза необходимые данные, как и данные о количестве абонентов Сети Интернет, как правило, разрозненны, но иногда могут быть найдены на сайтах национальных статистических служб.

1.3.3 Многостадийная модель охватов аудитории сети Интернет

В монографии В.Н.Цыгичко [286] рассмотрены три постановки задачи прогнозирования социально-экономического процесса, самой трудной из которых является третья, когда не задан оператор прогнозирования, т.е. отсутствует математическая (а иногда и вербальная) модель процесса. Известно много моделей, описывающих динамику количества пользователей нововведения (модель Басса, пороговые модели, модели байесовского обучения и т.п.), однако модели изменения интенсивности использования нововведения практически отсутствуют. Хотя Э.Роджерс [449] включил в свою модель принятия решения об использовании нововведения стадии применения (implementation) и подтверждения (confirmation), в фокусе его внимания была модификация нововведения самими пользователями (а не производителем) на этих стадиях. Нам же больше интересует вопрос о том, как определить, сколько индивидов будут пользоваться сетью Интернет и/или отдельными крупными информационными ресурсами [8] ежедневно, сколько – несколько раз в неделю, и сколько – лишь несколько раз в месяц, и т.д. Безусловно, можно попытаться поставить интенсивность использования Сети в зависимость от стоимости доступа к ней и благосостояния индивида и построить соответствующую микроэкономическую модель. Однако в настоящее время систематические данные о стоимости доступа, которые были бы собраны одним и тем же способом на протяжении нескольких лет, отсутствуют. Кроме того, даже члены одного и того же домохозяйства, которые платят за доступ к Интернету сообща, могут использовать Сеть с различной интенсивностью, таким образом, потребуется выделить несколько сегментов пользователей и исследовать поведение каждого.

Альтернативой может служить эмпирический подход, при котором для каждого вида данных (т.е. для “суточной”, “недельной” и др. аудиторий) независимо строится модель, например, логистическая. Подобный подход использовался М.И.Левиным и соавторами в работе [211], где впервые были построены модели интенсивности использования Интернета в России. К

сожалению, при обосновании такого способа приходится сделать предположения, которые не выглядят обоснованными. Например, в подобных моделях негласно предполагается, что на индивидов, которые будут использовать сеть ежедневно, могли повлиять только такие же активные пользователи, но не те, кто пользуется реже. Кроме того, сложно обеспечить взаимную согласованность таких моделей.

Ниже описана разработанная автором модель интенсивности (частоты) использования новых информационных технологий, которая преодолевает этот недостаток эмпирического подхода. Эта модель была применена в 2011 году автором и специалистами Фонда “Общественное мнение” для решения задачи прогнозирования интенсивности (частоты) использования Интернета.

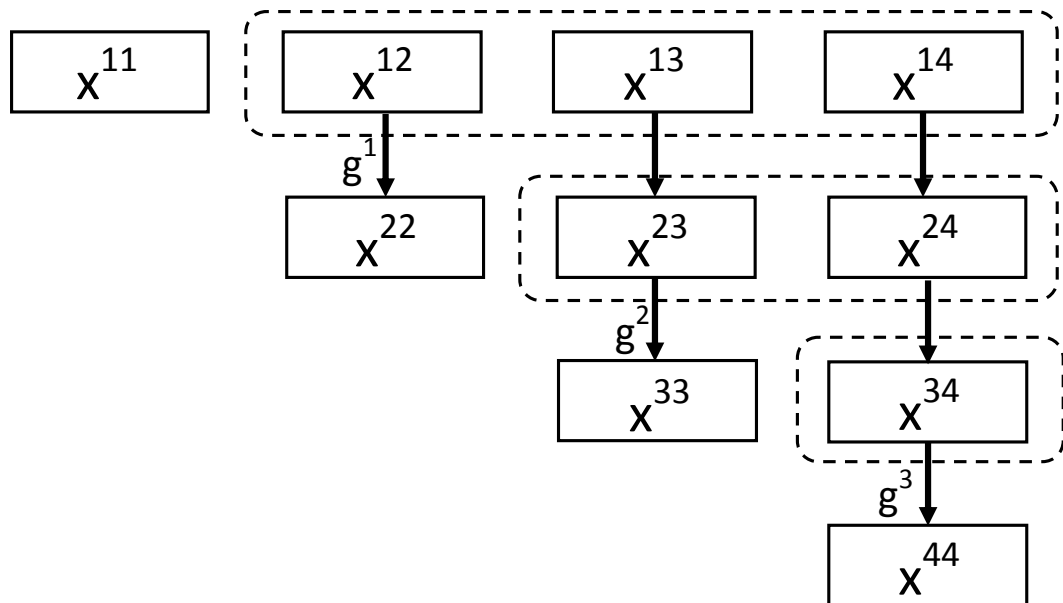


Рис.1.20. Смена состояний пользователей Интернета.

В предлагаемой нами модели предполагается, что для описания интенсивности использования Сети достаточно выделить небольшое число уровней использования и соответствующих этим уровням классов пользователей. Пусть, например, первому уровню соответствует использование Интернета *в среднем* один раз в сутки, второму – использование Интернета в среднем один раз в неделю; третьему – один раз в месяц; четвёртому – один раз в полгода. Всё

множество индивидов в каждый момент времени разбивается на сегменты по окончательной интенсивности использования Интернета. Представители первого сегмента всегда будут пользоваться Сетью редко (в среднем, не чаще, чем раз в полгода), представители второго – в перспективе могут перейти во второй класс по интенсивности, третьего – в третий и т.д.

Описанные сегменты изображены на рис.1.20 вертикальными последовательностями прямоугольников. Далее представители выделенных сегментов делятся по текущей интенсивности использования на несколько подмоществ. Между состояниями возможны переходы, например, $x^{13} \rightarrow x^{23} \rightarrow x^{33}$. Строки прямоугольников на рис.1.20 объединяют индивидов, использующих в данный (а не конечный) момент новую технологию с заданном уровнем интенсивности. Первый верхний индекс обозначает текущий уровень использования, а второй – окончательный уровень.

Систему уравнений баланса (уравнение 1 раздела **1.1**) дополним уравнениями для нескольких возможных состояний пользователя (опуская нижние индексы, отвечающие за пол и географию индивида):

$$\begin{aligned} \frac{\partial x^{1,1}}{\partial t} &= -\mu(t, \pi)x^{1,j} + f^{1,j}(t, \pi) \\ \frac{\partial x^{1,j}}{\partial t} &= -g^1(t, \pi)x^{1,j} - \mu(t, \pi)x^{1,j} + f^{1,j}(t, \pi) \quad \text{для } j > 1 \\ \frac{\partial x^{i,j}}{\partial t} &= g^{i-1}(t, \pi)x^{i-1,j} - g^i(t, \pi)x^{i,j} - \mu(t, \pi)x^{i,j} + f^{i,j}(t, \pi) \end{aligned} \quad (23)$$

$$\frac{\partial x^{n,j}}{\partial t} = g^{i-1}(t, \pi)x^{n-1,j} - \mu(t, \pi)x^{n,j} + f^{n,j}(t, \pi).$$

Здесь $x^{i,j}(t, \pi)$ – это доля тех, кто использует Интернет с интенсивностью уровня i и повысит её до j , $g^i(t, \pi)$ – функции риска смены уровня интенсивности с $i-1$ -го на i -й, $f^{i,j}(t, \pi)$ – источник новых пользователей, который начинают использовать Сеть с интенсивностью уровня i и в перспективе доберутся до j .

Требуется также сформулировать гипотезы относительно вида $f^{i,j}(t, \pi)$. Разумно было бы предположить, что

$$f^{i,j} = \rho^{i,j}(t, \pi)h(t, \pi)y(t, \pi), \quad (24)$$

$$\text{где } \rho^{i,j}(t, \pi) = \frac{x^{i,j}(t, \pi)}{\sum_{i,j} x^{i,j}(t, \pi)}, \quad (25)$$

т.е. что распределение по интенсивности новых пользователей пропорционально текущему распределению. В окончательных расчётах использовалась более простая модель: $\rho^{1,n}(t, \pi) = 1$, остальные $\rho^{i,j}(t, \pi) = 0$. Иными словами, предполагалось, что все новые пользователи начинают с минимальной интенсивности использования Сети, а заканчивают – максимальной. Таким образом, оказалось, что достаточно лишь одного сегмента – четвёртого.

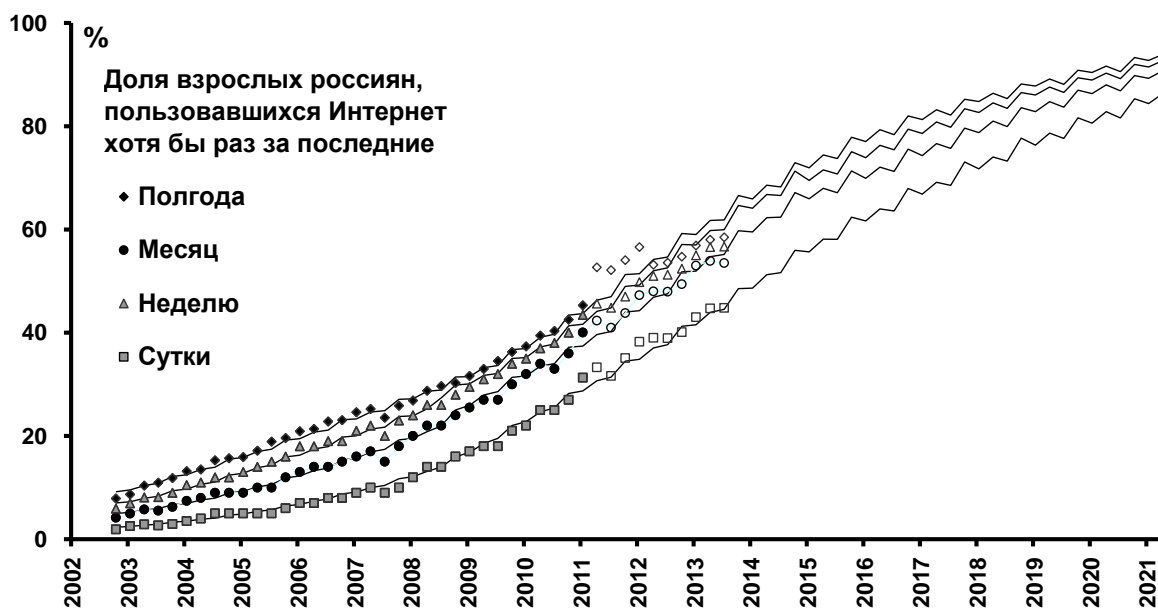


Рис.1.21. Прогноз охватов аудитории российского Интернета за сутки, неделю, месяц и полгода

Несмотря на очевидную грубость такого предположения, рис.1.21 иллюстрирует удовлетворительные прогнозы суточной и недельной аудиторий российского Интернета, сделанные на его основе. Здесь закрашенными значками изображены охваты аудиторий, измеренные ФОМ по зиму 2011 года включительно. Незакрашенные значки изображают данные, полученные после построения прогноза. Как показывает рис.1.21, за два с половиной года (с весны

2011 по лето 2012 года) максимальные ошибки суточного и недельного охватов аудитории составили 3 п.п., месячного охвата – 5 п.п..

Основной недостаток представленных здесь расчётов заключается в том, что модель применялась непосредственно к данным опросов Фонда “Общественное мнение”, которые, строго говоря, содержали не информацию о частоте использования Интернета, а информацию о моменте последнего использования Сети. Таким образом, мы применяли для описания и прогнозирования данных ФОМ модель, требующую других данных. Такой шаг был вынужденным, поскольку вспомогательная модель, предоставляющая “интерфейс” к данным ФОМ ещё не была разработана, а прогноз уже был востребован. Во второй главе данной работы представлены “интерфейсные” модели интенсивности использования Интернета в целом и отдельных информационных ресурсов, а также прогнозы, построенные на их основе.

Выводы

1. Разработана количественная модель распространения нововведения с учетом изменения его доступности и процессов воспроизводства населения, которая объясняют историческую динамику количества пользователей сети Интернет. На основе этой модели с использованием данных массовых опросов построены прогнозы доли пользователей Интернета в России, а также – отдельно в российских городах и сёлах. Для России в целом средняя квадратическая ошибка прогноза доли пользователей в возрасте 18 лет и старше за четыре года (с лета 2009 по весну 2013 г.) составила 3 процентных пункта, максимальная ошибка – 6 п.п. Максимальная ошибка в отдельной возрастной группе составила 12 п.п. – для возрастной группы от 18 до 24 лет.
2. Разработана многостадийная модель роста интенсивности использования нововведения, с помощью которой построены прогнозы суточного, недельного и месячного охватов аудитории российского Интернета. За два с половиной года (с весны 2011 по лето 2012 года) максимальные ошибки суточного и недельного охватов аудитории составили 3 п.п., месячного охвата – 5 п.п.

ГЛАВА 2. МОДЕЛИ ИНТЕНСИВНОСТИ ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И РЕСУРСОВ

Рассмотренные в первой главе данной работы модели распространения новых информационных технологий позволяют описать только динамику полного охвата аудитории, но не интенсивность использования Интернета различными группами пользователей, что существенно ограничивает их применение. В настоящее время организации, работающие в Интернет-индустрии, практически перестали использовать в своей операционной деятельности показатель полной аудитории Интернета. Сегодня индустрией востребованы оценки и прогнозы охватов аудитории Интернета и отдельных информационных ресурсов за месяц, неделю и сутки¹³.

Поэтому при внедрении разработанных моделей в практику нам потребовалось разработать специальную модель для описания интенсивности использования Интернета и информационных ресурсов, которая позволила бы прогнозировать охваты за сутки, неделю и месяц. Теоретические основы информационных ресурсов, методы их описания и классификации, организации и мониторинга рассмотрены в монографии А.Б.Антопольского и О.В.Шлыковой [8]. В разделе 1.3.3 мы описали разработанную нами многостадийную модель распространения интенсивности использования Интернета, в которой полугодовая аудитория выступала потенциалом для месячной, месячная – потенциалом для недельной и т.п. Такой подход, однако, не удовлетворяет критерию теоретической обоснованности. Во-первых, основанием классификации аудиторию на “месячную”, “недельную” и “суточную” является сложившаяся практика социологических организаций и индустрии, однако, не исключено, что пропущены какие-нибудь важные сегменты аудитории, например, индивиды, использующие Сеть приблизительно один раз в две недели. Во-вторых, в наших расчётах использовались данные измерений ФОМ, которые, по традиции

¹³ В итоге фонд “Общественное мнение” (ФОМ) перестал публиковать оценки полной аудитории, хотя и продолжает проводить её измерения и предоставляет необходимые данные автору данной работы. ВЦИОМ и Левада-центр продолжают публиковать оценку полной аудитории Интернета.

медиаизмерений имеют форму т.н. кумулятивных кривых охватов аудитории. Именно прогнозы этих кумулятивных кривых востребованы индустрией. Однако модели распространения нововведений описывают поведение индивидов, а не охваты аудитории.

Поэтому необходима дополнительная модель, обеспечивающая “интерфейс” между теорией распространения нововведений и теорией случайных процессов, объясняющей формирование кумулятивных кривых охвата аудитории. Такая модель разработана в первом разделе данной главы. Здесь рассмотрены различия в способах измерения аудитории Интернета и два способа аппроксимации распределения интенсивности использования Интернета. Методологической основой раздела служит теория случайных процессов, в частности – теория потоков событий. В первом случае плотности распределения интенсивности использования дискретны и отличаются лишь количеством “пиков” или “точечных масс”. Во втором случае распределения интенсивности аппроксимируются смесью гамма-распределений. Оба подхода позволяют аппроксимировать данные измерений ФОМ, РОМИР и ВЦИОМ. Выбор в пользу непрерывной плотности распределения обосновывается при помощи данных онлайн-панели, работа которой описана в разделе 2.2.

В разделе 2.2 предлагается модель распределения интенсивности использования информационных ресурсов Интернета в виде смеси бета-распределений. Эта модель позволяет аппроксимировать данные измерений онлайн-панели и отчасти обосновывает выбор непрерывного распределения и в разделе 2.1.

В разделе 2.3 представлена модель использования информационного ресурса WWW с учётом смены пользователями файлов cookies, позволяющая интерпретировать измерения Интернет-счётчиков, таких как LiveInternet, Rambler’s Top 100, Top.Mail.Ru и другие. Проведённая работа была инициирована критикой данных социологических опросов со стороны аналитиков Интернет-индустрии в 2007-2008 гг.

2.1 Количественная модель интенсивности использования сети Интернет

В 2002-2008 г. ежеквартально открыто публиковались результаты репрезентативных массовых опросов фонда “Общественное мнение” (ФОМ), которые сообщали, каков суточный, недельный, месячный, квартальный и полугодовой *охват* аудитории российского Интернета.

Поставим вопрос, какую информацию об *использовании* Интернета гражданами содержат данные об охватах? Можно ли с их помощью узнать, какое количество россиян использует Интернет *каждый день*? Или, например, *несколько раз раз в неделю*? Сколько людей использует Интернет реже, чем один раз в неделю? Именно в такой форме рекомендуют опрашивать респондентов инструкции поддерживаемых ООН проектов международных сопоставлений использования Интернета в различных странах. Очевидно, что прямого ответа на такие вопросы охваты не дают. Прямой ответ можно было найти в публикациях компании РОМИР за 2003-2007 гг., но в 2007 году публикация таких данных прекратилась. В настоящее время вопросы в похожей форме задают ВЦИОМ и Левада-центр, данные которых, к сожалению, публикуются не чаще двух раз в год. Возникает вопрос, нельзя ли получить показатели использования расчетным путем, используя данные об охватах аудитории?

В качестве методологической базы исследования в этой главе используется теории случайных процессов. *Случайным процессом* $N(t)$ называется процесс, значение которого при любом фиксированном $t = t_0$ является случайной величиной $N(t_0)$ [37]. *Потоком событий* называется последовательность однородных событий, появляющихся одно за другим в случайные моменты времени [37]. В этой главе использования информационных технологий и ресурсов индивидом моделируется при помощи потока событий. Количество событий $N(t, \tau)$, появившихся за промежуток времени длиной τ , есть случайный процесс (процесс накопления событий). Обозначим *интенсивность* потока [37]

событий буквой λ , это непрерывная неотрицательная величина. В общем случае интенсивность потока событий определяется как математическое ожидание числа событий в единицу времени

$$\bar{\lambda}(t) = \lim_{\tau \rightarrow 0} \frac{\langle N(t, \tau) \rangle}{\tau}.$$

Здесь угловые скобки означают математическое ожидание случайной величины. Для стационарного потока событий λ постоянна во времени. Для стационарного потока определим также вероятность того, что на промежутке $(t_0, t_0 + t)$ произошло хотя бы одно событие

$$w(t) = \sum_{k=1}^{\infty} \hat{v}_k(t) = 1 - \hat{v}_0(t).$$

Параметр потока событий определяется как предел

$$\bar{\lambda} = \lim_{t \rightarrow 0} \frac{w(t)}{t}.$$

Известно, что для любого стационарного потока событий $\lambda \geq \bar{\lambda}$, а для стационарных ординарных потоков, в частности, для простейшего потока событий, $\lambda = \bar{\lambda}$. Далее мы используем понятия интенсивности и параметра потока только для простейших потоков событий, поэтому в пределах этой главы обе величины всегда будут совпадать.

В первой главе данной работы основной интерес представляла величина $x^{i,j}(t, \pi)$ – плотность количества пользователей, которые сейчас используют Интернет с интенсивностью класса i , но со временем увеличат её и перейдут в класс j . Мы предполагаем, что значения интенсивности λ могут отличаться для различных пользователей. Определим в момент времени t плотность количества пользователей на единицу возраста, использующих Интернет с интенсивностью, меньшей λ

$$\tilde{X}^{i,j}(t, \pi, \lambda).$$

Тогда $0 \leq \tilde{X}^{i,j}(t, \pi, \lambda) \leq x^{i,j}(t, \pi)$, причём

$$x^{i,j}(t, \pi) = \lim_{\lambda \rightarrow \infty} \tilde{X}^{i,j}(t, \pi, \lambda).$$

Из определения $\tilde{X}^{i,j}(t, \pi, \lambda)$ следует, что эта величина – не убывает по λ , и её производная по λ неотрицательна

$$\tilde{x}^{i,j}(t, \pi, \lambda) = \frac{\partial \tilde{X}^{i,j}}{\partial \lambda} \geq 0.$$

Введённая здесь величина $\tilde{x}^{i,j}(t, \pi, \lambda)$ имеет смысл плотности количества индивидов на единицу возраста и на единицу интенсивности. Очевидно,

$$\int_0^\lambda \tilde{x}^{i,j}(t, \pi, \varphi) d\varphi = \tilde{X}^{i,j}(t, \pi, \lambda) \quad \text{и} \quad \int_0^\infty \tilde{x}^{i,j}(t, \pi, \lambda) d\lambda = x^{i,j}(t, \pi).$$

Мы предлагаем модель величины $\tilde{x}^{i,j}(t, \pi, \lambda)$ в виде

$$\tilde{x}^{i,j}(t, \pi, \lambda) = p_i(\lambda; t, \pi) \cdot x^{i,j}(t, \pi),$$

где $p_i(\lambda; t, \pi)$ – некоторая (заранее неизвестная) плотность вероятности случайной величины λ , определённой на $[0, \infty)$. Мы налагаем на $p_i(\lambda; t, \pi)$ требование конечности математического ожидания величины λ

$$\bar{\lambda} = \int_0^\infty \lambda \cdot p_i(\lambda; t, \pi) d\lambda < \infty,$$

которое обеспечивает конечность параметра (и интенсивности) итогового потока событий.

Кроме того, мы предполагаем, что

1. для интересующего нас периода измерения охватов T длиной полгода величину λ можно считать постоянной для каждого индивида;
2. для представителей i -го класса в момент времени t показатель λ есть случайная величина, распределённая по некоторому закону распределения;
3. закон распределения для каждого i -го класса не изменяется во времени;
4. законы распределения для всех классов принадлежат к одному семейству, хотя могут отличаться параметрами.

Такая модель не противоречит сформулированным выше определениям и свойствам $\tilde{x}^{i,j}(t, \pi, \lambda)$. Отметим, что $p_i(\lambda; t, \pi)$ зависит лишь от текущего класса интенсивности i , но не от индекса окончательного класса j .

В разделе 2.1.1. для краткости вместо $p_i(\lambda; t, \pi)$ используется обозначение $p(\lambda)$. Таким образом, $p(\lambda)$ есть вероятность использования Интернета с интенсивностью («частотой») λ раз в сутки ($\lambda \geq 0$). Величину $p(\lambda)$ можно интерпретировать и как *долю* аудитории, которая посещает Интернет с интенсивностью λ .

2.1.1 Разработка модели интенсивности (частоты) использования сети Интернет

Пусть аудитория Интернета (или некоторая её целевая часть) неоднородна и содержит как очень активных, так и менее активных пользователей. Объём выделенного сегмента аудитории будем предполагать постоянным на рассматриваемом интервале измерения длительности охватов и обозначим $R_0 = x^{i,j}(t, \pi)$. В работах [291,292] для относительных величин, аналогичных по смыслу $x^{i,j}(t, \pi) / k^{i,j}(t, \pi)$ используется термин "предельный охват".

Уточним, что подразумевается под использованием информационного ресурса с частотой "λ раз в сутки". Рассмотрим пользователей, для которых посещение информационного ресурса описывается *простейшим потоком событий* с интенсивностью λ . Простейшим называется ординарный стационарный поток случайных событий без последствия или, иными словами, стационарный пуассоновский поток. Для такого потока событий [37]:

1. пользователи появляются в Интернете поодиночке, а не группами (ординарность);
2. интенсивность использования Интернета не меняется со временем (стационарность);
3. вероятность выхода пользователя в Интернет не зависит от предшествующей истории его поведения (отсутствие последствия).

Состояния пользователя и переходы между ними для простейшего перехода событий изображены на рис.2.1 (А). Такую модель поведения индивидов в 1959

году Э.Эренберг предложил для описания статистики транзакций баз данных о поведении покупателей [359], этот подход и используется в моделях поведения потребителя. В литературе известен подход с использованием модели поведения пользователя и в виде потока Эрланга 2-го порядка [331], однако, как будет показано ниже, для описания и прогнозирования охватов аудитории Интернета, модель простейшего потока событий является достаточной. Правомерность предположения о стационарности потока обсуждается в конце раздела 2.1.

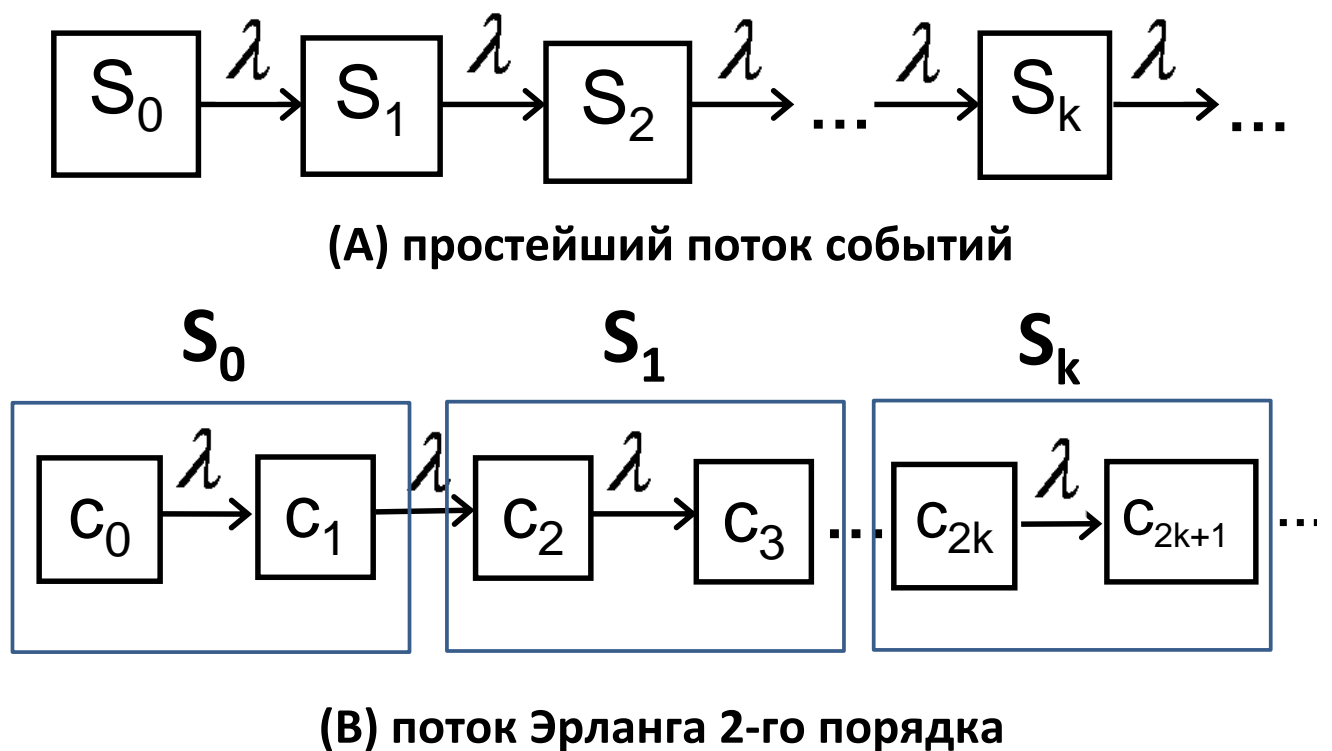


Рис.2.1. Состояния пользователя и переходы между ними: (А) – в простейшем потоке событий; (В) – в потоке Эрланга 2-го порядка.

Функция распределения (ф.р.) интенсивности использования информационного ресурса λ вычисляется при помощи интегрирования

$$P(\lambda) = \int_0^{\lambda} p(\varphi) d\varphi \quad (1)$$

и представляет собой вероятность того, что представитель такой аудитории использует Интернет *реже*, чем λ раз в сутки. Очевидно, что $P(\lambda \rightarrow \infty) \rightarrow 1$.

В этой главе нам потребуется распределение Эрланга непрерывной переменной $0 \leq \tau < \infty$, (имеющей смысл длительности промежутка времени, в течение которого измеряются охваты), зависящее от параметра λ , а именно

$$f_k(\tau; \lambda) = \lambda \frac{(\lambda\tau)^{k-1}}{(k-1)!} e^{-\lambda\tau} \text{ — плотность распределения Эрланга } k\text{-го порядка;} \quad (2)$$

$$F_k(\tau; \lambda) = 1 - e^{-\lambda\tau} \sum_{m=0}^{k-1} \frac{(\lambda\tau)^m}{m!} \text{ — функция распределения Эрланга } k\text{-го порядка.} \quad (3)$$

По определению функции распределения (ф.р.) и плотности распределения (п.р.)

$$\frac{dF_k(\tau; \lambda)}{d\tau} = f_k(\tau; \lambda). \quad (4)$$

Далее мы будем использовать следующее из (2) и (3) рекуррентное соотношение

$$f_k(\tau; \lambda) = \lambda(F_{k-1}(\tau; \lambda) - F_k(\tau; \lambda)). \quad (5)$$

Вероятность того, что индивид, характеризуемый параметром потока λ , за промежуток времени длиной τ обращался к ресурсу ровно k раз, описывает распределение Пуассона (распределение дискретной величины k)

$$\hat{u}_k(\tau; \lambda) = e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!} = \frac{1}{\lambda} f_{k+1}(\tau; \lambda). \quad (6)$$

Математическое ожидание количества индивидов, характеризуемых параметром λ , которые обращались к ресурсу ровно k раз за промежуток времени длиной τ , есть $R_0 p(\lambda) \hat{u}_k(\tau; \lambda)$.

Определим *абсолютный кумулятивный охват порядка k* аудитории как математическое ожидание количества представителей этой аудитории, которые использовали информационный ресурс *не менее k* раз за время τ . Тогда абсолютный кумулятивный охват или просто охват порядка k за время τ аудитории, использующей выбранный информационный ресурс с частотой λ , описывается формулой [452]:

$$R_k(\tau; \lambda) = R_0 \cdot F_k(\tau; \lambda) p(\lambda) \quad (7)$$

для $k = 1, 2, \dots$ и т.д. Действительно, количество тех, кто использовал ресурс *не менее k* раз, складывается из тех, кто делал это *ровно k* раз, $k+1$ раз и т.п.

$$R_k(\tau; \lambda) = R_0 \cdot p(\lambda) (\hat{u}_k(\tau; \lambda) + \hat{u}_{k+1}(\tau; \lambda) + \dots) = R_0 \cdot p(\lambda) F_k(\tau; \lambda). \quad (8)$$

Если же величина R_0 обозначает не общее число пользователей Интернета, а долю таковых в населении страны (или в иной целевой аудитории), то величину $R_k(\tau; \lambda)$ будем называть *относительным охватом* аудитории. В работах [291,292, 256] именно такая величина называется охватом аудитории.

Вероятность того, что произвольно выбранный индивид за время τ обращался к ресурсу *ровно* k раз, в данной модели составляет

$$\bar{u}_k(\tau) = \int_0^{\infty} \hat{u}_k(\tau; \lambda) p(\lambda) d\lambda = \frac{\tau^k}{k!} \int_0^{\infty} \lambda^k e^{-\lambda\tau} p(\lambda) d\lambda. \quad (9)$$

В частности вероятность, что пользователь не использовал ресурс, можно рассчитать при помощи преобразования Лапласа плотности $p(\lambda)$

$$\bar{u}_0(\tau) = \int_0^{\infty} p(\lambda) e^{-\lambda\tau} d\lambda. \quad (10)$$

Выражением (9) дается вероятностная смесь распределений Пуассона с параметром λ .

Дифференцируя k раз обе части (10) и сравнивая с (9), получаем формулу

$$\bar{u}_k(\tau) = \frac{\tau^k}{k!} \left(-\frac{d}{d\tau} \right)^k \bar{u}_0(\tau), \quad (11)$$

позволяющую найти любую $\bar{u}_k(\tau)$, если удалось найти выражение для $\bar{u}_0(\tau)$. Подчеркнём, что столь простая формула верна только в том случае, когда поведение индивида можно описать простейшим потоком событий.

Кумулятивный охват порядка k всей аудитории информационного ресурса может быть найден путем интегрирования всех охватов $R_k(\tau; \lambda)$ по λ

$$\bar{R}_k(\tau) = \int_0^{\infty} R_k(\tau; \lambda) p(\lambda) d\lambda = R_0 \int_0^{\infty} F_k(\tau; \lambda) p(\lambda) d\lambda. \quad (12)$$

Из (5), (10) и (12) вытекает рекуррентное соотношение

$$\bar{R}_k(\tau) = \bar{R}_{k-1}(\tau) - R_0 \bar{u}_{k-1}(\tau), \quad (13)$$

означающее, что разность k -го и $k-1$ -го охватов есть величина той части аудитории, которая использовала ресурс *ровно* k раз.

Разностное уравнение (13) следует дополнить граничным условием

$$\bar{R}_0(\tau) \equiv R_0. \quad (14)$$

Формулы (10), (11) и (13) позволяют рекуррентно рассчитать кумулятивные охваты аудитории ресурсов в следующих, важных для практики, случаях.

Пример 1. Отрицательное биномиальное распределение.

Зададим плотность распределения λ виде плотности гамма-распределения [37]:

$$p(\lambda; b, \alpha) = b \frac{(b\lambda)^{\alpha-1}}{\Gamma(\alpha)} e^{-b\lambda}, \quad (15)$$

где использована гамма-функция Эйлера

$$\Gamma(\alpha) = \int_0^{\infty} \theta^{\alpha-1} e^{-\theta} d\theta \quad (\alpha > 0). \quad (16)$$

Соответствующая функция распределения вероятностей имеет вид

$$P(\lambda; b, \alpha) = \frac{b^\alpha}{\Gamma(\alpha)} \int_0^\lambda \Lambda^{\alpha-1} e^{-b\Lambda} d\Lambda. \quad (17)$$

Заменим в (17) $y = b\Lambda$, тогда

$$P(\lambda; b, \alpha) = \frac{b^\alpha}{\Gamma(\alpha)} \int_0^\lambda \left(\frac{y}{b}\right)^{\alpha-1} e^{-y} \frac{1}{b} dy = \frac{1}{\Gamma(\alpha)} \int_0^{\lambda/b} y^{\alpha-1} e^{-y} dy. \quad (18)$$

В выражении справа стоит т.н. «регуляризованная» неполная гамма-функция. Подпрограммы для ее вычисления есть в стандартных статистических пакетах, в том числе и в прикладном пакете MS Excel (с помощью которого получен, в частности, рис.2.14).

Подставляя (17) в (10), получаем

$$\bar{u}_0(\tau) = \int_0^{\infty} p(\lambda) e^{-\lambda\tau} d\lambda = \frac{b^\alpha}{(b+\tau)^\alpha}. \quad (19)$$

Подставляя затем (19) в (11) находим вероятность того, что индивид за время τ обращался к ресурсу k раз

$$\bar{u}_k(\tau) = \frac{(\alpha)_k \tau^k}{k!} \frac{b^\alpha}{(b+\tau)^{\alpha+k}}, \quad (20)$$

где использован символ Похгаммера

$$(\alpha)_k = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)} = \prod_{i=1}^k (\alpha + i - 1), \quad (21)$$

таким образом, $(\alpha)_0 = 1$, $(\alpha)_1 = \alpha, \dots$, $(\alpha)_k = \alpha \cdot (\alpha + 1) \cdot \dots \cdot (\alpha + k - 1)$.

Распределение вероятностей целой величины k с параметрами $0 < p < 1$ и $\alpha > 0$

$$\Pr\{K = k\} = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)k!} p^\alpha (1 - p)^k, \quad k = 0, 1, 2, \dots \quad (22)$$

носит название *отрицательного биномиального распределения* (NBD). Очевидно,

формула (20) превращается в (22), если положить $p = \frac{b}{b + \tau}$.

Отрицательное биномиальное распределение использовалось в прикладных задачах уже в работах Дж.У.Юла в 1910 году [482] и М.Гринвуда и Дж.У.Юла в 1920 году [369]. В 1959 г. Эндрию Эренберг описал при помощи NBD статистику транзакций баз данных продуктовых сетей [359,332], позже применил модель NBD для описания охвата аудитории телеканалов. В работе Парка и Фейдера [439] отрицательное биномиальное распределение используется для моделирования аудитории WWW-сайтов. В наших работах [72,135] NBD применяется для моделирования аудитории российского Интернета в целом.

При помощи (13) находим все охваты аудитории информационного ресурса

$$\bar{R}_k(\tau) = \bar{R}_{k-1}(\tau) - R_0 \frac{(\alpha)_{k-1}}{(k-1)!} \left(\frac{b}{b + \tau} \right)^\alpha \left(\frac{\tau}{b + \tau} \right)^{k-1}. \quad (23)$$

В частности, при $k = 1$ получаем значения охвата

$$\bar{R}_1(\tau) = R_0 - R_0 \left(\frac{b}{b + \tau} \right)^\alpha. \quad (24)$$

Формула (24) используется нами ниже при разработке модели **D** использования информационного ресурса пользователем.

При этом охваты $\bar{R}_k(\tau)$ мы вычисляем по формулам (13) и (14), а для вычисления $\bar{u}_k(\tau)$ используем вытекающую из (20) рекуррентную формулу

$$\bar{u}_k(\tau) = \frac{\alpha + k - 1}{k} \frac{\tau}{b + \tau} \bar{u}_{k-1}(\tau). \quad (25)$$

Пример 2. В следующем разделе этой главы используется бета-распределение вероятности использовать информационный ресурс в течение промежутка времени Δt , в качестве которого обычно используются одни сутки. Это соответствует плотности распределения интенсивности использования ресурса

$$p(\lambda) = \frac{\Delta t}{\text{B}(\alpha, \beta)} (1 - e^{-\lambda \Delta t})^{\alpha-1} e^{-\beta \lambda \Delta t}.$$

Заменив переменную $\theta = e^{-\lambda \Delta t}$, получим $d\lambda = -\frac{1}{\Delta t} e^{\lambda \Delta t} dz$ и найдём аналог (20)

$$\bar{u}_0(\tau) = \frac{1}{\text{B}(\alpha, \beta)} \int_0^1 (1 - \theta)^{\alpha-1} \theta^{\tau/\Delta t + \beta - 1} d\theta = \frac{\text{B}(\alpha, \beta + \tau/\Delta t)}{\text{B}(\alpha, \beta)}.$$

Полученная формула позволяет рассчитать теоретические охваты

$$\bar{R}_1(\tau) = R_0 \left(1 - \frac{\text{B}(\alpha, \beta + \tau/\Delta t)}{\text{B}(\alpha, \beta)} \right),$$

т.е. все синтетические данные, которые мы можем сравнить с эмпирическими данными опросов Фонда “Общественное мнение”. Ценность полученного результата состоит в том, что он позволяет на основе одной и той же модели сопоставить данные массовых опросов и описанных в следующем разделе панельных исследований (например, опросов ФОМ и ВЦИОМ и панели TNS).

Для вычисления $\bar{u}_k(\tau)$ получаем рекуррентную формулу

$$\bar{u}_k(\tau) = \frac{1}{(\Delta t)^{k-1}} \frac{1}{k} \sum_{m=0}^{k-1} \frac{(-1)^{m+1} \tau^{m+1}}{m!} (\psi^{(m)}(\beta + \tau/\Delta t) - \psi^{(m)}(\alpha + \beta + \tau/\Delta t)) \cdot \bar{u}_{k-1-m}(\tau),$$

где $\psi^{(m)}(z) = \frac{d^m \psi^{(0)}(z)}{dz^m} = \frac{d^{m+1}}{dz^{m+1}} \ln \Gamma(z)$ – полигамма-функция, для вычисления которой во многих математических пакетах есть подпрограммы, и которую несложно вычислить рекуррентно при целых значениях β . После этого охваты $\bar{R}_k(t)$ можно рассчитать рекуррентно по формуле (13).

Смесь полученных распределений может стать важной альтернативой используемой далее смеси NBD, благодаря связи с обобщёнными полиномами Бернштейна, при помощи которых, как показали Хиршман и Уиддер [284], можно аппроксимировать широкий класс заданных на полупрямой плотностей распределения $p(\lambda)$.

Пример 3. Дискретные распределения

Рассмотрим дискретное распределение частоты λ ¹⁴. В этом случае функция распределения $P(\lambda)$ является разрывной кусочно-постоянной (ступенчатой)

$$P(\lambda) = \sum_{i=1}^M p_i \cdot \sigma(\lambda - \lambda_i), \quad (26)$$

где $\sigma(\lambda) = \text{sign}(\lambda) + 1/2$ – ступенчатая функция Хевисайда. Плотность такого распределения состоит из M дельта-функций Дирака:

$$p(\lambda) = \sum_{i=1}^M p_i \cdot \delta(\lambda - \lambda_i). \quad (27)$$

Подставив выражение для $p(\lambda)$ в (12), получим формулы охватов аудитории:

$$\bar{R}_k(\tau) = R_0 \sum_{i=1}^M p_i F_k(\tau; \lambda_i). \quad (28)$$

Обозначим $r_i = R_0 p_i$. Тогда формула охвата первого порядка за период τ для дискретных распределений λ приобретает вид

$$\bar{R}_1(\tau) = \sum_{i=1}^M r_i \cdot (1 - \exp(-\lambda_i \tau)). \quad (29)$$

Эта формула используется в следующем разделе для расчёта охватов $\bar{R}_1(\tau)$ в описанных ниже моделях **A1-A4**.

Далее для моделирования упоминавшихся выше данных РОМИР и ВЦИОМ об интенсивности использования Интернета мы используем величину

$$Q(\tau) = R_0(1 - P(1/\tau)), \quad (30)$$

представляющую размер аудитории, которая в среднем использует Интернет *не реже*, чем один раз за τ дней. Сказанное не следует трактовать как утверждение, что ни один пользователь никогда не пропустит τ дней – иногда он выходит в Интернет реже, а иногда чаще, но *в среднем* за достаточно большой промежуток

¹⁴ Соответствующее этой модели распределение *времени* между двумя обращениями произвольно выбранного индивида к информационному ресурсу носит название *гиперэкспоненциального*; плотность этого распределения имеет вид

$$\varphi(T) = \sum_{i=1}^M p_i \cdot \lambda_i e^{-\lambda_i T}.$$

времени T оказывается, что пользователь обращался к Интернету или информационному ресурсу не менее, чем T/τ раз.

Исследования РОМИР определяли недельную аудиторию следующим образом: “пользующиеся Интернетом не реже одного раза в неделю, но не ежедневно”. Сложив количество таких пользователей с количеством тех, кто пользуется Интернетом ежедневно, мы получаем $Q(\tau)$ для интервала времени длиной в одну неделю. Таким образом, РОМИР публиковала¹⁵ величину $Q(\tau)$ при нескольких значениях τ , равных суткам, неделе и месяцу. В частности, в четвёртом квартале 2005 г., который был выбран нами в [72] для исследования накопленной аудитории, $Q(30) = 0,17$.

К сожалению, РОМИР не сообщал, какова “полная” аудитория Интернета, то есть полное число использующих Сеть с ненулевой интенсивностью λ , поэтому для оценки этой полной аудитории мы использовали данные ФОМ. Согласно данным ФОМ, в 2008 году для взрослого населения России можно было считать, что 70% взрослого российского населения не использовали Интернет. Соответственно, для этого периода можно положить $Q(\infty) \approx 0,3$, а в четвёртом квартале 2005 г. в соответствии с данными ФОМ следовало положить $Q(\infty) \approx 0,2$.

Отметим, что остаётся не до конца исследованным вопрос о влиянии нестационарности, то есть о том, как влияет на итоги опросов респондент, который обычно пользуется Интернетом каждый день, но на последней неделе один день провел в командировке, и не проверил почту. Такие пользователи могут увеличить недельную аудиторию в ущерб ежедневной. Безусловно, среди пользователей найдутся и такие, которые пользуются Интернетом раз в неделю, но последнюю неделю провели на в туристическом походе или на курорте. Их ответы уменьшат недельную аудиторию. К сожалению, одни лишь данные массовых социологических опросов не позволяют оценить влияние нестационарности. Однако в следующем разделе мы приводим данные и

¹⁵ См., например, http://romir.ru/news/res_results/333.html

результаты моделирования аудитории онлайн-панелей, которые свидетельствуют, что влияние нестационарности на кумулятивные охваты аудитории невелико.

Основное предположение этого раздела состоит в том, что можно найти функции распределения $P(\lambda)$, способные удовлетворительно описать как данные ФОМ об охватах аудитории $\bar{R}_1(\tau)$, так и данные РОМИР о $Q(\tau)$. В таком случае, несмотря на небольшой объём данных, предоставляемый каждой из организаций, мы можем получить аппроксимацию $P(\lambda)$, достаточно точную для целей прогнозирования. Первоначально мы ожидали, что совместное использование различных данных массовых опросов позволяет найти весьма узкие границы для $P(\lambda)$ и рассчитывали найти одно простое распределение с малым числом параметров, пригодное для описания любых данных такого рода. Однако оказалось, что можно найти различные аппроксимации распределения $P(\lambda)$, которые существенно различаются по свойствам, но при этом с одинаковой точностью описывают все доступные нам данные массовых опросов. Сказанное проиллюстрировано ниже на примере ступенчатого распределения λ (модель **A** и **B**) и смеси гамма-распределений (модель **D**). Разрешающая способность данных о $\bar{R}_1(\tau)$ и $Q(\tau)$ при данном уровне погрешности измерений оказывается недостаточно высока, чтобы ограничить $P(\lambda)$. Мы пришли к выводу, для целей прогнозирования можно использовать различные аппроксимации $P(\lambda)$, в частности – в виде смеси гамма-распределений или в виде ступенчатых функций. Отметим, что задача аппроксимации $P(\lambda)$ при наличии конечного объёма данных оказывается непосредственно связанной с классической проблемой моментов [13,14] и, таким образом, всегда допускает бесконечное множество решений.

Для рассматриваемых далее моделей **A1-A4**, использующих дискретные распределения λ , теоретические охваты описывает формула (28), свободные параметры есть $\mathbf{m} = (r_1, \lambda_1, \dots, r_M, \lambda_M)$, а их общее количество составляет $2M$. Далее вместо $\bar{R}_1(\tau)$ будем использовать обозначение $\bar{R}_1(\tau; \mathbf{m})$, чтобы подчеркнуть зависимость значений теоретических охватов от свободных параметров модели.

В случае моделей **A1** и **A2** мы ищем такие параметры модели r_i, λ_i , чтобы расчетная кривая $\bar{R}_1(\tau; \mathbf{m})$ проходила как можно ближе к данным опросов d_k , полученным ФОМ. На параметры налагаются вытекающие из смысла задачи требования $r_i \geq 0, \lambda_i \geq 0$. Мы говорим, что параметры модели \mathbf{m}^* являются оптимальными \mathbf{m} , подразумевая, что они минимизируют функционал ошибок

$$\Phi(\mathbf{m}^*) = \min_{\mathbf{m}} \Phi(\mathbf{m}) \quad (31)$$

где

$$\Phi(\mathbf{m}) = \sum_{k=1}^{K_d} (d_k - \bar{R}_1(\tau_k; \mathbf{m}))^2, \quad (32)$$

– сумма квадратов отклонений расчетных (“синтетических”) данных $\bar{R}_1(\tau_k; \mathbf{m})$ от наблюдений ФОМ d_k , K_d – количество точек данных, в случае моделей **A1** и **A2** использовались только данных ФОМ, поэтому – $K_d = 5$ (сутки, неделя, месяц, квартал и полгода). Поскольку точность измерений ФОМ во всех пяти точках близка, мы не взвешиваем ошибки при помощи каких-либо весов. Для других моделей функционал ошибок имеет более сложный вид и будет описан в каждом случае отдельно.

Синтетические (расчётные) данные РОМИР, которые используются для оценки результата в моделях **A1** и **A2** и непосредственно для поиска оптимальных параметров моделях **A3** и **A4**, можно получить по формуле

$$Q(\tau) = \sum_{i=1}^M r_i \sigma(\tau - 1 / \lambda_i), \quad (33)$$

которую несложно вывести из (30) и (26). Эта формула верна только для дискретных распределений величины λ .

Модель A1. Однородная аудитория.

Рассмотрим сперва однородную модель, в которой все пользователи одинаковы и используют Интернет приблизительно с равной интенсивностью λ_1 (или выходят в Интернет в среднем каждые $1/\lambda_1$ суток). Такая модель (модель простейшего потока событий) чрезвычайно широко используются в технике, в

моделях массового обслуживания, в частности – при моделировании телетрафика и компьютерных сетей [171,40,201]. Эта модель оказалась чрезмерно упрощённой, даже для описания только лишь охватов $\bar{R}_1(\tau)$.

В случае модели **A1** есть только один сегмент пользователей $M=1$, и, соответственно два свободных параметра. Оптимальные значения параметров модели **A1**, минимизирующие функцию ошибки (32), оказались равны

$$\lambda_1^* = 0,19 \text{ суток}^{-1}, r_1^* = 18\% .$$

Иными словами, согласно модели **A1**, осенью 2005 г. общая аудитория Рунета составляла 18% взрослых россиян, каждый из которых выходил в Интернет приблизительно один раз в пять дней. При этом поведение каждого пользователя описывает простейший поток событий.

Даже при “оптимальных” значениях параметров модель однородной аудитории **A1** значительно (на несколько процентов населения России) отклоняется как от данных ФОМ, так и от данных РОМИР. Относительная ошибка превышает 10% для данных ФОМ и 30% для данных РОМИР. Такие расхождения многократно превышают нашу оценку [163] погрешности измерений ФОМ и РОМИР в этот период.

По нашей оценке [163] в рассматриваемый период относительная ошибка выборки ФОМ использовавшей 18 тысяч респондентов, составляла 1,7%. Выборка РОМИР-Мониторинг включала 1,6 тысяч респондентов, и относительная ошибка этой выборки осенью 2005 г. составляла 5-6%. Таким образом, отклонения предсказаний модели от данных пятикратно превосходят ошибки выборочных исследований, поэтому такая модель не подходит для аппроксимации данных социологических опросов об использовании Интернета

Модель A2. Два сегмента пользователей.

Удовлетворительного описания данных ФОМ об охватах $\bar{R}_1(\tau)$ можно описать при помощи моделей с двумя и большим числом сегментов пользователей. Самой простой из них является модель **A2**, в которой выделяется лишь два сегмента пользователей. Здесь будет показано, что модель **A2** способна

удовлетворительно аппроксимировать данные ФОМ об охватах $\bar{R}_1(\tau)$, хотя всё ещё не позволяет одновременно аппроксимировать и данные РОМИР о $Q(\tau)$.

Пусть представители первого сегмента используют Интернет в среднем λ_1 раз в сутки, а второй – λ_2 раз в сутки, а плотность распределения частоты использования Интернета имеет вид (27) с $M = 2$.

Тогда на вопросы интервьюеров РОМИР представители первого сегмента будут отвечать, что используют Интернет несколько раз за $t_1 = 1/\lambda_1$ дней, а второго – несколько раз за $t_2 = 1/\lambda_2$ дней.

Данные ФОМ описывает формула (28) при $M = 2$ с неизвестными нам пока параметрами $r_1, r_2, \lambda_1, \lambda_2$. Мы находим оптимальные параметры путем минимизации функционала ошибки (30); эти параметры оказываются равны:

$$\lambda_1 = 0,754 \text{ суток}^{-1}, r_1 = 10,9\%, \lambda_2 = 0,024 \text{ суток}^{-1}, r_2 = 8,8\%.$$

Оба сегмента оптимальной модели оказались приблизительно равны по количеству пользователей. Первый сегмент представляет активных пользователей, которые выходят в Интернет почти каждый день (один раз в 1,32 дня). Второй сегмент состоит из значительно менее активных пользователей, которые выходят в Интернет реже, чем раз в месяц (один раз в 40,1 дня). Модель **A2** с двумя сегментами пользователей значительно лучше описывает данные ФОМ об охватах $\bar{R}_1(\tau)$ (см. левую часть рис.2.2), чем модель **A1**. Ошибка аппроксимации $\bar{R}_1(\tau)$ теперь составляет менее одного процента населения РФ (и укладывается в пределы “трех сигм”). Если бы у нас не было данных РОМИР (см. правую часть рис.2.2), то мы не смогли бы отклонить гипотезу, что аудитория российского Интернета состоит всего лишь из двух однородных сегментов¹⁶. Отсюда следует вывод о том, что с помощью одних только данных ФОМ об $\bar{R}_1(\tau)$ мы вообще не можем доказать наличие аудитории, которая посещает Интернет

¹⁶ Далее будет показано, что двух сегментов аудитории достаточно для описания данных опросов, если каждый сегмент описать не вырожденным распределением λ , а гамма-распределением. Затем аналогичный результат будет получен для онлайн-панели, в этом случае также используются не вырожденные, а бета-распределения.

несколько раз в неделю, но не каждый день (“недельной” аудитории по терминологии РОМИР). Данные ФОМ об охвате аудитории можно было бы объяснить сочетанием активной (ежедневной) и малоактивной аудиторий, причем погрешность аппроксимации не превышала бы ошибку выборочных измерений.

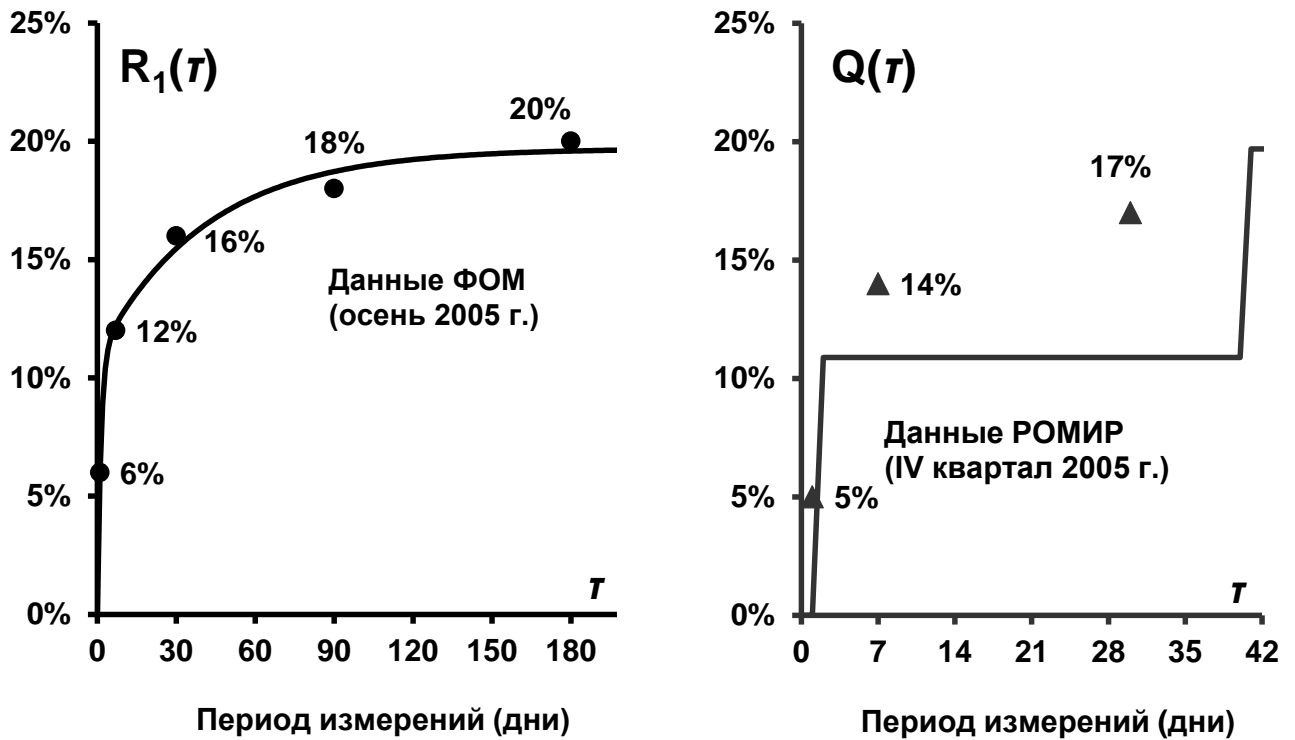


Рис.2.2. Результаты расчетов $\bar{R}_1(\tau)$ и $Q(\tau)$ при помощи модели аудитории **A2**, состоящей из двух сегментов.

Однако благодаря данным РОМИР о $Q(\tau)$, мы знаем, что действительная форма распределения $P(\lambda)$ сложнее, чем сумма двух ступенчатых функций Хевисайда. Правая часть рис.2.2 показывает, что величины $Q(\tau)$, предсказанные моделью **A2**, и реальные данные РОМИР расходятся весьма заметно: относительное отклонение составляет 50%. Данные РОМИР свидетельствуют, что некоторая часть аудитории пользуется Сетью несколько раз в неделю, но не каждый день.

Первый вывод из анализа модели **A2** состоит в том, что для эвристического описания данных ФОМ об охватах $\bar{R}_1(\tau)$ достаточно двух однородных сегментов пользователей: очень активного, и малоактивного. Модель **A2** можно было бы попытаться использовать для целей прогнозирования охватов аудитории по данным ФОМ, невзирая на то, что она противоречит данным РОМИР о $Q(\tau)$.

Второй вывод из нашего анализа состоит в том, что при наличии измерительной погрешности данные ФОМ о кумулятивном охвате аудитории $\bar{R}_1(\tau)$, по-видимому, можно с равной точностью описать с помощью очень широкого набора моделей, не все из которых способны одновременно описать данные других измерений интенсивности использования (например, измерений РОМИР).

Третий вывод состоит в том, что данные других измерений, в частности – данные РОМИР о $Q(\tau)$ позволяют отклонить часть моделей из упомянутого широкого набора, что и служит обоснованием поставленной нами задачи найти модель, описывающую совместно оба вида данных.

Модель **A3**. Три сегмента пользователей.

Чтобы включить в расчеты третий сегмент пользователей, достаточно добавить третий “спектральный пик” плотности распределения, и использовать формулу (27) с $M = 3$. Однако при этом число неизвестных параметров $2M = 6$ превышает объём данных ФОМ об охватах $\bar{R}_1(\tau)$, поэтому мы расширяем множество данных, включая в него и данные РОМИР о $Q(\tau)$. Функционал ошибки необходимо модифицировать, чтобы включить оба вида данных

$$\Phi(\mathbf{m}) = \alpha \sum_{k=1}^{K_d} (d_k - \bar{R}_1[\mathbf{m}](\tau_k))^2 + (1 - \alpha) \sum_{l=1}^{L_d} (\tilde{d}_l - Q[\mathbf{m}](\tau_l))^2. \quad (34)$$

Здесь постоянный коэффициент α является весом данных об охватах $\bar{R}_1(\tau)$, а $1 - \alpha$ является весом данных о $Q(\tau)$. Мы положили $\alpha = K_d / (K_d + L_d) = 5/8$. Постоянная $L_d = 3$ обозначает количество точек данных о $Q(\tau)$ для текущего момента времени.

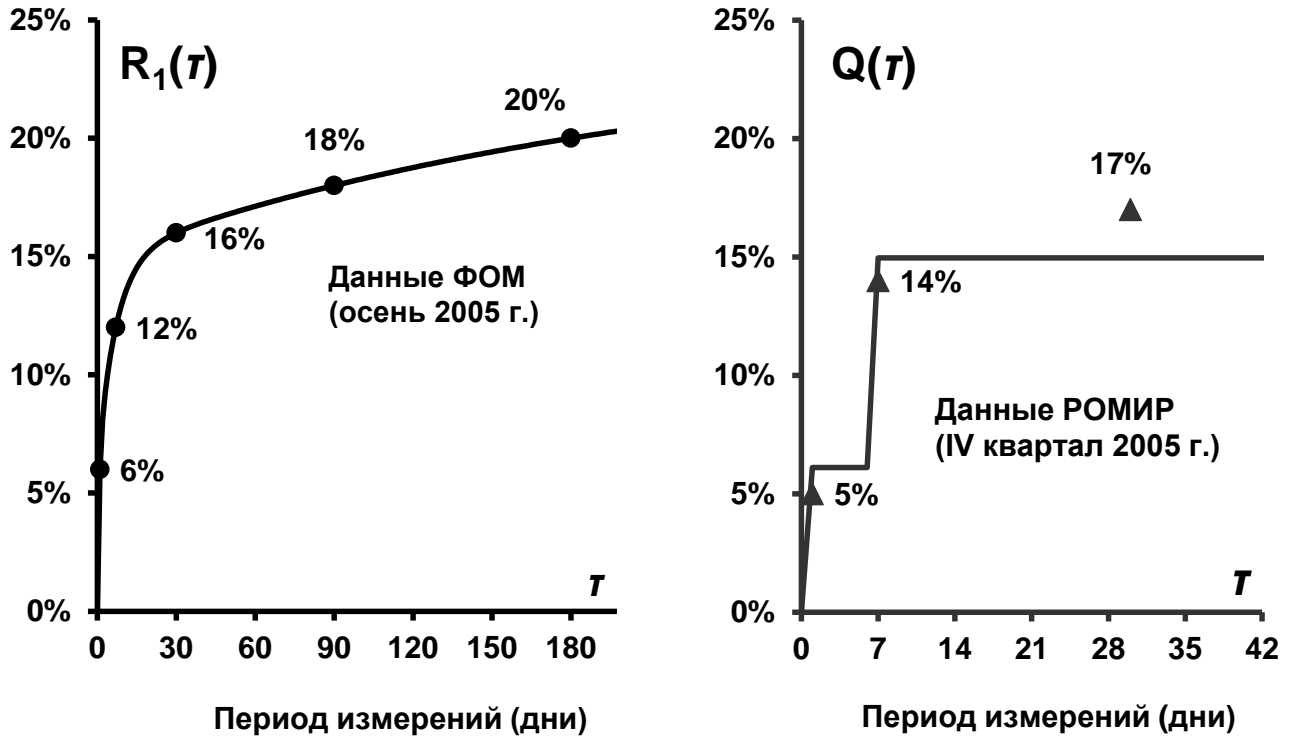


Рис.2.3. Результаты расчетов при помощи модели аудиторрии **A3**, состоящей из трех сегментов

Оптимальные параметры, минимизирующие функцию (34), оказались равны

$$\lambda_1 = 1,293 \text{ суток}^{-1}, r_1 = 6,1\%,$$

$$\lambda_2 = 0,129 \text{ суток}^{-1}, r_2 = 8,9\%,$$

$$\lambda_3 = 0,04 \text{ суток}^{-1}, r_3 = 8,9\%.$$

Число свободных параметров (шесть) приблизилось к объему данных, которыми мы располагаем: пятью измерениями ФОМ и тремя измерениями РОМИР. Если данные РОМИР (или аналогичные данные ВЦИОМ или Левада центра о $Q(\tau)$) для интересующего нас момента времени окажутся недоступны, то число неизвестных параметров модели превысит число точек данных. Таким образом, недостатком модели **A3** является то, что она не в любой момент времени может быть обеспечена данными массовых опросов. Поэтому в случае применения **A3** с целью прогнозирования охватов $\bar{R}_1(\tau)$ некоторые параметры придётся зафиксировать, например, можно зафиксировать $\lambda_1 = 1,293$ и $\lambda_3 = 0,04$.

Расчеты показали, что при помощи модели **A3** данные ФОМ за осень 2005 г. удается аппроксимировать с формально нулевой погрешностью (см. рис.2.3). Максимальное отклонение от данных РОМИР при этом составит всего лишь 2% населения России.

Модель A4.

В нашей работе [72] была исследована также модель **A4**, содержащая четыре сегмента пользователей. Было показано, что если добавить четвертый спектральный пик (с $\lambda < 1/30$), то можно найти такую его амплитуду, при которой отклонения от данных РОМИР равны нулю, а отклонения от данных ФОМ не превышают ошибки округления (рис.2.4). Критическим недостатком модели **A4**, однако, является тот факт, что число свободных параметров в ней (восемь параметров) сравнялось с числом точек данных массовых опросов, полученных в отдельный момент времени. При этом мы не имеем возможности снизить ошибку измерений за счёт осреднения, например, данных мониторинга за последовательные кварталы, поскольку нашей конечной целью является прогнозирование изменения аудитории во времени, и мы можем жертвовать разрешающей способностью. Нет практического смысла и в увеличении числа точек данных за счёт увеличения градаций τ , например, в вопросах респондентам, пользовались ли они Интернетом в последние две недели, три недели, два месяца и т.п., поскольку, на наш взгляд, респонденты не смогут дать точные ответы на такие вопросы. В [72] мы обратили внимание на сходство задачи определения параметров моделей **A1-A4** и задачи высокоразрешающего спектрального анализа, то есть определения амплитуд и фаз двух-трех синусоид по короткому отрезку данных при наличии шума. Такая задача известна чрезвычайно сильной неустойчивостью. Малые погрешности в данных наблюдений могут породить очень сильно различающиеся модели. Нелинейность по параметрам λ_i , значения которых заранее неизвестны и подбираются путём минимизации функционала ошибки (30), может приводить к значительной неустойчивости этих параметров, и эти параметры будут сильно

различаться в близкие моменты времени. Интерпретация моделей с резко изменяющимися во времени параметрами сегментов пользователей, на наш взгляд, настолько сложна, что полезность этих моделей снижается по сравнению с более простой **A3**. Таким образом, увеличение числа параметров моделей с дискретными плотностями $p(\lambda)$ свыше шести мы считаем нецелесообразным.

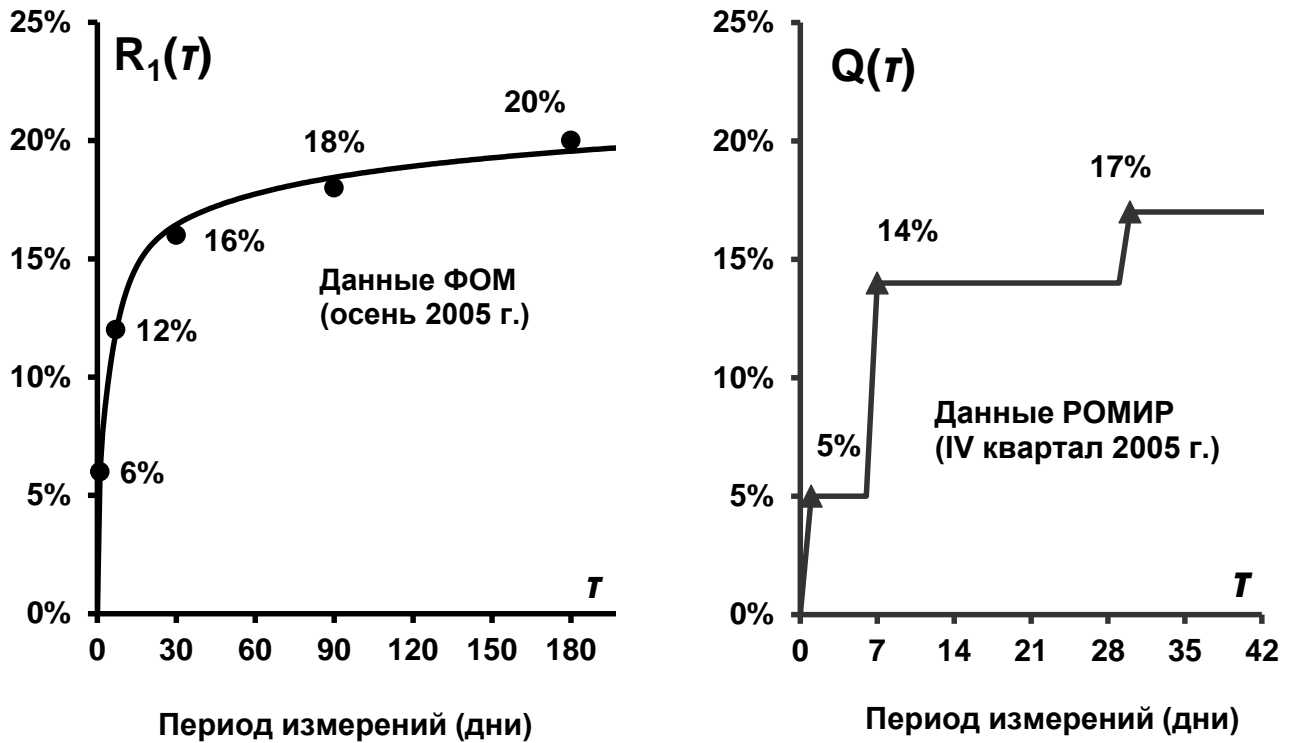


Рис.2.4. Результаты расчетов при помощи модели аудитории **A4**, состоящей из четырех сегментов

Еще более тесная связь существует между задачей аппроксимации $\bar{R}_1(\tau)$ и классической проблемой моментов [13,14] с ограничениями на функцию распределения в виде равенств в нескольких точках.

Модель С.

В работе [72] мы использовали также кусочно-постоянную модель плотности $p(\lambda)$ с разрывами в заранее фиксированных точках λ_i , при $i = 1, \dots, M$. Достоинством модели С является непрерывность полученной (кусочно-линейной) аппроксимации функции распределения $P(\lambda)$. При этом мы не претендовали на определение “оптимальных” частот λ_i , рассматривая модель лишь как способ

аппроксимации неизвестной нам $P(\lambda)$. Свободными параметрами \mathbf{m} модели **C** являются только амплитуды $r_i = R_0 p_i$, но не частоты λ_i . Для охвата аудитории использовалась формула, выведенная в [72]:

$$\bar{R}_1(\tau; \mathbf{m}) = \sum_{i=1}^M \left(\lambda_i - \lambda_{i-1} + \frac{\exp(-\lambda_i \tau) - \exp(-\lambda_{i-1} \tau)}{\tau} \right) \cdot r_i . \quad (35)$$

Недостаток подхода заключается в том, что для достижения приемлемой точности аппроксимации требуется весьма значительное число M свободных параметров, которое при неудачном выборе λ_i может превышать число точек данных. С целью “сглаживания” r_i в работе [72] мы добавили к функции ошибки (31) сумму квадратов первых разностей r_i :

$$\Phi(\mathbf{m}) = \alpha \sum_{k=1}^{K_d} (d_k - \bar{R}_1(\tau_k; \mathbf{m}))^2 + (1 - \alpha) \sum_{l=1}^{L_d} (d_l - Q(\tau_l; \mathbf{m}))^2 + \sum_{i=1}^{M-1} \varepsilon_i \cdot (r_{i+1} - r_i)^2 , \quad (36)$$

где ε_i – неотрицательные сглаживающие множители. Поскольку функция $\bar{R}_1(\tau; \mathbf{m})$ линейна по неизвестным параметрам r_i (см. формулу 35), но на r_i наложены ограничения $r_i \geq 0$, минимизация функции (36) не сводится к решению системы линейных уравнений, а является примером задачи квадратичного программирования. Модель **C** аппроксимирует данные приблизительно с той же точностью, что и модели **A3** и **A4** (см. приведённую ниже таблицу 2.1). В работе [72] нами показано, что полученные аппроксимации неизвестной функции распределения $P(\lambda)$ оказались “похожи” на аппроксимации при помощи моделей **A3** и **A4**.

К сожалению, выбор сглаживающих множителей ε_i существенно зависит от представлений исследователя о том, какого уровня колебания следует подавить в итоговой оценке $p(\lambda)$. Различные исследователи могут выбрать различные ε_i и, следовательно, получить различные “оптимальные” параметры модели. Так в процесс моделирования вносится субъективизм, который сегодня подвергается критике в литературе, посвящённых выбору наилучших моделей. По этой

причине мы далее отказываемся от использования модели **C** в пользу другого непрерывного распределения $P(\lambda)$, описанного в следующем разделе.

Модель **D**. Смесь отрицательных биномиальных распределений

Плотность распределения частоты использования ресурса в модели **D** описывается смесью двух гамма-распределений, соответствующих сегментам активных и малоактивных пользователей

$$p(\lambda; b_1, \alpha_1, b_2, \alpha_2, w) = w b_1 \frac{(b_1 \lambda)^{\alpha_1 - 1}}{\Gamma(\alpha_1)} e^{-b_1 \lambda} + (1 - w) b_2 \frac{(b_2 \lambda)^{\alpha_2 - 1}}{\Gamma(\alpha_2)} e^{-b_2 \lambda}. \quad (37)$$

Величина $r(\lambda) = R_0 \cdot p(\lambda)$ изображена в левой половине рис.2.5.

Соответствующая функция распределения имеет вид взвешенной суммы регуляризованных неполных гамма-функций (см. формулу (18))

$$P(\lambda; b_1, \alpha_1, b_2, \alpha_2, w) = w_1 \frac{1}{\Gamma(\alpha_1)} \int_0^{\lambda/b_1} y^{\alpha_1 - 1} e^{-y} dy + (1 - w_1) \frac{1}{\Gamma(\alpha_2)} \int_0^{\lambda/b_2} y^{\alpha_2 - 1} e^{-y} dy. \quad (38)$$

Из (24) выводим выражение для первого кумулятивного охвата аудитории

$$\bar{R}_1(\tau) = R_0 - R_0 w \left(\frac{b_1}{b_1 + \tau} \right)^{\alpha_1} - R_0 (1 - w) \left(\frac{b_2}{b_2 + \tau} \right)^{\alpha_2}. \quad (39)$$

Обозначив $r_1 = wR_0$, $r_2 = (1 - w)R_0$, получаем формулу для расчетов охвата

$$\bar{R}_1(\tau; \mathbf{m}) = r_1 - r_1 \left(\frac{b_1}{b_1 + \tau} \right)^{\alpha_1} + r_2 - r_2 \left(\frac{b_2}{b_2 + \tau} \right)^{\alpha_2}, \quad (40)$$

где $\mathbf{m} = (b_1, \alpha_1, r_1, b_2, \alpha_2, r_2)$ – совокупность свободных параметров модели.

Таблица 2.1. Сравнение ошибок моделей

Модель	Количество свободных параметров	Количество точек данных	Максимальное отклонение от данных ФОМ*	Максимальное отклонение от данных РОМИР*
A1	2	5	-2,9%	5,0%
A2	4	5	-0,7%	6,1%
A3	6	8	0,0%	2,0%
A4	4	8	0,6%	0,0%
C	20	8	1,2%	1,2%

D	6	8	0,5%	0,2%
---	---	---	------	------

* в процентах взрослого населения РФ

Минимизируя функцию ошибки вида (34) по m , мы получили параметры модели **D**, описывающие данные опросов ФОМ и РОМИР за осень 2005 года.

Оптимальные параметры модели “активного” сегмента:

$$r_1 = 14,4\% \text{ взрослого населения РФ, } b_1 = 1,8 \text{ суток, } \alpha_1 = 1,00.$$

Оптимальные параметры модели “малоактивного” сегмента:

$$r_2 = 7,8\% \text{ взрослого населения РФ, } b_2 = 75 \text{ суток, } \alpha_2 = 1,00.$$

Таким образом, в модели **D** активный сегмент составляет около двух третей всей аудитории и состоит из лиц, посещающих Интернет в среднем раз в 1,8 суток. Однако параметр λ не совпадает у всех представителей этого сегмента, как в моделях **A1-A4**, а распределён в широком диапазоне λ . Объём малоактивного сегмента немного превышает треть аудитории.

На рис.2.5 результаты расчетов по модели **D** сравниваются с данными ФОМ и РОМИР. В обоих случаях ошибка составляет менее процента населения РФ. Таким образом, модель **D** не уступает модели **A3** по точности аппроксимации эмпирических данных о $\bar{R}_1(\tau)$ и $Q(\tau)$. Вторым аргументом в пользу модели **D** является тот факт, что функция $r(\lambda) = R_0 p(\lambda)$ на рис.2.5 внешне оказалась похожа на приведённые в следующих разделах эмпирические данные об аудитории WWW-сайтов, которые получены при помощи онлайн-панелей лог-файлов серверной статистики. Как показывает табл. 2.1, два способа аппроксимации распределения частоты λ – при помощи дискретного распределения (модель **A3**) и при помощи непрерывного распределения (модель **D**), каждый из которых требует шесть свободных параметров, с равной точностью описывают данные опросов ФОМ и РОМИР. Решить вопрос о том, какой способ аппроксимации удачнее, т.е. позволяет приблизить истинное распределение частоты λ , используя меньшее число свободных параметров, можно было бы, если бы можно было измерить не только первую кумулятивную

кривую охвата аудитории Интернета $\bar{R}_1(\tau)$, но и несколько последующих – $\bar{R}_2(\tau)$, $\bar{R}_3(\tau)$ и т.п., которые изображены на рис.2.6.

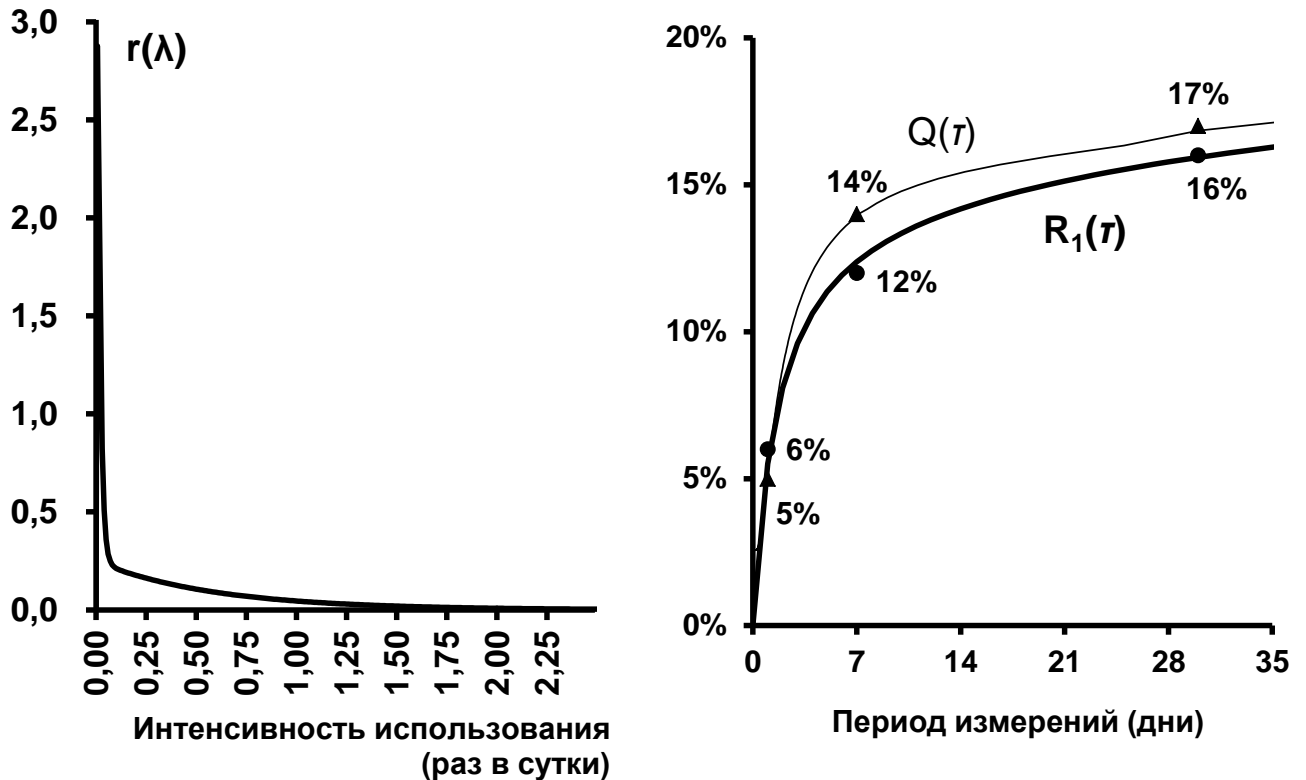


Рис.2.5. Результаты расчетов при помощи модели аудитории **D**, состоящей из двух сегментов с непрерывными распределениями $P(\lambda)$. Использованы данные РОМИР и ФОМ за осень 2005 г.

Рис. 2.6 иллюстрируют весьма заметные различия между кривыми охвата аудитории высших порядков. В кривых охвата для модели **A3**, использующей “дискретное” распределение частоты λ (левая половина рис.2.6), наблюдаются отчетливые “ступеньки” или “полочки”, происхождение которых детально обсуждается в нашей статье [72]. В кривых охвата для модели **D** (правая половина рис.2.6) ступеньки отсутствуют.

Если бы были доступны данные измерений $\bar{R}_k(\tau)$ для больших значений k , то для того, чтобы выяснить, какая модель, **A3** или **D**, лучше аппроксимирует неизвестное нам $P(\lambda)$, достаточно было бы измерить $\bar{R}_k(\tau_1)$ лишь при одном, достаточно большом, значении времени τ_1 . Подобные данные предоставляют так

называемые “онлайн-панели”, рассмотренные в следующем разделе. Однако наше изучение данных онлайн-панелей показало, что такие панели репрезентируют не всех пользователей Интернета, а лишь наиболее активную их часть.

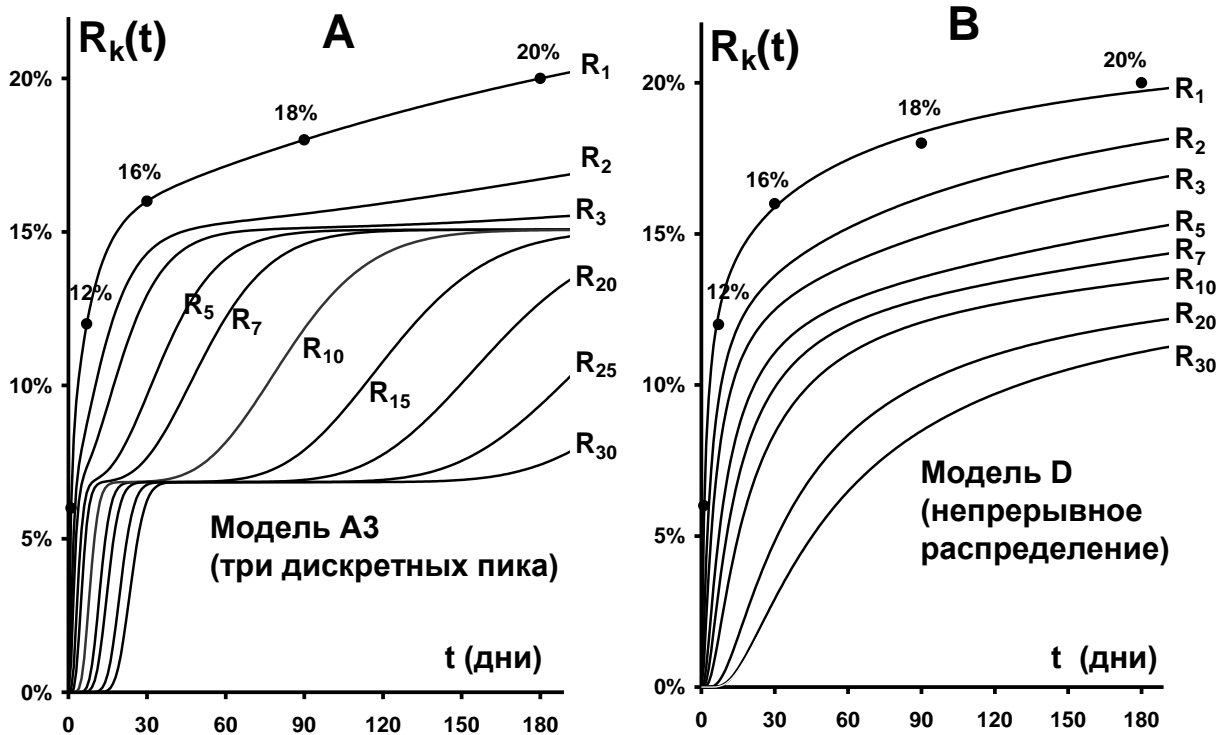


Рис.2.6. Кумулятивные кривые охвата аудитории $\bar{R}_k(\tau)$ для моделей с дискретным (А) и непрерывным (В) распределением λ .

К сожалению, при помощи массовых опросов измерить охваты аудитории Интернета высших порядков $\bar{R}_k(\tau)$ не представляется возможным, поскольку память респондентов не хранит необходимых данных. Действительно, респондентам сложно вспомнить, за какой период они выходили в Интернет два раза, три раза и так далее. Чтобы измерить $\bar{R}_k(\tau)$, можно было бы использовать дневниковые записи специально подготовленных членов панели, однако, такой способ измерений, вероятно был бы слишком дорог. Альтернативой являются данные онлайн-панелей, которые позволяют инструментально измерить аналоги кумулятивных кривых высших порядков. Действующие в России панели TNS, Gemius, КОМКОН-SpyLOG, LiveInternet и др. позволяют получить такие данные.

В следующем разделе будет продемонстрирован непараметрический способ оценки распределений интенсивности использования информационных ресурсов на основе данных онлайн-панелей.

Строго говоря, мы не можем отвергнуть предположение, что распределение частоты использования Интернета является дискретным. Однако, мы предполагаем, что люди ведут себя сходным образом при использовании Интернета и отдельных информационных ресурсов, поэтому будем аппроксимировать $p(\lambda)$ в обоих случаях смесями непрерывных распределений.

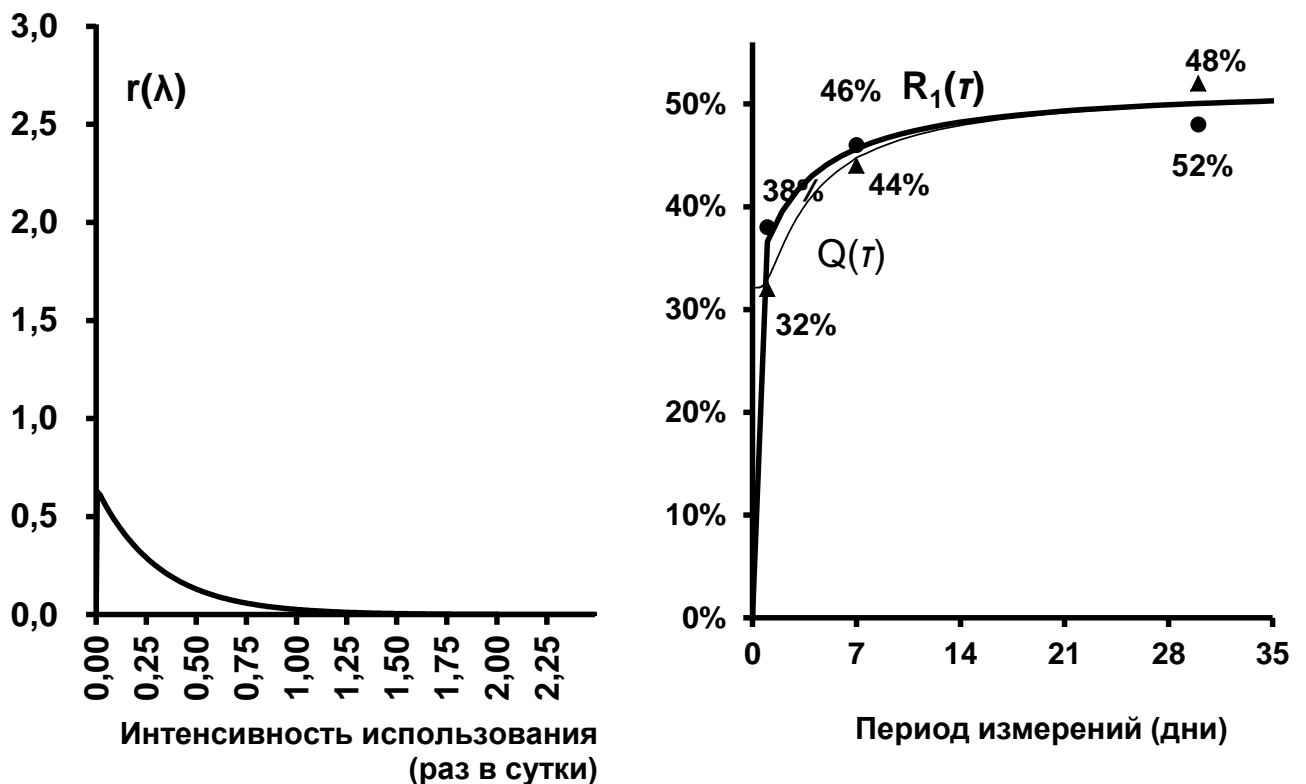


Рис.2.7. Слева – плотность $p(\lambda)$ в модели **D**. Справа – результаты расчетов при помощи модели **D**. Используются данные ФОМ (кружки) и ВЦИОМ (треугольники) за осень 2010 г.

Тестирование модели **D** на данных за 2008 и 2010 гг.

Как отмечалось выше с лета 2007 года публикация данных РОМИР о $Q(\tau)$ прекратилась. На секции IV социологического конгресса в 2012 году президент холдинга РОМИР, А.В.Милёхин, сообщил автору данной работы, что в связи с отсутствием коммерческого спроса прекращён и сбор данных об интенсивности

использования Интернета. Тем не менее в той же форме, что и РОМИР (или в весьма близкой форме), вопросы респондентам продолжают задавать две другие организации – ВЦИОМ и Левада-центр. Мы повторили расчёты по модели **D** для данных за осень 2008 г. и осень 2010 г., используя совместно данные ВЦИОМ о $Q(\tau)$ и ФОМ об охватах $\bar{R}_1(\tau)$. Эти тесты показали удовлетворительное согласование расчётных и эмпирических данных. Рис.2.7 позволяет удостовериться в удовлетворительном совпадении данных опросов за осень 2010 г. с результатами расчётов по модели **D**.

Таким образом, по итогам тестирования за разные моменты времени, отстоящие друг от друга на пять лет, что сравнимо с интересующим нас горизонтом прогнозирования, мы заключаем, что модель **D** способна одновременно описать как данные об охватах $\bar{R}_1(\tau)$, так и данные о $Q(\tau)$.

Таким образом, для прогнозирования данных массовых опросов об охватах $\bar{R}_1(\tau)$ мы выбираем модель **D**, а не **A3**, исходя из принципа совместимости:

- 1) модели с непрерывными распределениями $p(\lambda)$ лучше удовлетворяют данным панельных исследований крупных информационных ресурсов;
- 2) Гамма-распределение параметра простейшего потока, характеризующего индивидов, широко используется в моделях поведения потребителя и в медиаметрии уже более пятидесяти лет, начиная с пионерской работы Э.Эренберга [359,332].

2.1.2 Прогнозирование динамики охватов аудитории сети Интернет на основе разработанной количественной модели

Квазистационарная модель динамики $\bar{R}_1(\tau; m(t))$

Предыдущая часть этой главы была посвящена построению модели охвата аудитории $\bar{R}_1(\tau)$ в некоторый отдельно взятый момент времени t . В этом разделе мы построим модель охватов аудитории для последовательности моментов времени. Мы используем данные опросов ФОМ в Москве за каждый квартал с

2002 по 2010 г. Здесь разумно использовать для охватов обозначение $\bar{R}_1(\tau; \mathbf{m}(t))$, где τ обозначает интервал времени (сутки, неделю, месяц, квартал и полгода), а t – момент проведения опроса.

Аппроксимация исторических данных

Поскольку в описанной выше модели **D**, которую мы хотели бы использовать, шесть свободных параметров (см. формулу 40), а опросы ФОМ предоставляют лишь пять точек, мы зафиксировали значения трёх параметров: $b_1 = 0,08$ суток (12 раз в сутки), $\alpha_1 = 1,00$, $\alpha_2 = 1,00$. Таким образом, мы используем редуцированную версию модели **D**, которую будем обозначать **E**.

Свободными в модели **E** остаются лишь параметры: $r_1(t)$, $r_2(t)$ и b_2 . Первоначально мы позволяли параметру b_2 изменяться во времени используя упрощённую версию модели динамики установки индивида (предрасположенности к определённому виду поведения), развиваемую школой Ю.А.Гаврильца [44]. Однако итоговый прогноз, построенный на основе такого подхода, оказался менее точным (на трёхлетнем горизонте), чем прогноз, построенный в первой главе данной работы на основании эвристической многостадийной модели (см. рис.1.25). Мы предположили, что причина ухудшения точности прогноза состоит в чрезмерном количестве неизвестных величин, требующих прогнозирования. Поэтому в модели **E** параметр b_2 хотя и является заранее неизвестным, но должен быть один и тот же для всех моментов времени.

Плотность распределения λ в модели **E** является смесью двух экспонент

$$p(\lambda; b_1, b_2, w) = w(t)b_1e^{-b_1\lambda} + (1-w(t))b_2e^{-b_2\lambda}. \quad (46)$$

Обозначив $r_1(t) = w(t)R_0(t)$, $r_2(t) = (1-w(t))R_0(t)$, для каждого момента измерений t перепишем формулу (40) для расчета охвата аудитории в виде

$$\bar{R}_1(\tau; \mathbf{m}(t)) = \frac{\tau}{b_1 + \tau} r_1(t) + \frac{\tau}{b_2 + \tau} r_2(t), \quad (47)$$

где $\mathbf{m}(t) = (b_2, r_1(t), r_2(t))$ – совокупность свободных параметров модели, которые различаются для различных моментов времени t .

Положив, как указано выше, $b_1=0,08$ суток, мы фактически постулируем, что активная аудитория использует Сеть так часто, что её кривая кумулятивного охвата вырождается в ступень амплитудой $r_1(t)$, и формулу охвата можно приближённо записать в виде

$$\bar{R}_1(\tau; \mathbf{m}(t)) \approx r_1(t) + r_2(t) \left(1 - \frac{b_2}{b_2 + \tau} \right). \quad (48)$$

Этот приём сближает нашу модель с моделью Ж.-П.Шандона, описанной в следующем разделе данной главы, и имеет ту же цель: повысить устойчивость оценки параметров модели при малом объёме данных.

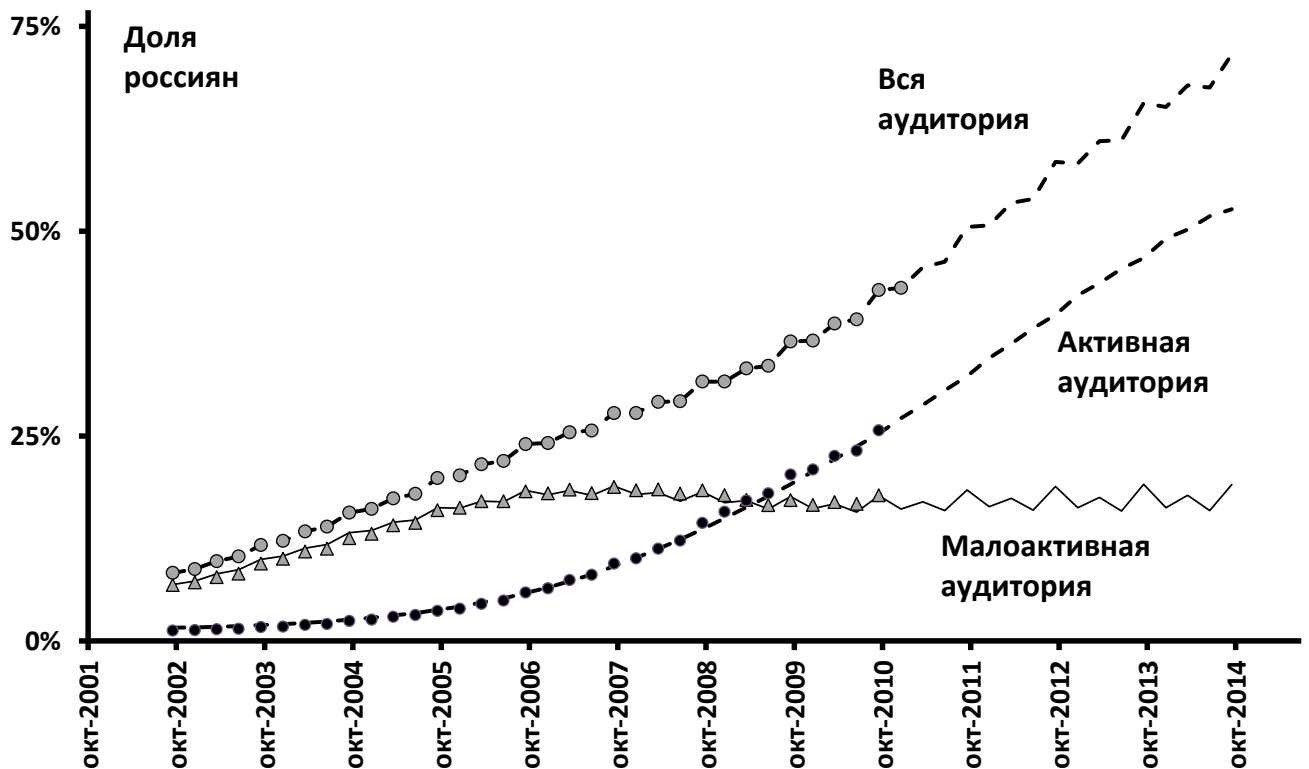


Рис.2.8. Динамика активной и малоактивной аудитории в России в 2001-2010 гг. и прогноз на основе модели **E**

Подход, который мы при этом используем, можно назвать *квазистационарным*, поскольку при выводе формул для охватов предполагалось, что параметры r_1 и r_2 не зависят от времени, а при моделировании данных мы позволяем этим параметрам принимать различные значения в различные моменты времени t . Нестационарные модели в медиаметрии чрезвычайно редки, отметим

здесь работу [453]. Мы считаем описанный подход допустимым, если общий объём аудитории изменяется лишь на несколько процентов на протяжении полугода, т.е. периода измерений кумулятивной кривой.

Кроме того, мы форсируем “плавность” колебаний параметров модели **EV** во времени, добавив к функционалу ошибки слагаемое, зависящее от амплитуды различий в параметрах, подобно тому, как это было сделано в модели **C**

$$\Phi(\tilde{\mathbf{m}}) = \sum_{i=1}^I \sum_{k=1}^{K_d} (d_{k,i} - \bar{R}_1(\tau_k; \mathbf{m}(t_i)))^2 + \sum_{i=1}^{I-1} \varepsilon_{i2} \cdot (r_1(t_{i+1}) - r_1(t_i))^2 + \sum_{i=1}^{I-1} \varepsilon_{i3} \cdot (r_2(t_{i+1}) - r_2(t_i))^2, \quad (49)$$

где вектор всех переменных параметров (включающих все моменты времени) обозначен $\tilde{\mathbf{m}} = (\mathbf{m}(t_1), \mathbf{m}(t_2), \dots, \mathbf{m}(t_I))$.

На рис.2.8 оценки объёма активной части аудитории $r_1(t)$ российского Интернета в 2002-2010 гг. показаны сплошными кружками, а малоактивной $r_2(t)$ – открытыми треугольниками. Сплошными линиями на рис.2.8 показаны результаты прогнозирования $r_1(t)$ и $r_2(t)$, и полной аудитории $x(t) = r_1(t) + r_2(t)$; методика соответствующей процедуры прогнозирования описана ниже.

Моделирование данных опросов ФОМ в России за 2002-2010 гг. при помощи модели **E** показывает (рис.2.8), что в 2002 г. малоактивная аудитория существенно превосходила активную, однако к середине 2009 года оба сегмента сравнялись по величине, после чего активная аудитория росла, а малоактивная – убывала. Таким образом, существует переток аудитории из малоактивного сегмента в активный. На рис.2.8 объём активной аудитории следует за объёмом полной аудитории, подобно тому, как на рис.2.3 количество абонентов Интернета следовало за количеством пользователей. Сделанное наблюдение позволяет предположить, что для моделирования динамики активной аудитории достаточно весьма простой многостадийной модели, которая может содержать существенно меньше параметров, чем модель, использованная в последнем разделе первой главы.

Прогнозирование охватов аудитории

Модель **Е** позволяет оценить объёмы активной и малоактивной аудитории в прошлом, но не содержит средств прогнозирования ни этих величин, ни охватов \bar{R}_1 . Для решения задачи прогнозирования охватов аудитории, построим модель **Г**, которая включает в себя модель **Е** и многостадийную моделью использования Интернета, рассмотренную в последнем разделе первой главы. В этой многостадийной модели оставим лишь два класса (две крайние слева вертикальные последовательности на рис.1.24), соответствующие малоактивным и активным пользователям.

Система уравнений баланса из первого раздела данной главы данной работы в данном случае будет содержать лишь три уравнения:

$$\begin{aligned}\frac{\partial x^{1,1}}{\partial t} &= -g^1(t, \pi)x^{1,1} - \mu(t, \pi)x^{1,1} + f^{1,1}(t, \pi) \\ \frac{\partial x^{1,2}}{\partial t} &= -g^1(t, \pi)x^{1,2} - \mu(t, \pi)x^{1,2} + f^{1,2}(t, \pi) \\ \frac{\partial x^{2,2}}{\partial t} &= g^1(t, \pi)x^{1,2} - \mu(t, \pi)x^{2,2} + f^{2,2}(t, \pi).\end{aligned}\tag{50}$$

Добавим несколько упрощающих предположений. Во-первых, в формуле для интенсивностей переходов $f^{i,j} = \rho^{i,j}(t, \pi)h(t, \pi)y(t, \pi)$, положим все $\rho^{i,j}(t, \pi)$ постоянными величинами, тогда

$$f^{i,j} = \rho^{i,j} \cdot h(t, \pi)y(t, \pi).\tag{51}$$

Пренебрегая воспроизводством населения, обозначим $x(t)$ – долю пользователей в населении рассматриваемого региона, а $v(t) = dx/dt$ – скорость притока новых пользователей. Затем, используя функции риска в форме $g(t, \pi) = \tilde{p} + \tilde{q}x^{2,2}(t)$, запишем уравнения для сегментов в виде

$$\begin{aligned}\frac{\partial x^{1,1}}{\partial t} &= \rho^{1,1}v(t) \\ \frac{\partial x^{1,2}}{\partial t} &= -(\tilde{p} + \tilde{q}x^{2,2}(t))x^{1,2} + \rho^{1,2}v(t) \\ \frac{\partial x^{2,2}}{\partial t} &= (\tilde{p} + \tilde{q}x^{2,2}(t))x^{1,2} + \rho^{2,2}v(t).\end{aligned}\tag{52}$$

Поскольку новая аудитория должна целиком распределяться по сегментам, неотрицательные величины $\rho^{i,j}$ связаны уравнением

$$\rho^{1,1} + \rho^{1,2} + \rho^{2,2} = 1. \quad (53)$$

Чтобы это условие выполнялось автоматически, введём вспомогательные параметры θ и φ , так чтобы выполнялись равенства

$$\rho^{1,1} = \sin^2 \theta, \quad \rho^{1,2} = \cos^2 \theta \sin^2 \varphi, \quad \rho^{2,2} = \cos^2 \theta \cos^2 \varphi. \quad (54)$$

Таким образом, в модели **F** лишь восемь свободных параметров:

$$\mathbf{m} = (\theta, \varphi, \tilde{p}, \tilde{q}, b_2, x^{1,1}(0), x^{1,2}(0), x^{2,2}(0)). \quad (55)$$

Легко видеть, что малоактивную аудиторию составляют два подмножества состояний, возможных в многостадийной модели ($x^{1,2}(t)$ и $x^{1,1}(t)$), а активную – только один ($x^{2,2}(t)$)

$$r_1(t) = x^{1,1}(t) + x^{1,2}(t) \quad (56)$$

$$r_2(t) = x^{2,2}(t).$$

Совокупность свободных параметров (55), скорость притока новых пользователей $v(t)$ (как внешняя переменная) и уравнения (47), (52), (54), (56) составляют математическое описание модели **F**. Несложно вывести уравнения непосредственно для объёмов активного и малоактивного сегмента. Достаточно заметить, что из первого уравнения системы (52) немедленно вытекает

$$x^{1,1}(t) = x^{1,1}(0) - \rho^{1,1}x(0) + \rho^{1,1}x(t). \quad (57)$$

Обозначив $c = x^{1,1}(0) - \rho^{1,1}x(0)$, из (52) и (56) получаем два уравнения

$$\frac{dr_2}{dt} = (\tilde{p} + \tilde{q}r_2)((1 - \rho^{1,1})x(t) + c - r_2) + \rho^{2,2}v(t), \quad (58)$$

$$\frac{dr_1}{dt} = -(\tilde{p} + \tilde{q}r_2)((1 - \rho^{1,1})x(t) + c - r_2) + (\rho^{1,1} + \rho^{1,2})v(t). \quad (59)$$

Поскольку $r_2(t) = x(t) - r_1(t)$, и $x(t)$ – известна, достаточно решить уравнение (58). Оно отличается от привычных уравнений многостадийных моделей, рассмотренных в первой главе, только наличием слагаемого $\rho^{2,2}v(t)$. Это слагаемое появляется, поскольку в модели **F** допускаются не только переходы

между соседними состояниями, но и “перескок” нового пользователя через “малоактивное” состояние напрямую в “активный” сегмент.

Оптимальные значения параметров модели были найдены при помощи минимизации функции ошибки вида

$$\Phi(\tilde{\mathbf{m}}) = \sum_{i=1}^I \sum_{k=1}^{K_d} (d_{k,i} - \bar{R}_1(\tau_k; \mathbf{m}(t_i)))^2. \quad (60)$$

Оптимальные значения параметров модели \mathbf{F} оказались равны:

$$\mathbf{m}^* = (\theta = 0,38, \varphi = 1,26, \tilde{p} = 0,0, \tilde{q} = 4,8(\text{год}^{-1}), b_2 = 5,7(\text{дней}), \\ x^{1,1}(0) = 6,9\%, x^{1,2}(0) = 0,0\%, x^{2,2}(0) = 1,6\%).$$

Рассмотрение значений параметров

$$\rho^{1,1} = 0,14, \quad \rho^{1,2} = 0,78, \quad \rho^{2,2} = 0,08,$$

подсказывает, что новые пользователи первоначально пополняют, в основном, малоактивный сегмент ($\rho^{1,2} = 0,78$), однако, благодаря значительному влиянию активных пользователей ($\tilde{q} = 4,8(\text{год}^{-1})$), быстро переходят в активный сегмент. Таким образом, сделанное в первой главе предположение, что все новые пользователи поступают в малоактивный сегмент, не является грубо ошибочным.

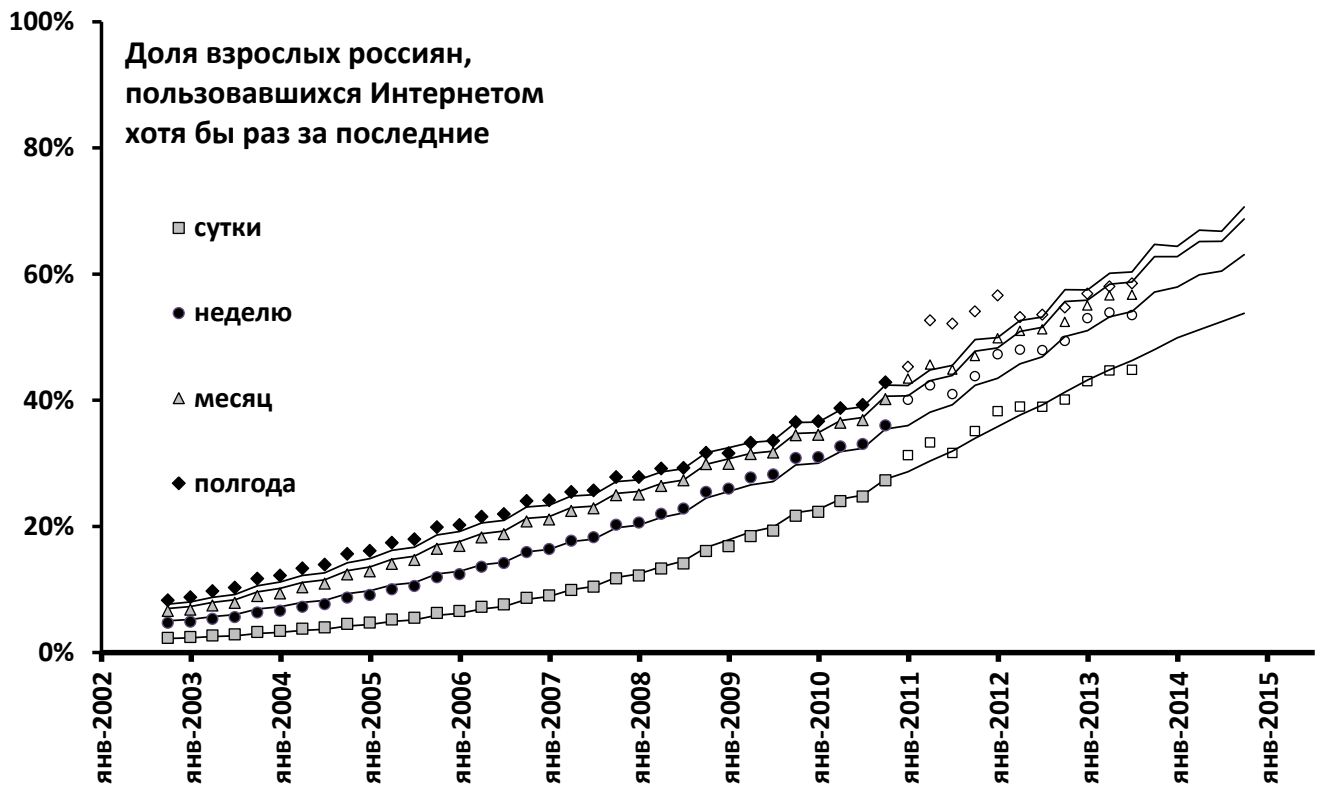


Рис.2.9. Прогноз полугодового, месячного, недельного и суточного охватов аудитории Интернета в России при помощи модели **F**.

Прогнозируемые значения $r_1(t)$ показаны тонкой линией на рис.2.8, $r_2(t)$ – жирной пунктирной линией. Найденные ранее $r_1(t_i)$ и $r_2(t_i)$ изображены открытыми треугольниками и закрашенными кружками соответственно.

Итогом применения модели **F** является прогноз охватов \bar{R}_1 суточной, недельной, месячной и полугодовой аудитории Интернета в России, представленный на рис.2.9. Сплошными символами представлены данные опросов ФОМ, которые были доступны автору на момент построения прогноза, а открытыми символами – новые данные.

Разработанная здесь модель **F** обладает двумя важными преимуществами по сравнению с моделью, описанной в последнем разделе первой главы. Во-первых, эта модель теоретически обоснована, поскольку описывает поведение пользователя при помощи потока событий, в то время как старая модель фактически подавала охваты аудитории на вход “чёрного ящика”. Во-вторых, в модели **F** пользователей достаточно разделить лишь на два сегмента, в то время как в прежней модели их было четыре. Соответственно, в модели **F** оказываются на один свободный параметр меньше, что делает её потенциально устойчивее. При этом, как показывает сравнение рис.1.25 и рис.2.9, прогнозы, построенные при помощи двух моделей, оказались чрезвычайно близки. Недостатком модели **F** (как и всех моделей данной главы) с точки зрения скорости её разработки и практического применения в условиях дефицита времени можно считать необходимость привлечения теории потоков событий, т.е. своего рода теоретический “барьер”, который необходимо преодолеть исследователю.

Пользователю модели, желающему избежать знакомства с теорией потоков событий, можно предложить следующий подход к прогнозированию охватов аудитории. Предполагается, что аудитория состоит из двух сегментов. Охваты аудитории этих сегментов пропорциональны неубывающим функциям $\hat{R}_{1,1}(\tau)$ и

$\hat{R}_{1,2}(\tau)$ ¹⁷. Объёмы сегментов, $r_1(t)$ и $r_2(t)$ меняются во времени и могут быть описаны при помощи описанной нами многостадийной модели. Такой подход может оказаться работоспособен и в том случае, когда поведение пользователя описывает не простейший поток, а поток с последствием, например, поток Эрланга второго порядка.

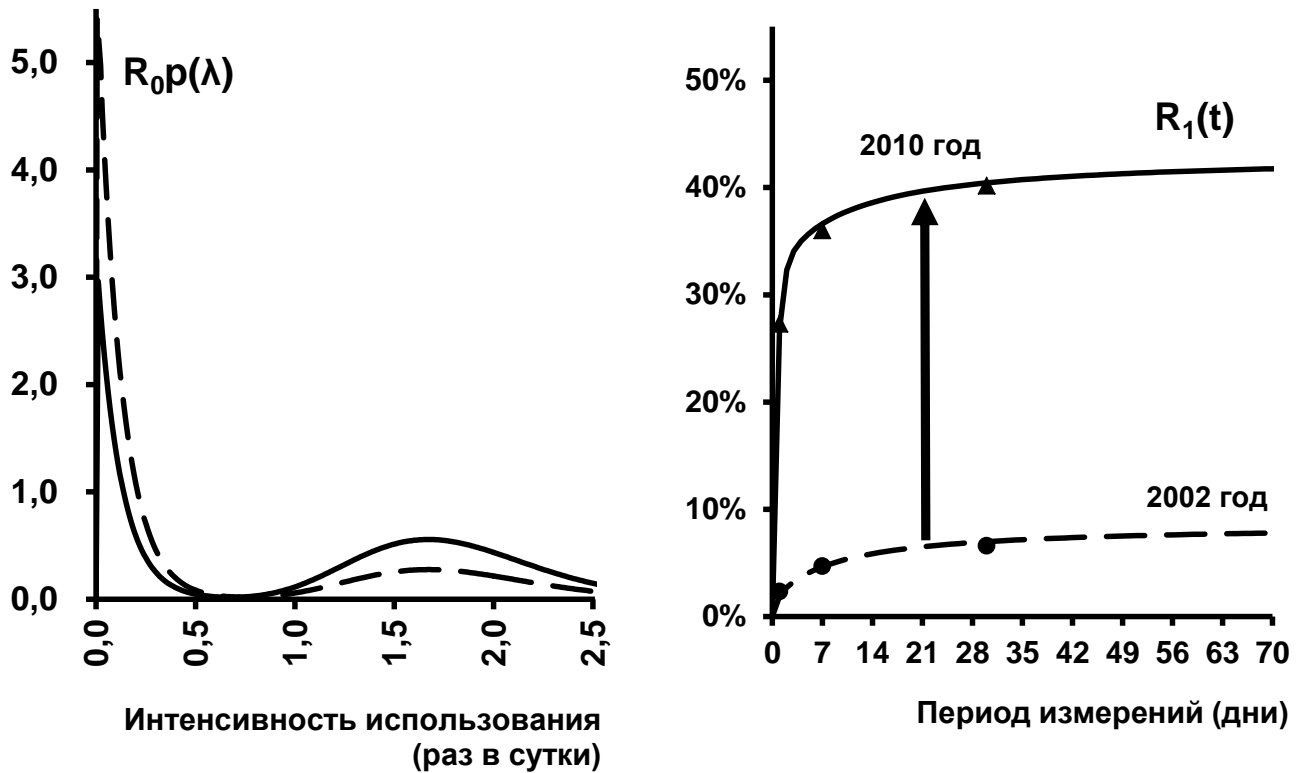


Рис.2.10. Слева – плотность $p(\lambda)$ в модели **Н**. Справа – результаты расчетов охватов аудитории при помощи модели аудитории **Н**. Используются данные ФОМ (кружки) за осень 2002 г. и осень 2010 г.

С нашей точки зрения, в условиях, когда наилучший способ аппроксимации $p(\lambda)$ неизвестен, для повышения надёжности прогнозирования можно использовать несколько аппроксимаций $p(\lambda)$, а затем осреднить полученные с их помощью прогнозы.

¹⁷ В частности, можно выбрать $\hat{R}_{1,i} = 1 - \exp(-\lambda_i \tau)$, и, таким образом, взять за базовую модель поведения пользователя модель **A2**. Как было показано выше, эта модель удовлетворительно описывает данные опросов ФОМ.

В качестве примера испытаем аппроксимацию $p(\lambda)$ при помощи той же смеси гамма-распределений (37), но зафиксируем и оставим свободными иные параметры, что в модели **F**. Положим в (40) $b_1 = b_2 = b = 8,5(\text{дней})$ и оставим свободными параметрами α_1 и α_2 . Формула для расчёта охватов приобретёт вид

$$\bar{R}_1(\tau; \mathbf{m}) = r_1 - r_1 \left(\frac{b}{b + \tau} \right)^{\alpha_1} + r_2 - r_2 \left(\frac{b}{b + \tau} \right)^{\alpha_2}. \quad (61)$$

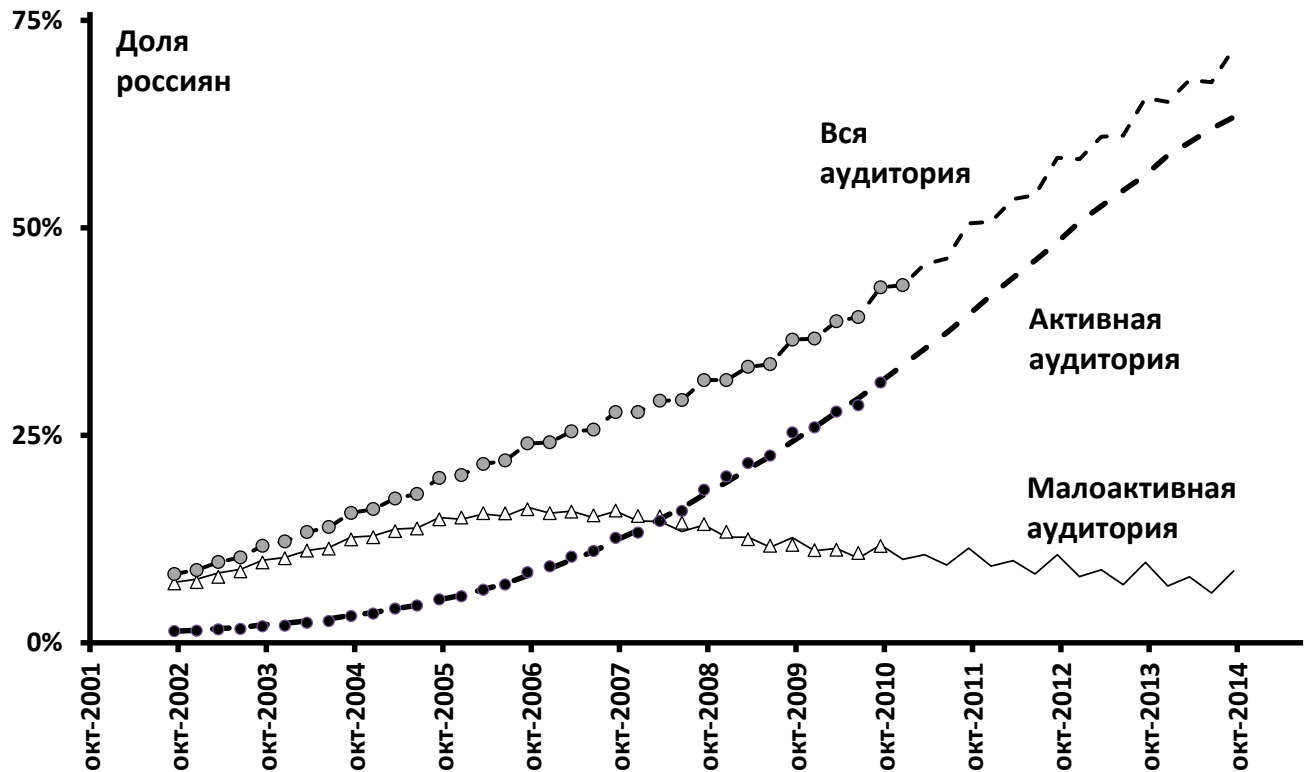


Рис.2.11. Динамика активной и малоактивной аудитории в России в 2001-2010 гг. и прогноз на основе модели **H**

В полученной модели **H** девять свободных параметров, т.е. на один больше, чем в модели **F**. Оптимальные значения параметров модели **H** оказались равны

$$\mathbf{m}^* = (\theta = 0,0, \varphi = 1,16, \tilde{p} = 0,0, \tilde{q} = 3,4(\text{год}^{-1}), \alpha_1 = 1,0, \alpha_2 = 15,2, \\ x^{1,1}(0) = 4,7\%, x^{1,2}(0) = 2,6\%, x^{2,2}(0) = 1,4\%).$$

Аппроксимация плотности $p(\lambda)$ показана на рис.2.10, здесь активный и малоактивный сегменты, в отличие от модели **F** почти не перекрываются (в сравнении со значительным их перекрытием на рис.2.5 и 2.7).

Новая аудитория распределяется в соответствии со значениями коэффициентов

$$\rho^{1,1} = 0,0, \quad \rho^{1,2} = 0,84, \quad \rho^{2,2} = 0,16.$$

Таким образом, в модели **Н** величина $x^{1,1}(t)$ не меняется во времени, составляя 4,7%. Объём малоактивного сегмента может составлять существенную величину, поскольку параметр влияния активных пользователей \tilde{q} – ниже, чем в модели **Н**, однако весьма быстро снижается до 4,7% населения уже к 2015 году (рис.2.11). Сравнение рис.2.11 с рис.2.8 показывает, что динамика объёма малоактивного сегмента $r_1(t)$ убывает значительно быстрее, чем в модели **Г**.

Несмотря на различия в оценках $r_1(t)$ и $r_2(t)$, результаты прогнозирования охватов \bar{R}_1 при помощи модели **Н**, представленные на рис.2.12, практически совпадают с полученными при помощи модели **Г**.

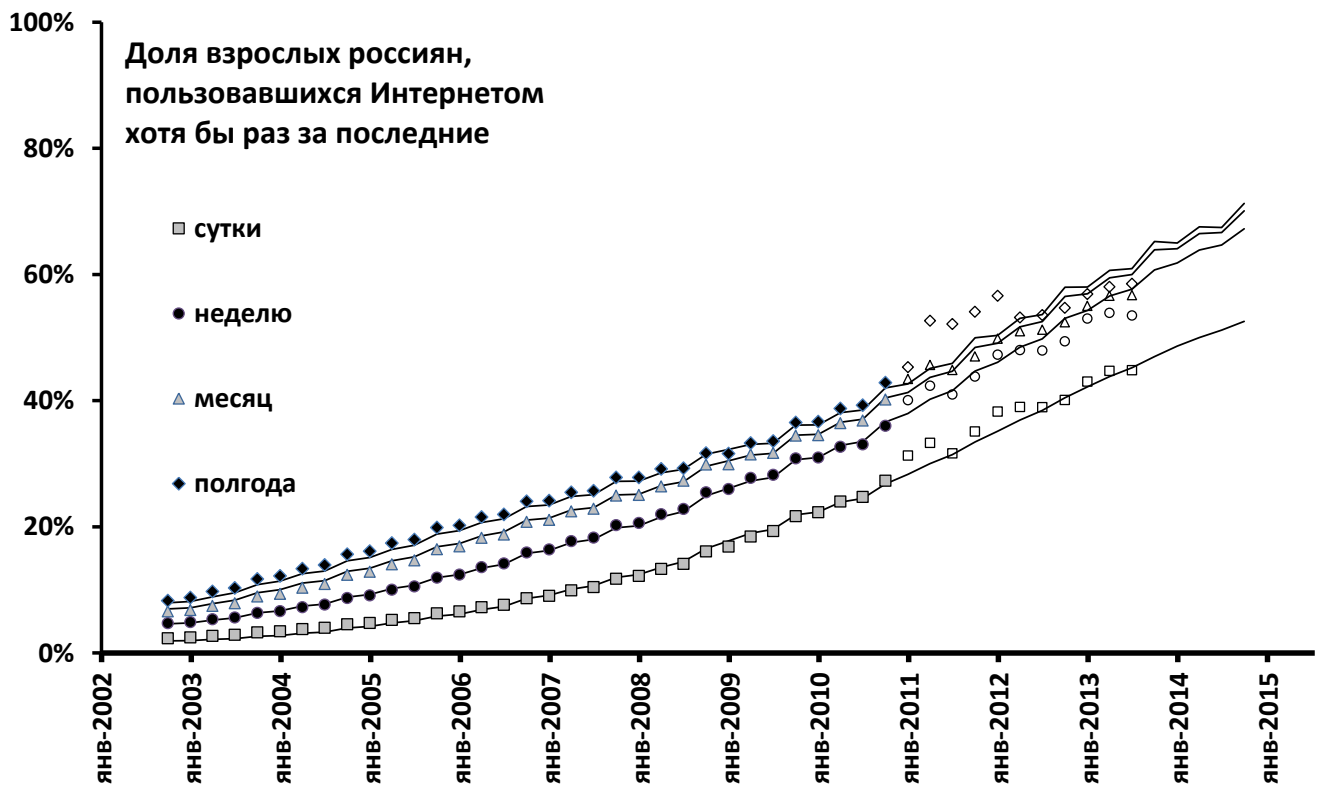


Рис.2.12. Прогноз полугодового, месячного, недельного и суточного охватов аудитории Интернета в России при помощи модели **Н**.

Мы заключаем, что вне зависимости от конкретной аппроксимации $p(\lambda)$ при использовании простой многостадийной модели динамики активного и

малоактивного сегмента пользователей, а также при небольшом (8-9) совокупном числе свободных параметров, разработанные в этом разделе модели пригодны для прогнозирования охватов аудитории Интернета с горизонтом пять лет.

2.2 Моделирование онлайн-панели при помощи смеси бета-биномиальных распределений

Бета-биномиальная модель широко используется в медиапланировании, в частности, для аппроксимации кумулятивного охвата аудитории газет и других медиа. Мы предприняли попытку применить эту модель к данным панельного исследования аудитории крупных Интернет-ресурсов [86,137,142]. К сожалению, классическая бета-биномиальная модель неудовлетворительно описывает данные онлайн-панели. Основных причин – две: различие в поведении “офисной” и “домашней” аудитории, и значительная доля малоактивных пользователей у большинства Интернет-ресурсов.

Количество дней, в течение которых пользователь использовал ресурс

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
1	5327																															
2	3991	2772																														
3	3794	1924	2125																													
4	3464	1886	1437	1783																												
5	3124	2446	1374	1257	1579																											
6	2496	2303	1812	1156	1088	1438																										
7	2427	1697	1808	1469	1041	982	1286																									
8	2314	1704	1627	1476	955	936	836	1149																								
9	2277	1597	1582	1339	996	876	763	786	1026																							
10	2137	1445	1302	1327	1204	954	743	733	725	956																						
11	2148	1439	1196	1051	1150	1113	848	714	672	672	873																					
12	2022	1408	1190	995	864	1121	970	793	665	657	593	811																				
13	1933	1434	1150	914	802	866	992	928	712	632	580	572	767																			
14	1863	1450	1070	960	698	777	787	919	841	723	569	564	497	721																		
15	1856	1414	1014	961	754	705	756	855	875	569	620	569	479	467	663																	
16	1859	1320	1078	879	752	737	733	838	826	602	500	593	485	457	427	621																
17	1784	1309	1036	863	682	723	691	783	717	790	548	498	493	478	426	397	606															
18	1724	1311	985	884	719	681	577	756	622	765	660	532	430	478	476	401	386	575														
19	1683	1333	975	887	669	676	603	575	678	651	643	627	508	394	459	455	383	375	542													
20	1635	1343	937	906	634	693	598	509	581	680	515	614	623	490	368	461	358	392	380	506												
21	1618	1299	912	874	667	656	633	483	522	569	624	466	590	596	475	339	403	386	370	372	462											
22	1606	1276	929	866	653	643	646	481	487	546	583	553	531	576	437	362	342	416	338	360	358	419										
23	1595	1264	881	863	653	649	630	499	465	525	586	537	552	543	416	340	366	346	399	298	364	327	391									
24	1606	1234	858	874	632	650	595	515	436	480	550	498	515	583	503	389	296	368	334	387	295	341	309	368								
25	1577	1239	833	862	604	615	639	503	436	426	541	441	517	517	527	450	374	306	350	311	386	275	322	303	355							
26	1557	1250	790	818	672	581	587	575	402	432	450	450	457	503	477	460	475	341	320	323	351	318	272	315	285	345						
27	1571	1225	796	800	670	579	557	566	468	357	449	375	500	392	507	417	471	465	323	298	342	317	300	244	314	265	343					
28	1579	1204	819	762	670	541	592	516	443	426	429	336	446	403	478	400	446	427	469	272	332	296	323	274	257	278	267	321				
29	1552	1213	826	728	701	548	550	517	460	422	405	351	422	410	434	400	485	411	400	387	250	278	302	310	265	220	289	235	309			
30	1571	1177	838	661	724	571	526	517	476	433	389	324	398	443	359	453	462	444	352	345	342	219	268	313	298	235	215	293	216	290		

Рис.2.13. Данные онлайн-панели об использовании почтового сервиса

Мы внесли два усовершенствования в бета-биномиальную модель. Первое является очевидным – мы вводим два сегмента пользователей, представители первого пользуется ресурсом относительно часто, а второго – редко. Второе усовершенствование заключается в отдельном моделировании поведения офисной и домашней аудитории. Использование информационных ресурсов домашней аудиторией, по-прежнему, описывает бета-биномиальная модель, но для описания поведения офисной аудитории мы исключаем выходные дни, т.е. используем суперпозицию нестационарных потоков.

Количество дней, в течение которых пользователь использовал ресурс

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Период наблюдений	1	291	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	386	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	450	34	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	557	43	18	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	705	93	23	11	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	861	75	62	9	11	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	963	118	32	52	6	11	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	1039	135	24	55	13	6	11	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	1103	146	39	57	7	11	12	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	1191	184	37	44	21	11	12	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	1334	241	52	22	27	28	8	8	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	1457	267	73	17	8	33	21	8	8	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	1644	282	75	34	3	27	15	24	8	4	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	1746	322	79	32	16	12	19	12	24	10	2	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	1802	362	81	33	18	4	15	24	19	8	8	2	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	1840	391	91	41	15	7	13	18	27	4	8	6	0	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	17	1895	434	100	45	15	14	14	0	21	28	0	8	6	0	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	1986	437	110	50	22	15	5	12	6	16	28	0	8	6	0	8	3	4	0	0	0	0	0	0	0	0	0	0	0	0
	19	1995	479	124	53	15	19	13	11	7	0	27	17	0	11	2	0	8	3	4	0	0	0	0	0	0	0	0	0	0	0
	20	2033	491	138	46	33	14	16	7	13	1	0	29	14	0	11	2	0	8	3	4	0	0	0	0	0	0	0	0	0	0
	21	2120	501	150	45	28	22	10	12	14	3	1	0	29	14	8	3	2	0	8	3	4	0	0	0	0	0	0	0	0	0
	22	2178	525	157	37	38	12	20	10	9	10	1	0	21	22	4	6	2	2	0	8	3	4	0	0	0	0	0	0	0	0
	23	2174	570	167	46	33	13	22	10	11	10	1	0	21	17	9	4	2	3	0	0	8	3	4	0	0	0	0	0	0	0
	24	2224	609	174	52	28	15	14	21	0	14	7	1	0	29	10	13	0	4	2	0	0	8	3	4	0	0	0	0	0	0
	25	2312	648	173	60	29	15	10	23	2	6	15	0	1	3	25	15	8	0	4	2	0	0	8	3	4	0	0	0	0	0
	26	2356	661	187	58	36	22	8	10	17	3	13	5	0	4	0	33	7	8	2	4	0	0	0	8	3	4	0	0	0	0
	27	2371	678	190	61	45	26	8	7	14	6	12	3	5	3	1	3	31	7	8	4	2	0	0	0	8	3	4	0	0	0
	28	2405	687	229	61	50	26	8	6	14	4	12	6	4	5	0	1	8	28	8	4	4	2	0	0	0	8	3	4	0	0
	29	2419	701	256	64	51	29	8	6	13	6	2	16	4	3	2	1	8	21	11	4	8	0	2	0	0	3	5	3	4	0
	30	2481	717	268	70	42	36	10	6	6	10	6	13	6	3	0	3	3	26	11	0	10	2	2	0	0	3	5	3	4	0

Рис.2.14. Данные онлайн-панели об использовании онлайн-СМИ

Начиная с 2006 года, несколько исследовательских компаний, в первую очередь – TNS Россия и Gemius, внедряют онлайн-панели в качестве инструмента исследования поведения аудитории крупных проектов российского Интернета.

Вне зависимости от того, какой способ используется – “user-centric” или “site-centric”, панель позволяет понять, как люди используют онлайн-ресурс. Панельные измерения сообщают, сколько людей, такого-то возраста и пола, пользовались сайтом за выбранный промежуток времени. При этом в практике медиаизмерений сложился способ, позволяющий избежать уточнения понятия “использования” и связанных с этим вопросов о моментах начала и окончания сессии, в течение которой пользователь использует ресурс. Когда определить такие моменты сложно, вместо этого фиксируют факт наличия или отсутствия использования ресурса в определённый день (или иной промежуток времени), подобно отметке в журнале учёта посещаемости занятий слушателями учебного заведения.

Данные панельных измерений, использованные в этой главе, приведены на рис.2.13 для популярного почтового сервиса и на рис.14 для крупного онлайн-СМИ.

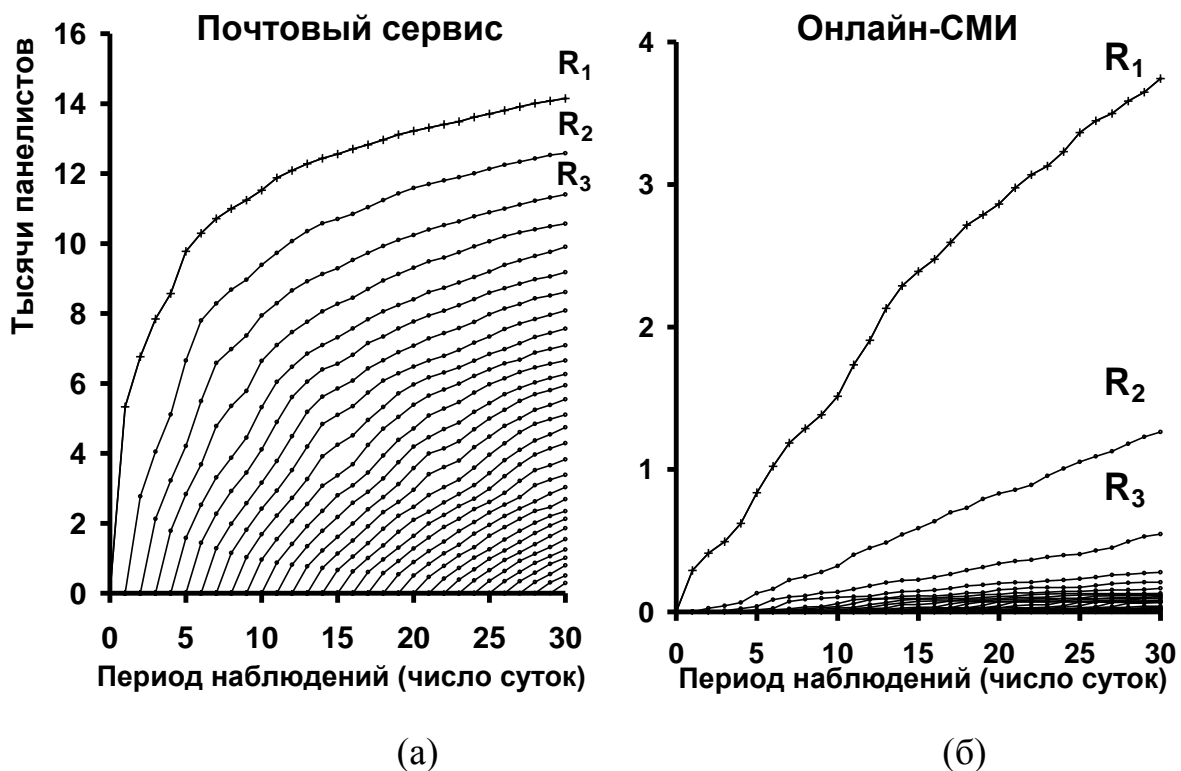


Рис.2.15. Кумулятивные охваты аудитории двух ресурсов по данным онлайн-панели: а) портал “большой тройки”, б) сайт ежедневной газеты.

На рис.2.15 приведены кумулятивные кривые охвата за период от 1 до 30 дней для двух крупных ресурсов Рунета – почтового сервиса и СМИ. По вертикали показано число членов панели, которые использовали ресурс за заданный период времени. Подписаны три первые кумулятивные кривые. В медиапланировании кумулятивные кривые используются с 50-х гг. XX века – в первую очередь для расчета охвата и частоты использования ресурсов при медиапланировании [424,160]. Рис.2.15 демонстрирует важные различия в поведении аудитории почтового портала и крупного онлайн-СМИ. Действительно, хотя аудитория СМИ за месяц составляет 26% аудитории почтового сервиса, по количеству тех, кто побывал на ресурсе два дня за месяц СМИ уступает уже в десять раз. По количеству побывавших три раза – в двадцать.

В наших работах [67,103] описана модель поведения пользователя, с учетом смены cookies. Она подробно описана в следующем разделе данной главы. Эта модель предназначена для интерпретации данных онлайн-счетчиков и довольно сложна. Но при использовании онлайн-панели эффект обновления cookies отсутствует, что значительно упрощает модель.

В марковской модели с дискретными состояниями и непрерывным временем процесс использования онлайн-ресурса пользователем представляет собой так называемый *простейший поток* событий. Иными словами:

1. Все сутки являются рабочими днями, а выходные и праздники отсутствуют.
2. Не возникает новых пользователей, а существующие пользователи никогда не отказываются от использования ресурса.
3. Процесс ординарен и не обладает “последствием”.

В предлагаемой модели поведения аудитории каждый пользователь характеризуется одним параметром – интенсивностью использования ресурса λ . Если измерения проводятся в дискретные моменты времени, например, один раз в сутки, то простейшей моделью поведения пользователя является

последовательность испытаний (схема Бернулли), в которой вероятность обращения к ресурсу в течение суток есть

$$\theta = 1 - \exp(-\lambda \Delta t), \quad (1)$$

где Δt - одни сутки.

Известно [424], что вероятность того, что член панели, характеризуемый параметром θ , будет зарегистрирован счетчиком на изучаемом ресурсе ровно k дней из n , описывается биномиальным распределением

$$v(k; n, \theta) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}. \quad (2)$$

Здесь $\theta \geq 0$, k и n - целые числа, $0 \leq k \leq n$, а $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ - биномиальные коэффициенты.

Предположим, что пользователи могут отличаться своими параметрами θ : одни используют информационный ресурс - чаще, другие - реже. Предположим, что членов панели так много ($R \rightarrow \infty$), что параметр θ можно считать непрерывно распределённым на отрезке $0 \leq \theta \leq 1$, и опишем плотность его распределения функцией $\varphi(\theta)$. Тогда математическое ожидание доли членов панели, которые зарегистрированы счетчиком на изучаемом ресурсе в точности k дней из n , составит

$$\hat{u}_k(n) = \binom{n}{k} \int_0^1 \theta^k (1 - \theta)^{n-k} \varphi(\theta) d\theta. \quad (3)$$

Современные онлайн-панели предоставляют наборы $\hat{u}_k(n)$ для весьма широкого диапазона значений n . В частности, мы располагаем данными для $n = 1, 2, \dots, 120$.

Непараметрический подход к оценке распределения частоты использования информационного ресурса

На рис.2.16 слева изображены охваты аудитории почтового сервиса $\tilde{R}_k(n)$. При помощи рамки выделены кумулятивные охваты за максимальный период - 30 дней. Стрелки, которые исходят из рамки, показывают, где эти охваты

изображены на графике справа. График справа изображает функцию, дополнительную к эмпирической функции распределения параметра θ . Интерпретировать график следует так: каждая точка изображает число членов панели y , которые использовали ресурс не менее x дней из 30. Теоретически один этот график справа полностью эквивалентен полному набору 30 кривых слева, но только в том случае, если верна гипотеза о том, что поведение каждого пользователя описывает простейший поток событий.

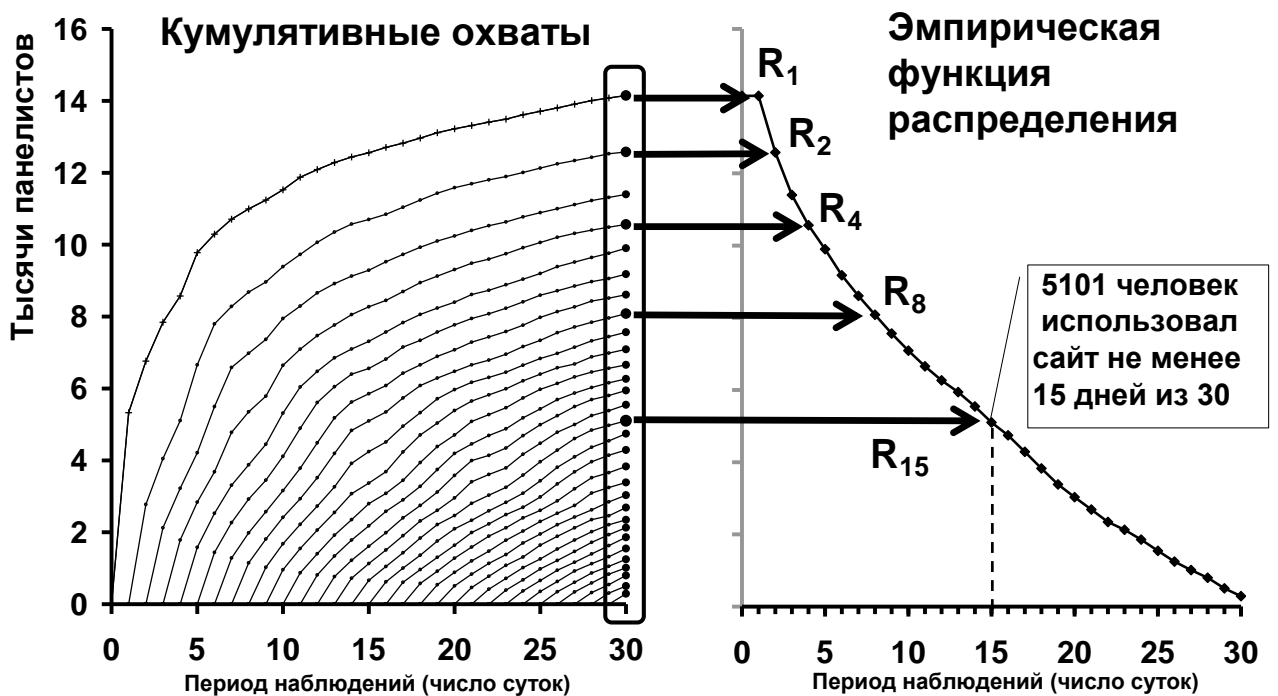


Рис.2.16. Кумулятивным охватам аудитории соответствует функция, дополнительная к эмпирической функции распределения интенсивности использования ресурса.

Рис.2.17 иллюстрирует результаты измерений за три периода – 5 суток, 15 суток и 30 суток. При этом масштаб горизонтальной оси приведен к единице. По вертикальной оси отложено количество членов панели (правильнее было бы отложить долю членов панели). Чем больше период наблюдений, тем слабее выражена разница между кривыми, что означает сходимость оценок.

Несмотря на простоту и наглядность непараметрического подхода, три его недостатка вынуждают использовать $R_k(n)$ при всех доступных k и n и прибегать к помощи параметрических моделей:

- 1) Эмпирические распределения не могут учесть эффект выходных дней.
- 2) Использование информационного ресурса пользователем может и не представлять собой простейшей поток. Если использование информационного ресурса пользователем не является простейшим потоком событий, а является, например, потоком Эрланга некоторого порядка, то график справа, разумеется, тоже будет связан со средними показателями процесса. Однако в этом случае эмпирические кривые $R_k(n)$ могут грубо расходиться с рассчитанными на основе формулы (3).
- 3) Аудитория информационного ресурса не постоянна во времени. Как в случае её возрастания, так и в случае оттока пользователей, по мере роста n эмпирические оценки функций распределения не будут сходиться.

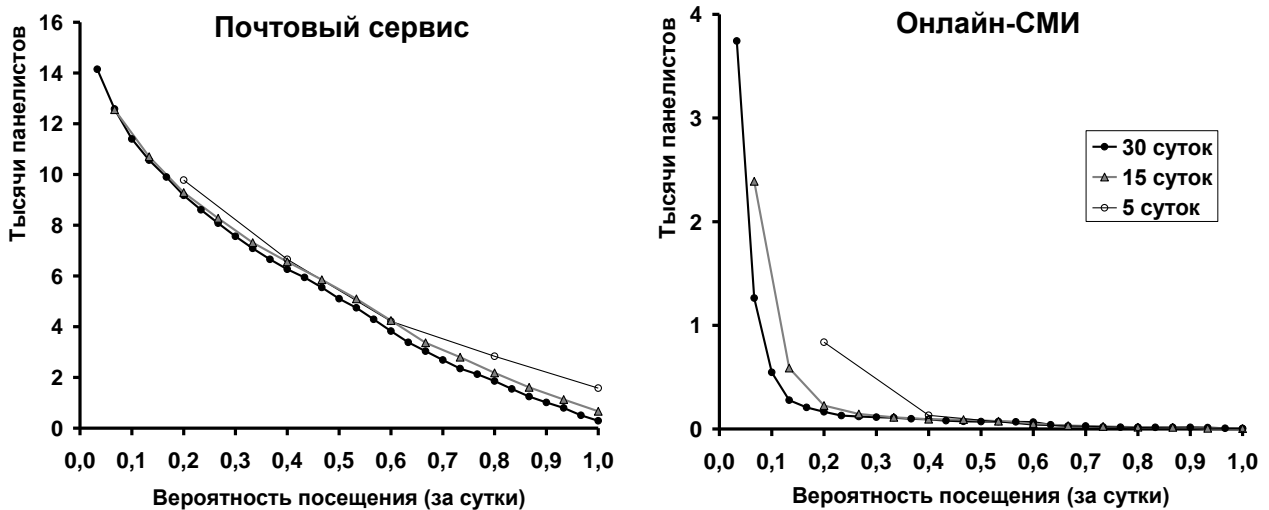


Рис.2.17. Эмпирические функции, дополнительные к функции распределения для почтового сервиса и онлайн-СМИ. Квадратики – охваты за 30 суток, треугольники – за 15, кружки – за 5 суток.

В силу указанных причин, так же как и в предыдущем разделе данной главы, мы используем параметрическую модель частоты использования информационного ресурса.

Параметрический подход к оценке распределения частоты использования информационного ресурса

Классическая бета-биномиальная модель была предложена для описания медиаметрических данных Дж.П.Хайетом в 50-е гг. XX века и стала широко известной благодаря статье Р.Мезерингэма [424]. Дж.П.Хайет предложил использовать в качестве смешивающей функции $\varphi(\theta)$ бета-распределение

$$\varphi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \quad (4)$$

где α и β – параметры бета-распределения, $B(\alpha, \beta)$ – бета-функция Эйлера.

Выражение для $\hat{u}_k(n)$ было получено Дж.Скелламом [460] в 1948 году

$$\hat{u}_k(n) = \binom{n}{k} \frac{B(k + \alpha, n - k + \beta)}{B(\alpha, \beta)}. \quad (5)$$

Это выражение носит название бета-биномиального распределения. Таким образом, для расчета теоретических кумулятивных кривых достаточно просуммировать величины, рассчитанные по формуле (5):

$$\tilde{R}_k(n) = R \cdot (\hat{u}_k(n) + \hat{u}_{k+1}(n) + \dots + \hat{u}_n(n)) \quad (6)$$

где $\tilde{R}_k(n)$ – k -й кумулятивный (накопленный) охват аудитории за n дней, то есть число членов панели, которые были зарегистрированы на изучаемом ресурсе *не менее* k дней из n , R – общий объём панели.

Ж.-Л.Шандон предложил [330] модификацию бета-биномиальной модели, названную им “двухсегментной бета-биномиальной моделью”, состоящую из двух сегментов: лояльной (“definite”) и случайной (“probable”) аудитории. Для лояльной аудитории $\theta=1$, таким образом $\lambda \rightarrow \infty$. Вместо формулы (4) в модели Шандона следует использовать выражение

$$\varphi(\theta) = \omega \cdot \delta(\theta-1) + (1-\omega) \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)},$$

где ω – доля лояльной аудитории, а вместо формулы (5) – выражение

$$\hat{u}_k(n) = \omega \cdot \delta_{kn} + (1 - \omega) \binom{n}{k} \frac{B(k + \alpha, n - k + \beta)}{B(\alpha, \beta)},$$

где использован символ Кронекера $\delta_{kn} = \begin{cases} 1 & \text{при } k = n \\ 0 & \text{при } k \neq n \end{cases}$.

Модель Шандона содержит лишь три параметра, однако учитывает важную особенность реальной аудитории: наличие лояльных пользователей. П.Данахер разработал алгоритм оценки параметров модели Шандона методом максимального правдоподобия и показал, что отсутствие члена, учитывающего лояльную аудиторию искажает оценки параметров модели [340]. Приведённая П.Данахером иллюстрация, однако, содержит заметный скачок амплитудой ω при $p = 1$. Фактически Ж.-Л. Шандон и П.Данахер обосновывают актуальность модели Шандона значительной величиной ω . Если бы ω была велика, то на рис.2.17 кривые спускались бы не к горизонтальной оси каждого из рисунков, а оставались несколько выше. В свою очередь на рис.2.15 и 2.16 в первый день амплитуда первой кумулятивной кривой должна была бы быть заметно выше, а все $\hat{u}_n(n) \geq \omega$, включая воспользовавшихся ресурсом 30 дней из 30, были бы велики. Мы заключаем, что данные онлайн-панелей не демонстрируют обособленный лояльный сегмент, как требует модель Ж.-Л.Шандона. Поэтому оба сегмента мы моделируем бета-распределениями, но с разными α_i и β_i .

Для построения удовлетворительной модели в нашей работе, написанной совместно с Т.А.Подлесной [160], и в работе [142] сделаны следующие предположения:

- (1) Аудитория каждого ресурса состоит из четырёх сегментов, которые мы условно назовём: 1) домашний активный; 2) домашний малоактивный; 3) офисный активный; 4) офисный малоактивный. Доли этих сегментов в общей аудитории ресурса обозначим соответственно w_1, w_2, w_3 и w_4 .
- (2) Каждый i -й сегмент характеризуется своими параметрами α_i и β_i ($i = 1..4$).
- (3) *Домашняя* аудитория состоит из пользователей, которые заходят на сайт со своего домашнего компьютера в любой день недели, в том числе – в выходные и

праздники. *Офисная* же аудитория по выходным и в праздники ресурсом не пользуется. Для офисного активного и офисного малоактивного сегментов в выражении для $\hat{u}_k(n)$ следует заменить полное число суток (включающее выходные) n на $m(n)$, то есть на число *рабочих дней*, пришедшееся на изучаемый период времени (n суток). Таким образом мы сталкиваемся с необходимостью использовать дополнительную ось времени, подобно тому, как при изучении эволюции возрастной структуры технологических систем приходится использовать “эксплуатационный возраст” системы, наряду с обычным понятием возраста [236].

Тогда вместо (3) получим

$$\hat{u}_k^i(n) = \begin{cases} \binom{n}{k} \int_0^1 \theta^k (1-\theta)^{n-k} \varphi_i(\theta) d\theta & \text{для } i = 1, 2 \\ \binom{m(n)}{k} \int_0^1 \theta^k (1-\theta)^{m(n)-k} \varphi_i(\theta) d\theta & \text{для } i = 3, 4 \end{cases},$$

где

$$\varphi_i(\theta) = w_i \frac{1}{\text{B}(\alpha_i, \beta_i)} \theta^{\alpha_i-1} (1-\theta)^{\beta_i-1}.$$

Итоговые формулы для частных охватов каждого сегмента $\hat{u}_k^{(i)}(n)$ ($i = 1..4$) даны в таблице 2.2.

Таблица 2.2. Формулы частных охватов $\hat{u}_k(n)$

	Активный	Малоактивный
Домашний	$\hat{u}_k^{(1)} = \binom{n}{k} \frac{\text{B}(k + \alpha_1, n - k + \beta_1)}{\text{B}(\alpha_1, \beta_1)}$	$\hat{u}_k^{(2)} = \binom{n}{k} \frac{\text{B}(k + \alpha_2, n - k + \beta_2)}{\text{B}(\alpha_2, \beta_2)}$
Офисный	$\hat{u}_k^{(3)} = \binom{m(n)}{k} \frac{\text{B}(k + \alpha_3, m(n) - k + \beta_3)}{\text{B}(\alpha_3, \beta_3)}$	$\hat{u}_k^{(4)} = \binom{m(n)}{k} \frac{\text{B}(k + \alpha_4, m(n) - k + \beta_4)}{\text{B}(\alpha_4, \beta_4)}$

Для доли членов панели, которые зарегистрированы счетчиком на изучаемом ресурсе в точности k дней из n , мы предлагаем формулу

$$\hat{u}_k(n) = w_1 \hat{u}_k^{(1)}(n) + w_2 \hat{u}_k^{(2)}(n) + w_3 \hat{u}_k^{(3)}(n) + w_4 \hat{u}_k^{(4)}(n). \quad (7)$$

Кумулятивные кривые охвата аудитории рассчитываются по формуле (6).

Доля активной аудитории почтового сервиса (рис.2.18) чрезвычайно велика. Значительную часть отклонений наблюдаемых данных от синтетических можно объяснить, если разделить аудиторию на две части, одна из которых (меньшая) не использует ресурс в выходные дни. Можно назвать ее “офисной” аудиторией. Таким образом, предположение о стационарности потока событий, не выполняется, поэтому ни классическая бета-биномиальная модель, ни смесь двух (или большего числа) бета-биномиальных распределений не описывают данные.

Требуется построить отдельную модель для “обычной” аудитории, а другую – для “офисной”. Как показывает рис.2.18, для описания данных этого уже вполне достаточно. В отличие от рис.2.15, сплошные линии на рис.2.18 – не просто соединительные линии, а теоретические кумулятивные кривые, рассчитанные при помощи моделей, приведенной на рис.2.17. Доля “офисной” аудитории для изученных нами ресурсов оказалась невелика (не более 20% от общей).

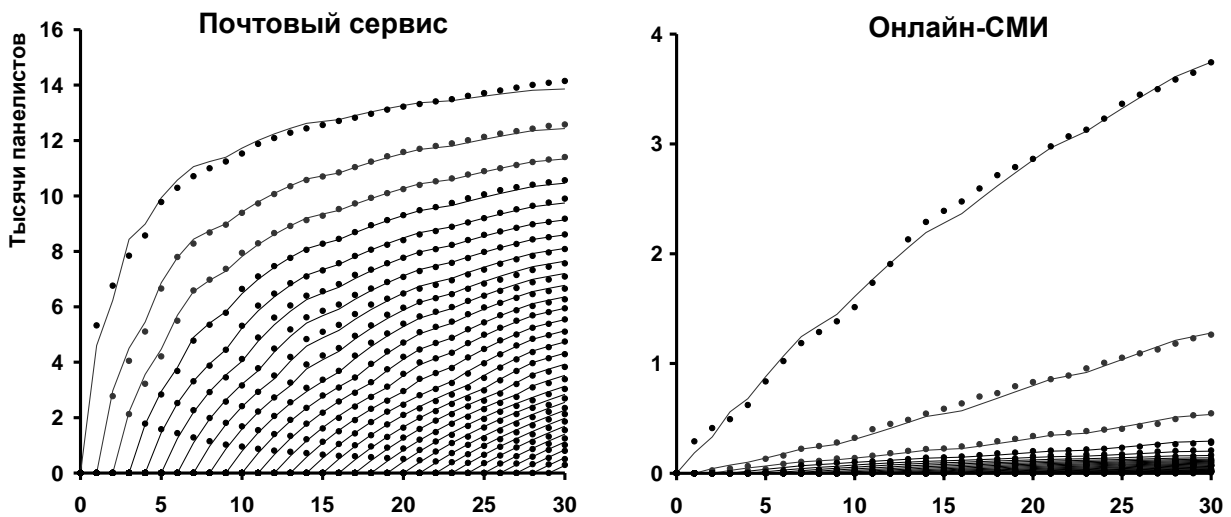


Рис.2.18 Кумулятивные охваты аудитории двух ресурсов: а) портал “большой тройки”, б) сайт популярной газеты.

Рис.2.19 позволяет сравнить “истинные” распределения параметра θ для аудитории почтового сервиса и онлайн-СМИ. Здесь сплошными линиями показаны результаты моделирования при помощи двух бета-распределений.

Тонкими линиями с точками – эмпирические функции распределения (те же, что и на рис.2.17).

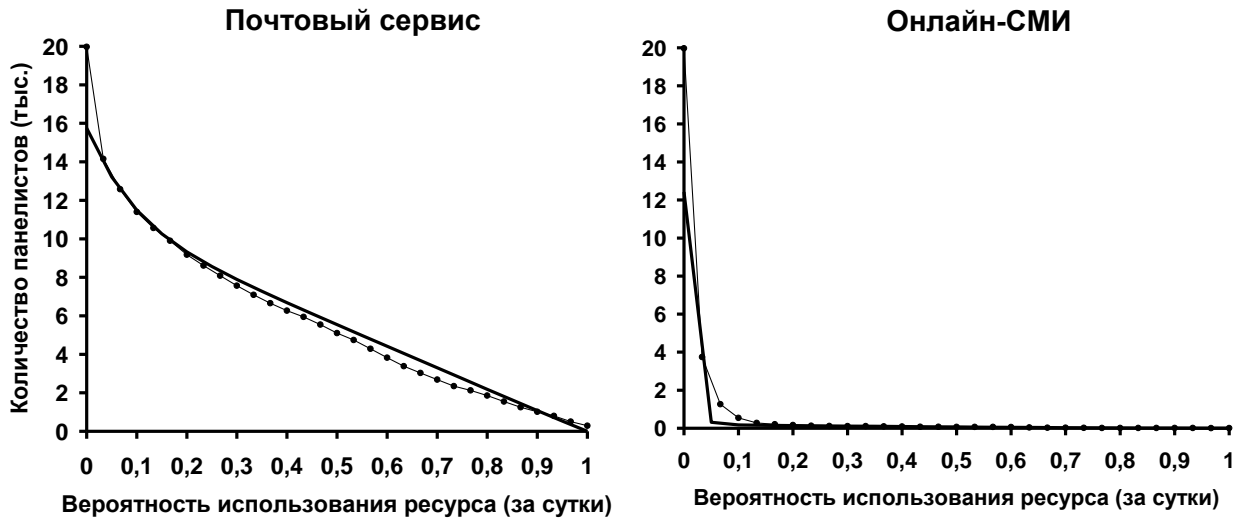


Рис.2.19. Теоретические (жирные линии) и эмпирические распределения (тонкие линии с кружками) интенсивности использования ресурса для почтового сервиса и онлайн-СМИ.

Поскольку эмпирические распределения практически совпадают с теоретическими, многие важные для управления аудиторией выводы можно сделать без математического моделирования. Однако эмпирические распределения не могут учесть эффект выходных дней. Кроме того, математические модели нужны для медиапланирования и для обоснованного принятия решений о развитии аудитории сайта.

Мы ожидаем, что моделирование кумулятивных кривых охвата аудитории Интернет-ресурсов при помощи смеси бета-биномиальных распределений войдет в состав программного обеспечения для медиапланирования, поэтому проведенные в нашей работе сравнения станут стандартной практикой.

2.3 Модель охвата аудитории информационных ресурсов WWW

В 2002–2006 гг. абсолютные величины охвата аудитории Интернета, полученные путем социологических опросов, широко использовались для моделирования и прогнозирования развития российского сегмента Сети

[63,48,148,197]. В то же время краткосрочную динамику аудитории Интернета в целом и отдельных ресурсов (т.н. “сайтов”) удобнее было изучать при помощи интернет-счетчиков или программ статистической обработки серверных лог-файлов [162,147]. В 2007-2008 г. социологический и инструментальный способы измерений охвата аудитории начали резко противоречить друг другу. Счетчики и внутренние статистические системы демонстрировали темпы прироста аудитории — от 40% до 70%, а результаты опросов Фонда «Общественное мнение» (ФОМ), ВЦИОМ, Левада-центра, TNS, не подтверждали такие темпы [48].

В работе [67] нами было проведено сравнение темпов прироста охвата аудитории одного из крупнейших российских Интернет-проектов, вычисленных по данным его собственной статистической системы и с помощью онлайн-панели TNS за 2007-2008 гг. Различия в поведении темпов прироста оказались чрезвычайно велики: по данным панельных исследований они быстро падали, а по данным внутренней статистической системы оставались постоянными и чрезвычайно высокими — около 70% в год. Такие расхождения представляли серьезные затруднения для руководства Интернет-компаний.

Непосредственное сопоставление результатов опросов и измерений счетчиков затруднено, в частности, тем, что пользователи время от времени меняют компьютеры, операционные системы, стирают файлы cookies, записанные на их жесткие диски счетчиками [389]. Файлы cookies и есть главный источник сведений об аудитории сайтов. Если период измерений достаточно длинный, например, месяц, то один и тот же пользователь, обновив cookie, может быть учтен счетчиком несколько раз. Ситуация еще более осложняется использованием нескольких браузеров, различных способов выхода в Интернет, а также тем, что на одном компьютере работают разные люди и т.д. Некоторые пользователи отключают cookies, и счетчики пытаются учесть их при помощи IP-адресов и браузеров. Несмотря на то, что пользователей немного, счетчики могут учитывать их заново каждый день, и такой «вклад» в увеличение месячной аудитории окажется заметным.

Попытку согласовать разнородные данные предпринял исследовательский комитет ассоциации «Интернет и бизнес» [183,197]. По оценке комитета, по сравнению с концом 2006 г. количество российских пользователей Интернета, посещающих сайты Рунета хотя бы раз в месяц, выросло на 29%. При оценке роста аудитории Интернета в России исследовательский комитет учел значительное число факторов («многобраузерность», большое число мест доступа и прочее) и ввел ряд поправок. В частности, поскольку порталы не публиковали динамику российской аудитории, комитет оценил темпы ее прироста при помощи онлайн-опросов. Прделанная комитетом работа, с нашей точки зрения, важна в первую очередь как иллюстрация реального процесса информационного обеспечения принятия решений, – лица, принимающие решения ожидают от аналитиков разработку таких показателей, которые наиболее выгодным образом описывают деятельность предприятий.

Как показано в нашей работе [67], кумулятивные кривые охвата аудитории, построенные по данным социологических опросов и интернет-счетчика, значительно различаются. Кумулятивная кривая, соответствующая результатам опросов, быстро выходит на асимптоту (см. рис. 2.10). Для данных за осень 2005 г. месячная аудитория составляет более 80% полной аудитории Рунета. Отметим, что в более новых данных, собранных весной 2008 г., доля месячной аудитории превысила 90%. Иными словами, в аудитории, измеренной путем опросов, очень мало «редких» пользователей. Напротив, кумулятивная кривая, построенная на основе данных статистики крупного портала, явно далека от выхода на уровень насыщения. Такая кривая продолжает расти с постоянной скоростью. Отметим, что данные популярного счетчика LiveInternet (www.liveinternet.ru) о частоте посещения крупных Интернет-ресурсов, свидетельствуют, что три четверти представителей аудитории портала проводят на страницах даже самых крупных проектов не больше одного дня в месяц.

2.3.1 Модель поведения пользователей информационного ресурса WWW

В этом разделе мы ограничимся исследованием только одной проблемы из числа перечисленных ранее: эффекта удаления файлов cookies. В связи с этим мы построили простейшую модель кумулятивного охвата аудитории, которая учитывала бы эффект смены cookies и удовлетворяла данным наблюдений [83,109], т.е. теоретические кумулятивные кривые для счетчиков.

Рассмотрим пользователей, для которых (рис. 2.20)

- посещение информационного ресурса описывается простейшим (стационарным пуассоновским) потоком событий с интенсивностью λ ;
- уничтожение файлов cookies описывается также простейшим потоком с интенсивностью μ .

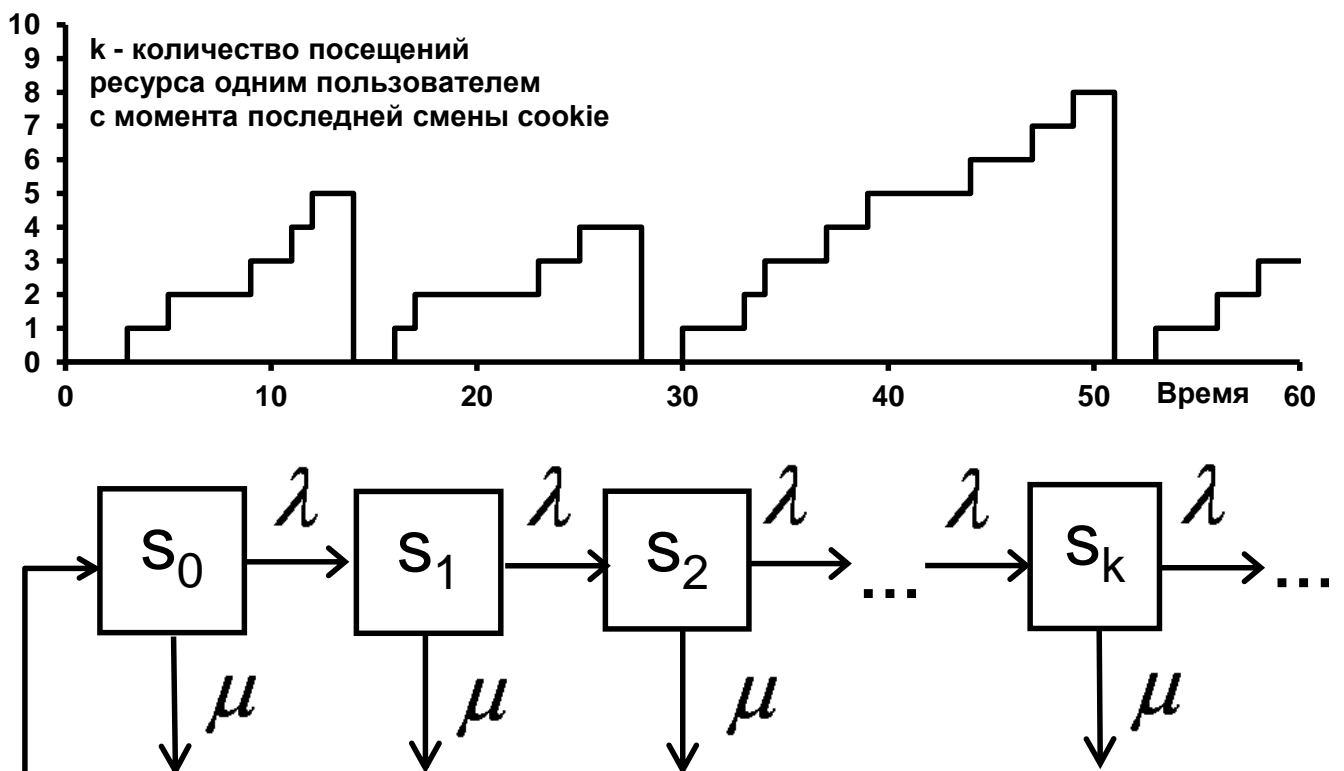


Рис.2.20. Вверху – количество посещений ресурса одним пользователем с момента последней смены cookie. Внизу – граф состояний актуальных уникальных пользователей.

Для простейшего потока событий вероятность наступления нового события не зависит от предшествующей истории, интенсивность потока не меняется со временем, а события происходят поодиночке, а не «пачками» [37]. Мы используем такую модель поведения пользователей по причине ее простоты и возможности легко получить явные выражения для абсолютного охвата аудитории $R_k(t; \lambda, \mu)$.

Отметим, что рассматриваемый случайный процесс **не** образует потока однородных событий, поскольку использование ресурса и смена cookies являются разнородными событиями. Дискретная случайная величина k , равная количеству посещений информационного ресурса пользователем после последней смены cookie, может не только возрастать, но и падать скачком до нуля, а потом возрастать снова (см. верхнюю часть рис.2.20). Таким образом теория входящих потоков однородных событий, которая использовалась в предыдущих разделах нашей работы, не применима к данной задаче.

Обозначим $p(\lambda, \mu)$ плотность распределения (п.р.) числа пользователей, посещающих ресурс с интенсивностью λ и стирающих cookie с интенсивностью μ . Сначала мы должны найти кумулятивные кривые $R_k(t; \lambda, \mu)$ для заданных значений λ , μ периода измерений t и числа визитов k . Эти кумулятивные кривые описывают численность уникальных пользователей (далее — у.п.), т.е. cookies, пользовательских профилей и иным способом определенных пользователей (например, при помощи IP-адреса и браузера), которые в течение времени t были зарегистрированы статистической системой не менее, чем k раз.

Общие кумулятивные кривые являются линейной суперпозицией таких решений:

$$\bar{R}_k(t) = \int_0^{\infty} \int_0^{\infty} p(\lambda, \mu) R_k(t; \lambda, \mu) d\lambda d\mu. \quad (1)$$

Пусть общая аудитория российского Интернета составляет R пользователей и постоянна во времени (решение для монотонно неубывающей аудитории может быть получено при помощи свертки во временной области).

Чтобы рассчитать абсолютные охваты аудитории $R_k(t; \lambda, \mu)$, требуются вспомогательные величины $u_k(t; \lambda, \mu)$ — численность актуальных у.п. (уникальных пользователей, cookies, пользовательских профилей), которые:

- в течение времени t были зарегистрированы счетчиком ровно k раз;
- «живы» в момент t .

В частности, $u_0(t; \lambda, \mu)$ — это численность актуальных у.п., которые за время t еще не были зарегистрированы счетчиком; $u_1(t; \lambda, \mu)$ — численность актуальных у.п., которые за время t были зарегистрированы счетчиком ровно один раз, и т.д. Все эти профили актуальны («живы») в момент t . В отличие от $u_k(t; \lambda, \mu)$, охваты $R_k(t; \lambda, \mu)$ включают и те cookies, которые после регистрации были уничтожены и больше уже не появятся.

Множества сгруппированных таким образом актуальных у.п. не пересекаются. В любой момент времени верны уравнения баланса

$$\sum_{k=0}^{\infty} u_k(t; \lambda, \mu) = R. \quad (2)$$

В начальный момент времени:

$$\begin{aligned} u_0(0; \lambda, \mu) &= R \\ u_1(0; \lambda, \mu) &= u_2(0; \lambda, \mu) = \dots = u_k(0; \lambda, \mu) = \dots = 0 \end{aligned} \quad (3)$$

Численности актуальных у.п. связаны посредством уравнений (с точностью до амплитуды совпадающих с уравнениями Колмогорова [37] для вероятностей):

$$\begin{aligned} \frac{du_0}{dt} &= \mu \cdot R - (\lambda + \mu)u_0 \\ \frac{du_k}{dt} &= \lambda u_{k-1} - (\lambda + \mu)u_k \\ &\dots \end{aligned} \quad (4)$$

Отметим, что величины $\hat{u}_k(t; \lambda, \mu) = \frac{1}{R} u_k(t; \lambda, \mu)$ есть вероятности регистрации пользователя (из данной группы, характеризуемой параметрами λ и μ) счетчиком ровно k раз за время t .

Для начала проинтерпретируем второе уравнение. Число актуальных у.п., зарегистрированных ровно один раз, прирастает за счет регистрации «новых», ранее не регистрировавшихся актуальных у.п. и убывает как за счет повторной регистрации части этих «уникальных пользователей» (с интенсивностью λ), так и за счет уничтожения другой их части (с интенсивностью μ). Аналогично интерпретируются третье и все последующие уравнения.

Первое уравнение интерпретируется несколько иначе. Прирост «свежих», не зарегистрированных актуальных у.п. обеспечивается «смертностью» всех групп актуальных у.п. (т.е. уничтожением cookies). Некоторая часть «новорожденных» актуальных у.п. уничтожается еще до регистрации. В частности, пользователь может установить новую операционную систему, а затем обнаружить сбой в работе программного обеспечения и переустановить систему снова, не будучи зарегистрированным счетчиком.

Связи между состояниями а.у.п. иллюстрирует нижняя половина рис.2.20.

Вывод решений для а.у.п. при помощи преобразований Лапласа дан в нашей работе [72]. Решения можно выразить в следующем компактном виде

$$u_k(t; \lambda, \mu) = R \cdot \frac{\lambda^k}{(\lambda + \mu)^{k+1}} (f_{k+1}(t; \lambda + \mu) + \mu F_{k+1}(t; \lambda + \mu)), \quad (5)$$

$$\text{где } f_k(t; \lambda + \mu) = (\lambda + \mu) \times \frac{((\lambda + \mu)t)^{k-1}}{(k-1)!} \times e^{-(\lambda + \mu)t} \quad (6)$$

— плотность распределения (п.р.) Эрланга k -го порядка;

$$F_k(t; \lambda + \mu) = 1 - e^{-(\lambda + \mu)t} \sum_{m=0}^{k-1} \frac{((\lambda + \mu)t)^m}{m!} \quad (7)$$

— функция распределения (ф.р.) Эрланга k -го порядка.

Выражение (5) легко вывести при помощи вероятностного подхода. Вероятность того, что а.у.п. не сменит cookie за время t составляет $e^{-\mu t}$. Вероятность того, что а.у.п. за время t зарегистрирован счётчиком k раз, **при условии**, что эта а.у.п. не сменил cookie, составляет

$$\frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Соответственно, вероятность того, что а.у.п. за время t зарегистрирован счётчиком k раз и не сменил cookie, вычисляется как произведение

$$\frac{(\lambda t)^k}{k!} e^{-\lambda t} \times e^{-\mu t} = \frac{\lambda^k}{(\lambda + \mu)^{k+1}} f_{k+1}(t; \lambda + \mu).$$

Умножив на R , получаем первое слагаемое в правой части формулы (5).

Однако а.у.п. может и сменить cookie после обращения к ресурсу. На малом отрезке времени $[t - \tau, t - \tau + \Delta\tau]$ один а.у.п. успеет сменить cookie $\mu\Delta\tau$ раз. Вероятность того, что а.у.п. сменил cookie на отрезке $[t - \tau, t - \tau + \Delta\tau]$, а затем в период $[t - \tau, t]$ был зарегистрирован счётчиком k раз, уже не меняя cookie, есть

$$\mu\Delta\tau \times \frac{\lambda^k}{(\lambda + \mu)^{k+1}} f_{k+1}(\tau; \lambda + \mu).$$

Проинтегрируем это выражение по τ и умножим полученную сумму на R . Результатом является второе слагаемое в правой части формулы (5).

Перейдем к расчету охватов $R_k(t; \lambda, \mu)$. Численность у.п., зарегистрированных не менее k раз, пополняется только за счет актуальных у.п., зарегистрированных не менее $k - 1$ раз:

$$\frac{dR_k}{dt} = \lambda u_{k-1}(t; \lambda, \mu) \quad (11)$$

Для вычисления R_k , перепишем выражение для ф.р. Эрланга k -го порядка

$$F_k(t; \lambda + \mu) = 1 - \frac{1}{\lambda + \mu} \sum_{m=1}^k f_m(t; \lambda + \mu) \quad (12)$$

Число «живых» cookies, зарегистрированных ровно k раз за время t , приобретает вид

$$u_k(t; \lambda, \mu) = R \frac{\lambda^k}{(\lambda + \mu)^{k+1}} \left(\mu - \frac{\mu}{\lambda + \mu} \sum_{m=1}^k f_m(t; \lambda + \mu) + \frac{\lambda}{\lambda + \mu} f_{k+1}(t; \lambda + \mu) \right) \quad (13)$$

Используя связь между п.р. и ф.р., искомые охваты запишем в виде

$$R_k(t; \lambda, \mu) = R \frac{\lambda^k}{(\lambda + \mu)^k} \left(\mu t - \frac{\mu}{\lambda + \mu} \sum_{m=1}^{k-1} F_m(t; \lambda + \mu) + \frac{\lambda}{\lambda + \mu} F_k(t; \lambda + \mu) \right) \quad (14)$$

Эта формула использовалась нами в работах [67,83,103,109], позже мы нашли гораздо более простое выражение, которое приводим здесь без доказательства

$$R_k(t; \lambda, \mu) = R \cdot \frac{\lambda^k}{(\lambda + \mu)^k} \left((1 + \mu t) F_k(t; \lambda + \mu) - k \frac{\mu}{\lambda + \mu} F_{k+1}(t; \lambda + \mu) \right).$$

Покажем, что знать одну лишь кумулятивную кривую охвата $R_1(t)$ недостаточно, ибо одну и ту же кривую $R_1(t)$ можно описать разными моделями.

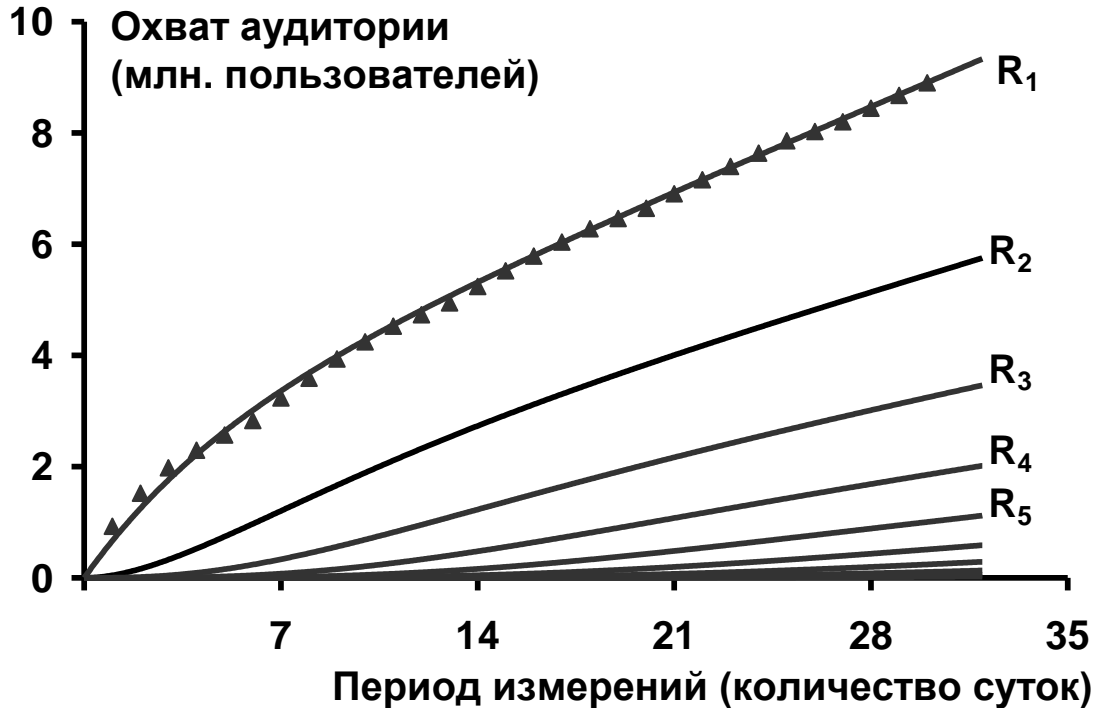


Рис.2.21. Накопленные охваты зарегистрированных не менее k раз за период t «уникальных пользователей» $R_k(t)$. Сплошная линия — результаты расчетов, треугольники — реальные данные за 2005 г.

Однородная модель

Кумулятивный охват аудитории, соответствующий модели, состоящей из одного сегмента пользователей, имеет вид

$$R_1^{(1)}(t; \lambda, \mu) = R \frac{\lambda \mu}{\lambda + \mu} t + R \frac{\lambda^2}{(\lambda + \mu)^2} (1 - \exp(-(\lambda + \mu)t)). \quad (15)$$

Рис. 2.21. иллюстрирует накопленные охваты аудитории $R_k(t)$, рассчитанные для значений параметров $\lambda = 1/7$ суток⁻¹, $\mu = 1/16$ суток⁻¹, $R = 5,0$ млн. человек.

Однако полученные значения параметров представляются неправдоподобными. Особенно неправдоподобно частой выглядит смена cookies, производимая каждые две недели, что противоречит опубликованным данным [183,389].

Модель с двумя сегментами пользователей без смены cookies

Кумулятивный охват аудитории, состоящей из двух сегментов: “активного” сегмента и сегмента “бесконечно редких” пользователей, которые используют ресурс бесконечно редко, может быть представлен в виде:

$$R_1^{(2)}(t; l, m) = \left(A \frac{lm}{l+m} + B \right) \cdot t + A \frac{l^2}{(l+m)^2} (1 - \exp(-(l+m)t)) \quad (16)$$

где $\{A, l, m\}$ - параметры активного сегмента, B - численность сегмента “редких” пользователей. Найдем параметры такой модели из двух сегментов, чтобы выполнялось тождество $R_1^{(1)}(t; \lambda, \mu) \equiv R_1^{(2)}(t; l, m)$. Очевидно, такие параметры должны удовлетворять трем условиям

$$A \frac{lm}{l+m} + B = R \frac{\lambda\mu}{\lambda+\mu}, \quad l+m = \lambda+\mu, \quad \text{и} \quad A \frac{l^2}{(l+m)^2} = R \frac{\lambda^2}{(\lambda+\mu)^2}. \quad (17)$$

Если в качестве свободного параметра выбрать параметр m , то тождество $R_1^{(1)}(t; \lambda, \mu) \equiv R_1^{(2)}(t; l, m)$ окажется истинным для любых значения времени при выполнении равенств:

$$l = \lambda + \mu - m \quad A = \frac{\lambda^2}{l^2} R \quad B = \frac{R\lambda\mu - Alm}{\lambda + \mu}. \quad (18)$$

Из первого и третьего равенств и очевидного требования $B \geq 0$ следует дополнительное ограничение $m \leq \mu$.

Таким образом, для произвольных неотрицательных параметров однородной модели λ, μ, R можно найти бесконечно много таких неотрицательных параметров l, m, A, B модели из двух сегментов, что кривые $R_1(t)$ для двух моделей совпадут.

В силу аддитивности $R_1(t)$ решение задачи идентификации параметров модели (численности R и распределения $p(\lambda, \mu)$) по кумулятивному охвату $R_1(t)$

не единственно: существует бесконечное число различных моделей с совпадающими $R_1(t)$ [98]. Рис.2.22 иллюстрирует накопленные охваты аудитории $R_k(t)$, рассчитанные для модели, включающей:

- 1) “активную” аудиторию с интенсивностью использования $l = 0,20$ суток⁻¹ численностью $A = 2,4$ млн. человек, которые вообще не меняют cookie ($m = 0$ суток⁻¹), и
- 2) “бесконечно редкую” аудиторию с суточным охватом в $B = 187,6$ тыс. уникальных пользователей в сутки, бесконечно малой интенсивностью использования ресурса и произвольной частотой смены cookies.

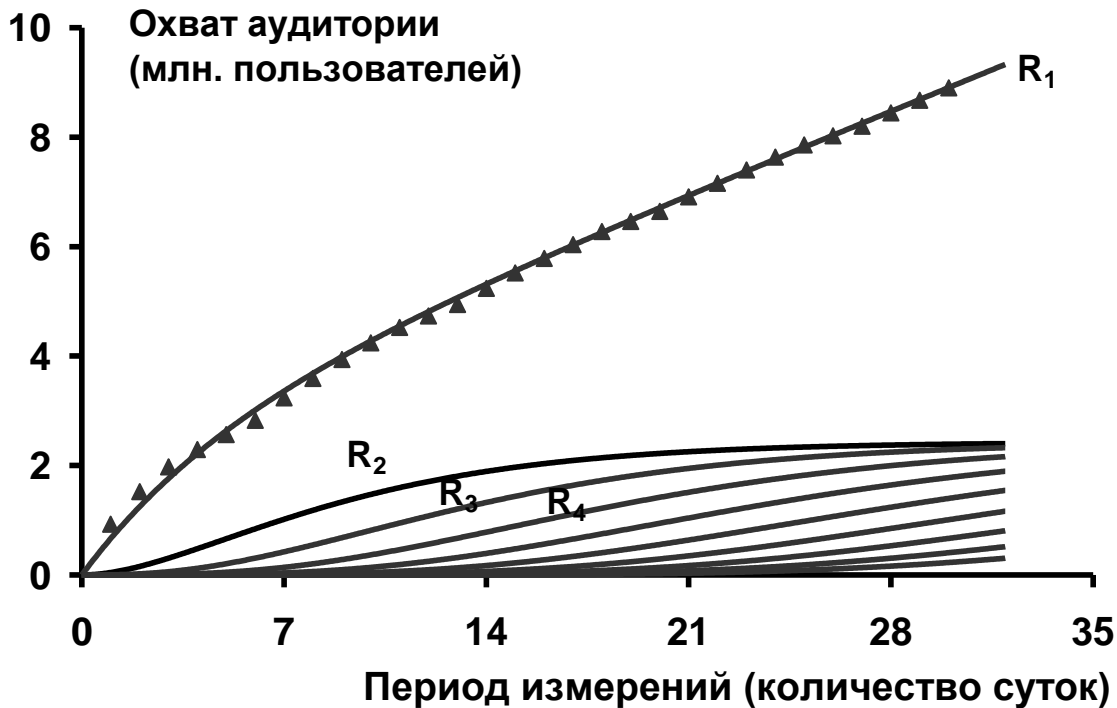


Рис.2.22. Накопленная аудитория зарегистрированных не менее k раз «уникальных пользователей». Для расчетов использована модель с двумя сегментами пользователей, не меняющих cookies.

Модель с двумя сегментами пользователей и сменой cookies

Результаты исследований поведения пользователей, проведенные компаниями Jupiter Research [389], Nielsen/NetRatings, Atlas Solutions и Gemius

[183] свидетельствуют, что от 39 до 43% пользователей WWW меняет cookies в течение одного месяца. В России по словам Марины Соловьевой, руководителя российского отделения Gemius: “мы фиксируем ежемесячный процент обновления на уровне 41%” [183]. Если уничтожение cookies описывается простейшим потоком, то его интенсивность должна составлять $m = 1/57 \text{ суток}^{-1}$ (пользователь меняет cookie в среднем один раз в 57 суток).

Результаты расчетов по такой, наиболее реалистичной из рассмотренных, модели показаны на рис.2.23. Здесь $l = 0,19 \text{ суток}^{-1}$, $A = 2,9$ млн. человек, $m = 0,018 \text{ суток}^{-1}$, $B = 169,5$ тыс. уникальных пользователей в сутки. Мы предполагаем, что в реальности охваты высших порядков $R_k(t)$ для рассматриваемого ресурса должны были бы выглядеть так, как показано на рис. 2.23.

На рис.2.23 можно видеть, что кривые охвата $R_k(t)$ высших порядков ($k \geq 2$) существенно различаются. Таким образом, остается возможность однозначно определить R и $p(\lambda, \mu)$ по нескольким кривым $R_k(t)$. К сожалению, в отличие от классического случае чтения газет, в случае аудитории WWW охваты $R_k(t)$ высших порядков не могут быть надежно измерены, поскольку трудно определить, когда закончилось первое событие, состоящее в использовании ресурса, и началось следующее.

Отметим, что аналогичные решения можно было бы найти для любых $0 \leq m \leq \mu = 1/16$. Более того, взвешенные средние трех рассмотренных моделей породят ту же самую $R_1(t)$.

На рис.2.21-2.23 можно видеть, что теоретические кривые охвата $R_k(t)$ высших порядков ($k \geq 2$) существенно различаются. Требуется выяснить, можно ли восстановить распределение интенсивности по нескольким кривым $R_k(t)$ (предполагая, что часть необходимой информации, например, зависимость p от μ , известна).

К сожалению, в отличие от классического случае чтения газет, в случае аудитории WWW охваты $R_k(t)$ высших порядков не могут быть надежно измерены, поскольку трудно определить, когда закончилось первое событие, состоящее в использовании ресурса, и началось следующее.

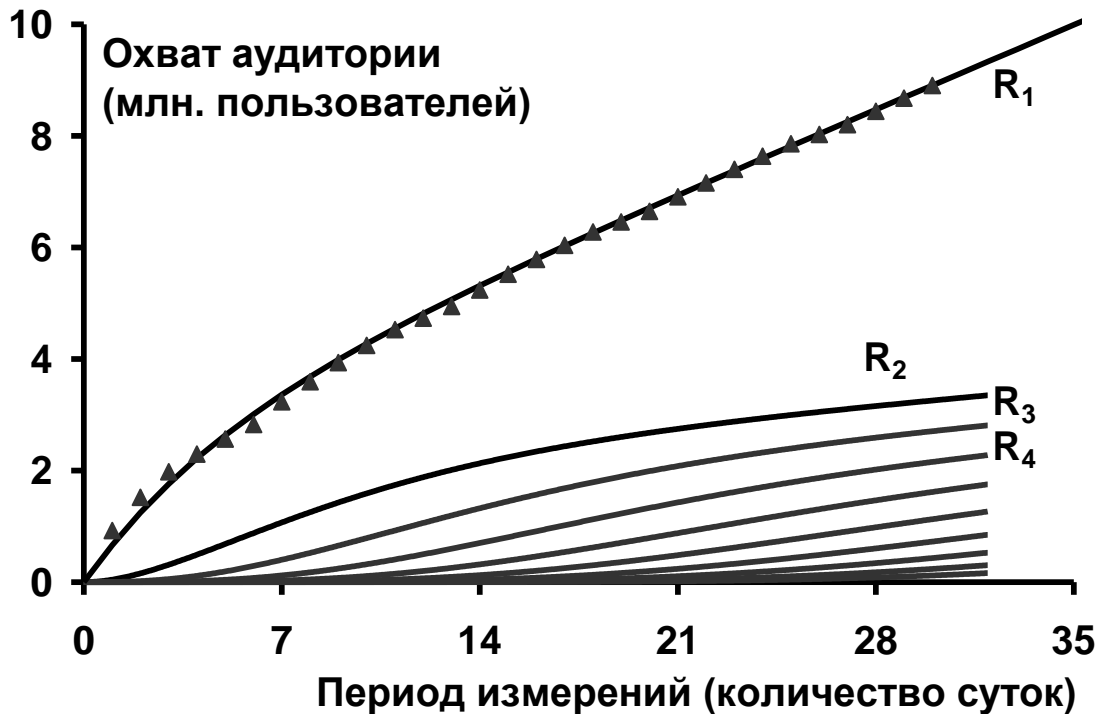


Рис.2.23. Накопленная аудитория зарегистрированных не менее k раз «уникальных пользователей». Для расчетов используется модель с двумя сегментами пользователей и сменой cookies.

Практическое решение проблемы расчета кумулятивных охватов высших порядков известно: счетчики измеряют и публикуют информацию о количестве пользователей, которые были отмечены на сайте только в один день месяца, только в два дня месяца и т.д.. Такие измерения разные системы способны выполнять единообразно. Поэтому мы предлагаем исследовать возможность ограничить допустимые решения при помощи других кривых — $\tilde{R}_k(n)$, которые описывают количество уникальных пользователей, посетивших ресурс в течение k дней из n , и которые рассмотрены в последнем разделе данной главы.

Дополнительная сложность состоит в том, что в реально наблюдаемых данных ярко выражен эффект выходных дней, когда часть аудитории не пользуется Интернетом. В этой работе мы наметим путь учета этого эффекта для кривых $R_k(t)$. Для нестационарной задачи специального вида:

$$\begin{aligned}\lambda(t) &= s(t) \cdot \lambda \\ \mu(t) &= s(t) \cdot \mu\end{aligned}\quad (19)$$

с помощью подстановки $dS = s(t)dt$ можно получить решения достаточно простого вида:

$$u_k(t; \lambda, \mu) = R \cdot \frac{\lambda^k}{(\lambda + \mu)^{k+1}} (f_{k+1}(S(t); \lambda + \mu) + \mu F_{k+1}(S(t); \lambda + \mu)) \quad (20)$$

$$R_k(t; \lambda, \mu) = R \left(\frac{\lambda}{\lambda + \mu} \right)^k \left(\mu S(t) - \frac{\mu}{\lambda + \mu} \sum_{m=1}^{k-1} F_m(S(t); \lambda + \mu) + \frac{\lambda}{\lambda + \mu} F_k(S(t); \lambda + \mu) \right) \quad (21)$$

Это позволяет разделить аудиторию на два сегмента: “рабочий” и “домашний”, соответствующих использованию рабочего и домашнего компьютеров. При этом можно предположить, что в выходные дни почти все “рабочие” компьютеры выключены. Такая, более сложная, модель позволяет точнее описать данные.

Возвращаясь к исходной проблеме оценки динамики роста российского Интернета, отметим, что «редкие» пользователи вкупе с эффектом уничтожения cookies чрезвычайно сильно влияют на кумулятивный месячный охват аудитории и маскируют его реальный рост. В настоящее время онлайн-статистика показывает нам не реальную динамику месячного охвата аудитории, а совокупность большого числа эффектов, которые нельзя учесть, не зная их механизма; ряд таких эффектов упомянут в статье [197].

Аудитория многих крупных российских информационных WWW-ресурсов растет уже более десяти лет, и это делает актуальной задачу учета ее роста, поставленную нами в [90,99]. В этих работах нами были получены теоретические кумулятивные кривые охвата аудитории WWW-ресурса с учетом уничтожения записей cookies для случая неубывающего во времени полного количества пользователей. Для некоторых видов зависимости полного объема аудитории от

времени, в частности для квазимногочленов, точные величины охватов можно получить с помощью известных алгоритмов. Такой способ расчета теоретических кривых $R_k(t)$ может быть использован в качестве подпрограммы решения прямой задачи при численном решении обратной задачи восстановления динамики полной аудитории $x(t; \lambda, \mu)$ по наблюдаемым кривым охвата $R_k(t)$.

Кроме того, в работах [102,96] нами была предложена модель для расчета теоретических кумулятивных кривых охвата аудитории WWW-ресурса с учетом уничтожения записей cookies при наличии не только притока новых, так и оттока существующих пользователей.

В случае аудитории WWW-ресурсов в настоящее время существуют способы измерения только кривой первого порядка $R_1(t)$, причём лишь в некоторые моменты времени, а для кривых высшего порядка доступны лишь аналоги, моделирование которых требует другого теоретического аппарата и рассмотрено в следующем разделе этой главы.

Однако упомянутые здесь теоретические результаты могут найти применение при анализе баз данных электронной коммерции, если под $R_k(\tau)$ подразумевать не кумулятивные охвата аудитории, а количество потребителей услуг, совершивших не менее k транзакций за период времени τ . В приложении Б к диссертации приведен акт от 03.03.2014 г. о внедрении в ООО “Юду” разработанных нами моделей.

Модель потока заявок Интернет-портала услуг

В отличие от Интернет-счётчиков, статистические системы учёта электронных порталов услуг (так же как и аналогичные системы электронных магазинов) регистрируют своих пользователей, которые используют имя аккаунта (login) и пароль (password) для аутентификации. Для упрощения работы пользователя и такие системы используют механизм “cookies”, однако, если пользователь стёр cookie, то система спросит у него пароль, определит, что он уже зарегистрирован и не станет учитывать его многократно.

Как показывает практический опыт таких систем, отток пользователей, разочаровавшихся в товарах или услугах, оказывается столь существенен, что им нельзя пренебречь. Будем считать, что отток пользователей можно описать простейшим потоком событий с параметром μ , таким образом он аналогичен смене cookies. Переходы между состояниями пользователей такого портала можно изобразить при помощи диаграммы 2.24. Отличие этой диаграммы от нижней части рис. 2.20 состоит в том, что пользователи не возвращаются в состояние S_0 , а переходят в терминальное состояние S^* .

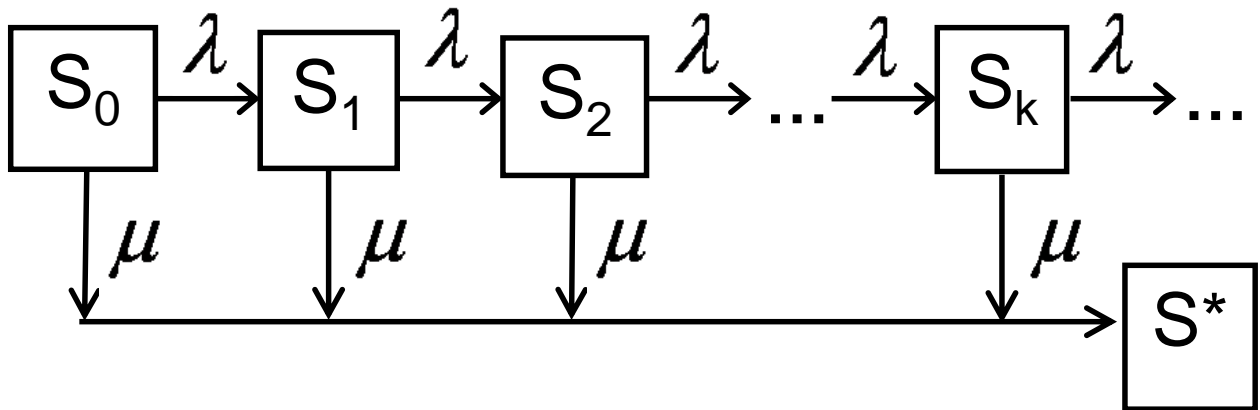


Рис.2.24. Граф состояний пользователя системы заявок

Система уравнений такой модели оказывается однородной и её решения имеют более простой вид, чем полученные в предыдущем разделе

$$\hat{u}_k(\tau; \lambda, \mu) = \frac{\lambda^k}{(\lambda + \mu)^{k+1}} f_{k+1}(\tau; \lambda + \mu) \quad \text{и} \quad R_k(\tau; \lambda, \mu) = R \cdot \frac{\lambda^k}{(\lambda + \mu)^k} F_k(\tau; \lambda + \mu).$$

Здесь $\hat{u}_k(\tau; \lambda, \mu)$ есть вероятность того, что пользователь отправил ровно k заявок за время τ , $R_k(\tau; \lambda, \mu)$ – кумулятивные охваты пользователей (количество

пользователей, отправивших не менее k заявок за время τ). Поскольку неизвестно, остаются ли индивиды, сделавшие заказ, пользователями, или они уже отказались от дальнейшего использования системы, величину $\hat{u}_k(\tau; \lambda, \mu)$ непосредственно измерить нельзя. Однако, в отличие от случая счётчика, измерению поддаются все охваты $\bar{R}_k(\tau)$, рассчитываемые по формуле (1).

В частности, в работе Д.Шмитляйна, Д.Морисона и Р.Коломбо [454] в качестве смешивающего распределения выбрано двумерное гамма-распределение с плотностью

$$p(\lambda, \mu; b, \alpha, c, \beta) = bc \frac{(b\lambda)^{\alpha-1}}{\Gamma(\alpha)} \frac{(c\mu)^{\beta-1}}{\Gamma(\beta)} e^{-b\lambda - c\mu},$$

что позволяет получить распределение дискретной переменной k (т.е. количества заявок за время τ)

$$\bar{u}_k(\tau) = \frac{\Gamma(k + \alpha)}{k! \Gamma(\alpha)} \frac{b^\alpha \tau^k}{(b + \tau)^{k+\alpha}} \frac{c^\beta}{(c + \tau)^\beta},$$

зависящее от четырёх параметров b, α, c, β , которое Д.Шмитляйн и соавторы назвали Парето-NBD распределением. Хотя кумулятивные охваты $\bar{R}_k(\tau)$ и не выражаются через элементарные функции, для производных охватов по времени, интегрируя формулу (11), можно получить нужное выражение

$$\frac{d\bar{R}_k}{d\tau} = R_0 \frac{k}{\tau} \bar{u}_k(\tau).$$

Тогда $\bar{R}_k(\tau)$ можно рассчитать при помощи численного интегрирования.

Среднее число заявок в единицу времени выражается формулой

$$\bar{k}(\tau) = R_0 \frac{\alpha}{b} \frac{c^\beta}{(c + \tau)^\beta}.$$

При практическом моделировании статистики заявок онлайн-ресурса мы использовали смесь двух Парето-NBD распределений. Первое распределение описывало пользователей, которые быстро разочаровывались в ресурсе, а второе – долговременных пользователей. В качестве данных ООО «ЮДу» предоставило нам ряд измерений кумулятивных охватов $\bar{R}_k(\tau_i)$ и соответствующих $\bar{k}(\tau_i)$. Разработанная нами модель была использована в деятельности ООО «ЮДу», что

подтверждается актом о внедрении, приведённом в приложении Б к данной работе.

Мы рекомендуем операторам Интернет-счётчиков классифицировать а.у.п., использующих информационные ресурсы Интернета в течение месяца на два класса: 1) а.у.п., уже зарегистрированные счётчиком ранее, 2) а.у.п., которым cookie была впервые выдана в этом месяце. Статистические данные для первой группы пользователей можно было бы описать при помощи как моделей, рассмотренных в данном разделе, так и активно развивающихся альтернатив [362]. Таким образом, при помощи весьма недорогой модификации способа сбора данных, можно было бы значительно упростить процесс моделирования.

2.3.2 Способ расчёта охватов аудитории информационных ресурсов WWW по данным счетчиков

Как отмечено выше, при помощи счётчика удаётся измерить лишь охват первого порядка $R_1(t)$ в дискретные моменты времени t_0, t_1, \dots, t_n . Так же как и в случае онлайн-панелей, вместо охватов высшего порядка $R_k(t)$ на практике приходится использовать их дискретные аналоги $\tilde{R}_k(n)$. Для их описания используется марковский случайный процесс с дискретными состояниями и дискретным временем (цепь Маркова). Возможные состояния пользователей и переходы между ними изображены при помощи орграфа в верхней части рис.2.25.

В описанной далее модели процесс посещения информационного ресурса каждым из пользователей описывается простейшим потоком событий с параметром λ , а процесс уничтожения cookie – простейшим потоком событий с параметром μ [67]. Пусть R_0 – полное число пользователей, а $v_k(t, \lambda, \mu)$ – это количество актуальных уникальных пользователей (а.у.п.), обладатели которых, которые обращались к информационному ресурсу k дней из t .

Пусть измерения охватов аудитории проведены в моменты $t_k = kT$, $k = 0, 1, \dots, n+1$, тогда будем использовать краткую запись $v_k(n)$ в качестве синонима $v_k(n; \lambda, \mu)$.

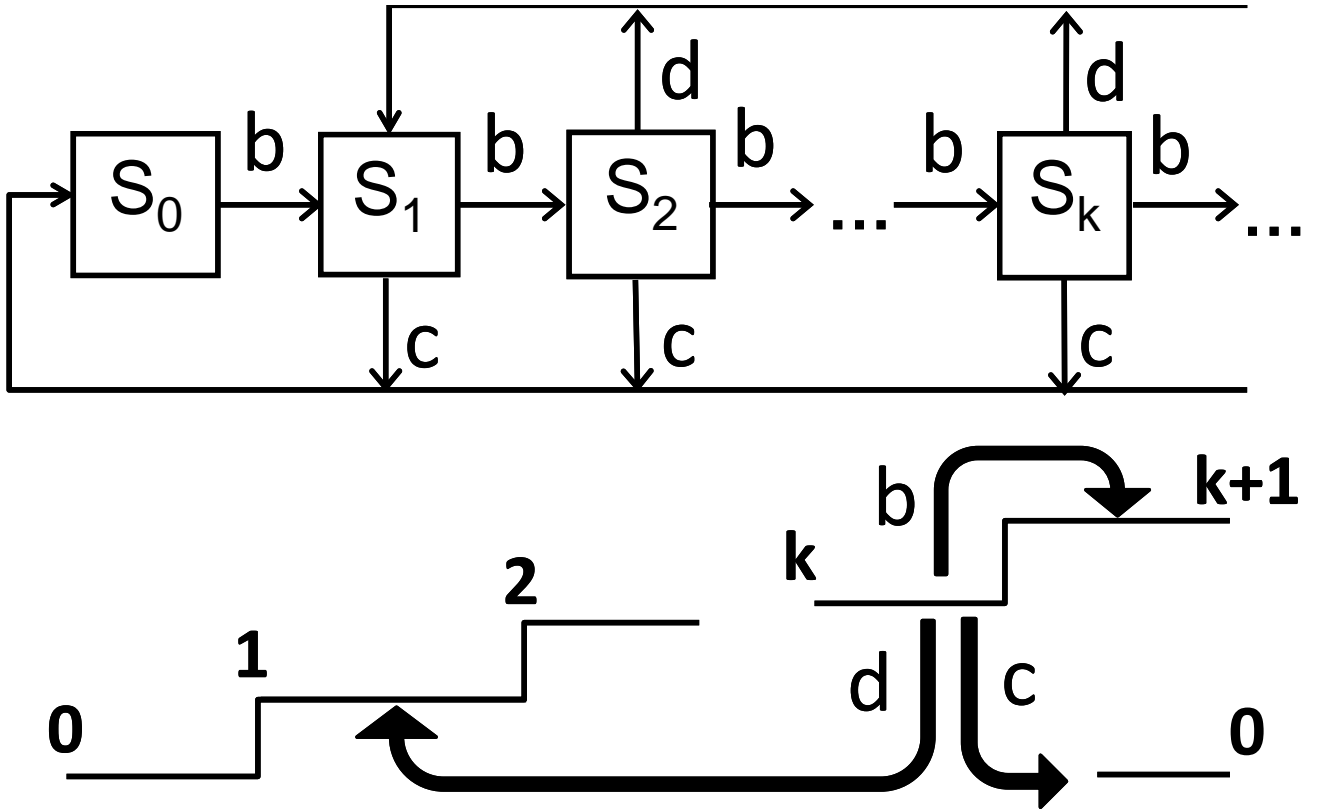


Рис.2.25. Вверху: граф состояний актуальных уникальных пользователей (а.у.п.). Внизу: схема “случайных блужданий” а.у.п.

Нами сформулировано и доказано следующее **Утверждение 1**.

Величины $\hat{v}_k(n)$ удовлетворяют линейным разностным уравнениям

$$\hat{v}_0(n+1) = a\hat{v}_0(n) + c, \quad (22)$$

$$\hat{v}_1(n+1) = b\hat{v}_0(n) + a\hat{v}_1(n) + d,$$

$$\hat{v}_k(n+1) = b\hat{v}_{k-1}(n) + a\hat{v}_k(n) \quad \text{для} \quad 2 \leq k \leq n,$$

$$\hat{v}_{n+1}(n+1) = b\hat{v}_n(n),$$

$$\text{с начальными условиями } \hat{v}_0(0) = 1 \quad \text{и} \quad \hat{v}_k(0) = 0 \quad \text{для} \quad 1 \leq k \leq n, \quad (23)$$

где использованы константы

$$a = e^{-(\lambda+\mu)T}, \quad b = (1 - e^{-\lambda T})e^{-\mu T}, \quad c = \frac{\mu}{\lambda + \mu}(1 - e^{-(\lambda+\mu)T}), \quad d = 1 - a - b - c. \quad (24)$$

Действительно, полную группу событий, которые могут произойти с индивидом за время T , образуют следующие четыре несовместных события:

1) Индивид может не менять cookie и не использовать информационный ресурс. В итоге он останется в прежнем состоянии s_k . Вероятность этого события есть $a = e^{-(\lambda+\mu)T}$, поскольку смена cookie и использование ресурса независимы.

2) Индивид может не менять cookie, но использовать ресурс. Вероятность такого события, то есть условная вероятность перехода из состояния s_k в состояние s_{k+1} есть $b = (1 - e^{-\lambda T})e^{-\mu T}$.

3) Индивид может стереть cookie в промежутке между τ и $\tau + \Delta\tau$, образно говоря, “родиться заново” и с тех пор ни разу не использовать ресурс. Тогда он совершит переход из состояния s_k в состояние s_0 . Рассчитаем вероятность такого перехода. В каждый момент времени вероятность смены cookie индивидом за малый промежуток $\Delta\tau$ составляет $\mu\Delta\tau$. Вероятность того, что произвольный индивид, сменив cookie в промежутке между τ и $\tau + \Delta\tau$, больше её не менял и не использовал ресурс, есть $c = \mu \int_0^T e^{-(\lambda+\mu)\tau} d\tau = \frac{\mu}{\lambda + \mu} (1 - e^{-(\lambda+\mu)T})$.

4) Индивид, стерев cookie (в последний раз) в промежутке между τ и $\tau + \Delta\tau$, может с тех пор воспользоваться информационным ресурсом (один раз или более). В данном случае это означает переход в состояние s_1 . Вероятность перехода из s_k в s_1 получим $d = \mu \int_0^T (1 - e^{-\lambda\tau}) e^{-\mu\tau} d\tau = 1 - c$.

Система (50) даёт компактное описание матрицы переходов марковской цепи, описывающей состояния индивида. В нижней части рис.2.25 переходы индивида изображены в виде “случайных блужданий” между состояниями.

Решения уравнений (22) получены нами при помощи метода производящих функций и выражаются через распределение Паскаля. На практике, однако, вычисления гораздо проще осуществляются непосредственно по формулам (22).

Найдём способ расчёта охватов $\tilde{R}_k(n)$ аудитории. Как и в случае онлайн-панелей (см. раздел 2.2) под охватом k -го порядка подразумевается количество пользователей, зарегистрированных на ресурсе не менее k дней из n .

Для расчёта охватов $\tilde{R}_1(n)$ можно использовать формулу (15) для $t = nT$

$$\tilde{R}_1(n; \lambda, \mu) \equiv R_1(n; \lambda, \mu) = R_0 \frac{\lambda\mu}{\lambda + \mu} nT + R_0 \frac{\lambda^2}{(\lambda + \mu)^2} (1 - \exp(-(\lambda + \mu)nT)),$$

ибо обе величины измеряют количество уникальных пользователей, которые не менее одного раза использовали информационный ресурс за время t .

Охваты $\tilde{R}_k(n)$ при $k \geq 2$ можно рассчитать по рекуррентной формуле

$$\tilde{R}_k(n+1; \lambda, \mu) = \tilde{R}_k(n; \lambda, \mu) + R_0 \cdot \tilde{\lambda} \cdot T \cdot \hat{v}_{k-1}(n; \lambda, \mu), \quad (25)$$

где $\hat{v}_{k-1}(n; \lambda, \mu)$ рассчитываются при помощи системы (22) и введена константа

$$\tilde{\lambda} = \frac{1 - e^{-(\lambda + \mu)T}}{(\lambda + \mu)T} \lambda.$$

(26)

Доказательство. Пусть к началу отрезка времени было $v_{k-1}(n; \lambda, \mu)$ а.у.п., зарегистрированных $k-1$ раз. Найдём, сколько индивидов воспользуются ресурсом (и увеличат \tilde{R}_k) до истечения отрезка времени T . Для этого нужно, чтобы индивид не сменил cookie и не обратился к ресурсу до некоторого момента $0 \leq t \leq T - \Delta t$, а затем обратился к ресурсу в интервале от t до $t + \Delta t$. Вероятность обращения к ресурсу на этом интервале есть $\lambda \Delta t$, откуда находим

$$\tilde{R}_k(n+1; \lambda, \mu) - \tilde{R}_k(n; \lambda, \mu) = R_0 \lambda \int_0^T \exp(-(\lambda + \mu)\theta) d\theta \times \hat{v}_{k-1}(n; \lambda, \mu) = R_0 \lambda \frac{1 - e^{-(\lambda + \mu)T}}{\lambda + \mu} \hat{v}_{k-1}(n; \lambda, \mu).$$

Справедливость формулы (25) доказана.

Нам не удалось получить явные выражения для кумулятивных охватов аудитории при использовании непрерывных распределений λ и μ ¹⁸. Поэтому мы возвращаемся к дискретным распределениям λ , использованным в разделе **2.1.1**, и лишь одним $\mu = \mu_1$. Мы аппроксимируем плотность распределения суммой

$$p(\lambda, \mu) = \sum_{i=1}^M p_i \cdot \delta(\lambda - \lambda_i) \delta(\mu - \mu_1). \quad (27)$$

Тогда расчётные охваты аудитории имеют вид сумм

$$\tilde{R}_k(n) = \sum_{j=1}^J r_j \tilde{R}_k(n; \lambda_j, \mu_1), \quad (28)$$

¹⁸ Мы не исключаем, что явные выражения охватов можно получить, используя не распределения λ и μ , а распределения $v = \lambda + \mu$ и λ/v , следуя подходу, использованному в работе [362].

где, как и прежде, используется обозначение $r_i = R_0 p_i \geq 0$. Частоты λ_j растут в геометрической прогрессии $\lambda_j = \lambda_0 s^j$. Свободными параметрами модели являлись $\mathbf{m} = (r_1, r_2, \dots, r_J)$. Как и в модели С (раздел 2.1), число свободных параметров велико ($J = 15$) и в функцию ошибки добавлено слагаемое, сглаживающее колебания r_i

$$\Phi(\mathbf{m}) = \sum_{n=1}^N \sum_{k=1}^n w_{ki} (d_{kn} - \tilde{R}_k(n; \mathbf{m}))^2 + \sum_{j=1}^{J-1} \varepsilon_j \cdot (r_{j+1} - r_j)^2. \quad (29)$$

Здесь ε_j – неотрицательные сглаживающие множители; n – количество суток; d_{kn} – данные счётчика о k -м охвате за n суток (нам удалось собрать 211 точек); w_{ik} – веса (которые полагались равны единице или нулю в зависимости от наличия или отсутствия данных счётчика); N – максимальное значение n (в нашей работе – от 28 до 31 суток, в зависимости от месяца).

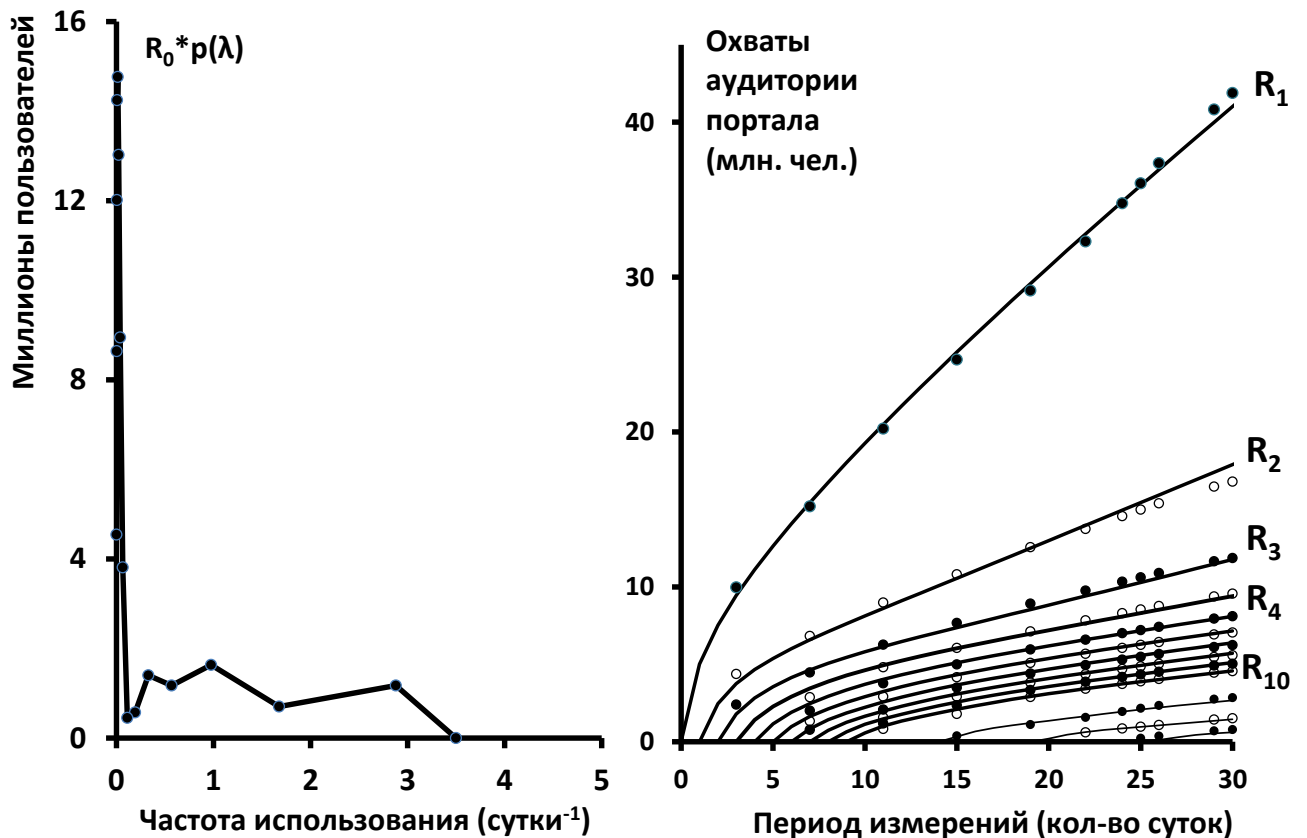


Рис.2.26. Слева – $R \cdot p(\lambda, \mu)$. Справа – расчётные охваты $\tilde{R}_k(n)$ (линии) и данные статистической системы Rambler (кружки) об аудитории крупного Интернет-портала.

На рис.2.26 слева представлена полученная аппроксимация величины $r_i = R_0 p_i$ для крупного Интернет-портала, пропорциональная плотности распределения p_i частоты использования этого информационного ресурса. Здесь и на последующих рисунках точки (λ_i, r_i) соединены отрезками прямых линий, чтобы график имеет форму т.н. полигона, хотя распределение является дискретным, и, более корректным (но и менее удобным для восприятия) было бы представление в виде столбиковой диаграммы. Справа на рисунке показаны данные d_{kn} об охватах, которые были измерены в ноябре 2008 года.

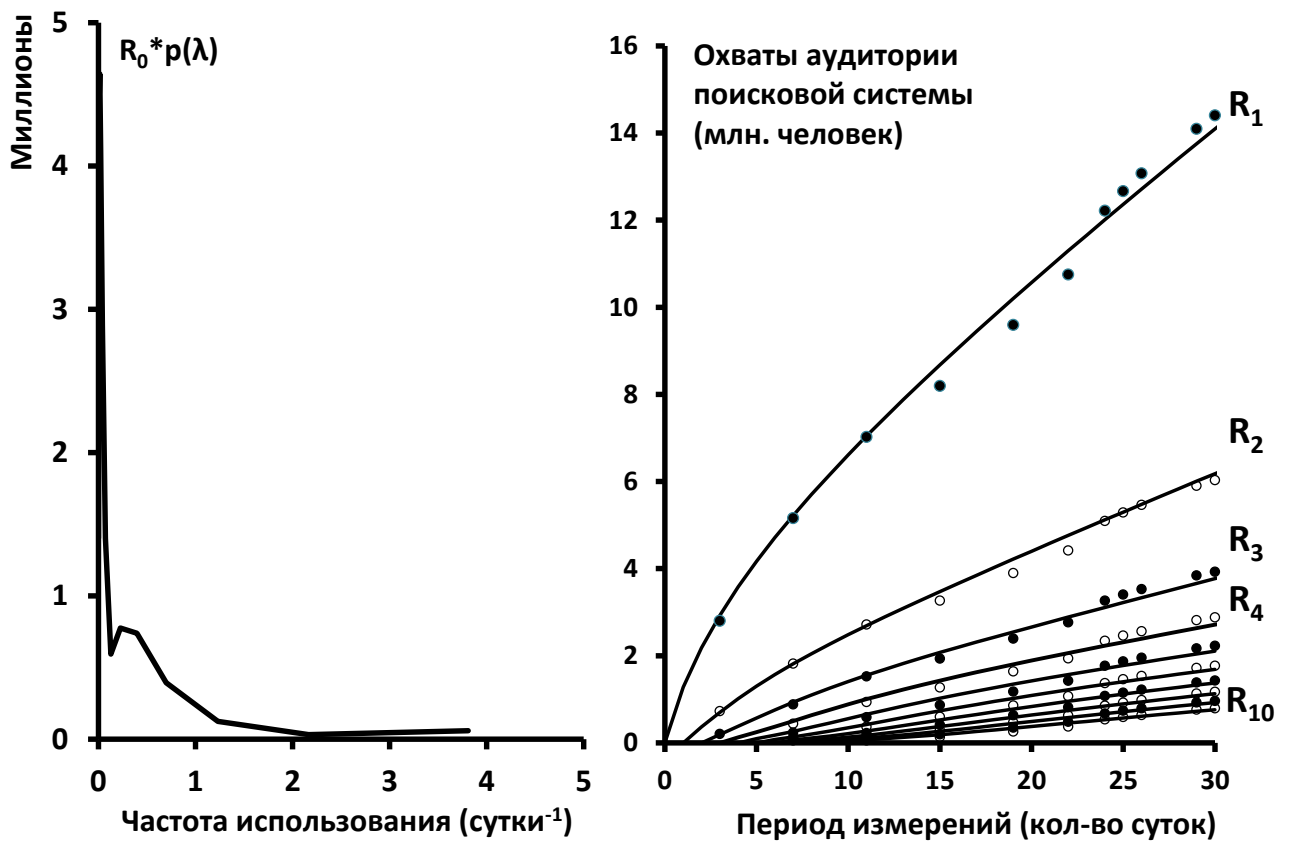


Рис.2.27. Слева – плотность $p(\lambda, \mu)$. Справа – расчётные охваты $\tilde{R}_k(n)$ (линии) и данные статистической системы Rambler (кружки) об аудитории популярной поисковой машины.

Здесь, в отличие от случая онлайн-панелей (раздел 2.2), аудитория не сегментировалась на “рабочую” и “домашнюю”, поскольку нас интересовало

моделирование кумулятивных охватов $\tilde{R}_k(n)$ в условиях смены cookies, а эффект выходных дней, как показали результаты раздела 2.2, не является значительным.

На рис.2.27 в аналогичном формате представлены результаты моделирования и исходные данные для популярной поисковой системы, также относящиеся к ноябрю 2008 года.

Из сравнения рис.2.26 и рис.2.27 можно заключить, что использование поисковой системы является более однородным. В случае портала можно наблюдать с одной стороны – значительный пик малоактивной аудитории, а с другой – существенную долю пользователей, посещающих ресурс чаще одного раза в сутки.

Выводы

1. Разработана модель поведения онлайн-панели в виде смеси бета-биномиальных распределений, которая описывает все кумулятивные кривые охвата аудитории информационного ресурса. Модель применена для аппроксимации кумулятивных кривых крупного портала и популярного онлайн-СМИ.
2. При помощи моделирования данных онлайн-панели обосновано использование вероятностной смеси непрерывных, а не дискретных распределений частоты использования Интернета для прогнозирования суточного, недельного и месячного охватов аудитории.
3. Разработанная в разделе 2.3 модель может быть применена для восстановления $p(\lambda, \mu)$ небольших информационных ресурсов, аудиторию которых не удаётся измерить при помощи панельных исследований, а также для восстановления $p(\lambda, \mu)$ некоммерческих ресурсов, для которых панельные исследования слишком дороги. Следует сделать практическое замечание о том, что для того, чтобы получить оценки плотности распределения частоты использования менее популярных ресурсов, чем исследованные здесь, может потребоваться осреднение данных за несколько месяцев.

ГЛАВА 3. СОВМЕСТНОЕ МОДЕЛИРОВАНИЕ ПАРКА И РЫНКА ТЕХНИЧЕСКИХ СИСТЕМ

Традиционно количество персональных компьютеров (ПК), которые до недавнего времени служили основным устройством доступа к Интернету, рассматривалось, как одно из основных ограничений, препятствующих распространению Интернета¹⁹ [199,262]. Статистические данные, собранные в развитых странах, подтверждали эту точку зрения – парк ПК в этих странах служил (и по сей день является) верхним ограничителем для числа пользователей Интернета. В России, так же как и во многих развивающихся странах, эта эмпирическая закономерность не выполнялась, поскольку ПК, будучи относительно дорогими устройствами, совместно использовались несколькими индивидами; но зачастую применялась при построении прогнозов распространения Интернета. Критикуя указанную практику, автор настоящего исследования ставил задачу построения модели российского парка ПК, которая согласовывалась бы со всеми доступными данными о распространении компьютеров, включая их продажи, наличие в домохозяйствах и в индивидуальном пользовании.

Данные о количестве или доле пользователей нововведения, несмотря на их высокую полезность для моделирования, в настоящее время не являются наиболее распространенным и легко доступным видом статистических данных. Причина состоит в сложности проведения репрезентативных социологических опросов, к тому же, на ранних этапах распространения нововведений в новых отраслях экономики доля пользователей в населении столь незначительна, что не может быть измерена путем массовых опросов. В этот период в новых индустриях складывается традиция измерения рынка при помощи таких показателей, которые

¹⁹ В данном 06.04.2009 г. интервью закрытому сегодня изданию “Газета” генеральный директор крупнейшего российского онлайн-ресурса Mail.Ru, Дмитрий Гришин, называет доступность персональных компьютеров одним из двух главных факторов роста российской Сети: “Главный фактор – это покупательная способность населения приобретать компьютеры и то, как крупные компании и государство будут развивать доступ к интернету в регионах. покупательную способность населения приобретать компьютеры” (см. <http://corp.mail.ru/press/pubs/724>).

могут быть получены оперативно, что позволяет использовать их для текущих задач управления предприятием, а также для оценки стоимости бизнеса. В случае рынка мобильной связи в качестве таких показателей укоренились суммарное количество абонентов (SIM-карточек) и поквартальные продажи телефонных трубок, в случае мобильных приложений – число “скачиваний”, в случае Интернета – количество IP-адресов и файлов “cookies”, в случае широкополосного Интернета – число подключенных квартир, в компьютерной индустрии – объём продаж или поставок компьютеров.

Такие данные отнюдь не бесполезны при изучении распространении нововведений, более того, первые количественные модели диффузии инноваций, такие как модель Ф.Басса, были разработаны именно для объема продаж товаров длительного пользования (в натуральном выражении), а не для количества потребителей таких товаров. При этом в ранних исследованиях предполагалось, что повторные продажи устройств “на замену” вышедшим из строя или морально устаревшим товарам длительного пользования несущественно искажают общую картину. Проблема учета продаж “на замену” была в целом решена в работах Дж.Ольсона и С.Чои [437] и В.Камакуры и Ш.Баласубраманьяна [392], однако примеры применения моделей продаж “на замену” остаются немногочисленными [36,186,187]. Отметим теоретический вклад отечественных исследователей, которым удалось формализовать задачи с переменной во времени функцией дожития технического устройства [53,54,236,264].

В то время как уровень информатизации российских домохозяйств регулярно анализируется в российской научной литературе [19,215], включая и сопоставления с индивидуальным использованием ИКТ [190,189], математическое моделирование с целью прогноза при таких сопоставлениях ранее не привлекалось. Результаты, обсуждаемые в этой главе, опубликованы в наших работах, посвящённых моделированию распространению мобильной связи ([65,82,71,74,70,84,87,95,111,116,138,112,122,118]) и парка компьютеров ([134,113,125,136,133,143,179]). Нами разработана модель роста парка

технических систем с учетом данных исследований о размере рынка и онлайн-опросов о сроке службы устройства, которая применена для оценке размера и прогнозирования российского парка персональных компьютеров.

Наиболее важным нам представляется тот факт, что два вида показателей – о продажах и о парке эксплуатируемых устройств, каждый из которых может содержать случайные и систематические ошибки, могут использоваться совместно для построения более надежных моделей и прогнозов процессов распространения нововведений, чем позволяет каждый из показателей по отдельности. В данной главе попытка такого совместного использования проведена при помощи данных о продажах и парке персональных компьютеров и мобильных телефонов.

Разработку моделей осложняют два факта:

- несколько индивидов, составляющих домохозяйство, могут приобретать и использовать устройство совместно;
- один индивид может использовать несколько устройств (тем более несколько устройств может использовать домохозяйство).

Таким образом, прежде, чем приступать к моделированию парка и рынка технических систем, необходимо иметь в арсенале, во-первых, модели распространения нововведений в домохозяйстве и, во-вторых, модели приобретения дополнительных устройств. Впервые модель, в которой явно учитывались различия в способах распространения информации внутри и вне домохозяйства, построил В.А.Шведовский в 1981 году [290]. В обзоре Б.Ратчфорда и др. [447] рассмотрены способы решения второй задачи на основе классической диффузионной модели, в которой владельцы дополнительных устройств влияют друг на друга. С нашей точки зрения обе задачи можно решать в рамках единого подхода, который изложен в двух последующих разделах на примере распространения мобильных телефонов. Покупка дополнительного устройства при этом рассматривается как покупка устройства для “виртуального” члена домохозяйства.

3.1 Модели распространения мобильной связи среди индивидов и в домохозяйствах

Мобильная телефонная связь, которая лишь в 1981 году начала распространение на массовом рынке, к 2009 году охватила 4,6 миллиарда абонентов. Как отмечают Р.Перес, Э.Мюллер и В.Махайан [442], мобильные телефоны не только чрезвычайно широко распространены, но и наиболее выпукло иллюстрируют многие сложные проблемы моделирования распространения инноваций. В частности, в мобильной телефонии *услуга* связи объединена с *продуктом*, который включает как устройство, так и программное обеспечение. Структура рынка подвижной электросвязи сложна, число игроков рынка регулируется государством, так что в одних регионах рынок является конкурентным, а в других может быть монополизирован. Конкуренция сопровождается ценовыми войнами и интенсивными рекламными кампаниями в традиционных и новых СМИ. Постоянным источником беспокойства компаний является отток пользователей к конкурентам. В течение жизненного цикла мобильной связи возникли несколько поколений технологий, а рынок завоевывали новые стандарты передачи данных. Эволюционировали и управляющие маркетинговые переменные – стоимость минуты разговора, цена телефона, способы и места оплаты трафика, преобладающие каналы распространения, дизайн и функции телефона, аксессуары [442].

3.1.1 Модели с постоянным потенциалом распространения

В этом разделе на основе данных массовых опросов о распространении мобильной связи в России проведено сравнение классической модели распространения нововведения в атомизованном обществе и новой модели, в которой решающей единицей является не индивид, а **домашнее хозяйство**²⁰. В

²⁰ Под *домашним хозяйством* понимается совокупность лиц, проживающих в одном жилом помещении или его части, как связанных, так и не связанных отношениями родства, совместно

отличие от классической, новая модель успешно объясняет три вида данных социологических опросов: о личном владении сотовыми телефонами, об использовании сотовой связи, о наличии сотового телефона в домохозяйстве.

Важно отметить, что область применения нашей модели не ограничивается мобильной связью. При помощи этой модели можно изучать распространение персональных компьютеров, планшетов, цифровых фотоаппаратов, навигаторов, устройств для чтения электронных книг и любых иных нововведений, которые первоначально весьма дороги и сложны и приобретаются для домохозяйства в единственном экземпляре, после чего по мере снижения цен и упрощения приобретаются дополнительные экземпляры для каждого члена домохозяйства или для некоторых его членов. Основные результаты этого раздела опубликованы в наших работах ([65,82,71,74,70,84,87,95,111,116,138,112,122,118]).

Данный раздел ставит и решает следующие задачи

- 1) поставить вопрос о способности модели атомизованного общества (т.е. классической модели распространения нововведений) объяснить результаты комплексных социологических опросов о распространении новых информационных и телекоммуникационных технологий;
- 2) построить модель распространения нововведения в домохозяйствах, формализующую вербальную модель Э.Роджерса распространения мобильной связи [449]
- 3) построить модель распространения нововведения в домохозяйствах, идентифицировать её параметры, позволяющие удовлетворить данным комплексных опросов
- 4) найти целовую эластичность, которая позволит моделировать динамику потенциала российского рынка мобильной связи;
- 5) предложить социологический эксперимент, обработать его данные и проверить предсказания разработанной нами модели.

обеспечивающих себя пищей и всем необходимым для жизни, т.е. полностью или частично объединяющих и расходующих свои средства. Домашнее хозяйство может состоять из одного человека, живущего самостоятельно [52].

Обнаружив отток абонентов в 1999 году, крупные российские операторы сотовой связи приняли решение о выводе услуги мобильной связи на массовый рынок и радикально снизили стоимость минуты разговора. В результате, рост числа абонентов возобновился, а доля владельцев сотовых телефонов среди россиян в 2000 году достигла 2%, что позволило начать изучение динамики распространения мобильной связи путем социологических опросов. Начиная с этого времени, публикуются уникальные в плане разнообразия и полноты данные:

- о личном владении мобильным телефоном (“Левада-Центр”, ФОМ, ВЦИОМ, РОМИР);

- о наличии мобильных телефонов и их числе в домохозяйствах (“Левада-Центр”, ФОМ, ВЦИОМ, компания “Гфк Русь”);

- об использовании мобильной связи взрослыми россиянами (РОМИР, ФОМ, компания “Башкирова и партнеры”).

Таким разнообразием данных мы обязаны отчасти конкуренции исследовательских агентств, которые избегают непосредственных сравнений заказчиками своих услуг и поэтому формулируют вопросы разными способами, а отчасти – отсутствию единого подхода к изучению распространения нововведений при помощи массовых опросов. Помимо неудобства, вызванного невозможностью непосредственно сравнивать данные различных источников, обнаружился важнейший плюс отмеченного разнообразия – данные описывают различные стороны процесса и позволяют построить модели распространения нововведения не только в атомизованном обществе (которое само по себе является чрезмерным упрощением), но в обществе, состоящем из домохозяйств.

Высокая скорость и, соответственно, сравнительно короткое характерное время процесса распространения мобильной связи позволяют использовать для моделирования диффузии этого нововведения сравнительно простые модели, не требующие учета процессов воспроизводства населения, что, как показано в первых двух главах данной работы, оказывается необходимым при исследовании более медленных процессов, таких, как процесс распространения Интернета [301].

Воспользовавшись этим преимуществом, мы построили количественную модель распространения сотовой связи в российском обществе, которая объясняет все три описанных выше вида данных социологических опросов.

Отметим, что собранные аналитическими агентствами данные о динамике стоимости минуты разговора, позволяют также учесть влияние стоимости использования нововведения (которая складывается из тарифов и стоимости сотовых телефонов) на скорость диффузии. Детальное изучение этого вопроса проведено в следующем разделе.

Модель АС: Атомизованное общество с постоянным потенциалом рынка

На основе данных опросов “Левада-Центра” и Фонда “Общественное мнение” (ФОМ) за 2000-2006 годы нами были построены и опубликованы модель и прогноз распространения мобильной связи в России [301]. В этой работе мы ставили целью проиллюстрировать наличие российских данных, которые можно было бы описать при помощи классической модели распространения нововведений Ф.Басса, а также продемонстрировать возможность прогнозирования таких данных на основе классической теории. В качестве показателя уровня распространения мобильной связи была выбрана доля взрослых (старше 16 лет) россиян, *владеющих* личным мобильным телефоном. Источником данных послужили массовые опросы Левада-центра и ФОМ, которые измеряли долю владельцев сотовых телефонов методом в течение нескольких лет, не изменяя методики измерений. В то же время изменения в способе подсчета *абонентов* (активных SIM-карт) операторами мобильной связи бывали весьма значительными. В частности, в январе 2008 года в результате изменения способа учета абонентов крупным оператором число абонентов мобильной связи в России уменьшилось сразу на восемь миллионов [254]. Поскольку предсказать учетную политику операторов мы не в состоянии, мы отказались от попыток построить модели количества абонентов, хотя именно этот показатель наиболее широко использовался как в научной, так и в производственной литературе. Мы

ограничились моделированием доли пользователей и владельцев мобильных телефонов в населении.

Данные социологических опросов, а также данные Росстата о численности абонентов в 1992-2000 годах [253,252] позволили нам провести сравнение двух наиболее популярных моделей распространения нововведений – Ф. Басса и Б. Гомперца [315,408].

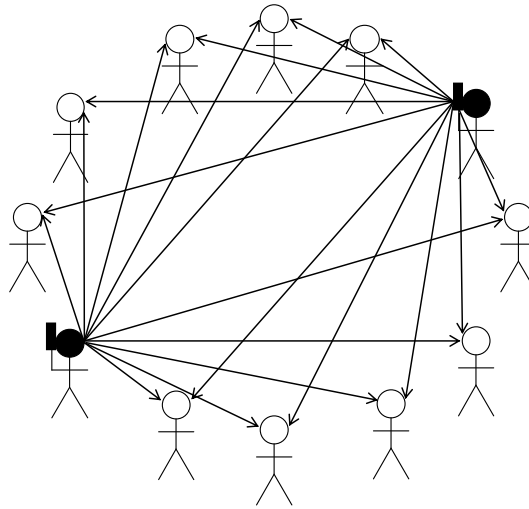


Рис.3.1. Граф влияния владельцев сотовых телефонов на остальных индивидов в модели атомизированной клики.

В модели Ф. Басса общество однородно и представляет собой так называемую *атомизованную клику* (см. рис.3.1), в которой все единицы, принимающие решение (индивиды, домохозяйства, организации), одинаковы, независимы (“атомарны”), и каждый с равной вероятностью воздействует на каждого. На решение субъекта о начале использования нововведения оказывают влияние существующие пользователи и средства массовой информации, причем воздействия складываются. Иными словами, в этой модели два коммуникационных канала, через которые субъект получает информацию, влияющую на его решение; первым является межличностное общение, а вторым – масс-медиа (телевидение, газеты, книги и т.п.). Согласно современным

представлениям первый коммуникационный канал включает не только межличностное общение, но также рыночные сигналы и экстерналии [442].

Модель Ф.Басса может быть описана нелинейным обыкновенным дифференциальным уравнением первого порядка

$$\frac{dX}{dt} = \left(p + q \frac{X}{M} \right) (M - X) \quad (1)$$

с начальным условием $X(0) = N_0 \geq 0$. Здесь $X(t)$ – количество пользователей нововведения, M – потенциал распространения нововведения, q – параметр “внутреннего” воздействия или “имитации”, p – параметр “внешнего” воздействия (СМИ и рекламы).

В.Махайан и Р.Петерсон с целью подчеркнуть наличие двух каналов влияния предложили называть модель, описываемую уравнением (1), моделью смешанных (комбинированных) воздействий (“mixed influence”) [408].

Модель Ф. Басса [315] является частным случаем модели смешанных воздействий, в ней начальное условие нулевое, то есть $X(0) = 0$, а момент начала распространения нововведения предполагается известным.

Отметим, что потенциал распространения нововведения M , то есть полное число людей, способных воспользоваться инновацией, включающее всех текущих и будущих ее пользователей, заранее нам не известен. Потенциал M в общем случае не совпадает со взрослым населением России K , которое, по данным Росстата, составляет 110 933 868 человек.

Отметим также, что в описываемой в этом разделе модели **АС** (а также рассмотренной далее **AV**) величина $X(t)$ – это количество *владельцев* сотовых телефонов. В моделях **АС** и **AV** оно предполагается тождественным количеству *пользователей* мобильной связи. Различия между этими величинами будут рассмотрены при обсуждении модели **ВС**.

Уравнение (1) имеет явное решение в виде логистической кривой

$$X(t) = \frac{1 - p\theta \exp(-(p+q)t)}{1 + q\theta \exp(-(p+q)t)} M, \quad (2)$$

где $\theta = \frac{M - N_0}{pM + qN_0}$ – постоянная, имеющая размерность времени.

Для описания и прогнозирования динамики количества абонентов и пользователей мобильной связи наиболее часто использовались два частных случая.

В первом случае, полагая $N_0 = 0$, получаем модель Ф. Басса:

$$X(t) = \frac{1 - \exp(-(p+q)t)}{1 + \frac{q}{p} \exp(-(p+q)t)} M. \quad (3)$$

Во втором случае, предполагая доминирование межличностных коммуникаций, полагаем интенсивность внешних воздействий $p = 0$ и получаем логистическую модель (модель “внутренних воздействий” [408])

$$X(t) = \frac{M}{1 + \left(\frac{M}{N_0} - 1 \right) \exp(-qt)}. \quad (4)$$

Модель Ф. Басса была применена Райтом и др. [481] для описания динамики числа абонентов мобильной связи в Новой Зеландии. Грубер и Вербовен [374] использовали логистическую модель для исследования распространения мобильной связи в странах Европейского союза. Согласно их выводам, наиболее сильно на процесс диффузии повлияли переход от аналоговых стандартов к цифровым и конкуренция операторов связи. Оба выделенных фактора способствовали снижению цен на услуги мобильной связи и, соответственно, их большей доступности. Кроме того, авторы попытались подтвердить эффект “наверстывания”, т.е. более высокой скорости распространения мобильной связи в тех странах, где процесс начался позже.

Франк [366] использовал логистическую функцию (4) с переменным во времени параметром $q(t)$ для моделирования распространения мобильной связи в Финляндии. Хотя такой подход не является корректным, поскольку указанная функция не удовлетворяет исходному уравнению (1), заслуживает внимания найденная этим автором корреляция между приростом числа абонентов и вариациями темпов прироста среднедушевого ВВП Финляндии.

Вызывают интерес работы, посвященные распространению мобильной связи в странах Южной Европы, где этот процесс не столь сильно опережал российский. В частности, А.Ботельо и Л.Пинто применили логистическую модель для объяснения динамики распространения мобильной связи в Португалии [325], а Х.Михалакелис и др. [426] построили модели динамики сотовой телефонии в Греции при помощи большого количества S-образных кривых. Изучению распространения мобильной связи в различных странах мира посвящены коллективная монография факультета журналистики МГУ им. М.В.Ломоносова [172] и монография М.Кастельса и его соавторов [327].

Изучая эффект “наверстывания”, Джiovанис и Скиадас [367] предложили первую многостадийную модель распространения мобильной связи

$$\frac{dX}{dt} = \left(p + q \frac{X(t-T)}{M} \right) (M - X(t)), \quad (5)$$

в которой между началом использования нововведения и началом влияния на других пользователей существует временной лаг T , аналогичный известному в эпидемиологии “латентному” периоду, когда больной еще не заразен, то есть не способен инфицировать окружающих.

В перечисленных работах использовалась статистика количества абонентов, предоставляемая операторами мобильной связи, Организацией экономического сотрудничества и развития (OECD), Международным телекоммуникационным союзом (ITU), а также публикации государственных статистических служб. Эвристическое применение моделей распространения нововведений в обществе, состоящем из *людей*, к прогнозированию статистики *абонентов* не дало положительных результатов, поскольку один индивид часто обладает несколькими мобильными телефонными номерами, и, наоборот, в развивающихся и бедных странах один сотовый телефон может использовать большое домохозяйство и даже целый поселок. В частности, Грубер и Вербовен [374] определили потенциал проникновения мобильной связи в Европе в 62% численности населения, а прогноз Л. Франка для Финляндии составил 92% [366]. Однако в последующие годы количество абонентов мобильной связи в

исследуемых этими авторами странах превысило численность населения. Сегодня статистические службы европейских стран вместо статистики операторов используют данные панельных исследований и опросов об использовании мобильной связи населением.

Классические трех- и четырехпараметрические модели распространения нововведений, и их расширения, использовались для моделирования и прогнозирования мобильной связи в России и других странах в работах Л. Е. Варакина [31,32], Ю. В. Деарт и соавторов [61,60], Л. С. Тимошенко [272], Т. А. Кузовковой и Л. С. Тимошенко [206], С. Ю. Казанцева и И. Э. Фролова [177,178], А. В. Абилова [1], К. Е. Афанасьевой и В. И. Ширяева [12,11].

В этих работах диффузионные модели применялись к данным операторов мобильной связи о количестве абонентов (активных SIM-карт) и базирующимся на них статистическим данным Росстата. Проблема различия между количеством пользователей и числом абонентов сотовой связи обсуждалась в статье И. С. Березина [24] (см. также статьи С.Казанцева и И.Э.Фролова [177,178]). И.С.Березин отметил, что количество активных SIM-карт не является корректным способом операционализации понятия проникновения мобильной связи, и предложил в качестве альтернативы данные опросов РОМИР о доле населения, использующей мобильную связь.

В этой главе мы не ставим задачу моделирования численности абонентов (активных SIM-карт), но применяем теорию диффузии нововведений для моделирования динамики количества *пользователей* и *владельцев* мобильных телефонов в России. Подчеркнем, что в модели **АС** (и ее расширении на случай переменного потенциала рынка – модели **AV**) указанные две величины совпадают, однако в моделях **ВС** и **BV** они различаются.

Поскольку при моделировании данных опросов исходными данными является доля владельцев мобильных телефонов во взрослом населении страны, проведем в уравнении (3) замену

$$x(t) = \frac{X(t)}{K} \quad \text{и} \quad \mu = \frac{M}{K}$$

и получим

$$x(t) = \frac{1 - \exp(-(p+q)t)}{1 + \frac{q}{p} \exp(-(p+q)t)} \cdot \mu. \quad (6)$$

Задача идентификации модели была поставлена нами следующим образом: найти такие параметры $\hat{\mathbf{m}}$ модели, при которых достигается глобального минимума функционал ошибки (отклонения реальных данных d_k от синтетических $x(t_k)$)

$$\Phi(\hat{\mathbf{m}}) = \min_{\mathbf{m}} \Phi(\mathbf{m}),$$

при условии, что на параметры модели $\mathbf{m} = (\mu, p, q)$ наложены ограничения: $0 \leq \mu \leq 1, p > 0, q > 0$. В качестве синтетических данных использована расчетная доля владельцев телефонов в населении страны.

Мы выбрали функционал ошибки, который соответствует мультипликативной модели с логнормальным распределением ошибок

$$\Phi(\mathbf{m}) = \sum_{i=0}^K w_i \left| \frac{1}{\sigma_i} \ln \left(\frac{d_i}{\tilde{d}_i(\mathbf{m})} \right) \right|^2,$$

где $\tilde{d}_i(\mathbf{m}) = x(t_i)$ – синтетические (расчетные) данные; d_i – данные социологических опросов; σ_i – стандартные ошибки; w_i – весовые коэффициенты (мы использовали веса, убывающие в геометрической прогрессии в направлении убывания оси времени).

В качестве начального момента распространения мобильной связи в России мы выбрали сентябрь 1991 года. Известно, что первый пользовательский мобильный звонок в России совершил 9 сентября 1991 г. мэр Санкт-Петербурга А. А. Собчак, который позвонил в США.

Поскольку в процессе распространения мобильной связи в России можно выделить две фазы – до и после 1998 года, – правомерно было бы построить две модели, отдельно для каждой фазы. Вторым аргументом в пользу построения двух моделей является характер данных: период до 1998 года покрывают только данные Росстата о количестве абонентов, и лишь с 2000 года появились данные социологических опросов. Однако та модель, которую мы построили по данным

опросов, удовлетворительно аппроксимирует и исторические данные за 1992-1999 годы, поэтому мы ограничимся одной моделью.

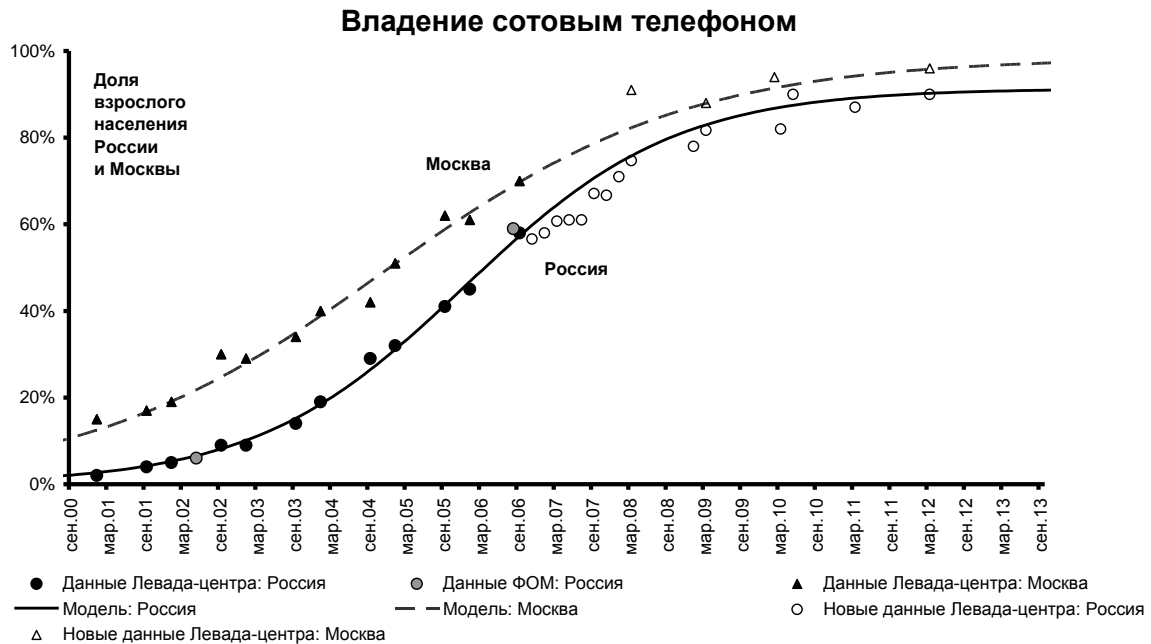


Рис.3.2. Динамика наличия личных сотовых телефонов в Москве и в России. Открытыми символами показаны данные опросов, проведенных после построения прогноза.

В работе [301] Э.А.Юриной и автором этой работы получены следующие параметры модели, минимизирующие функционал ошибки: $\mu = 0,91$, $p = 1,9 \cdot 10^{-5}$ лет⁻¹, $q = 0,71$ лет¹. Как показывает рис.3.2, модель Басса удовлетворительно описывает исторические данные об использовании мобильной связи в России и в Москве в 2000-2006 годах²¹. На рис.3.2 представлены также новые данные за 2007-2012 годы, опубликованные “Левада-Центром” и позволяющие визуально оценить точность прогноза.

Однако такая модель объясняет не все данные массовых опросов. В частности, обнаруживаются систематические различия между долей пользователей и владельцев сотовых телефонов в населении. Результаты расчетов

²¹ Прогноз использовал данные за 2000-2006 г. и был опубликован в 2008 году. Результаты впервые были доложены автором на конференции “Социология инноватики” в декабре 2007 года, см. [75]. В этой работе параметры модели немного отличались (например, $\mu = 0,89$), поскольку в качестве момента начала распространения нововведения был выбран 1997 год.

доли владельцев телефонов среди взрослых россиян $x(t)$ на рис.3.3 показаны совместно с данными опросов “Левада-Центра”, ФОМ, ВЦИОМ, компаний ГФК, РОМИР, “Башкирова и партнеры”. Доля взрослых россиян, лично владеющих сотовыми телефонами, изображена сплошными кружками. Треугольниками показана доля взрослых россиян, у которых мобильный телефон есть в домохозяйстве, а крестиками – доля заявивших, что пользуются сотовой связью (респондентов не спрашивали, личный или чужой телефон они используют).

Рис.3.3 показывает, что доля россиян, у которых мобильный телефон есть в домохозяйстве, выше, чем доля имеющих мобильный телефон в личном пользовании [301]. Это наблюдение представляется тривиальным, однако не учитывается в модели атомизованного общества. Ни модель Ф. Басса, ни общая модель смешанного воздействия [408], ни модель Гомперца, к описанию которой мы переходим, не предоставляют инструментария для совместного моделирования количества пользователей и владельцев нововведения.

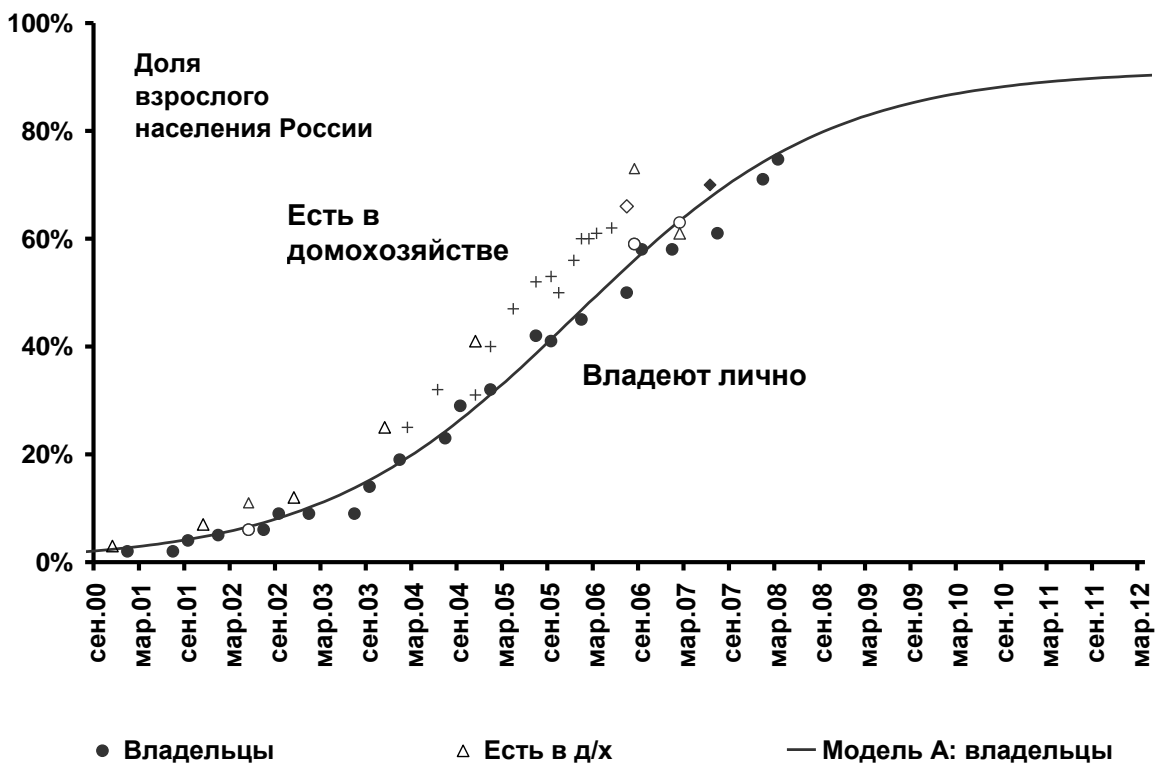


Рис.3.3. Динамика наличия сотового телефона в домохозяйстве и в личном пользовании [301]

Модель Гомперца

Популярной моделью распространения нововведений является модель Гомперца [348,408,421,426,472,473], которую можно описать следующим дифференциальным уравнением

$$\frac{dX}{dt} = bX \ln\left(\frac{M}{X}\right), \quad (7)$$

решение которого имеет вид

$$X(t) = M \exp[-\exp(-b(t - \tau))]. \quad (8)$$

В модели Гомперца три параметра, так же, как и в модели Басса. Параметр M описывает потенциал распространения нововведения, b характеризует скорость роста, а τ является координатой точки перегиба ($X''(t = \tau) = 0$).

Для распространения сотовой связи в России получены следующие параметры модели Гомперца, минимизирующие функционал ошибки: $\mu = 1,0$, $b = 0,384$, $\tau = (8/03/2005)$.

Модель Гомперца является эвристической, и попытки ее применения к процессам в неоднородном обществе автору неизвестны. В то же время, модель Гомперца весьма точно описывает данные Международного союза электросвязи (ITU) о числе абонентов мобильной связи в США на протяжении 20 лет [473]. Кроме того, как отмечено в ряде работ [421,472,473], на практике прогнозы, полученные при помощи модели Гомперца, зачастую оказываются точнее.

В дальнейших расчетах мы используем модель Ф. Басса, однако практически те же результаты для периода 2000-2008 гг. могли бы получить и для модели Гомперца. Выводы, полученные в следующем разделе, справедливы для обеих моделей.

Модель распространения нововведения в домохозяйствах

Мы намерены показать, что модель атомизованного общества не способна одновременно объяснить динамику двух видов статистических показателей: доли россиян, владеющих мобильными телефонами, и доли домохозяйств, в которых есть такие устройства.

Мы предлагаем альтернативную модель **В**, которая объясняет оба вида показателей. Модель **В** формализует наблюдение Э.Роджерса [449], который отметил, что первоначально сотовый телефон приобретался домохозяйством для совместного использования несколькими членами, и только потом наступила эпоха личных, индивидуальных сотовых телефонов.

Для построения нужной нам количественной модели требуются данные о распределении домохозяйств по размеру. Возьмем домохозяйство, в котором живет j взрослых людей. Упорядочим этих людей по времени приобретения личного сотового телефона. Полные численности взрослых граждан, проживающих в домохозяйствах из j взрослых, которые стоят i -ми в гипотетической семейной очереди на личный сотовый телефон, обозначим K_{ij} . Поскольку Росстат публикует данные о домохозяйствах, численность взрослых в которых не превышает шести, условимся считать, что в домохозяйствах не бывает более шести взрослых. Таким образом, верхнюю границу индекса j обозначим J и положим равной 6.

Таблица 3.1

Доли взрослых членов домохозяйств k_{ij} (%)

Число членов домохозяйства	Порядковый номер члена домохозяйства					
	1	2	3	4	5	6
1	13,0					
2	21,4	21,4				
3	9,0	9,0	9,0			
4	3,2	3,2	3,2	3,2		
5	0,6	0,6	0,6	0,6	0,6	
6	0,2	0,2	0,2	0,2	0,2	0,2

Разделив K_{ij} на K , получим долю $k_{ij} = \frac{K_{ij}}{K}$ жильцов с номером i , проживающих в домохозяйствах размера j . Эти величины опубликованы

Росстатом и показаны нами в таблице 3.1. Сумма всех элементов табл. 3.1 равна единице (значения элементов даны в процентах).

Для построения количественной модели **В**, объясняющей все описанные ранее данные социологических опросов, используем четыре допущения:

1. Условный “глава домохозяйства” первоначально приобретает новинку для себя, становится ее официальным владельцем.

2. Период времени между приобретением домохозяйством мобильного терминала и началом использования этого устройства всеми взрослыми членами домохозяйства пренебрежимо мал. Это предположение позволяет считать эквивалентными данные об использовании новинки и о наличии новинки в домохозяйстве.

3. На решение взрослого россиянина о покупке нововведения влияет доля других *пользователей* среди взрослого населения, которую он оценивает в процессе общения (рис.3.4).

4. Пользователи-дети не влияют на решения взрослых об использовании нововведения.

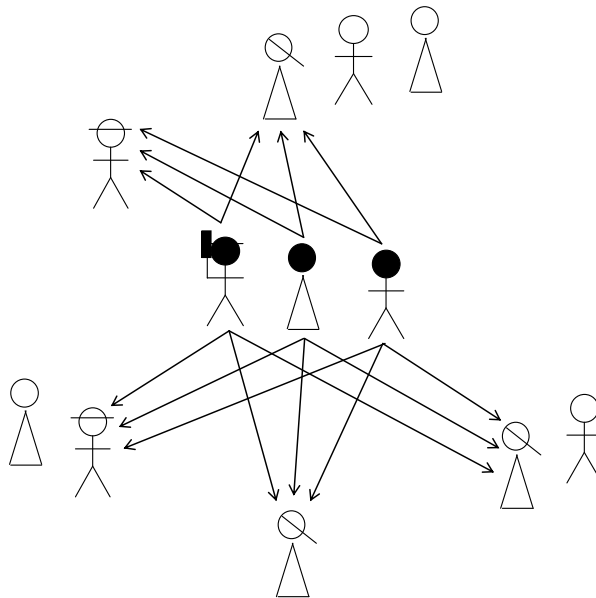


Рис.3.4. Влияние пользователей сотовых телефонов на глав домохозяйств в модели **В**. Глав домохозяйств можно распознать по “шляпкам”.

Необходимость в последнем допущении вызвана неполнотой опубликованных данных опросов, которые изначально не были предназначены для обеспечения моделирования распространения инноваций. Как показывают некоторые исследования, наличие ребенка в семье часто стимулировало покупку сотового телефона. Однако у нас нет достаточных данных об использовании сотовых телефонов российскими детьми.

Предполагается, что позже, по мере накопления средств, домохозяйство приобретает мобильные устройства и для других жильцов. Возникает вопрос, не является ли более правдоподобным сценарий, когда глава домохозяйства по мере появления денег приобретает новые устройства для себя, а поддержанную трубку дарит другому члену семьи. Такое поведение чрезвычайно распространено и подтверждается онлайн-опросами. Однако математические модели обоих сценариев совпадают. При использовании тех показателей, которые нам сегодня доступны, эти два сценария неразличимы. Подробнее этот вопрос обсуждается в конце этого раздела, там же приведены и данные онлайн-опросов.

Рис.3.4 иллюстрирует структуру межличностных коммуникаций в модели **В**. Отметим, что в нашей работе [138] было проведено моделирование возможных способов влияния домохозяйств друг на друга, позволяющее учесть более высокую информированность крупных домохозяйств, однако для апробации этих моделей у нас недостаточно данных.

Рассмотрим домохозяйство, включающее трех взрослых (и произвольное число несовершеннолетних). Примером домохозяйства из трех взрослых (старше 18 лет) людей могут служить совместно проживающие жена, муж и совершеннолетняя дочь. Добавив свекровь, получим домохозяйство из четырех взрослых, добавив совершеннолетнего сына – из пяти.

Упорядочим всех членов домохозяйств, включающих трех взрослых людей, по времени приобретения ими (или для них) личного мобильного телефона.

Обозначим численность обладающих личными мобильными телефонами глав домохозяйств из трех взрослых – X_{31} , вторых взрослых членов домохозяйств

из трех взрослых – X_{32} ; и, соответственно, третьих взрослых членов таких домохозяйств – X_{33} .

Таким образом, первый индекс обозначает количество взрослых в домохозяйстве, а второй – очередность приобретения взрослым человеком личного мобильного телефона по сравнению с другими взрослыми членами домохозяйства

Аналогично поступим с домохозяйствами других размеров. В общем случае численность владеющих личными сотовыми телефонами j -х членов домохозяйств, включающих i взрослых, обозначим X_{ij} .

Не все взрослые члены домохозяйств владеют сотовыми телефонами, их численность в каждом сегменте удовлетворяет ограничению $X_{ij}(t) \leq K_{ij}$. Уточним разницу между K_{ij} и X_{ij} . Величина X_{ij} включает только тех взрослых, у которых телефон уже есть, а K_{ij} – и тех, у которых телефон однажды будет, и даже тех, у кого телефона нет и не будет, но кто стоит в воображаемой семейной очереди i -м.

Опишем процесс приобретения первого мобильного телефона домохозяйством (или его “главным” представителем) общей моделью распространения нововведений [408]

$$\frac{dX_{i1}}{dt} = h(t, U) \cdot (M_{i1} - X_{i1}(t)). \quad (9)$$

Здесь $U(t)$ – полное количество взрослых пользователей сотовых телефонов,

M_{i1} – потенциал распространения нововведения в домохозяйствах численностью i (количество домохозяйств, которые рано или поздно начнут использовать сотовый телефон, и которые состоят из i членов).

Функция $h(t, U)$ носит название “функция риска” и характеризует вероятность приобретения нововведения тем, у кого его еще нет. В зависимости от выбора функции риска, уравнение (9) превращается в уравнение модели смешанных воздействий (1), Гомперца (7) и т.п.

Мы постулируем, что функция риска $h(t, U)$

1. Не зависит от размера домохозяйства, т.е. не выделяем отдельных $h_i(t,U)$, но используем единую $h(t,U)$;

2. функционально зависит от текущего количества пользователей сотовой связи $U(t)$ (а не, например, числа владельцев телефонов, как в модели **A**, не числа самих трубок, и не числа домохозяйств, в которых есть телефоны, и т.д.).

Если дополнительно предположить, что все взрослые члены домохозяйства начинают использовать мобильную связь одновременно с покупкой первого (общего) сотового телефона, то, как показано в последнем разделе этой главы, для численности *пользователей* сотовой связи справедливо уравнение

$$\frac{dU}{dt} = h(t,U)(M - U), \quad (10)$$

а численности обладающих сотовыми телефонами глав домохозяйств размером i , есть линейные функции от числа пользователей

$$X_{i1}(t) = a_i + b_i U(t), \quad (11)$$

где

$$a_i = \frac{X_{i1}(0)M - U(0)M_{i1}}{M - U(0)}, \quad b_i = \frac{M_{i1} - X_{i1}(0)}{M - U(0)}. \quad (12)$$

Вывод уравнений (10)-(12) дан в нашей монографии [111] и не приводится здесь из-за значительного объёма.

Если в качестве функции риска выбрать линейную функцию от $U(t)$, то, как показывает уравнение (10), в отличие от модели **A**, в модели **B** смещенная логистическая кривая описывает не численность *владельцев* телефона, а численность *пользователей* мобильной связи.

Предположим, что домохозяйства приобретают личные сотовые телефоны для второго, третьего и прочих взрослых не сразу, а с некоторой, распределенной во времени, задержкой $g_{ij}(t)$. Мы предполагаем, что форма распределения задержки остаётся неизменной на протяжении всего исследуемого периода времени. Тогда численность обладающих личными мобильными телефонами j -х

взрослых членов домохозяйств, где проживает i взрослых, можно описать сверткой во временной области:

$$X_{ij}(t) = g_{ij}(t) \otimes X_{i1}(t). \quad (13)$$

В частности, если плотности распределения задержки выбраны в виде δ -импульсов Дирака

$$g_{ij}(t) = \beta_{ij} \delta(t - \tau_{ij}) \quad (14)$$

(это самый простой для вычислений способ), то покупка сотового телефона для j -го члена домохозяйства отстает по времени от покупки телефона для главы домохозяйства на время задержки τ_{ij} ($\tau_{ij} \geq 0$):

$$X_{ij}(t) = \beta_{ij} X_{i1}(t - \tau_{ij}), \quad (15)$$

где $0 \leq \beta_{ij} \leq 1$.

Будем использовать β_{ij} и τ_{ij} следующего вида:

1. Для $j \geq 2$ задержка пропорциональна порядковому номеру взрослого члена домохозяйства

$$\tau_{ij} = (j-1)\tau. \quad (16)$$

2. Для $j \geq i \geq 2$ множитель для j -х членов домохозяйства составляет

$$\beta_{ij} = \gamma \beta^{j-2} \quad (17)$$

с ограничениями: $0 \leq \beta \leq 1$, $0 \leq \gamma \leq 1$.

Такая модель охватывает и случаи, когда телефоны приобретаются не для всех взрослых. В частности, при $\gamma = 0$ личными телефонами обзаведутся только главы домохозяйств, а при $\beta \rightarrow 0$ – только первый (глава) и второй члены домохозяйства.

Выражение для численности владельцев аппаратов приобретает вид

$$X_{ij}(t) = \gamma \beta^{j-2} X_{i1}(t - (j-1)\tau) \quad (18)$$

Полное число владельцев сотовых телефонов в момент времени t :

$$X(t) = \sum_{i=1}^J \sum_{j=1}^i X_{ij}(t). \quad (19)$$

Полученные формулы достаточны для расчета теоретических значений всех величин, которые собраны в ходе опросов, но содержат довольно много неизвестных параметров. С целью сократить количество неизвестных параметров, введем несколько дополнительных предположений.

Во-первых, сократим количество неизвестных потенциалов. Предположим, что потенциалы распространения мобильных телефонов среди глав домохозяйств пропорциональны численности таких домохозяйств:

$$M_{i1} = c \cdot K_{i1}. \quad (20)$$

Иными словами, мы выдвигаем предположение, что доля домохозяйств, в которых уже есть или однажды появится сотовый телефон, не зависит от численности домохозяйств. Вообще говоря, считается, что в крупных домохозяйствах (за исключением очень бедных семей) обычно больше новинок, а сильнее всех в среднем отстают в освоении нововведений одинокие люди. Однако, во-первых, это означает скорее различие в параметрах скорости распространения q , чем в потенциалах, а во-вторых, в нашем распоряжении нет необходимых для моделирования таких различий российских данных. По этой причине мы и сокращаем число параметров.

Умножая обе части последнего равенства на i и суммируя, убеждаемся, что

$$M = \sum_{i=1}^J iM_{i1} = c \cdot \sum_{i=1}^J iK_{i1} = cK, \text{ откуда } c = \frac{M}{K} \text{ и, наконец,}$$

$$M_{i1} = \frac{K_{i1}}{K} M = k_{i1} M. \quad (21)$$

Величины k_{i1} возьмем из таблицы 3.1, таким образом, вместо шести неизвестных M_{i1} остался один неизвестный параметр M .

Во-вторых, сократим аналогичным способом количество независимых начальных условий. Потребуем, чтобы для всех i

$$X_{i1}(0) = \frac{M_{i1}}{M} U(0), \quad (22)$$

иными словами, чтобы и в начальный момент времени наличие телефонов у глав домохозяйств не зависело от размера домохозяйств.

С учетом предыдущего предположения (20), получаем

$$X_{i1}(0) = k_{i1}U(0). \quad (23)$$

Теперь вместо шести $X_{i1}(0)$ неизвестным остался только $U(0)$.

Тогда

$$a_i = 0, \quad b_i = \frac{M_{i1} - k_{i1}U(0)}{M - U(0)}$$

и

$$X_{i1}(t) = b_i U(t). \quad (24)$$

Настало время выбрать конкретный вид S-образной кривой для моделирования $U(t)$. Выберем линейную модель функции риска (модель Ф.Басса):

$$h(t, U) = p + q \frac{U}{M}, \quad (25)$$

и, кроме того, учитывая, что до сентября 1991 года сотовой связи в России не было, используем нулевое начальное условие

$$U(0) = 0.$$

Тогда уравнение (10) превращается в уравнение (1), а $U(t)$ выражается при помощи логистической функции (являющейся решением уравнения (4))

$$U(t) = \frac{1 - \exp[-(p+q) \cdot t]}{1 + \frac{q}{p} \exp[-(p+q) \cdot t]} \cdot M \quad (26)$$

Вследствие такого выбора функции риска и начальных условий получаем $b_i = k_{i1}$, $X_{i1}(t) = k_{i1}U(t)$ и окончательную формулу для количества владельцев сотовых телефонов

$$X(t) = \lambda_1 U(t) + \gamma \sum_{j=2}^J \beta^{j-2} \lambda_j U(t - (j-1)\tau). \quad (27)$$

$$\text{где } \lambda_j = \sum_{i=j}^J k_{i1}. \quad (28)$$

Такое число взрослых людей на вопрос “*Есть ли у вас лично мобильный сотовый телефон?*” может искренне дать утвердительный ответ.

Оптимальные параметры моделей

Модель	μ (%)	q (лет ⁻¹)	p (лет ⁻¹)	τ (лет)	β	γ
В	85,8	0,80	$1,8 \cdot 10^{-5}$	1,09	1,0	1,0
А	91,3	0,71	$2,9 \cdot 10^{-5}$	-	-	-

Наилучшим образом (в предположении логнормального распределения ошибок) данным опросов удовлетворяют параметры модели **В**, указанные в таблице 3.2. В этой же таблице для сравнения приведены оптимальные параметры модели **А**.

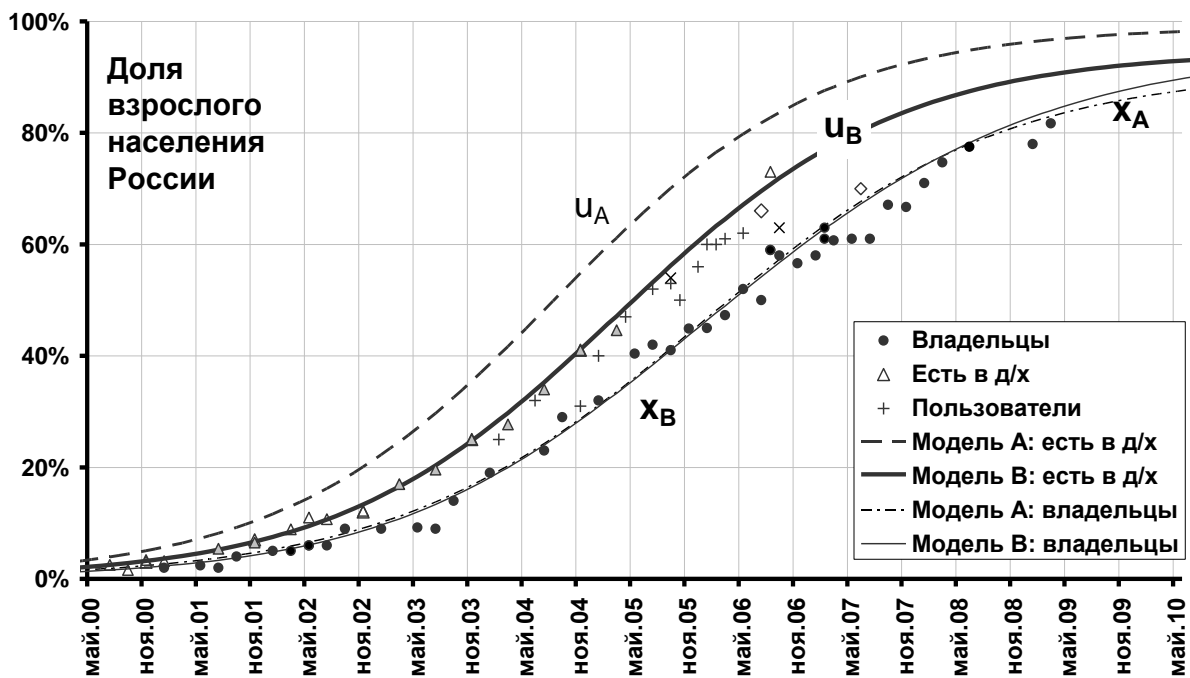


Рис.3.5. Сравнение моделей А и В

На рис.3.5 вместе с данными социологических опросов представлены результаты расчетов по модели **В**:

доли взрослых владельцев сотовых телефонов $x_B(t) = X(t)/K$ (сплошная тонкая линия);

доли взрослых пользователей мобильной связи $u_B(t) = U(t)/K$ (сплошная жирная линия);

доли респондентов, отвечающих, что в домохозяйстве есть хотя бы один сотовый телефон, которая также равна $u_B(t)$.

Здесь использованы данные ФОМ, ВЦИОМ, “Левада-Центра”, “Ромир”, компаний “ГФК Русь” и “Башкирова и партнеры”.

Синтетическая кривая $u_A(t)$, рассчитанная по модели **A** при помощи формул комбинаторики (см. таблицу 3.3), показана на рис.3.5 пунктирной линией. При расчете $u_A(t)$ тоже предполагалось, что как только какой-нибудь член домохозяйства приобретает сотовый телефон, пользоваться им начинает все домохозяйство. Подчеркнём, что в моделях **A** и **B** величины $u(t)$ и $w(t)$ (доля респондентов, подтверждающих, что в домохозяйстве есть сотовый телефон) предполагаются тождественными.

Однако в модели **A** все члены домохозяйства принимают решение о приобретении телефона независимо, другие члены домохозяйства влияют на это решение не больше, чем произвольный посторонний. На рис.3.5 эта независимость проявляется в значительном превышении $u_A(t)$ над $x_A(t)$. В атомизованном обществе многие респонденты отвечали бы, что в домохозяйстве у них имеется сотовый телефон, но лично у респондента мобильного нет, ибо наличие телефона в домохозяйстве не влияет на собственное решение респондента (точнее влияет не более, чем наличие телефона у случайного встречного – а с членами своего домохозяйства в атомизованном обществе видятся не чаще, чем со случайными встречными). Этот эффект напрямую проявляется в значительном превышении пунктирной кривой над сплошной на рис.3.5 (см. аналогичный эффект на рис.3.6(б) и рис.3.7). Он, однако, вызван спецификой мобильного телефона, который, как отметил Э.Роджерс [449], проникает в домохозяйство как общий для всех инструмент, но позже становится личным. Другие устройства (например, ADSL-модем, Wi-Fi роутер, принтер, сканер, копировальная машина, холодильник, стиральная машина, швейная машинка) могут не приобретаться в личное пользование, в таком случае взаимное расположение пунктирной и сплошной кривой на рисунках будет иным. При этом

взаимное положение $u_A(t)$ и $x_A(t)$ всегда будет одинаковым, а взаимное положение $u_B(t)$ и $x_B(t)$ может отличаться.

Рис.3.5 демонстрирует, что в то время как $x_A(t)$ и $x_B(t)$ почти совпадают, кривые $u_A(t)$ и $u_B(t)$ отличаются весьма заметно. Модель **A** предсказывает гораздо более высокий уровень проникновения сотовой связи в домохозяйствах, чем имеющий место в реальности. По-видимому, общество не является атомизованным, и наличие в домохозяйстве хотя бы одного сотового телефона существенно влияет на решение о покупке личных аппаратов для остальных членов домохозяйств.

Как иллюстрирует рис.3.5, модель **B** удовлетворительно объясняет все три вида данных: о личном владении сотовым телефоном, о наличии его в домохозяйстве и об использовании мобильной связи.

Таблица 3.3

Формулы расчета вероятностей различных сценариев в моделях А и В

Событие	Символ	Модель A	Модель B
1.	$x(t)$	$x_A(t) = X_A(t) / K$	$x_B(t) = X_B(t) / K$
2.	$x_{all}(t)$	$\sum_{i=1}^J ik_{i1} x_A^i$	$\sum_{i=1}^J ik_{i1} x_{ii,B}$
3.	$x_{some}(t)$	$u_A - x_{all,A}$	$u_B - x_{all,B}$
4.	$x_{none}(t)$	$\sum_{i=1}^J ik_{i1} (1 - x_A)^i$	$1 - u_B$
5.	$x_+(t)$	$u_A = 1 - x_{none,A}$	u_B

Дополнительные вопросы респондентам

В феврале 2007 г. специалисты Фонда “Общественное мнение” задали респондентам и более сложные вопросы, охватывающие пять событий:

1. Респондент владеет личным мобильным телефоном.
2. Все члены домохозяйства владеют личным мобильным телефоном.
3. Некоторые, но не все члены д/х-ва владеют личным мобильным телефоном.

4. Ни один член домохозяйства не владеет личным мобильным телефоном.
5. В домохозяйстве есть хотя бы один мобильный телефон.

В таблице 3.3. приведены формулы расчета вероятностей этих ответов для модели **A** и модели **B**.

Статистику ответов на первый и пятый вопрос мы подробно исследовали выше. Ответы на вопросы 2, 3, 4 были собраны Фондом “Общественное мнение” в ходе опроса 15 февраля 2007 г. Эти данные вместе с соответствующими теоретическими величинами, указанными в таблице 3.3, представлены на рис.3.6. В этой таблице используются обозначения $x_A(t) = X_A(t)/K$, $x_B(t) = X_B(t)/K$, $x_{ij,A}(t) = X_{ij,A}(t)/K$, $u_A(t) = U_A(t)/K$ и т.п. Пунктирными линиями показаны синтетические кривые $x_{none}(t)$, $x_{some}(t)$, $x_{all}(t)$, соответствующие модели **A**, сплошными линиями – аналогичные кривые, соответствующие модели **B**.

При расчете параметров моделей данные этого опроса учтены не были, поэтому их можно считать контрольными. Рис.3.6 также свидетельствует в пользу модели **B**.

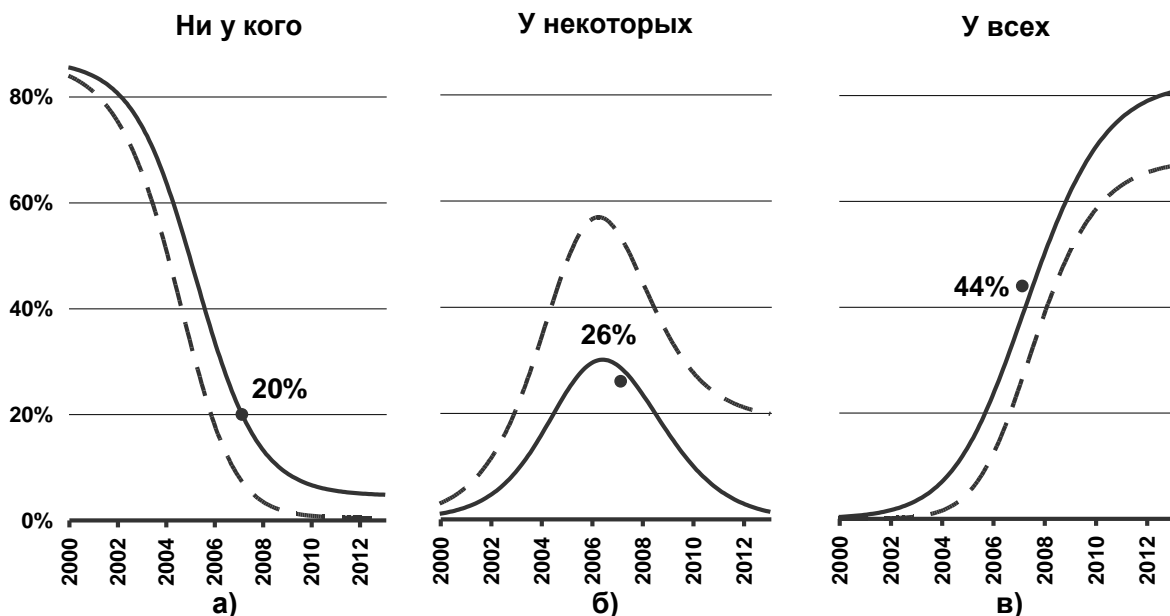


Рис.3.6. Наличие мобильных телефонов у членов домохозяйства: а – ни у кого; б – у некоторых; в – у всех

Мы заключаем, что модель **В**, в которой решающей единицей является домохозяйство, более полно описывает процесс распространения мобильной связи среди россиян, чем классическая модель **А**. Модель **В** способна одновременно объяснить несколько видов данных, в то время как модель **А** – только один.

Анализ данных опросов ФОМ 2002, 2006 и 2011 гг.

Данные другого опроса ФОМ, который проводился в 2002 и 2006 гг. (рис.3.7, закрашенные кружки), тоже лучше согласуются с моделью **В**. Сравнение сплошных и пунктирных кривых на рис.3.6 и рис.3.7 позволяет предположить, что дополнительный опрос, проведенный в 2010-2012 годах, смог бы предоставить чрезвычайно важные данные для обоснования выбора наиболее адекватной модели.

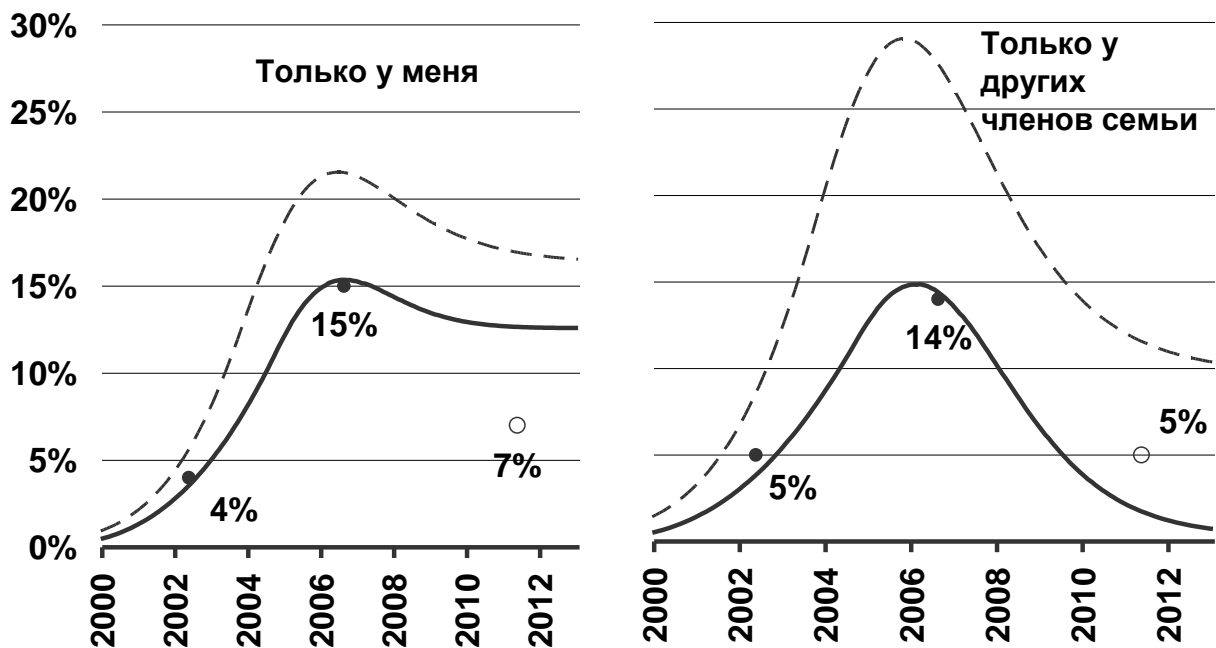


Рис.3.7. Наличие мобильных телефонов у респондента и у членов его семьи: а – только у респондента; б – только у других членов семьи. Сплошная линия – предсказания модели **В**, пунктирная – модели **А**.

По просьбе автора данной работы Фонд “Общественное мнение” включил вопросы, соответствующие рис.3.7, в опрос населения об использовании

мобильной связи и мобильного Интернета, который был проведён в конце мая – начале июня 2011 года. Результаты изображены на рис.3.7. незакрашенными кружками и располагаются заметно ближе к предсказаниям модели **В**.

Замечание 1.

Отметим, что в наших работах [95,111] рис.3.7(б) содержал ошибку: пунктирная кривая, соответствующая $x_{other}(t)$ была рассчитана с использованием $x_B(t)$ вместо $x_A(t)$. Расхождение между ошибочной и истинной кривыми незначительно, поскольку $x_B(t)$ и $x_A(t)$ очень близки.

Замечание 2.

Важно отметить, что в случае модели **А** выражения для всех изучаемых величин ($x_{all}(t)$, $x_{some}(t)$ и т.п.) являются функциями $x_A(t)$ и не включают каких-либо параметров диффузионной модели (p , q и т.п.), хотя включают параметры структуры домохозяйств k_{i1} . Это означает, что, если предполагается, что общество атомизовано, то эти величины на практике определяются наблюдаемыми значениями d_i , и не зависят от того, какая именно модель распространения нововведений в атомизованном обществе нами выбрана. Приведём два примера.

Пример 1.

Вместо модели риска Ф.Басса мы можем использовать модель Б.Гомперца, чтобы аппроксимировать те же данные d_i . Если полученные прогнозные значения модели Гомперца $x_{A,Gompertz}(t)$ окажутся близки к значениям, которые предсказывает модель Басса $x_{A,Bass}(t)$, то близки будут и прогнозируемые $x_{all}(t)$, $x_{some}(t)$ и т.п.

Пример 2.

Если мы усложним модель **А**, рассмотрев атомизованное, но неоднородное общество, например, используя расширяющийся со временем потенциал распространения нововведения (см. модель **АС**, описанную ниже), и применим усложненную модель к тем же самым данным наблюдений, то прогнозируемые $x_{all}(t)$, $x_{some}(t)$ тоже почти не изменятся.

Заметим, что при моделировании $U(t)$ мы не учитывали резкое снижение стоимости минуты разговора (MoU) и стоимости мобильного терминала, которые имели место в 1999-2003 годах, и, безусловно, оказывали влияние на решения индивидов и домохозяйств о начале использования мобильной связи. Однако мы предполагаем, что изменение цен во времени не нарушит связь между теоретическими кривыми $x(t)$ и $u(t)$ и не отменяет выводов относительно необходимости использования домохозяйств в качестве решающей единицы. Возможно, линейную зависимость (16) потребуется заменить более сложной, в которой задержки между последовательными покупками телефонов будут сокращаться. Модели, учитывающие снижение MoU обсуждаются в следующем разделе.

Отметим, что подобно многим процессам в социально-экономических системах, процесс распространения нововведений, по-видимому, может быть описан с удовлетворительной точностью и при помощи других моделей. В частности, в нашей модели не учтены различия в использовании мобильной связи различными возрастными группами. “Глава домохозяйства”, который формально используется в нашей модели **В**, возможно, соответствует представителю возрастной группы 18-25 лет, которая осваивала мобильную связь наиболее активно. При этом различия между использованием мобильной связи в домохозяйствах различного размера, возможно, удастся полностью объяснить наличием или отсутствием в домохозяйстве молодых людей, а не самими размерами домохозяйств. Однако модели, учитывающие возраст членов домохозяйств, требуют для идентификации гораздо более детальные данные, чем те, которыми мы располагаем.

Замечание о механизме приобретения и передачи устройств

В рассмотренной модели **В** предполагается, что первоначально глава домохозяйства приобретает сотовый телефон для себя, а позже, по мере накопления средств, домохозяйство приобретает мобильные устройства и для других жильцов.

Как отмечалось выше, возникает вопрос, не является ли более правдоподобным сценарий, когда глава домохозяйства по мере появления денег приобретает новые устройства для себя, а подержанную трубку дарит другому члену семьи.

С целью выяснить, насколько распространена практика дарения сотовых телефонов, автор настоящей работы в 2007 году провел опрос на Интернет-портале Рамблер. Результаты представлены на рис.3.8.

Дарили (отдавали) ли вы мобильный телефон?



Рис.3.8. Итоги онлайн-опроса о дарении или передаче мобильных телефонов

Оказалось, что практика дарения сотовых телефонов весьма распространена, по крайней мере среди пользователей Интернета. Итоги проведенного нами онлайн опроса подтверждают, что общество (или та его часть, которая использует Интернет и активно участвует в онлайн-опросах), в отношении приобретения сотовых телефонов не является атомизованным. Более правомерно назвать его

обществом с префигуративным типом культуры [221,141]. Более 80% участников опроса в прошлом дарили или отдавали свои сотовые телефоны. При этом мы снова наблюдаем возрастную неоднородность того же характера, что изучаемая в первой главе настоящей работы – чаще всего респонденты дарили мобильный телефон родителям. Отметим, что родители могут вести домохозяйство отдельно от детей, таким образом дарение устройств может пересекать границы домохозяйств.

Дополнительный вопрос (рис.3.9) уточнял механизм передачи сотового телефона. Как показали результаты онлайн-опроса, среди тех респондентов, которые дарили телефоны, приблизительно половина приобретали их специально, а другая половина дарила собственный. Таким образом, в реальности имеют место оба упомянутых механизма.

Вы отдали/подарили этому человеку собственный мобильный телефон или специально приобрели новый мобильный телефон в подарок?

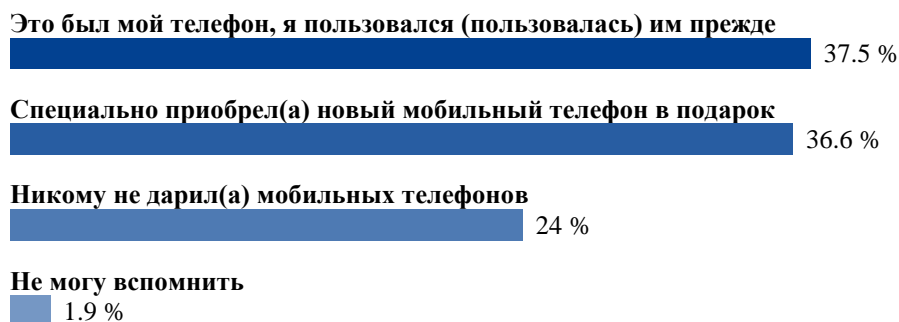


Рис.3.9. Итоги онлайн-опроса о дарении или передаче мобильных телефонов

Безусловным недостатком проведенного нами опроса является его нерепрезентативность. Не исключено, что индивиды, не склонные делать подарки, попросту не примут участие в таком опросе. Поэтому представленные данные приведены в качестве иллюстрации, а не доказательства.

К сожалению, в 2007 году, когда был проведен опрос, математическая модель распространения нововведения еще не была построена, поэтому опрос не

служил инструментом ее верификации. К тому моменту, когда автор построил модель процесса, он уже потерял возможность проводить онлайн-опросы.

Таким образом, оба рассмотренных способа поведения главы домохозяйства распространены и подтверждается онлайн-опросами. Как уже отмечалось, математические модели для выбранных нами показателей совпадают для обоих случаев.

3.1.2 Модели с линейно растущим потенциалом распространения

В этом разделе мы сосредоточимся на двух аспектах процесса распространения сотовой связи. Мы продемонстрируем необходимость выделения по меньшей мере двух этапов распространения сотовой связи в России и построим количественную динамическую модель диффузии для второго этапа, которая учитывает информацию о снижении стоимости минуты разговора. Наш подход отличается от предыдущих исследований распространения мобильной связи в России, в которых либо использовались регрессионные модели, не включающие время в качестве переменной [203], либо динамические модели не логистического семейства [249], либо одна диффузионная модель для всего периода роста числа абонентов [61,60,95,111,177,178,206,301].

Динамика проникновения сотовой связи в России

Известно, что первый пользовательский мобильный звонок в России совершил 9 сентября 1991 г. мэр Санкт-Петербурга А. А. Собчак, который позвонил в США. Вплоть до 1998 года пользователями мобильной связи являлись преимущественно организации и состоятельные индивиды [248,268]. Второй этап распространения этого нововведения в России начинается в 1999 году, когда, обнаружив отток абонентов, крупные операторы приняли решение о выводе услуги мобильной связи на массовый рынок и начали снижать стоимость минуты разговора. В итоге рост числа абонентов возобновился, а доля владельцев сотовых телефонов среди россиян в 2000 году достигла 2%.

Ю.А.Бакман и В.С.Панфилов полагают, что началом этапа бурного роста систем сотовой связи в России можно считать 1999 г., когда количество

пользователей превысило первый миллион человек [20]. О.Панкратова отмечает более высокую скорость роста распространения мобильной связи в России по сравнению с европейскими странами, так что расширение абонентской базы с 40% до 100% произошло в России за 2 года [237].

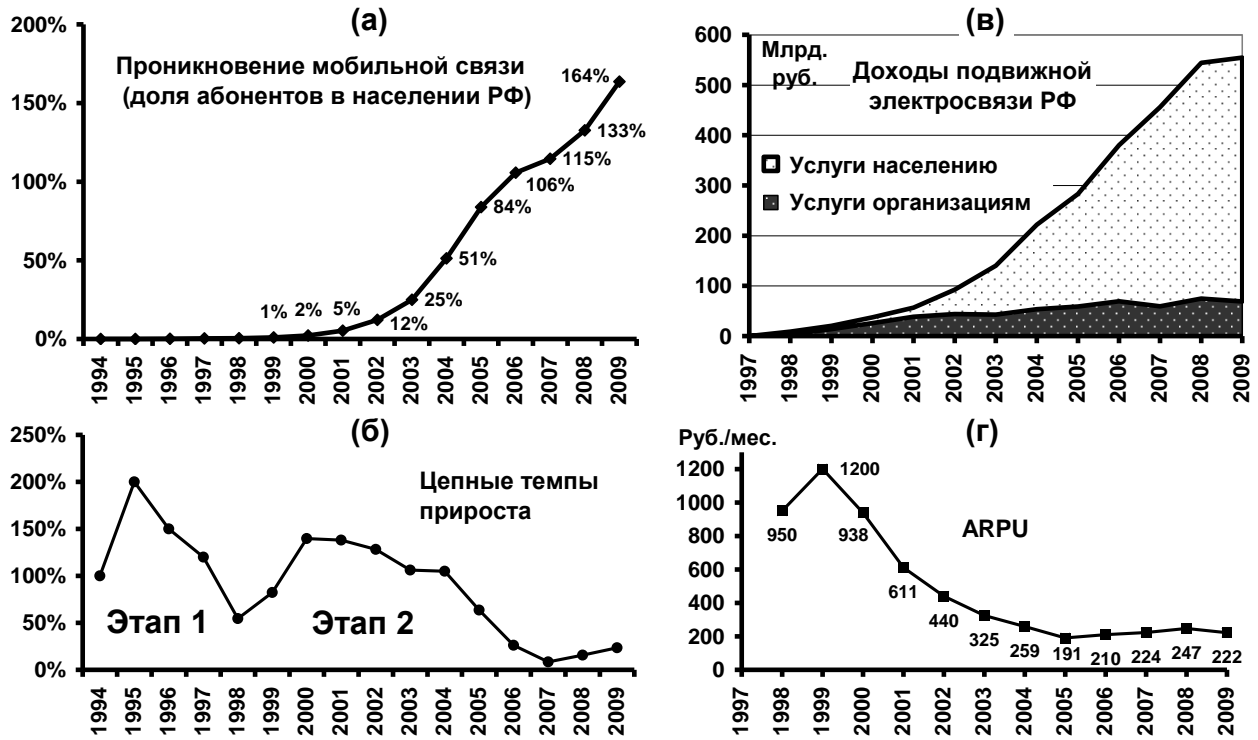


Рис.3.10. Показатели распространения мобильной связи в РФ: а) относительное число абонентов, б) цепные темпы прироста числа абонентов за год, в) доходы подвижной электросвязи за год, г) среднемесячная выручка с абонента (ARPU). Использованы данные Росстата.

В 2010 г. Международный Союз Электросвязи (МСЭ) отметил, что подобный “рывок” был характерен и для стран СНГ в целом. Мы полагаем, что причиной столь высокой скорости диффузии стал быстрый рост потенциала рынка, вызванный снижением средней стоимости минуты разговора (т.н. *аррт*), к которому привела интенсивная конкуренция операторов сотовой связи. Как показывает рис.3.10(б), высокие цепные темпы прироста количества абонентов $R(t) = (X(t + \Delta t) - X(t)) / X(t)$ наблюдались и в 1991-1998 г., однако к 1999 г. потенциал первого этапа распространения мобильной связи в РФ был исчерпан.

Объем услуг связи населению в денежном выражении в этот период не превышал половины объема услуг организациям. Однако он быстро рос, и в 2009 году семикратно превысил объем услуг организациям (см. рис.3.10.(в)). Поэтому в период после 1999 года мы ограничиваемся моделированием распространения мобильной связи среди населения.

Начиная с 2000 года, публикуются уникальные в плане разнообразия и полноты данные массовых опросов о личном владении мобильным телефоном (“Левада-Центр”, ФОМ, ВЦИОМ, РОМИР). Мы присоединяемся к мнению И.С.Березина [24] о том, что количество активных абонентов (SIM-карточек), которое во многих исследованиях используется в качестве данных, не является надежным статистическим показателем, поскольку операторы связи не раскрывают методику вычисления активной абонентской базы, так что известен случай, когда крупный российский оператор “списал” сразу восемь миллионов абонентов. Поэтому в качестве данных для количественного моделирования мы выбираем данные массовых социологических опросов, хотя для иллюстрации этапов распространения сотовой связи привлекаем данные Росстата о количестве абонентов.

При помощи рис.3.11 можно сравнить показатели распространения мобильной связи в России и Тайване. Тайвань является первым крупным регионом, где количество абонентов превзошло численность населения [335]. Это произошло в 2002 году. Рис.3.11(а) и (в) демонстрируют сходство динамики относительного числа абонентов (количество активных абонентов, отнесенного к численности населения). Здесь треугольниками изображены данные Росстата и МСЭ, а квадратами – данные о стоимости минуты разговора [237,278,335].

На рис.3.11(б) и (г) показаны результаты расчета цепных темпов прироста количества абонентов с шагом Δt , равным одному году. Как в Тайване, так и в России на графике темпов прироста отчетливо выделяются две фазы убывания темпов, обозначенные цифрами “1” и “2”, обведенными в кружок. Сравнение поведения темпов прироста $R(t)$ с динамикой стоимости минуты разговора $аррт$

позволяют предположить, что кратковременный быстрый рост $R(t)$ является результатом быстрого снижения $аррт$. Действительно, в случае Тайваня в 1997 году на рынок в дополнение к монополисту фиксированной связи было допущено пять новых компаний, что повлекло 22-кратное снижение $аррт$ в течение двух лет и 32-кратное за весь период.

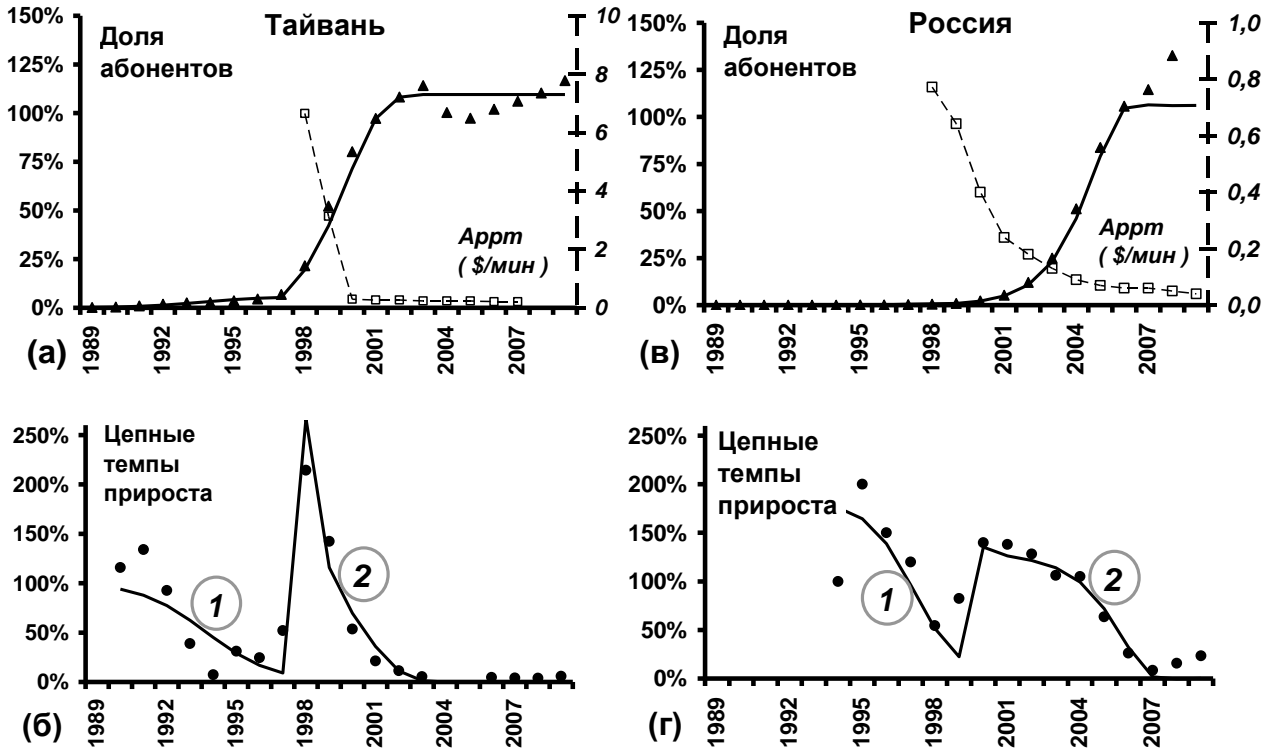


Рис.3.11. Динамика показателей распространения мобильной связи в Тайване: а) – относительное число абонентов, б) – цепные темпы прироста числа абонентов за год; аналогичные показатели в России: в) – относительного числа абонентов, г) – цепные темпы прироста. Пунктирными линиями показана стоимость минуты разговора ($аррт$), которой соответствуют правые оси (а) и (в). Цифры 1 и 2 в кружках обозначают две фазы распространения мобильной связи. Использованы данные [237,278,335] и МСЭ (www.itu.int/ITU-D/ICTEYE/).

В 1999 году снижение цен начинается и в России, но не столь стремительное, хотя в итоге стоимость минуты разговора падает почти в 20 раз (см.табл.3.5). Темпы прироста абонентской базы здесь были не столь высоки, так что кривая

относительной доли абонентов на рис.3.11(в) отстает от кривой на рис.3.11(а) приблизительно на пять лет.

С целью объяснения поведения цепных темпов прироста, сплошными линиями на рис.3.11 показаны результаты наших расчетов [123] по двухсегментной модели распространения нововведений П.Стеффенса и Д.Мурти [461]. Поскольку снижение цен в Тайване произошло практически “мгновенно”, мы полагаем, что двухсегментная модель удовлетворительно описывает динамику показателей мобильной связи в Тайване. Однако в России процесс снижения цен занял несколько лет, в течение которых показатель *аррт* монотонно падал, поэтому для второго этапа распространения мобильной связи в нашей стране мы считаем более подходящей модель с линейным, а не скачкообразным ростом потенциала рынка.

Уравнение используемой в наших работах модели комбинированных воздействий с линейно растущим потенциалом рынка имеет вид [116,112,122]

$$\frac{dX}{dt} = \left(p + q \frac{X(t)}{K} \right) (M(t) - X(t)). \quad (1)$$

с начальным условием $X(0) = N_0$.

Здесь используются следующие обозначения: $X(t)$ – текущая численность пользователей, т.е. тех, кто уже использует нововведение в момент времени t ; $M(t)$ – потенциал распространения инновации, т.е. полная численность *потенциальных* пользователей, включающая как тех, кто уже пользуется, так и тех, кто способен когда-либо воспользоваться нововведением, p – параметр внешних воздействий (параметр инновации), q – параметр внутренних воздействий (параметр имитации), K – полное число коммуницирующих индивидов (например, численность популяции).

При этом потенциал $M(t)$ может быть функцией дохода индивидов, *аррт* и других переменных. Предполагается, что $X(t) \leq M(t)$ и $dM/dt \geq 0$.

Решение уравнения (1) для случаев, когда потенциал $M(t)$ линейно растет со временем $M(t) = \mu + \lambda \cdot t$, или постоянен, получено нами в работах

[131,116,112,122] при помощи т.н. *интеграла Доусона* – специальной функции, для вычисления которой с высокой точностью существует большое количество алгоритмов и программ. В настоящей работе моделирование распространения мобильной связи в условиях линейно растущего потенциала рынка $M(t)$ проведено при помощи методов, описанных в [122].

Таблица 3.4.

Типы моделей распространения нововведения

Единица, принимающая решение	Потенциал $M(t)$	
	Постоянный	Растущий
Индивид	АС	AV
Домохозяйство	ВС	BV

Четыре модели диффузии нововведений, результаты применения которых ко второму этапу распространения мобильной связи в России обсуждаются ниже, перечислены в табл.3.4. Модель **АС** с постоянным потенциалом рынка и индивидом в качестве единицы, принимающей решение, является наиболее распространенной [61,60,111,177,178,301], однако, на наш взгляд, недопустимо упрощает действительность. Модель **AV** с переменным потенциалом рынка рассмотрена в наших работах [112,122]. Модель **ВС**, в которой решения принимают домохозяйства, изучена в наших работах [70,84,87,95,111], и, наконец, **BV**, в которой решения принимают домохозяйства и потенциал рынка растет со временем, предложена в работах [116,112,118].

Модель AV

Новизна предлагаемой в настоящей статье модели **AV** состоит в том, что в ней учтен рост потенциала $M(t)$, который в классической модели **АС** считался постоянным. В реальности доступность мобильной связи, которую отражает $M(t)$, росла во времени. Сперва $M(t)$ был мал по сравнению с численностью населения страны. Так, Ю.Федорова отмечает, что “всего 10 с небольшим лет назад сотовый телефон считался предметом роскоши и далеко не каждый мог себе позволить

оплачивать счета, составлявшие в среднем 2-3 сотни долларов в месяц» [278]. Сходное мнение высказывает В.В.Радаев: “Мобильные телефоны относились к разряду экзотики еще в середине 1990-х годов (всего лишь 5-8 лет назад), служили знаком статусного превосходства, элементом престижного потребления” [248].

Однако уже в августе 2006 г. большинство респондентов ФОМ полагало, что мобильный телефон может иметь каждый, кто захочет (57%). Еще 17% респондентов связали наличие сотового телефона с уровнем дохода или наличием доступа. Несмотря на указанную возможность, в этот период только 54% взрослых россиян владели личным сотовым телефоном. Следовательно, потенциал $M(t)$ уже в 2006 г., по-видимому, составлял не менее 83% взрослого населения, и был заметно выше, чем текущий уровень проникновения $X(t)$. В июне 2004 г. было опубликовано мнение представителя оператора, согласно которому в Перми в результате “ценовых войн” операторов мобильная связь стала доступна даже малообеспеченным слоям населения, в том числе тем, кто был готов тратить на нее лишь 1-2 доллара в месяц [268]. При этом проникновение сотовой связи в Перми составляло 50%, а в Пермской области – 25-30% (этот уровень совпадал со среднероссийским). Эти данные подтверждают предположение о том, что в 2004 году, потенциал рынка достиг максимальной величины ($M(t) = M_{\max}$), и заметно превысил проникновение $X(t)$.

Таблица 3.5. Динамика АРРМ в России

в 2000-2009 гг., долл. США по данным [237,278]

Год	АРРМ	Год	АРРМ	Год	АРРМ
	(\$/min)		(\$/min)		(\$/min)
1998	0,77	2002	0,18	2006	0,06
1999	0,64	2003	0,13	2007	0,06
2000	0,40	2004	0,09	2008	0,05
2001	0,24	2005	0,07	2009	0,04

Как показывают данные табл.3.5 (см. также рис.3.12), данные о стоимости минуты разговора (*аррт*) в период между 1999 и 2009 гг. можно удовлетворительно описать при помощи гиперболического тренда

$$P(t) = \frac{P_0}{t - T} \quad (2)$$

параметры которого, определенные методом наименьших квадратов, оказались равны: $P_0 = 223,7$ (долл.*сут.⁻¹), $T = 04/10/1998$. (В качестве единицы времени приняты сутки). Данные об *аррт* взяты из работ [237,278] и изображены серыми треугольниками на рис.3.12, а результаты расчетов по формуле (2) показаны сплошной серой линией.

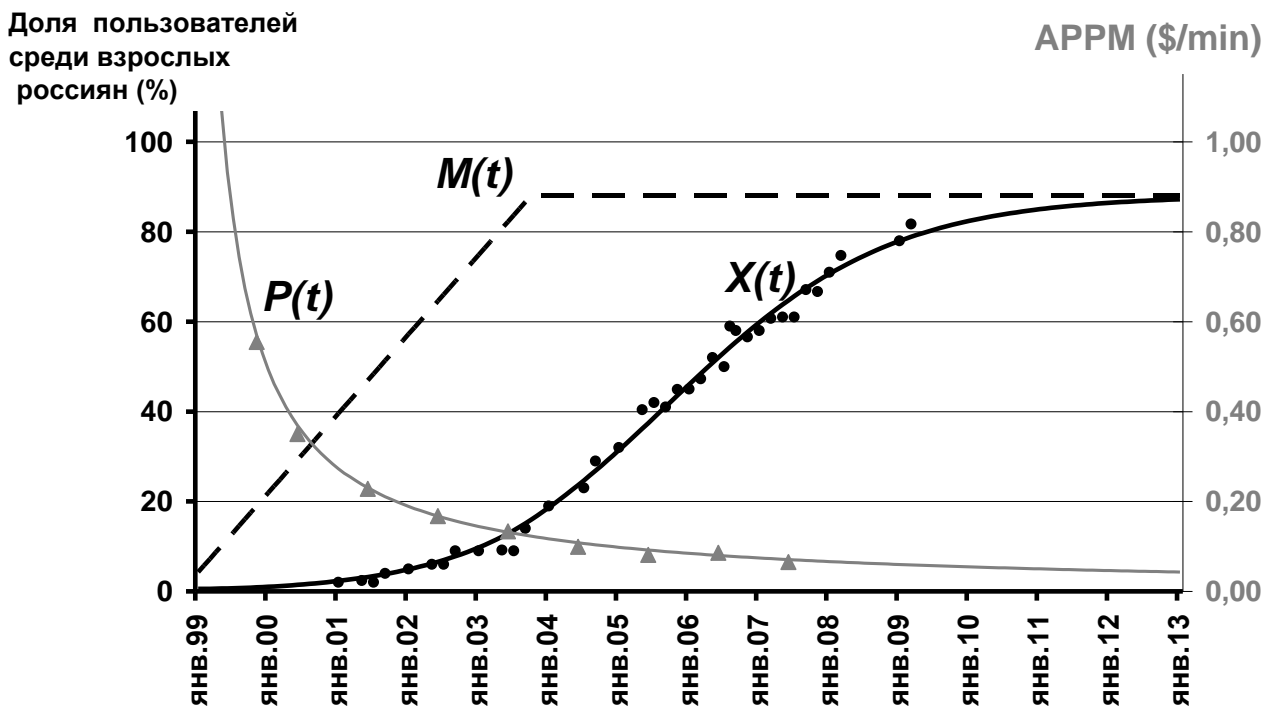


Рис.3.12. Динамика наличия сотового телефона в личном пользовании по данным Левада-Центра – кружки, динамика *аррт* по данным [237,278] – треугольники.

Чтобы упростить задачу мы игнорируем рост доходов россиян и снижение стоимости сотового телефона в рассматриваемый период, полагая, что снижение *аррт* было ключевым фактором. Примем во внимание лишь снижение средней стоимости минуты разговора (*аррт*) и рассмотрим модель, в которой только *аррт*

определяет потенциал $M(t)$. Используем для указанного периода обратную зависимость потенциала распространения новой услуги от ее стоимости:

$$M(t) = \frac{A}{P(t)} \text{ пока } M(t) < M_{\max}, \quad \text{далее } M(t) = M_{\max}. \quad (3)$$

Можно интерпретировать такую зависимость как единичную эластичность спроса по цене. Подставляя (2) в (3), получаем *линейно растущий потенциал* распространения нововведения

$$M(t) = \frac{A}{P_0}(t-T), \text{ пока } M(t) < M_{\max}. \quad (4)$$

Мы зафиксировали начальную долю владельцев телефонов N_0/K равной 0,54% в начале 1999 г., и предположили, что модель **AV** справедлива, начиная с этого момента времени. Таким образом, последние два параметра модели **AV** табл.3.6 были заданы нами заранее. Остальные четыре параметра табл.3.6 были определены нами методом наименьших квадратов при помощи данных Левада-центра и явного выражения для $X(t)$, (использующего интеграла Доусона) полученного в нашей работе [122].

Таблица 3.6

Параметры моделей **AC** и **AV**

Модель	M/K (%)	q (лет ⁻¹)	p (лет ⁻¹)	A (чел. * руб./сут.)	N_0/K (%)	t_0
AC	91,3	0,71	$2,9 \cdot 10^{-5}$	–	0	1.09.1991
AV	88,1	1,52	$8,4 \cdot 10^{-2}$	10,81	0,54	1.01.1999

Модель **BV**

Модель атомизованного общества объясняет лишь часть доступных данных опросов – данные о личном владении сотовым телефоном, но неспособна объяснить данные о наличии этого устройства в домохозяйстве. В 2001-2006 гг. доля россиян, у которых мобильный телефон есть в домохозяйстве, систематически превышала долю имеющих мобильный телефон в личном пользовании (см.рис.3.3 и рис.3.5).

Это наблюдение представляется тривиальным, однако специально не учитывается в моделях атомизованного общества **АС** и **AV**. С целью построения количественной модели, объясняющей все описанные ранее данные социологических опросов, нами была изучена модель **BC** [70,111]. В этой модели единицей, принимающей решение, является домохозяйство, а не индивид, однако потенциал распространения нововведения $M(t)$ предполагался постоянным, как и в модели **АС**. Последнее предположение игнорирует имевшее место в реальности значительное удешевление услуг мобильной связи, отмеченное в работах [237,248,249,278]. Модель **BV**, учитывающая такое удешевление, предлагается в наших работах [116,112,118].

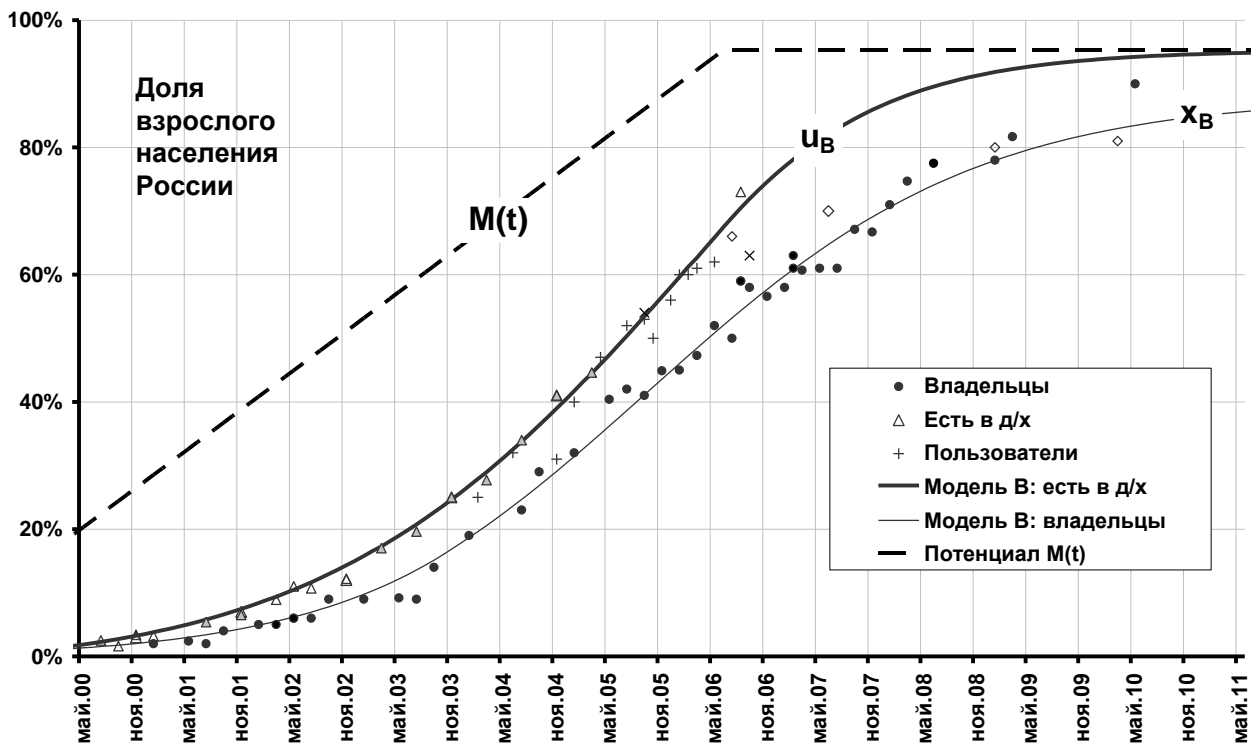


Рис.3.13. Динамика наличия сотового телефона в домохозяйствах (по данным Левада-Центра, ФОМ, ВЦИОМ).

На рис.3.13 представлены результаты расчета при помощи модели **BV** с переменным потенциалом $M(t)$. В этом случае уравнению (1) удовлетворяет не количество владельцев телефонов $X(t)$, а количество *пользователей* сотовой связи $U(t)$ (возможно, использующих устройства, принадлежащие другим членам

семьи). Количество владеющих мобильными телефонами домохозяйств численностью i членов выражается формулой

$$X_{i1}(t) = \alpha_i U(t) + (X_{i1}(0) - \alpha_i U(0)) e^{-H(t)}, \quad (5)$$

которая является аналогом формулы (41) из предыдущего раздела.

Здесь используется интегральная функция риска

$$H(t) = pt + \int_0^t q \frac{U(\theta)}{K} d\theta. \quad (6)$$

В случае линейной функции риска (см. формулу 11 предыдущего раздела) выражения (5) и (6) можно выразить через интеграл Доусона [122] (см. раздел 1.2.2 данной диссертации). Здесь мы рассматриваем лишь наиболее простой для расчетов частный случай пропорциональных начальных условий ($X_{i1}(0) = \alpha_i U(0)$).

Таблица 3.7

Параметры моделей **BC** и **BV**

Модель	M/K (%)	q (лет ⁻¹)	p (лет ⁻¹)	τ (лет)	β	γ	A (чел. * руб./сут.)
BC	85,84	0,8	$1,77 \cdot 10^{-5}$	1,09	1,0	1,0	–
BV	95,00	0,8	0,11	1,09	1,0	1,0	10,81

На рис.3.13 потенциал $M(t)$ изображен пунктирной линией, доли пользователей сотовой связи среди взрослых россиян $u_B(t) = U(t)/K$ и доля владельцев телефонов среди взрослых россиян $x_B(t) = U(t)/K$ – жирной и тонкой линиями. Отметим, что в модели **BV** потенциал $M(t)$ возрастает вплоть до 2006 г., а не до 2004 г., как в модели **AV**.

Мы полагаем, что данные опросов совместно с данными об *аррт* не устраняют неопределенность в поведении потенциала $M(t)$, поэтому нельзя утверждать, что модель **BV** “лучше” описывает данные, чем модель **BC**. Можно лишь утверждать, что модель **BV** тоже описывает данные приблизительно с той же точностью. Однако модель **BV** учитывает снижение *аррт*, в то время как модель **BC** этот факт полностью игнорирует.

Отметим также, что, хотя различие в поведении потенциала на рис.3.12 и рис.3.13 кажется не слишком значительным, оно отвечает существенно различному поведению пользователей. В модели **AV**, соответствующей рис.3.12, потенциал вырос резко (почти скачкообразно), однако индивиды проявили определенную инертность, медленно принимая решение о приобретении телефона. В модели **BV**, которой соответствует рис.3.13, операторы, постепенно снижая стоимость минуты разговора, плавно увеличивают потенциал $M(t)$, после чего домохозяйства существенно быстрее, чем в **AV**, приобретают сотовые телефоны. В монографии Э.А.Разроева [249] описаны маркетинговые исследования, при помощи которых операторы, снижая цены, пытались определить уровень доходов пользователей, которым доступна услуга, тем самым измеряя потенциал рынка $M(t)$. Как показывает табл.3.6, в модели **BV** оказывается высок параметр внешних воздействий p , а значит, на решения домохозяйств значительное влияние оказывала реклама операторов.

Заметим, что приведенный в табл.3.6 набор параметров – лишь один возможных наборов, дающих приблизительно одинаковую ошибку аппроксимации данных. Изучаемая проблема характеризуется обилием правдоподобных решений, поэтому мы имеем возможность указать лишь на еще одно из возможных объяснений, но, безусловно, не можем выделить единственное решение. Подобная неоднозначность вообще характерна для исследований диффузии нововведений, тем более, что потенциал $M(t)$, играющий важную роль в модели (1), гораздо сложнее измерить, чем количество пользователей, владельцев или абонентов. Мы полагаем, что однозначные результаты могут быть получены лишь в том случае, если исследователи будут с самого начала планировать опрос как источник данных для построения диффузионных модели распространения нововведения.

Замечание о функции распределения задержки $g_{ij}(t)$

В то время как выбранный нами вид плотности распределения задержки $g_{ij}(t)$ в виде дельта-функции (и, соответственно, ступенчатой функции дожития)

упрощает расчёты, более реалистичной альтернативой, по-видимому, является функция дожития в виде гамма-распределение, о чём свидетельствует рис.3.14, полученный при помощи онлайн-опросов, проведённых автором на портале Рамблер.

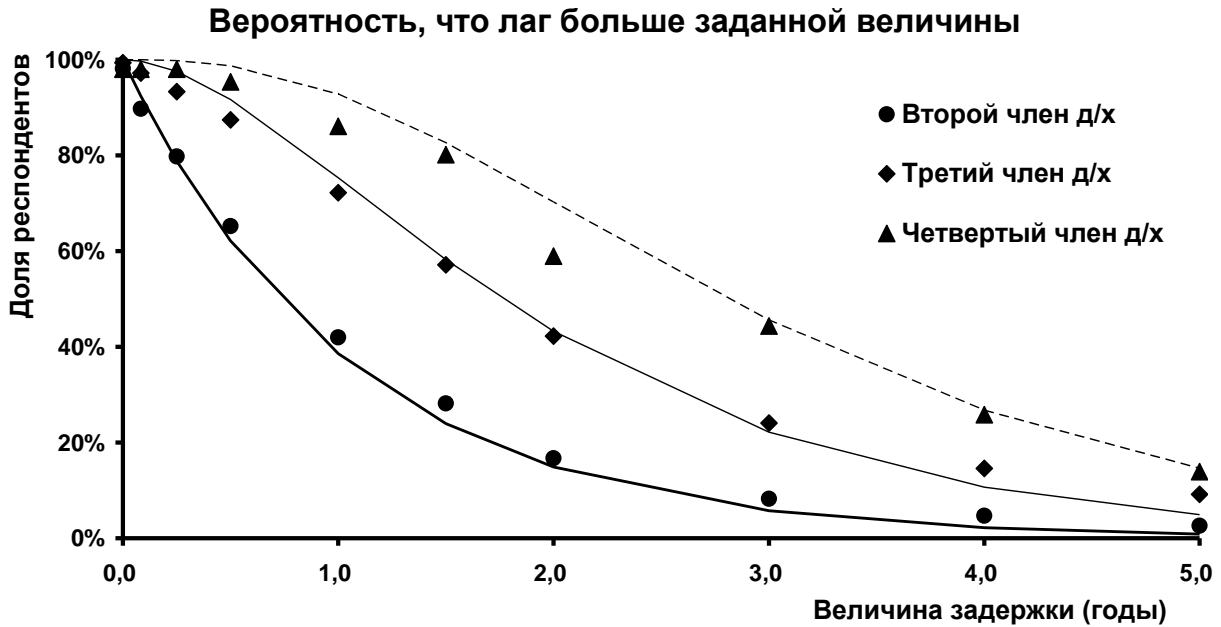


Рис.3.14. Распределение величины задержки до приобретения личного сотового телефона (по данным онлайн-опроса).

Несмотря на сделанное замечание, мы полагаем, что форма распределения задержки не оказывает существенного влияния на результаты моделирования и не меняет сделанных в этой работе выводов.

3.2 Совместное моделирование компьютерного рынка и парка компьютеров

Количество персональных компьютеров (ПК) на душу населения страны — один из ключевых показателей развития информационного общества, используемый при составлении страновых рейтингов ООН и Международного союза электросвязи (МСЭ). В данной работе представлены расчетные оценки размеров российского компьютерного парка и компьютерного рынка России в натуральных величинах. Построенные нами ранее прогнозы динамики компьютерного парка до 2012 г. [134,113,125,179] скорректированы и

сопоставлены с новыми данными о продажах компьютеров, и прогнозами Минкомсвязи и Министерства экономического развития (МЭР) [246].

3.2.1 Модель распределения срока службы компьютера

Под *эксплуатационным компьютерным парком* мы понимаем совокупное количество компьютеров (настольных, мобильных и серверов), которые эксплуатируются пользователями. Компьютерный парк РФ по известным нам оценкам насчитывает от 40 до 63 миллионов устройств.

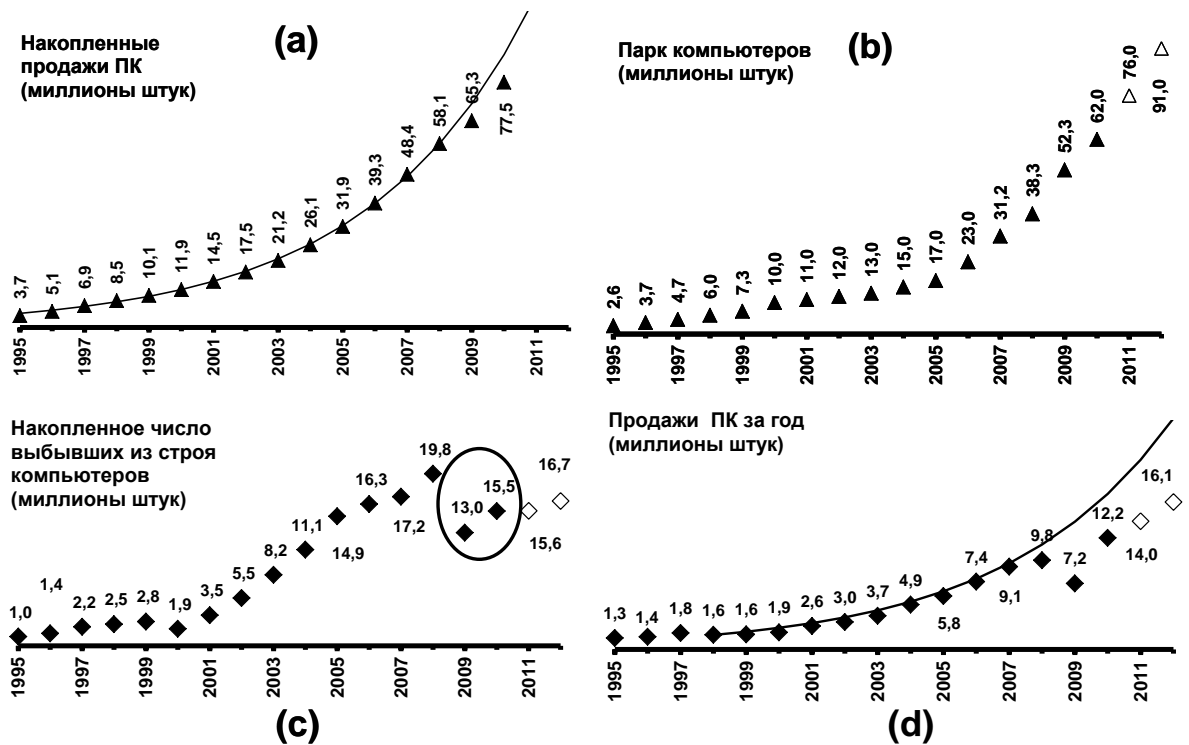


Рис.3.15. Показатели продаж и парк ПК в РФ в 1995-2010 гг.

При анализе *годового объема продаж* ПК в России в натуральном выражении мы используем осреднение данных трех источников, в числе которых исследовательские компании IDC Russia, IT-Research и Gartner-Dataquest, публикующие статистику продаж новых компьютеров (см., например, [385,386,280,306]). Компания IDC подразумевает под *рынком ПК* поставки производителей дистрибьюторам персональных компьютеров, включая настольные, портативные и серверы стандартной архитектуры, сходную методику

использует и Gartner. Компания IT-Research оценивает величину рынка иным способом, подсчитывая отгрузку дистрибьюторов в розницу. Все три организации включают в свои оценки не только продажи крупных поставщиков, но и т.н. ‘самосбор’ ПК из комплектующих мелкими фирмами и непосредственно потребителями. Осредненные продажи изображены на рис.3.15(d), вместе с хорошо аппроксимирующей их в 1999-2007 гг. экспонентой. Добавлена наша оценка продаж за 2010 год и наш наиболее оптимистический прогноз на 2010-2012 гг.

Отметим указанную ранее проблему необходимости разработки “вспомогательных” моделей, своего рода интерфейсов к данным [273,146]²², при попытке применения моделей распространения нововведений для изучения распространения новых информационных технологий. Безусловно, между поставкой компьютеров дистрибьюторам и продажей конечному потребителю протекает некоторое время, таким образом эти показатели отнюдь не тождественны. Различия не сводятся к простой задержке, поскольку в периоды падения спроса в магазинах могут накапливаться нераспроданные устройства. В этом разделе мы, так же как и перечисленные выше компании-измерители, предполагаем, что в стабильной экономической ситуации различия между поставками и продажами незначительны, поэтому считаем их совпадающими.

Запишем уравнение баланса парка компьютеров [36,264]

$$Q(t) = N(t) + R(t). \quad (1)$$

Здесь $Q(t)$ – продажи ПК (в “штуках”), накопленные за весь период существования рынка (в РФ мы условно начинаем отсчет с 1986 года); $N(t)$ – парк ПК, который мы определяем как количество компьютеров, которые используются владельцами; $R(t)$ – накопленные продажи ПК *на замену* выбывшим из строя, они равны количеству вышедших из строя ПК.

²² Ю.Н.Толстова [273] отмечает: “Социальные явления настолько сложны, что, каким бы хорошим метод ни был, всегда требуется определенная его подгонка под конкретную ситуацию. Любой метод всегда обрастает методическими наработками, связанными с его адаптацией, привязкой к реальной социологической задаче. Чаще всего упомянутая подгонка осуществляется за счет создания нетривиальных методик сбора данных и интерпретации получаемых с помощью выбранного метода результатов. При этом каждая конкретная задача требует своих собственных и теоретических, и практических разработок.”

На рис.3.15(a) изображены накопленные продажи $Q(t)$, полученные суммированием продаж за каждый год (рис.3.15(d)), а на рис.3.15(b) показан парк ПК $N(t)$ за период с 1995 до 2010 г. и прогнозные значения МЭР на 2011 и 2012 годы. При построении рис.3.15(a) и (d) использованы данные IDC, IT-Research, Gartner Dataquest, рис.3.15(b) - данные Минкомсвязи РФ и прогноз МЭР. Данные Минкомсвязи доступны на сайте Minsvyaz.ru в разделе “Деятельность министерства/Статистика отрасли”, исторические данные опубликованы в монографии [206]. На рис.3.15(c) все данные использованы совместно. Открытыми символами показаны прогнозные значения. Сплошной линией на рис.3.15(a) и (d) показаны экспоненты, наилучшим образом приближающие продажи за 1999-2007 гг. и отражающие известный факт роста российского рынка ПК устойчивыми темпами прироста около 21% в эти годы.

Переписав уравнение баланса парка компьютеров в виде

$$R(t) = Q(t) - N(t), \quad (1a)$$

мы получаем формулу для вычисления парка вышедших из строя компьютеров. Вычисления приводят нас к неожиданному результату, который показан на рис.3.15(c). А именно, в 2009 году *накопленное* количество сломанных компьютеров $R(t)$ якобы уменьшилось на 6,8 миллионов штук. Однако величина $R(t)$ не может убывать, поскольку является суммой последовательности положительных чисел, представляющих количество вышедших из строя компьютеров за последовательные годы (или иные интервалы времени). Даже если бы в 2009 году не сломался ни один компьютер, величина $R(t)$ оставалась бы постоянной, но не могла бы снизиться. Единственным объяснением снижения накопленного числа сломанных компьютеров был бы массовый ремонт вышедших из употребления устройств. Однако для объяснения столь крупных расхождений потребовался бы внезапный ремонт 6,8 миллионов ПК в 2009 году, а столь масштабные работы вряд ли остались бы незамеченными²³. Другой

²³ В.В. Ключков обратил внимание автора на возможность простоя значительного парка технических устройств, в частности, пригодные к эксплуатации самолёты могут не использоваться годами [185]. В случае отсутствия средств для покупки новых аппаратов, возможно возвращение в эксплуатацию старых. Мы признаём, что

причиной расхождений могли бы быть большие отрицательные ошибки в оценке рынка ПК аналитическими агентствами, однако падение рынка ПК, начиная с последнего квартала 2008 года, отмечалось не только в России. Более вероятным нам кажется предположение о том, что претерпела изменения методика Минкомсвязи оценки величины парка ПК, причем смыкание данных было произведено без учета совместимости с данными о продажах ПК.

Указанные расхождения побуждают нас построить модель и составить прогнозы компьютерного рынка и компьютерного парка Российской Федерации до 2012 г. При построении модели мы опираемся на подход В.Камакуры и Ш.Баласубраманьяна [392], который учитывает функцию распределения вероятности времени покупки “на замену”. Действительно, часть компьютерного парка ежегодно устаревает и нуждается в замене. Гипотеза, которая лежит в основе модели таких покупок, состоит в том, что “возраст замены” есть случайная величина, которая подчиняется распределению вероятностей, аналогичному распределению вероятности поломки продукта. В основе расчетов при этом лежит уравнение теории восстановления

$$\frac{dN}{dt} = q(t) - r(t) = q(t) - \int_{t_0}^t q(\theta)g(t - \theta)d\theta, \quad (2)$$

где $N(t)$ – парк компьютеров, $q(t) = \frac{dQ}{dt}$ – продажи ПК в единицу времени (в

“штуках”), $r(t) = \frac{dR}{dt}$ – продажи ПК “на замену” в единицу времени, $g(t)$ –

плотность распределения времени жизни ПК; t_0 - момент начала продажи компьютеров в изучаемом регионе.

Разумеется, распределение срока службы компьютеров меняется от года к году. Например, в год финансового кризиса фирма может отложить замену служебных компьютеров своих сотрудников, а частные пользователи – повременить с покупкой нового компьютера. Однако у нас недостаточно данных, чтобы учесть такие флуктуации. Поэтому мы считаем функцию отклика процесса устаревания

доступные данные не позволяют нам отвергнуть теоретическую возможность реализации такого сценария в 2009 году.

стационарной во времени. Ошибки, возникающие в результате такого упрощения, не мешают моделировать динамику компьютерного парка РФ с достаточной для наших целей точностью.

С целью определить неизвестную нам функцию $g(t)$ в конце 2004 года мы провели масштабный опрос о сроке службы компьютера на Интернет-портале Rambler. Основной вопрос звучал так: "В каком году был выпущен компьютер, с которого вы сейчас вышли в Интернет?" Вычитая из года проведения опроса год, который назвал респондент, мы вычисляли возраст компьютера. Такой подход позволил отчасти избежать округления возраста респондентами до величин, кратных 5, характерного для ответов на вопрос о *возрасте* предмета потребления [6], а не о моменте его приобретения. В ходе опроса 56 тысяч респондентов смогли сообщить год выпуска своих компьютеров. Мы учитывали приобретение только новых компьютеров, однако, как показали опросы, доля компьютеров, приобретенных поддержанными, невелика.

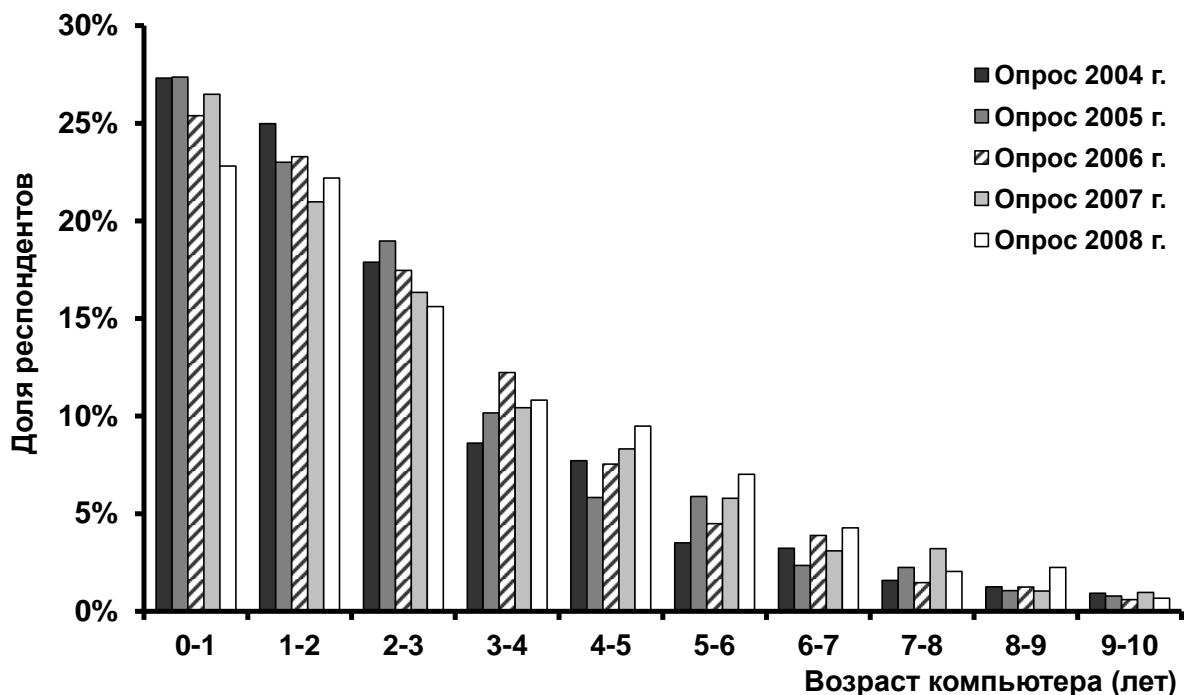


Рис.3.16. Итоги пяти онлайн-опросов о возрасте используемого компьютера.

В дальнейшем аналогичный опрос мы повторяли ежегодно в 2005-2008 гг., итоги пяти опросов представлены на рис.3.16. В каждом случае число респондентов измерялось десятками тысяч. Отметим, что частоты (частости), представленные на рис.3.16, оказались весьма близки к полученным в 2004 году результатам дневникового исследования КОМКОН, в котором панелистов спрашивали, как долго они пользуются одним компьютером.

Как показывает рис.3.16, итоги опросов за различные годы весьма близки, существенно отличаются лишь ответы за 2008 г., когда доля респондентов, купивших ПК непосредственно в год опроса, была заметно меньше, чем в предыдущие годы. Мы связываем это с резким падением продаж ПК в России в IV квартале, которое (в натуральном выражении) составило 38%. Мы исключили результаты опросов за 2008 год из процедуры определения распределения срока службы компьютера, поскольку использование таких данных требовало бы знания точной формы продаж $q(t)$ в период их быстрых изменений.

Обозначим $\Delta R(t)$ – количество ПК, вышедших из строя между t и $t + \Delta t$. Эта величина зависит от возраста компьютеров, и, таким образом, от истории продаж $q(t)$

$$\Delta R(t) = \int_t^{t+\Delta t} \frac{dL(T-\theta)}{d\theta} \cdot q(\theta) d\theta, \quad (3)$$

где $L(a)$ – т.н. “функция дожития” или *функция надежности* устройства, представляющая собой вероятность работоспособности устройства, возраст которого равен a ; $q(t) = \frac{dQ(t)}{dt}$ – интенсивность продаж (в единицу времени).

При помощи вероятностных рассуждений нетрудно показать, что модель для определения функции надежности $L(t)$ по данным опроса имеет вид

$$\Delta v_i = c \cdot \int_{t_{i-1}}^{t_i} L(T-\theta) \cdot q(\theta) d\theta, \quad (4)$$

где i – год, в частности 1998, 1998, ..., 2008; t_i – момент окончания этого года; T – момент проведения опроса; Δv_i – теоретическое значение количества

респондентов, ответивших, что компьютер, которым они пользуются, заполняя нашу анкету, приобретен новым в i -м году; c – неизвестная нам постоянная, равная доле наших респондентов среди владельцев ПК.

Если $L(t)$ и $q(t)$ между t_i и t_{i+1} меняются медленно, то для расчета Δv_i можно использовать приближение

$$\Delta v_i \approx c \cdot L(T - t_i + \delta) \cdot \Delta Q_i, \quad (5)$$

где ΔQ_i – продажи за год i (или иной период измерения) в натуральном выражении (штуках); $t_i - \delta$ – середина года опроса (данные за 2004 г. мы относим к середине 2004 г. и т.п.). Далее мы не будем использовать это приближение, поскольку для выбранного вида нами функции надежности существует точное аналитическое выражение для Δv_i , которое здесь и используется.

Для описания функции надежности В.Камакура и Ш.Баласубраманьян применили усеченное нормальное распределение [392], а в наших работах [134,113,125,133] использовалась усеченная логистическая функция

$$L(x) \equiv L(x; \mu, \tau) = \frac{1 + \exp(-\mu\tau)}{1 + \exp(\mu(x - \tau))} \quad (6)$$

с параметрами положения τ и масштаба μ . Однако при экспоненциально растущих продажах (которые и имели место в 1999-2007 гг.)

$$q(t) = \alpha \exp \beta t \quad (7)$$

наиболее удобным видом функции надежности оказывается гамма-распределение $\gamma(t; \mu, z)$ с параметрами μ и z и плотностью

$$g(t) = \frac{d\gamma}{dt} = \mu \frac{(\mu t)^{z-1}}{\Gamma(z)} e^{-\mu t}. \quad (8)$$

Подставляя $L(t) = \gamma(t; \mu, z)$ и $q(t)$ вида (7) в формулу (4) находим формулу количества респондентов

$$\Delta v_i = c \cdot e^{\beta T} (f(T - t_{i-1}) - f(T - t_i)), \quad (9)$$

где

$$f(a) = \frac{1}{\beta} \left\{ 1 - [1 - \gamma(a; z, \mu)] e^{-\beta a} - \frac{\mu^z}{(\beta + \mu)^z} \gamma(a; z, \beta + \mu) \right\}. \quad (10)$$

Поскольку расчет гамма-распределения $\gamma(t; \mu, z)$ является стандартной функцией большинства вычислительных пакетов (включая офисный пакет MS Excel), в случае экспоненциально растущего рынка величину (4) можно рассчитать с высокой точностью, не используя численное интегрирование. Для усеченного логистического распределения (6) также можно выразить число респондентов при помощи специальных функций, однако итоговая формула имеет еще более громоздкий вид. Поэтому в данной работе мы используем гамма-распределение. В более общем случае, когда функция продаж имеет произвольную форму, или при желании работать с поквартальными продажами, уравнение (4) следует интегрировать численно.

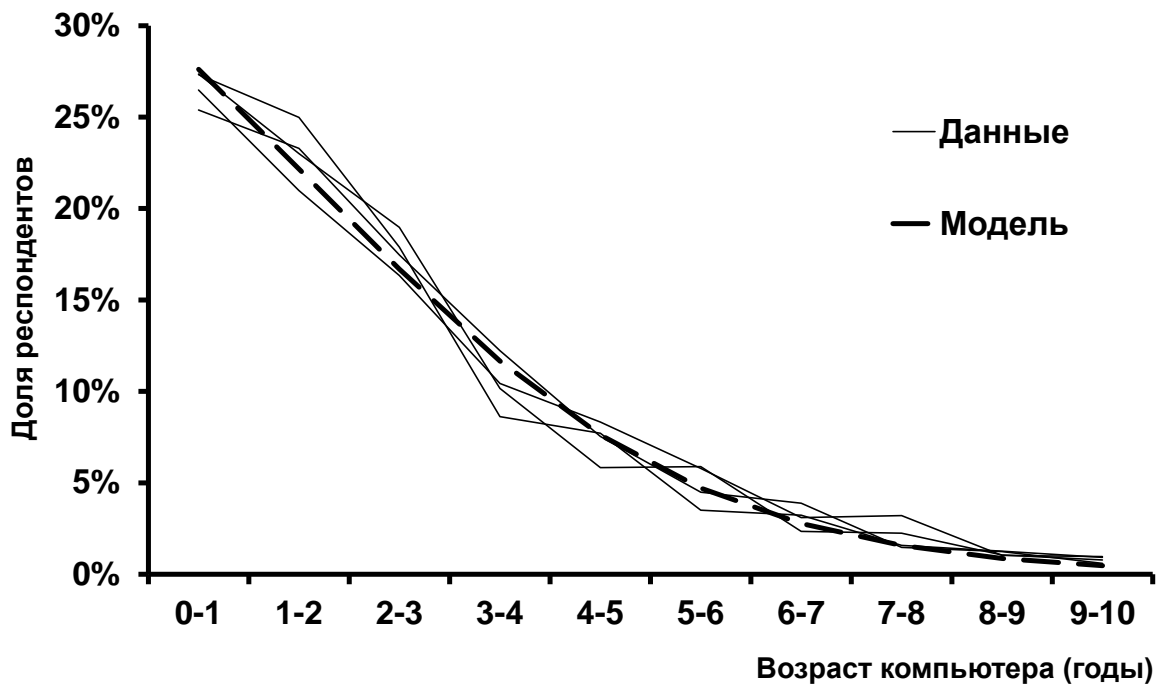


Рис.3.17. Итоги моделирования данных опросов о возрасте компьютера Δv_i .

Для определения параметров μ , z и c мы использовали метод максимального правдоподобия. На рис.3.17. тонкими линиями изображены данные опросов, жирной пунктирной линией – итоги расчетов при помощи формул (9-10) и оптимального набора параметров $\mu = 0,56$, $z = 3,4$. Параметр c различен для каждого года, поскольку число респондентов различалось, в

дальнейших вычислениях он не требуется. Средний срок службы компьютера при таком наборе параметров составляет 6,12 года.

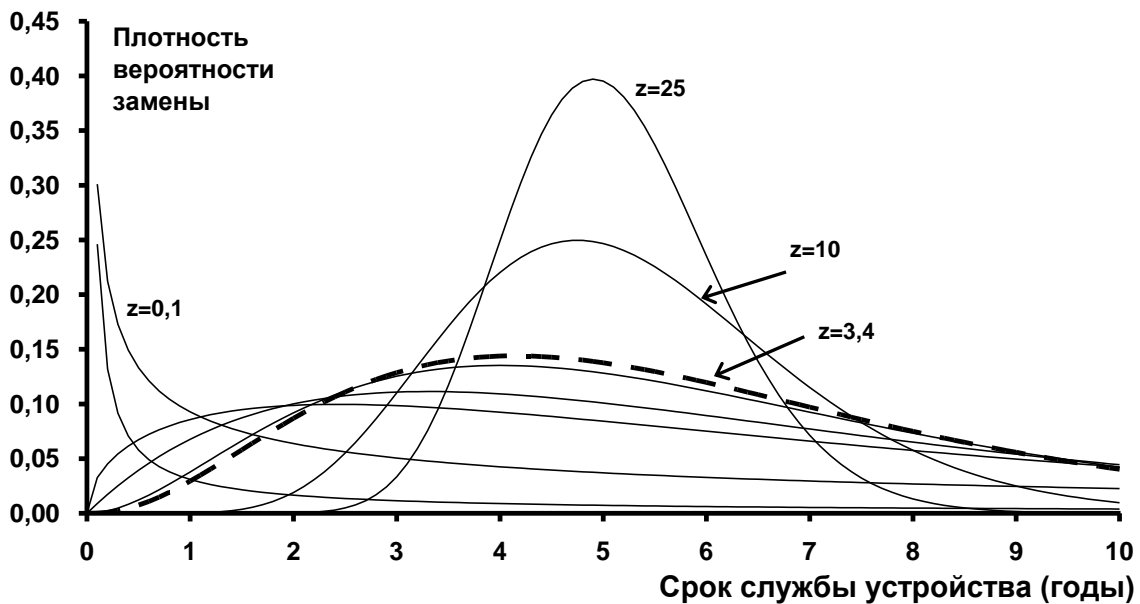


Рис.3.18. Теоретические плотности распределения срока службы компьютера.

Плотность гамма-распределения (8) изображена жирной пунктирной линией на рис.3.18 и весьма близка к плотности усеченного логистического распределения, использованного нами ранее [134,113,125,133]. Заметим, что при росте рынка с постоянными темпами (7), который имел место в 1999-2007 гг., любая из плотностей распределения на рис.3.18 продуцирует одинаковую кривую продаж 'на замену' (3), которая при темпах прироста рынка в 21% задержана по сравнению с первыми продажами на 5,18 года. Сделанное замечание обосновывает необходимость измерения плотности распределения при помощи опросов, поскольку в условиях экспоненциального роста рынка найти $g(t)$ при помощи известных продаж и парка ПК по формуле (2) невозможно, но для вычисления парка при более сложной зависимости $q(t)$ при помощи (2) потребуется знать точную форму $g(t)$.

В 2004-2008 гг. мы проводили опросы о моменте покупки компьютера в конце года, в декабре. Кроме того, дополнительные опросы были проведены в весенние, летние и осенние месяцы, при этом респондентам был задан вопрос о

том, в каком месяце они приобрели ПК, что помогло нам определить сезонные множители. Здесь мы ограничимся замечанием о том, что при определении срока службы компьютера предлагаемым нами методом важен месяц проведения опроса, поскольку сезонные колебания продаж весьма сильны, и существуют задержки между покупкой компьютера и подключением его к Интернету, а также между “поставкой” компьютера (которые подсчитывают маркетинговые агентства) и непосредственно продаж. Дальнейшие детали подробно обсуждаются в нашей работе [136].

Парк компьютеров РФ: сценарий F2010

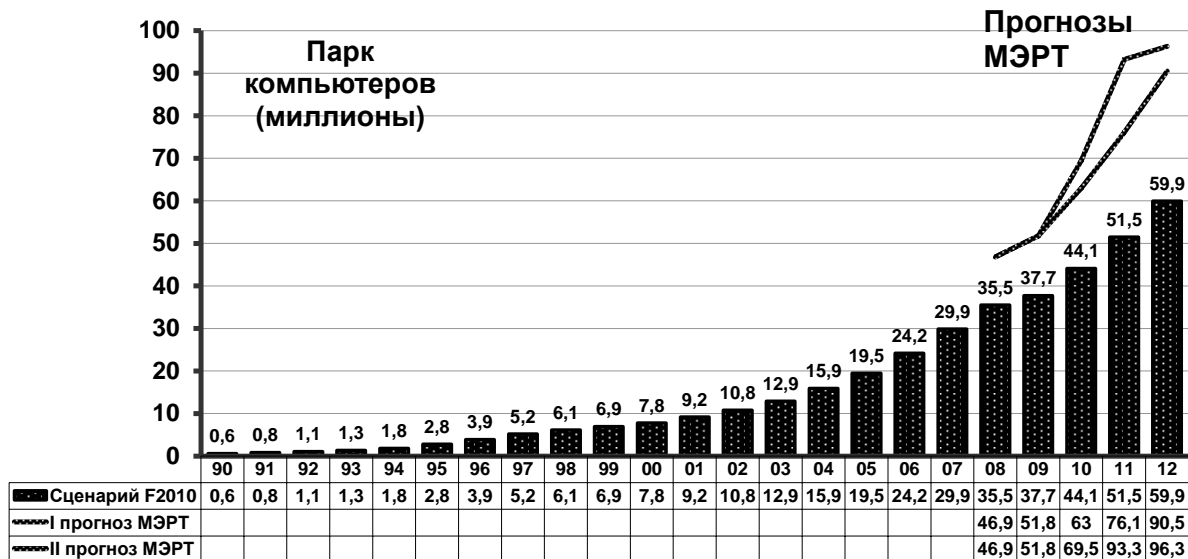


Рис.3.19. Компьютерный парк РФ: модель и прогнозы.

Нами рассчитаны несколько сценариев роста компьютерного рынка и сопутствующего роста парка компьютеров, при этом проанализированы несколько опубликованных стратегий [134]. В качестве оптимистического сценария роста компьютерного парка нами был предложен представленный на рис.3.19 сценарий, основанный на опубликованных прогнозах продаж с темпами прироста рынка персональных компьютеров в 15% в 2011-12 гг. В этом случае уровень продаж в 2012 г. должен составить чрезвычайно высокую величину – 16

миллионов компьютеров, указанную на рис.3.15(d). Мы намеренно используем столь высокое значение, чтобы подчеркнуть, что оно не смягчает диапазон расхождений наших оценок с опубликованными прогнозами МЭР.

Опубликованный весной 2010 г. Министерством Экономического развития прогноз парка ПК объемом в 63 миллионов штук [246], подтвержденный позже в публикациях Минкомсвязи оценки в 62 миллиона устройств более, чем на 40% превосходит размеры, которые мы получили путем расчетов по модели (2). Расхождения прогноза Минэкономразвития с нашим прогнозом 'F' на 2011 г. являются почти двукратными (см.рис.3.19). По расчетам министерства к 2012 году по сравнению с 2008 годом количество ПК возрастет более чем в 2 раза, достигнув 97 миллионов. Оптимистический и консервативный прогнозы Минэкономразвития изображены серыми линиями в правом верхнем углу рис.3.19, результаты наших расчетов показаны темными прямоугольниками.

Безусловно, одной из причин расхождений может быть грубая ошибка в оценке объема компьютерного рынка агентствами IDC, Gartner и IT-Research, в частности недоучет т.н. "самосбора". Однако методика IT-research использует подсчет количества ввезенных в РФ жестких дисков, которые не производятся внутри страны и не могут быть собраны самостоятельно. Более того, по утверждению компаний, их оценки включают и собранные самостоятельно компьютеры. Расхождения в оценках между агентствами невелики, и использование максимальной оценки вместо средней не привело к существенному снижению обсуждаемых расхождений в размере парка ПК. Наконец, причиной различий может быть ошибка в определении нами плотности вероятности выхода из строя компьютера. Средний срок службы ПК может оказаться выше, чем оценка в 6,12 года, полученная нами путем моделирования ответов пользователей лишь тех компьютеров, которые используются для выхода в Интернет. Не исключено, что компьютеры, не подключенные к Интернету, а используемые в производстве или для ведения личных финансов и набора текстов, используются в среднем значительно дольше. Заметим, однако, что наша оценка отнюдь не

является низкой: в работе [199] предполагалось, что период обновления парка ПК составляет 5 лет, а согласно работе [36] серверный парк обновляется в среднем тоже один раз в 5 лет.

Альтернативную оценку компьютерного парка можно получить, опираясь на итоги ежегодного федерального государственного статистического наблюдения, которые доступны на сайте Росстата Gks.Ru в сборниках 'Российский статистический ежегодник' за 2001-2009 гг. Это наблюдение охватывает крупные и средние организации, однако, к сожалению, не учитывает субъекты малого предпринимательства, поэтому его итоги, вместе с результатами выборочного обследования домохозяйств Росстатом, позволяют вычислить лишь нижнюю оценку парка ПК. Она оказывается немного ниже нашей оценки, представленной на рис.3.19, причем последнее доступное значение относится к 2009 году и составляет 37,2 млн. компьютеров.

Несмотря на сделанные замечания относительно численных значений, наша модель обладает преимуществами внутренней согласованности, и не допускает расхождений между продажами и парком ПК, продемонстрированных на рис.3.19.

Мы полагаем, что совместное моделирование данных о продажах компьютеров в РФ и размере компьютерного парка позволяет получать более реалистичные прогнозы обеих величин, избежать грубых рассогласований, подобных продемонстрированным на рис.3.15(с)), а также исследовать влияние регулирования рынка ПК на парк компьютеров.

3.2.2 Прогнозирование парка и рынка ПК, планшетных компьютеров и сотовых телефонов

Здесь разработанная в предыдущем разделе модель модифицирована и применена для практического прогнозирования парка и рынка ПК в России. Главное отличие заключается в том, что при прогнозировании используется не оценка продаж $q(t)$, выполненная представителями компаний-производителей компьютеров, а величина $\frac{dN}{dt}$, оцененная при помощи модели распространения нововведения в домохозяйствах, разработанной в предыдущей главе.

Основным фактором роста компьютерного рынка за последние десять лет выступает спрос домашних хозяйств (а не организаций), которые в условиях роста доходов смогли позволить себе использование Интернета, компьютерных игр, видео и иных услуг, предоставляемых растущим числом поставщиков. По нашим расчётам в 2003 году домохозяйства имели приблизительно столько же компьютеров, сколько и корпорации, однако в 2012 году во владении домохозяйств находится 53-56 миллионов ПК, в то время как корпорации используют лишь около 10 миллионов ПК. Таким образом, хотя корпорации обновляют свой парк чаще, покупки частных лиц составляют основную часть продаж компьютерного рынка.

Российский рынок ПК в 1996-2007 гг. демонстрировал устойчивые темпы прироста около 21% в эти годы, в последнем квартале 2008 года финансовый кризис вызвал резкое падение продаж, и в 2009 г. рынок провалился на 30%, однако провал был вскоре компенсирован резким ростом на 48% в 2010 году. Далее последовали относительно благополучные для компьютерных компаний 2011 и 2012 год, однако в последнем квартале последнего началось замещение ноутбуков, и в первую очередь – нетбуков, планшетными компьютерами. Наш прогноз на 2012 год, составивший 13,9 миллионов устройств оказался посередине между данными измерений Gartner (13,6 млн.) и IT-Research (14,0 млн.)

Мы предполагаем, что в 2013-2016 г. рынок по-прежнему будет определяться спросом со стороны домохозяйств, и мы продолжаем оценивать потенциал проникновения компьютеров (включая любые будущие устройства, способные использовать Интернет, предоставлять возможность играть в цифровые игры, смотреть видео и обрабатывать документы) также в 90%.

В основе наших расчетов компьютерного рынка лежит модель динамики распространения ПК в российских домохозяйствах и организациях, аналогичная модели распространения мобильных телефонов, разработанной в предыдущих разделах данной главы. Для каждого из сегментов рынка – домашнего и корпоративного мы строим отдельную модель компьютерного парка, при помощи

которой затем рассчитываем величину рынка. Компьютерный парк РФ по нашей оценке в 2012 году составил 52,5 миллионов устройств, и в 2013 году составит 59,5 устройств (см. рис.3.21).



Рис.3.20. Компьютерный рынок России: осредненные опубликованные данные IDC, IT-Research и Gartner, а также наши прогнозы.

При моделировании спроса домохозяйств на компьютеры мы используем широко применяемую модель диффузии инноваций Ф.Басса и подбираем ее параметры таким образом, чтобы удовлетворить ежегодно публикуемым данным опросов Левада-центра о наличии или отсутствии компьютера в российских семьях. Кроме того, как свидетельствуют результаты обследований бюджетов домохозяйств Росстатом, некоторая доля российских семей, владеющих ПК, с течением времени приобретает второй и третий компьютеры. Это следует из того, что на 100 домохозяйств в 2012 году приходилось 86 компьютеров, в то время как компьютерами обладали лишь 61% домохозяйств. По нашим оценкам 50% домохозяйств однажды приобретает второй компьютер в дополнение к уже работающему первому, 25% - приобретает и третий, а соответствующие

временные лаги – 3 года (для второго дополнительного компьютера) и 6 лет (для третьего). Покупки дополнительных компьютеров также включены в нашу модель, что позволило нам учесть как данные Левада-центра о доле ПК в семьях, так и данные Росстата.

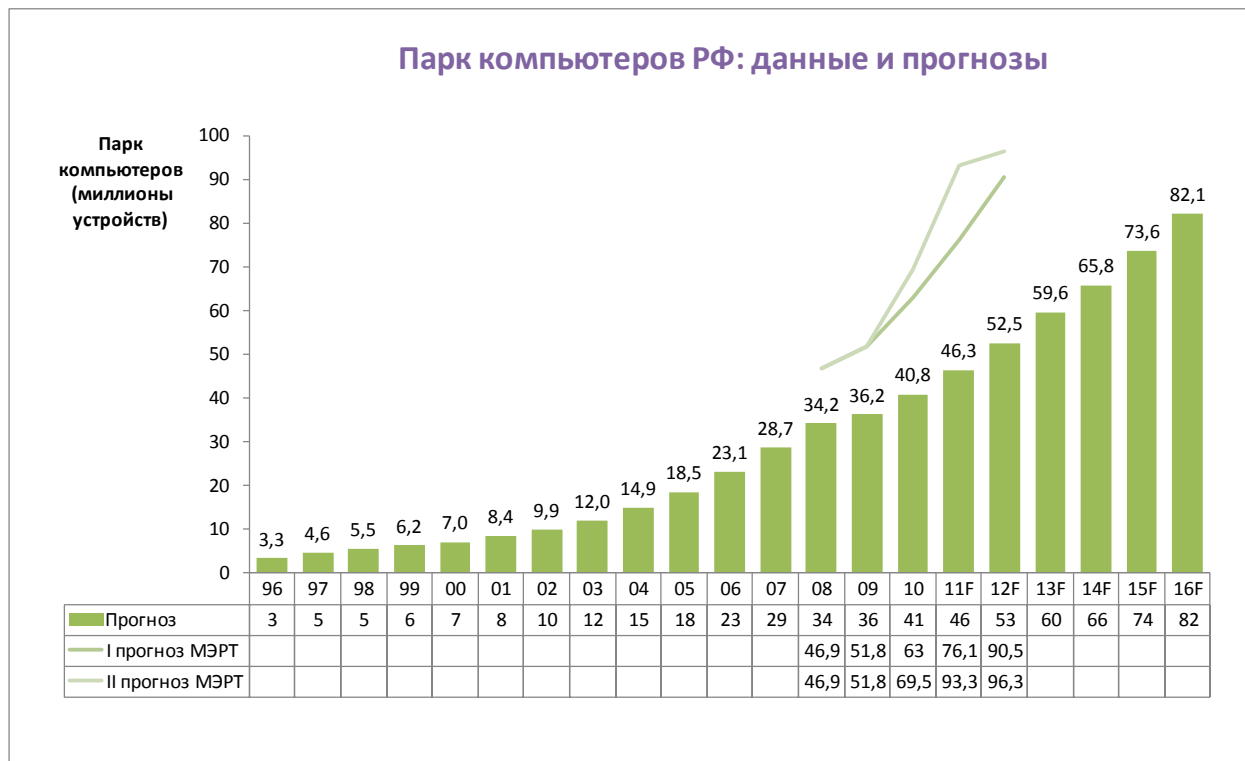


Рис.3.21. Парк компьютеров России: данные, прогнозы Министерства экономического развития и наш прогноз.

Что касается корпоративного сегмента рынка, то мы предполагаем, что спрос на компьютеры со стороны российских организаций продолжит расти, однако понижаем темпы прироста с 16% в год до 7% в год (на основе собственной экспертной оценки).

Суммируя парки домашних и корпоративных компьютеров, мы получаем итоговый прогноз парка ПК, изображенный на рис.3.21. Заметим, что этот парк не включает планшетные компьютеры, парк которых, по нашим расчётам в 2016 году составит 33,6 миллионов устройств, что в 8 раз выше, чем по итогам 2012 года. Итогом расчетов является прогноз продаж компьютеров на российском рынке до 2016 года, который изображен на рис.3.20.

На рис.3.20 представлен также наш прогноз рынка планшетных компьютеров. Мы исходили из предположения, что ранними покупателями планшетных компьютеров являются обеспеченные лица, которые способны без ущерба для своего бюджета приобретать дополнительные устройства. Таким образом, их покупки расширяют рынок на 3-4 миллиона устройств. Покупка планшетного компьютера не влияет на их решения о покупке дополнительных ноутбуков или настольных компьютеров. Однако в дальнейшем покупки совершают менее обеспеченные лица, которые ради приобретения планшетного компьютера отказываются от приобретения ноутбука. Таким образом, покупки планшетов снижают объём продаж ноутбуков, хотя почти не сказываются на общем размере компьютерного парка. По нашим прогнозам основной рост продаж планшетных компьютеров придётся на 2013 год, после чего продажи планшетников в значительной степени стабилизируются.

Продажи сотовых телефонов в России

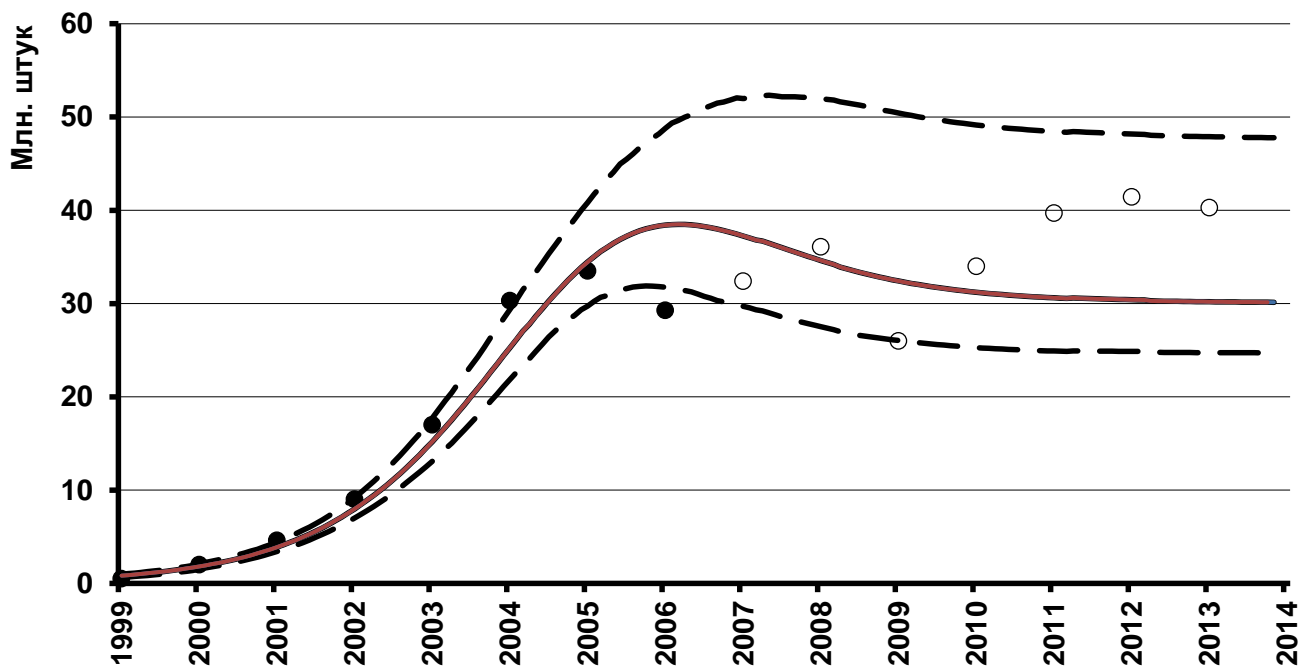


Рис.3.22. Продажи сотовых телефонов в России: данные компании “Евросеть” и наш прогноз. Пунктиром показаны 95%-е доверительные интервалы.

Наконец, на рис.3.22 представлен наш прогноз рынка сотовых телефонов. Сплошными кружками показаны данные по 2006 г. включительно, которые были опубликованы компанией “Евросеть” и использованы нами при идентификации параметров модели. Открытыми кружками показаны аналогичные данные, опубликованные в 2007-2013 гг. Сплошной линией показан наш прогноз продаж сотовых телефонов в России. Пунктирными линиями показаны 95%-е доверительные интервалы, рассчитанные методом Монте-Карло; способ расчёта описан в работах [3,344] и в кандидатской диссертации автора [163].

На рис.3.20 и 3.22 обращают на себя внимание значительные “провалы” рынков ПК и сотовых телефонов: “обрушения” рынка могут достигать 25% от среднего значения. Такие краткосрочные эффекты не учитываются в моделях распространения нововведений, хотя они чрезвычайно важны для игроков компьютерного рынка. Мы полагаем, что полезные для практиков краткосрочные прогнозы могут быть получены путём применения моделей, распространения нововведений с экспертными поправками за влияние краткосрочных факторов.

Описанные здесь прогнозы парков и рынков технических устройств были использованы ЗАО ФИНАМ при подготовке ряда аналитических отчётов.

Выводы

1. Разработанные модели, учитывающие наличие в обществе не только индивидов, но и домохозяйств, способны объяснить данные массовых социологических опросов об использовании мобильной связи и персональных компьютеров. Единицей, принимающей решение о приобретении нового устройства и первым его владельцем первоначально является глава домохозяйства, а последующие члены домохозяйства получают устройство в личное владение с задержкой.
2. Разработана модель роста парка технических систем с учетом данных маркетинговых исследований о размере рынка и онлайн-опросов о сроке службы устройства, которая применена для прогнозирования российского парка компьютеров, планшетов и сотовых телефонов.

ГЛАВА 4. МОДЕЛЬ ВЫБОРА НОВЫХ ИКТ ДЛЯ ИНВЕСТИРОВАНИЯ ВЕНЧУРНЫМ ФОНДОМ

При венчурном инвестировании в инновационные проекты в качестве единицы, принимающей решение об использовании нововведения, выступает не индивид или домохозяйство, а организация специального вида – *венчурный фонд*. Л.В.Ивина и В.А.Воронцов предлагают следующее толкование этого термина: “«фонд венчурного капитала» – фонд (организация), аккумулирующая частные финансовые средства для инвестирования в высокорисковые высокотехнологичные инновационные проекты. Фонд вкладывает средства путём приобретения акций новой компании на ограниченный срок (до тех пор, пока рынок не будет способен вернуть вложенный капитал с существенной прибылью, но обычно не более чем на 7 лет)” [175]. Те же авторы определяют *startup* (start-up) как только что возникшую компанию [175].

В наших работах [76,77,81,91,89,88,158] предложен инструмент выбора (отбора) инновационных проектов в области ИКТ, построенный на основе матрицы первичной оценки бизнес-идеи. П.Л.Виленский и др. определяют *проект* как “комплекс допустимых действий (работ, услуг, управленческих операций и решений), обеспечивающих достижение определенных целей (получение определенных результатов)” [39]. Авторы определения отмечают, что в теории управления проектами под проектом понимается совокупность действий, исполнителей и средств по выработке определенных целей и их достижению.

В данной главе изучена две проблемы, регулярно встречающиеся в деятельности венчурных фондов:

1. проблема низкого качества фильтра нижнего уровня системы (воронки) отбора высокотехнологичных проектов, которая решена при помощи методов теории распознавания образов;
2. проблема длительного обслуживания заявок на нижнем уровне системы отбора инновационных проектов, которая решена при помощи инструмента

поддержки принятия решений, позволяющего аналитику обрабатывать заявки и давать заявителям предварительные ответы в течении одного часа.

Мы применяем к задаче отбора проектов (стартапов) известный в теории распознавания образов метод *обучения с учителем*; в роли последнего в нашей работе выступает портфель проектов фонда, в отношении которых лица, принимающие решения, уже сделали выбор. В работе [153] автором и Н.С.Мациевским при помощи этого инструмента проанализированы проблемы венчурного инвестирования в проекты предоставления программного обеспечения в качестве услуги (“software-as-a-service”).

Результаты исследования внедрены ЗАО ФИНАМ при отборе венчурных проектов в сфере ИКТ в 2006-2013 гг., что подтверждается Актом о внедрении от 24.12.2013 (см. приложение Б-4).

4.1 Модель принятия решения об инвестировании в информационные и телекоммуникационные технологии венчурным фондом

Венчурные фонды, наряду с центрами трансфера технологий, технопарками, бизнес-инкубаторами и центрами подготовки кадров, составляют инфраструктуру национальной инновационной системы. В Концепции долгосрочного социально-экономического развития Российской Федерации на период до 2020 года [198] развитие механизмов венчурного финансирования в сфере информационно-коммуникационных технологий рассматривается как часть стратегии развития науки, национальной инновационной системы и технологий. Ожидается, что высокие прибыли венчурных инвесторов будут стимулировать привлечение внебюджетных средств в науку и инновации.

Венчурное инвестирование можно определить как долгосрочное предоставление средств молодым компаниям, находящимся на ранних стадиях развития, в обмен на долю в этих компаниях [182]. Поскольку будущее новых компаний, в особенности, – их финансовые показатели, характеризуются

чрезвычайно высокой неопределенностью, потенциальные инвесторы редко обладают всей информацией, необходимой для использования традиционных инструментов оценки эффективности инвестиций. Для успешного вложения средств венчурному инвестору необходимо информационно-аналитическое обеспечение, в первую очередь – системно-аналитические методы отбора²⁴ проектов ранних стадий (“стартапов”). В этой главе мы не выйдем за рамки задач венчурного фонда и не будем делать различий между терминами “проект” и “стартап”, некоторые определения которых были даны выше.

В настоящей работе рассмотрен инструмент отбора инновационных проектов, построенный на основе матрицы первичной оценки бизнес-идеи (МПОБ) [15,16,27] с учетом специфики проектов, поступающих на рассмотрение венчурного фонда. Фонд ФИНАМ ИТ, который послужил базой для результатов, излагаемых в этой работе, фокусируется на молодых компаниях и проектах, работающих в области информационных и телекоммуникационных технологий (ИКТ). На рассмотрение инвестиционного комитета фонда каждый год поступает несколько сотен проектов, основатели которых рассчитывают в перспективе получить высокую прибыль благодаря успешной реализации новой идеи. Наиболее значимыми активами в структуре фонда в 2006-2008 гг., к которым относятся данные, послужившие экспериментальной базой разработанного нами построения инструмента, являлись доли в компаниях “Мамба” и “Бука”. Краткое описание этих компаний позволяет составить представление о характерном масштабе стартапов и проиллюстрировать случаи, в которых фонд считает задачу выбора удачно решенной.

²⁴ Термин “отбор” не следует ассоциировать с “отъемом”, “захватом”, “отчуждением” или экспроприацией. Толковый словарь Д.Н.Ушакова определяет отбор как действие по глаголу “отобрать”, но не в значении принудительного изъятия (отнять), а в значении выделения, выбора некоторого количества из общего числа. В толковом словаре В.И.Даля приведены синонимы – “выбирать”, “отделять”, “сортировать” или “класть особо”. В качестве примера употребления приведён “отборный товар”. Среди синонимов и сходных по смыслу выражений можно назвать “выбор”, “выделение”, “селекцию”, “кастинг” и “скрининг”. Последний термин часто употребляется в отношении стартапов.

Пример инвестированной компании: “Бука”

“Бука” (Buка Ltd.) – издатель компьютерных игр, входил в тройку лидирующих игроков рынка. Объем выручки компании “Бука” в 2007 г. составил около 40 млн долларов, показатель EBITDA – более 14 млн долларов. Годовые темпы прироста оборота компании в 2007 г. превышали 50%. В 2008 г. фонд продал “Буку” компании “1С”, лидеру российского рынка компьютерных игр.

Пример инвестированной компании: ЗАО “Мамба”

ЗАО “Мамба” - интерактивный сервис онлайн-знакомств с широкой партнерской сетью, где пользователи оплачивают услуги (в основном – по рекламе своих анкет) при помощи SMS-сообщений, отправляемых через операторов сотовой связи. Основные партнеры – проекты Mail.Ru и Rambler. Оборот компании в 2007 г. составил 405 миллионов рублей, темпы прироста оборота - около 50% в год, рентабельность EBITDA – от 47% до 50%. Фонд планирует продажу “Мамбы” после достижения компанией стоимости в 4-5 млрд рублей.

Необходимость разработки формализованной системы отбора стартапов вызвана ростом интенсивности потока проектов, поступающих на рассмотрение инвестиционного комитета фонда. В частности, обсуждаемый в этой работе инструмент отбора (скрининга) был применён при обработке полутора тысяч инвестиционных заявок, поступивших на рассмотрение фонда ФИНАМ ИТ в 2006-2012 гг. Ни руководители фонда, ни члены инвестиционного комитета не имеют возможности тщательно рассмотреть все заявки. Поэтому венчурный фонд нанимает одного или несколько аналитиков, которые при отборе заявок должны репрезентировать инвестиционный комитет и выбирать проекты так, как это делал бы комитет, если бы располагал необходимым для этого временем. На практике аналитики зачастую имеют существенно иное видение перспектив развития индустрии ИКТ и иные предпочтения [261]. Необходимо подчеркнуть, что аналитик не является и не может являться экспертом во всех направлениях ИКТ, поскольку спектр этот чрезвычайно широк. Наиболее популярные сейчас

направления деятельности стартапов – электронная коммерция [149], социальные сети, поисковые, “облачные”, рекламные технологии, технологии видеоконференций, дистанционное образование – слишком различаются, чтобы один эксперт мог глубоко разбираться в соответствующих технологиях и бизнес-моделях. Поэтому требуется разработка методики, которая позволяла бы аналитику руководствоваться не только личным опытом и соображениями “здорового смысла”, но, в первую очередь, – предпочтениями инвестиционного комитета. Сами эти предпочтения могут не быть изложены в виде инструкций, не быть сформулированы или быть сформулированы расплывчато, однако их можно выявить путем анализа портфеля проектов, в которые фонд инвестировал ранее.

Многие предложения, являясь инновационными в широком смысле, не соответствуют интересам венчурных инвесторов. Частичная формализация показателей и критериев, которыми руководствуется венчурный фонд при отборе проектов, делает его деятельность более понятной для предпринимателей и изобретателей, которые предлагают свои проекты на рассмотрение. В результате, предложение адаптируется к спросу, и работа обеих сторон становится эффективнее.

Система отбора инновационных проектов венчурным фондом (рис.4.1) включает организационную и информационную составляющие. Организационная подсистема включает мероприятия по сбору и обработке данных (в том числе первичный отбор, т.н. *первые встречи* и последующие встречи с авторами предложений), обсуждение проектов на совещаниях инвестиционного комитета, процедуру принятия решения о необходимости сбора дополнительных данных или о переходе к стадии подписания соглашения о конфиденциальности и тщательного изучения компании (“due diligence”). На каждой из изображённых на рис.4.1 стадий число рассматриваемых фондом проектов, сокращается на порядок, что иллюстрируют при помощи воронки или “ракеты Дрейпера” [182].

На каждой стадии с проектом работают различные специалисты: на этапе предварительного изучения – это бизнес-аналитики, на этапе тщательного

изучения возрастает роль юристов и специалистов по бухгалтерскому учету, экспертов по ИКТ, решения принимаются инвестиционным комитетом [150].

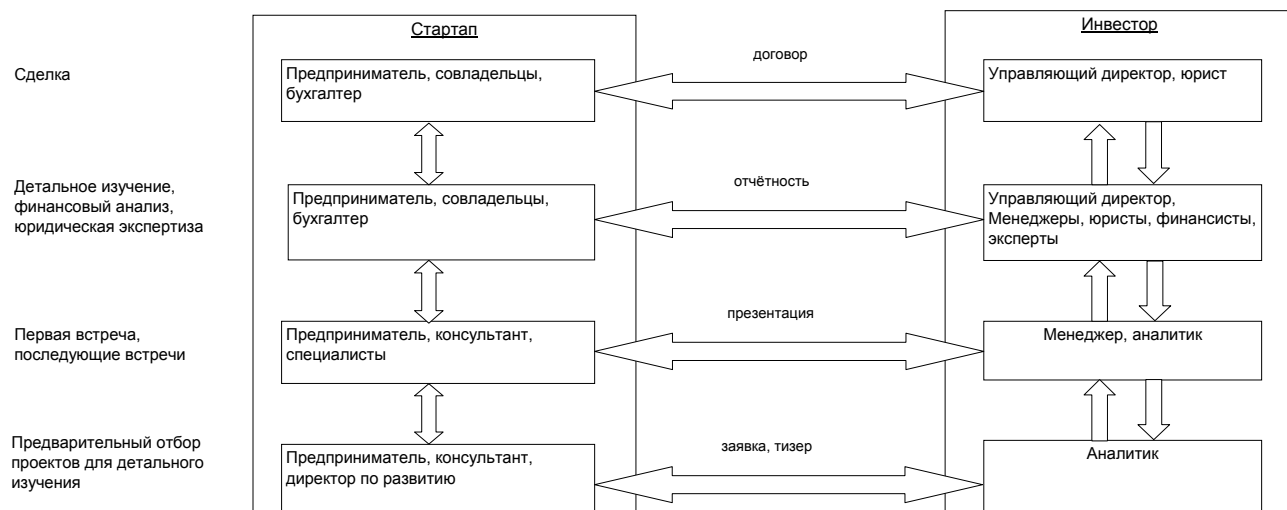


Рис.4.1. Система отбора инновационных проектов

Указанные группы специалистов, хотя отчасти и пересекаются, но не совпадают, поэтому важно, чтобы их действия были согласованы. Наиболее важно, чтобы первичный отбор проектов репрезентировал предпочтения инвестиционного комитета, а не собственные симпатии аналитиков²⁵.

Матрица предварительной оценки бизнес-идеи

Метод отбора проектов на основе матрицы предварительной оценки бизнес-идеи (МПОБ) принадлежит к классу т.н. балльных моделей и ставит целью быстро, силами малого числа подготовленных аналитиков оценить перспективность большого числа деловых идей [15,16,27]. Процесс расчета

²⁵ Возникает вопрос, действительно ли предварительным отбором предложений занимается или должен заниматься специальный работник (аналитик), а не руководство фонда? Можно заметить, что на мероприятиях, посвященных открытому отбору стартапов, или на специальных телевизионных программах, судьями на конкурсах инновационных проектов выступают узнаваемые “первые лица” – руководители фондов. Однако список финалистов, как правило, составляют отборочные комиссии. Вряд ли слишком смелым будет предположение, что успешные руководители рационально расходуют свое время и умело делегируют полномочия. В общем случае, руководитель работает сам там, где высока “цена вопроса”. В случае венчурного фонда можно считать *срочными и важными* задачи инвестирования крупных денежных средств, и другие вопросы, которые не будучи решенными, повлекут большие потери или значительную упущенную выгоду. Согласно модели делегирования полномочий в виде так называемой “матрицы Эйзенхауэра”, популяризованной С.Кови [188], руководитель передает сотрудникам задачи, которые классифицирует как *срочные, но не важные*. Предварительный отбор (отсев) предложений можно считать срочной, но не важной задачей, и в венчурных фондах его зачастую поручают аналитику. В отличие от венчурной компании, т.н. “бизнес-ангел” – ветеран индустрии, инвестирующий в стартапы личные средства, отсеивает предложения самостоятельно.

интегральной оценки привлекательности стартапа по качественным показателям на основе МПОБ можно изобразить при помощи блок-схемы (рис.4.2).

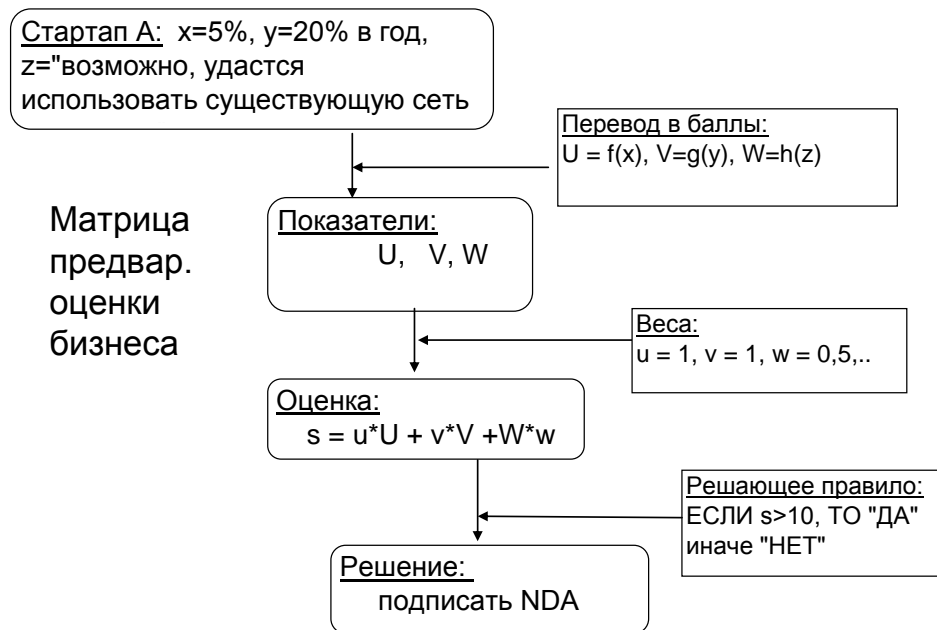


Рис.4.2. Алгоритм предварительного отбора проекта

Методика предварительной оценки бизнес-идеи требует расчета шести агрегированных показателей, каждый из которых является суммой 3-5 более простых (элементарных) показателей. В таблице 4.1 перечислены агрегированные показатели, и для двух из них – “привлекательности рынка” и “синергии бизнеса” – указаны все элементарные показатели. Остальные элементарные показатели подробно описаны в статье [15] и гл. 3 монографии С.Я. Бабаскина [16]. Элементарные показатели могут быть измерены, оценены или рассчитаны экспертом. В этой работе мы будем для краткости называть элементарные показатели “признаками”, хотя они не всегда являются результатами непосредственных измерений.

Первоначально набор из 16 признаков, разделенных на 6 групп (по 2-3 признака на группу), был опубликован в работе А.Бретта [27] и предназначался для оценки факторов технологического риска потенциальных коммерческих возможностей. А.Бретт допускал возможность исключения или замены некоторых

признаков и добавления небольшого числа новых. Двенадцать новых признаков были добавлены в работах С.Я.Бабаскина [15,16]. Расширенный С.Я.Бабаскиным набор признаков внедрен в корпоративной системе “4И” сбора и обработки предложений сотрудников корпораций, разработанной под руководством В.В.Матюхина в корпорации Текора [219].

Таблица 4.1

Матрица предварительной оценки бизнеса [16,27]

№	Показатель		Диапазон изменений
1	Привлекательность рынка (5 признаков)		-10..10
	1.1.	Рыночные перспективы	-2..2
	1.2.	Темпы прироста рынка	-2..2
	1.3.	Барьеры для вхождения на рынок	-2..2
	1.4.(*)	Уровень конкуренции	-2..2
	1.5.(*)	Прибыльность сегмента рынка	-2..2
2	Выгоды пользователя (5 признаков)		-10..10
3	Обоснованность идеи (5 признаков)		-10..10
4	Синергия бизнеса (4 признака)		-8..8
	4.1.	Соответствие миссии компании	-2..2
	4.2.	Возможность использования существующей клиентской базы	-2..2
	4.3.	Достаточность знаний и опыта	-2..2
	4.4.(*)	Возможность использования существующего производства и системы распределения	-2..2
5	Достаточность ресурсов (4 признака)		-8..8
6	Охрана идеи (3 признака)		-6..6

Заметим, что иерархические системы показателей, каждый из которых представляет собой сумму или среднее арифметическое элементарных признаков,

часто встречаются в задачах отбора или отсева. Так, например, в работе Ю.И.Погорлецкой, посвященной кредитному скорингу в системе банковского риск-менеджмента [243], используются пять показателей, которые характеризуют 1) финансовое состояние эмитента долговой бумаги; 2) экономическое состояние эмитента, 3) деловую репутацию, 4) маркетинговые позиции, 5) организационную структуру и уровень управления. При оценке финансового состояния агрегируются семь групп признаков. Остальные четыре показателя рассчитываются при помощи 4-6 элементарных признаков по формуле средней арифметической. По данным С.В.Уланова в кредитном скоринге количество признаков составляет от 6 до 18 [275]. Можно поставить вопрос о минимально необходимом или оптимальном количестве признаков. По-видимому, 20-30 признаков на практике оказываются достаточными для решения задач отбора/отсеивания компаний или проектов в разумные сроки²⁶.

Инвестиционный комитет венчурного фонда, выступающий в роли лица, принимающего решение (ЛПР), или, точнее, единицы, принимающей решение (ЕПР), может использовать МПОБ в качестве одного из инструментов отбора проектов. (В этой работе мы не принимаем во внимание различия между ЛПР и ЕПР, предполагая, что инвестиционные комитеты венчурных фондов состоят из единомышленников). Инструмент должен преобразовывать оценки отдельных признаков, выполненные аналитиками, таким образом, чтобы полученная интегральная оценка соответствовала предпочтениям ЛПР.

Модели функции полезности, используемой ЛПР

Помимо показателей, необходимо определить вектор весов \vec{w} , который описывает важность отдельных показателей. Традиционный подход к задаче выбора использует линейную аддитивную модель функции полезности, в которой общая полезность бизнес-идеи представлена в виде взвешенной суммы оценок полезности отдельных показателей:

²⁶ Предлагаемая ниже диаграмма “Привлекательность – неопределенность бизнес-идеи” использует лишь 19 признаков из 28 признаков, предложенных С.Я.Бабаскиным, причем один из признаков (“отрицательные последствия для окружающей среды”) в сфере ИКТ используется “вхолостую”. Таким образом, в нашем способе отбора проектов венчурным фондом “работают” только 18 признаков.

$$\begin{aligned}
 U(\mathbf{X}) &= \sum_{k=1}^K \alpha_k U_k(\mathbf{X}) \\
 U_k(\mathbf{X}) &= \sum_{i=1}^N w_i u_{ki} \\
 u_{ki} &= \sum_{j=1}^{M_i} U_{kij}(x_{ij})
 \end{aligned} \tag{1}$$

где $U(\mathbf{X})$ – суммарная оценка полезности бизнес-идеи \mathbf{X} ; $U_k(\mathbf{X})$ – полезность бизнес-идеи \mathbf{X} , измеренная k -м экспертом; u_{ki} – оценка привлекательности проекта по i -му агрегированному показателю, данная k -м экспертом; x_{ij} – величина j -го признака (элементарного показателя) в составе i -го агрегированного показателя; M_i – количество признаков, образующих i -й агрегированный показатель; $U_{kij}(x_{ij})$ – полезность признака x_{ij} , измеренная k -м экспертом; w_i – вес i -го показателя; α_i – вес мнения i -го эксперта ($\sum_{i=1}^K \alpha_i = 1$); N – число агрегированных показателей, K – число экспертов.

Описанная модель явным образом учитывает различия важности мнений отдельных экспертов, при этом, однако, мнению каждого эксперта о полезности каждого частного признака присваивается одинаковый вес. Вес не всегда отражает реальные предпочтения, – в частности, финансист и технолог могут иметь противоположные точки зрения на полезность скорой окупаемости проекта, и это различие во взглядах нередко завершается “изгнанием” технолога из компании. Некоторые авторы рекомендуют учитывать ценность эксперта при помощи векторной функции, а не одного лишь скалярного веса, поскольку эксперты способны точно измерять лишь часть признаков [277].

Подчеркнем, что сами признаки x_{ij} объективно существуют, не зависят от экспертов и становятся доступны экспертам в случае успешного сбора данных. Однако функции полезности $U_{kij}(x_{ij})$ различных экспертов, безусловно, могут отличаться, что привносит субъективизм отдельных оценок в традиционные балльные методы и требует погашения отклонений путем осреднения оценок

экспертов. Сами эксперты сообщают ЛПР не оценки значений признаков x_{ij} (в миллионах рублей, годах или процентах), а только свои оценки их полезности $U_{kij}(x_{ij})$ в баллах, причем форма функций полезности может быть неизвестна лицу, принимающему решения.

Обычно используется один из двух основных критериев выбора предложений:

1. Принимается предложение \mathbf{X} с максимальной полезностью:

$$U(\mathbf{X}) = \max(U(\mathbf{Y}), U(\mathbf{Z}), \dots). \quad (2)$$

2. Предложение \mathbf{X} принимается, если его полезность превосходит заранее заданную допустимую (“пороговую”) величину

$$U(\mathbf{X}) > U_{crit}. \quad (3)$$

Отсортировав проекты по $U(\mathbf{X})$, можно получить т.н. “рейтинг”, который часто используется в качестве инструмента поддержки принятия решений.

Особенности предлагаемого подхода

Отметим три основных отличия нашего подхода от традиционного [15,16,27]. Во-первых, мы “настраиваем” систему показателей под нужды венчурного ИКТ-фонда, установив соответствие между значениями элементарных показателей (признаков) и их балльными оценками. Мы предположили, что такая процедура позволяет ослабить влияние субъективизма оценок, а также случайные вариации оценок во времени, их зависимость от кратковременных предпочтений и настроений аналитика (который в марте может быть захвачен одной “горячей” идеей, а в апреле – уже совершенно другой). В работе автора, написанной совместно с Е.А.Кобзаревой обсуждается попытка экспериментальной проверки этой процедуры [158]. Мы также предполагаем, что для более или менее сходных проектов заранее определенное соответствие между значениями признаков и количеством баллов может облегчить работу эксперта, при этом если исходные признаки x_{ij} близки, то близки окажутся и балльные оценки различных экспертов $U_{kij}(x_{ij})$.

Во-вторых, мы визуализируем структуру системы показателей в виде двумерной карты и отмечаем стартапы точками на этой карте (см.рис.4.4,4.5,4.7).

В-третьих, мы отказываемся от нормативного задания критерия отбора. Мы рассматриваем задачу отбора как задачу *обучения с учителем*, где в качестве обучающих множеств выступают выборка проектов, ранее отобранных фондом для тщательного изучения (“due diligence”), и выборка проектов, отклоненных фондом без детального изучения²⁷. Целью применяемого в данной работе метода когнитивного моделирования является построение модели процесса принятия решений менеджментом [419]. Зачастую опытный инвестор руководствуется опытом и интуицией, не формализуя критерии выбора альтернатив, либо не объясняя эти критерии аналитикам. В то же время, аналитикам доступна обучающая выборка проектов, которые ЛПР уже принял к рассмотрению. Построенная на основе этой выборки модель принятия решений становится инструментом прогнозирования будущих решений ЛПР, и, в случае задачи отбора проектов, позволяет оценить целесообразность тщательного изучения проекта или дополнительного сбора информации о проекте. С течением времени, по мере накопления больших массивов данных, возможно выделение из первой обучающей выборки подмножества финансово успешных проектов.

Перевод экспертных оценок в баллы

Традиционный способ предназначен для отбора чрезвычайно сильно различающихся проектов, при этом от экспертов требуются глубокие и специфические знания, а балл, присваиваемый экспертом, имеет самостоятельную ценность. Однако венчурный фонд сталкивается с задачей выбора проекта среди десятков однотипных, очень похожих друг на друга представителей общего класса, для которого уже набрана статистика показателей. Мы предлагаем

²⁷ Связь между распознаванием образов и принятием решений обсуждалась уже в классическом труде К.Фукунаги “Интуитивно ясно, что до некоторой степени процесс принятия решений имеет отношение к распознаванию образов; например, в шахматной игре следующий ход делается в зависимости от ситуации (образа), сложившейся в данный момент времени на шахматной доске; решение о том, покупать или продавать акции на бирже также принимается в результате анализа сложного информационного образа. Поэтому целью создания теории распознавания образов являлось выявление сложных механизмов процессов принятия решений, а также автоматизация этих процессов с помощью средств вычислительной техники. Однако ввиду сложности проблемы распознавания образов, основные исследования в этой области были сосредоточены на более реальных задачах, таких как распознавание букв латинского алфавита и классификация кривых” [283].

устанавливать соответствие между значениями отдельных признаков и “общепринятыми” баллами, которые добавил бы за такие значения “типичный” компетентный специалист, репрезентирующий мнение инвестиционного комитета фонда. Соответствие имеет вид:

$$U_{kij} = U_{ij}(\hat{x}_{kij}), \quad (4)$$

где \hat{x}_{kij} – оценка признака проекта, измеренная k -м экспертом (которая может быть количественной или атрибутивной и измеряется в денежных единицах, годах, процентах и т.п.); U_{ij} – полезность значения признака, выраженная в баллах одинаковым способом для всех аналитиков.

Так, например, оценивая проект по показателю “темпы прироста рынка” (см. таблицу 4.1), мы добавим два балла проектам, работающим на рынках с годовыми темпами прироста более 100%, один балл – при темпах от 50 до 100%, ноль баллов – от 20 до 50%; мы вычтем один балл при темпах в 10-20% и два балла - при годовых темпах прироста менее 10%.

Оценка функции полезности ЛПР приобретает вид

$$U(\mathbf{X}) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^{M_i} w_{ik} U_{ij}(x_{kij}). \quad (5)$$

Теперь, если два аналитика полагают, что рынок, на который намерен выйти со своей услугой автор проекта, растет на 15% в год, то оба вычтут из оценки привлекательности рынка один балл. При отсутствии единой системы соответствия один эксперт мог бы счесть темпы прироста рынка “высокими” и добавить проекту один балл, а другой – “чрезвычайно низкими”, и вычесть у проекта два балла. Общие U_{ij} устраняют такой разброс.

Цель введения общих U_{ij} состоит в том, чтобы от субъективных оценок (таких как “рынок растёт чрезвычайно быстро”), которые могут различаться у различных экспертов, перейти к количественным показателям, которые разные эксперты при наличии одинаковой информации оценят единообразно. Разумеется, при таком подходе может потеряться “здоровый” разброс мнений экспертов

относительно того, сколько баллов следует присваивать тому или иному значению признака. Составитель руководства по измерению функций полезности может навязать экспертам свою собственную функцию U_{ij} , не близкую к соответствующей функции ЛПР. Поэтому руководства по измерению функций полезности должны составляться с участием ЛПР.

Анализ обучающих данных

При помощи МПОБ нами были проанализированы результаты отбора 150 идей, проектов и компаний, рассмотренных венчурным фондом ФИНАМ ИТ в первой половине 2007 г. Средняя оценка $U(X)$ коммерческой привлекательности предложений, отобранных фондом для тщательного изучения («due diligence»), оказалась более чем вдвое выше, чем средняя оценка произвольного предлагаемого проекта. Таким образом, оценка, полученная с помощью МПОБ, в целом согласуется с решениями инвестиционного комитета венчурного фонда.

Факторный анализ матриц МПОБ позволил нам разделить агрегированные показатели на две группы, соответствующие двум главным компонентам. На рис.4.3 первой главной компоненте соответствует вертикальная ось, второй – горизонтальная. Обнаружилось, что показатели “привлекательность рынка” и “выгоды для пользователя” имеют высокие значения первой главной компоненты и низкие значения второй. Напротив, показатели “обоснованность идеи” и “синергия бизнеса” имеют высокие значения второй главной компоненты и низкие – первой.

Показатель “достаточность ресурсов” имеет высокие значения обеих главных компонент. Это логично, поскольку недостаток ресурсов, с одной стороны, не позволяет преодолеть барьеры для входа на рынок и использовать экономию от масштаба, а с другой – увеличивает вероятность провала проекта. Показатель “охрана идеи” показал отрицательную корреляцию с показателями обоснованности идеи и синергии бизнеса. Мы объясняем это тем, что в интернет-индустрии, к которой принадлежит значительное число рассмотренных проектов, быстро клонируются все полезные (и бесполезные) идеи.

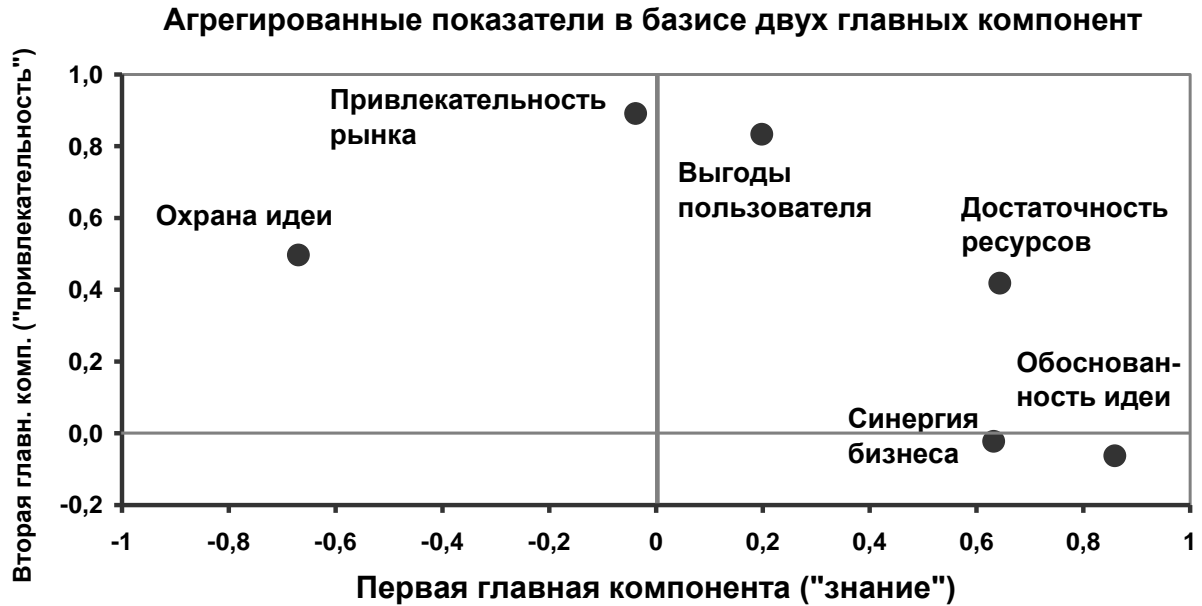


Рис.4.3. Агрегированные показатели в базисе главных компонент

Близость четырёх агрегированных показателей к главным компонентам корреляционной матрицы, иллюстрируемая рис.4.3, обосновывает возможность использовать два комплексных²⁸ показателя вместо одного и сравнить полученный результат с обучающей выборкой, состоящей из проектов, отобранных венчурным фондом для тщательного изучения.

Карта неопределенность-привлекательность

Предлагаемые два комплексных показателя проекта рассчитываются по формулам:

$$S_{mk}(\mathbf{X}) = \sum_{i=1}^N w_{mi} u_{ki}$$

$$u_{ki} = \sum_{j=1}^{M_i} U_{ij}(x_{kij}) \quad (6)$$

²⁸ Под *комплексным показателем* здесь понимается взвешенная сумма показателей табл.4.1, а не комплексное число, состоящее из действительной и мнимой части. Аналогично в теории надежности под комплексным показателем понимают “показатель надежности, характеризующий несколько свойств, составляющих надежность объекта” [ГОСТ 27.002-89]. В качестве примера можно привести коэффициент готовности, который рассчитывается при помощи времени наработки на отказ и среднего времени восстановления работоспособности изделия.

где $S_{mk}(\mathbf{X})$ – m -й комплексный показатель проекта \mathbf{X} , полученный k -м аналитиком; w_{mi} – матрица весов; $i = 1..N$, $m = 1..M$.

Ограничимся только двумя комплексными показателями ($M = 2$). Тогда матрица весов w_{mj} приобретает чрезвычайно простой вид (табл. 4.2).

Таблица 4.2

Матрица весов

i	Показатель	$m = 1$	$m = 2$
1	Привлекательность рынка	1	0
2	Выгоды для пользователя	1	0
3	Обоснованность идеи	0	1
4	Синергия бизнеса	0	1
5	Достаточность ресурсов	0	0
6	Охрана идеи	0	0

Условно назовем первый комплексный показатель (S_{1k}) показателем “**привлекательности идеи**”, на рис.4.4 ему соответствует вертикальная ось. Второй комплексный показатель (S_{2k}) отражает субъективную²⁹ меру экономической *безопасности* инвестора. Можно интерпретировать его как показатель “компетентности” инвестора, подразумевая, что он измеряет достаточность опыта и знаний фонда для инвестирования в предлагаемый проект³⁰. На рис.4.4 этот показатель отложен по горизонтальной оси влево. Противоположный ему показатель можно считать мерой воспринимаемой “**неопределённости**” или риска проекта; на рис.4.4 он откладывается по горизонтальной оси вправо³¹.

²⁹ Подчеркнем, что одной из целей работы является разработка такого инструмента, чтобы субъектом выступал ЛПР, а не аналитик. С целью подчеркнуть субъективный характер оценок, выделенные показатели можно назвать воспринимаемой (субъективной) привлекательностью и воспринимаемой неопределенностью.

³⁰ В финансовой литературе используется термин “кредитная надёжность” заёмщика [243], который, на наш взгляд, близок по смыслу. В отличие от венчурного инвестирования, ожидаемая прибыль кредитора заключена в относительно узких диапазонах. В отличие от стартапа, должник не может принести кредитору десятикратный возврат на инвестиции, поэтому при кредитовании показатель “привлекательности” почти постоянен, и двумерное распределение сводится к одномерному.

³¹ Выбранные нами названия – условны. Мы не утверждаем, что рассчитываемые нами показатели S_{1k} и S_{2k} (и составляющие их признаки) операционализуют концепции “привлекательности” и “неопределенности”, которые

Отметим, что ни в один из двух комплексных показателей мы не включили показатель “достаточности ресурсов”, присутствующий в МПОБ, поскольку считаем, что его следует рассматривать уже после прохождения первичного отбора. Действительно, если проект интересный, но ресурсоемкий, то венчурный инвестор может собрать “пул” инвесторов, либо предложить инноватору разбить проект на несколько стадий. Кажется неразумным, чтобы перспективный проект был без учета мнения ЛПР отклонен аналитиком на стадии предварительного изучения по причине ресурсоемкости.

В оригинальном варианте [15,16] метод МПОБ решает задачу классификации при помощи линейной дискриминантной функции. Мы предлагаем использовать двумерную карту оценок с использованием двух агрегированных показателей (рис.4.4). Каждому стартапу на такой карте соответствует точка. При классификации таких точек, можно не ограничиваться областями в форме полуплоскостей (как в линейном дискриминантном анализе), но использовать области более сложной формы, например, угол. На рис.4.4 прямоугольная трапеция³², в которую попали проекты, отобранные инвестиционным комитетом, помечена надписью “Тщательное изучение”. Тупой угол между правой боковой стороной и основанием трапеции можно описать системой двух неравенств

$$\begin{aligned} S_{1k}(\mathbf{X}) &\geq 5 \\ S_{2k}(\mathbf{X}) &\geq 10 - S_{1k}(\mathbf{X}) \end{aligned} \quad (7)$$

используют ЛПР или аналитики, или определяют другие исследователи. Это связано с тем, что приоритет в нашей работе отдается разработке инструмента поддержки принятия решений, который может использоваться и как “чёрный ящик”. Обоснование валидности измерения “привлекательности” и “неопределенности” при помощи матрицы МПОБ А.Бретта и С.Я.Бабаскина в нашей работе не проведено. На этот недостаток внимание автора обратил д.э.н., профессор О.Г.Голиченко. Мы не исключаем, что валидные показатели собственно “привлекательности” и “неопределенности” в понимании ЛПР, могут оказаться как более, так и менее эффективными при попытке предсказания его решений, чем показатели S_{1k} и S_{2k} .

³² В работах [76,77,81,91,89,88,153] мы писали, что проекты, отобранные инвестиционным комитетом попадают в угол. Высказывание было бы точным, если бы показатели S_{1k} и S_{2k} не были ограничены. Однако, строго говоря, изображающая область тщательного изучения геометрическая фигура на рис.4.4 представляет собой не угол, а четырехугольник (прямоугольную трапецию), поскольку ограничена также максимальными значениями показателей: $S_{1k} \leq 20$ и $S_{2k} \leq 20$. Четырехугольник описывается четырьмя неравенствами, однако последние два из них при использовании МПОБ выполняются автоматически. Область отклоненных предложений представляет собой пятиугольник, а область сбора информации – семиугольник.

Область заведомо неинтересных проектов помечена надписью “отклонение”. В нее попадают точки, удовлетворяющие хотя бы одному из неравенств

$$\begin{aligned} S_{1k}(\mathbf{X}) &\leq 0 \\ S_{1k}(\mathbf{X}) &\leq (10/25) \cdot (5 - S_{2k}(\mathbf{X})) \end{aligned} \quad (8)$$

В случае попадания идеи в область, лежащую между двумя, описанными выше, нужно предпринять сбор дополнительной информации, если такой сбор не требует существенных затрат.

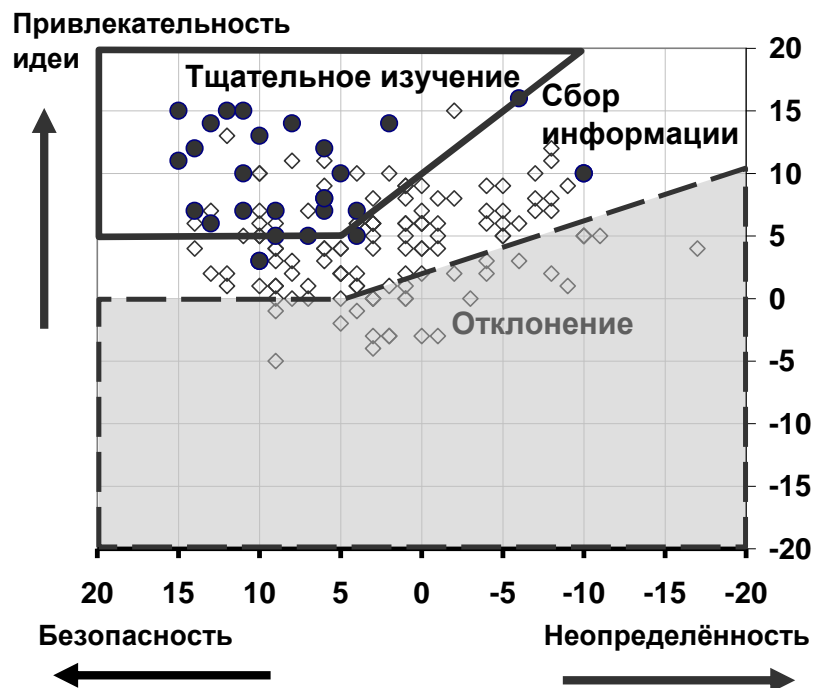


Рис.4.4. Диаграмма “неопределённость – привлекательность бизнес-идей”

Задача отбора стартапа для тщательного изучения превращается в задачу классификации с обучением, при этом требуется предварительное построение решающих поверхностей, исходя из цены ошибки первого и второго рода. На рис.4.4 все рассмотренные проекты изображены ромбами, а проекты, отобранные фондом для тщательного изучения, – закрашенными кружками. Сплошная линия ограничивает область повышенной вероятности принятия решения о проведении тщательного изучения проекта. Пунктирная линия ограничивает область идей,

которые со столь высокой вероятностью будут отклонены фондом, что дополнительный сбор информации не требуется. На данном этапе исследований мы ограничились кусочно-линейными решающими поверхностями.

Таким образом, в данной главе решены две проблемы, препятствующие деятельности венчурных фондов:

1. Проблема низкого качества фильтра нижнего уровня системы (воронки) отбора высокотехнологичных проектов, вызванная отсутствием прямой коммуникации между руководством фонда и аналитиками (на рис.4.1 между руководством и аналитиками находится несколько звеньев), решена при помощи метода обучения с учителем;

2. Проблема длительного обслуживания заявок решена при помощи разработки стандартизованного инструмента обработки заявки, который позволяет аналитику обработать заявку и сообщить автору заявки предварительный ответ, репрезентирующий предпочтения руководства фонда, в течении одного часа.

Применение инструмента при организации первых встреч

Разработанный инструмент целесообразно использовать уже при подготовке т.н. *первых встреч* с авторами идей и проектов, поскольку он помогает определить направление поиска дополнительной информации. Необходимость в дополнительной информации возникает тогда, когда итоговые оценки положения проекта на карте “Неопределенность – привлекательность”, определенные несколькими аналитиками, резко различаются, причем некоторые, но не все из получившихся оценок попадают в зону “тщательного изучения”.

Дополнительная информация должна помочь экспертам определить значения признаков проекта x_{ij} точнее. Если эта дополнительная информация решает проблему разброса оценок, то после ее сбора оценки различных экспертов x_{kij} будут мало отличаться от среднего значения признака. Возможна и ситуация, в которой дополнительная информация не поможет согласовать оценки экспертов. В последнем случае ресурсы на поиск информации и повторную

оценку расходуется впустую. На практике стоимость поиска дополнительной информации о каждом отдельном проекте ничтожно мала по сравнению с выгодой, упущенной при пропуске перспективной инновации. Однако венчурный фонд работает с многими сотнями проектов. Очевидно, что при среднем потоке заявок в одно предложение в рабочий день один специалист не может тратить на изучение одного проекта в среднем более одного дня. Поэтому когда дополнительное исследование проектов превышает возможности штата специалистов, требуется нанять дополнительного квалифицированного специалиста, а в этом случае совокупные расходы на зарплату, офисное пространство и технику и т.п. уже вполне сравнимы с инвестициями в небольшой стартап. Таким образом, даже на уровне предварительного отбора, аналитикам фонда приходится отсеивать некоторые проекты без дополнительного сбора информации.

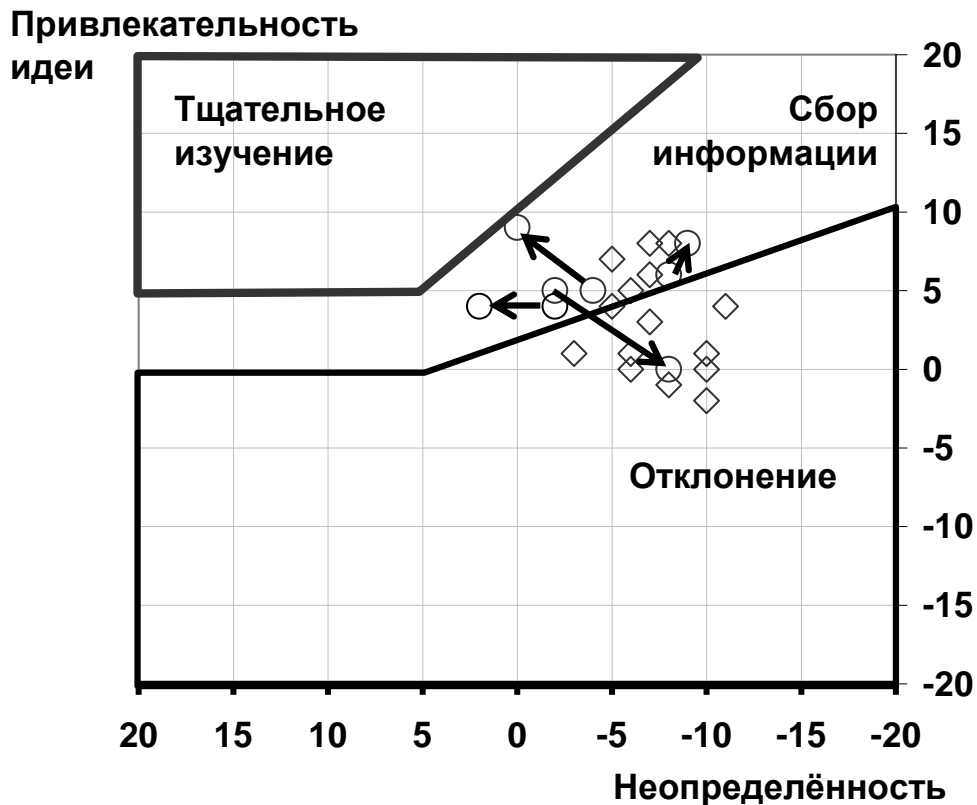


Рис.4.5. Изменение предварительной оценки проектов SaaS по итогам “первых встреч”.

На рис.4.5 стрелками, соединяющими кружки, показано, как изменились представления аналитика фонда о четырех проектах по итогам “первых встреч” с представителями трех молодых компаний, разрабатывающих решения SaaS (software-as-a-service) – услуг по предоставлению возможности использования программного обеспечения (ПО). Этот инновационный вид услуг в последние годы становится популярным направлением деятельности российских разработчиков ПО, и многие молодые коллективы предлагают решения в этой области.

Дополнительно на рис.4.5 ромбами показаны еще 15 проектов, разрабатывающих или внедряющих ПО сходного назначения, авторы которых предоставляют или рассчитывают предоставлять ПО как сервис. Каждый из четырех проектов по итогам первоначального изучения предложений попал в “промежуточную” зону, поэтому с целью сбора дополнительной информации были проведены “первые встречи”. Как иллюстрирует рис.4.5, информация, которую аналитики фонда приобретают в ходе “первых встреч” может смещать предварительную оценку проекта в различных направлениях, однако амплитуда смещений обычно недостаточно велика, чтобы вывести проекты в область тщательного изучения. В целом, область SaaS, с точки зрения изучаемого в этой работе венчурного фонда, выглядит малопривлекательной и сравнительно малознакомой.

Проверка гипотезы о снижении разброса оценок

На протяжении шести лет с 2007 по 2012 гг. автору удалось собрать оценки ста проектов по показателям МПОБ, выставленные менеджерами и аналитиками фонда ФИНАМ ИТ. В эксперименте приняли участие два менеджера фонда и аналитик, которые оценили 24, 12 и 64 поступивших в фонд проекта соответственно. Кроме того, всем проектам независимо выставил оценки автор данной работы, работающий в фонде ФИНАМ ИТ в должности аналитика.

Суммарное рабочее время, затраченное коллегами автора (исключая самого автора) на участие в эксперименте составило более 50 часов, т.е. больше полной

рабочей недели. По этой причине эксперимент не мог быть проведён быстро, ибо автор не имел возможности оторвать коллег от рабочих обязанностей на длительный промежуток времени. С точки зрения автора, большое количество оценок, выставленных одним специалистом, помимо плюсов, имеет и недостатки. Во-первых, при большом объеме работы, которая не рассматривается как приоритетная (работа принадлежит к нижнему уровню рис.4.1), оценки могут быть выставлены наспех и неаккуратно. Во-вторых, просьба оценить большое количество проектов может быть воспринята как попытка оценить профессиональную компетентность.

Полученные данные позволили автору в совместной работе с Е.А.Кобзаревой [158] проанализировать разброс в оценках привлекательности и неопределенности, u_{ki} , полученных различными специалистами. Несмотря на описанную выше попытку стандартизации путем описания методики перевода оценок в баллы, разброс оказался весьма велик, как показывают гистограммы на рис.4.5 (а). Сплошными линиями на этом рисунке изображены теоретические плотности нормальных распределений.

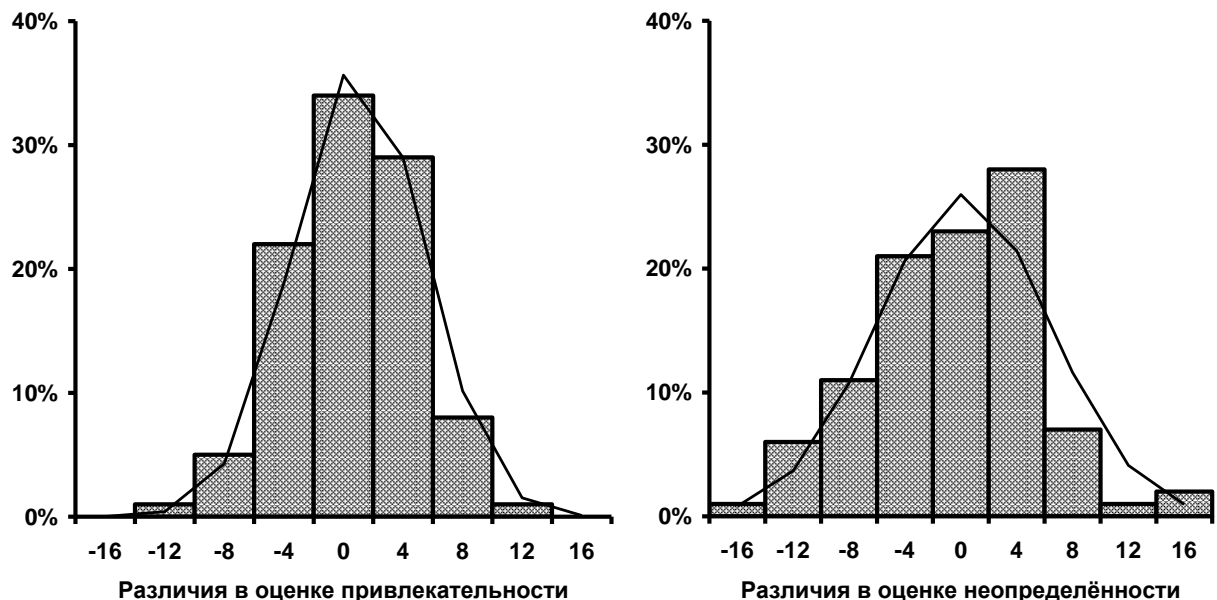


Рис.4.6. Различия в оценках привлекательности и неопределённости проектов, сделанных различными специалистами

Оказалось, что разности оценок двух аналитиков δ_1 и δ_2 распределены по нормальному закону, что проиллюстрировано на рис.4.6. Проверка нормальности распределения была проведена при помощи критерия χ^2 . Средние значения и среднеквадратические отклонения разностей вычислены по исходным наблюдениям и приведены в таблице 4.3.

Таблица 4.3.

	Различия в оценке привлекательности δ_1	Различия в оценке неопределенности δ_2
Среднее значение	1,02	0,18
Среднеквадратическое отклонение	$\sigma = 4,20$	$\sigma = 6,05$

Как показывает таблица, разница средних значений оценок оказалась невелика, таким образом, “в среднем” специалисты выставляют оценки с незначительными систематическими смещениями. Автор работы (Специалист 1) оценивает показатель привлекательности проектов в среднем на один балл ниже, чем его коллеги. Систематическое смещение в оценке показателя неопределенности практически отсутствует. Однако показатель неопределенности обладает существенно более высоким значением среднеквадратического отклонения. Здесь можно предложить два объяснения. Первое состоит в том, что автору не удалось стандартизовать признаки, составляющих этот показатель, т.е. дать точные инструкции относительно присвоения баллов. В основном, выставление баллов по этим признакам было оставлено на усмотрение эксперта. Таким образом, меньшее значение среднеквадратического отклонения первого показателя является результатом более удачной стандартизации процедуры оценки. Можно, однако, предположить, что показатель неопределенности набора проектов, поступивших на рассмотрение фонда объективно обладает большим разбросом. В таком случае, большой разброс второго показателя – свойство потока предложений, а не дефект процедуры. Оба рассуждения, на наш взгляд,

должны стимулировать попытку стандартизации оценки признаков, составляющих показатель неопределённости.

Исследование дополнительных показателей

Мы протестировали дополнительный показатель новизны бизнес-идеи. Для каждого из изученных проектов эксперт-аналитик оценил степень новизны проектов по пяти возможным признакам, выделенным Й. Шумпетером [298]: новый продукт, новый рынок, новый материал, внедрение новой технологии, новая организация производства. Агрегированный показатель новизны суммировал эти пять оценок. Как показал корреляционный анализ, линейный коэффициент корреляции с выделенным нами ранее фактором “неопределенности” составил 0,65, что выше, чем коэффициенты корреляции с показателями обоснованности идеи и синергии бизнеса по отдельности (оба оказались равны 0,54), а корреляция с фактором привлекательности и его составляющими практически отсутствует. Чем новее изобретение, тем оно непонятнее и “дальше” от инвестора, поэтому возросшая неопределенность должна компенсироваться высокой привлекательностью инновации.

Можно усмотреть соответствие между выделенными нами факторами привлекательности и неопределенности и другими наборами показателей, в частности, показателями полезности и простоты использования, на которых основаны модели TAM [342] и UTAUT [475], разработанные Ф. Дэвисом, В.Венкатешем и их соавторами при исследовании принятия решений о выборе информационных технологий. Важно также изучить соответствие между выбранными нами показателями и признаками, выделенными Э. Роджерсом [449] и позволяющими прогнозировать относительную скорость распространения инновации, такими как относительное превосходство нововведения, его совместимость, наглядность, возможность попробовать и простота. Предположительно, показатель относительного превосходства соответствует выделенному нами показателю привлекательности, а остальные показатели Э. Роджерса детализируют показатель неопределенности. Мы сопоставили также

показатель “неопределённости” с коэффициентом риска, использованным Н.Л.Карданской [180], который тоже вычисляется в виде взвешенной суммы. Нашу диаграмму неопределенность-привлекательность (рис.4.4) можно разделить на четыре вертикальные полосы равной ширины, соответствующие (слева направо) областям минимального, повышенного, критического и недопустимого риска. Как показывает рис.4.4, послуживший экспериментальной базой для нашей работы фонд ФИНАМ ИТ не выбрал для тщательного изучения ни одно из предложений, попадающих в область недопустимого риска. Безусловно, аналоги используемых нами агрегированных показателей не исчерпываются приведёнными примерами и могут быть найдены в любой области, где индивиды или организации принимают решения в условиях неопределённости, включая столь далёкие от венчурного инвестирования как искусствоведение [241] и науковедение³³.

С нашей точки зрения, более высокая точность отсева проектов достигается путем включения третьей координаты S_{3k} , – “человеческого фактора”, агрегирующего информацию об источниках информации о проекте, о рекомендациях известных экспертов, предыдущих успешных проектах ключевых лиц стартапа, об истории сотрудничества предпринимателя-инноватора с фондом. Такие характеристики не учитывались в методе МПЮБ, который изначально был предназначен для отбора идей, а не компаний или коллективов. По-видимому, социальные сети лиц, принимающих инвестиционные решения, процесс обмена информацией между членами этих сетей и механизмы их влияния друг на друга должны служить главными источниками сведений, позволяющих совершенствовать инструменты отбора. Не исключено, что для этого потребуются вновь принять во внимание индивидуальные полезности, однако измеряющие уже не оценки *аналитиков*, осуществляющих предварительный отбор, а предпочтения *экспертов*, предположительно оказывающих влияние на принятие решения

³³ Профессор Д.А.Гусев (Университет Purdue, США) обратил внимание автора на возможность проведения сопоставлений между агрегированными показателями МПЮБ и выделенными Л.Б.Баженовым [19] методологическими регулятивами естественнонаучной теории, в число которых входят принципиальная проверяемость, максимальная общность, предсказательная сила, принципиальная простота и системность.

(например, известных в индустрии предпринимателей или руководителей конкурирующих венчурных фондов). В этом отношении методика предварительного отбора подобна массовому стандартизованному тестированию школьников, которое решает задачу отсева двоечников, но не позволяет распознавать будущих талантливых ученых. Для решения задачи отбора лучших предназначена стадия тщательного изучения “due diligence”.

Анализ итогов деловой игры на основе разработанного инструмента отбора

С целью выяснить границы применения к задачам отбора феномена “мудрости толпы” [299], состоящего в возможности замены эксперта коллективом потребителей, на основе МПОБ нами была разработана деловая игра, позволяющая участникам составить представление о том, как венчурный фонд отбирает проекты в сфере ИКТ с целью инвестиций³⁴. В ходе игры каждому из тридцати трех студентов-старшекурсников МГУКИ, специализирующихся на мультимедийных и информационных системах, было предложено представить себя в роли руководителей крупных компаний сферы ИКТ, венчурных фондов – частных и с государственным участием, - бизнес-ангелов и т.п. Все участники игры получили одну и ту же информацию о SaaS-проекте Roomix, опубликованную в статье А.Гусакова в журнале “Секрет фирмы” [57], и независимо заполнили соответствующие элементы МПОБ.

³⁴ Заметим, что деловые игры применялись для изучения распространения информации со времён первых исследований С.К.Додда [349], в том числе и в нашей стране [276].

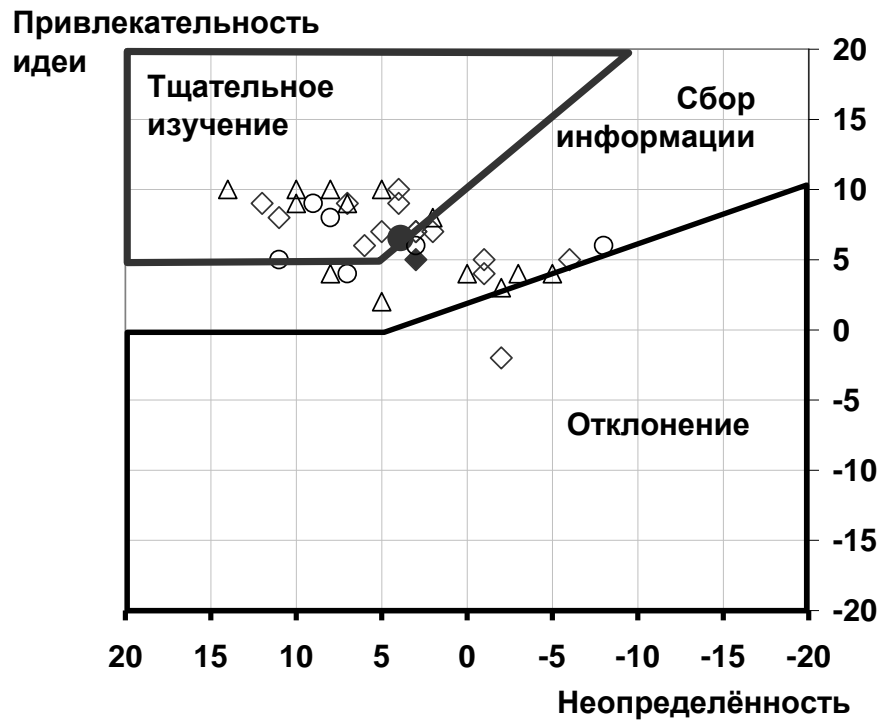


Рис.4.7. Разброс оценок одного и того же SaaS-проекта, полученных тридцатью тремя участниками деловой игры

SaaS-проект Roomix являлся аналогом (“клонем”) английского проекта MyDeco и позволял моделировать размещение мебели в квартире при помощи средств трехмерной визуализации. Авторы российского проекта рассчитывали предоставлять эту услугу пользователям Интернета бесплатно и взимать абонентскую плату с мебельных магазинов, чьи предложения размещаются на сайте Roomix.ru. Этот проект не был поддержан инвестиционным комитетом ФИНАМ ИТ, однако осенью 2008 г. другой венчурный фонд, AddVenture, инвестировал в него некоторые средства (не превышающие, по словам представителей фонда, 50 тыс. долларов). Таким образом, мнения венчурных фондов по поводу перспектив одного и того же проекта существенно разошлись.

Как показывает рис.4.7, агрегированные оценки, данные проекту студентами по предлагаемой нами методике, также характеризуются весьма существенным разбросом оценок. При этом, по горизонтальной координате разброс оказался значительно выше, чем по вертикальной. Мы связываем это с

тем, что для оценок “привлекательности” проекта, которые агрегируются в вертикальной координате, мы составили строгие количественные инструкции, в то время как для оценок обоснованности и синергии, которые агрегируются в горизонтальной координате, предложили использовать пятибалльную шкалу согласия Лайкерта.

Средняя оценка участников игры изображена на рис.4.7 закрашенным кружком, а оценка эксперта фонда ФИНАМ ИТ – закрашенным треугольником. В среднем, участники игры оценили привлекательность проекта выше, чем эксперт фонда, тем не менее, их средний результат остался в зоне дополнительного сбора информации, – там же, где и оценка специалиста. Сначала у автора сложилось мнение, что получен обнадеживающий пример описанного Дж.Шуровьески феномена “мудрости толпы” [299], и аналитика-профессионала позволительно заменить достаточно многочисленной группой студентов – потенциальных пользователей проекта. Однако все последующие деловые игры, проведённые автором после 2010 года, давали в итоге систематически более высокие оценки проекту (который уже был закрыт!). Мы предполагаем, что чем менее квалифицирован индивид, выставяющий оценку, тем более он подвержен “обаянию” продавца, в роли которого в данном случае выступает преподаватель. Нельзя также исключить, что студенты высоко оценивают перспективы проекта, поскольку надеются угодить преподавателю, предполагая, что, представляя проект, преподаватель тем самым уже его рекомендует и, соответственно, ожидает для него высоких оценок. Мы заключаем, что ни в практических, ни в исследовательских задачах отбора венчурных проектов не следует опираться на массовые неквалифицированные оценки. С другой стороны, в игровой форме можно отработывать процедуры применения МПОБ, которые позже могут быть предложены, например, сообществам предпринимателей и бизнес-ангелов.

Что касается непосредственно Roomix, то в 2009 году, по сообщениям прессы со ссылкой на инвесторов проекта, проект якобы привлек дополнительные 300 тысяч долларов инвестиций. Однако основатели проекта в блогах Интернета

утверждали, что в реальности сумма инвестиций была в пять раз меньше, к тому же выражалась не в деньгах, а в рекламной поддержке. В мае 2010 года проект Roomix закрылся “на реструктуризацию” и с тех пор не предоставляет услуг. Стартовавшие позже проекты-аналоги, основанные другими предпринимателями также закрыты, либо сменили направление деятельности.

4.2 Проблемы венчурного инвестирования в новый способ предоставления программного обеспечения

В этом разделе, посвященном применению разработанного нами инструмента отбора стартапов и подвергающей определённой ревизии статью, написанную нами в 2011 году в соавторстве с Н.С.Мациевским, изложены итоги анализа формирующегося в настоящее время в России рынка SaaS (“software as a service”), то есть программного обеспечения, предоставляемого в качестве удаленной услуги. Представлены результаты анализа сорока SaaS-проектов [153] при помощи описанного выше инструмента отбора инновационных проектов [76-88], на основе матрицы первичной оценки бизнес-идеи (МПОБ) [15,16,27].

В.И. Дрожжинов определяет SaaS как такую модель развертывания ПО, в которой поставщик выдает потребителю лицензию на приложение для использования в качестве услуги по требованию [166]. При этом В.И.Дрожжинов выделяет следующие виды SaaS: государственные приложения, Интернет-услуги; блоги, обзоры и социальные сети; совместное использование информации и знаний (Wiki); связь (e-mail); совместная работа (e-meeting); офисные приложения; пакеты автоматизации проектирования и управления (ERP, CRM и др.). Проанализировав ряд определений SaaS, А.А.Колесов предложил следующее: «SaaS — это бизнес-модель отношений в цепочке “разработчик — поставщик — пользователь”, при которой поставщик (сервис-провайдер) самостоятельно управляет приложением, предоставляя заказчику доступ к нему через Интернет, а разработчик создает приложение, предназначенное именно для такой схемы работы, передавая право на его развертывание поставщику» [191].

А.А.Колесов предлагает считать 2008-й год отправной точкой российского SaaS и называет компании “Мегаплан” и “МойСклад” в качестве первопродцев, сумевших вырасти до ведущих независимых поставщиков SaaS-решений [192,193].

По оценке издания Cnews, иаиболее востребованными решениями SaaS в России сейчас являются антивирусы, решения для организации совместной работы, а также бухгалтерские системы, CRM, HR-продукты.

В качестве целевых потребителей разработчики видят, как правило, средний и малый бизнес (СМБ), которым предлагаются недорогие решения, стоимость аренды которых составляет несколько сотен рублей в месяц. Главными достоинствами SaaS считаются экономия средств и времени, простота использования, легкость изменений, отсутствие затрат на поддержку и обслуживание.

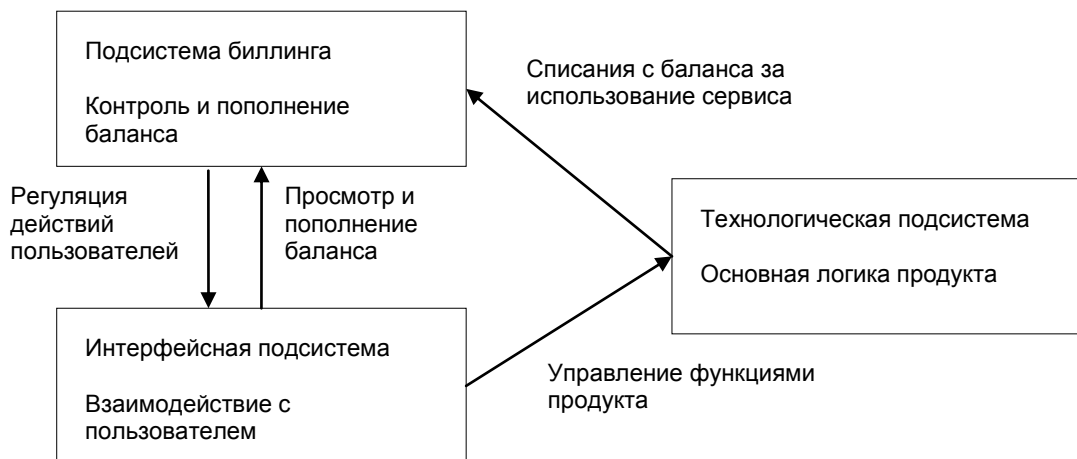


Рис.4.8. Схема работы SaaS проекта по Н.С.Мациевскому [153]

Схема работы SaaS-проекта с пользователями включает подсистему биллинга, отвечающую за финансовую логику; интерфейсную подсистему, обеспечивающую взаимодействия с пользователями; технологическую подсистему, которая обеспечивает функционирование логики продукта [153].

Приведем теперь некоторые оценки объема рынка SaaS в мире и в России (табл.4.4). Величина мирового рынка SaaS уже составляет миллиарды долларов: в

2009 году компания Gartner оценила его объем в 7,9 млрд. долларов. По предварительной оценке Gartner, в 2010 г. рынок вырос на 15,7% и составил 9,2 млрд. [245]. Крупнейший игрок рынка, компания Salesforce, была создана в 1999 году, а в 2010 году ее выручка составила 1,66 млрд. долларов. В настоящее время капитализация этой компании, акции которой с 2004 года торгуются на Нью-Йоркской фондовой бирже, составляет 17,8 миллиардов долларов.

Таблица 4.4

Оценки глобального и российского рынка SaaS

Рынок SaaS	2007	2008	2009	2010	2011 (прогноз)
Глобальный (млрд.долл)	5,0	6,4	7,5	9,2	10,7
Российский (млн.долл)	-	1,8	4,8	3,0	-

Оценка рынка SaaS-решений в России в 2008 г. составила 1,8 млн. долларов [191]. В 2009 году оценка IDC составила 5,7 млн. долларов, а в 2010 г. основатель компании «Тилби» В.Шабат оценил рынок в 3 млн. долларов [289]. Приведенные оценки разнородны и не позволяют оценить темпы прироста рынка. Активисты индустрии, ссылаясь на прогноз IDC, считают эти темпы чрезвычайно высокими и предрекают взрывной роста рынка до 161 млн. долларов в 2013 году.

В качестве препятствий для развития SaaS в России называют высокую монополизацию рынка делового ПО, недостаток средств и опыта в маркетинге и продажах, скептицизм потенциальных заказчиков [191,192]. Аналитическое агентство CNews Analytics отмечает также недостаточную осведомленность потребителей о возможностях модели SaaS и опасения за сохранность данных. Наконец, одним из основных барьеров для распространения SaaS называют нехватку инвестиций. Чтобы выяснить причины отсутствия у венчурных фондов к инвестированию в SaaS, рассмотрим механизм принятия инвестиционных решений.

В таблице 4.5 представлен фрагмент заполненной матрицы предварительной оценки бизнеса, при этом каждая колонка соответствует одному из SaaS-проектов, поступивших на рассмотрение венчурного фонда. Наблюдается систематическое отрицательное смещение показателя “Синергия бизнеса”, в то время, как показатель “Выгоды для пользователя”, как правило, положителен. Таким образом, эксперты венчурного фонда не отрицают, что SaaS-решения позволяют потребителям сокращать издержки. В то же время, не имея опыта инвестирования в SaaS, фонд не находит синергии между предлагаемыми проектами и уже сделанными приобретениями (портфельными проектами).

Таблица 4.5

Фрагмент заполненной матрицы предварительной оценки бизнеса

ПОКАЗАТЕЛИ	Проект 1	Проект 2	Проект 3	Проект 4	Проект 5	Проект 6	Проект 7	Проект 8	Проект 9	Проект 10	Проект 11	Проект 12
Привлекательность рынка	-1	1	2	4	-2	3	3	5	-2	1	-3	0
Выгоды для пользователя	5	4	3	5	3	4	3	3	2	3	2	3
Обоснованность идеи	4	1	-1	2	0	1	0	-1	-2	0	-2	-1
Синергия бизнеса	-6	-7	-3	-2	-6	-7	-7	-7	-6	-5	-6	-6
Потребность в ресурсах	1	-3	-2	-2	-1	2	-1	-2	-6	-4	-6	-1
Охрана идеи	-1	0	1	0	-2	-3	-2	-1	0	-2	-1	-2

Основные отличия нашего подхода к обработке МПОБ от традиционного описаны в работах [77,88] и в первом разделе данной главы. Мы изображаем структуру системы показателей в виде двумерной карты и отмечаем стартапы точками на этой карте. Кроме того, следуя подходу когнитивного моделирования [419], мы отказываемся от нормативного задания критерия отбора, но рассматриваем задачу отбора как задачу *обучения с учителем*, где в качестве обучающих множеств выступают выборка проектов, ранее отобранных фондом для тщательного изучения (“due diligence”), и выборка проектов, отклоненных фондом без детального изучения.

На рис.4.9 на карте, аналогичной рис.4.4, изображены итоги анализа сорока проектов SaaS. Среднее значение показателя привлекательности для этих проектов составило “+2,9”, показателя неопределенности – “-4,6”. Таким

образом, средний проект SaaS с точки зрения данного венчурного фонда выглядит не слишком привлекательным и малознакомым.

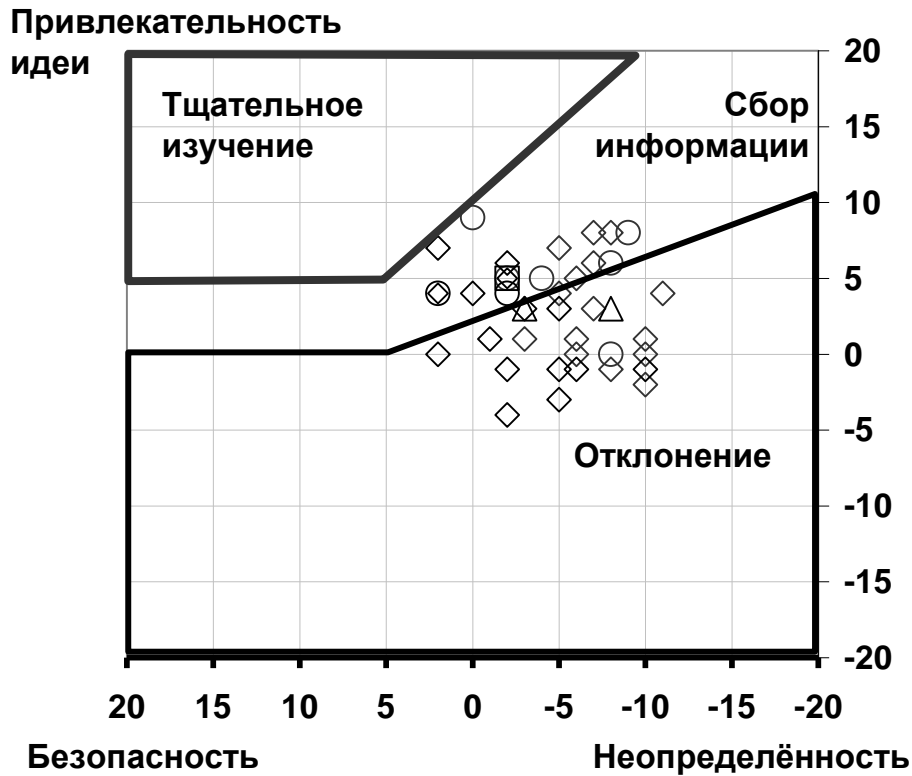


Рис.4.9. Показатели предварительной оценки сорока проектов SaaS

С целью сбора дополнительной информации были проведены “первые встречи” с представителями шести проектов SaaS, оценки которых попали в промежуточную зону. Оказалось однако, что информация, которую аналитики фонда приобрели в ходе “первых встреч” может смещать предварительную оценку проекта в различных направлениях, но амплитуда смещений недостаточно велика, чтобы переместить проекты в область тщательного изучения.

Детальный анализ агрегированных показателей позволяет уточнить модель представлений венчурного фонда о SaaS-проектах. Шесть средних значений агрегированных показателей изображены на рис.4.10.

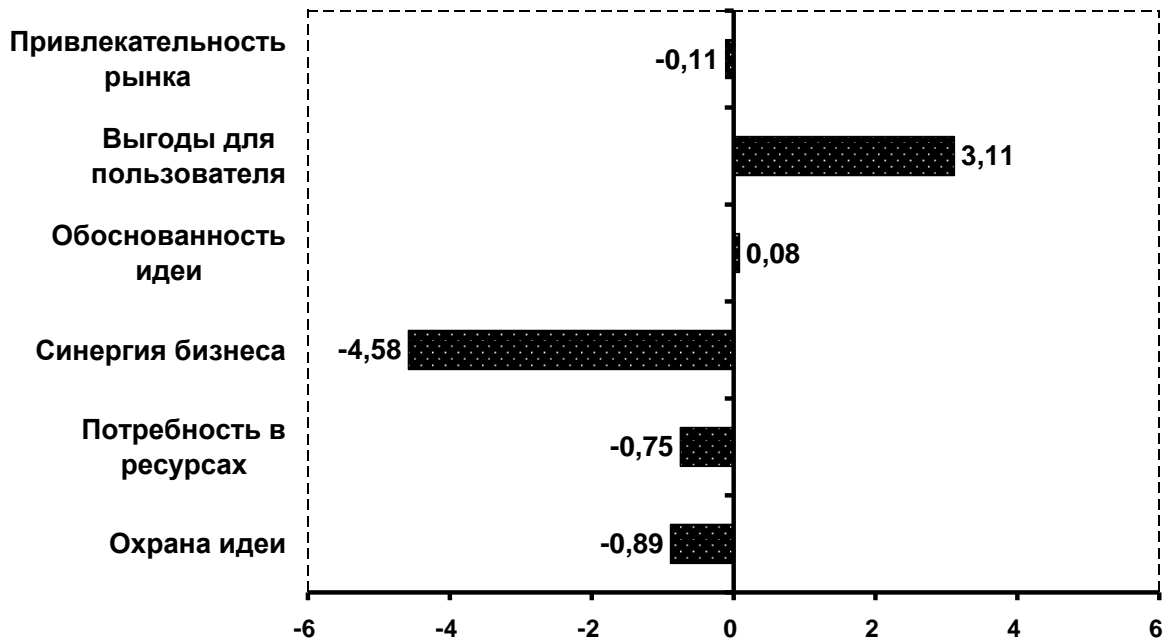


Рис.4.10. Средние агрегированные показатели проектов SaaS

Как показывает рис.4.10, средние показатели привлекательности рынка и обоснованности идеи близки к нулю, то есть к среднему значению по всем изученным проектам. Показатель потребности в ресурсах оказался слабо отрицательным, как и показатель охраны идеи. Последний низок, поскольку многие идеи в сфере SaaS имеют западные прототипы, что препятствует выходу россиян на зарубежные рынки. Наиболее важным является высокое отрицательное значение показателя “синергии бизнеса”, что говорит не только об отсутствии у фонда опыта на рынке продаж программного обеспечения, но и о том, что фонд не находит выгод для уже проинвестированных компаний от добавления к портфелю SaaS-проекта. Иными словами, становясь на позиции среднего и малого бизнеса, эксперты фонда не видят радикальных преимуществ от использования SaaS портфельными компаниями.

Показатель “Выгода для пользователя” умеренно положителен, однако не столь высок, чтобы компенсировать риски. Поэтому венчурные фонды воздерживаются от инвестиций в SaaS, дожидаясь первых отечественных историй успеха.

Можно заключить, что, в отличие от разработчиков, венчурные фонды не ожидают в ближайшем будущем большого рынка для отечественных разработок SaaS. По традиции в России компании-интеграторы внедряют решения, разработанные крупными (пока преимущественно – западными) компаниями, что дает потребителю преимущество надежности. Важным барьером для внедрения SaaS является и отсутствие длительной традиции использования легального программного обеспечения. Многие малые и средние компании используют нелегальное программное обеспечение, что лишает SaaS его основного конкурентного преимущества – дешевизны. Чтобы осознать выгоды от использования SaaS, компании должны иметь опыт использования легального программного обеспечения большим числом сотрудников. Поэтому пионерами в использовании SaaS выступают крупные компании, однако российские стартапы, как правило, ориентируются на малый бизнес, который зачастую вообще не готов платить за программное обеспечение. В силу сказанного, актуальным представляется исследование текущих практик использования программного обеспечения средним и малым бизнесом и механизма принятия решений о потреблении программного обеспечения компаниями СМБ [342]. На наш взгляд, относительно высокие шансы на успех могут иметь SaaS-проекты, которые предлагает услуги компаниям, ведущим бизнес непосредственно в Интернете. Такие компании регулярно используют программное обеспечение и способны оценить эффект сокращения затрат на него; кроме того, быстро меняющаяся внешняя среда требует от них постоянной адаптации программного обеспечения, что создает спрос на дополнительные сервисы разработчиков.

Осенью 2013 г. А.А.Колесов констатировал, что в течение прошедших с 2008 года пять лет за исключением трёх игроков рынка, включающих, помимо названных ранее компаний “Мой склад” и “Мегаплан” компанию “1С-Битрикс”, не появилось новых лидеров рынка [194]. При этом А.А.Колесов отметил, что все три компании уже являются дочерними или совместными предприятиями традиционных производителей ПО или других крупных компаний. Мы

закключаем, что традиционные производители ПО оказались не столь “неповоротливыми”, как рассчитывали первопроходцы рынка SaaS, и в течении пяти лет освоили канал распространения ПО как услуги, опередив венчурных инвесторов.

Выводы

В данной главе изучены и решены две проблемы, регулярно встречающиеся в деятельности венчурных фондов:

1. проблема низкого качества фильтра нижнего уровня системы (воронки) отбора высокотехнологичных проектов, которая решена при помощи методов теории распознавания образов;

2. проблема длительного обслуживания заявок на нижнем уровне системы отбора инновационных проектов, которая преодолена при помощи инструмента поддержки принятия решений, позволяющего аналитику обрабатывать заявки и давать заявителям предварительные ответы в течении одного часа.

Результаты данной главы внедрены ЗАО ФИНАМ при отборе венчурных проектов в сфере ИКТ в 2006-2013 гг., что подтверждается **Актом о внедрении от 24.12.2013** (см. приложение Б-4).

Развивая подход, основанный на матрице предварительной оценки бизнес-идеи, мы опробовали новый инструмент поддержки принятия решения – диаграмму “неопределённость - привлекательность”, которая построена на основе суммирования двух групп показателей. Предложенный нами инструмент отсева инновационных проектов, построенный на основе матрицы первичной оценки (МПО) бизнес-идеи с учетом специфики проектов, поступающих на рассмотрение венчурного фонда, является достаточно простым и эффективным для применения в практике работы венчурных фондов. Диаграмма “неопределённость – привлекательность” способна помочь предпринимателям понять логику принятия венчурным фондом решений по предложенным проектам, обосновать необходимость сотрудничества предпринимателя и потенциального инвестора в ходе изучения инновационных предложений.

На основе этого подхода мы обосновываем выбранное нами в этой работе определение нововведений, данное Э.Роджерсом [449] и позволяющее рассматривать диффузию инноваций, как процесс принятия решений большим числом агентов (индивидов, домохозяйств, организаций).

Безусловно, данной работой направления исследований информационных процессов, протекающих в венчурных фондах, далеко не исчерпаны. Перспективными нам представляются следующие задачи:

- рассмотреть каналы распространения информации (заявки, презентации, встречи), их пропускную способность, виды помех, способы повышения пропускной способности каналов;
- разработать меры по снижению интенсивности ошибок во внутренних и внешних каналах, включая развитие коммуникативных навыков и формата сообщений;
- разграничить потери информации благодаря отсутствию экспертизы и благодаря неполноте представленной информации;
- обосновать важность предлагаемой автором методики низкой пропускной способностью канала нижнего уровня;
- провести сравнительный анализ эффективности передачи информации через традиционные каналы и новых способов коммуникаций, таких как социальные сети венчурных инвесторов;
- разработать информационные критерии эффективности “первых встреч” и других стадий отбора венчурных проектов;
- обосновать фокус исследования лишь на нижнем уровне системы - уровне работы аналитика, описать интерфейсы предоставления информации выше- и нижележащим уровням иерархии;
- исследовать обратную связь в системе и взаимное обучение;
- рассмотреть систему отбора проектов как самообучающуюся нейронную сеть, состоящую из двух или трёх слоёв.

ЗАКЛЮЧЕНИЕ

Данная диссертационная работа является завершенным научным исследованием, актуальность которого обусловлена интенсивным внедрением новых информационных и телекоммуникационных технологий, ростом числа электронных изданий, разнообразия информационных ресурсов и услуг, предлагаемых через Интернет. Как показал опыт 2006-2013 гг., данные социологических опросов и маркетинговых исследований существенно отклоняются от предсказаний классических логистических моделей. Выявленная проблема делает изучение общих закономерностей, позволяющих объяснить и прогнозировать распространение инфокоммуникационных инноваций, актуальной научной задачей.

В исследовании получены следующие **научные результаты**:

1. Построена количественная модель распространения нововведения в неоднородном (по полу и географии проживания) обществе с учетом демографических процессов и изменения стоимости новинки во времени. Разработанная модель апробирована на данных массовых репрезентативных социологических опросов. Построен и подтвержден прогноз распространения Интернета в России и крупных географических единицах.
2. Разработан способ решения уравнений распространения нововведений в стабильной популяции и проанализированы случаи обучения использованию нововведения в семье и в школе, не поддающиеся решению на основе ранее использованных моделей.
3. Разработаны многостадийные модели распространения Интернета, которые апробированы на данных массовых социологических опросов. На этой основе построены прогнозы распространения Интернета в России до 2023 г.
4. Разработаны и апробированы количественные модели, позволяющие совместно анализировать статистику распространения нововведений в домохозяйствах и среди индивидов.

5. Разработана и апробирована модель роста охвата аудитории Интернета в целом по данным социологических опросов и модели роста охвата аудитории крупных информационных ресурсов (поисковых и почтовых сервисов и масс-медиа) по данным онлайн-панелей и счетчиков.
6. Разработаны модель и прогноз распространения компьютерного парка России на основании данных маркетинговых исследований и онлайн-опросов. Обоснована важность системного подхода к моделированию компьютерного парка и необходимость совместного использования данных маркетинговых исследований и онлайн-опросов. Построены прогнозы компьютерного парка и рынка Российской Федерации до 2016 г.
7. Разработан и апробирован метод выбора проектов в сфере ИКТ для венчурного инвестирования. На его основе проанализировано новое направление в предоставлении ИТ-услуг (SaaS), выработаны рекомендации для инвесторов и ИТ-предпринимателей.

Реализация результатов исследования.

- Выводы, модели и прогнозы, полученные в ходе работы, используются руководством ЗАО ФИНАМ для целей стратегического менеджмента.
- Опубликованные автором данные и прогнозы использованы в инвестиционном холдинге ФИНАМ, ЗАО «Юду», а также руководством официального портала Олимпийских Игр Сочи-2014.
- Построенные модели нашли применение в учебном процессе при чтении курсов «Социология виртуального пространства», «Поисковые системы Интернета» для студентов специализации «Интернет-журналистика» в МГУ им. Ломоносова и для студентов специализации «Технологии электронного бизнеса», «Технологии информационного бизнеса» в Московском государственном университете культуры и искусств.

Соответствующие акты о внедрении приведены в приложении **Б** к диссертации.

Перспективы дальнейшей разработки темы

Основными направлениями развития темы исследования в среднесрочной перспективе, с нашей точки зрения, могут стать:

1. Применение разработанных моделей для анализа сценариев и прогнозирования распространения новых устройств (планшеты, смартфоны, 3-D принтеры), информационных технологий и ресурсов (онлайн-обучение, мобильные библиотеки, геоинформационные системы, онлайн-заявки) и практик (мобильный доступ к Интернету, использование государственных услуг онлайн).

2. Разработка алгоритма численного решения уравнений многостадийной модели распространения новых информационных технологий с учётом воспроизводства населения, и соответствующего программного кода.

3. Разработка модели распространения нововведения, учитывающей как возраст индивида, так и размер домохозяйства, членом которого индивид является, позволила бы интегрировать результаты первой и третьей главы данной работы, и использовать совместно данные о распространении Интернета среди индивидов и в домохозяйствах.

4. Разработка моделей принятия и реализации решений при управлении информатизацией с целью более надёжного прогнозирования потенциала распространения новых технологий.

5. Применение результатов, полученных при изучении классической проблемы моментов, с целью аппроксимации неизвестных распределений интенсивности использования информационных ресурсов.

6. Апробация полученных в четвёртой главе моделей принятия решений венчурным фондом на данных о решениях нескольких инвестиционных фондов с известной специализацией.

7. Применение разработанной модель отбора стартапов, для решения задачи выбора количественных моделей распространения новых технологий из множества моделей-кандидатов.

Рекомендации об использовании результатов

Теоретические и практические результаты исследования могут использоваться при:

1. Совершенствовании государственной информационной политики,
2. Управлении инфокоммуникационными проектами для прогнозирования распространения новых продуктов и услуг,
3. Медиапланировании крупных информационных ресурсов WWW,
4. Отборе проектов для инвестирования инвестиционными комитетами венчурных фондов и конкурсными комиссиями
5. Разработке учебных планов и образовательных программ по специализации «Технологии электронного бизнеса», «Технологии информационного бизнеса», «Интернет-журналистика».

СПИСОК ЛИТЕРАТУРЫ

1. Абилов, А.В. Закономерности развития регионального инфокоммуникационного комплекса / А.В. Абилов. – М.: Горячая линия-Телеком, 2008. – 264 с.
2. Абрамов, А. Победитель конкурса интернет-проектов Web Ready 2010 нашел инвестора [Электронный ресурс] / А. Абрамов. – Электрон. текстовые дан. – Режим доступа: <http://www.spbit.ru/news/n77717/>. – Загл. с экрана.
3. Айвазян, С.А. Прикладная статистика. Основы эконометрики: учеб. для вузов. В 2 т. 2-е изд., испр. - Т.2. / С.А. Айвазян. – М.: ЮНИТИ-ДАНА, 2001. – 432 с.
4. Алексеева, О.А. Уровень информатизации домашних хозяйств в России / О.А. Алексеева, О.Н. Махрова // Экономическая наука современной России. – 2009. – №2. – С.117-127.
5. Алексеева, О.А. Гендерные аспекты информатизации / О.А. Алексеева, Н.А. Ермакова, В.Н. Жеребин // Народонаселение. – 2013. – №2(60). – С.113-124.
6. Ананьев, А.А. Определение возрастной структуры эксплуатационного парка товаров длительного пользования / А.А. Ананьев, И.М. Ильенков // Экономика и математические методы. – 1980. – т. XVI, вып.3. – С.571-577.
7. Андерсон, Р. Инфекционные болезни человека. Динамика и контроль / Р. Андерсон, Р. Мэй. – М.: Мир, 2004. – 784 с.
8. Антопольский, А.Б. Информационные ресурсы России. Учебное пособие по специальности "Прикладная информатика (в менеджменте)". Вып.1. Информационные ресурсы инновационного развития. / А.Б. Антопольский, О.В. Шлыкова. – М.: ИПКИР-МГУКИ, 2006. – 270 с.
9. Аньшин, В.М. Инновационный менеджмент / В.М. Аньшин, А.А. Дагаев (ред.). – М.: Дело, 2003. – 528 с.
10. Афанасьев, В.Н. Анализ временных рядов и прогнозирование / В.Н. Афанасьев, М.М. Юзбашев. – М.: Финансы и статистика, 2001. – 228 с.
11. Афанасьева, К.Е. Идентификация состояния и прогнозирование регионального рынка / К.Е. Афанасьева, В.И. Ширяев // Проблемы управления. – 2007. – №3. – С.63-65.
12. Афанасьева, К.Е. Прогнозирование региональных рынков сотовой связи / К.Е. Афанасьева, В.И. Ширяев // Проблемы прогнозирования. – 2007. – №5. – С.97-105.
13. Ахиезер, Н.И. Классическая проблема моментов и некоторые вопросы анализа, связанные с нею / Н.И. Ахиезер. – М.: Физматлит, 1961. – 312 с.
14. Ахиезер, Н.И. О некоторых вопросах теории моментов / Н.И. Ахиезер, М.Г. Крейн. – Харьков: ГОНТИ, 1938. – 256 с.
15. Бабаскин, С.Я. Использование аппарата нечетких множеств при отборе бизнес-идей / С.Я. Бабаскин // Машиностроитель. – 2005. – №4. – С.36-43.
16. Бабаскин, С.Я. Инновационный проект. Методы отбора и инструменты анализа рисков / С.Я. Бабаскин. – М.: Дело АНХ, 2009. – 240 с.
17. Багатурова, О.С. Математические модели маркетинга Аналитический обзор / О.С. Багатурова, А.Г. Мамиконов // Автоматика и телемеханика. – 1991. – 8. – С.3-35.
18. Багриновский, К.А. Современные методы управления технологическим развитием / К.А. Багриновский, М.А. Бендииков, Е.Ю. Хрусталева. – М.: Росспэн, 2001. – 272 с.
19. Баженов, Л.Б. Строение и функции естественнонаучной теории / Л.Б. Баженов. – М.: Наука, 1978. – 232 с.

20. Бакман, Ю.А. Рынок сотовой связи России: от экстенсивного развития к интенсивному / Ю.А. Бакман, В.С. Панфилов // Проблемы прогнозирования. – 2008. – №5. – С.36-54.
21. Бароян, О.В. Моделирование и прогнозирование эпидемий гриппа для территории СССР. / О.В. Бароян, Л.А. Рвачев, Ю.Г. Иванников. – М.: ИЭМ им. Н.Ф. Гамалеи, 1977. – 546 с.
22. Бартоломью, Стохастические модели социальных процессов / Д.Бартоломью. – М.: Финансы и статистика, 1985. – 296 с.
23. Барютин, Л.С. Основы инновационного менеджмента. Теория и практика: Учебник. – 2-е изд., перераб. и доп. / Л.С. Барютин и др.. – М.: ЗАО: "Издательство "Экономика", 2004. – 518 с.
24. Березин, И.С. Сколько в России абонентов мобильной связи? / И.С. Березин // РОМИР-Мониторинг. – 2007. – 19 февраля.
25. Бешенков, С.А. Моделирование и формализация. Методическое пособие / С.А. Бешенков, Е.А. Ракитина. – М.: Лаборатория Базовых Знаний, 2002. – 336 с.
26. Болдырев, Р.Л. Моделирование диффузии инноваций с учетом ценового фактора / Р.Л. Болдырев, И.Б. Руссман, И.Н. Щепина // Труды конф. "Математическое моделирование систем. Методы, приложения и средства", Воронеж, 1998.
27. Бретт, А. Оценка коммерциализуемости технологий / А. Бретт // Коммерциализация технологий. Мировой опыт – российским регионам / сост. и общ. Ред. Н.М.Фонштейн. – Пер. с англ.. – 1995. – С.7-47.
28. Булгак, В.Б. Новые методы прогнозирования развития телекоммуникаций и их применение в отрасли "Связь Российской Федерации" / В.Б. Булгак, Л.Е. Варакин, Н.Н. Каледина, В.Д. Москвитин, Л.Ф. Шамаева. – М.: МАС, 2000. – 105 с.
29. Вабищевич, П.Н. Численные методы. Вычислительный практикум / П.Н. Вабищевич. – М.: Книжный дом "Либроком", 2010. – 320 с.
30. Вайдлих, В. Социодинамика. Системный подход к математическому моделированию в социальных науках. Пер. с англ. С.Л.Котлярова, под ред. Ю.С.Попкова, А.Е.Семечкина. Изд. 2-е, стереотипное / В. Вайдлих. – М.: Едиториал УРСС, 2005. – 480 с.
31. Варакин, Л.Е. 25 лет инфокоммуникационной революции / Л.Е. Варакин, ред.. – М.: МАС, 2006. – 264 с.
32. Варакин, Л.Е. Информационно-экономический закон. Взаимосвязь инфокоммуникационной инфраструктуры и экономики / Л.Е. Варакин. – М.: МАС, 2006. – 160 с.
33. Вартанова, Е.Л. Ч.2. Северная модель: государство всеобщего информационного благоденствия? / Е.Л. Вартанова // Национальные модели информационного общества / Отв. ред. и сост. Е.Л.Вартанова, науч.ред. Н.В.Ткачева. – М.: ИКАР. – 2004. – . – С.83-112.
34. Вартанова, Е.Л. Ч.1. Новые неравенства большой Европы / Е.Л. Вартанова, Я.Н. Засурский // Национальные модели информационного общества / Отв. ред. и сост. Е.Л.Вартанова, науч.ред. Н.В.Ткачева. – М.: ИКАР. – 2004. – С.83-112.
35. Варшавский, Л.Е. Исследование динамики показателей рынков наукоемкой продукции (на примере рынков компонентов персональных компьютеров) / Л.Е. Варшавский // Экономика и математические методы. – 2004. – т.40, №1. – С.101-116.
36. Васильев, В.В. Мониторинг информатизации: показатели, методология оценки и прогнозирования / В.В. Васильев, Т.Ю. Салюткина. – М.: Палеотип, 2005. – 160 с.

37. Вентцель, Е.С. Теория случайных процессов и ее инженерные приложения: Учеб. пособие для втузов. — 2-е изд. / Е.С. Вентцель, Л.А. Овчаров. — М.: Высшая школа, 2000. — 383 с.
38. Вершинская, О.Н. Международные индексы готовности стран к информационному обществу / О.Н. Вершинская, О.А. Алексеева // Труды института системного анализа Российской академии наук. — 2011. — Т.61, №2. — С.19-25.
39. Виленский, П.Л. Оценка эффективности инвестиционных проектов: Теория и практика: учеб. Пособие. — 4-е изд., перераб. и доп / П.Л. Виленский, В.Н. Лившиц, С.А. Смоляк. — М.: Дело АНХ, 2008. — 1104 с.
40. Вишневский, В.М. Теоретические основы проектирования компьютерных сетей / В.М. Вишневский. — М.: Техносфера, 2003. — 512 с.
41. Влэдуц, Г.Э. Научная и техническая информация как одна из задач кибернетики / Г.Э. Влэдуц, В.В. Налимов, Н.И. Стяжкин // Успехи физических наук. — 1959. — Т. 69. — № 1. — С.13–56.
42. Вольтерра, В. Математическая теория борьбы за существование / В. Вольтерра. — Москва-Ижевск: Институт компьютерных исследований, 2004. — 288 с.
43. Вороненкова, Г.Ф. Интернет для всех: германский путь в “e-society” ? / Г.Ф. Вороненкова // Национальные модели информационного общества / Отв. ред. и сост. Е.Л.Вартанова, науч.ред. Н.В.Ткачева. — М.: ИКАР. — 2004. — С.50-64.
44. Гаврилец, Ю.Н. Моделирование динамики социальной установки (отношение к коммерческой рекламе на телевидении) / Ю.Н. Гаврилец, Ю.В. Фомина // Социология: 4М - методология, методы, математические модели. — 2002. — №15. — С.110-134.
45. Гаврилов, Л.А. Биология продолжительности жизни / Л.А. Гаврилов, Н.С. Гаврилова. — М.: Наука, 1991. — 276 с.
46. Галицкий, Е.Б. Заветные десять процентов / Е.Б. Галицкий // Независимая газета. — 2003. — 4 июля.
47. Галицкий, Е.Б. Почти 10 миллионов / Е.Б. Галицкий // Независимая газета. — 2003. — 23 мая.
48. Галицкий, Е.Б. Не исключено, что нас ждет пессимистический сценарий развития Интернета / Е.Б. Галицкий // Интернет-маркетинг. — 2007. — №1(37). — С.10-25.
49. Гельман, Л.М. Модели инновационных процессов (обзор зарубежной литературы) / Л.М. Гельман, М.И. Левин // Экономика и математические методы. — 1989. — Т.25, №6.
50. Гиляревский, Р.С. Основы информатики: курс лекций / Р.С. Гиляревский. — М.: Экзамен, 2003. — 319 с.
51. Голиченко, О.Г. Современная технологическая революция и новые возможности инновационного развития "догоняющих" стран / О.Г. Голиченко // Инновации. — 2010. — 3. — С.12-22.
52. Госкомстат России, / Госкомстат России. — М.: Госкомстат России, 2001. — 397 с.
53. Григоров, Ю.Н. Система интерактивного проектирования программ развития парков технических систем / Ю.Н. Григоров, Ю.Н. Павловский, В.Б. Соколов // Вопросы кибернетики: моделирование человеко-машинных кибернетических систем.. — 1984. — С.133-145.
54. Григоров, Ю.Н. Проблемы использования технологии математического моделирования в народном хозяйстве (на примере моделей возрастных структур парков технических

- систем) / Ю.Н. Григоров, П.В. Данилов, Ю.Н. Павловский // Математическое моделирование. – 1991. – т.3. - №4. – С.57-66.
55. Громыко, Г.Л. Теория статистики: Учебник. - 2-е изд., / Г.Л. Громыко, ред.. – ИНФРА-М, 2006. – 476 с.
 56. Губанов, Д.А. Социальные сети: модели информационного влияния, управления и противоборства / Д.А. Губанов, Д.А. Новиков, А.Г. Чхартишвили. – М.: Изд-во физико-математической литературы, 2010. – 228 с.
 57. Гусаков, А. Трехмерный уют / А. Гусаков // Секрет фирмы. – 2008. – №41 (273). – С.48-49.
 58. Гуц, А.К. Математическая социология / А.К. Гуц, Л.А. Паутова, Ю.В. Фролова. – Омск: ОМГУ, 2003. – 191 с.
 59. Дворкина, М.Я. Как долго длится инновация / М.Я. Дворкина // Библиосфера. – 2009. – №4. – С.15-18.
 60. Деарт, Ю.В. Анализ тенденций развития сетей связи как объектов диффузии инноваций / Ю.В. Деарт // Вопросы экономических наук. – 2005. – №3. – С.24-28.
 61. Деарт, Ю.В. Прогноз развития сотовой связи в России / Ю.В. Деарт // Вестник связи. – 2005. – №4. – С.32-34.
 62. Деарт, Ю.В. Прогноз количества пользователей Интернет в России / Ю.В. Деарт, А.Ю. Цым, И.В. Бурцев // Электросвязь. – 2005. – №6. – С.16-18.
 63. Делицын, Л.Л. Динамика роста российского Интернета и крупных Интернет-порталов / Л.Л. Делицын // Информационно-аналитическая деятельность в России: состояние и перспективы. Материалы Всероссийской научно-практ. конф. аналитических работников / Под. ред. А.И. Селиванова, Н.А. Слядневой. – М.: ИПК Госслужбы, 2004. – Ч.2. – С. 79-81.
 64. Делицын, Л.Л. Проблема цифрового неравенства и потенциал развития Интернета в России / Л.Л. Делицын // Информационные процессы. – 2006. – Т.6. -№2. - С. 124-130.
 65. Делицын, Л.Л. Две модели распространения сотовой связи в России / Л.Л. Делицын // Вопросы структуризации экономики. - №3. – Махачкала: ИСЭИ ДНЦ РАН. – 2008. – С. 36-40.
 66. Делицын, Л.Л. Количественный анализ распространения Интернета среди мужчин и женщин / Л.Л. Делицын // Сборник статей V Междунар. научно-практ. конф. «Женщина в политике и обществе»/Под ред. А.И. Кравченко, Т.И.Королевой. – Пенза. Приволжский Дом знаний. – 2008. – С.18-22.
 67. Делицын, Л.Л. Кумулятивные кривые охвата и проблема динамики аудитории российского Интернета / Л.Л. Делицын // Интернет-Маркетинг. – 2008. – №5. – С. 280 — 292.
 68. Делицын, Л.Л. Моделирование распространения Интернета в России с учетом цен и демографических процессов / Л.Л. Делицын // Материалы VI всерос. научно-практ. конф. “Проблемы информатизации социальных систем: региональный аспект”, Чебоксары, 24-26 апреля 2008г./ Отв. редактор проф. Н.В.Софронова. - Чебоксары. – 2008. – С. 67-70.
 69. Делицын, Л.Л. Моделирование распространения Интернета с учетом демографических процессов и стоимости доступа / Л.Л. Делицын // Интернет и современное общество. Сборник тезисов докладов объединенной конференции 28-30 октября 2008 г. / СПб.: Изд-во СПбГУ. – 2008. – С. 48-51

70. Делицын, Л.Л. Моделирование распространения нововведения с учетом распределения размеров домохозяйств / Л.Л. Делицын // Материалы 5-й междунар. заочной научн. конф. «Синергетика природных, технических и социально– экономических систем». - Тольятти. ТГУС. - 2008. – С.189-195.
71. Делицын, Л.Л. Модель распространения нововведений в домохозяйствах на примере мобильной связи // / Л.Л. Делицын // Информационно-вычислительные технологии и их приложения: сб. статей VIII Междунар. научно-техн. конф./ Под ред. проф. А.Н.Кошева. – Пенза: РИО ПГСХА. – 2008. – С. 130-133.
72. Делицын, Л.Л. Можно ли согласовать данные социологических опросов об интенсивности использования Интернета? / Л.Л. Делицын // Интернет-Маркетинг. – 2008. – №6. – С. 334-352.
73. Делицын, Л.Л. О неединственности представления логистической модели распространения инноваций в форме вероятностной смеси / Л.Л. Делицын // Высокие технологии, фундаментальные и прикладные исследования, образование. Сб. трудов Шестой междунар. научно-практ. конф. “Исследование, разработка и применение высоких технологий в промышленности” :16-17 октября 2008 г./ Под ред. А.П. Кудинова, Г.Г. Матвиенко. – СПб.: Изд-во Политехн. ун-та. – 2008. – С.32-34.
74. Делицын, Л.Л. Об одном применении теории вероятностей для сравнения моделей распространения инноваций / Л.Л. Делицын // Инновационные процессы в менеджменте: сб. статей V Междунар. научно-практ. конф./ Под ред. Л.С. Егоровой, Г.С. Широкаловой, Г.В.Синцова. – Пенза. Приволжский Дом знаний. – 2008. – С. 15-19.
75. Делицын, Л.Л. Развитие информационных и телекоммуникационных технологий в России как процесс распространения инноваций в неоднородном обществе / Л.Л. Делицын // Социология инноватики: Социальные механизмы формирования инновационной среды. Материалы II международной конференции / Под ред. Ю.А. Карповой. – 2008. – С.293-313.
76. Делицын, Л.Л. Разработка инструмента отбора стартапов для финансирования венчурным фондом / Л.Л. Делицын // Высокие технологии, фундаментальные и прикладные исследования, образование. Т.12: Сб. трудов Пятой междунар. научно-практ. конф. “Исследование, разработка и применение высоких технологий в промышленности”/ Под ред. А.П. Кудинова, Г.Г. Матвиенко. – СПб.: Изд-во Политехн. ун-та. – 2008. – С. 8-9.
77. Делицын, Л.Л. Распознавание стартапов для финансирования венчурным фондом / Л.Л. Делицын // Информационные технологии в профессиональной деятельности и научной работе: сб. материалов всерос. научно-практ. конф. Ч.2./ – Йошкар-Ола: Марийский государственный технический университет. – 2008. – С. 29-31
78. Делицын, Л.Л. Распространение инноваций с учетом цен и демографических процессов: модель с непрерывным временем / Л.Л. Делицын // Информационно-вычислительные технологии и их приложения: сб. статей VIII Междунар. научно-техн. конф./ Под ред. проф. А.Н.Кошева. – Пенза: РИО ПГСХА. – 2008. – С. 134-137.
79. Делицын, Л.Л. Распространение Интернета в неоднородном обществе / Л.Л. Делицын // Высокие технологии, фундаментальные и прикладные исследования, образование. Т.12: Сб. трудов Пятой междунар. научно-практ. конф. “Исследование, разработка и применение высоких технологий в промышленности”/ Под ред. А.П. Кудинова, Г.Г. Матвиенко. – СПб.: Изд-во Политехн. ун-та. – 2008. – С. 68-69.

80. Делицын, Л.Л. Распространение информации в социальных сетях как случайный процесс / Л.Л. Делицын // Высокие технологии, фундаментальные и прикладные исследования, образование. Сб. трудов Шестой междунар. научно-практ. конф. «Исследование, разработка и применение высоких технологий в промышленности»: 16-17 октября 2008 г./ Под ред. А.П. Кудинова, Г.Г. Матвиенко. – СПб.: Изд-во Политехн. ун-та. – 2008. – С.34-36.
81. Делицын, Л.Л. Система отбора инновационных проектов венчурным фондом / Л.Л. Делицын // Материалы IV международной научной конференции «Инновационное развитие и экономический рост». Москва. РУДН, 6-7 ноября 2008 г./ М.:РУДН – 2008. – С.466-470.
82. Делицын, Л.Л. Совместное моделирование численности пользователей и владельцев нового товара / Л.Л. Делицын // Составляющие научно-технического прогресса: Сб. материалов 4-й междунар. научно-практ. конф.: 23-24 апреля 2008 г. / – Тамбов: Издательство ТАМБОВПРИНТ. 2008. – С. 27-30.
83. Делицын, Л.Л. Способ расчета теоретических кумулятивных кривых охвата аудитории WWW-сайтов / Л.Л. Делицын // Высокие технологии, фундаментальные и прикладные исследования, образование. Сб. трудов Шестой междунар. научно-практ. конф. «Исследование, разработка и применение высоких технологий в промышленности»: 16-17 октября 2008 г./ Под ред. А.П. Кудинова, Г.Г. Матвиенко. – СПб.: Изд-во Политехн. ун-та. – 2008. – С.88-90.
84. Делицын, Л.Л. Сравнение моделей распространения нововведений на основе исторических рядов динамики / Л.Л. Делицын // Инновационные процессы в менеджменте: сб. статей V Междунар. научно-практ. конф./ Под ред. Л.С. Егоровой, Г.С. Широкаловой, Г.В.Синцова. – Пенза. Приволжский Дом знаний. – 2008. – С. 19-24.
85. Делицын, Л.Л. Чтобы поднять аудиторию, надо уронить цены / Л.Л. Делицын // Информкурьерсвязь. - №11. - 2008. – С.45-47.
86. Делицын, Л.Л. Бета-биномиальная модель поведения онлайн-панели / Л.Л. Делицын // Информационные технологии и математическое моделирование (ИТММ-2009): Материалы VIII Всероссийской научно-практической конференции с международным участием (13-14 ноября 2009 г.). - Томск: Изд-во Томского ун-та, 2009.
87. Делицын, Л.Л. Вероятностные модели распространения мобильной связи в России / Л.Л. Делицын // Труды междунар. конф. «Перспективы развития телекоммуникационных систем и информационные технологии». Под ред. д-ра экон. наук, проф. А.В.Бабкина, д-ра техн. наук, проф. В.А.Кежаева. – СПб.: Изд-во Политехн. ун-та. – 2009. – С.378-394.
88. Делицын, Л.Л. Инструментарий для отбора венчурным фондом инновационных проектов / Л.Л. Делицын // Научно-технические ведомости СПбГПУ. Экономические науки. – 2009. - №4. – С.296-307.
89. Делицын, Л.Л. Когнитивное моделирование принятия решений инвестиционным комитетом венчурного фонда / Л.Л. Делицын // Когнитивный анализ и управление развитием ситуацией (CASC'2009): Труды международной конференции (17-19 ноября 2009 г., Москва). – М.: ИПУ РАН. – 2009. – С.169-172.
90. Делицын, Л.Л. Кумулятивные кривые охвата аудитории WWW-ресурса для неубывающего количества пользователей / Л.Л. Делицын // Тезисы доклада на междунар. научно-практ. конф. «Наука и производство – 2009». – Брянск, БГТУ. – 2009. – С131-133.

91. Делицын, Л.Л. Методика отсева инновационных проектов в сфере ИКТ венчурным фондом / Л.Л. Делицын // Труды всерос. конф. «Инновационная экономика и промышленная политика региона». / Под ред. д-ра экон. наук, проф. А.В.Бабкина. – СПб.: Изд-во Политехн. ун-та. – 2009. – С.58-74.
92. Делицын, Л.Л. Многостадийные модели распространения ИКТ в России / Л.Л. Делицын // Системное моделирование социально-экономических процессов: труды 32-й Междунар. научной школы-семинара, Вологда, 5-10 октября 2009 г.: в 3 ч./под ред. В.Г.Гребенникова, И.Н.Щепиной, В.Н.Эйтингона. – Воронеж: ИПЦ ВГУ. – 2009. – С. 145-151.
93. Делицын, Л.Л. Многостадийные модели распространения Интернета в России / Л.Л. Делицын // Высокие технологии, фундаментальные исследования, образование. Т2: Сб. трудов Седьмой междунар. научно-практ. конф. “Исследование, разработка и применение высоких технологий в промышленности” / Под ред. А.П. Кудинова, Г.Г. Матвиенко. – СПб.: Изд-во Политехн. ун-та. – 2009. – С54-56.
94. Делицын, Л.Л. Моделирование данных социологических опросов о распространении мобильной связи в России / Л.Л. Делицын // Научно-технические ведомости СПбГПУ, 2009. - №1. – С.96-104.
95. Делицын, Л.Л. Моделирование данных социологических опросов о распространении мобильной связи в России / Л.Л. Делицын // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. – 2009. – №1. – С.96-104.
96. Делицын, Л.Л. Моделирование охвата аудитории информационного ресурса с учетом оттока пользователей / Л.Л. Делицын // Высокие технологии, фундаментальные исследования, образование. Т2: Сб. трудов Седьмой междунар. научно-практ. конф. “Исследование, разработка и применение высоких технологий в промышленности”/ Под ред. А.П. Кудинова, Г.Г. Матвиенко. – СПб.: Изд-во Политехн. ун-та. – 2009. – С.57-59.
97. Делицын, Л.Л. Моделирование распространения нововведений при положительном скачке потенциала / Л.Л. Делицын // Инноватика - 2009: Труды междунар. конф. – Ульяновск: УлГУ. – 2009. – С. 310-311.
98. Делицын, Л.Л. Неединственность моделей охвата аудитории WWW-ресурсов в условиях замены файлов cookies / Л.Л. Делицын // Современные проблемы информатизации в проектировании и информационных системах. Сб. трудов. Вып. 14/ Под ред. д.т.н. проф. О.Я.Кравца – Воронеж: “Научная книга”. – 2009. – С.487-491.
99. Делицын, Л.Л. Некоторые вопросы расчета кумулятивных кривых растущего охвата аудитории WWW-ресурсов / Л.Л. Делицын // «Актуальные проблемы и инновации в экономике, управлении, образовании, информационных технологиях». Матер. междунар. научн. конф. (12-15 мая 2009 г., Ставрополь-Кисловодск). – В 4 т. Вып.5.т.3. – Ставрополь. – 2009. – С.89-92.
100. Делицын, Л.Л. Некоторые иерархические многостадийные модели распространения нововведений / Л.Л. Делицын // Научно-технические ведомости СПбГПУ. Экономические науки. – 2009. – №5. – С.279-287.
101. Делицын, Л.Л. Необходимые условия роста использования Интернета в регионах России в период мировой экономической рецессии / Л.Л. Делицын // Доклад на Второй ежегодн. межрег. научно-практ. конф. “Инфокоммуникационные технологии в региональном развитии”. – Смоленск: Администрация Смоленской области. – 2009. – Электрон. текстов. ресурс. - http://expo.smolensk.ru/dokald_09/delicin.doc

102. Делицын, Л.Л. О проблемах использования кумулятивных кривых аудитории WWW-ресурсов при оттоке пользователей / Л.Л. Делицын // Сб. матер. IV-й Междунар. научно-практ. конф. «Маркетинг и общество». – Казань: КГФЭИ. – 2009. – С.87-90.
103. Делицын, Л.Л. О проблеме согласования различных способов подсчета охвата аудитории WWW-ресурсов / Л.Л. Делицын // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. – 2009. – №1. – С.164-170.
104. Делицын, Л.Л. Применение многостадийной модели диффузии инноваций для исследования процесса распространения Интернета в России / Л.Л. Делицын // Маркетинг взаимодействия в инновационной экономике. Материалы междунар. научн. конф. 27-29 сент. 2009. Ч.1 / Под ред. д.э.н., проф. И.А.Максимцева. – СПб.: СПбГУЭФ. – 2009. – С.10-16.
105. Делицын, Л.Л. Применение многостадийной модели диффузии нововведений для объяснения показателей распространения Интернета в России / Л.Л. Делицын // «Синергетика природных, технических и социально–экономических систем»: Сб. статей VI междунар. заочн. научн. конф. (май 2009) - Тольятти. Изд-во ПВГУС. – 2009. – С.112-119.
106. Делицын, Л.Л. Прогнозирование распространения Интернета в России при помощи модели диффузии нововведений / Л.Л. Делицын // Матер. V междунар. научно-практ. конференции “Экономическое прогнозирование: модели и методы”. – Воронеж. – 2009. – С.112-120.
107. Делицын, Л.Л. Простейшая многостадийная модель распространения Интернета в России и в Москве / Л.Л. Делицын // Сб. статей VI междунар. научно-практ. конф. «Инновационные процессы в менеджменте». - Под ред. А.И.Кравченко, Г.В.Синцова, Г.С. Широкаловой. – Пенза: Приволжский Дом Знаний. – 2009. – С.11-15.
108. Делицын, Л.Л. Распространение информационных и телекоммуникационных технологий в России как многостадийный процесс диффузии нововведения в неоднородном обществе / Л.Л. Делицын // Социология инноватики: Человек в инновационном мире. Материалы III междунар. конф. / Под ред. Ю.А. Карповой. – М.: Инион РАН. – 2009. – С.68-73.
109. Делицын, Л.Л. Стохастическая модель охвата аудитории WWW-ресурсов в условиях замены файлов cookies / Л.Л. Делицын // Труды междунар. конф. «Перспективы развития телекоммуникационных систем и информационные технологии». / Под ред. д-ра экон. наук, проф. А.В.Бабкина, д-ра техн. наук, проф. В.А.Кежаева. – СПб.: Изд-во Политехн. ун-та. – 2009. – С.195-212.
110. Делицын, Л.Л. Эпидемическая SIR-модель жизненного цикла нововведения / Л.Л. Делицын // Сб. статей VII междунар. научно-практ. конф. «Инновационные технологии научных исследований социально-экономических процессов». - Под ред. д.п.н. В.П. Линьковой. – Пенза: Приволжский Дом Знаний. – 2009. – С.38-43.
111. Делицын, Л.Л. Количественные модели распространения нововведений в сфере информационных и телекоммуникационных технологий / Л.Л. Делицын. – М.: МГУКИ, 2009. – 106 с.
112. Делицын, Л.Л. Анализ динамики проникновения сотовой связи в домохозяйствах и среди индивидов на основе модели диффузии нововведений с растущим потенциалом проникновения / Л.Л. Делицын // Посткризисные очертания инновационных процессов: Материалы Десятых Друкеровских чтений // Под ред. Р.М.Нижегородцева. М. – Новочеркасск: ЮРГТУ (НПИ), 2010. – с.320-329.

113. Делицын, Л.Л. Анализ итогов совместного прогнозирования компьютерного рынка и парка компьютеров РФ на основе новых статистических данных / Л.Л. Делицын // Труды всерос. конф. «Экономика и инновации в промышленности». Под ред. д-ра экон. наук, проф. А.В.Бабкина. – СПб.: Изд-во Политехн. ун-та. – 2010. – С.18-25.
114. Делицын, Л.Л. Вычисление количества пользователей инновации с экспоненциально растущим потенциалом при помощи обобщенного интеграла Доусона / Л.Л. Делицын // Приоритетные направления современной российской науки. Сб.матер. междунар. научно-практ. конф. 25-26.01.2010. – Тамбов. Тамбовпринт. – 2010. – С.30-35.
115. Делицын, Л.Л. Демографическое обоснование модели диффузии нововведения с линейно растущим потенциалом / Л.Л. Делицын // «Маркетинг и общество»: Сб. матер. V Междунар. научно-практ. конф. – Казань: Изд-во КГФЭИ. – 2010. – С.57-59.
116. Делицын, Л.Л. Интерпретация данных массовых опросов об использовании мобильной связи при помощи модели диффузии нововведений с растущим потенциалом распространения / Л.Л. Делицын // Сб. статей VII междунар. научно-практ. конф. «Инновационные процессы в менеджменте». – Под ред. А.И.Кравченко, Г.С. Широкаловой. – Пенза: Приволжский Дом Знаний. – 2010. – С.22-29.
117. Делицын, Л.Л. Использование специальных функций для вычисления количества пользователей инновации с растущим потенциалом распространения / Л.Л. Делицын // «Актуальные вопросы современной науки». Матер. VII Междунар. Интернет-конф. (Таганрог, 31 янв. 2010 г.); Сб.научн. трудов/ под ред. д.пед.н., проф.Г.Ф.Гребенникова. – М.:Спутник+. – 2010. – С.341-344.
118. Делицын, Л.Л. Количественная модель распространения сотовой связи как инновации с линейно растущим потенциалом распространения / Л.Л. Делицын // Научные труды вольного экономического общества России. – 2010. – т.137. – С.884-900.
119. Делицын, Л.Л. Моделирование распространения Интернета с учетом неоднородности структуры влияния в обществе, процессов воспроизводства населения и снижения стоимости доступа / Л.Л. Делицын // //Труды Третьей Междунар. конф. “Математическое моделирование социальной и экономической динамики (MMSED-2010). 23-25 июня 2010 г. / Науч.ред. М.Г.Дмитриев, А.П.Петров, Н.П.Третьяков. – М.:ЛЕНАНД. – 2010 г. – С.87-89/.
120. Делицын, Л.Л. Моделирование распространения Интернета в Москве / Л.Л. Делицын // Математика. Компьютер. Образование. Тезисы докл. XVII междунар. конф. г.Дубна 25-30 янв. 2010 / - М.-Ижевск. - 2010. – С.335.
121. Делицын, Л.Л. Модель диффузии нововведений в обществе, неоднородном по склонности реципиентов распространять информацию / Л.Л. Делицын // Шестые Курдюмовские чтения Международная междисциплинарная научная конференция «Синергетика в естественных науках» с элементами научной школы для молодежи. – Тверь: Твер. Гос. Ун-т. – 2010 г. – С.87-90.
122. Делицын, Л.Л. Некоторые модели диффузии инноваций в маркетинге инфокоммуникационных услуг / Л.Л. Делицын // Труды вольного экономического общества России. – 2010. – т.130. – С.412-420.
123. Делицын, Л.Л. О близости результатов расчетов при помощи двух моделей распространения нововведений при помощи двух принципиально различных моделей / Л.Л. Делицын // Высокие технологии, фундаментальные исследования, промышленность. Т4: Сб. трудов Девятой междунар. научно-практ. конф. “Исследование, разработка и

- применение высоких технологий в промышленности” 22-23.04.2010, СПб./ Под ред. А.П. Кудинова. – СПб.: Изд-во Политехн. ун-та. – 2010. – С.98-103.
124. Делицын, Л.Л. О верхней оценке количества пользователей инновации с экспоненциально растущим потенциалом / Л.Л. Делицын // Системный анализ в проектировании и управлении: сб. науч. трудов XIV междунар. научно-практ. конф. Ч.1. СПб.: Изд-во Политехн. ун-та, 2010. – С. 291-298.
 125. Делицын, Л.Л. Опыт совместного прогнозирования компьютерного рынка и парка компьютеров Российской Федерации / Л.Л. Делицын // Материалы VI междунар. научно-практ. конференции “Экономическое прогнозирование: модели и методы”, 6 апреля 2010: в 2.ч./под общ.ред.В.В. Давниса, В.И.Тиняковой. – Воронеж: ИПЦ ВГУ – 2010. – С.130-135.
 126. Делицын, Л.Л. Применение интегральной показательной функции для расчета количества пользователей инновации с экспоненциально растущим потенциалом / Л.Л. Делицын // Методы и алгоритмы прикладной математики в технике, медицине и экономике: матер. IX Междунар. научно-практ. конф. – 26/02/2010. – Новочеркасск: ЮРГТУ (НПИ). – 2010. – С.50-59.
 127. Делицын, Л.Л. Применение модели диффузии нововведений для стратегического планирования распространения Интернета в России / Л.Л. Делицын // Тезисы доклада на XI всерос. симпозиуме “Стратегическое планирование и развитие предприятий”. – М.: ЦЭМИ РАН. – 13-14 апр. 2010. – С.68-70.
 128. Делицын, Л.Л. Прогнозирование распространения Интернета в России при помощи модели диффузии нововведений / Л.Л. Делицын // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. – 2010. - №1. – С.74-82.
 129. Делицын, Л.Л. Прогнозы распространения Интернета в США и Швеции на основе модели диффузии инновации с линейно растущим потенциалом / Л.Л. Делицын // Приоритетные направления современной российской науки. Сб.матер. междунар. научно-практ. конф. 25-26.01.2010. – Тамбов. Тамбовпринт. – 2010. – С.15-18.
 130. Делицын, Л.Л. Распространение инноваций в однородном обществе как стационарный поток событий / Л.Л. Делицын // Инноватика-2010. Труды междунар. конф. Под ред. Булярского С.В. – Ульяновск: УлГУ, 2010. – С.233-234.
 131. Делицын, Л.Л. Расчет количества пользователей инновации с линейно растущим потенциалом распространения при помощи интеграла Доусона / Л.Л. Делицын // «Теория и практика управления инновационным развитием социально-экономических систем». Труды междунар. научно-практ. конф. Ростов-на-Дону (24-25 дек.2009). Часть I. – Ростов-на-Дону: СКАГС, 2010. – С.48-55.
 132. Делицын, Л.Л. Скоро ли пользователь Интернета станет похож на среднего россиянина? / Л.Л. Делицын // Онлайн-исследования в России 2.0. Под ред. Шашкина А.В., Девятко И.Ф., Давыдова С.Г. – М.: РИЦ “Северо-Восток”. – 2010. – С. 31-42.
 133. Делицын, Л.Л. Совместное моделирование компьютерного рынка и парка компьютеров на основе системного подхода / Л.Л. Делицын // Системный анализ в экономике. Материалы научно-практ.конф. Москва, 24-25 ноября 2010. / Под ред. чл.-корр.РАН, Г.Б.Клейнера. – М.: ЦЭМИ РАН, 2010. – с.54-56.
 134. Делицын, Л.Л. Совместное прогнозирование компьютерного рынка и парка компьютеров при помощи модели распространения нововведения / Л.Л. Делицын // Труды вольного экономического общества России. – 2010. – т.130. – С.421-429.

135. Делицын, Л.Л. Согласование данных массовых опросов об интенсивности использования Интернета при помощи отрицательного биномиального распределения вероятностей / Л.Л. Делицын // Интернет и современное общество. Труды XVIII Всероссийской объединенной конференции, 19-22 октября 2010 г. / СПб.: Изд-во СПбГУИТМО. – 2010. – С. 23-26.
136. Делицын, Л.Л. Способ оценки индекса сезонности продаж при помощи онлайн-опросов для оценки распределения продолжительности жизни товара / Л.Л. Делицын // Труды всерос. конф. «Экономика и инновации в промышленности». / Под ред. д-ра экон. наук, проф. А.В.Бабкина, д-ра экон. наук, проф. В.В.Кобзева. – СПб.: Изд-во Политехн. ун-та, 2010. – С.18-25.
137. Делицын, Л.Л. Сравнение бета-биномиальной и дискретной моделей поведения онлайн-панелей / Л.Л. Делицын // Современные проблемы информатизации в экономике и обеспечении безопасности: Сб. трудов. Вып. 15 / под ред. д.т.н. проф. О.Я.Кравца. – Воронеж: Научная книга. – 2010. – С.71-76.
138. Делицын, Л.Л. Сравнение способов контакта групп, принимающих решение, при моделировании распространения мобильной связи в домохозяйствах / Л.Л. Делицын // Труды Третьей Междунар. конф. “Математическое моделирование социальной и экономической динамики” (MMSED-2010). 23-25 июня 2010 г. / Науч.ред. М.Г.Дмитриев, А.П.Петров, Н.П.Третьяков. – М.:ЛЕНАНД. – 2010 г. – С.89-91.
139. Делицын, Л.Л. Сравнительный анализ распространения Интернета в России и некоторых европейских странах / Л.Л. Делицын // Вестник МГУ. Серия журналистика. – №4, 2010. – С.23-39.
140. Делицын, Л.Л. Статистика распространения Интернета в Москве и в Швеции: чего ждать дальше? / Л.Л. Делицын // Интернет маркетинг. – 2010. – №2. – С.66-82.
141. Делицын, Л.Л. Элементы префигуративного типа культуры в системе неопросветительства / Л.Л. Делицын // Феномен неопросветительства / Под ред. А.Ю.Алексеева, С.Ю.Карпук. – М.: ИИнтелл, 2010. – С.158-164.
142. Делицын, Л.Л. Моделирование охвата аудитории WWW-сайтов при помощи бета-биномиальной вероятностной смеси / Л.Л. Делицын // Междунар. конф. Информационное общество: Состояние и тенденции межгосударственного обмена научно-технической информацией в СНГ. – М.: ВИНТИ РАН, 2011. – С.45-47.
143. Делицын, Л.Л. Совместное моделирование компьютерного рынка и парка компьютеров на основе статистических данных / Л.Л. Делицын // Научно-технические ведомости СПбГПУ. Экономические науки. – 2011. - №1. – С.265-272.
144. Делицын, Л.Л. Количественная модель распространения Интернета среди городского и сельского населения России / Л.Л. Делицын // Вестник Костромского государственного университета им. Н.А. Некрасова. — 2012. — Т. 18, № 6. — С. 52–56.
145. Делицын, Л.Л. Об одном обобщении модели CGS распространения нововведения / Л.Л. Делицын // Системы компьютерной математики и их приложения, 2013. № 14. С. 148-150.
146. Делицын, Л.Л. Разработка и применение количественных моделей распространения новых информационных технологий / Л.Л. Делицын // Научно-техническая информация. Серия 1: Организация и методика информационной работы. – 2014. – №5. – С.24-32.
147. Делицын, Л.Л. Состояние российского Интернета на сегодняшний день / Л.Л. Делицын, И.И. Засурский // Маркетинг в России и за рубежом. – 2003. – №2. – С.80 - 86.

148. Делицын, Л.Л. Состояние российского Интернета на сегодняшний день / Л.Л. Делицын, И.И.Засурский // Маркетинг в России и за рубежом. — 2003. — №2. — С. 80–86.
149. Делицын, Л.Л. Традиционный бизнес пошёл в Интернет / Л.Л. Делицын, И.И. Засурский // Интернет-Маркетинг. — 2003. — №3. — С.2-10.
150. Делицын, Л.Л. Проблемы оценки стоимости бизнеса малых и средних компаний / Л.Л. Делицын, С.А. Харченко // Фактор риска. — 2006. — Сентябрь-октябрь. — С.18-19.
151. Делицын, Л.Л. Моделирование распространения нововведения в неоднородной социально-экономической системе при помощи свободного программного обеспечения / Л.Л. Делицын, С.А. Белоцерковский // Открытое образование. — 2011. — №2(86). — Ч.2 — С.159-262.
152. Делицын, Л.Л. Модель распространения нововведения в обществе с функцией дожития первого типа / Л.Л. Делицын, С.А. Белоцерковский // Вторая всерос. научно-практ. Конф. "Проблемы анализа и моделирования региональных социально-экономических процессов", 21-22 апреля 2011 г.— Казань: Изд-во КГФЭИ. — 2011. — С.124-127.
153. Делицын, Л.Л. Некоторые вопросы информационно аналитической поддержки венчурного инвестирования в новый способ предоставления программного обеспечения / Л.Л. Делицын, Н.С. Мациевский // Вестник Академии экономической безопасности МВД России. - 2011. - № 7. — С. 76-83.
154. Делицын, Л.Л. Об одном способе обобщения модели CGS распространения нововведения / Л.Л. Делицын, Т.А. Подлесная // Седьмые Курдюмовские чтения Международная междисциплинарная научная конференция «Синергетика в естественных науках» с элементами научной школы для молодежи. — Тверь: Твер. Гос. Ун-т. — 2011 г.
155. Делицын, Л.Л. Решение уравнений CGS распространения нововведения в стабильной популяции / Л.Л. Делицын, С.А. Белоцерковский // Математическое моделирование и краевые задачи. Труды восьмой Всероссийской научной конференции с международным участием. Ч.3.: Дифференциальные уравнения и краевые задачи. - Самара: СамГТУ, 2011. - С.46-49.
156. Делицын, Л.Л. Точные решения уравнений распространения нововведения в стабильной популяции с функцией дожития Гомперца-Мейкема / Л.Л. Делицын, С.А. Белоцерковский // XI Международная науч.-практ.конф. "Моделирование. Теория, методы и средства". - Новочеркасск: ЮРГТУ (НПИ) - 2011. - С.55-63.
157. Делицын, Л.Л. / Л.Л. Делицын, С.А. Белоцерковский // Некоторые точные решения уравнений распространения нововведений в стабильной популяции с возрастной структурой / Методы и алгоритмы прикладной математики в технике, медицине и экономике: матер. XI Междунар. научно-практ. конф. — 26/02/2011. — С.23-29.
158. Делицын, Л.Л. Скрининг венчурных проектов в отрасли ИКТ / Л.Л. Делицын, Е.А. Кобзарева // Научные труды вольного экономического общества России. — 2012. — Т.174.
159. Делицын, Л.Л. Прогнозирование глобальной аудитории Интернета на основе динамики ее региональных сегментов / Л.Л. Делицын, Н.С. Матвеева // Научные труды вольного экономического общества России. — 2013. — Т.179. — С.248-256.
160. Делицын, Л.Л. Моделирование охвата аудитории WWW-сайтов при помощи смеси бета-биномиальных распределений / Л.Л. Делицын, Т.А. Подлесная // Труды вольного экономического общества России. — 2011. — т.155. — С.120-128.
161. Делицын, Л.Л. Стохастические модели распространения новых идей в социальных сетях / Л.Л. Делицын, Т.А. Подлесная // Открытое образование, 2011. №2(86). — Ч.2 — С.242-244.

162. Делицын, Л.Л. Коллективный разум / Л.Л. Делицын, И.И. Засурский, И.Г. Куринной // Информационное общество. – 2003. – вып. 3. – С.32-49.
163. Делицын, Л.Л. Развитие информационных и телекоммуникационных технологий в России как процесс распространения инноваций. Дисс. ...канд. техн. наук / Л.Л. Делицын. – МГУКИ, 2006. – 293 с.
164. Делицын, Л.Л. Обобщенная модель CGS распространения нововведения в растущей популяции / Л.Л. Делицын // Труды вольного экономического общества России. – 2010. – т.164. – С.412-420.
165. Дмитриев, М.Г. Интегро-дифференциальные краевые задачи, возникающие в модели "Власть-общество-экономика" / М.Г. Дмитриев, Г.И. Лаптев // Ученые записки РГСУ. – 2009. – №13. – С.216-221.
166. Дрожжинов, В.И. Что дает госсектору внедрение SaaS, PaaS и IaaS? [Электронный ресурс] Доклад на круглом столе «SaaS: расширение сервиса», организованном CNews Analytics и CNews Conferences [Электронный ресурс] / В.И. Дрожжинов. – Электрон. текстовые дан. – Режим доступа: http://www.cnews.ru/reviews/ppt/04_03_2010/drozhzhinov.ppt. – Загл. с экрана.
167. Дуброва, Т.А. Статистические методы прогнозирования / Т.А. Дуброва. – М.: ЮНИТИ, 2003. – 208 с.
168. Ершова, Т.В. Использование ИКТ домохозяйствами и населением России: результаты исследований и актуальных рейтингов / Т.В.Ершова // Экономика, Статистика и Информатика. Вестник УМО: Научно-практический журнал. – 2011. – №5. – С.152-156.
169. Ершова, Т.В. Опрос населения и домохозяйств как инструмент развития информационного общества / Т.В. Ершова, С.Б. Шапошник // Информационное общество: научно-аналитический журнал. – 2011. – №6. – С.12-31.
170. Жеребин, В.М. Информационная составляющая в бюджете российской семьи / В.М. Жеребин // ФЭС: Финансы. Экономика. Стратегия. – 2010. – №8. – С.35-40.
171. Жожикашвили, В.А. Сети массового обслуживания. Теория и применение к сетям ЭВМ / В.А. Жожикашвили, В.М. Вишневский. – М.: Радио и связь, 1988. – 192 с.
172. Засурский, Я.Н. К мобильному обществу: утопия и реальность / Я.Н.Засурский (ред.). – М.: МГУ, 2009. – 304 с.
173. Земсков, А.И. Электронные библиотеки: учеб. для вузов / А.И. Земсков, Я.Л. Шрайберг. – М.: Либерия, 2003. – 351 с.
174. Иванов, И.Ф. Использование логистической кривой для оценки стоимости компании на развивающемся рынке / И.Ф. Иванов // Корпоративные финансы. – 2008. – №1(5). – С.47-62.
175. Ивина, Л.В. Венчурный бизнес. Толковый словарь труднопереводимых англоязычных терминов. / Л.В. Ивина, В.А. Воронков. – М.: Финансы и статистика; Финансовый контроль, 2006. – 232 с.
176. Измоденова, К.В. Об оптимальном управлении процессом распространения информации / А.П. Михайлов, К.В.Измоденова // Математическое моделирование. – 2005. – Т.17, №6. – С.67-76.
177. Казанцев, С.Ю. Состояние и потенциал развития инфокоммуникационного комплекса России / С.Ю. Казанцев, И.Э. Фролов // Проблемы прогнозирования. – 2005. – №3. – С.17-40.

178. Казанцев, С.Ю. Условия и потенциал развития российского инфокоммуникационного комплекса / С.Ю. Казанцев, И.Э. Фролов // Проблемы прогнозирования. – 2006. – №4. – С.80-97.
179. Караева, Э.А. Прогнозы проникновения информационно-коммуникационных технологий в России / Э.А. Караева, Л.Л. Делицын // Макромодели Василия Леонтьева и перспективы развития российской и мировой экономики. Материалы XX междисциплинарной дискуссии / Под ред. В.И. Кушлина, Г.Ю. Ивлевой, Ю.В. Яковца. - М.: Проспект. - 2006. - с.225-230.
180. Карданская, Н.Л. Управленческие решения / Н.Л. Карданская. – М.: Юнити-ДАНА, 2009. – 440 с.
181. Карпова, Ю.А. Введение в социологию инноватики: Учебное пособие / Ю.А. Карпова. – СПб: Питер, 2004. – 192 с.
182. Каширин, А.И. Венчурное инвестирование в России / А.И. Каширин, А.С. Семенов. – М.: Вершина, 2007. – 320 с.
183. Кепман, М. АИБ ищет россиян среди кукисов [Электронный ресурс] / М. Кепман. – Электрон. текстовые дан. – Режим доступа: http://telnews.ru/Mihail_Kepman/c37571 . – Загл. с экрана.
184. Клейнрок, Л. Теория массового обслуживания. Пер. с англ. / Пер. И.И. Грушко; ред. В.И. Нейман. / Л. Клейнрок. – М.: Машиностроение, 1979. – 432 с.
185. Клочков, В.В. Проблемы прогнозирования спроса на перспективные гражданские самолеты российского производства / В.В. Клочков, Т.М. Гусманов // Проблемы прогнозирования. – 2007. – 2. – С.16-31.
186. Кобленц-Мишке, Ю.А. Пороговые модели спроса на предметы длительного пользования с насыщаемой потребностью. Модели анализа и планирования показателей уровня жизни. / Ю.А. Кобленц-Мишке // Пороговые модели спроса на предметы длительного пользования с насыщаемой потребностью // Модели анализа и планирования показателей уровня жизни. – М.: ЦЭМИ АН СССР. – 1987. – с.31-42
187. Кобленц-Мишке, Ю.А. Моделирование зависимости потребления предметов длительного пользования от динамики распределения доходов // Дисс. ...канд. экон.наук / Ю.А. Кобленц-Мишке. – М.: ЦЭМИ РАН, 1994.
188. Кови, С. Семь навыков высокоэффективных людей: Мощные инструменты развития личности / С. Кови. – М.: Альпина Паблишер, 2012. – 374 с.
189. Колесин, И.Д. Математические модели субкультур / И.Д. Колесин. – СПб.: изд. СПбГУ, 2007. – 134 с.
190. Колесин, И.Д. Математические модели самоорганизации в социокультурных системах / И.Д. Колесин. – СПб.: изд. СПбГУ, 2009. – 148 с.
191. Колесов, А.А. Есть ли SaaS в России? [Электронный ресурс] / А.А. Колесов. – Электрон. текстовые дан. – Режим доступа: <http://www.itoday.ru/16070.html>. – Загл. с экрана.
192. Колесов, А.А. Затянувшиеся роды SaaS в России [Электронный ресурс] / А.А. Колесов. – Электрон. текстовые дан. – Режим доступа: <http://www.itoday.ru/32982.html>. – Загл. с экрана.
193. Колесов, А.А. Анатомия SaaS: истоки и суть модели [Электронный ресурс] / А.А. Колесов. – Электрон. текстовые дан. – Режим доступа: <http://www.pcweek.ru/its/article/detail.php?ID=147721>. – Загл. с экрана.

194. Колесов, А.А. Российские SaaS-разработчики смотрят в будущее с оптимизмом [Электронный ресурс] / А.А.Колесов. – Электрон. текстовые дан. – Режим доступа: <http://www.pcweek.ru/its/article/detail.php?ID=157353>. – Загл. с экрана.
195. Колин, К.К. Информационное неравенство – новая проблема XXI века // Социология, социальность, современность. - М.:Союз. - 1998. / К.К. Колин // Информационное неравенство – новая проблема XXI века // Социология, социальность, современность. – М.:Союз. – 1998.
196. Колин, К.К. Фундаментальные основы информатики: социальная информатика: Учебное пособие для вузов / К.К. Колин. – М.: Академический Проект; Екатеринбург: Деловая книга., 2000. – 350 с.
197. Количество российских пользователей интернета выросло в 2007 году на 29% [Электронный ресурс] / . – Электрон. текстовые дан. – Режим доступа: <http://a-ib.ru/node/54>. – Загл. с экрана.
198. Концепция долгосрочного социально-экономического развития Российской Федерации на период до 2020 года. Утверждена распоряжением Правительства Российской Федерации от 17 ноября 2008 г. № 1662-р
199. Конявский, В.А. Прогноз развития российского сегмента сети Интернет до 2010 г. / В.А. Конявский // Управление защитой информации. – 2003. – Т.7. – №3.
200. Корн, Г. Справочник по математике для научных работников и инженеров / Г. Корн, Т. Корн. – М.: Наука, 1984. – 831 с.
201. Корнышев Ю.Н. Теория телетрафика / Ю.Н. Корнышев, А.П. Пшеничников, А.Д. Харкевич. – М.: Радио и связь, 1996. – 272 с.
202. Коротков, А.В. Преодоление цифрового неравенства как стратегия развития информационного общества. Дисс. ... канд. филол. наук /А.В. Коротков. – М.: 2003. –183с.
203. Крамин, Т.В. К оценке потенциала и структуры спроса на услуги сотовой связи в регионах России / Т.В. Крамин // Проблемы современной экономики. – 2004. – №1-2. – С.152-155.
204. Кристенсен, К.М. Дилемма инноватора. Как из-за новых технологий погибают сильные компании / К.М. Кристенсен. – Альпина Бизнес Букс, 2004. – 240 с.
205. Кристенсен, К.М. Решение проблемы инноваций в бизнесе. Как создать растущий бизнес и успешно поддерживать его рост / К.М. Кристенсен, М.Е. Рейнер. – Альпина Бизнес Букс, 2004. – 290 с.
206. Кузовкова, Т.А. Анализ и прогнозирование развития инфокоммуникаций / Т.А. Кузовкова, Л.С. Тимошенко. – М.: ё, 2009. – 224 с.
207. Лаврентьева, Н. Российский стартап получил \$500 тыс. на “облачное” ПО [Электронный ресурс] / Н. Лаврентьева. – Электрон. текстовые дан. – Режим доступа: <http://www.cnews.ru/news/top/index.shtml?2011/03/05/430977>. – Загл. с экрана.
208. Лахути, Д.Г. Милль, Пирс и Поппер о логике научного открытия / // Вопросы философии. – 2012. – №3. – С.101-109.
209. Лебедев, П.А. Российский Интернет в марте 2009: базовые параметры, медиапредпочтения и Интернет-сервисы / П.А. Лебедев, Е.Б.Галицкий // Доклад на объединенной конференции РИФ+КИБ. – 2009. – 22.04.2009.
210. Лебедева, Н.М. Культура и инновации: к постановке проблемы / Н.М. Лебедева, Е.Г.Ясин // Форсайт. – 2009. – №2(10).

211. Левин, М.И. Модели развития Интернета в России / М.И. Левин, Е.Б. Галицкий, В.Ю. Ковалева, И.Н. Щепина // Интернет-маркетинг. – 2006. – №4. – С.194-207.
212. Линдеман, Е.В. Современные тенденции и проблемы использования комплекса информационных ресурсов в библиотечных сервисах / Е.В. Линдеман // Современные тенденции и проблемы использования комплекса информационных ресурсов в библиотечных сервисах // Доклад на XVII международной конференции «LIVCOM-2013: Информационные технологии, компьютерные системы и издательская продукция для библиотек». – Суздаль. – 2013.
213. Линдеман, Е.В. Информационные сервисы библиотеки с электронными ресурсами: как не нарушить закон / Е.В. Линдеман, М.В. Гончаров, К.А. Колосов // Университетская книга. – 2012. – 8. – С.58-65.
214. Лисин, Б.К. Стратегический ресурс инноваций / Б.К. Лисин. – М.: Институт стратегических инноваций, 2000. – 48 с.
215. Лопатина, Н.В. Социология управления информатизацией: методологический статус и позиции в структуре современной социологии / Н.В. Лопатина // Вестник МГУКИ. – 2008. – 2. – С.78-82.
216. Лопатина, Н.В. Управление информатизацией как приоритетная задача социального развития / Н.В. Лопатина // Информационные ресурсы России. – 2005. – №1. – С.23-25.
217. Максиянова, Т.В. Международные показатели интенсивности использования Интернета в системе Интернет-экономики / Т.В. Максиянова // Вопросы современной науки и практики. Университет им. В.И.Вернадского. – 2012. – №4(42). – С.272-280.
218. Марчук, Г.И. Математические модели в иммунологии. Вычислительные методы и эксперименты / Г.И. Марчук (ред.). – М.: Наука, 3-е изд., 1991. – 276 с.
219. Матохин, В.В. Корпоративная система сбора и обработки предложений ("Информационный Инкубатор Инновационных Идей")[Электронный ресурс]«Система 4И» [Электронный ресурс] / В.В. Матохин. – Электрон. текстовые дан. – Режим доступа: <http://www.4ionline.ru/>. – Загл. с экрана.
220. Медынский, В.Г. Инновационный менеджмент / В.Г. Медынский. – М.: Инфра-М, 2008. – 295 с.
221. Мид, М. Культура и мир детства. Избранные произведения / Сост. и предисл. И.С.Кона. / М. Мид. – М.: Наука, 1983. – 429 с.
222. Михайлов, А.И. Основы информатики / А.И. Михайлов, А.И. Черный, Р.С. Гиляревский. – М.: Наука, 1968. – 756 с.
223. Михайлов, А.П. Модели информационной борьбы / А.П. Михайлов, Н.А. Маревцева // Математическое моделирование. – 2011. – Т.23, №10. – С.19-32.
224. Михайлов, А.П. Поведенческие гипотезы и математическое моделирование в гуманитарных науках / А.П. Михайлов, А.П. Петров // Математическое моделирование. – 2011. – Т.23, №6. – С.18-32.
225. Михайлов, А.П. О пределах инновационных заимствований / А.П. Михайлов, В.Ж. Келле, В.А. Шведовский // Социология: 4М - методология, методы, математические модели. – 2001. – №13. – С.114-123.
226. Московкин, В.В. Основы концепции диффузии инноваций / В.М. Московкин // Бизнес-информ. – 1998. – №17-18. – С.41-48.

227. Мышкис, А.Д. Смешанные функционально-дифференциальные уравнения / А.Д. Мышкис // Новые проблемы теории функционально-дифференциальных уравнений, СМФН. – 2003. – 4. – С.5-120.
228. Национальные модели, Национальные модели информационного общества / Отв. ред. и сост. Е.Л. Вартанова, науч. ред. Н.В.Ткачева. – М.: Икар, 2004. – 412 с.
229. Национальные модели, Национальные модели информационного общества / Отв. ред. и сост. Е.Л. Вартанова, науч. ред. Н.В.Ткачева. – М.: Икар, 2004. – 412 с.
230. Нижегородцев, Р.М. Логистическое моделирование экономической динамики, Ч.І. / Р.М. Нижегородцев // Проблемы управления. – 2004. – 1. – С.46-53.
231. Нижегородцев, Р.М. Логистическое моделирование экономической динамики, Ч.ІІ. / Р.М. Нижегородцев // Проблемы управления. – 2004. – 2. – С.52-58.
232. Нижегородцев, Р.М. Модели клеточных автоматов в теории диффузии инноваций / Р.М. Нижегородцев, В.Д. Секерин, С.В. Лисафьев // Вопросы новой экономики. – 2012. – 3.
233. Операторы предполагают снижать объемы строительства сетей ШПД, Операторы предполагают снижать объемы строительства сетей ШПД [Электронный ресурс] / . – Электрон. текстовые дан. – Режим доступа: <http://www.telecomru.ru/article/?id=5090&PHPSESSID=811b2506b328a071e3228cd2e3818bd3> . – Загл. с экрана.
234. Ослон, А.А. Интернет: реалии и возможности // Доклад на круглом столе Института современного развития [Электронный ресурс] / А.А. Ослон. – Электрон. текстовые дан. – Режим доступа: 1. <http://bd.fom.ru/pdf/int0209.pdf>. – Загл. с экрана.
235. Ослон, А.А. Интернет: реалии и возможности // Доклад на круглом столе Института современного развития [Электронный ресурс] / А.А. Ослон. – Электрон. текстовые дан. – Режим доступа: 1. <http://bd.fom.ru/pdf/int0209.pdf>. – Загл. с экрана.
236. Павловский, Ю.Н. Имитационное моделирование / Ю.Н. Павловский, Н.В. Белотелов, Ю.И. Бродский. – М.: Академия, 2008. – 240 с.
237. Панкратова, О. Рынок мобильной индустрии России и СНГ / О. Панкратова // Доклад на форуме РУССОФТ . – 2009. – 11 июня.
238. Перфильев, Ю.Ю. Территориальная организация российского интернет-пространства / Ю.Ю. Перфильев // Интернет и российское общество / Под ред. И. Семенова; Моск. Центр Карнеги – М.: Гендальф, 2002. – С.21-47.
239. Перцев, Н.В. Исследование решений одной системы интегродифференциальных уравнений, возникающей в моделях динамики популяций / Н.В. Перцев // Вестник Омского университета. – 1996. – №1. – С.24-26.
240. Перцев, Н.В. Об асимптотическом поведении одной системы линейных дифференциальных уравнений с последействием / Н.В. Перцев // Известия высших учебных заведений. Математика.. – 1996. – №9. – С.48-52.
241. Петров, В.М. Количественные методы в искусствознании / В.М. Петров. – М.: Академический проект, 2004. – 432 с.
242. Плотинский, Ю.М. Модели социальных процессов / Ю.М. Плотинский. – М.: Логос, 2001. – 296 с.
243. Погорлецкая, Ю.И. Кредитный скоринг в системе банковского риск-менеджмента// Автореф. Дисс. на соискание ученой степени канд. экон. Наук / Ю.И. Погорлецкая //

244. Полтерович, В.М. Эволюционная модель взаимодействия процессов создания и заимствования технологий / В.М. Полтерович, Г.М.Хенкин // Экономика и математические методы. – 1988. – Т. XXIV, вып. 6. – С.1071-1083.
245. Попсулин, С. Мировой рынок SaaS растет на 15% в год [Электронный ресурс] / С. Попсулин. – Электрон. текстовые дан. – Режим доступа: <http://www.cnews.ru/news/top/index.shtml?2010/12/14/420016>. – Загл. с экрана.
246. Прогноз социально – экономического развития российской федерации на 2010 год и на плановый период 2011 и 2012 годов, Прогноз социально – экономического развития российской федерации на 2010 год и на плановый период 2011 и 2012 годов [Электронный ресурс] / . – Электрон. текстовые дан. – Режим доступа: <http://www.economy.gov.ru/minec/activity/sections/macro/prognoz/doc1254407742765>. – Загл. с экрана.
247. Пэлфри, Дж. Дети цифровой эры / Джон Пэлфри, Урс Гассер. – М.: Эксмо, 2011. – 368 с.
248. Радаев, В.В. Обычные и инновационные практики в деятельности российского среднего класса / В.В. Радаев // Мир России. – 2003. – Т.12. - №4. – С.89-119.
249. Разроев, Э.А. Инфокоммуникационный бизнес: управление, технологии, маркетинг / Э.А. Разроев. – СПб.: Профессия, 2003. – 352 с.
250. Родионов, И.И. Рынок информационных услуг и продуктов / И.И. Родионов, Р.С. Гиляревский, В.А. Цветкова, Г.З. Залаев. – М.: МК-Периодика, 2002. – 509 с.
251. Романов, П.С. Проблемы адаптации научных библиотек к инновационным процессам в зарубежном библиотековедении / П.С. Романов // Библиосфера. – 2010. – 1. – С.42-45.
252. Россия в цифрах, Россия в цифрах. 2001: кратк. стат. сб. / М.: Госкомстат России, 2001. – 397 с.
253. Россия в цифрах, Россия в цифрах. 2005: кратк. стат. сб. / М.: Росстат, 2005. – 477 с.
254. Россия: Сотовых телефонов стало на 8 млн меньше, Россия: Сотовых телефонов стало на 8 млн меньше / // Cnews. – 2008. – 17 февраля.
255. Росстат, Число крестьянских (фермерских) хозяйств и индивидуальных предпринимателей [Электронный ресурс] / Росстат. – Электрон. текстовые дан. – Режим доступа: http://www.gks.ru/bgd/regl/b08_47/IssWWW.exe/Stg/html/2-14.htm. – Загл. с экрана.
256. Рязанов, Ю.Г. Медиапланирование / Ю.Г. Рязанов, Г.А. Шматов, ред.. – Екатеринбург: Изд. Уральский рабочий, 2002. – 312 с.
257. Самарский, А.А. Математическое моделирование / А.А. Самарский, А.П. Михайлов. – М.: Наука, 2001. – 320 с.
258. Сахал, Д. Технический прогресс: концепции, модели, оценки / Пер. с англ.; Под ред. и вступ. статья А.А. Рывкина / Д. Сахал. – М.: Финансы и статистика, 1985. – 366 с.
259. Секерин, В. Д. Инновации в маркетинге / В.Д. Секерин, Р.М. Нижегородцев, А.Е. Горохова, Д.В. Секерин. – Изд. МГУИЭ, 2013. – 252 с.
260. Сладкова, О.Б. Количественные методы в исследовании социокультурного пространства / О.Б. Сладкова // Личность, культура, общество: Междисциплинарный научно-практический журнал социальных и гуманитарных наук. – 2007. – Том 9., вып. 3. – С.113-124.
261. Сляднева, Н.А. Информационно-аналитическая деятельность: проблемы и перспективы / Н.А. Сляднева // Информационно-аналитический журнал "Факт". – 2006. – №6.

262. Смолян, Г.Л. Интернет в России. Перспективы развития / Г.Л. Смолян, В.Н. Цыгичко, Д.Д. Хан-Магомедов. – М.: Едиториал УРСС, 2004. – 200 с.
263. Соколов, А.В. Общая теория социальной коммуникации: Учебное пособие / А.В. Соколов. – СПб.: Изд-во Михайлова В.А., 2002. – 461 с.
264. Соколов, В.Б. Модели динамики парков искусственных объектов / В.Б. Соколов // Математическое моделирование. – 1990. – Т.2.-№5. – С.28-39.
265. Староверов, О.В. Азы математической демографии / О.В. Староверов. – М.: Наука, 1997. – 158 с.
266. Столяров, Ю.Н. Сущность информации / Ю.Н. Столяров. – М.: МАИ, 2000. – 120 с.
267. Столяров, Ю.Н. Библиотекосведение, библиографоведение и книговедение как единая научная специальность: Полный курс лекций для аспирантов и соискателей / Ю.Н. Столяров. – Орёл: Орловский гос. Ин-т культуры и искусств. ООО ПФ "Оперативная полиграфия", 2007. – 266 с.
268. Сухомесова, И. Сотовое перемирие / И. Сухомесова // Новый Компаньон. Пермская деловая и политическая газета. – 2004. – №21 (314) от 22 июня .
269. Тараканов, К.В. Информатика / К. В. Тараканов. – М.: Книга, 1986. – 304 с.
270. Тараканов, К.В. Аналитическое моделирование систем / К. В. Тараканов, Л.А. Овчаров, А.Н. Тырышкин. – М.: Советское Радио, 1974. – 240 с.
271. Тард, Ж.-Г. Законы подражания / Ж.-Г. Тард. – М.: Академический Проект, 2011. – 304 с.
272. Тимошенко, Л.С. Стратегия устойчивого роста / Л.С. Тимошенко // Информкурьерсвязь. – 2004. – №12. – С.43-48.
273. Толстова, Ю.Н. Основы многомерного шкалирования: учебное пособие / М.: КДУ, 2006. – 160 с.
274. Тявкин, И.В. Математическая модель информационного поиска и оценка эффективности поисковой системы / И.В. Тявкин, В.М. Тютюнник // Вестник Тамбовского государственного технического университета. – 2007. – Т.14,№3. – С.630-643.
275. Уланов, С.В. Скоринговые модели и средства управления рисками для поддержки принятия кредитных решений. Дисс. ...канд.экон.наук / С.В. Уланов. – Ижевск: Ижевский ГТУ, 2007. – 152 с.
276. Ухин, Ю.Ю. Деловая игра "Избирательное распространение информации" / Ю.Ю. Ухин, О.П. Щеглова. – М.: ИПКИР, 1981. – 87 с.
277. Файзуллин, Р.Т. Система оперативного принятия решений на основе экспертных оценок / Р.Т. Файзуллин, Б.И. Ефимов // Информационные системы моделирования и управления. – 2006. – 4(29). – С.527-535.
278. Федорова, Ю. Игра по правилам и без / Ю. Федорова // Связьинвест. Единая техническая политика. – 2010. – №2 (92). – С.57-58.
279. Федорюк, М.В. Обыкновенные дифференциальные уравнения. – 2-е изд., перераб. и доп. / М.В. Федорюк. – М.: Наука. Главная ред. физ.-мат. литературы, 1985. – 448 с.
280. Филиппов, С. Барская любовь / С. Филиппов // Компания. – 2002. – №221.
281. Фонд "Общественное мнение", Интернет в России, Выпуски 1-33 [Электронный ресурс] / Фонд "Общественное мнение". – Электрон. текстовые дан. – Режим доступа: <http://bd.fom.ru/report/map/projects/internet/>. – Загл. с экрана.

282. Фонд "Общественное мнение", СМИ и Интернет [Электронный ресурс] / Фонд "Общественное мнение". – Электрон. текстовые дан. – Режим доступа: <http://fom.ru/SMI-i-internet/>. – Загл. с экрана.
283. Фукунага, К.Е. Введение в статистическую теорию распознавания образов. Пер. с англ / К. Фукунага. – М.: Наука. Главная ред. физ.-мат. литературы, 1979. – 368 с.
284. Хиршман, И.И. Преобразования типа свертки. Пер.с англ. В.П.Потапова под ред.А.О.Гельфонда. / И.И. Хиршман, Д.В. Уиддер. – М.: Изд-во иностранной литературы, 1958. – 312 с.
285. Хмелевский, Ю.И. Самовоспроизводящиеся системы / Ю.И. Хмелевский. – М.: Наука, 1991. – 432 с.
286. Цыгичко, В.Н. Прогнозирование социально-экономических процессов: научное издание. – Изд. 2-е, перераб. и доп. / В.Н. Цыгичко. – М.: КомКнига, 2007. – 240 с.
287. Чугунов, А.В. Российская интернет-аудитория в зеркале социологии / А.В. Чугунов. – СПб.: Изд-во С-Пб. Университета, 2006. – 320 с.
288. Чугунов, А.В. Социология Интернета: методика и практика исследований Интернет-аудитории: учеб. пособ. / А.В. Чугунов. – СПб.: Факультет филологии и искусств СПбГУ, 2007. – 130 с.
289. Шабат, В. SaaS-проекта в России – репортаж с передовой. [Электронный ресурс] / В. Шабат. – Электрон. текстовые дан. – Режим доступа: <http://www.slideshare.net/vshabat/saa-s-5741368>. – Загл. с экрана.
290. Шведовский, В.А. Моделирование распространения информации в смежных социальных группах / В.А. Шведовский // Математические методы в социологическом исследовании. - М.: Наука, 1981. - С.109-167.
291. Шматов, Г.А. Основы медиапланирования. Эвристический подход. Учеб. Пособие / Г.А. Шматов. – Екатеринбург: Изд-во Урал. ун-та, 2005. – 332 с.
292. Шматов, Г.А. Математическая теория медиапланирования / Г.А. Шматов. – Екатеринбург: Институт экономики УРО РАН, 2009. – 330 с.
293. Шрайберг, Я.Л. Первое десятилетие информационного века: влияние информационно-электронной среды на роль и позицию библиотек в развивающемся обществе ежегодный доклад конференции «Крым». Год 2010 / Я.Л. Шрайберг // Судак; Москва: изд.ГПНТБ России, 2010. - 77 с.
294. Шрайберг, Я.Л. Интеграция библиотек в развивающееся информационное общество: что нас ждет впереди?: ежегодный доклад конференции "Крым", год 2012 / Я.Л. Шрайберг // Москва : ГПНТБ России, 2012. - 63 с.
295. Шрайберг, Я.Л. Электронная информация, библиотеки и общество: что нам ждать от нового десятилетия информационного века? : ежегодный доклад Конференции "Крым", год 2011 / Я.Л. Шрайберг // Научн. и техн. б-ки, - 2012. - 63 с.
296. Шрайберг, Я.Л. Электронная книга, будущее библиотеки и общественное сознание: попытка осмысления и предвидения: ежегодный доклад Конференции "Крым", год 2013 / Я.Л. Шрайберг // Москва : ГПНТБ России, 2013. - 71 с.
297. Шрайберг, Я.Л. Интернет-ресурсы и услуги для библиотек. В 2 ч. / Я.Л. Шрайберг, М.В. Гончаров, О.В. Шлыкова. – М.: Либерия, 2001 - 2002. – Ч.1. – 72 с. – (С компьютером на «ты»: Учебно-справочное пособие. Вып. 6). Ч.2. – 104 с. – (С компьютером на «ты»: Учебно-справочное пособие. Вып. 7). с.

298. Шумпетер, Й.А. Теория экономического развития. Капитализм, социализм и демократия / Й.А. Шумпетер. – М.: Эксмо, 2007. – 864 с.
299. Шуровьески, Д. Мудрость толпы. Почему вместе мы умнее, чем поодиночке, и как коллективный разум формирует бизнес, экономику, общество и государство. / Д.Шуровьески. – М.: 2007. – 304 с.
300. Щепина, И.Н. Модели диффузии и замещения нововведений (обзор литературы) / И.Н. Щепина. – М.: ЦЭМИ РАН, 1990.
301. Юрина, Э.А. “Эпидемиологические” модели распространения мобильной связи и Интернета в России / Э.А. Юрина, Л.Л. Делицын // Интернет-маркетинг. – 2008. – №1(43). – С.2-15.
302. Яблонский, А.И. Математические модели в исследовании науки / А.И. Яблонский. – М.: Мысль, 1986. – 352 с.
303. Яковец, Ю.В. Эпохальные инновации XXI века / Ю.В.Яковец. – М.: Междунар. Ин-т П. Сорокина – Н. Кондратьева. – М.: ЗАО “Издательство “Экономика”, 2004. – 444 с.
304. Янч, Э. Прогнозирование научно-технического прогресса / Э. Янч. – М.: Прогресс, 1974. – 568 с.
305. Abramowitz, M. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables / M. Abramowitz, and I. Stegun. – New York: Dover, 1964.
306. Andreev, A. Russian Technology: In the crucible [Электронный ресурс] / A. Andreev. – Электрон. текстовые дан. – Режим доступа: Russian Technology: In the crucible. – Загл. с экрана.
307. Arbesman, S. 2011 preview: Peak Internet comes into view [Электронный ресурс] / S. Arbesman, R. Courtland. – Электрон. текстовые дан.
308. Armstrong, J.S. Sales forecast for existing consumer products and services: Do purchase intentions contribute to accuracy? [Электронный ресурс] / J.S. Armstrong, V.G. Moritz, V. Kumar. – Электрон. текстовые дан. – Режим доступа: <http://www-marketing.wharton.upenn.edu/ideas/pdf/armstrong-salesintentions.pdf>. – Загл. с экрана.
309. Ausubel, J.H. Elektron: Electrical systems in retrospect and prospect / J.H. Ausubel, C. Marchetti // Daedalus. – 1996. – 125(3). – С.139-169.
310. Ausubel, J.H. Toward green mobility: The evolution of transport / J.H. Ausubel, C. Marchetti, P.S. Meyer // European Review. – 1998. – 6(2). – С.143-162 .
311. Bagozzi, R.P. The legacy of the technology acceptance model and a proposal for a paradigm shift / R.P. Bagozzi // Journal of the Association for Information Systems. – 2007. – 8(4). – С.244-254.
312. Bagozzi, R.P. Development and test of a theory of technological learning and usage / R.P. Bagozzi, F.D. Davis, P.R. Warsaw // Human relations. – 1992. – 45(7). – С.660-686.
313. Bailey, N.T.J. The elements of stochastic processes with applications to natural sciences / N.T.J. Bailey. – New York: Wiley, 1964. – 268 с.
314. Banks, R.B. Growth and Diffusion Phenomena: Mathematical Frameworks and Applications (Texts in Applied Mathematics) / R.B. Banks. – Springer, 1993. – 455 с.
315. Bass, F.M. A New Product Growth for Model Consumer Durables / F.M. Bass // Management Science. – 1969. – vol. 15. – С.215-227.
316. Bass, F.M. Heterogeneity in purchase intention, diffusion and exogenous influence / F.M. Bass // In Research Traditions in Marketing, ed. Gilles Laurent, Gary L. Lilien, and Bernard Pras., – 1994. – . – С.224-226.

317. Bass, F.M. Modeling the marketing-mix influence in new-product diffusion / F.M. Bass, D. Jain, T. Krishnan // *New-product diffusion models. International series in quantitative marketing.* – 2000. – vol. 11. – С.99-122.
318. Bass, F.M. Modeling the marketing-mix influence in new-product diffusion / F.M. Bass, D. Jain, T. Krishnan // *International series in quantitative marketing.* – 2000. – vol. 11. – С.99-122.
319. Bass, P.I. Generational Diffusion Model Data: The IBM Mainframe Case [Электронный ресурс] / P.I. Bass. – Электрон. текстовые дан. – Режим доступа: <http://www.basseconomics.com/Downloads/Papers/Bass%202003.pdf>. – Загл. с экрана.
320. Bass, P.I. Multiple Generation Diffusion Model Estimation Using Adaptive Nonlinear Least Squares Algorithm NL2SOL. [Электронный ресурс] / P.I. Bass, J. Isaacson. – Электрон. текстовые дан. – Режим доступа: <http://www.basseconomics.com/Downloads/Papers/Bass%20and%20Isaacson%202003.pdf>. – Загл. с экрана.
321. Bass, P.I. IT Waves: Two Completed Generational Diffusion Models [Электронный ресурс] / P.I. Bass, F.M. Bass. – Электрон. текстовые дан. – Режим доступа: <http://www.basseconomics.com/Downloads/Papers/Bass%20and%20Bass%202004.pdf>. – Загл. с экрана.
322. Bemmaor, A.C. Modeling the diffusion of new durable goods: word-of-mouth effect versus consumer heterogeneity / A.C. Bemmaor // *In Research Traditions in Marketing*, ed. Gilles Laurent, Gary L. Lilien, and Bernard Pras,. – 1994. – С.201-223.
323. Berndt, E.R. Econometric estimates of price indexes for personal computers in the 1990's / E.R. Berndt, Z. Griliches, N.J. Rappoport // *Journal of Econometrics.* – 1995. – Vol.68, Issue 1. – С.243-258.
324. Bewley, R. A flexible logistic growth model with applications in telecommunications / R. Bewley, Fiebig D.G. // *International journal of forecasting.* – 1988. – 4. – С.177-192.
325. Botelho, A. The diffusion of cellular phones in Portugal / A. Botelho, L.C. Pinto // *Telecommunications Policy.* – 2004. – vol.28 (5-6). – С.427-437.
326. Brauer, F. Mathematical models in population biology and epidemiology / F. Brauer, C. Castillo-Chavez. – New York: Springer Verlag, 2001. – 416 с.
327. Castells, M. A. Mobile communication and society: a global perspective / A project of the Annenberg Research Network on international communication / M. Castells, M. Fernandez-Ardevol, J.L. Qiu, A. Sey. – The MIT Press, 2007. – 331 с.
328. Centrone, F. Bass-type model for a dynamic market with logistic growth / F. Centrone, E. Salinelli // *Aspects of mathematical modeling* ed. By R.J.Hosking and E.Venturino. *Mathematics and Biosciences in Interaction.* – 2008. – . – С.343-362.
329. Centrone, F. Demographic processes in a model of innovation diffusion with a dynamic market / F. Centrone, F. Goia, E. Salinelli // *Technological Forecasting and Social Change.* – 2007. – 74. – С.247-266.
330. Chandon, J.-L. J. comparative study of media exposure models. Ph.D. dissertation / J.-L. J. Chandon. – 1976.
331. Chatfield, C. A Consumer Purchasing Model with Erlang Inter-Purchase Times / C.Chatfield, G.J. Goodhardt // *Journal of the American Statistical Association.* – 1973. – vol.68, No.344.
332. Chatfield, C. Progress on a Simplified Model of Stationary Purchasing Behavior / C. Chatfield, A.S.C. Ehrenberg, G.J. Goodhardt // *J.R. Statist. Soc. Series A (General).* – 1966. – Vol.129, No.3. – С.317-367.

333. Chow, S.G. Technological Change and the Demand for Computers / S.G. Chow // *The American Economic Review*. – 1967. – Vol. 57, No. 5. – C.1117-1130.
334. Christensen, C.M. Seeing what's next?: using the theories of innovation to predict industry change / C.M. Christensen, C.D. Anthony, E.A. Roth. – Boston, MA: Harvard Business School Publishing, 2004. – 312 c.
335. Chu, W. Diffusion of mobile telephony: An empirical study in Taiwan / W. Chu, F. Wu, K. Kao, D. Yen // *Telecommunications Policy*. – 2009. – vol.33. – C.506-520.
336. Cohen, M.A. Operation planning in the presence of innovation diffusion dynamics / M.A. Cohen, T.H. Ho, H. Matsuo // *International series in quantitative marketing*. – 2000. – vol. 11. – C.237-259.
337. Coleman, J. The diffusion of an Innovation Among Physicians / J. Coleman, E. Katz, H. Menzel // *Sociometry*. – 1957. – Vol.20, No.4. – C.253-270.
338. Cortes, J.-C. Age-structured mathematical modeling approach to short-term diffusion of electronic commerce in Spain / J.-C. Cortes, I.-C.Lombana, R.-J. Villanueva // *Mathematical and computer modeling*. – 2010. – Vol.52, Issues 7-8. – P.1045-1051.
339. Cox, D.R. The Theory of Stochastic Processes / D.R. Cox, H.D. Miller. – Chapman and Hall/CRC, 1977.
340. Danaher, P.J. Three-Parameter Generalization of the Beta-Binomial Distribution with Applications. FSU Technical report M760, AFOSR Technical Report 87-208 / P.J. Danaher. – 1987. - 19 p.
341. Danaher, P.J. Optimal Internet Media Selection / P.J. Danaher, J. Lee, L. Kerbache // *Marketing Science*. – 2010. – Vol.29, No.2, March-April. – P.336-347.
342. Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology / F.D. Davis // *MIS Quarterly*. – 1989. – 13(3).
343. De Roos, A.M. Demographic analysis of continuous-time life-history models / A.M. de Roos // *Ecol. Lett.*. – 2008. – 11(1). – P.1-15.
344. Debecker, A. Determination of the uncertainties in S-curve logistic fits / A. Debecker, T. Modis // *Technol. Forecast. Soc. Change*. – 1994. – 46. – P.153-173.
345. Delitsin, L.L. Dynamics of Internet Diffusion in Russia and some European Countries: a comparative analysis / L.L. Delitsin // *World of Media. Yearbook of Russian Media and Journalism Studies*.. – 2011. – №2. – P.273-293.
346. Diekmann, O. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation / O. Diekmann, J.A.P. Hesterbeek. – Wiley, 2000. - 320 p.
347. Dijkstra, D. A continued fraction expansion for a Generalization of Dawson's Integral / D. Dijkstra // *Mathematics of Computation*. – 1977. – Vol.31, No. 138. – P.503-510.
348. Dixon, R. Hybrid corn revisited / R. Dixon // *Econometrica*. – 1980. – Vol. 48, No.6. – P.1451-1461.
349. Dodd, S.K. Formulas for Spreading Opinions / S.K. Dodd // *The Public Opinion Quarterly*. – 1958. – Vol. XXII. - No.4. – P.537-554.
350. Dodson, J.A. Models for new product diffusion through advertising and word of mouth / J.A. Dodson, E.Muller // *Management Science*. – 1978. – №24. – P.1568-1578.
351. Durlauf, S.N. Social Dynamics: Economic Learning and Social Evolution / S.N. Durlauf, H. Peyton Young, eds.. – London: The MIT Press, 2001. - 238 p.
352. Dutton, W.H. The Internet in Britain: 2009 / W.H. Dutton, E.J. Helsper, M.M. Gerber. – Oxford Internet Institute, University of Oxford., 2009.

353. Easingwood, C.J. A nonsymmetric responding logistic model for technological substitution / C.J. Easingwood, E.Muller, V.Mahajan // *Technological Forecasting and Social Change*. – 1981. – vol. 20. – P.199-213.
356. Easingwood, C.J. A nonuniform influence innovation diffusion model of new product acceptance / C.J. Easingwood, E.Muller V.Mahajan // *Marketing Science*. – 1983. – vol. 2. – P.273-295.
357. Economidies, N. Critical mass and network size with application to the US FAX market [Электронный ресурс] / N. Economidies, C. Himmelberg. – Электрон. текстовые дан.
358. Efron, B. An Introduction to the Bootstrap / B. Efron, R.J. Tibshirani. – New York, NY: Chapman and Hall, 1993.
359. Ehrenberg, A. The pattern of consumer purchases / A. Ehrenberg // *Applied Statistics*. – 1959. – 8. – P.26-41.
360. Ehrenberg, A. Repeat-buying. Facts, Theory and Applications / A. Ehrenberg. – New York: Oxford University Press, 1988.
361. Elderkin, R.H. Seed dispersal in a patchy environment with global age-dependence / R.H. Elderkin // *J. Math. Biol.* – 1982. – 13. – P.283-303.
362. Fader, P.S. "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model / P.S. Fader, B.G.S. Hardie, K.L. Lee // *Marketing Science*. – 2008. – 24 (Spring). – P.275-284.
363. Findahl, O. The Internet in Sweden / O. Findahl. – World Internet Institute, 2008.
364. Fisher, J. A simple substitution model of technological change / J. Fisher, R. Pry // *Technological Forecasting and Social Change*. – 1971. – vol. 3. – P.75-88.
365. Fornerino, M. Internet adoption in France / M. Fornerino // *The Service Industries Journal*. – 2003. – Vol. 23, Issue 1. – P.119-135.
366. Frank, L. An analysis of the effect of the economic situation on modeling and forecasting the diffusion of wireless communications in Finland / L. Frank // *Technological Forecasting and Social Change*. – 2004. – Vol.71, Issue 4, May 2004. – P.391-403.
367. Giovanis, A.N. A new modeling approach investigating the diffusion speed of mobile telecommunication services in EU-15 / A.N. Giovanis, C.H. Skiadas // *Computational economics*. – 2007. – vol.29, issue 2. – P.96-106.
368. Goldenberg, J. Talk of the network: a complex systems look at the underlying process of word-of-mouth / J. Goldenberg, B. Libai, E. Muller // *Marketing Letters*. – 2001. – 12(3). – P.209-221.
369. Greenwood, M. An Enquiry into the Nature of the Frequency Distributions Representative of Multiple Happenings, with Particular Reference to the Occurrence of Multiple Attacks of Disease or Repeated Accidents / M. Greenwood, G.U. Yule // *J.R. Statist. Soc.A.* – . – 83. – P.255-279.
370. Griliches, Z. Hybrid Corn: An Exploration in the Economics of Technological Change / Z. Griliches // *Econometrica*. – 1957. – vol. 25. – P.501-522.
371. Griliches, Z. Hybrid corn revisited. A reply / Z. Griliches // *Econometrica*. – 1980. – Vol.48, No.6. – P.1463-1465.

372. Griliches, Z. Hedonic price indexes for personal computers: Intertemporal and interspatial comparisons / Students, Z.Griliches, D.S. Hammermesh // *Economic letters*. – 1994. – Vol.44, Issue 4. – P.353-357.
373. Gruber, H. The economics of mobile telecommunications / H.Gruber. – New York: Cambridge University Press, 2005. - 323 p.
374. Gruber, H. The diffusion of Mobile Telecommunications Services in the European Union / H. Gruber, F. Verboven // *European Economic Review*, Elsevier. – 2001. – vol.45(3). – P.577-588.
375. Grübler, A. The Rise and Fall of Infrastructures / A. Grübler. – New-York: Springer-Verlag, 1990. - 305 p.
376. Grübler, A. Time for a change: On the patterns of diffusion of innovation / A. Grübler // *Technological Trajectories and the Human Environment* – Washington, DC: National Academies Press, 1997.
377. Grübler, A. Technology and Global Change / A. Grübler. – Cambridge University Press, 1998.
378. Gurtin, M.E. Nonlinear age-dependent population dynamics / M.E. Gurtin, R.C. MacCamy // *Arch. Rat. Mech. Anal.*. – 1974. – 54. – P.281-300.
379. Gurtin, M.E. Some simple models for nonlinear age-dependent population dynamics / M.E. Gurtin, R.C. MacCamy // *Mathematical Biosciences*. – 1979. – Vol.43, Iss.3-4. – P.199-211.
380. Helbing, D. Quantitative Sociodynamics: Stochastic Methods and Models of Social Interaction Processes / D. Helbing. – Springer, 2nd ed., 2010. - 350 p.
381. Hethcote, H.W. An age-structured model for pertussis transmission / H.W. Hethcote // *Math. Biosci.*. – 1997. – 145. – P.89-136.
382. Hoppensteadt, F.C. Mathematical theories of populations: demographics, genetics and epidemics / F.C. Hoppensteadt. – SIAM, 1975. - 72 p.
383. Huang, C.-Y. Global Digital Divide: A Dynamic Analysis based on the Bass Model / C.Y. Huang, H.N. Chen // *J. of Public Policy and Marketing* 2010. – 2010. – Vol.29, No.2. – P.248-264.
384. Ianelli, M. Gender Structured Population Modeling: Mathematical Methods, Numerics, and Simulations / M. Ianelli, M. Marche, F.A. Milner. – Philadelphia: SIAM, 2005. - 175 p.
385. IT-research, IT Research: российский рынок десктопов "просел" на 30% [Электронный ресурс] / . – Электрон. текстовые дан. – Режим доступа: <http://www.itoday.ru/news/11109.html>. – Загл. с экрана.
386. IT-research, IT-research подводит итоги 2009 г. и дает прогноз на 2010 г. [Электронный ресурс] / . – Электрон. текстовые дан. – Режим доступа: <http://www.itresearch.ru/press-center/news/detail.php?ID=13310>. – Загл. с экрана.
387. Jain, D. Effect of price on the demand for durables / D. Jain, R.C. Rao // *Journal of business and economic statistics*. – 1990. – vol. 8. – P.163-170.
388. Jeuland, A. heterogeneity into parsimonious models of diffusion of innovation. Working paper. / A. Jeuland. – Chicago : University of Chicago, Center for Research in Management, 1981.
389. Jupiter Communications, Accurate web site visitor measurement crippled by cookie blocking and deletion, Jupiter research finds [Электронный ресурс] / Jupiter Communications. – Электрон. текстовые дан. – Режим доступа:

- <http://www.jupitermedia.com/corporate/releases/05.03.14-newjupresearch.html>. – Загл. с экрана.
390. Kalish, S. A new-product Adoption Model with Price, Advertising and Uncertainty / S. Kalish // *Management Science*. – 1985. – 31 (December). – P.1569-1585.
 391. Kamakura, W.A. Long-term forecasting with innovation diffusion models: the impact of replacement purchases / W.A. Kamakura, S.K. Balasubramanian // *Journal of forecasting*. – 1987. – vol. 6. – P.1-19.
 392. Kamakura, W.A. Long-term forecasting with innovation diffusion models: the impact of replacement purchases / W.A. Kamakura, S.K. Balasubramanian // *Journal of forecasting*. – 1987. – vol. 6. – P.1-19.
 393. Kermack, W.O. Contributions to the mathematical theory of epidemics / W.O. Kermack, A.G. McKendrick // *of the Royal Society of London*. – 1927. – Series A, 115. – P.700-721.
 394. Keyfitz, N. Applied Mathematical Demography (Statistics for Biology and Health) / N. Keyfitz, H. Caswell. – Springer, 3rd edition, 2005. - 558 p.
 395. Kiiski, S. Cross-country diffusion of the Internet / S. Kiiski, M. Pohjola // *Information Economics and Policy*. – 2002. – Vol.14, Issue 2.. – P.297-310.
 396. Kremer, M. Population growth and technological change: one million B.C. to 1990. / M. Kremer // *The quarterly journal of economics*. – 1993. – vol. 108. – P.681-716.
 397. Krishnan, T.V. Impact of a late entrant on the diffusion of a new product/service / T.V. Krishnan, F.M. Bass, V. Kumar // *Management Science*. – 2000. – vol. 45. – P.1650-1663.
 398. Lasswell, H.D. The structure and function of communication in society / H.D. Lasswell // In Lyman Bryson (ed.) *The communication of ideas*. – 1948. – . – P.37-51.
 399. Lawrence, K.D. Application of Diffusion Models: Some Empirical Results / K.D. Lawrence, W.H. Lawton // *New Product Forecasting* / Y. Wind, V. Mahajan, and R.C. Cardozo, eds., – Lexington Books: Lexington, 1981. – P.529-541.
 400. Lee, J. Internet Media Planning: An Optimization Model / J. Lee, L. Kerbache // *KDI School of Pub Policy & Management*. 1. . Paper No. 05-02.
 401. Lekvall, P. A Study of Some Assumptions Underlying Information Diffusion Functions / P. Lekvall, C. Wahlbin // *The Swedish Journal of Economics*. – 1973. – Vol.75. - No.4. – P.362-377.
 402. Linz, P. Analytical and Numerical Methods for Volterra Equations / P. Linz. – Philadelphia, Pa.: SIAM Studies in Appl. Math., SIAM, 1985.
 403. Loomis, D.G. Forecasting the Internet. Understanding the Explosive Growth of Data Communications / D.G. Loomis, L.D. Taylor (Eds). – Springer, 2002. - 272 p.
 404. Lotka, A.J. Elements of Physical Biology / A.J. Lotka. – Baltimore MD: Williams and Wilkins, 1925.
 405. Macklin, B. The New European Internet Hot Spots. [Электронный ресурс] / B. Macklin, K. Von Abrams. – Электрон. текстовые дан. – Режим доступа: <http://www.emarketer.com/Article.aspx?R=1005935>. – Загл. с экрана.
 406. Mahajan, V. Innovation diffusion in a dynamic potential adopter population / V. Mahajan, R.A. Peterson // *Management Science*. – 1978. – No.24. – P.1589-1597.
 407. Mahajan, V. Innovation diffusion and new product growth models in marketing / V. Mahajan, E. Muller // *Journal of marketing*. – 1979. – vol. 43. – P.55-68.
 408. Mahajan, V. Models for Innovation Diffusion / V. Mahajan, R. Peterson. – Beverly Hills, CA: Sage, 1985.

409. Mahajan, V. Timing, diffusion and substitution of successive generations of technological innovations: The IBM mainframe case / V. Mahajan, E. Muller // *Technological Forecasting and Social Change*. – 1996. – vol. 51. – P.109-132.
410. Mahajan, V. Determination of adopter categories by using innovation diffusion models / V. Mahajan, E. Muller, R.K. Srivastava // *Journal of marketing research*. – 1990. – vol. 27 (1). – P.37-50.
411. Mahajan, V. New-product diffusion models in marketing: a review and directions for research / V. Mahajan, E. Muller, F.M. Bass // *Journal of marketing*. – 1990. – vol. 54. – P.1-26.
412. Mahajan, V. New-product diffusion models: from theory to practice / V. Mahajan, E. Muller, R. Wind, eds. // *International series in quantitative marketing*. – 2000. – vol. 11. – P.3-26.
413. Mansfield, E. Technical Change and the Rate of Imitation / E. Mansfield // *Econometrica*. – 1961. – vol. 29. – P.741-766.
414. Marchetti, C. Society as a Learning System / C. Marchetti // *Technological Forecasting and Social Change*. – 1980. – vol. 18. – P.267-282.
415. Marchetti, C. The Future / C. Marchetti // *Sinergetica ed Instabilita Dinamiche (Synergetics and Dynamical Instabilities)* / G.Caglioti, and H. Haken (eds) – Bologna, Italy: Societa Italiana di Fisica, 1988.
416. Marchetti, C. The dynamics of energy systems and the logistic substitution model / C. Marchetti, N. Nakicenovic // *IASA Research Report RR-79-13* – Laxenburg, Austria: 1979.
417. Marchetti, C. Human Population Dynamics Revisited with the Logistic Model: How Much Can Be Modeled and Predicted? / C. Marchetti, P.S. Meyer, J.H. Ausubel // *Technological Forecasting and Social Change*. – 1996. – vol. 52. – P.1-30.
418. Martino, J.P. *Technological Forecasting for Decision Making* / J.P. Martino. – New York: American Elsevier, 1975.
419. Martino, J.P. *R&D project selection* / J.P. Martino. – New York: J. Wiley & Sons, 1995.
420. McKendrick, A.G. Applications of mathematics to medical problems / A.G. McKendrick // *Proc. Edin. Math. Soc.* – 1926. – vol. 14. – P.98-130.
421. Meade, N. Technological forecasting – Model selection, model stability, and combining models / N. Meade, T. Islam // *Management Science*. – 1998. – Vol.44, No.8. – P.1115-1130.
422. Meade, N. Modelling and forecasting the diffusion of innovation - a 25-year review / N. Meade, T. Islam // *International journal of forecasting*. – 2006. – Vol. 22. – P.519-545.
423. Messina, R. Riccati equation-based generalization of Dawson's integral function / R. Messina, M.A. Jivulescu, A. Messina, A. Napoli // *Mathematical Methods in the Applied Science*. – 2007. – 30. – P.2055-2064.
424. Metheringham, R.A. Measuring the Net Cumulative Coverage of a Print Campaign / R.A. Metheringham // *Journal of Advertising Research*. – 1964. – 4. – P.23-28.
425. Meyer, P.S. Bi-Logistic Growth / P.S. Meyer // *Technological Forecasting and Social Change*. – 1994. – vol. 47. – P.89-102.
426. Michalakelis, C. Diffusion models of mobile telephony in Greece / C. Michalakelis, D. Varoutas, T. Spichopoulos // *Telecommunications Policy*. – 2008. – vol.32. – P.234-245.
427. Mickens, R.E. *Nonstandard finite difference models of differential equations* / R.E. Mickens. – World Scientific Publishing Company, 1994. - 249 p.
428. Midgley, D.F. A simple mathematical theory of innovative behavior / D.F. Midgley // *J.Consum. Res.* – 1976. – 3(1). – P.31-41.

429. Modis, T. The end of the Internet Rush / T. Modis // *Technological Forecasting and Social Change*. – 2005. – Vol.72, Issue 8. – P.938-943.
430. Moldovan, S. Cellular automata modeling of resistance to innovations: Effects and solutions / S. Moldovan, J.Goldenberg // *Technological Forecasting & Social Change*. – 2004. – 71. – P.425-442.
431. Morrison, D.G. Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort? / D.G. Morrison, D.C. Schmittlein // *Journal of business and economic statistics*. – 1988. – 6 (April). – P.145-159.
432. Moshier, S.L. *Methods and Programs for Mathematical Functions* / S.L. Moshier. – Chichester, England: Ellis Horwood, 1989.
433. Nakicenovic, N. Software package for the logistic substitution model / N. Nakicenovic // *IIASA Research Report RR-79-12* – Laxenburg, Austria: 1979.
434. Nakicenovic, N. *Growth to Limits: Long Waves and the Dynamics of Technology* / N. Nakicenovic. – Vienna, Austria: 1984.
435. Nakicenovic, N. *Diffusion of Technologies and Social Behaviour* / N. Nakicenovic. – Berlin, Germany: Springer-Verlag, 1991.
436. Nielsen, J. Bass curves for the diffusion of innovations [Электронный ресурс] / J. Nielsen. – Электрон. текстовые дан. – Режим доступа: <http://www.useit.com/alertbox/basscurves.html>. – Загл. с экрана.
437. Olson, J. A product diffusion model incorporating repeat purchases / J. Olson, S. Choi // *Technological forecasting and social change*. – 1985. – vol. 27. – P.385-3987.
438. Ozga, S.A. Imperfect markets through lack of knowledge / S.A. Ozga // *The Quarterly Journal of Economics*. – 1960. – Vol.74. - No.1. – P.29-52.
439. Park, Y.-H. Modeling browsing behavior at multiple sites / Y.-H.Park, P.S. Fader // *Marketing Science*. – 2004. – Vol.23, No.3. – P.280-303.
440. Parker, P. Aggregate diffusion forecasting models: a critical review / P. Parker // *International journal of forecasting*. – 1994. – vol. 10. – P.353-380.
441. Pearl, R On the rate of growth of the population of the United States since 1790 and its mathematical representation / R. Pearl, L.J. Reed // *Proceedings of the National Academy of Sciences*. – 1920. – vol. 6. – P.275-288.
442. Perez, R. Innovation diffusion and new product growth models: a critical review and research directions / R. Perez, E. Muller, V. Mahajan // *Intern. J. of Research in Marketing* . – 2010. – vol.27. – P.91-106.
443. Pessemier, E.A. *Product Management, Strategy and Organization* / E.A. Pessemier. – New York: John Wiley and sons, 1977.
444. Peterka, V. The dynamics of energy systems and the logistic substitution model. Volume 2: Theoretical part / V. Peterka, F. Fleck // *Administrative Report AR-78-1C*, International Institute for Applied Systems Analysis, 1978.
445. Press, W.H. *Numerical Recipes in Fortran: The Art of Scientific Computing* / W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. – Cambridge, UK: Cambridge University Press, 1992.
446. Putsis, W.P. Estimation techniques for macro diffusion models / W.P. Putsis Jr., S. Srinivasan // *International series in quantitative marketing*. – 2000. – vol. 11. – P.263-291.

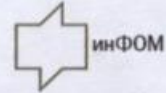
447. Ratchford, B.T. Diffusion models with replacement and multiple purchases / B.T. Ratchford, S.K. Balasubramanian, W.A. Kamakura // International series in quantitative marketing. – 2000. – vol. 11. – P.123-140.
448. Roberts, J.H. Disaggregated level diffusion models / J.H. Roberts, J.M. Lattin // International series in quantitative marketing. – 2000. – vol. 11. – P.207-236.
449. Rogers, E.M. Diffusion of Innovations, 5th Edition / E.M. Rogers. – Free Press, 2003. - 576p.
450. Rose, R. / R. Rose. – Oxford: Oxford Internet Institute Research Report. – 2006.
451. Rossiter, J.R. Advanced Media Planning / J.R. Rossiter, P.J. Danaher. – Kluwer, 1998. - 112p.
452. Rust, R. Advertising media models. A practical guide / R. Rust. – 1986. - 159 p.
453. Sabavala, D.J. A Nonstationary Model of Binary Choice Applied to Media Exposure / D.J. Sabavala, D.G. Morrison // Management science. – 1981. – 27 (June). – P.637-657.
454. Schmittlein, D.C. Counting Your Customers: Who They Are and What They Will Do Next? / D.C. Schmittlein, D.G. Morrison, R. Colombo // Management science. – 1987. – 33. – P.1-24.
455. Sciadas, C.H. Two generalized models for forecasting innovation diffusion / C.H. Skiadas // Technological Forecasting and Social Change. – 1985. – vol.27. – P.39-61.
456. Sciadas, C.H. Innovation diffusion models expressing asymmetry and/or positively or negatively influencing forces / C.H. Skiadas // Technological Forecasting and Social Change. – 1986. – vol.30. – P.313-330.
457. Sharif, M.N. A generalized model for forecasting technological substitution / M.N. Sharif, C. Kabir // Technological Forecasting and Social Change. – 1976. – vol. 8. – P.353-364.
458. Sharif, M.N. The Weibull distribution as a general model for technological forecasting / M.N. Sharif, Islam M.N. // Technological Forecasting and Social Change. – 1980. – vol. 18. – P.247-256.
459. Sharif, M.N. Polynomial innovation diffusion models / M.N. Sharif, K.Ramanathan // Technological Forecasting and Social Change. – 1981. – vol. 21. – P.301-323.
460. Skellam, J.G.) A probability distribution derived from the binomial distribution by regarding the probability of success as a variable between sets of trials. J.R.Statist.Soc. B / J.G. Skellam // . – 1948. – 10. – P.257-265.
461. Steffens, P. A mathematical model for new product diffusion: the influence of innovators and imitators / P. Steffens, D. Murthy // Mathematical and computer modeling. – 1992. – Vol.16, No.4. – P.11-26.
462. Steffens, P.R. An aggregate sales model for consumer durables incorporating a time varying mean replacement age / P.R. Steffens // Journal of forecasting. – 1991. – vol. 20. – P.63-77.
463. Stoneman, P. The economics of technological diffusion / P. Stoneman. – Wiley-Blackwell, 2002. - 328 p.
464. Tarde, G. The laws of imitation / G. Tarde. – New York: Holt, Rinehart and Winston Inc., 1903.
465. Teng, J.T.C. Information technology innovations: general diffusion patterns and its relationship to innovation characteristics / J.T.C. Teng, V. Grover, W. Guttler // IEEE transactions on engineering management. – 2002. – vol. 49. – No. 1. – P.13-27.
466. Thieme, H.R. Mathematics in Population Biology (Princeton Series in Theoretical and Computational Biology) / H.R. Thieme. – NY: Princeton University Press, 2003. - 568 p.

467. Thirtle, C.G. The role of demand and supply in the generation and diffusion of technical change / C.G. Thirtle, V.W. Ruttan. – Harwood academic publishers, 1987, 1994. - 172 p.
468. Tsoularis, A. Analysis of logistic growth models / A. Tsoularis // Res. Lett. Inf. Math. Sci.. – 2001. – Vol.2. – P.23-46.
469. Valente, T. Network models of the diffusion of innovations / T. Valente. – Hampton press, 1995, 1999. - 171 p.
470. Van den Bulte, C. Social contagion and income heterogeneity in new product diffusion: a meta-analytic test / C. Van den Bulte, S. Stremersch // Management science. – 2004. – vol.23. – No.4. – P.530-544.
471. Van den Bulte, C. New product diffusion with Innovators and Imitators / C. Van den Bulte, Y.V. Joshin // Marketing Science. – 2007. – №3. – P.400-421.
472. Vanston, L.K. / L.K. Vanston, J.H. Vanston. – Technology Futures Inc., 1996.
473. Vanston, L.K. Technology forecasting for telecommunications / L.K. Vanston, R.L. Hodges // Teletronikk. – 2004. – N.4. – P.32-42.
474. Varian, H.R. Economics of information technology [Электронный ресурс] / H.R. Varian. – Электрон. текстовые дан. – Режим доступа: <http://www.sims.berkeley.edu/~hal/Papers/mattioli/mattioli.html>. – Загл. с экрана.
475. Venkatesh, V. User Acceptance of Information Technology: Toward a Unified View / V.Venkatesh, M.G.Morris, G.B.Davis, F.D.Davis // 27(3). – 2003. – MIS Quarterly. – P.425-478.
476. Verhulst, P.-F. Notice sur la loi que la population suit dans son accroissement / P.-F. Verhulst // Correspondance Mathematique et Phusique, publiee par A. Quetelet . – 1838. – vol. 10. – P.113.
477. Verhulst, P.-F. Recherches Mathematiques sur la loi d'accroissement de la population / P.-F. Verhulst // Nouveaux Memoires de l'Academie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique . – 1845. – vol. 18. – P.1-32.
478. Vishvanath, A. / A. Vishvanath, G.A. Barnett, eds. // The Diffusion of Innovations: A Communication Science Perspective / edited by A.Vishwanath, G.A.Barnett / Peter Lang Publishing Inc. – 2011. – 239 p.
479. Von Foerster, H. Some Remarks on Changing Populations / H. Von Foerster // The Kinetics of Cellular Proliferation / F. Stohlman Jr. (Hg.) – New York: Grune and Stratton, 1959. – P.382–407.
480. Von Foerster, H. Friday, November 13, AD 2026 / H. Von Foerster // Science. – 1960. – vol. 132. – P.1291–1295.
481. Wright, M. A validation of the Bass new product diffusion model in New Zealand / M. Wright // Marketing Bulletin. – 1997. – vol. 8. – P.15-29.
482. Yule, G.U. On the distribution of deaths with age when the causes of death act cumulatively, and similar frequency distributions / G.U. Yule // J.R. Statist. Soc.. – 1910. – 73. – P.26-38.

ПРИЛОЖЕНИЕ А.

ООО «инФОМ»

123242, Россия, Москва, пер. Капринова, 3
телефон : (+7 495) 745-87-65
факс : (+7 495) 745-89-03
www.fom.ru fom@fom.ru



01.12.2013

РАЗРЕШЕНИЕ

на использование данных исследований
Общества с ограниченной ответственностью
«Институт фонда «Общественное мнение»

Настоящим Общество с ограниченной ответственностью «Институт фонда «Общественное мнение» (далее – ООО «инФОМ») дает разрешение Делицыну Леониду Леонидовичу на использование данных исследований ООО «инФОМ» по следующим темам:

- проникновение Интернета в России и Москве по возрастным группам до лета 2013 года включительно;
- полугодовая, месячная, недельная и суточная аудитория Интернета в России и Москве до лета 2013 года включительно.

Указанные данные разрешается использовать для апробации моделей, изложенных в диссертационной работе Делицына Л.Л. «Разработка и применение количественных моделей распространения новых информационных технологий», представленной на соискание ученой степени доктора технических наук.

Также сообщаем, что разработанная Делицыным Л.Л. многостадийная модель динамики интенсивности использования сети Интернет и модель распространения сети Интернет в населённых пунктах различного типа, учитывающая демографические процессы и изменение стоимости доступа во времени, может применяться в деятельности ООО «инФОМ» по проведению социологических исследований и наблюдений.

Настоящий документ направляется в диссертационный совет Д 210. 010. 01 при ФГОУ «Московский государственный университет культуры и искусств».

Генеральный директор
ООО «инФОМ»

А.А. Ослон

ПРИЛОЖЕНИЕ Б.

articulproduction

УТВЕРЖДАЮ

Генеральный директор
Куликова О.Ю.

Дата 10.05.2011



А К Т

о внедрении результатов
докторской диссертационной работы

Комиссия в составе:

председатель – Куликова О.Ю., генеральный директор ООО «Артикул Продакшн»,
члены комиссии:

Иванов П.А. – технический директор ООО «Артикул Продакшн»,

Лисняк М.Я. – директор по НИОКР ООО «Артикул Продакшн»,

составили настоящий акт о том, что результаты диссертационной работы ДЕЛИЦЫНА Леонида Леонидовича «Разработка и применение количественных моделей распространения новых информационных технологий», представленной на соискание ученой степени доктора технических наук, а именно, разработанные автором:

- модель распространения Интернета в регионе, учитывающая процесс воспроизводства населения,
- модель интенсивности использования информационного ресурса в виде смеси бета-биномиальных распределений

внедрены в практику деятельности ООО «Артикул Продакшн».

Использование указанных результатов позволило ЗАО «Артикул Продакшн» рассчитать прогноз суточного охвата аудитории и количество посещений официального сайта Олимпийских Игр 2014 («Сочи 2014») пользователями Интернета из России, стран СНГ, Европы, США и Канады, и из Китая. Построенный прогноз позволил предоставить руководству официального сайта «Сочи 2014» данные, необходимые для расчета объема входящего и исходящего трафика, и рассчитать необходимые затраты на аренду каналов доступа. Экономический эффект от применения разработанных Л.Л. Делицыным моделей оценивается нами в 6 миллионов рублей.

Акт направляется в диссертационный совет Д 210. 010. 01 при ФГОУ «Московский государственный университет культуры и искусств».

Председатель комиссии

Члены комиссии:

Члены комиссии:

О.Ю. Куликова

М.Я. Лисняк

П.А. Иванов



В диссертационный совет Д 210. 010. 01
при ФГОУ «Московский государственный
университет культуры и искусств»

Сопроводительное письмо

Настоящим сопроводительным письмом подтверждаю, что при составлении прогнозов аудитории и количества просмотренных страниц официального сайта Олимпийских Игр 2014 года в Сочи и его разделов, выполненных ООО «Артикул Продакшн», были применены результаты диссертационной работы Делицына Леонида Леонидовича «Разработка и применение количественных моделей распространения новых информационных технологий», представленной на соискание ученой степени доктора технических наук.

Олимпийские игры традиционно вызывают чрезвычайно высокий интерес пользователей Интернета во многих странах мира, что потребует от сайта Сочи 2014 устойчивости к экстремально высоким нагрузкам. Поэтому для решения задачи обеспечения бесперебойной работы этого информационного ресурса необходимы надежные прогнозы количества его пользователей за месяц и в среднем за сутки. Построение таких оценок осложняется быстрым ростом аудитории Интернета, поскольку за период между последовательными зимними Играми аудитория Интернета более чем удваивается. Следовательно, при расчетах нельзя опираться только на статистику сайта предыдущих Игр, и требуется достоверный прогноз аудитории Интернета в ряде регионов мира и в целом.

Развитие широкополосного доступа к Всемирной Сети позволяет использовать мультимедийные материалы и просматривать больше страниц сайта, чем это удавалось посетителям предыдущих Игр, поэтому необходимы также прогнозы количества просмотренных страниц основных разделов сайта Сочи 2014. Главными показателями, необходимыми для эффективного управления ресурсом, являются среднесуточное количество пользователей и среднесуточное количество вызовов страниц сайта, которые позволяют оценить размер необходимого серверного парка и объем трафика.

При расчете необходимых показателей ООО «Артикул Продакшн» применило модели динамики аудитории Интернета и интенсивности использования информационных ресурсов, разработанные Л.Л. Делицыным в ходе диссертационного исследования.

Результаты диссертационной работы Л.Л. Делицына представляются нам новыми, научно обоснованными и практически значимыми. Разработанные Л.Л. Делицыным модели применимы для повышения эффективности систем управления информационными ресурсами и улучшения на этой основе качества и эффективности решений, принимаемых в управленческой деятельности.

Руководитель управления интернет-проектов
АНО «Оргкомитет «Сочи 2014»

М.Н.Чеканов

Подпись М.Н.Чеканова заверяю

Старший менеджер отдела
кадрового администрирования



Н.С.Ванина

ООО «Юду»

ООО «Юду»
г. Москва,
Фрунзенская наб.16,
корпус 1, помещение IV
тел. (499) 271-0553

УТВЕРЖДАЮ

Генеральный директор

Лоскутов А.В.

Дата 03.03.2014

А К Т

о внедрении результатов
докторской диссертационной работы



Комиссия в составе:

председатель – Лоскутов А.В., генеральный директор ООО «Юду»:

Гидирим А.А – заместитель генерального директора

Кутергин Д.С. – исполнительный директор

Кордубайло А.А. – технический директор

составили настоящий акт о том, что результаты диссертационной работы Делицына Леонида Леонидовича «Научные основы разработки и применение количественных моделей распространения новых информационных технологий», представленной на соискание ученой степени доктора технических наук по специальности 05.25.05, внедрены в ООО «Юду».

Разработанные Л.Л. Делицыным модели для интерпретации данных онлайн-панели в виде смеси бета-биномиальных распределений и для частоты использования информационного ресурса в виде смеси Парето-NBD распределений применены для прогнозирования количества заказов по отдельным когортам, которые используются в деятельности компании.

Акт направляется в диссертационный совет Д 210.010.01 при ФГОУ «Московский государственный университет культуры и искусств».

Председатель комиссии

/А.В. Лоскутов

Члены комиссии:

/ А.А. Гидирим

Члены комиссии:

/ Д.С. Кутергин

Члены комиссии:

/А.А. Кордубайло



ЗАО «Инвестиционный холдинг «ФИНАМ»
109240, г. Москва,
ул. Верхняя Радищевская, д.4, стр.3-4-5
Телефон (495) 796-93-88
Факс (495) 796-93-89



УТВЕРЖДАЮ

Президент – Председатель
Правления ЗАО
«Инвестиционный холдинг
«ФИНАМ»
В.В. Кочетков

Дата 24.12.2013
Гербовая печать

А К Т

о внедрении результатов
докторской диссертационной работы

Комиссия в составе:

председатель – Кочетков В.В., Президент ЗАО «Инвестиционный холдинг «ФИНАМ»

Колоусов Д.В. – генеральный директор ООО «Инвестиционная компания «ФИНАМ», к.ф.-м.н.,

Исаев В.Ю. – руководитель отдела ЗАО «ФИНАМ», к.и.н.

составили настоящий акт о том, что результаты диссертационной работы Делицына Леонида Леонидовича «Разработка и применение количественных моделей распространения новых информационных технологий», представленной на соискание ученой степени доктора технических наук, внедрены в практику деятельности ЗАО «Инвестиционный холдинг «ФИНАМ».

Разработанная Л.Л. Делицыным модель выбора нововведений в сфере информационных и телекоммуникационных технологий венчурным фондом использовалась ЗАО «Инвестиционный холдинг «ФИНАМ» в 2007-2013 гг. С применением этой модели было обработано более полутора тысяч инвестиционных заявок. В результате прошёл процедуру тщательного исследования и получил инвестиции коллектив предпринимателей, основавший три инновационных компании, которые в настоящее время входят в инвестиционный портфель FINAM Global. Экономический эффект от применения разработанных Л.Л. Делицыным моделей оценивается нами в 18 миллионов рублей.

Акт направляется в диссертационный совет Д 210. 010. 01 при ФГОУ «Московский государственный университет культуры и искусств».

Председатель комиссии

Члены комиссии:

Члены комиссии:

подпись /В.В. Кочетков

подпись /Д.В. Колоусов

подпись /В.Ю. Исаев



**МОСКОВСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА
(МГУ)**

ФАКУЛЬТЕТ ЖУРНАЛИСТИКИ

125009 Москва, ул. Моховая, д. 9
Телефон: 629-74-35, Факс: 629-49-54

СПРАВКА

о внедрении в учебном процессе результатов
докторской диссертационной работы Л.Л. Делицына
«Разработка и применение количественных моделей
распространения новых информационных технологий»

Результаты диссертационной работы Делицына Леонида Леонидовича «Разработка и применение количественных моделей распространения новых информационных технологий», представленной на соискание ученой степени доктора технических наук, внедрены в учебном процессе факультета журналистики МГУ им. М.В. Ломоносова.

Разработанные Делицыным Л.Л. модели кумулятивного охвата аудитории онлайн-СМИ и поисковых систем Интернета используются в учебных курсах

1. «Методы компьютерного анализа данных в гуманитарных науках» магистратуры первого года обучения;
2. «Организация, хранение и поиск информации в Интернете» вариативной части бакалавриата (Профессиональный модуль-1 «Интернет-журналистика»);
3. «Информационные ресурсы и поисковые системы Интернета» программы профессиональной переподготовки «Интернет-журналистика»

Разработанные Делицыным Л.Л. прогнозы распространения Интернета в России и в других странах составили основу лекции «Прогноз развития интернета в России» зимней научной школы «Телевидение и интернет в современном медиапространстве: взаимодействие и взаимовлияние» для аспирантов и молодых преподавателей.

Справка направляется в диссертационный совет Д 21001001 при ФГОУ ВПО «Московский государственный университет культуры и искусств».

Декан
факультета журналистики
профессор

Заместитель декана по УМО
доцент



Е.Л.Вартанова

М.М. Лукина