

Федеральное государственное образовательное бюджетное учреждение
высшего образования
«Финансовый университет при Правительстве Российской Федерации»

На правах рукописи

Кораблев Юрий Александрович

ЕМКОСТНЫЙ МЕТОД АНАЛИЗА РЕДКИХ СОБЫТИЙ В ЭКОНОМИКЕ

5.2.2. Математические, статистические и инструментальные методы
в экономике

ДИССЕРТАЦИЯ
на соискание ученой степени
доктора экономических наук

Научный консультант

Бывшев Виктор Алексеевич,
доктор технических наук, профессор

Москва - 2023

Оглавление

Введение.....	4
Глава 1 Обзор существующих методов анализа редких событий	18
1.1 Обзор терминологии и задач, в которых используется термин «редкие события».....	18
1.2 Существующие методы анализа редких событий	28
1.3 Анализ недостатков существующих методов анализа редких событий	57
Глава 2 Ёмкостный метод анализа и прогнозирования редких событий	68
2.1 Формулирование методики для исследования редких событий.....	68
2.2 Примеры использования ёмкостного метода.....	80
2.3 Формулирование условий, предпосылок, допущений.....	115
Глава 3 Восстановление функции по выборке событий	129
3.1 Разработка оптимизационной задачи восстановления функции по интегралам	129
3.2 Выбор коэффициента сглаживания.....	146
3.3 Алгоритм восстановления функции по интегралам	149
3.4 Реализация на R.....	153
3.5 Пример применения алгоритма на языке R	156
3.6 Реализация в MS Excel	162
Глава 4 Оценка точности ёмкостного метода	171
4.1 Погрешность наблюдений, связанная с дискретизацией времени	171
4.2 Исследование вопросов точности метода в зависимости от удаленности от конечного потребителя с помощью моделирования .	179
4.3 Математическое обоснование падения точности при удалении от конечного потребителя	189
4.4 Изучение влияния конкуренции или потери части данных на точность восстановления исходной функции	214
4.5 Изучение влияния неопределенности внутри самого процесса, протекающего в источнике, на точность метода	225

4.6 Сравнение емкостного метода с другими методами.....	243
4.7 Методика проверки адекватности прогнозных моделей, построенных с помощью емкостного метода.....	253
Глава 5 Обобщение идеи емкостного метода	261
5.1 Использование емкостного метода для анализа исторических, социальных и политических процессов.....	261
5.2 Восстановление функции по разным функционалам для анализа и прогнозирования редких событий, вызванных произвольными процессами.....	270
5.3 Разработка универсального метода исследования произвольных процессов	282
5.4 Разработка метода исследования неизвестных событий	303
Заключение	308
Список литературы	312
Приложение А Алгоритм Куина и Фернандеса на языке R	326
Приложение Б Дополнительные примеры применения метода.....	329
Приложение В Расчет интеграла 1	333
Приложение Г Расчет интеграла 2	334
Приложение Д Расчет интеграла 3	335
Приложение Е Математическое обоснование падения точности от потери части данных	338

Введение

Актуальность темы исследования. Способность анализировать и предсказывать будущие события всегда имело и будет иметь большое значение для любой области деятельности. Издревле мореплаватели пытались предсказать погоду перед выходом в открытое море, а фермеры определить наилучшее время и вид высаживаемых культур, чтобы урожай не побил морозами и не спалило солнцем. В современной экономике это особенно актуально. Знание о грядущем событии позволит грамотно к нему подготовиться, тем самым извлечь определенную выгоду или уменьшить возможные убытки. На торгах информация о грядущем спаде или скачке цен позволяет некоторым счастливицам немислимо обогатиться. В логистике подготовить нужный объем запасов на складах или на полках магазинов. Трудно переоценить значимость методов, предназначенных для анализа и прогнозирования будущих событий.

Тем не менее хоть методы анализа и прогнозирования событий постоянно развиваются, в существующих методах имеются определенные недостатки при анализе редких событий. Редкие события отличаются низкой частотой или вероятностью появления, отличаются нерегулярностью и особой сложностью при прогнозировании. Подробнее о том, чем редкие события отличаются от частых событий будет сказано в первой главе Исследование существующих методов показало, что недостатки можно связать с тем, что практически все методы ограничиваются лишь статистическими моделями, в которых анализируется либо частота появления событий за интервал времени, либо закон распределения интервалов времени между появлением событий. Другие методы предполагают построение по собранной статистике классификационных моделей, которые по наблюдаемым сопутствующим признакам относят наблюдения к одной из возможных категорий. Недостатком таких методов является то, что, анализируя лишь статистику, методы никак не используют информацию о механизмах образования этих

событий. Это приводит к тому, что недостаток информации необходимо компенсировать большим объемом обучающей выборки, что недоступно при анализе редких событий. Другим же недостатком существующих применяемых в экономике статистических методов анализа редких событий является то, что все полученные оценки параметров моделей рассматриваются как статичные, а не динамичные, и в них не учитываются какие-либо возможные изменения этих оценок со временем. В итоге построенные модели прогнозирования редких событий могут быть неадекватны в новых изменяющихся условиях, которые могут наступить уже к очередному событию. Основываясь на вышесказанном, можно констатировать, что разработка новых методов анализа и прогнозирования редких событий в экономике является актуальной темой исследования.

В данной диссертационной работе разработан метод, краеугольным камнем которого является как раз информация о механизме (процессе) образования событий, а не только их статистика. Тем не менее статистические методы в данной работе никак не принижаются, они, наоборот, активно используются, но не для построения статистических моделей, а для определения параметров механизма образования событий. Во второй главе формулируется основная идея нового общего подхода анализа и прогнозирования редких событий. В той же главе в качестве частного случая рассматриваются процессы образования событий, схожие с процессами опустошения или наполнения некоторой емкости, которые определили название метода («ёмкостный метод»). Благодаря этому можно предсказывать будущие дискретные покупки или моменты приобретения услуг конкретных неподконтрольных клиентов. В третьей главе разрабатывается математический метод восстановления параметров механизма образования событий. Четвертая глава посвящена изучению вопросов точности. В пятой главе происходит возвращение от частного случая к общему и выполнено обобщение идей ёмкостного метода на произвольные процессы образования событий. Так как отправной точкой для анализа и прогнозирования событий

является информация о механизме образования событий, возможно кого-то разочарует, что так называемые «черные лебеди» Н. Талеба остаются за рамками рассмотрения. Однако если когда-то удастся построить математическую модель образования этих «черных лебедей», то описанная в данной диссертационной работе методика позволит прогнозировать и их.

Степень разработанности темы исследования. Существует большое количество различных и даже совсем не похожих друг на друга задач, в которых используется понятие «редкие события», и в каждой из этих задач редкие события понимают немного по-разному, теория используемых методов тоже значительно отличается. Редкие события изучаются в большом количестве областей. В данном диссертационном исследовании мы не будем затрагивать такие области как физика ядерных процессов, астрономия, геология и предсказание землетрясений, которые также оперируют понятием редких событий.

Вероятностное моделирование редких событий опирается на фундамент, который заложили Байес Т., Бернулли Д., Парето В., Пирсон К., Пуассон С.Д., Фишер Р., Госсет В., Марков А.А., Колмогоров А.Н. и др. К редким событиям иногда относят отказы технических устройств и информационных систем в теории надежности систем, развитие которой связано с работами Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д., Ушакова И.А., Кокса Д., Барлоу Р., Проскан Ф. и др. Вероятности отказов или степень надежности устройств или структур определяют методами статистических испытаний (Монте-Карло), методами надежности первого (First-order reliability method, FORM) [1] и второго порядка (SORM) [2], выборка по линиям (Line sampling) [3; 4] и их модификациями. Данные методы развиваются в работах Hasofer A.M., Lind N.C., Ang S.H., Thoft-Christensen P., Madsen H.O., Baker M.J., Kiureghian A., Beck J.L., Au S.K., Lin H.Z., Hwang S.J., Schueller G.I., Pradlwarter H.J., Angelis M., Patelli E., Wang Z., Broccardo M., Melchers R.E. и др.

Очень часто редкие события рассматривают с точки зрения потоков событий (точечных процессов), развитие которых осуществлено в работах Пуассона С.Д., Пальма К., Вейбулла В., Эргланга А., Феллера В., Кингмена Дж., Кокса Д., Башарина Г.П., Наумова В.А. и др. В экономике теория потоков событий находит очень широкое применение у российских исследователей: при моделировании собственных объемов страховых запасов - Истомина А.А., Ян Л.А., Замалетдинова Д.А.; в сельском хозяйстве - Белякова А.Ю., Петрова С.А.; надежности технических систем гражданской инфраструктуры - Постников И.В., Скопинцев В.А.; при оценке стоимости акций – Кожевников А.С.; при моделировании экологических катастроф – Борисов В.В.; при расчете рисков негативных событий – Саченко Л.А. и многие другие.

Другой часто используемый подход предполагает использование методов бинарной или множественной классификации, где по наблюдаемым сопутствующим признакам наблюдения относят к одному из классов, тем самым распознавая определенный тип событий. Для этого используются хорошо известные методы машинного обучения, такие как логистическая и пробит регрессия, метод ближайших соседей, метод опорных векторов, нейронные сети и другие методы, основы которых были заложены в работах Берксона Дж., Блисса Ч., Вапника В., Червоненкиса А., Босера Б., Бравермана Э., Cortes С., Cover Т., Singh S., Choe. W., McCullagh P., Nelder J. и др. Методы классификации имеют очень широкое применение в работах отечественных и зарубежных исследователей: для определения устойчивости (дефолта/банкротств) компаний или банков – Гусятников П.В., Могилат А.Н., Туктарова П.А., Заиченко Е.М., Рыгин В.Е., Биджоян Д.С., Kumar K., Tan С., Jardin P., Johnsen T. и многих других (где дефолты/банкротства определяются как редкие события); в информационной безопасности для определения атак или эксплойтов – Лифанов К.А, Зайцев К.С.; в прогнозировании ненулевого спроса – Пивкин К.С.; при определении потенциальных покупателей – Заказчикова Н.А.; и во многих других областях экономики.

Отдельно для задачи прогнозирования ненулевого спроса (прерывистого спроса, *intermittent demand forecast*), в которой также используют понятие редких событий, существуют специфические статистические методы прогнозирования, позволяющие рассчитать объемы собственных страховых запасов. В этом направлении работали исследователи Croston J.F., Syntetos A.A., Boylan J.E., Leven E., Segerstedt A., Vinh D.Q., Willemain T.R., Smart C. N., Hua Z. S., Kaya G.O., Pince C. и другие.

Как было сказано выше, во всех перечисленных методах никак не учитываются механизмы образования событий, в то же время все получающиеся модели прогнозирования оказываются статичными. Практически полное отсутствие исследований вопросов использования информации о механизме образования редких событий в методах анализа и прогнозирования будущих событий в экономике формирует актуальность и цель данного диссертационного исследования.

Цель исследования - разработка метода анализа и прогнозирования редких событий в экономике, базирующегося на рассмотрении событий с точки зрения процессов, протекающих внутри источников этих событий и приводящих к образованию этих событий.

Задачи исследования:

– Обзор существующих методов анализа и прогнозирования редких событий, применяемых в экономике, а также исследование и выявление причин их недостатков.

– Разработка общего подхода для анализа и прогнозирования редких событий в экономике с использованием информации о процессах образования этих событий.

– Разработка метода анализа и прогнозирования редких событий, процесс образования которых схож с процессами наполнения/опустошения некоторой емкости, и который может использоваться для моделирования процесса потребления или процесса накопления возмущений при образовании событий.

– Разработка математического метода для определения параметров процесса образования событий, схожего с процессом наполнения / опустошения емкости.

– Рассмотрение примеров и демонстрация применения разработанного емкостного метода для анализа и прогнозирования редких событий в экономике на основе как данных моделирования, так и реальных данных.

– Формулировка набора условий, предпосылок, допущений для того, чтобы разработанный емкостный метод анализа и прогнозирования редких событий можно было использовать на практике для прогнозирования экономических событий, а также описание возможных последствий нарушения этих предпосылок.

– Сравнение разработанного емкостного метода анализа и прогнозирования редких событий с другими существующими методами.

– Исследование точности разработанного емкостного метода, определение причины возникновения погрешностей, а также разработка методики проверки адекватности получающихся моделей прогнозирования редких событий.

– Создание инструментальных средств для анализа редких событий в экономике с помощью разработанного емкостного метода.

– Обобщение емкостного метода на другие виды событий, такие как социальные, политические и исторические события.

– Обобщение идей разработанного метода анализа и прогнозирования редких событий на процессы образования событий, которые используют значения первых и вторых производных, а также интегралов для образования событий, и разработка соответствующего математического метода восстановления параметров таких процессов.

– Обобщение идей разработанного метода анализа и прогнозирования редких событий на процессы, заданные произвольными алгоритмами, и разработка метода восстановления параметров процесса образования событий, заданного в виде алгоритмической модели. Предложение методики

исследования неизвестных событий с помощью подбора соответствующей алгоритмической модели процесса их образования.

Объект исследования – редкие события в экономике и методы их анализа и прогнозирования.

Предмет исследования – процессы, протекающие в источниках редких событий, способы моделирования и восстановления их параметров с целью анализа и дальнейшего прогнозирования формирующихся событий.

Область исследования. Диссертация подготовлена в соответствии с пунктами: 3. «Разработка и развитие математических и эконометрических моделей анализа экономических процессов (в т.ч. в исторической перспективе) и их прогнозирования», 4. «Разработка и развитие математических и компьютерных моделей и инструментов анализа и оптимизации процессов принятия решений в экономических системах» Паспорта научной специальности: 5.2.2. Математические, статистические и инструментальные методы в экономике (экономические науки).

Научная новизна исследования заключается в разработке метода анализа и прогнозирования редких событий в экономике. Метод базируется на использовании информации о процессах образования редких событий, связанных с наполнением/опустошением некоторых материальных или нематериальных ёмкостей. Параметры таких процессов восстанавливаются методом сплайновой коллокации со штрафом на нелинейность (шероховатость) по наблюдаемым с погрешностью функционалам. После экстраполяции на будущие периоды времени восстановленных параметров эти параметры используются при моделировании самого процесса образования событий для получения прогноза будущих моментов времени возникновения этих событий и соответствующих им искомым количественных значений.

Положения, выносимые на защиту:

1) Предложен общий метод анализа и прогнозирования редких событий в экономике, использующий информацию о процессе образования событий. Суть метода заключается в 5 шагах: разделение событий в зависимости от

источников, где они образованы; составление модели процесса образования событий; восстановление параметров процесса по выборке редких событий; экстраполяция значений параметров процесса на будущее; запуск процесса с заданными параметрами и прогноз будущих событий (С. 68–72).

2) Разработан емкостный метод анализа и прогнозирования редких событий, которые образованы процессами, схожими с процессами наполнения/опустошения емкости (С. 72–80, примеры применения С. 81–115). Разработан математический метод восстановления параметров этих процессов на основе сплайновой коллокации (восстановления функции в виде сплайна по последовательности интегралов, наблюдаемых с погрешностью) (С. 76–78; 129–146). Проведено исследование по выбору оптимального коэффициента сглаживания с помощью метода L-кривой, кросс-валидации и принципа невязки Морозова (С. 146–149).

3) Разработан программный инструментарий на языке R, реализующий соответствующий математический метод восстановления параметров процесса образования событий в виде сплайна (восстановление функции по интегралам, наблюдаемым с погрешностью) (С. 149–162). Показано как реализовать аналогичные вычисления средствами Excel (С. 162–169).

4) Получены формулы расчета дополнительной погрешности наблюдений, возникающей вследствие дискретности измерения времени (С. 171–178) и в зависимости от удаленности от конечного потребителя (С. 179–213). Проведено моделирование и выявлено как растет погрешность восстановления параметров процесса при потере части данных (С. 214–224). Проведено моделирование и выявлено как растет погрешность восстановления параметров процесса вследствие неопределенности внутри самого процесса образования событий (С. 225–243).

5) Разработана методика проверки адекватности для сконструированных с помощью емкостного метода моделей прогнозирования редких событий (С. 253–258).

б) Произведено обобщение емкостного метода для анализа социальных и исторических событий на примере «цветных революций» (С. 261–266) и русско-турецких войн (С. 266–270).

7) Разработан математический метод восстановления параметров процессов образования событий по разным функционалам, не только по интегралам, но также по первым и вторым производным, с помощью сплайновой коллокации (С. 270–275, пример С. 277–282), а также для этой задачи получены формулы для выбора оптимального параметра сглаживания с помощью кросс-валидации (С. 275–277).

8) Разработан метод восстановления параметров процесса образования событий, представленного в виде алгоритмической модели, которая может быть произвольным алгоритмом (С. 282–298, пример С. 298–303). Данный метод является универсальным и может использоваться для анализа и прогнозирования событий, образованных произвольным процессом, для которого можно составить алгоритмическую модель.

9) Разработан метод анализа и прогнозирования совсем неизвестных событий с помощью подбора алгоритмических моделей процесса их образования (С. 303–307), который базируется на ранее указанном универсальном методе восстановления параметров процесса образования событий.

Теоретическая и практическая значимость работы. Теоретические положения диссертационного исследования формируют фундамент для целого направления в анализе и прогнозировании редких событий, в котором основным элементом является информация о процессах образования событий. На основе этого фундамента можно сконструировать целое множество моделей, предсказывающих появление новых событий в различных прикладных областях. В то же время восстановление динамически изменяющихся параметров процессов образования событий открывает возможность к исследованию еще неизвестных зависимостей и закономерностей, что представляет собой отдельный теоретический интерес.

Созданная теория исследований редких событий образует новое направление и в будущем может привести к образованию нового направления обучения студентов, которые бы специализировались на построении и оценке параметров моделей процессов формирования событий в экономике.

Практическая значимость исследования заключается как раз в том, что с помощью разработанного метода можно анализировать и прогнозировать редкие события в экономике, что позволит подготовиться к будущим событиям и извлечь определенную выгоду. Разработанным методом, например, могут пользоваться аналитики отделов продаж, отделов планирования производства и др. Это дает возможность более эффективно сформировать план производства, план закупки сырья, план пополнения запаса, оптимизировать логистические расходы и т.д. Благодаря разработанному методу можно уменьшить объемы продуктовых запасов товаров, лежащих на складах. В результате минимизировать различные виды издержек, например издержки на хранение, организовать более высокий уровень сервиса, добиться большего удовлетворения клиентов. Применение разработанного метода в маркетинге позволит заблаговременно предупреждать клиентов о том, что подходит время пополнять запасы продукции у правильного продавца/поставщика. Разработанный математический метод анализа редких событий может использоваться в других областях, а не только в экономике, выявлять скрытые закономерности, давать прогноз возникновения будущих событий.

Методология и методы исследования. Формирование моделей процессов образования событий опирается на экономическую теорию, системный анализ, теорию управления запасами.

Восстановление параметров процессов образования редких событий опирается на теорию вероятностей, математический анализ, сплайновую коллокацию (восстановление функции по функционалам), методы регуляризации, такие как кросс-валидация, метод L-кривой и невязки Морозова. Восстановление параметров алгоритмических моделей базируется

на имитационном моделировании, методах оптимизации черного ящика, теории вероятностей.

Исследование вопросов точности опирается на математический анализ, теорию вероятностей, имитационное моделирование, метод Монте-Карло.

Степень достоверности, апробация и внедрение результатов исследования. Достоверность исследования подтверждается грамотным использованием математического аппарата, используемых моделей и методов. Корректность полученных в исследовании математических выражений подтверждается на представленных в работе численных примерах. Справедливость предложенных методик обосновывается компьютерным моделированием и примерами, основанными на реальных данных.

Результаты исследования были доложены и обсуждены на ряде международных/всероссийских научных конференциях/круглых столах/научных семинарах: III Всероссийской научной конференции «Россия 2030 глазами молодых ученых» (Москва, ИНИОН РАН, 26 апреля 2012 г.); VI Международной научно-практической Интернет-конференции «Современные проблемы моделирования социально-экономических систем» (г. Харьков, Харьковский национальный экономический университет, 3–12 апреля 2014 г.); VII Международной научно-практической Интернет-конференции «Современные проблемы моделирования социально-экономических систем» (г. Харьков, Харьковский национальный экономический университет, 2–10 апреля 2015 г.); Международном круглом столе «Системная экономика, социально-экономическая кибернетика, мягкие измерения в экономике – 2017» (Москва, Финансовый университет, 8 июня 2017 г.); Научном семинаре на кафедре «Системный анализ в экономике» (Москва, Финансовый университет, 13 марта 2018 г.); Ежегодном международном круглом столе «Системная экономика, социально-экономическая кибернетика, мягкие измерения в экономике – 2018» (Москва, Финансовый университет, 6 июня 2018 г.); Всероссийском межвузовском круглом столе «Мягкие измерения в научной и учебной деятельности»

(Москва, Финансовый университет, 17 сентября 2018 г.); V Международной научно-практической конференции-биеннале «Системный анализ в экономике» - 2018 (Москва, Финансовый университет, 21-23 ноября 2018 г.); Международной научно-практической конференции «Анализ данных, принятие решений и Финансовые технологии», посвященной 100-летию Финансового университета (Москва, Финансовый университет, 23 мая 2019 г.); Ежегодной международной конференции «Системная экономика, социально-экономическая кибернетика, мягкие измерения в экономике – 2019» (Москва, Финансовый университет, 7 июня 2019 г.); Международной научно-практической конференции «Системная экономика, социально-экономическая кибернетика и мягкие измерения в экономике» (Москва, Финансовый университет, 20 мая 2020 г.); VII Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 25 июня 2020 г.); VI Международной научно-практической конференции-биеннале «Системный анализ в экономике – 2020», (Москва, Финансовый университет, 9–11 декабря 2020 г.); Научном семинаре в Департаменте анализа данных и машинного обучения на Факультете информационных технологий и анализа больших данных (Москва, Финансовый университет, 20 января 2021 г.); VIII Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 28 мая 2021 г.); Международной научно-практической конференции «Системная экономика, социально-экономическая кибернетика и мягкие измерения в экономике» (Москва, Финансовый университет, 24–25 июня 2021 г.); Научном семинаре в ЦЭМИ – «Прикладная статистика и моделирование реальных процессов» (Москва, ЦЭМИ, 8 декабря 2021 г.); Научном семинаре на секции «Эконометрика» в Департаменте математики (Москва, Финансовый университет, 20 января 2022 г.); IX Международной научно-практической конференции «Современная

математика и концепции инновационного математического образования» (Москва, Финансовый университет, 9 июня 2022 г.); Международной научно-практической конференции «Системная экономика, социально-экономическая кибернетика и мягкие измерения в экономике» (Москва, Финансовый университет, 1 июля 2022 г.); VII Международной научно-практической конференции-биеннале «Системный анализ в экономике – 2022» (Москва, Финансовый университет, 7–8 декабря 2022 г.).

Результаты исследования были использованы при выполнении НИР по одноименной теме проекта исследований гранта РФФИ «Емкостный метод анализа редких событий в экономике» в 2019–2021 г., № 19-010-00154.

Результаты исследования были использованы при выполнении НИР Финуниверситета (Государственное задание, приказ от 20.03.2020 № 0564/о) по теме «Системные атрибуты цифровой экономики как среды развития инновационных процессов в России».

Получено Свидетельство о государственной регистрации программы для ЭВМ №2020617973 «Программный продукт, реализующий метод восстановления функции по последовательности интегралов, которые наблюдаются с погрешностью, основанный на интегральных сплайнах со штрафной функцией».

Материалы диссертационной работы внедрены в практическую деятельность компании ООО «АУМЕД» в целях улучшения планирования логистических процессов. Разработанные методики и программный инструментарий применяется в текущей деятельности для анализа и прогнозирования заказов отдельных корпоративных клиентов, возникновения спроса, планирования запасов, что способствует определенности в принятии решений.

Материалы диссертационной работы внедрены в практическую деятельность компании ООО «Квайссер Фарма» в целях совершенствования процессов планирования реализации фармацевтической продукции. Разработанная методика анализа редких событий позволяет определять

будущие потребности в поставках. Отдельные положения применялись при обосновании спада или повышения спроса на биологически активные добавки в период пандемии COVID-19.

Материалы докторской диссертации используются кафедрой «Системный анализ в экономике» в преподавании учебных дисциплин: «Экономико-математическое моделирование логистики», «Математические модели и методы в логистике».

Апробация и внедрение результатов исследования подтверждены соответствующими документами.

Публикации. Материалы исследования опубликованы в 30 научных публикациях общим объемом 41,51 п.л. (авторский 38,57 п.л.), в том числе в: 20 статьях в рецензируемых научных изданиях, определенных ВАК при Минобрнауки России, общим объемом 26,96 п.л. (авторский объем – 24,02 п.л.); 3 статьях, входящих в цитатно-аналитическую базу RSCI общим объемом 6,65 п.л. (весь объем авторский); 1 статье в журнале второго квартиля (Q2) международной цитатно-аналитической базы «Scopus», авторский объем 2,0 п.л.; 1 статье в журнале первого квартиля (Q1) международной цитатно-аналитической базы «Scopus», авторский объем 2,1 п.л..

Структура и объем диссертации. Диссертация состоит из введения, 5 глав, заключения, списка литературы, состоящего из 148 наименований, 6 приложений. Текст диссертации изложен на 342 страницах, содержит 44 таблицы и 130 рисунков.

Глава 1

Обзор существующих методов анализа редких событий

1.1 Обзор терминологии и задач, в которых используется термин «редкие события»

Термин «событие» в русском языке имеет давнее происхождение. Согласно историку-языковеду и академику РАН Виноградову В.В. [5] слово «событие» встречается аж в 1096 году в церковной письменности. Слово «событие» в церковнославянском языке имеет значение «исполнение» и «осуществление». В словаре Даля [6] определяется как «происшествие, что сбылось». Но в теории вероятностей термин «событие» уже понимается как результат некоторого опыта или испытания. В данном диссертационном исследовании термин «событие» понимается именно в широком смысле, что что-то произошло или случилось, не обязательно связанное со статистикой, хоть речь ниже и будет вестись о различных статистических методах. Отмечу, что в результате некоторого испытания или измерения полученное значение само по себе может являться интересующим исследователя явлением, то есть быть событием. При использовании термина «событие» мы не будем себя ограничивать природой того, что может «произойти» или «сбыться». В то же время данное исследование носит экономический характер, поэтому в основном мы будем рассматривать интересующие нас с экономической точки зрения события (мы не будем рассматривать физические или геологические события, такие как испускание фотонов или землетрясения, хотя они также имеют огромное значение для экономики).

Среди существующих методов анализа и прогнозирования отдельно выделяют методы для работы с частыми и редкими событиями. Под частотой или редкостью событий понимают либо насколько количество одних событий за наблюдаемые интервалы времени превосходит количество других событий, либо насколько вероятности возникновения этих событий за промежуток времени отличаются друг от друга. Говоря о частоте или редкости как о

количестве событий, следует понимать, что это относительные характеристики. Для одних систем даже тысяча событий в секунду может быть недостаточной, чтобы такие события назывались частыми, и с другой стороны, для некоторых систем события, которые возникают раз в несколько месяцев, по-прежнему могут являться частыми событиями. Зачастую под редкими событиями понимают также более важные и значимые события, имеющие серьезные последствия, хотя это не всегда может быть так. Николас Талеб называет события, которые никто не ожидал (аномальные), обладающие огромным воздействием и заставляющие человека искать объяснение их возникновению, как «Черные лебеди» [7]. В теории информации говорится о том, что редкие события уже из-за того, что они редкие, несут большую информацию [8]. Между частыми и редкими событиями сложно провести четкую грань. Где заканчиваются частые события и начинаются редкие события – это известная философская проблема, относящаяся ко второму закону диалектики «закон перехода количества в качество» (Энгельс Ф.). Различия могут быть определены только в контексте исследуемой проблемы.

В исследовании [9] авторы, делая обзор множества (около 200) вышедших международных статей, анализируют используемую другими авторами этих статей терминологию при работе с редкими событиями. В исследовании авторы выявляют, что редкие события зачастую называют по-разному, подразумевая одно и то же, а также наоборот, называя одинаково, подразумевают совсем разные понятия. Самыми частыми терминами, по их мнению, являются: *rare event*, *anomaly*, *novelty and outlier* (редкие события, аномалии, новизна, выбросы). Задачи, в которых используется эта терминология, немного отличаются.

Среди анализируемых статей авторы выделяют следующие задачи, в которых используется термин «редкие события»:

- классификация редких событий во временных рядах;
- классификация редких событий по наблюдаемым признакам;

- определение вероятности возникновения события в заданном периоде времени;
- определение вероятности выхода за границы допустимых условий.

В работах, в которых используется терминология «обнаружение аномалий», также решается задача классификации событий по наблюдаемым признакам. Термин «обнаружение новизны» встречается в работах, в которых анализируется выборка наблюдений, принадлежащих только одному классу, и новые наблюдения, которые слишком отличаются от них, классифицируются как «новизна», одновременно с этим формируя новый класс наблюдений, если таких наблюдений накопится большое количество. Термин «обнаружение выбросов» в основном используется при работе с сильно зашумленными данными, где задача заключается в том, чтобы определить, что является выбросом, а что нет.

В упомянутом исследовании не говорится о еще одном термине, а именно «инцидент» или «анализ инцидентов», который используется в том же контексте, что и термин «событие» [10]. Стоит обратить внимание, что в упомянутом исследовании, в котором изучаются другие существующие работы, посвященные редким событиям, а также в других работах, изученных автором данного диссертационного исследования, задача анализа редких событий с точки зрения процессов, порождающих редкие события, ранее не встречалась.

Кратко дадим пояснения упомянутым задачам.

1) Задача классификации редких событий во временных рядах. Рассматривается некоторый временной ряд, в котором отображается изменение одной наблюдаемой величины в течение времени. Обнаружение редкого события подразумевает выявление в данных этого же временного ряда определенных характерных картин, которые появляются сравнительно редко. Ключевой особенностью является то, что распознавание происходит на основе все того же временного ряда, а не какого-либо другого. Примером таких

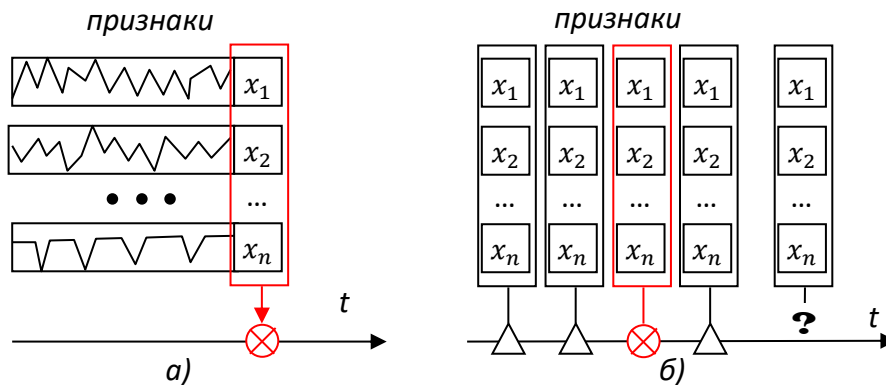
редких событий может являться кратковременный скачок, рисунок 1.1, например, скачок цен или напряжения в сети.



Источник: составлено автором по материалам [9].

Рисунок 1.1 – Классификация редких событий во временном ряду

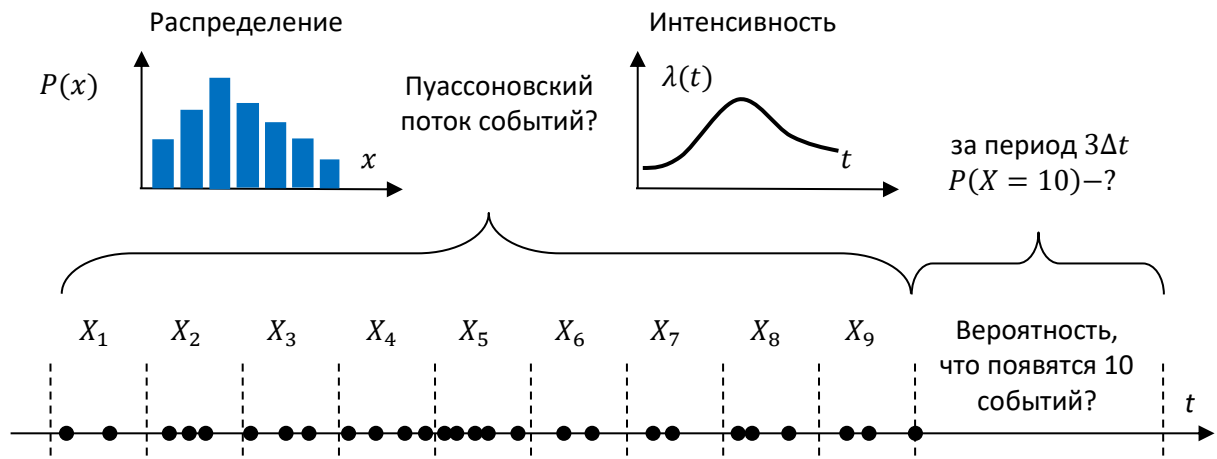
2) Задача классификации редких событий по наблюдаемым признакам. Можно дать несколько интерпретаций, либо наблюдаются значения сразу нескольких временных рядов, показывающих значения некоторых признаков (характеристик) определенного объекта или явления, по которым надо идентифицировать наступление редкого события, либо фиксируются события, обладающие набором признаков, по которым необходимо определить, относится ли оно к обычным или редким. В первом случае событий как таковых еще нет, но факт наблюдения определенной комбинации признаков сам становится событием, рисунок 1.2а. Во втором наоборот, события уже есть, и их надо отнести к обычным или необычным, рисунок 1.2б. Примером в первом случае может являться обнаружение аномальных погодных условий, во втором случае выявление среди потока машин велосипедистов. Данная задача является более общей по сравнению с первой задачей.



Источник: составлено автором по материалам [9].

Рисунок 1.2 – Классификация редких событий по наблюдаемым признакам

3) Задача определения вероятности возникновения события в заданном периоде времени. Здесь по статистическим данным определяется частота, закон распределения или закон и интенсивность возникновения событий за единицу времени. Тогда для периода времени заданной длины определяется вероятность возникновения заданного количества событий, рисунок 1.3. Примером может служить определение вероятности появления сразу 10 обращений за 15 минутный интервал времени.



Источник: составлено автором.

Рисунок 1.3 – Определение вероятности возникновения событий в заданном промежутке времени

4) Задача определения вероятности выхода за границы допустимых условий. В данной задаче задается граница допустимых условий в виде системы сложных нелинейных уравнений в многомерном пространстве параметров, закон распределения которых также задается определенным образом. Для определения вероятности выхода за границу допустимых условий требуется рассчитывать многомерные интегралы по сложным фигурам, образованным пересечением границ, что аналитически не рассчитывается, а рассчитывается приближенно с помощью метода Монте-Карло, метода надежности первого порядка (First-order reliability method, FORM) [1], метода надежности второго порядка (SORM) [2] (подробный обзор [11; 12]), выборки по линиям (Line sampling) [3; 4] и их модификаций. Вероятности редких событий, которые связаны с тем, что комбинация параметров отклонится в определенную очень узкую область, может быть

очень сложно рассчитать даже с помощью численных методов, особенно когда параметров сотни или тысячи. Для этого требуются особенные математические методы [13; 14]. Данная задача достаточно специфичная, она сильно отличается от ранее перечисленных. Термин «редкие события» тут используется как аналог понятию возникновения неисправности в некоторых технических устройствах или системах.

В первых двух задачах решается задача классификации и распознавания. Однако наиболее полезно не просто классифицировать события, а предсказывать моменты их появления. Поэтому прежде всего нас будет интересовать задача предсказания моментов появления редких событий. Обратим внимание, что если в задачах классификации в качестве признаков использовать характеристики, отнесенные к предшествующим моментам времени (лаговые значения), то получающиеся классификационные модели могут обладать способностью предсказывать наступление будущих событий по наблюдаемым признакам. В третьей задаче определяются статистические параметры потока событий и затем определяются вероятности возникновения заданного количества событий в следующий период, однако моменты времени появления событий не предсказываются. В четвертой задаче лишь определяется вероятность возникновения редкого события, когда случайная величина окажется в/вне нужной области многомерного пространства при разыгрывании самой случайной величины. К предсказанию моментов времени возникновения событий данная задача отношения не имеет, поэтому мы не будем ее разбирать.

Также надо упомянуть следующие задачи, которые могут встретиться при работе с событиями.

Иногда события анализируют лишь для более компактного и наглядного изображения. Перед исследователем в этом случае стоит задача подобрать такую математическую функцию, которая бы основываясь на этих данных, выглядела бы как можно красивее. Тогда перед исследователем может возникнуть необходимость сглаживания данных, а в некоторых случаях даже

фильтрации данных, чтобы отфильтровать некрасивые аномальные значения (выбросы).

В более сложном случае исследователю необходимо анализировать события с целью определения как одни действия влияют на наблюдаемый результат, а как влияют другие. Перед исследователем тогда стоит задача определения показателей зависимости одних наблюдаемых величин с другими наблюдаемыми величинами, как связаны одни и другие величины. Например, такой задачей может быть определение зависимости количества совершаемых покупок от количества пришедших покупателей, определение цены на некоторые товары от количества покупок этих товаров, или наоборот. Перед исследователем возникает необходимость определения математической модели и ее параметров, определяющей эту взаимосвязь, которая в общем виде может быть непростой. Похожей задачей может являться определение взаимозависимости группы наблюдаемых переменных, где каждая наблюдаемая переменная зависит от многих других переменных, в том числе и от самой себя, взятой за предшествующий период наблюдения.

Самой же сложной задачей является задача прогнозирования будущих значений интересующих исследователя величин. Примером таких задач может служить прогнозирование количества покупок в следующем месяце или спрос за заданный временной промежуток. Прогнозирование может быть не связано с временем, например, когда исследователя интересует, чему будет равна интересующая его величина в рамках определенной им ранее модели при подстановке других входных значений, от которых зависит искомая величина. Тем не менее, так как человек, зная прошлое, живя в настоящем, всегда интересуется будущим, большинство прогнозов так или иначе связаны с фактором времени. Основная сложность при прогнозировании заключается в том, что будущие значения в общем случае могут не соответствовать тем закономерностям, которые были определены по прошлым наблюдениям. Причин несколько, первая в том, что исследователь мог банально ошибиться при построении модели закономерности. Вторая, что при выявлении

закономерности рассматриваемые наблюдения соответствовали лишь узкому периоду времени наблюдений, то есть исследователь выявлял закономерность лишь в локальной области, а не глобальной. Третья, закономерность изменения интересующих исследователя величин может зависеть от таких внешних факторов, которые исследователь даже не предполагает рассматривать, изменение этих факторов может кардинальным образом повлиять на ход всего процесса изменения наблюдаемой величины (бифуркация).

При прогнозировании будущих событий исследователь вынужден принимать предположение о том, что будущие события подчиняются закономерности, выявленной им по прошедшим наблюдениям. Чисто математические модели, зависящие только от фактора времени, часто критикуют, говоря о том, что это «как ехать на машине вперед, смотря лишь в зеркало заднего вида». Однако модели могут также учитывать другие количественные и даже качественные факторы. Поэтому такая критика не всегда справедлива. Некоторые поведенческие модели, прогнозируя будущее, опираются не на зависимости от времени, а на моделирование поведения элементов (участников) изучаемой системы, когда время используется лишь в качестве шагов, на которых происходят изменения состояния элементов под воздействием других элементов системы.

Таким образом, анализ событий может использоваться для выявления закономерностей и прогнозирования будущих событий. Примерами наиболее распространенных задач в экономике (например, из области производства и логистики некоторой компании), в которых чаще всего приходится использовать прогнозирование будущих событий и для которых это имеет большое значение, являются следующие.

- 1) Есть некоторое предприятие, выпускающее определенную продукцию. Для того, чтобы спланировать производство, руководитель предприятия требует спрогнозировать продажи в каждый из следующих нескольких месяцев.

2) Есть некоторый торговый центр, реализующий продукцию многочисленным покупателям. Для того чтобы определить размер страхового запаса, требуется спрогнозировать суммарные продажи за несколько следующих месяцев.

3) В отдел оптовых продаж периодически через большие интервалы времени обращаются оптовые покупатели. Требуется определить вероятность того, что за следующую неделю будет заданное количество обращений.

4) В том же отделе продаж требуется определить конкретный момент времени, когда позвонит определенный клиент.

5) Определить вероятность возникновения особого события при заданном стечении обстоятельств.

Решение этих задач может серьезно влиять на экономическое состояние компаний. На основе прогноза может строиться текущее функционирование компании и даже вся ее будущая политика развития. Первая, вторая и пятая задача может быть отнесена к задачам классификации редких событий во временном ряду или по наблюдаемым признакам, упомянутым выше. Третья задача относится к задаче определения вероятности возникновения события в заданном периоде времени. Четвертая задача только в редких частных случаях частично может быть отнесена к задаче классификации с использованием лаговых переменных, и то только когда будут наблюдаться значения этих переменных. В общем, случае если ничего не наблюдается, такая задача не может быть решена методами классификации. Методы работы с пуассоновскими потоками событий также не могут предсказать сам момент наступления события, а потоки Пальма только стационарны. Существование других методов, позволяющих решить эту задачу, не выявлено (в данном диссертационном исследовании будет разработан метод, который может решать все эти четыре задачи, включая и определение момента наступления следующего события).

Вместе с тем надо обратить внимание, что любой прогноз может естественно не сбыться, причем либо частично, либо полностью. Если было

определено, что спрос будет расти, а на самом деле он упал, то компания несет значительные убытки. Если компания рассчитывала, что спрос будет падать, а на самом деле он возрос, то компания теряет прибыль и, более того, позволяет конкурентам занять часть рынка. Это были примеры, когда действительность полностью не соответствовала прогнозируемым ожиданиям компании. Однако даже тогда, когда действительность соответствует ожиданиям компании, небольшие отклонения могут иметь серьезное влияние на финансовое состояние компании. Если компания предполагала, что спрос возрастет на 5%, а он возрос на 4% или 6%, то компания будет опять нести убытки в результате перепроизводства или упущенной выгоды. Опять же, пусть уже маленькими, но все же ошибками могут воспользоваться конкуренты, что может ослабить позицию компании.

Точность прогнозов имеет большое значение. Наличие неопределенности заставляет компании резервировать часть средств для страхования рисков, вместо того чтобы эти средства направлять на развитие. Увеличение точности методов анализа и прогнозирования может уменьшить эти риски, вследствие чего освободить и пустить на развитие часть зарезервированных средств. Уменьшение рисков может иметь серьезное социальное значение, так уменьшение рисков в ипотечном кредитовании позволяет снизить процентную ставку по кредиту и сделать жилье более доступным [15].

Это приводит к тому, что надо выбирать наиболее точные методы прогнозирования. Причем некоторые методы хорошо работают с одним типом задач и плохо работают с другим типом задач. Следует подбирать методы в соответствии с решаемой задачей. Далее рассмотрим различные методы анализа и прогнозирования, а еще ниже попробуем соотнести эти методы с возможными задачами.

1.2 Существующие методы анализа редких событий

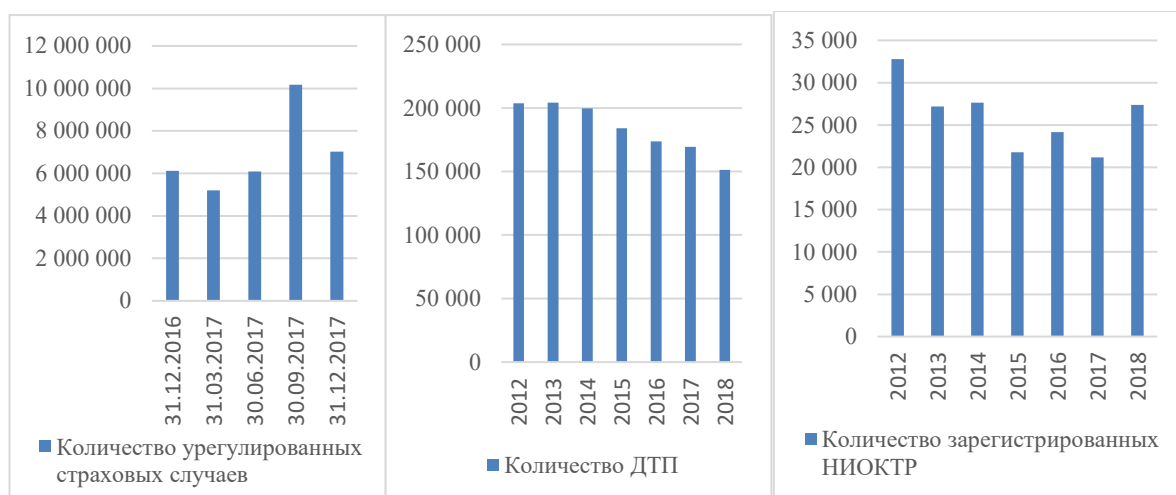
Часть математических методов базируются на статистике и теории вероятностей, пытаясь лишь найти закономерности в данных о количестве появления событий вместо того, чтобы анализировать каждое событие индивидуально и разбираться в причинах его появления. В таких методах события, будь они частые или редкие, представляются в виде временного ряда ¹). Для этого подсчитывают количество событий на каждом последовательно расположенном интервале времени. Причем, то, что интервалы времени могут быть разной ширины, например, когда в месяцах разное количество дней, как правило игнорируют. Примерами таких временных рядов могут являться: количество покупок, совершенных каждый месяц; количество пассажиров, пользующихся общественным транспортом; количество страховых случаев в каждом квартале; количество дорожно-транспортных происшествий; количество людей, прошедших вакцинацию за год; количество работ, подаваемых на конкурс за год и так далее. Эти данные представляют либо в табличном виде, например, таблица 1.1, либо в графическом виде, используя либо столбиковые диаграммы, рисунок 1.4, либо график из точек, иногда соединенных между собой отрезками прямых линий.

Таблица 1.1 - Пример построения временного ряда из событий

Количество урегулированных страховых случаев (за квартал)		Количество ДТП (в год)		Количество зарегистрированных НИОКТР (в год)	
31.12.2016	6 121 658	2012	203 597	2012	32 797
31.03.2017	5 198 440	2013	204 068	2013	27 201
30.06.2017	6 085 614	2014	199 720	2014	27 636
30.09.2017	10 172 470	2015	184 000	2015	21 791
31.12.2017	7 024 209	2016	173 700	2016	24 160
-	-	2017	169 432	2017	21 168
-	-	2018	151 291	2018	27 366

Источник: составлено автором по данным [16; 17; 18].

¹) Ряд последовательных значений, характеризующих изменение показателя во времени (Справочник технического переводчика и Экономико-математический словарь).



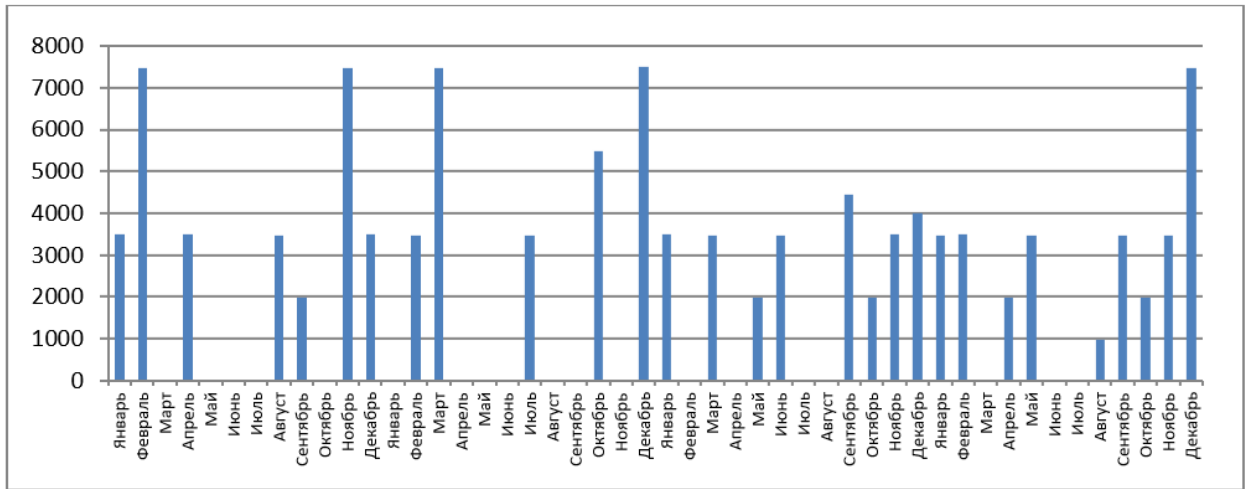
Источник: составлено автором по данным [16; 17; 18].

Рисунок 1.4 - Изображение частых событий в виде диаграмм

Далее полученные таким образом временные ряды анализируются. Для анализа данных частых событий используют проверенные и зарекомендовавшие себя математические методы: методы фильтрации и сглаживания, спектральный анализ, фазовый анализ, методы аппроксимации, методы исследования временных рядов, факторный анализ и адаптивные методы и др. Эти методы, если получится правильно выявить закономерность из этих данных, могут предсказывать уровни соответствующего временного ряда в будущих периодах.

В отличие от частых событий редкие события, если представлять их по-прежнему в виде временного ряда, как правило будут содержать огромное количество нулевых значений. Например, определяя спрос по месяцам при оптовых продажах нескольким клиентам, получаем график, рисунок 1.5. Причем от выбора ширины интервала времени, за которое подсчитывается количество событий, зависит количество нулевых значений. Чем уже интервалы времени, тем реже события в них попадают и тем чаще количество нулевых значений. Конечно, можно делать интервалы шире, чтобы в них попало как можно больше значений, но при этом уменьшается количество самих интервалов, меняется структура временного ряда. Более того выбор ширины интервалов времени часто продиктован смыслом рассматриваемых значений, исходя из целей дальнейшего использования. Бессмысленно

объединять интервалы в кварталы или в года, если в дальнейшем потребуется прогноз спроса за месяц.



Источник: составлено автором.

Рисунок 1.5 - Представление редких событий в виде временных рядов

Конечно, во временных рядах для частых событий иногда также встречаются нулевые значения, но это происходит в основном тогда, когда общий порядок наблюдаемых значений разбросан в районе нуля. Также стоит обратить внимание, что наличие нулевых значений ни в коем случае не надо путать с отсутствием значений, с чем очень часто приходится сталкиваться в анализе больших данных с помощью машинного обучения. Отсутствующие данные всего лишь говорят, что за выбранный период измерения не проводились, не фиксировались (поменялась форма анкеты, которую ранее заполняли, поэтому в старых записях часть значений отсутствует). В отличие от отсутствующих значений нулевые значения говорят о том, что измерения за соответствующий период проводились, но значения были равны нулю.

Для анализа такого вида временных рядов, содержащих большое количество нулевых значений, классические статистические и математические методы анализа данных, которые успешно работают с данными частых событий, к редким событиям плохо подходят. Существуют специальные методы, некоторые из них более наивные, они используют достаточно жесткие мало реалистичные предположения, другие менее наивные, они используют более реалистичные предположения. В последних

работах [19; 20] делается широкий обзор множества подходов для прогнозирования или работы в условиях редкого (прерывистого, *intermittent*) спроса (на основе метода Кростона или бутстреппинга, а также эвристические подходы). Однако ничего подобного, что предлагается в данной диссертационной работе, там не упоминается. Далее рассмотрим метод Кростона и метод Виллемейна (бутстреппинга), которые чаще всего используются в логистике. Данные методы особенным образом работают со значениями временных рядов. Также рассмотрим популярные методы классификации, такие как метод ближайших соседей и логистической регрессии. Помимо этого мы рассмотрим теорию случайных процессов, работающую не с временными рядами, а с потоками событий.

Метод Кростона

В 1972 году Кростон в работе [21] предложил очень простую методику для прогнозирования редких продаж (редкого спроса), но тем не менее базирующуюся на жестких предположениях. Предполагается, что ненулевой спрос Y_i в каждый интервал наблюдения возникает случайно и независимо с одинаковой вероятностью p (распределение Бернулли). В то же время размер ненулевого спроса распределен по нормальному закону с одинаковым математическим ожиданием m и дисперсией σ^2 , причем тоже независимо. Хотя сами случайные величины и соответствуют одному закону распределения, но характеристики закона могут быть нестационарными, т.е. вероятность возникновения ненулевого спроса p , математическое ожидание ненулевого спроса m и дисперсия σ^2 могут меняться со временем. Поэтому для определения этих характеристик он предлагает использовать экспоненциальное сглаживание.

Метод заключается в следующем, Исходные данные разбиваются на два ряда: ряд ненулевых значений Y_i и ряд длительности времени (количества интервалов наблюдения) до появления ненулевого значения τ_i . Далее эти два ряда сглаживаются экспоненциально описанным ниже способом

(коэффициент сглаживания α Кростон рекомендует брать в районе 0.1–0.2), и получаются ожидаемые значения ненулевого спроса Z_i и количества интервалов между ненулевым спросом T_i , формулы (1.1) и (1.2)

$$Z_i = (1 - \alpha)Z_{i-1} + \alpha Y_i, \quad (1.1)$$

$$T_i = (1 - \alpha)T_{i-1} + \alpha \tau_i. \quad (1.2)$$

Тогда вероятность появления ненулевого спроса определяется как $p = 1/T_i$.

Далее Кростон определяет превышение «ожидаемого» значения спроса над фактическим $e_i = Z_i - Y_i$ (формирует новый ряд), опять сглаживает полученные значения экспоненциально $\sigma_i = (1 - \alpha)\sigma_{i-1} + \alpha|e_i|$ и определяет уровень страховых запасов в трехкратном объеме $S_i = k\sigma_i$, где $k = 3$. Величина заказа определяется с учетом «ожидаемого» спроса и страхового запаса $R_i = Z_i + k\sigma_i$. Также Кростон вводит специальные условия: $(1 - p)^{t-t_n} < 0.01$ – при получении нулевого спроса сигнализирует, что очень долго нет спроса; $(t - t_n)p < 0.2$ – при получении ненулевого спроса сигнализирует о том, что спрос наступил слишком рано (где $t - t_n$ показывает сколько единиц времени прошло с последнего ненулевого спроса).

Пример. Пусть дан временной ряд таблица 1.2. Из этого временного ряда формируются два ряда Y_i и τ_i , таблица 1.3.

Таблица 1.2 – Временной ряд значений спроса y_t

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	3500	7480	0	3486	0	0	0	3477	1986	0	7456	3489
t	13	14	15	16	17	18	19	20	21	22	23	24
y_t	0	3470	7464	0	0	0	3474	0	0	5469	0	7486
t	25	26	27	28	29	30	31	32	33	34	35	36
y_t	3487	0	3471	0	1976	3480	0	0	4456	1989	3484	3996
t	37	38	39	40	41	42	43	44	45	46	47	48
y_t	3471	3483	0	1985	3476	0	0	978	3474	1995	3471	7456

Источник: составлено автором.

Таблица 1.3 – Ряд ненулевых значений спроса Y_i и длительности между ненулевым спросом τ_i

i	Y_i	i	Y_i	i	τ_i	i	τ_i
1	3500	16	3480	1	1	15	1
2	7480	17	4456	2	2	16	3
3	3486	18	1989	3	4	17	1
4	3477	19	3484	4	1	18	1
5	1986	20	3996	5	2	19	1
6	7456	21	3471	6	1	20	1
7	3489	22	3483	7	2	21	1
8	3470	23	1985	8	1	22	2
9	7464	24	3476	9	4	23	1
10	3474	25	978	10	3	24	3
11	5469	26	3474	11	2	25	1
12	7486	27	1995	12	1	26	1
13	3487	28	3471	13	2	27	1
14	3471	29	7456	14	2	28	1
15	1976						

Источник: составлено автором.

Начиная с $Z_1 = Y_1$ и $T_1 = \tau_1$ проводим экспоненциальное сглаживание с коэффициентом сглаживания $\alpha = 0.15$ и получаем $Z_{29} = 3705,381$ и $T_{28} = 1,3489$. Тогда вероятность возникновения ненулевого спроса определяется как $p = 1/T_{28} = 0,7414$. Сглаженное значение ошибки прогноза $\sigma = 1285,553$. Величина страхового запаса $S = 3856,66$. Сигналом о том, что слишком долго нет ненулевого спроса будет прошествие 4 интервалов времени (3,405 если не округлять). Сигналом о том, что наступление ненулевого спроса возникло слишком рано, будет прошествие всего 0,2698 от интервала, что невозможно для временных рядов.

Хоть данный метод и используется в основном в логистике для определения уровня страховых запасов и величины заказа, однако этот метод, по сути, пытается предсказать момент наступления будущего события через $1/p$ интервалов времени. Конечно же данный метод очень прост, что не может не сказаться на адекватности всех рассчитываемых прогнозов. Наибольшей критике подвергается предположение о независимости наблюдаемых значений спроса и количеством интервалов между их возникновением. Однако из-за своей простоты он стал очень популярным, с момента выхода

публикации (1972) до текущего времени (конец 2020) количество международных статей, цитирующих эту работу, составляло 368, а запрос в обычном поисковике выдавал более 4 млн результатов на различных сайтах. Конечно же существует множество модификаций данного метода, в работе [22] делается обзор и краткий анализ существующих версий этого метода.

Метод Виллемейна (бутстреппинга)

В 1979 году Бернанд Эфрон опубликовал в статье [23] новый метод анализа статистической информации и построения доверительных интервалов, а в последующих работах развил его. Метод получил название «Bootstrap», что можно приблизительно по смыслу перевести как «Самовытягивание». Наиболее популярными были работы [24; 25], которые на текущий момент уже были процитированы безумное количество раз (2278 и 3817 соответственно) в научных статьях, а в интернете поиск выдает 9 млн результатов. Наиболее полно метод представлен в 449 страничной монографии [26], в которой метод представлен с разных сторон.

Метод преимущественно предназначен для построения более точного доверительного интервала для некоторой интересующей исследователя характеристики θ . Вместо того, чтобы откладывать от оценки среднего значения $\hat{\theta}$, оцененной по некоторой выборке $X = \{x_1, \dots, x_n\}$, заданное количество (например, $t_{кр}$ - критическое значение распределения Стьюдента для установленной доверительной вероятности) раз оценок среднего квадратичного отклонения $\hat{\sigma}$, в результате чего получается симметричный доверительный интервал, метод предполагает получить из имеющейся выборки множество случайных выборок $X^* = \{x_1^*, \dots, x_n^*\}$, с помощью случайного выбора элементов из исходной выборки, причем элементы могут выбираться повторно. Количество случайным образом сформированных выборок X^* берется как можно больше, например, 100 000. По каждой такой выборке X^* определяется соответствующая оценка θ^* интересующей исследователя характеристики. Полученные значения оценок θ^* сортируются

по возрастанию (по сути, строится эмпирическая функция распределения $G(\theta)$). Наконец, доверительный интервал для доверительной вероятности $1 - 2\alpha$, например, для вероятности 95%, определяется как интервал между значениями, стоящими на позиции, определяющими, что доля α всех значений оказались меньше и больше соответственно (т.е. $\theta \in [G^{-1}(\alpha); G^{-1}(1 - \alpha)]$). Например, между значениями θ^* , стоящими на позициях 2500 и 97500 в отсортированном списке из 100 000 значений (аналогично построению доверительного интервала методом Монте-Карло для законов распределения, для которых проще генерировать случайные числа, чем искать обратную функцию).

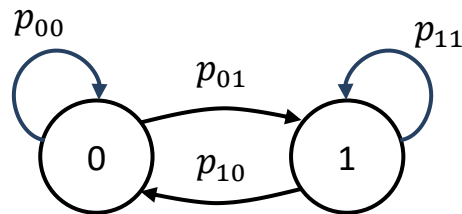
Получающийся доверительный интервал хорош тем, что он может учитывать асимметрию, в отличие от стандартных методов, таких как доверительный интервал на основе критерия Стьюдента. В случае, если исходная выборка была репрезентативна¹⁾, то набор бутстреп оценок, полученных по случайным выборкам, также будет репрезентативным. Вариация оценок при увеличении количества случайных выборок X^* асимптотически сходится к вариации оценки, заключенной в исходной выборке. Однако в случае, если в исходную выборку попал какой-то выброс, то при построении бутстреп оценок этот выброс возьмется с той же вероятностью, что и любое другое наблюдение, что приведет к размножению этого выброса и в результате к смещению границ доверительного интервала. Метод бутстреппинга очень удобен и его легко реализовать, его можно применить к любому набору данных, не беспокоясь о выполнении предположения о нормальности.

В 2004 году в работе [27] Томас Виллемейн адаптировал этот метод для анализа и прогноза редкого (прерывистого) спроса и внес несколько модификаций. Виллемейн постарался исправить недостаток, присущий как методу Кростона, так и обычному методу бутстреппинга при анализе спроса,

¹⁾ Соответствие характеристик, полученных в результате выборочного наблюдения, показателям, характеризующим всю генеральную совокупность - Большой Энциклопедический словарь [146].

а именно, неспособность учитывать возможную корреляцию между следующими нулевыми и ненулевыми значениями. Метод, предложенный Виллемайном, предполагал сначала оценить вероятности переходов для марковского процесса с двумя состояниями, нулевой и ненулевой спрос, рисунок 1.6. Другими словами, вначале требуется оценить следующие вероятности:

- появления ненулевого спроса, если предыдущий спрос был нулевым;
- появления нулевого спроса, если спрос был нулевым;
- появления ненулевого спроса, если предыдущий спрос был ненулевым;
- появление нулевого спроса, если спрос был ненулевым.



Источник: составлено автором по материалам [27].

Рисунок 1.6 – Марковский процесс с двумя состояниями – нулевой и ненулевой спрос, определению подлежат вероятности перехода между состояниями

Далее с определенными вероятностями генерируется по заданной схеме марковского процесса набор случайных величин, причем количество генерируемых случайных величин ограничено временем исполнения заказа, например, на три месяца вперед. Следующим шагом сгенерированные ненулевые элементы заменяются на значения ненулевого спроса, выбранные случайным образом (равновероятно) из имеющейся исходной выборки.

Затем Виллемайн, рассматривая основной недостаток метода бутстеппинга, а именно возвращать только те значения, которые есть в выборке, предложил на ненулевые значения дополнительно накладывать небольшой разброс по нормальному закону распределения. Новое ненулевое значение определяется по формуле (1.3)

$$X^{*'} = 1 + \lfloor X^* + z\sqrt{X^*} \rfloor, \quad (1.3)$$

где X^* – выбранное из выборки ненулевое значение;

z – сгенерированная стандартная случайная величина;

$\sqrt{X^*}$ – задает разброс новой случайной величины.

Причем, если новая величина $X^{*'}$ получается отрицательной, то разброс убирается $X^{*'} = X^*$. Данная процедура в отличие от обычного метода бутстреппинга гарантирует то, что случайным образом отобранные величины получаются уникальными.

Следующим шагом подсчитывается суммарный сгенерированный спрос на интересующем исследователя интервале времени, к примеру за 3 месяца. Затем вся процедура повторяется многократно, например 100 000 раз, чтобы получить объемную выборку из значений суммарного сгенерированного спроса. Т.е. в ранее использованных обозначениях для обычного метода бутстреппинга определяется оценка $\theta^* = \sum_i X^{*'}_i$, затем генерируется очередная выборка и вновь определяется оценка θ^* , и так многократно пока не будет получена выборка из самих оценок θ^* как можно большего объема. Наконец, по выборке значений θ^* о суммарном спросе за следующие периоды определяется доверительный интервал для заданной доверительной вероятности.

С помощью этой процедуры можно легко определять объем заказа с необходимым уровнем удовлетворения спроса (уровень сервиса). Т.е. чтобы удовлетворить 99% спроса достаточно установить объем заказа как 99% квантиль функции распределения случайной величины θ^* (т.е. $G^{-1}(0.99)$). Сравнивая точность предложенного метода на реальных данных о прерывистом спросе в промышленности, Виллемейн определил, что его метод значительно превосходит по точности обычное экспоненциальное сглаживание и метод Кростона. Причем на реальных данных оказалось, что метод Кростона был иногда даже хуже, чем обычное экспоненциальное сглаживание, хотя в предыдущей статье [28] метод Кростона оказался лучше.

По сути метод Виллемейна осуществляет прогноз редких событий по оцененным переходным вероятностям в ранее представленной схеме

марковского процесса. Можно рассчитать ожидаемое количество следующих нулевых интервалов до появления первого ненулевого как $1/p_{01}$, что можно интерпретировать как время будущего события. Однако в самом методе Виллемайна обращается внимание только на распределение для накопленного суммарного спроса за интересующий исследователя период времени.

Метод ближайших соседей

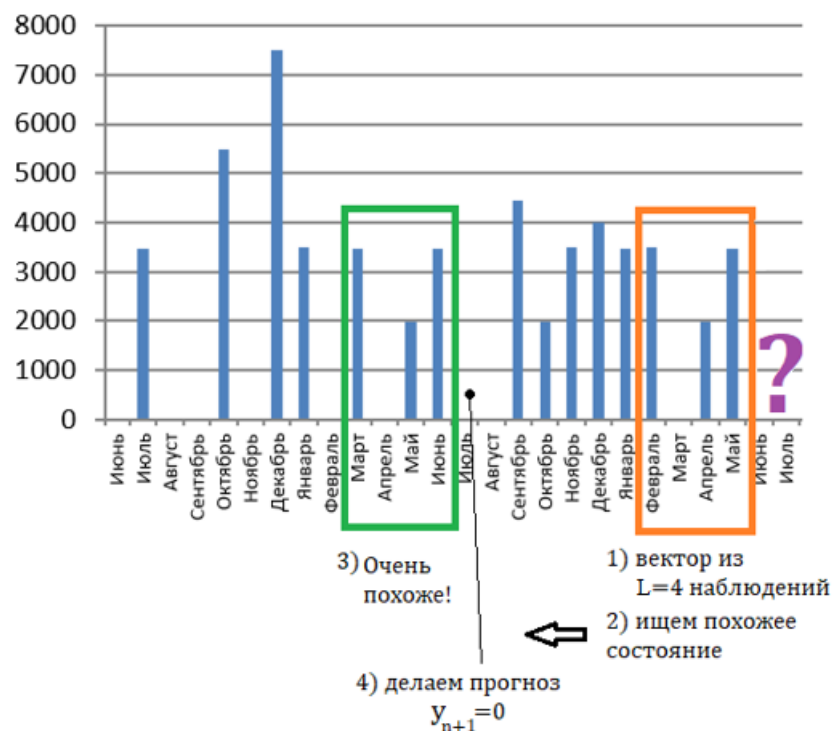
Идея метода ближайших соседей достаточно проста. С одной стороны метод использует элементы скользящей средней, с другой стороны метод использует расчет ближайшего расстояния, из-за этого непросто отследить истоки этого метода. В привычном виде метод ближайших соседей появился в работе [29], но по заявлению авторов правило «ближайшего соседа» появилось в работе [30]. По заверению авторов сайта [31], которые пытались найти создателя метода ближайших соседей, концепция минимального расстояния схожа с концепцией «бритвы Окама»¹⁾, жившего в 14 веке, однако с этим можно не согласиться. Марселло Пелилло из университета Ка Фоскари в Венеции в работе [32] в специальном исследовании установил, что идеи метода ближайших соседей встречались у египетского математика Ибн аль-Хайсама (латинизированное имя Альхазен) в работах по оптике (перевод [33]) в идеях по визуальному распознаванию предметов, жившего в период с 965 по 1039 г.

Метод ближайших соседей является простейшим методом классификации, но он также подходит для непараметрического прогнозирования временных рядов, имеющих сложную структуру. Для задачи классификации метод относит исследуемый объект к тому классу, к которому относится большинство его ближайших соседей. Например, если объект лежит в корзине с яблоками, то он скорее всего тоже яблоко. Под близостью часто понимают эвклидово расстояние, но можно использовать произвольную меру,

¹⁾ «Не следует множить сущее без необходимости», философ-номиналист Уильям из Оккама (1285—1349).

если она лучше подходит для задачи. При этом отбор соседей ограничивается либо максимальным расстоянием, либо количеством уже найденных соседей.

Для задачи прогнозирования временных рядов в качестве соседей выступают дни (или другие интервалы), которые наиболее похожи на сегодняшний день. Для определения «похожести» (близости) выбирают последовательность предшествующих наблюдений фиксированной длины, например, последовательность из $L=4$ наблюдений. Далее просматривают всю последовательность наблюдений и находят такие дни, последовательность наблюдений у которых близка со сравниваемой. По-другому можно сказать, что по ряду данных скользит окно заданной ширины, и если в это окно попадет похожая последовательность, то это и будет наш сосед. Далее находят всех таких соседей, ограничиваясь либо максимальным расстоянием, либо количеством соседей, и затем определяют прогнозное значение как среднее (либо взвешенное по расстоянию) значение следующих за этими соседями наблюдений. Идею метода ближайших соседей, если ограничиться только одним соседом, можно схематично изобразить на рисунке 1.7.



Источник: составлено автором.

Рисунок 1.7 – Идея метода ближайших соседей

В более сложных моделях при поиске соседей сравнивают последовательности не только наблюдаемых значений, но и приращения этих значений, а также бинарный вектор, который состоит из нулей и единиц, сигнализирующих, что произошло увеличение или уменьшение значения, и др. Другими словами, каждому положению окна приписывается состояние, которое может характеризоваться множеством признаков, сформированным определенным образом из данных временного ряда. При использовании одновременно нескольких типов признаков появляется дополнительная трудность с оценкой «близости» состояний. Например, при расчете евклидова расстояния, когда исходные значения измеряются миллионами, а другой признак лишь принимает значения ноль или один, нужны соответствующие масштабирующие коэффициенты. Так, если $x_{i,j}$ – это j -тый признак i -того наблюдения, то, задав масштабирующий множитель C_j для каждого типа признака, расстояние d_i между состояниями, состоящими из L наблюдений, можно оценить по формуле (1.4)

$$d_i = \sqrt{\sum_j C_j \sum_{l=0}^{L-1} (x_{i-l,j} - x_{n-l,j})^2}. \quad (1.4)$$

Мерой расстояния необязательно брать евклидово расстояние, можно брать любую другую, например, сумму абсолютных значений разницы составленных из наблюдений признаков, формула (1.5)

$$d_i = \sum_j C_j \sum_{l=0}^{L-1} |x_{i-l,j} - x_{n-l,j}|. \quad (1.5)$$

При поиске соседей можно найти всех соседей, которые находятся ближе, чем заданное расстояние $d_{\text{макс}}$, можно из них отобрать заданное количество ближайших. Однако просмотр всей последовательности может быть не желателен, тогда можно прекратить поиск соседей, когда их будет найдено требуемое количество.

Прогноз следующего значения ряда y_{n+1} получается на основе множества найденных соседей K как среднее значение следующих за соседом наблюдений, либо как взвешенное среднее, где веса обратно пропорциональны расстоянию, формула (1.6)

$$y_{n+1} = \frac{1}{|K|} \sum_{i \in K} y_{i+1} \quad \text{или} \quad y_{n+1} = \frac{\sum_{i \in K} y_{i+1}/d_i}{\sum_{i \in K} 1/d_i}. \quad (1.6)$$

Модификаций метода можно придумать большое множество. Изменять можно наборы признаков, т.е. что будет сравниваться, конструируя их всевозможными способами. Можно изменять способ расчета расстояния до соседа и как это расстояние потом будет сравниваться. Изменять можно метод получения прогноза, конструируя различные функции от следующих за соседом значений y_{i+1} . Иногда адаптируют технологии нечетких множеств для метода ближайших соседей [34].

Метод ближайших соседей относят к непараметрическим методам, так как в нем заранее не надо предполагать исходный вид модели, возвращающей прогнозное значение. Однако параметры у метода имеются, это ширина окна. Оптимальное значение ширины окна как правило подбирается на практике эмпирическим путем.

Описанный метод можно также назвать методом распознавания образов в имеющемся ряде данных. Прогнозное значение берется как среднее по следующим значениям за каждым из образов. Метод будет обладать неплохой точностью, если в исходном ряде данных проявляется тенденция, когда образы периодически повторяются, и при этом следующие за этими образами значения также имеют тенденцию повторяться.

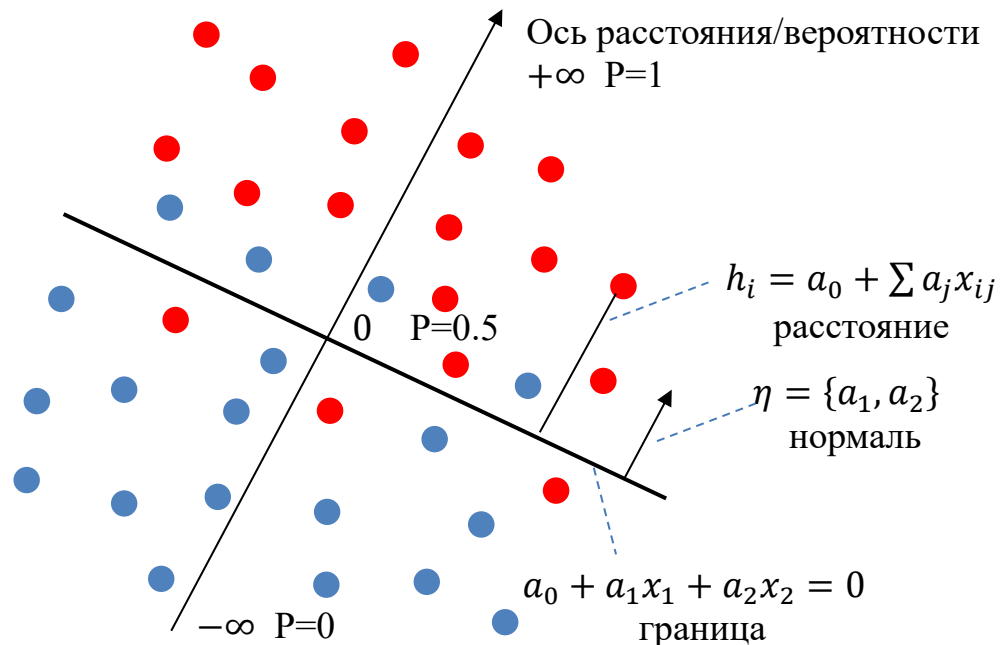
Метод логистической регрессии

В ранее описанном методе ближайших соседей поиск знакомых образов происходит среди данных все того же временного ряда, но зачастую могут наблюдаться сразу несколько значений разнообразных признаков, то есть могут быть известны значения сразу нескольких временных рядов. Редкое

событие может быть привязано к определенной комбинации значений этих наблюдаемых признаков. Другими словами, предполагается, что определенные значения признаков сигнализируют о наступлении (наблюдении) редкого события, в то время как другие значения признаков сигнализируют об отсутствии данного редкого события. Задача может заключаться в том, чтобы определить (классифицировать) какие комбинации значений наблюдаемых признаков соответствуют возникновению редкого события, а какие не соответствуют. Для этих целей разработаны специальные методы бинарной классификации. Один из таких методов – это метод логистической регрессии [35; 36; 37; 38].

Согласно исследованию истории возникновения логистической регрессии [39], осуществленной датским экономистом Жаном Соломоном Крамером, логистическая регрессия (логит модель) была предложена американским математиком, занимающимся статистикой, Джозефом Берксоном [40] в 1944 году, как более эффективная с точки зрения вычислений модификация пробит регрессии (пробит модели), предложенной американским биологом, занимающимся статистическими исследованиями, Чарльзом Игтнер Блиссом [41; 42] в 1934 году.

Идея логистической регрессии заключается в том, чтобы в многомерном пространстве признаков провести плоскость, которая бы отделяла наблюдения одного класса от наблюдений другого класса. Графическая интерпретация метода логистической регрессии в двумерном пространстве представлена на рисунке 1.8.



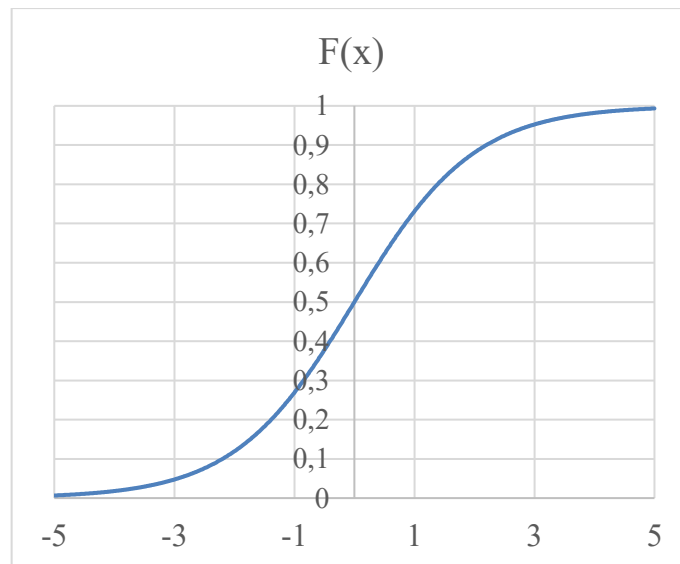
Источник: составлено автором как обобщение сути метода, подчерпнутой из разных открытых источников.

Рисунок 1.8 – Графическая интерпретация метода логистической регрессии на плоскости

Для того чтобы провести прямую в двумерном пространстве необходимо задать уравнение прямой $a_0 + a_1x_1 + a_2x_2 = 0$ (в n -мерном случае необходимо задать гиперплоскость $a_0 + a_1x_1 + \dots + a_nx_n = 0$). Другими словами, необходимо определить неизвестные коэффициенты a_0, a_1, a_2 . Причем вектор нормали к этой прямой определяется как раз через эти параметры $\eta = \{a_1, a_2\}$. Эта нормаль (как градиент) показывает на сколько растёт «расстояние» от границы, то есть на сколько нарушается уравнение прямой. Подставляя координаты x_1 и x_2 конкретного наблюдения в уравнение границы получаем «расстояние» до границы $h_i = a_0 + \sum a_j x_{ij}$ (чтобы получить геометрическое расстояние, надо h_i умножить на $1/\sqrt{a_1^2 + a_2^2}$). Далее это «расстояние» до границы конвертируется в вероятность принадлежности текущего наблюдения к одному из двух классов. Чем больше расстояние от границы, тем больше вероятность принадлежности к заданному классу. Наблюдения, лежащие на самой границе, с вероятностью 0.5 могут принадлежать как к одному классу, так и к другому. Для упрощения формул в вектор признаков добавляют еще один признак, который для каждого

наблюдения всегда принимает значение 1, например, для многомерного случая $X_i = (1, x_{i1}, x_{i2}, \dots, x_{in})$, тогда расстояние до границы h_i можно рассчитывать как скалярное произведение вектора неизвестных параметров $A = (a_0, a_1, \dots, a_n)$ и вектора X_i , т.е. $h_i = AX_i = \sum a_j x_{ij}$.

Для конвертации этого «расстояния» от границы в вероятность можно использовать различные функции, которые преобразовывали бы величины, принимающие значения от $-\infty$ до $+\infty$, в величины от 0 до 1. Для этих целей хорошо подходят функции распределения $F(x)$ случайных величин с бесконечной областью определения. В логистической регрессии используется логистическое распределение $F(x) = \frac{1}{1+e^{-(x-m)/s}}$, рисунок 1.9, с нулевым математическим ожиданием $m = 0$ и масштабом s , который связан с коэффициентами a_0, a_1, \dots, a_n уравнения границы (например, уравнение $1x_1 + 1x_2 = 1$ и $100x_1 + 100x_2 = 100$ задают одну ту же линию, но у второго расстояние h_i получается в 100 раз больше, т.е. коэффициенты a_j могут содержать масштабирующий множитель).



Источник: составлено автором по формуле логистического распределения.

Рисунок 1.9 – Функция распределения логистического закона распределения с нулевым математическим ожиданием m и единичным масштабом s

В результате вероятность принадлежности наблюдения к одному из классов можно рассчитать как, формула (1.7)

$$P(X_i) = P(h_i) = F(h_i) = \frac{1}{1 + e^{-h_i}} = \frac{1}{1 + e^{-\sum a_j x_{ij}}}. \quad (1.7)$$

Определение значений неизвестных коэффициентов $A = (a_0, a_1, \dots, a_n)$ (т.е. построение границы) происходит по обучающей выборке, содержащей уже классифицированные наблюдения. Для каждого наблюдения i , обладающего набором признаков $X_i = (1, x_{i1}, x_{i2}, \dots, x_{in})$, задано значение $y_i \in \{0, 1\}$ соответствия определенному классу, 1 – соответствует, 0 – не соответствует. Для наблюдений, соответствующих заданному классу, определяем вероятность этого соответствия $P(Y = 1|X_i) = P(X_i)$, для наблюдений, несоответствующих заданному классу, определяем вероятность этого несоответствия $P(Y = 0|X_i) = 1 - P(X_i)$. Можно составить универсальное выражение для вероятности правильной классификации наблюдения одновременно для двух классов, формула (1.8)

$$P(Y = y_i|X_i) = P(X_i)^{y_i}(1 - P(X_i))^{1-y_i} = F(h_i)^{y_i} (1 - F(h_i))^{1-y_i}, \quad (1.8)$$

где в случае, если $y_i = 1$, второе выражение возводится в 0 степень, а в случае, если $y_i = 0$, то первое выражение.

Так как распределение $F(x)$ симметричное, то $1 - F(x) = F(-x)$ и выражение упрощается, формула (1.9)

$$P(Y = y_i|X_i) = F(h_i)^{y_i} F(-h_i)^{1-y_i}. \quad (1.9)$$

Далее по методу максимального правдоподобия максимизируем произведение вероятности для всех наблюдений обучающей выборки $\prod_i P(Y = y_i|h_i) \rightarrow \max$, формула (1.10)

$$\prod_i F(h_i)^{y_i} F(-h_i)^{1-y_i} \rightarrow \max. \quad (1.10)$$

Логарифмируя, получаем, формула (1.11)

$$\sum_i y_i \ln(F(h_i)) + (1 - y_i) \ln(F(-h_i)) \rightarrow \max. \quad (1.11)$$

Использование логистического распределения $F(x) = \frac{1}{1 + e^{-x}}$ (сигмоиды), вместо нормального распределения, хорошо тем, что можно

значительно упростить полученное выражение. Также логистическое распределение хорошо тем, что для него достаточно легко найти обратную функцию $x = \text{Ln}(F(x)/1 - F(x))$.

Используя то, что $F(x) = (1 + e^{-x})^{-1}$, получаем формулу (1.12)

$$\sum_i -y_i \ln(1 + e^{-h_i}) - (1 - y_i) \ln(1 + e^{+h_i}) \rightarrow \max, \quad (1.12)$$

где максимизация отрицательных величин эквивалентна минимизации положительных, в результате имеем, формула (1.13)

$$\sum_i y_i \ln(1 + e^{-h_i}) + (1 - y_i) \ln(1 + e^{+h_i}) \rightarrow \min. \quad (1.13)$$

В этом выражении первое слагаемое обращается в ноль при $y_i = 0$, а второе при $y_i = 1$, для каждого наблюдения активно только одно из двух слагаемых. Причем для наблюдений $y_i = 1$ степень у экспоненты отрицательная, а для наблюдений $y_i = 0$ – положительная. В итоге окончательно получается следующая оптимизационная задача, формула (1.14)

$$\sum_i \ln(1 + e^{-I_i h_i}) = \sum_i \ln(1 + e^{-I_i \sum_j a_j x_{ij}}) \rightarrow \min, \quad (1.14)$$

где $I_i = \begin{cases} -1, & y_i = 0 \\ 1, & y_i = 1 \end{cases}$ (если бы исходный класс наблюдений $y_i \in \{0,1\}$ обозначался как $y_i \in \{-1,1\}$, то имели бы $\sum_i \ln(1 + e^{-y_i \sum_j a_j x_{ij}}) \rightarrow \min$).

Полученная оптимизационная задача дальше не упрощается, из-за своей структуры она является нелинейной, поэтому нахождение искомым параметров $A = (a_0, a_1, \dots, a_n)$ производится средствами нелинейного программирования, чаще всего с помощью градиентных методов. После определения коэффициентов уравнения границы $A = (a_0, a_1, \dots, a_n)$ (успешного обучения модели логистической регрессии) для следующих еще неклассифицированных наблюдений по имеющимся признакам можно определить вероятность принадлежности к одному из двух классов $P(X_i) = (1 + e^{-AX_i})^{-1}$. Для задач классификации достаточно проверять больше ли эта вероятность чем 0.5 или нет. Зачастую для принятия и реализации важных

дорогостоящих решений недостаточно вероятности прогноза лишь немногим превышающей половину, часто решение принимается при вероятностях только больше, чем 0.8 или даже 0.9.

Пару слов надо сказать о функции распределения $F(x)$, служащей для преобразования расстояния от границы в вероятность. Если это функция распределения, то для какой случайной величины? Если мы возьмем производную от функции распределения, то естественно получим плотность вероятности логистического распределения, которая будет иметь симметричную колоколообразную форму с центром около нуля. То есть у нас относительно границы разбросана случайная величина. Этой случайной величине можно дать интерпретацию. Если при отклонении от границы в одну сторону у нас растет вероятность принадлежности наблюдения к одному из классов, то, следовательно, при таком отклонении уменьшается вероятность непринадлежности этого наблюдения к тому же классу. Если же мы отклоняемся в другую сторону от границы, то уменьшается вероятность принадлежности к выбранному классу, однако в то же время растет вероятность принадлежности к другому классу, а когда растет вероятность принадлежности к другому классу, уменьшается вероятность непринадлежности к этому другому классу. Таким образом, относительно границы разбросана случайная величина, символизирующая непринадлежность к любому классу, т.е. ошибка классификации. Эта ошибка классификации разбросана в соответствии с выбранным законом распределения: логистическое распределение в логистической регрессии или нормальное распределение в пробит регрессии. Такая интерпретация логистической регрессии через ошибку классификации носит название модели скрытой переменной (латентной переменной).

В том случае, если разброс ошибки классификации уменьшается, то функция сигмоиды $F(x)$ сужается, а при нулевом разбросе сигмоида вырождается в ступенчатую линию (плотность распределения становится импульсной). Такое наблюдается в случае, когда наблюдения можно линейно

разделить так, чтобы все наблюдения одного класса гарантированно лежали с одной стороны от границы, а другого класса с другой стороны. Это приводит к тому, что параметры распределения становятся очень большими (коэффициент масштаба s очень маленький), а при нулевом разбросе коэффициенты уходят в бесконечность. Градиентный метод работать не будет, так как производная будет не существовать. В этом случае прибегают к регуляризации, когда в оптимизационную задачу добавляют слагаемое $\sum_j |a_j|$ или $\sqrt{\sum_j a_j^2}$, что заставляет алгоритм минимизации уменьшать также сами значения параметров, формула (1.15)

$$\sum_i \ln(1 + e^{-I_i \sum_j a_j x_{ij}}) + \lambda \sqrt{\sum_j a_j^2} \rightarrow \min, \quad (1.15)$$

где λ – коэффициент, показывающий насколько много внимания надо уделять растущим параметрам по сравнению с основной задачей.

Параметр λ задается исследователем вручную еще до запуска алгоритма, но существуют и автоматические методы его подбора.

Из-за того, что граница имеет линейный вид (прямая или гиперплоскость), логистическую регрессию относят к методам линейной классификации. Однако граница, разделяющая наблюдения на два класса, может быть нелинейной, например, когда в двумерном пространстве граница принимает вид круга, эллипса или еще более сложной формы. Например, для эллипса нужно определить параметры уравнения $(x_1 - x_1^0)^2/a + (x_2 - x_2^0)^2/b = 1$. Расстояние до границы уже не будет линейной функцией от входных признаков, но можно прибегнуть к линеаризации. Раскрыв скобки уравнения, будем иметь уравнение вида $a_0 + a_{11}x_1 + a_{12}x_1^2 + a_{21}x_2 + a_{22}x_2^2 = 0$. Если теперь вместо входных признаков $X_i = (1, x_1, x_2)$ использовать $X_i = (1, x_1, x_1^2, x_2, x_2^2)$, то получится по-прежнему использовать линейную форму модели для определения неизвестных коэффициентов. Однако такая линеаризация приводит к мультиколлинеарности, когда

объясняющие переменные (столбцы) становятся линейно зависимыми, что сильно влияет на погрешность определения искомым параметров. Если это влияние становится очень сильным, то логистическая регрессия не подходит, должны использоваться более сложные методы нелинейной классификации. Для нелинейного случая также надо отметить то, что расстояние от границы не обязательно будет всегда изменяться от $-\infty$ до $+\infty$. Например, для границы в виде нашего эллипса единичная вероятность может соответствовать точке, лежащей в середине этой области, т.е. на некотором конечном расстоянии от границы. Тогда вместо логистического распределения, имеющего бесконечную область определения, следует выбирать другие распределения с фиксированной с одной стороны областью определения.

В случае, если в обучающей выборке, по которой подбираются неизвестные параметры уравнения границы, наблюдений одного класса значительно больше, чем наблюдений другого класса, то есть имеет место несбалансированность выборки, то определяемые коэффициенты будут смещены. В этом случае граница будет расширяться в сторону от подавляющего класса к меньшему классу. Например, если граница имеет замкнутую область, когда мы увеличиваем объем обучающей выборки дублируя наблюдения, принадлежащие подавляющему классу, то визуально это будет приводить к тому, что область с редкими событиями, отделенная границей, сужается. Изменение положения границы в свою очередь приведет к тому, что вероятность классификации наблюдений из меньшего класса будет занижена. А это в свою очередь может иметь серьезные последствия, так как эти редкие события могут иметь очень высокое влияние, и ошибка их классификации может стоить очень дорого.

Существуют способы для работы с несбалансированными выборками. В работе [43] произведен достаточно обширный обзор существующих методов классификации в условии несбалансированной выборки. Все методы авторы делят на две группы подходов, внутренние и внешние методы. В группе внутренних методов происходит изменение самого алгоритма классификации.

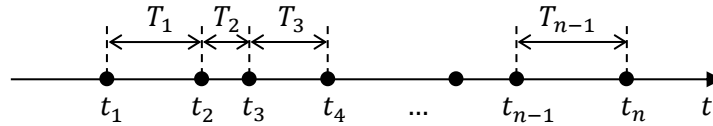
Одним из таких подходов является введение весов для ошибок классификации пропорционально стоимости этой ошибки (редким событиям ставится больший вес). Другие внутренние методы по-своему учитывают несбалансированность выборок. В отличие от них группа внешних методов исправляет несбалансированность самой обучающей выборки. Такие методы изменяют количество наблюдений в выборке, либо дублированием наблюдений, принадлежащих к меньшему классу, либо удалением части наблюдений подавляющего класса наблюдений. Какие из наблюдений удалять или дублировать можно вбирать либо случайно, либо по какому-то правилу. Конечно же можно комбинировать внутренние и внешние методы при работе с несбалансированными выборками, подробнее об этом в [43].

Метод логистической регрессии самый популярный, но не единственный метод классификации. Существуют другие методы, например, на основе нейронных сетей [44; 45]. При использовании методов классификации можно по наблюдаемым признакам определить какие события относятся к редким, а какие нет. В то же время наблюдая за динамикой изменения некоторых параметров, например, за значениями временного ряда, когда событий как таковых нет, можно зафиксировать наступление специфических условий, относящихся к редким событиям, и тем самым сформировать редкое событие.

Потоки событий

Помимо анализа редких событий с точки зрения анализа временных рядов или задач классификации существует еще один подход. В теории случайных процессов редкие события представляются в виде потоков событий [46]. Поток событий представляет собой последовательность моментов времени возникновения событий t_i , графически это выглядит как нанесение точек на ось времени, рисунок 1.10. Отличительной особенностью является то, что события появляются одно за другим через случайные периоды времени T_i , причем вероятность появления одновременно двух событий пренебрежимо

мала (ординарность). Если же такие события могут возникать пачками, например, машина поставяет сразу партию готовой продукции, то поток будет неординарным.



Источник: составлено автором по материалам [46].

Рисунок 1.10 - Поток событий

Интервалы между событиями T_i предполагаются независимыми друг от друга, но соответствующие одному закону распределения. Если закон распределения стационарен, т.е. его характеристики не зависят от времени, то такой поток называют потоком Пальма. Также поток может быть нестационарным и его характеристики могут меняться. В случае, если интервалы времени между событиями T_i соответствуют экспоненциальному закону распределения, то такой поток называют потоком Пуассона (стационарный поток Пуассона является частным случаем потока Пальма). В пуассоновском потоке количество событий, выпавших на один интервал времени, полностью не зависит от того, сколько событий выпало на любой другой не пересекающийся период времени. Данное свойство называют отсутствием последействия. Для потоков, обладающих отсутствием последействия, справедливо то, что появление следующего события полностью не зависит от того, сколько времени прошло от предыдущего события. Поток Пальма называют потоком с ограниченным последействием, так как вероятность появления следующего события зависит от того, сколько времени прошло спустя предыдущее событие.

Основной характеристикой потоков событий является интенсивность $\lambda(t)$ ¹⁾, которая показывает среднее количество событий, возникающих за единицу времени в момент времени t . Для стационарного потока, чьи

¹⁾ $\lambda(t) = \lim_{\Delta t \rightarrow 0} E[X(t, t + \Delta t)]/\Delta t$, где $X(t, t + \Delta t)$ количество событий, выпавших на интервал от t до $t + \Delta t$, а $E(\cdot)$ – математическое ожидание случайной величины.

характеристики не меняются, интенсивность потока постоянна $\lambda(t) = \lambda = \text{const}$ (у пуассоновского потока интенсивность является единственной характеристикой, другие характеристики отсутствуют). Для пуассоновского потока эта же интенсивность λ стоит как степень затухания вероятности в формуле экспоненциального закона распределения для интервалов времени между поступлениями $f(t) = \lambda e^{-\lambda t}$, $F(t) = 1 - e^{-\lambda t}$.

Так как интервалы времени в потоках событий являются случайными величинами, то количество событий, выпадающих на заданном периоде времени, также будет случайной величиной. Среднее количество событий $X(t, t + \Delta t)$, выпавших на интервал от t до $t + \Delta t$, для стационарных потоков определяется как $E[X(t, t + \Delta t)] = \lambda \Delta t$, а для нестационарных как $E[X(t, t + \Delta t)] = \int_t^{t+\Delta t} \lambda(t) dt$. Для пуассоновских потоков количество событий X , выпадающих на заданном интервале времени, соответствует распределению Пуассона $P(X = x) = a^x e^{-a} / x!$, где a – ожидаемое количество событий на этом интервале ($a = \lambda \Delta t$ или $a = \int_t^{t+\Delta t} \lambda(t) dt$), для единичного интервала времени и постоянной интенсивности $a = \lambda$. Распределение Пуассона является предельным случаем Биномиального распределения¹⁾, когда вероятность успеха стремится к нулю, а количество испытаний стремится к бесконечности. Распределение Пуассона называют законом редких событий, а пуассоновский поток называют потоком редких событий, так как события возникают по одному через ненулевые интервалы времени. Как было сказано в предыдущем параграфе, частота событий является относительной характеристикой, иногда в практических работах потоки событий, где время между событиями измеряется микросекундами, по-прежнему называют потоками редких событий.

В отличие от пуассоновских потоков для потоков Пальма вероятность появления событий и, следовательно, их количество на заданном периоде

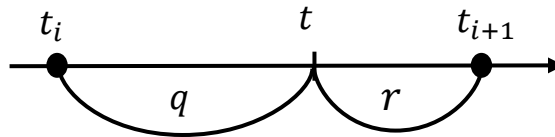
¹⁾ Определяет вероятность появления m успехов в серии из n испытаний $P_m = C_n^m p^m (1 - p)^{n-m}$, где p – вероятность успеха, а $C_n^m = n! / m! (n - m)!$ – количество сочетаний.

времени зависит от времени, прошедшего после предыдущего события. Время до следующего события будет случайной величиной, закон распределения которой рассчитывается следующим образом, формула (1.16)

$$f_R(r|q) = \frac{f(q+r)}{1-F(q)}, \quad (1.16)$$

где $f_R(r|q)$ – плотность условного закона распределения оставшегося времени r до следующего события при известном значении времени q , прошедшего спустя предыдущее событие;

$f(t)$ и $F(t)$ – плотность и (интегральный) закон распределения для самих интервалов времени между событиями, рисунок 1.11.



Источник: составлено автором по материалам [46].

Рисунок 1.11 – Оставшееся время до следующего события r зависит от прошедшего времени после предыдущего события q

Обычные пуассоновские процессы часто используются для прогноза спроса на заданный период времени и определения размера запаса (в условиях прерывистого/редкого спроса). Происходит это следующим способом. В начале определяется интенсивность пуассоновского процесса поступления $\lambda(t)$. Делать это можно разными способами. В простейшем случае интенсивность определяется как суммарный спрос за несколько интервалов, поделенный на количество таких интервалов [47; 48; 49], т.е. как средний спрос за интервал времени. Причем такая интенсивность получается стационарной. Иногда используют скользящее среднее или ядерные функции¹⁾ для определения нестационарной интенсивности [50; 51; 52].

¹⁾ На непрерывной оси времени в момент времени t определяется сглаженное значение $\hat{y}(t) = \frac{\sum_{i=1}^n y_i \cdot K\left(\frac{t_i-t}{h}\right)}{\sum_{i=1}^n K\left(\frac{t_i-t}{h}\right)}$, где y_i - наблюдение в момент времени t_i , h - параметр сглаживания (радиус сглаживания), $K(x)$ одна из ядерных функций, задающая веса наблюдениям. В качестве ядерной функции может использоваться стандартное нормальное распределение $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Предложены Надарай и Уотсоном в 1964 году [147; 148].

Далее, зная значение интенсивности λ (в нестационарном случае надо сделать прогноз функции интенсивности $\lambda(t)$ на интересующий исследователя период времени), по формуле вероятности для распределения Пуассона $P(x) = a^x e^{-a} / x!$, где a – ожидаемое количество событий на интересующем периоде времени, рассчитывают вероятности возникновения определенного количества событий $P(0), P(1), P(2), \dots$. Тогда размер запаса устанавливается на том уровне, чтобы обеспечить с заданной вероятностью нужное количество товара (спрос). Для этого подсчитывают сумму вероятностей $P(0) + P(1) + P(2) + \dots$ до тех пор, пока не будет превышена доверительная вероятность, после чего запас устанавливается на уровне последнего значения, чья вероятность была добавлена в эту сумму последней (запас $S = k: \sum_k P(k) > P_{\text{дов}}$).

Это был простейший способ использования Пуассоновских потоков для прогноза количества будущих событий (спроса). Причем представление о самих дискретных событиях как о потоке событий, по сути, тут не используется. Используются лишь характеристики Пуассоновского закона распределения, хотя Пуассоновский закон описывает как раз поток дискретных событий.

Иногда используют потоки с интервалами времени между событиями, соответствующими закону распределения Эрланга, которое показывает, как распределена сумма из k экспоненциальных случайных величин. Потоки, в которых время между событиями соответствует распределению Эрланга, также называют прореженным пуассоновскими потоками, когда из потока берется только каждое k -тое событие. У закона распределения Эрланга уже два параметра, это степень затухания λ плотности вероятности суммируемых экспоненциальных величин (интенсивность λ исходного потока) и количество самих суммируемых величин k (при $k = 1$ распределение совпадает с экспоненциальным, параметр k задает порядок распределения). Средний интервал следования событий в потоке Эрланга составляет k/λ , а количество

событий за единицу времени λ/k . В работе [53], посвященной использованию потоков Эрланга для прогнозирования редкого спроса, средний спрос за единицу времени определяется как $E(y_t) = \lambda\mu/k$, где μ – среднее значение для ненулевого спроса (величины μ и λ/k определяются из данных), также в этой работе определяется дисперсия $Var(y_t)$, но уже более сложным образом. Рассчитав $E(y_t)$ и $Var(y_t)$, можно определить величину запаса, необходимого для удовлетворения спроса на необходимом уровне. В работе [54] также используется распределение Эрланга, но при этом утверждается, что так как оценка количества событий за единицу времени λ/k определяется через оценку интервалов времени между событиями, оценка спроса за единицу времени $\lambda\mu/k$ получается смещенной, и надо вводить корректирующий множитель $\frac{k-1}{k}$ (при небольших значениях порядка распределения Эрланга k эта корректировка может быть значительной).

Авторы этих работ утверждают, что потоки Эрланга более естественны, так как спрос у клиентов может возникать под действием других потоков, что в результате похоже на прореживание потоков, когда k событий одного потока порождают одно события другого потока. В описанном выше подходе поток событий Эрланга используется только для перехода к закону распределения Эрланга и дальнейшему использованию лишь формул закона распределения. Сама же идея рассмотрения потока событий с точки зрения того, что эти события являются дискретными и по-своему уникальными, теряется в обобщающих формулах закона распределения. Использование формул законов распределения для оставшегося времени до следующего события из теории потоков Пальма не встречается в практических работах (в той же логистике), посвященных анализу и предсказыванию редких событий.

Селективные методы

Иногда используются селективные методы [55], которые в зависимости от входных данных переключают модель работы с редкими событиями на одну из нескольких возможных, с выбором самого метода по значению

ошибки прогноза на предыдущем шаге. В упомянутой работе осуществляются манипуляции над все теми же временными рядами, содержащими большое количество нулевых значений. Сам подход предполагает производить прогнозирование одновременно по нескольким методам, но опираться на тот метод, который на прошлом шаге имел наименьшую ошибку прогноза. Причем переключаемые методы могут быть одинаковыми, но иметь различающиеся параметры. То есть среди методов могут быть несколько методов Кростона, с разным коэффициентом сглаживания, несколько методов Виллемейна с разной шириной окна, несколько видов классификационных методов с разным значением коэффициента регуляризации, несколько методов, основанных на потоках событий Пуассона или Эрланга разного порядка. Данный подход по-своему является универсальным, его можно описать следующей фразой: «Если не знаешь какой метод использовать, используй их все сразу» - однако же за это приходится платить большим объемом вычислений, зачастую выполняемых в холостую, а также низким быстродействием.

Эвристические методы

Конечно же исследователи могут использовать собственный опыт и определенные знания, которые свойственны конкретной прикладной области при определенных условиях. Как таковых общих идей и методов тут выявить нельзя. Опыт и знания некоторых фирм и компаний являются ценностью пока сохраняются в секрете, поэтому вполне вероятно некоторые успешные методы могут умышленно держаться в тайне и не публиковаться, чтобы сохранить преимущество перед конкурентами.

1.3 Анализ недостатков существующих методов анализа редких событий

Перечисленные методы конечно же можно применять в той или иной степени для анализа редких событий, однако они обладают существенными недостатками.

На мой взгляд, метод Кростона (который, прогнозирует следующее ненулевое значение, как результат экспоненциального сглаживания ненулевых значений, через ожидаемое количество интервалов между ненулевыми значениями, сглаженными экспоненциально) очень абстрактен. Предположение о возникновении ненулевого спроса по распределению Бернулли является немотивированным. Во время экспоненциального сглаживания не происходит никакого поиска закономерности, которые безусловно должны иметься, неправильно предполагать, что ненулевое значение появится через ожидаемое количество нулевых значений. Величина ненулевого спроса экспоненциально сглаживается, причем совершенно независимо от интервалов между этими ненулевыми значениями. На практике величина ненулевого спроса просто обязана быть связана с интервалами между ними. Метод Кростона, по сути, прогнозирует появление ненулевого спроса через ожидаемое количество интервалов времени, причем это ожидаемое значение пересчитывается каждый раз при возникновении ненулевого спроса всего лишь с помощью экспоненциального сглаживания. Никакая динамика не учитывается. Данный метод может работать только тогда, когда события стремятся возникать регулярно, через одинаковые периоды времени. Если эти периоды времени меняются по некоторому сложному правилу, то метод, сглаживая их экспоненциально, не заметит это правило. Аналогично и с самой величиной ненулевого значения, если она меняется по некоторому правилу, то экспоненциальное сглаживание также не позволит выявить это правило. Вернее, оно способно что-то дать, если

величина ненулевого значения не меняется или меняется монотонно (в последнем случае будет временной лаг).

По моему мнению, причина основного недостатка метода Кростона кроется в том, что метод работает с редкими событиями, ограничиваясь лишь представлением этих событий в виде простейшей сухой статистической информации, когда любая неопределенность заменяется лишь вероятностью без попытки использования более сложных моделей на основе имеющихся знаний.

Метод Виллемейна (бутстреппинга) частично освещает мглу неопределенности метода Кростона, когда использует схему марковского процесса с двумя состояниями. Однако затем переходы между состояниями все равно определяются сухой вероятностью, определяемой из статистики. Вновь все покрывает мгла неопределенности. Также из описания метода сам марковский процесс предполагается стационарным, т.е. вероятности переходов предполагаются неизменными во времени. Опять возможная динамика и закономерность ее изменения будет закопана в одной общей куче стационарной вероятности. После определения вероятности переходов и генерации самого процесса с помощью случайных чисел и равномерного закона распределения, случайным образом выбираются значения ненулевого спроса из имеющейся выборки. Этот выбор я также считаю наивным. Исходное распределение значений во временном ряду (выборке) может быть неслучайным или значения будут коррелированными, а слепой отбор этих значений с помощью равномерного распределения не сможет этого учесть. Например, если в наблюдаемые периоды времени спрос менялся по некоторому закону, то случайная выборка значений за эти периоды только скроет возможную закономерность. Нет никаких оснований считать, что значения ненулевого спроса возникают абсолютно независимо и случайно, особенно независимо от интервалов следования (интервалы следования должны зависеть от величины ненулевого спроса). Также случайная выборка значений из прошлых данных не способна учесть того, что в этих отобранных

данных отсутствует, но могло проглядываться как тенденция в прошлых данных. Например, если имеется монотонный рост значений, то естественно предположить, что и будущие значения будут завышенными, а случайная выборка из прошлых значений отберет только имеющиеся значения, которые не успели возрасти (случайный разброс, который добавляет Вилемейн, хоть и позволяет учесть другие условия, которых не было в исходных данных, но не позволяет учесть возможные закономерности и долгосрочную динамику).

Хоть метод Вилемейна и пытается искусственно размножить выборку, основным недостатком на мой взгляд является то, что он фиксирует сложившуюся ситуацию (исходный набор данных) как неизменную статичную картину, а эта картина могла быть динамичной. Также считаю, что при генерировании событий (моделировании марковского процесса) возникновение событий лишь по вероятности является слишком упрощенным.

Методы классификации, такие как метод ближайших соседей или метод логистической регрессии и др., очень популярны для анализа редких событий, но они также содержат свои недостатки. Во-первых, большое влияние может оказывать дискретизация времени, с которой происходят наблюдения. Редкое событие может быть пропущено (если оно не фиксируется своим собственным моментом времени), если в моменты наблюдений за значениями некоторых характеристик не попали краткосрочные изменения этих характеристик, которые произошли с большей частотой, чем частота наблюдений. Например, если проверять звенит ли будильник раз в час, заглядывая на секунду в комнату, где он стоит, то скорее всего мы упустим момент, когда он прозвенел. Проверять надо каждую минуту, а не каждый час. Также измерение каждого сопутствующего признака может наблюдаться со своей собственной частотой, что усложнит анализ. Например, если частота наблюдений одного признака в десять раз меньше, чем другого признака, то момент возникновения редкого события будет не раньше, чем обновится значение самого медленно обновляемого признака. Если события возникают по сложной схеме, например, когда один класс событий

сосредоточен не в одной области пространства признаков, а сразу в нескольких несвязанных областях, могут потребоваться сложные методы нелинейной классификации.

Методы классификации могут предсказывать моменты появления будущих событий, если в качестве признаков выбираются лаговые переменные, наблюдаемые несколько периодов назад. Однако методы классификации опять делают как бы неподвижный снимок текущей ситуации и ищут эту же картину среди признаков следующих наблюдений. Осуществляя оценку параметров модели классификации из имеющихся данных, метод не определяет, как должны меняться эти параметры со временем, чтобы классификация оставалась корректной. То есть опять отсутствует учет возможной динамики и выявление закономерностей изменения параметров нестационарной изменяющейся картины.

Существенным же недостатком является то, что некоторым редким событиям не предшествуют никакие наблюдаемые характерные изменения признаков, признаки изменяются не до возникновения события, а вовремя или после наступления его. В этом случае простое наблюдение за признаками, даже если они рассматриваются как лаговые переменные у еще не наступившего события, не поможет в предсказании этого редкого события. Такая ситуация может возникать достаточно часто. Также хочется обратить внимание, что задача классификации и распознавания немного отличается от более общей задачи анализа данных. В таких задачах мы определяем наборы признаков (характерный образ), которые свидетельствуют о возникновении редкого события, после чего ищем знакомые образы в новых данных, в то время как в задаче анализа данных мы пытаемся выявить сложные закономерности в изменяющихся динамично наборах параметров. То есть опять задача классификации является статичной, когда анализ данных не ограничивается статичными моделями. Еще хоть методы классификации и могут рассказать о вероятности возникновения события в будущий период при заданных значениях признаков, однако они не заглядывают внутрь самого

процесса образования события, вероятность определяется лишь статистически.

С помощью Пуассоновских потоков можно найти вероятность появления заданного количества событий за некоторое время, но при этом нельзя определить моменты времени возникновения событий. Конечно же можно предположить, что событие возникнет через ожидаемый интервал времени (равный $1/\text{интенсивность}$), однако особенность Пуассоновских потоков в том, что для них наблюдается отсутствие последействия. Из-за отсутствия последействия вероятность наступления событий никак не зависит от того, сколько времени прошло от прошлого события, а это как раз говорит о том, что бессмысленно ожидать следующее событие, откладывая любой интервал времени. С помощью потоков Пальма уже можно определить моменты времени наступления следующих событий, отложив математическое ожидание интервалов между событиями от предыдущего события. Однако так как закон распределения интервалов времени в потоках Пальма стационарен, все будущие события будут идти с одинаковым шагом, что не всегда соответствует действительности. Использование нестационарных потоков с ограниченным последействием, то есть таких, где время между событиями соответствует нестационарному не экспоненциальному закону, для прогнозирования будущих событий не встречается в литературе. Кроме того, определение параметров для нестационарных потоков может быть непростой задачей. Например, когда предполагается, что события происходят через интервалы времени с нормальным законом распределения, лишь из имеющихся данных о произошедших событиях надо определить нестационарные значения как математического ожидания, так и дисперсии для интервалов между событиями (наблюдаемое изменение интенсивности потока может быть связано с изменением дисперсии, а не с изменением математического ожидания).

Возвращаясь к задачам анализа редких событий, можем сказать, что метод Кростона, Виллемейна и элементы теории потоков могут быть

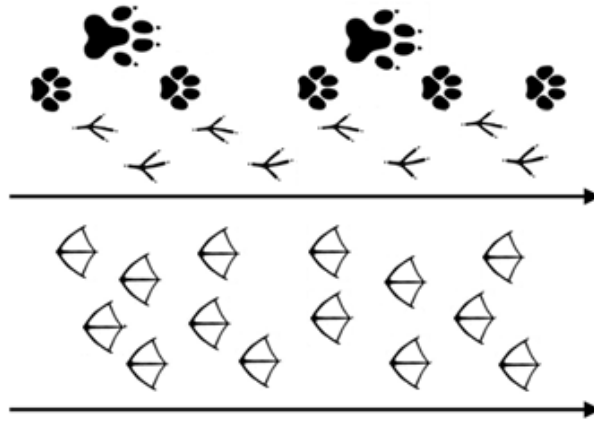
использованы для прогнозирования количества событий на заданном периоде времени, причем в научных исследованиях отдают предпочтение методу Виллемейна. С помощью теории потоков можно определить вероятность возникновения заданного количества событий в заданном периоде времени. Методы классификации могут определить вероятность наступления только одного события, но только с точки зрения вероятности правильной классификации (при предположении, что собранная по другим событиям статистическая информация распространяется на это новое событие). Однако все рассмотренные методы не способны предсказывать сами дискретные моменты времени наступления событий (методы могут предсказывать только регулярные события, которые следуют через равные периоды времени). Вместе с тем из-за неспособности выявлять закономерности данные методы неприменимы для прогнозирования значений, которые несут эти события. Другие методы, способные прогнозировать моменты наступления событий, не встречаются. Поэтому разработка новых методов анализа редких событий является актуальной задачей.

Однако, прежде всего хочется обратить внимание на один существенный недостаток у существующих методов анализа редких событий. В них совершенно не рассматривается механизм образования событий, предполагается, что события возникают сами по себе без какой-либо причины, как данность. Или может причина и есть, но она полностью игнорируется в существующих математических методах. В данных методах, которые рассматривают редкие события как поток случайных событий или анализируют данные редких событий как временной ряд, предполагается, что интенсивность событий или их количество на интервалах времени зависит либо от времени, либо от внешних факторов. Конечно, такая связь есть, но она не появляется из ниоткуда. Такая связь на самом деле возникает из-за того, что время или внешние факторы влияют на элементы механизма образования событий. Данный механизм в зависимости от внешних факторов начинает генерировать события либо реже, либо чаще, из-за чего изменяется их

интенсивность (как у потока событий) или появляется тренд у количества событий за интервалы времени (как у временного ряда). Игнорируя механизм образования событий, мы лишаемся возможности проведения более детального анализа и осуществления более точного прогнозирования. Если мы хотим научиться определять, когда могут наступать события в будущем, мы должны обратить внимание именно на внутренние процессы, которые приводят к возникновению этих событий.

Трудность анализа редких событий может быть вызвана следующим. Когда из данных редких событий, изначально представленных дискретно с подробной информацией по каждому событию, конструируется временной ряд, то есть происходит простой подсчет количества событий на интервалах времени, теряется практически вся информация о событиях. Аналогично происходит в момент объединения событий в один общий поток. Например, при таком агрегировании в один поток теряется информация, связанная с источником ¹⁾, где образовалось событие, а ведь механизм образования событий в каждом источнике может быть разным (пример, покупки совершаются разными покупателями, причины покупки у которых совсем разные). Дальнейший анализ такого смешанного потока событий вызовет много неточностей. Анализ агрегированного потока можно сравнить с картиной, когда следопыт, изучая следы животных на тропе, вместо того, чтобы различать следы от разных животных, начинает считать, что все следы одинаковые, рисунок 1.12.

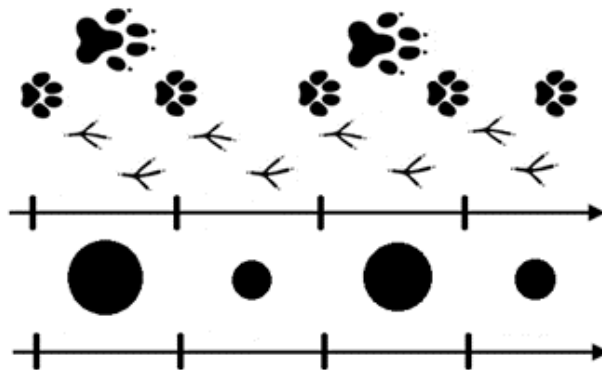
¹⁾ Под источниками события понимаются некоторые объекты или системы, в которых происходят какие-то процессы, приводящие к возникновению этих событий.



Источник: составлено автором.

Рисунок 1.12 – При получении агрегированного потока событий источники событий не различаются

Если же из данных событий строится временной ряд, то дополнительно теряются моменты времени возникновения каждого события. В этом случае анализ редких событий по временным рядам можно сравнить с картиной, когда следопыт начинает суммировать площадь следов на определенных участках, а затем пытаться выявить закономерность в последовательности получившихся площадей, рисунок 1.13.



Источник: составлено автором.

Рисунок 1.13 - Агрегирование данных по интервалам эквивалентно определению суммарной площади следов

Если нам встретится такой следопыт и нам доведется наблюдать его работу, то мы очень сильно удивимся его поведению и скажем ему, что следы надо анализировать отдельно. Аналогично мы можем утверждать, что агрегирование данных редких событий в один поток или во временной ряд также неправильно. В современную эпоху цифровых технологий обмена и

хранения данных это представляется крайне расточительным, теряется очень много ценной информации.

Тем не менее несправедливо утверждать, что все исследователи, занимающиеся анализом агрегированных данных такие же странные как наш следопыт. Дело в том, что исследователи анализируют те данные, которые им предоставлены, и выбор метода анализа данных зависит от того, как изначально представлены эти данные. Начальное представление данных зависит от того, какие методы анализа данных известны собирающему данные исследователю. *Именно знание методов дальнейшего анализа данных является основополагающей отправной точкой ко всему исследованию.* Если бы все исследователи заранее знали наилучший метод анализа данных, то и исходный набор данных был бы сформирован и представлен соответствующим образом. Однако, очень часто получается так, что сбором данных занимается один исследователь, а анализом данных занимается уже другой исследователь, причем первый не знает, как второй будет их анализировать. Из-за исходного представления данных первым исследователем, второй исследователь становится ограниченным в выборе метода анализа. Получается, что именно на первого исследователя ложится груз ответственности за первичное формирование набора данных и дальнейший выбор метода анализа этих данных, которые он же представил и структурировал определенным образом.

Есть несколько способов решения этой проблемы. Первый, собирать данные должен тот исследователь, который их же и будет анализировать. Второй способ, оба исследователя должны быть на связи друг с другом и понимать какие методы исследования будут использоваться. Однако эти два способа возможны только в том случае, когда само исследование запланировано раньше, чем этап сбора данных, что на практике не всегда происходит (обычно сперва собираются данные, а затем спустя некоторое время исследователи подключаются к их анализу). Третий, наверное, самый универсальный, собирать данные в первоначальном виде, чтобы никакая

информация не терялась, то есть никак не агрегировать их. Именно третий способ следует применять, тогда никакая информация не потеряется. В нынешнюю цифровую эпоху такой сбор данных не должен быть затруднительным, проблем с хранением данных не должно возникнуть, существующие хранилища данных весьма вместительны и доступны.

Выводы по главе 1

1) Проведено исследование существующих научных работ, в которых используется термин «редкое событие». Данное понятие применяется в задачах: классификации резких изменений во временных рядах; классификации наблюдений по наблюдаемым признакам; определения вероятности возникновения события за определенный период времени; определения вероятности выхода за границы допустимых условий; обнаружения выбросов и появление нового класса в задачах кластеризации. Делается вывод, что в существующих методах решения таких задач никак не используется информация о способах появления самих этих событий, то есть о процессах образования событий.

2) Сделан вывод, что наибольшее значение для экономики составляет выявление закономерностей и прогнозирование появления событий, что является не простой задачей.

3) Произведен обзор существующих методов, применяемых для анализа и прогнозирования редких событий в экономике. Делается вывод, что существующие методы ограничиваются лишь статистическими методами, в том смысле, что анализируют лишь появление или не появление событий, но не способ появления событий. Вместе с тем получающиеся модели оказываются статическими, хотя образование событий может происходить в динамически изменяющихся условиях. Делается вывод о необходимости обратить внимание на внутренние процессы, приводящие к образованию событий.

4) Трудности при анализе редких событий могут быть связаны с первичной обработкой данных. Делается вывод, что необходимо отказаться от агрегирования информации при фиксации событий, так как при агрегировании теряется индивидуальная информация (надо уметь отличать в каких источниках образованы события).

Глава 2

Емкостный метод анализа и прогнозирования редких событий

Ранее при рассмотрении существующих методов анализа редких событий мы выявили существенные недостатки, которые заключаются в том, что в этих методах никак не анализируется и не задействуется механизм образования событий. В них появление или не появление следующего события определяется лишь вероятностью по схеме испытания Бернулли, когда для фиксированного интервала времени или для интервала произвольной длительности определяется вероятность появления события либо только из имеющейся статистики ранее возникших событий, либо из учета сопутствующих этим событиям внешних признаков. Другим недостатком существующих методов является неспособность учитывать динамику изменения характеристик, связанных с образованием событий. На самом деле у всех событий, очевидно, есть причины их возникновения, причем эти «причины» могут динамически изменяться. Материалы этой главы были опубликованы¹⁾ в следующих работах автора [56; 57; 58; 59; 60; 61; 62; 63; 64].

2.1 Формулирование методики для исследования редких событий

Выявленные недостатки многих существующих методов в основном связаны с тем, что исходные данные предварительно агрегируются определенным образом с потерей большей части полезной информации. Потерянную информацию в то же время пытаются компенсировать статистическими механизмами, утверждая, что наблюдаемые различия в признаках происходят под действием некоторой случайности, которая в одном наблюдении приводит к отклонению в одну сторону, а в другом наблюдении в другую сторону. Причем эта неопределенность в дальнейшем не

¹⁾ Корректнее сказать, что данная глава была составлена по результатам работы на протяжении ряда лет, которые опубликованы в виде научных статей и тезисов конференций.

раскрывается, оставаясь как правило лишь идеализированной абсолютно случайной величиной.

И тут задаемся вопросом, разве отклонения в наблюдениях являются случайными, или почему процесс возникновения события представляется случайным процессом? Неужели совсем нет никакой информации о том, как возникают эти события? Откуда взялась эта случайность? Что такое эта случайность, если случайность — это лишь мера неопределенности или отсутствия знания. Добавив некоторые знания и представления, можно избавиться от большей степени существующей неопределенности. В предлагаемом подходе как раз вносится дополнительная информация, которая элиминирует большую часть неопределенности. Для этого предлагается следующая схема этапов для анализа и прогнозирования редких событий.

1) Первым шагом предлагается отказаться от агрегирования данных в тех случаях, когда теряется информация, чтобы в дальнейшем была возможность изучать всю имеющуюся информацию. Другими словами, предлагается разделять источники¹⁾ событий, а не смешивать их, после чего изучать процесс образования событий в каждом источнике. В каждом источнике образования событий может быть свой собственный уникальный процесс образования событий, когда мы смешиваем все события в одну кучу, мы не сможем разобраться в этих процессах. На примере торговли нужно отказаться от агрегирования продаж от разным покупателям. Вместо этого каждого покупателя следует изучать отдельно.

2) На втором шаге надо использовать как можно больше имеющейся информации о характере протекающего в источнике процесса, приводящего к возникновению событий. На основе этой информации надо предположить механизм образования событий внутри этого источника. Наверное, это самый главный и ответственный этап. Чем точнее предположение окажется к действительности, тем точнее будет результат в дальнейшем. Другими

¹⁾ Под источниками событий понимаются некоторые объекты или системы, в которых происходят какие-то процессы, приводящие к возникновению этих событий.

словами, надо предложить модель, механизм которой приводит к тем самым событиям, которые мы наблюдаем в источнике. В случае, если информации недостаточно, можно попробовать предположить в начале один процесс образования событий, а затем его уточнить или предложить следующий. Даже, если не получится полностью угадать процесс, а только частично, все равно это будет лучше, чем грубо обозвать абсолютно все неопределенности случайностью, так как любая дополнительная информация уменьшит в итоге неопределенность (главного врага исследователя).

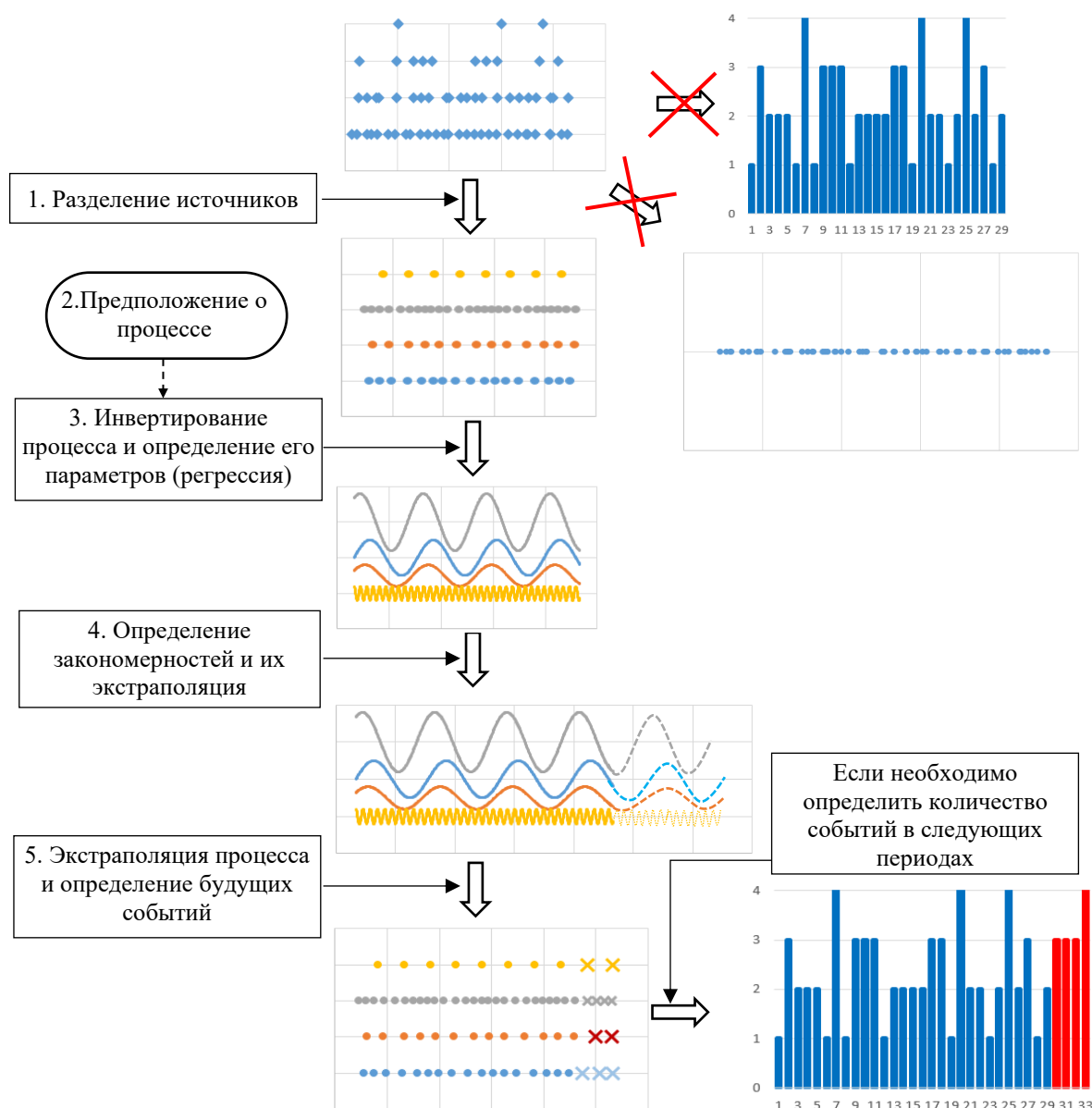
3) На третьем шаге для каждого источника после того, как процесс образования событий установлен, необходимо из имеющихся данных восстановить параметры этого процесса, то есть выполнить регрессию параметров процесса из данных редких событий. Необходимо определить значения параметров механизма/модели образования событий. Однако, никто не говорит, что параметры процесса обязаны быть стационарными, наоборот, на практике чаще всего имеют дело именно с изменяющимися со временем значениями параметров. Тогда нужно будет восстановить динамику изменения параметров процесса из истории редких событий.

4) На четвертом шаге, когда значения параметров процесса были восстановлены, следует произвести поиск закономерностей в том, как эти параметры изменялись со временем или в зависимости от других внешних факторов. Цель этого этапа произвести экстраполяцию параметров процесса на будущее. На этом шаге надо построить соответствующие модели изменения параметров со временем или от изменения внешних факторов, после чего по этим моделям получить прогноз будущих значений параметров процесса в условиях ожидаемых будущих значений внешних факторов. Для этого можно использовать любые известные математические или статистические методы. В общем случае для каждого источника событий для каждого параметра процесса образования событий следует строить собственные модели, по которым в дальнейшем будет произведена экстраполяция, причем для каждой

отдельной модели можно использовать совсем разные математические или статистические методы.

5) Пятым шагом будет запуск самого процесса образования редких событий с экстраполированными значениями параметров, определенными на четвертом шаге. Необходимо построить и запустить алгоритм, реализующий механизм формирования событий в источнике. В результате этого этапа получается прогноз будущих событий, возникающих в конкретном источнике. Другими словами на этом шаге происходит моделирование редких событий по построенной модели, параметры которой были восстановлены из данных редких событий.

Такие действия необходимо выполнить с каждым источником. Затем, если в дальнейшем потребуется получить прогноз будущих событий в виде временного ряда, не составит труда просуммировать эти же события за выбранные интервалы времени. Схема этого метода анализа и прогнозирования редких событий представлена на рисунке 2.1.



Источник: составлено автором.

Рисунок 2.1 - Схема анализа и прогнозирования редких событий

Необходимо отметить, что процесс, протекающий в каждом источнике, может быть полностью уникальным, а не только различаться параметрами. Однако, если все источники событий имеют одну природу, то это существенно упрощает задачу. Достаточно восстановить параметры процесса каждого источника, действуя по единой схеме.

Типов процессов, протекающих в источниках, может существовать большое множество, но в экономике, на мой взгляд, самым распространенным процессом является процесс потребления и расхода продукции, когда запас ведет себя как опустошающаяся емкость. Другим таким процессом может

быть процесс накопления некоторого воздействия до определенного уровня, после чего срабатывает событие. В обоих случаях источники событий можно моделировать как емкости, предложенный метод анализа и прогнозирования редких событий я называю «ёмкостным методом». Например, в торговле обращение клиента за товаром может быть вызвано исчерпанием запаса, когда запасы товара опускаются ниже критического уровня, формируется событие как заказ на пополнение запасов (или на производство). Другим примером может являться обращение клиента за услугой при возникновении соответствующей потребности. Например, потребность в стрижке волос, когда накопление воздействия, то есть рост волос, заставляет клиента обратиться в парикмахерскую, чтобы вновь постричься. В примере обращения клиента за товаром событие образовывается вследствие опустошения емкости, в примере обращения клиента за услугой событие образовывается вследствие наполнения емкости. Обращаю внимание, что в этих примерах мы рассматриваем не собственные запасы/длину волос, а мы моделируем как емкости именно чужие источники событий, то есть рассматриваем запасы/длину волос наших клиентов (не нас самих, в отличии от систем управления запасами).

Какими параметрами обладает такой процесс? Например, в системах управления запасами [65] выделяют такие параметры как максимальный уровень запаса, критический уровень запаса и другие параметры. Однако, у такого процесса самым важным и единственным необходимым параметром для дальнейшего прогноза события является функция $f(t)$ скорости (интенсивности¹⁾) расхода запаса (или накопления воздействия). Достаточно знать только лишь эту функцию $f(t)$ для того, чтобы затем уметь моделировать процесс возникновения следующего событий. Примером такой функции может являться: 1) функция спроса от времени, которая вытягивает

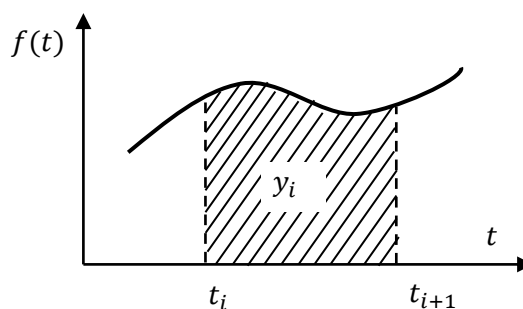
¹⁾ Так как в действительности потребление происходит не непрерывно, а дискретно, причем не одинаковыми порциями, а разыми, то логично назвать эту функцию $f(t)$ не скоростью, а интенсивностью, где под интенсивностью понимается средняя скорость расхода продукции. Однако в дальнейшем для более ясного изложения по-прежнему будем часто называть ее именно как скорость.

товар из запасов на складе промежуточного распространителя, который пополняет свои запасы у нас (у производителя); 2) индивидуальная скорость потребления продукции конечного потребителя, который закупается у нас (у владельца магазина); 3) интенсивность покупок неизвестных покупателей у выбранного клиента (не путать с интенсивностью покупок у нас самих), который закупается у нас (у оптовика) и т.д. Эта функция $f(t)$ является минимальным необходимым параметром процесса, чтобы определить момент времени, когда произойдет следующее событие. Однако, чтобы определить величину будущего события (объем покупки), могут потребоваться другие параметры. Можно предположить, что объем будущей покупки будет соответствовать некоторой зависимости между предшествующими покупками и функцией расхода запаса $f(t)$, и построить соответствующую модель (в простейшем случае можно предположить, что объем покупки будет равен объему предшествующей покупки или среднему значению предшествующих покупок). Подробнее об этом будет сказано ниже и рассмотрено на некоторых примерах. Тем не менее ключевым моментом является именно определение функции скорости расхода запаса (или накопления воздействия) $f(t)$. Это был второй шаг предложенной схемы анализа и прогнозирования редких событий, мы предположили, что редкие события в экономике могут быть смоделированы как процесс опустошения/накопления некоторой емкости и выявили основные параметры этого процесса.

Далее мы переходим к третьему шагу предложенной методики, а именно к восстановлению/регрессии параметров этого процесса. Необходимо восстановить динамику изменения функции $f(t)$ скорости изменения запаса или накопления воздействия из имеющихся данных редких событий. Для этого инвертируем процесс образования событий, чтобы получить значения параметров. В нашем примере инвертируем процесс потребления продукции, получаем обратную задачу к задаче управления запасами (алгоритм в минус первой степени), когда по имеющимся данным редких событий (t_i, y_i) - моментах времени и величинах воздействия события (покупок) определяется

скорость изменения $f(t)$. Для нашего процесса потребления или накопления воздействия используем **основное предположение** – величина совершенного события y_i есть интеграл от исходной функции $f(t)$ за время от момента совершения этого события t_i до момента времени совершения следующего события t_{i+1} , рисунок 2.2. То есть следующее $(i + 1)$ -е событие происходит в тот момент времени, когда запас, сделанный во время текущего i -того события, закончился. С точки зрения накопления возмущения можно дать следующую интерпретацию, предположив, что следующее $(i + 1)$ -е событие возникнет тогда, когда накопленное воздействие достигнет критической точки, если это возмущение было сброшено во время текущего i -того события.

$$y_i = \int_{t_i}^{t_{i+1}} f(t) dt , \quad (2.1)$$



Источник: составлено автором.

Рисунок 2.2 - Основное предположение

Это предположение справедливо, оно негласно используется в логистике в теории управления запасами [65] во время моделирования собственных запасов при определении точки заказа, страховых запасов в связи с неопределенностью спроса и функционального цикла, а также при расчете функции потерь. Отдельно надо заметить, что любое изменение предпочтений потребителей между товарами должно выражаться через изменение функции $f(t)$, а не через нарушение этого предположения. Более подробно по поводу этого предположения, необходимых условий для его соблюдения, а также последствий его нарушения будет рассказано в параграфе 2.3. Конечно же для отдельных видов товаров или услуг это предположение может не всегда выполняться, но пока будем считать, что предположение выполняется.

С использованием этого предположения задача восстановления (регрессии) $f(t)$ представляет из себя задачу восстановления неизвестной функции $\hat{f}(t)$, для которой известна последовательность интегралов за известные непересекающиеся периоды времени. Как будет показано ниже в главе 4, оказывается, как бы странно это не звучало, наблюдаемые значения y_i содержат в себе погрешность $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$ (причем не наоборот, это не функция определяется с погрешностью относительно наблюдений, как в эконометрике). Задачу восстановления функции $f(t)$ по ее интегралам можно сформулировать в виде минимизации суммы квадратов отклонений

$$\sum_{i=1}^{n-1} w_i \left(y_i - \int_{t_i}^{t_{i+1}} \hat{f}(t)dt \right)^2 + C \int_{t_1}^{t_n} (\hat{f}''(t))^2 dt \rightarrow \min, \quad (2.2)$$

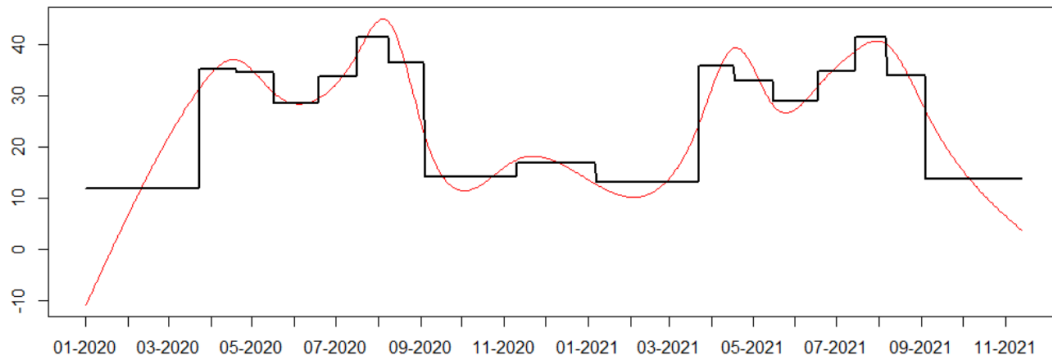
где второе слагаемое - штраф на нелинейность;

C – произвольная большая константа, влияющая на сглаживание;

w_i – вес наблюдения;

n – номер последнего события.

Решение этой задачи получено достаточно сложными математическими методами в виде особого кубического интегрального сглаживающего (аппроксимирующего) сплайна. Пример восстановления функции по интегралам представлен на рисунке 2.3. Подробно о постановке этой оптимизационной задачи, параметрах и о способе ее решение написано в главе 3.



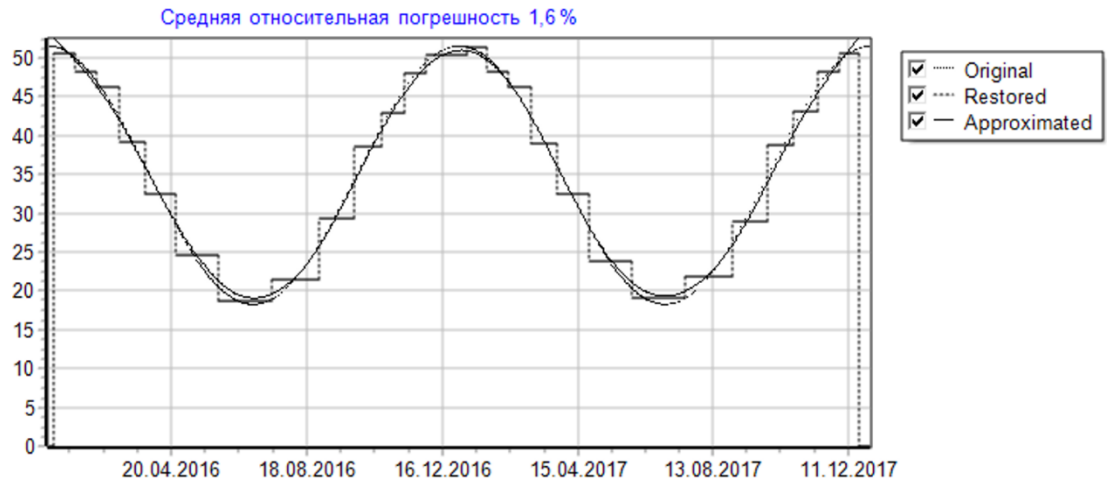
Ступеньки (горизонтальные отрезки) показывают среднее значение функции на этом интервале $y_i/(t_{i+1} - t_i)$, площадь под каждой из ступенек (до оси ноль) – наблюдаемый интеграл y_i . Функция подбирается таким образом, чтобы площадь под функцией приблизительно соответствовала площади ступеньки с учетом штрафа на нелинейность. Подробнее в главе 3.

Источник: составлено автором.

Рисунок 2.3 – Пример восстановления функции $f(t)$ по интегралам y_i , которые наблюдаются с погрешностью $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$, шт. в день

Если возможности осуществлять сложных вычислений нет, можно предложить упрощенный способ, где достаточно ограничиться рассмотрением лишь средней скорости $y_i/(t_{i+1} - t_i)$. Последовательность этих средних скоростей будет представлять из себя ступенчатую функцию, рисунок 2.3, которую можно сгладить любым известным методом для дальнейшего анализа. Для сглаживания могу порекомендовать воспользоваться обычным кубическим регрессионным сплайном, который реализуется в R Project [66] функцией `smooth.spline` базового пакета `stats` или на классических языках программирования с помощью библиотек `Alglib` [67]. Однако, такое сглаживание ступенчатой линии можно использовать только если исходная функция (которая изначально не известна) изменяется достаточно медленно, чтобы изменения этой функции можно было выявить из изменения этих средних скоростей (ступенек), пример рисунок 2.4. Для дальнейшего использования такого сглаживания может быть достаточно, причем по результатам моделирования точность восстановления упрощенным способом оказывается все еще очень высокой, если частота наблюдений значительно превышает частоту изменения исходной функции. Однако такое обычное сглаживание не учитывает основную предпосылку, сглаживаются не интегралы, а ступенчатая функция, построенная по ним. Другими словами,

обычное сглаживание является не совсем обоснованным, сглаженная функция не согласовывается с интегралами от исходной функции. Если же функция изменяется достаточно быстро, а средние скорости почти не изменяются (усредняют эти изменения), как на предыдущем рисунке 2.3, то такое сглаживание будет создавать большую дополнительную ошибку.



Источник: составлено автором.

Рисунок 2.4 - Упрощенный способ восстановления $f(t)$, шт. в день

Даже такое упрощенное усреднение превосходит классический подход на основе временных рядов. Не будем подробно останавливаться на этом, математическое обоснование представлено в работе [64].

Восстановление исходной функции $f(t)$ из данных редких событий позволит в дальнейшем определить момент наступления будущего события. Однако для определения величины воздействия, которое несет это событие, в нашем примере это объем покупки, могут понадобиться другие параметры, такие как критический уровень и максимальный объем запасов. Общей схемы для определения этих параметров нет. Определение этих параметров возможно при условии введения дополнительных предположений. В следующем параграфе будет рассмотрен пример, где определяются такие параметры из данных редких событий.

Четвертый этап экстраполяции параметров процесса на будущее никак не привязан к типу процесса, конкретно на этом этапе предполагается полная свобода и полная ответственность исследователя, можно использовать любые

известные математические или статистические средства. Также на этом этапе помимо данных прошедшей истории можно привлечь дополнительные источники информации, например, мнение экспертов, соцопросы, ожидаемые тенденции из-за политических происшествий или из-за принятия важных законов и регламентов. В нашем примере предполагается сделать прогноз как будет вести себя функция скорости расхода продукции $f(t)$ в будущем.

Наконец на пятом этапе запускается процесс формирования событий. Для этого в нашем примере реализуется процесс потребления так же, как в моделях управления запасами. Начиная с последнего момента пополнения запаса каждый день вычитается значение экстраполированной функции $\hat{f}(t)$ из запаса, который изначально равен объему последней покупки. Когда значение запаса опустится ниже нуля, значит пришло время пополнять запас, в этот момент получаем время возникновения будущего события. Однако для прогноза величины будущего события необходимо будет сделать несколько дополнительных предположений о том, как будет меняться наблюдаемая величина заказа. Рассмотрим эти этапы более подробно на примере в следующем параграфе.

Отдельно несколько слов надо сказать о системах управления запасами [65], в которых на первый взгляд может показаться есть что-то похожее. Во-первых, в системах управления запасами моделируется расход собственных запасов или запасов подконтрольных нам пунктов распространения, а в предложенном емкостном методе анализируется запас неподконтрольных нам покупателей. Во-вторых, в задачах управления запасами прогноз спроса уже задан в каждом подконтрольном пункте распространения, когда здесь наоборот происходит определение скорости расхода (потребления) продукции у покупателей, а после чего определяются дискретные моменты возникновения спроса (покупок) каждого покупателя. В-третьих, используя данный метод мы, не управляем ни чьими запасами, не считаем издержки и не оптимизируем точку заказа. В-четвертых, нам не надо знать ни максимальную величину заказа, ни критическую точку запасов, ни

величину страховых запасов, чтобы определить момент времени, когда последний раз пополненный запас закончится (правда, придется сделать предположение об изменении наблюдаемой величины заказа).

Предполагаемый метод является как бы обратным алгоритмом, алгоритмом в минус первой степени, по сравнению с системами управления запасами. Получается некий регрессионный анализ, цель которого определить скорость расхода продукции, причем индивидуально по каждому нашему покупателю. С другой стороны, этот метод можно отнести к непараметрическим методам прогнозирования. В четвертой главе мы поговорим о точности данного метода. В пятой главе мы попробуем использовать данный метод для других видов редких событий, для которых выполняется основное предположение, не только для анализа продаж (например, социальных, политических и др. событий). Также в пятой главе мы разработаем метод для анализа событий, которые образуются произвольным процессом, а не только процессами, схожими с процессами опустошения/наполнения емкости.

2.2 Примеры использования емкостного метода

Чтобы продемонстрировать работу метода и его эффективность, вначале воспользуемся данными, полученными в результате моделирования, а не реальными данными. Если мы воспользуемся реальными данными, мы не сможем оценить эффективность метода, так как у нас не будет истинных значений, то есть у нас не с чем будет сравнивать полученные результаты. Поэтому сперва получим данные редких событий на основе моделей управления запасами, в которых мы самостоятельно сможем априори задать истинную функцию скорости расхода продукции, которую затем будем восстанавливать, и с которой мы будем затем производить сравнение. Затем следующим шагом произведем прогнозирование будущих событий и сравним прогнозные даты и значения будущих событий с теми, которые бы получились, если бы мы продолжили моделирование. После примеров,

построенных на данных моделирования, приведем пример использования емкостного метода на реальных данных. Сравнение емкостного метода с другими методами будет осуществлено в главе 4.

**Пример на основе модели управления запасами № 1, события
чаще чем раз в месяц**

Смоделируем простейшую систему управления запасами для трех клиентов при достаточно частых событиях (чаще чем раз в месяц), получим набор данных о событиях пополнения запаса этими клиентами. Для каждого клиента зададим индивидуальные значения скорости расхода продукции $f(t)$ (пусть это будут гармонические функции), максимальный запас, начальный запас, критический уровень запасов. Усложним для третьего клиента логику поведения: пусть, когда он видит, что запас кончается быстро (быстрее чем за 15 дней), он повышает уровень максимального запаса в 2 раза, а если он видит, что запас кончается медленно (дольше чем 20 дней), наоборот, сокращает максимальный запас в два раза. Предполагаем, что заказ движется мгновенно, нет неопределенности спроса и функционального цикла (если заказ движется не мгновенно или есть неопределенность, то это приведет лишь к увеличению критического уровня, после которого полагается делать заказ). В результате моделирования в течение двух лет будем иметь следующие данные о времени и объемах покупок наших клиентов, таблица 2.1.

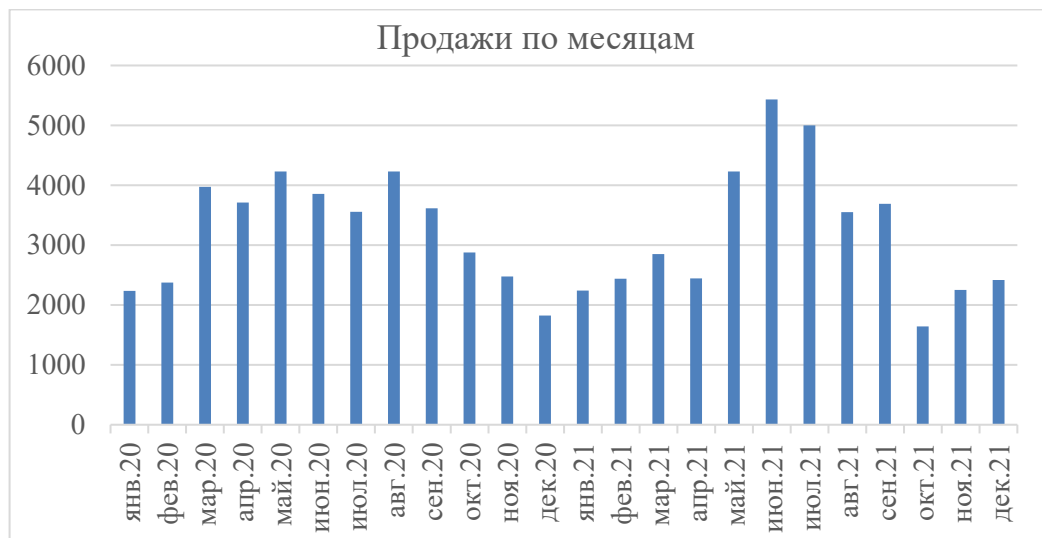
Таблица 2.1 - Данные результатов моделирования системы управления запасами для дальнейшего восстановления исходной зависимости емкостным методом

Клиент 1		Клиент 2		Клиент 3	
Дата t_i	Объем y_i	Дата t_i	Объем y_i	Дата t_i	Объем y_i
11.01.2020	1114,115	11.01.2020	514,4274	24.01.2020	604,0634
08.02.2020	1141,975	05.02.2020	504,1815	22.02.2020	221,7311
04.03.2020	1112,811	29.02.2020	508,4942	01.03.2020	624,9489
29.03.2020	1102,088	27.03.2020	519,2279	19.03.2020	613,2965
26.04.2020	1110,112	20.04.2020	524,9776	03.04.2020	634,499
31.05.2020	1119,101	08.05.2020	524,0971	16.04.2020	1443,4
13.07.2020	1108,96	23.05.2020	528,5013	10.05.2020	603,186
21.08.2020	1108,222	06.06.2020	537,2679	20.05.2020	1456,799
21.09.2020	1129,002	19.06.2020	500,297	09.06.2020	1408,732
17.10.2020	1107,017	03.07.2020	506,6361	28.06.2020	1409,383
11.11.2020	1120,577	19.07.2020	518,7472	17.07.2020	1418,921
07.12.2020	1116,422	06.08.2020	524,808	06.08.2020	1438,24
06.01.2021	1113,698	24.08.2020	514,8002	28.08.2020	642,8028
13.02.2021	1106,693	10.09.2020	512,5303	08.09.2020	1446,487
28.03.2021	1100,366	27.09.2020	526,6422	07.10.2020	619,5263
03.05.2021	1119,241	15.10.2020	515,2883	24.10.2020	633,1191
01.06.2021	1130,962	08.11.2020	515,2761	15.11.2020	220,5331
27.06.2021	1140,962	18.12.2020	502,2907	25.11.2020	620,2834
22.07.2021	1114,795	19.01.2021	510,2025	31.12.2020	205,048
18.08.2021	1111,353	12.02.2021	508,186	14.01.2021	614,4976
20.09.2021	1124,737	09.03.2021	512,3159	15.02.2021	219,7579
01.11.2021	1120,714	04.04.2021	501,4128	23.02.2021	605,2369
12.12.2021	1104,397	26.04.2021	525,8809	15.03.2021	634,6847
-	-	13.05.2021	532,0703	30.03.2021	603,8455
-	-	27.05.2021	511,3226	12.04.2021	1416,788
-	-	09.06.2021	502,9095	07.05.2021	620,8652
-	-	23.06.2021	532,941	17.05.2021	1445,331
-	-	08.07.2021	526,4685	07.06.2021	664,2932
-	-	25.07.2021	529,3331	16.06.2021	1459,679
-	-	12.08.2021	515,0923	05.07.2021	1419,928
-	-	30.08.2021	522,3064	24.07.2021	1407,215
-	-	16.09.2021	521,3989	13.08.2021	1403,277
-	-	03.10.2021	520,4432	05.09.2021	640,3159
-	-	22.10.2021	508,9938	16.09.2021	1404,882
-	-	19.11.2021	505,0699	18.10.2021	611,4484
-	-	30.12.2021	500,1612	07.11.2021	625,5173
-	-	-	-	05.12.2021	213,1149
-	-	-	-	17.12.2021	600,6923

Источник: составлено автором.

Пусть вас не смущает то, что объемы покупок не являются целыми числами. Это произошло из-за того, что функции $f(t)$ были заданы непрерывными, а критический уровень запасов и максимальный запас целыми числами. Можно конечно было округлять результат, но для упрощения и для большей точности не будем ничего отбрасывать.

Далее, пусть эти трое клиентов покупали продукцию у нас, а мы были производителем или оптовым поставщиком, других данных об этих клиентах у нас нет. Попробуем на основе этих данных определить будущие продажи. Сразу отметим, что в случае, если мы агрегируем данные и получаем временной ряд продаж по месяцам, то получится картина как на рисунке 2.5. В этом случае мы столкнемся со всеми теми проблемами, которые были описаны в параграфе 1.3, в результате чего поиск закономерностей и прогнозирование из мешанины данных будет весьма затруднительным.



Источник: составлено автором.

Рисунок 2.5 – Агрегированные продажи по месяцам, шт. в месяц

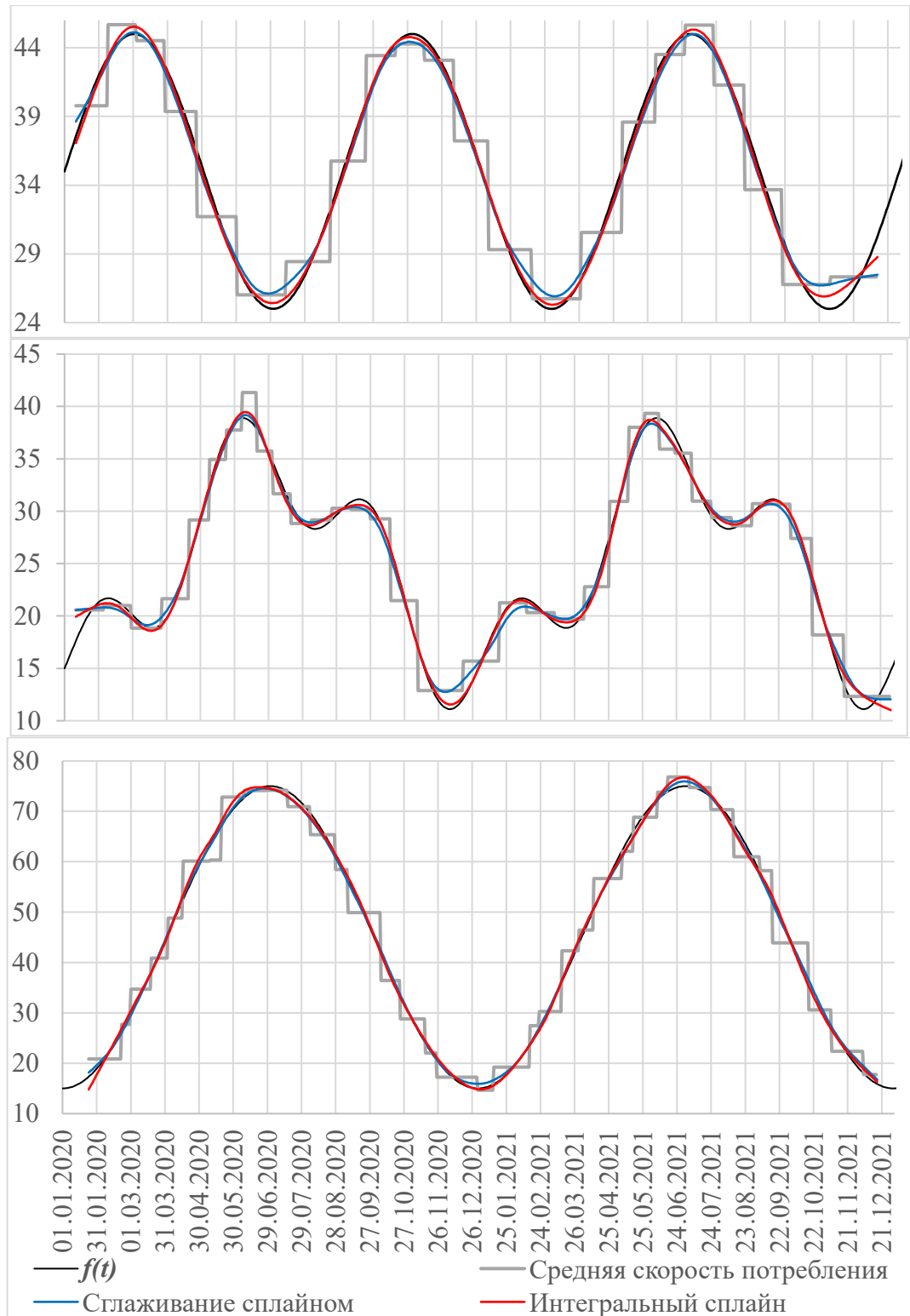
Теперь воспользуемся предложенным методом анализа и прогнозирования редких событий.

Шаг 1. Разделяем источники событий. Так как в исходных данных покупки каждого клиента и так были разделены по разным столбцам, то этот шаг автоматически выполнен.

Шаг 2. Предполагаем, что процесс образования событий соответствует процессу наполнения и опустошения емкости, и моделируем его как емкость, то есть используем описанное выше предположение.

Шаг 3. Проводим регрессию параметров процесса. Мы пытаемся восстановить скорость $\hat{f}(t)$, с которой расходовался запас каждого клиента из данных редких событий таблицы 2.1. Для демонстрации будем использовать

оба способа, как простейший, где определяется средняя скорость $y_i/(t_{i+1} - t_i)$, которая затем сглаживается обычным регрессионным сплайном, так и более сложным, где строится интегральный сплайн особым образом, который минимизирует сумму квадратов отклонения наблюдений y_i и интегралов $\int_{t_i}^{t_{i+1}} \hat{f}(t) dt$, подробнее в главе 3. Еще раз обращаю внимание, что скорость расхода продукции $\hat{f}(t)$ восстанавливаем для каждого клиента в отдельности, рисунок 2.6. Черная сплошная линия показывает исходную функцию $f(t)$, заложенную при моделировании. Серая ступенчатая линия показывает среднюю скорость расхода запаса клиентов, площадь под каждой ступенькой равняется самой покупке y_i . Синяя сплошная линия сглаживает эту ступенчатую функцию обычным сплайном и приблизительно показывает скорость расхода запаса каждый день. Даже такой простой подход неплохо восстанавливает исходную функцию $f(t)$. Однако такой простой способ лишь аппроксимирует точки ступенчатой функции средних скоростей, площадь под этой функцией может сильно не совпадать с объемом покупки y_i . Красная линия – это ежедневная скорость расхода запаса, которая восстановлена из данных о покупках клиентов с помощью интегрального сплайна, эта линия очень хорошо приближается к исходной функции. Площадь под сплайн функцией и площадь под ступенькой (объем покупок) согласуются, но так как площадь под исходной функцией и объемы покупок могут немного различаться, восстановление проходит с погрешностью. Благодаря использованию штрафа на нелинейность $C \int_{t_1}^{t_n} (\hat{f}''(t))^2 dt$ скачки интегралов сглаживаются, восстановление улучшается, подробнее в главах 3 и 4.



Источник: составлено автором.

Рисунок 2.6 – Скорость расхода продукции $\hat{f}(t)$ первого, второго и третьего клиента соответственно, шт. в день

Среднее относительное отклонение восстановленной функции от исходной с помощью интегрального сплайна составляет всего 0,95%, 1,9% и 1,5% для первого, второго и третьего клиента соответственно. Причем

наибольшее отклонение замечено на концах сплайна, это классическая проблема всех сплайнов, на концах интервала из-за недостатка информации о поведении функции сплайн не знает куда стремиться (подробнее в главе 3). Если же не учитывать концы сплайнов и отбросить по 30 точек, то средняя относительная погрешность уменьшится и составит 0,088%, 1,19% и 1,28%. Увеличивающуюся на концах погрешность желательно учесть на следующем шаге при построении модели и экстраполяции.

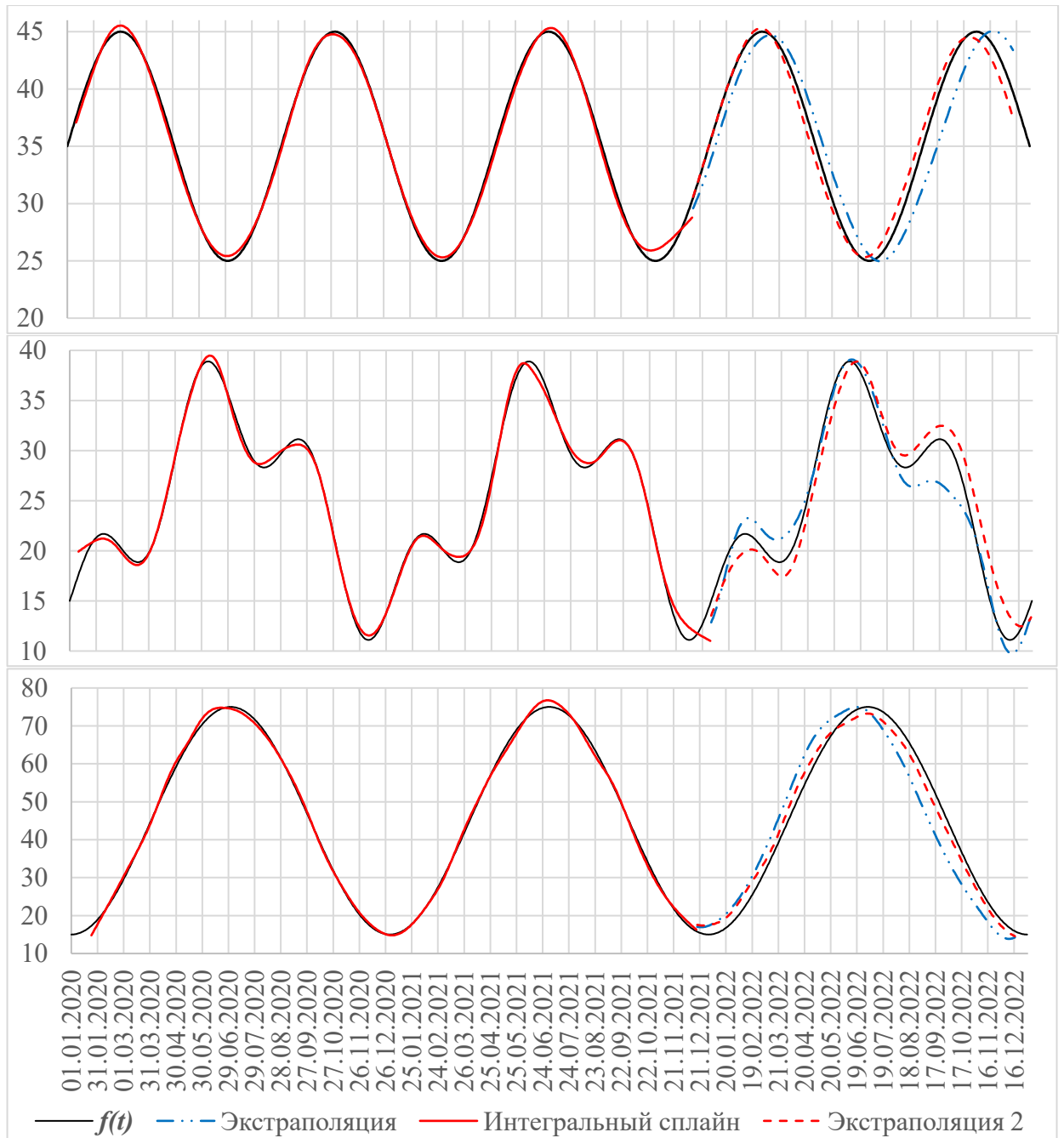
Шаг 4. Строим модель и проводим экстраполяцию параметров процесса, то есть пытаемся определить, как поведет себя функция скорости расхода продукции $\hat{f}(t)$ в будущем. Мы наблюдаем некоторые колебания (при моделировании мы их такими и заложили), если бы у нас имелись знания из соответствующей предметной области о том, как поведет себя эта функция в будущем, мы бы строили модель на основе их. Тут мы будем считать, что предметная область нам подсказывает, что функция состоит из суммы гармонических колебаний. Тогда попробуем разложить эту функцию на соответствующие гармоники и экстраполируем ее на будущее.

Стоит сказать, что обычное разложение в ряд Фурье не подходит для экстраполяции, так как в нем весь рассматриваемый интервал времени (выборка наблюдений) предполагается периодом всей функции, и при экстраполяции это приводит к тому, что копируются первые значения функции и значение на конце интервала скачком становится равным значению в начале интервала (хорошо об этом написано в статье [68]). Для экстраполяции хорошо подходит метод Куинна и Фернандеса (Quinn-Fernandes algorithm) [69; 70], который представляет исходные данные в виде суммы ограниченного числа гармонических функций, которые можно с легкостью экстраполировать. Алгоритм получил свою популярность в области прогнозирования финансовых показателей на рынках Forex. Подробно описывать этот алгоритм не будем, в интернете можно найти версию алгоритма, выложенную пользователем Vladivir (за авторством некоторого

gpwr) [71]. В приложении А представлена адаптированная под язык R версия алгоритма.

Благодаря использованию алгоритма Куинна и Фернандеса раскладываем восстановленную функцию скорости расхода продукции $\hat{f}(t)$ на конечное количество гармоник (применяем алгоритм к значениям сплайна). Тем самым мы строим модель для экстраполяции значений, после чего экстраполируем значения на год вперед. Из-за того, что на концах участка сплайн обладает большой погрешностью, построение модели может быть менее точным по сравнению с моделями, которые строятся без учета точек на концах сплайна. На рисунке 2.7 синей пунктирной линией показаны экстраполированные значения функции, оцененной по всей выборке (по значениям восстановленной функции), красной пунктирной линией показаны точки модели, построенной по выборке без учета некоторого количества начальных и конечных точек (минус 50 точек для клиента 1 и 2, минус 20 точек для клиента 3). Экстраполированные значения по модели, которая получена не по всей выборке, а только по ее части, без некоторого количества точек с начала и с конца, оказались ближе к истинным значениям.

Из-за того, что восстановление исходной функции по данным редких событий происходило все же не идеально, а с небольшой погрешностью (всего около 1%), даже эта небольшая погрешность приводит к тому, что в нашем примере при оценке моделей параметры этих моделей, такие как амплитуда, частота и фаза, определяются с незначительной, но все же погрешностью. Из-за этого со временем значения, полученные по модели, и истинные значения, заложенные в самом начале моделирования покупок, будут расходиться. В результате чем дальше мы экстраполируем значения функции, тем все меньшей и меньшей точностью мы обладаем. Не следует пытаться получать прогноз на слишком далекую перспективу.



Источник: составлено автором.

Рисунок 2.7 – Экстраполяция скорости расхода продукции первого, второго и третьего клиента соответственно, шт. в день

Шаг 5. Запуск процесса образования редких событий. Так как прогноз скорости расхода продукции уже получен, не составит труда определить, в какую дату закончится последний раз сделанный запас, а вот объем покупки предстоит дополнительно спрогнозировать. В нашем примере для первого и второго клиента это сделать достаточно просто. Из данных Таблицы 1 мы видим, что объемы покупок приблизительно на одном уровне, но все же немного различаются. Можно предположить, что покупатель каждый раз

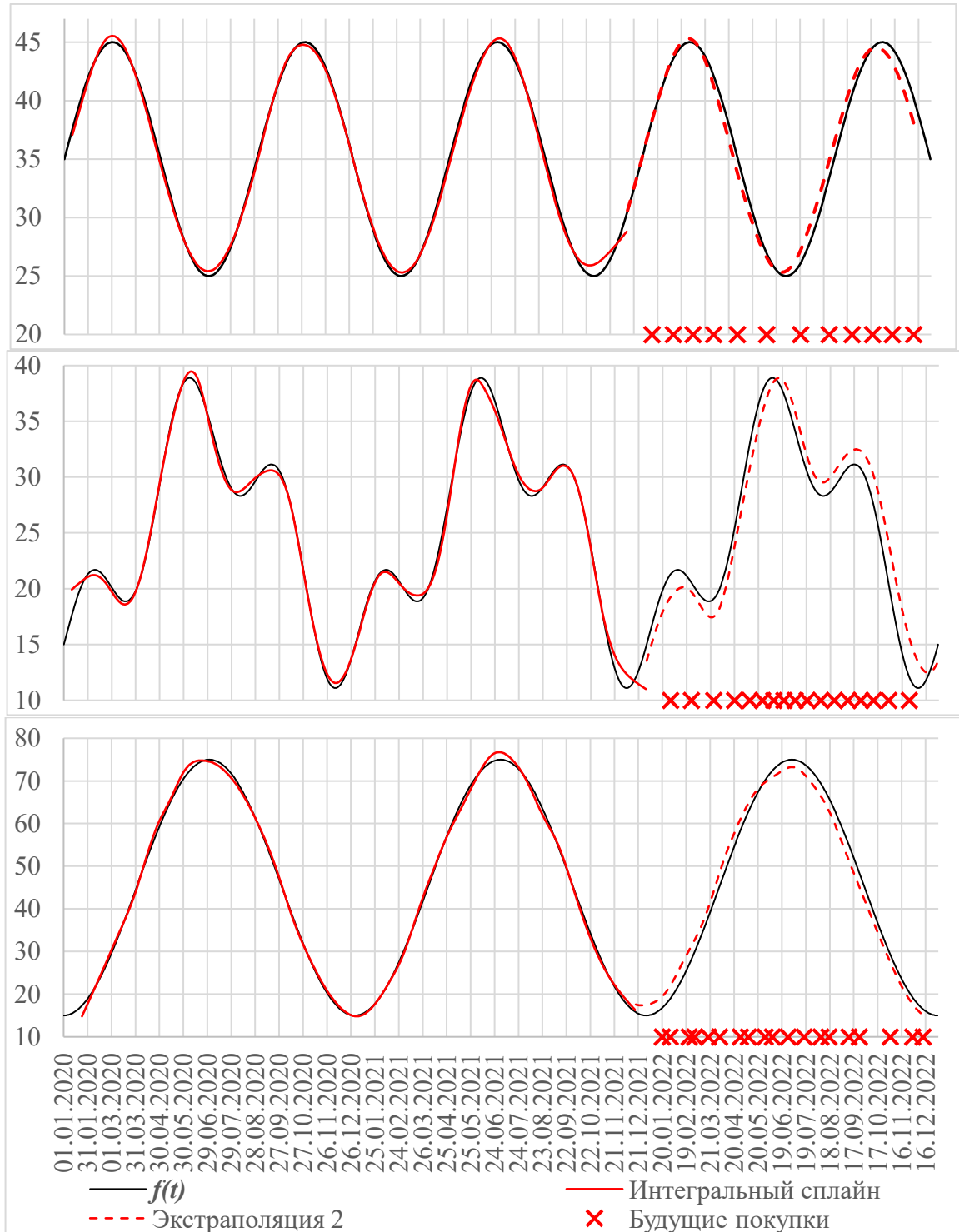
делает заказ до определенно постоянного максимума в тот момент, когда запасы опускаются ниже критического уровня. Из-за как раз этого уменьшения запаса ниже критического уровня объемы покупок не постоянны, так как эта величина имеет некоторый случайный характер, в момент перед покупкой запасы выше критического уровня, затем происходит вычитание $f(t_i)$, запасы становятся ниже критического уровня. Объем покупки без учета компенсации этого занижения (разницу между максимальным и критическим уровнем запаса) приблизительно можно определить как $\frac{1}{n} \sum_{i=1}^n (y_i - \frac{1}{2} \hat{f}(t_i))$.

Для третьего клиента, который меняет максимальный уровень в зависимости от длительности расхода предыдущей покупки, это сделать немного тяжелее. Из данных замечаем, что объем покупки резко уменьшается, когда время с предыдущей покупки равно 29, 24, 22, 29, 22, 36, 32, 25, 21, 23, 32, 28 дней, минимальное из них 21. Наоборот, объем покупки резко увеличивается, когда время между покупками 8, 13, 10, 11, 10, 14, 8, 13, 10, 9, 12 дней, максимальное из них 14. Далее видим, что объемы покупок после вычета $\hat{f}(t_i)/2$ находятся в районе трех групп со средними значениями 205, 602, 1394. Можем определить критический уровень запаса предполагая, что максимальный уровень увеличивается в 2 раза, составляя уравнение, $2(x + 205) = x + 602$ и $2(x + 602) = x + 1394$, из первого получаем $x = 208$ из второго $x = 190$, среднее между ними 199 (при моделировании задавали 200). Тем самым определили недостающие параметры¹⁾.

Далее, когда запускаем процесс расхода продукции для каждого клиента, когда запас продукции клиента опустился ниже критического уровня, объем покупок определяется как ожидаемый объем покупок плюс разница между критическим и текущим уровнем запасов. После этого, зная объем покупки, вновь определяем дату, когда сделанный запас снова закончится, затем определяем объем очередной покупки, и так далее. В результате

¹⁾ Данные действия следовало отнести к шагу 4, так как на нем определяются параметры.

получаем прогноз будущих покупок покупателей на год вперед для трех покупателей, рисунок 2.8, таблица 2.2.



Источник: составлено автором.

Рисунок 2.8 – Экстраполяция и прогноз будущих покупок

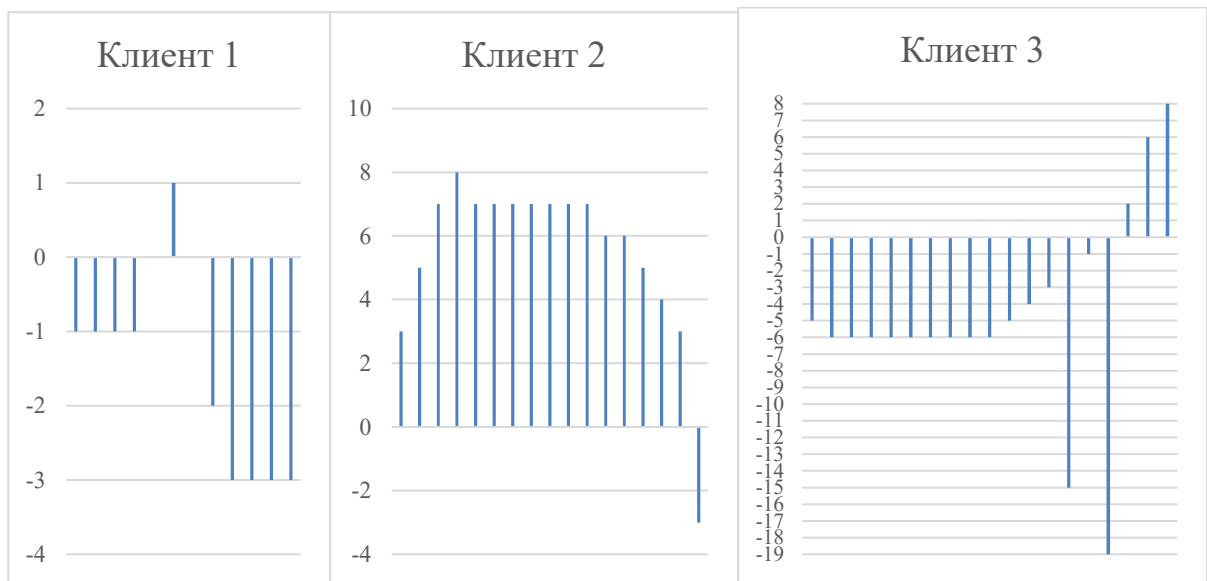
Таблица 2.2 – Прогноз будущих покупок трех клиентов

Клиент 1		Клиент 2		Клиент 3	
Дата t_i	Объем y_i	Дата t_i	Объем y_i	Дата t_i	Объем y_i
13.01.2022	1101,329	30.01.2022	520,1686	20.01.2022	209,5944
09.02.2022	1116,927	25.02.2022	517,3576	30.01.2022	606,8991
06.03.2022	1123,24	25.03.2022	511,5398	23.02.2022	230,899
01.04.2022	1131,209	20.04.2022	520,1159	02.03.2022	622,3613
01.05.2022	1124,655	09.05.2022	521,644	19.03.2022	621,5748
07.06.2022	1099,579	25.05.2022	536,4689	02.04.2022	1425,791
20.07.2022	1112,967	08.06.2022	524,1783	28.04.2022	641,5266
25.08.2022	1112,185	21.06.2022	504,0039	08.05.2022	1430,414
23.09.2022	1114,02	05.07.2022	521,8361	29.05.2022	622,216
19.10.2022	1128,388	20.07.2022	504,6267	07.06.2022	1432,721
13.11.2022	1103,398	06.08.2022	515,2753	27.06.2022	1442,342
10.12.2022	1101,441	23.08.2022	507,4128	17.07.2022	1455,057
-	-	09.09.2022	533,6898	07.08.2022	648,4121
-	-	25.09.2022	518,2384	17.08.2022	1444,974
-	-	11.10.2022	503,0449	11.09.2022	628,1787
-	-	30.10.2022	517,8003	24.09.2022	1423,856
-	-	25.11.2022	504,0123	02.11.2022	601,4933
-	-	-	-	30.11.2022	213,2429
-	-	-	-	13.12.2022	609,2449

Источник: составлено автором.

Отклонение прогнозных дат от фактических (если продолжать моделирование) для первого клиента составило 0-3 дня (среднее отклонение ~ 1,58 дня). Причем отклонение самых первых дат составляло всего 1 день в меньшую сторону, в середине отклонение стало 1 день в большую сторону, а последних на 3 вновь в меньшую сторону, рисунок 2.9. Отклонение объемов составило в среднем лишь 1% (объем пополнения запаса, когда запас пополняется до постоянного максимума очень легко восстанавливается). Для второго клиента, так как экстраполированная функция получилась как будто сдвинутой и в начале значения функции оказались заниженными, прогнозные даты в начале оказались сдвинутыми относительно фактических на 3-8 дней в большую сторону, в середине отклонение составляло 7 дней, затем отклонения уменьшались и в конце даты стали сдвинуты на 3 дня в меньшую сторону. Отклонение объемов также составило в среднем 1%. Для третьего клиента, у которого экстраполированная функция оказалась как будто сдвинутой влево и значения которой в начале оказались завышенными, отклонение первых дат

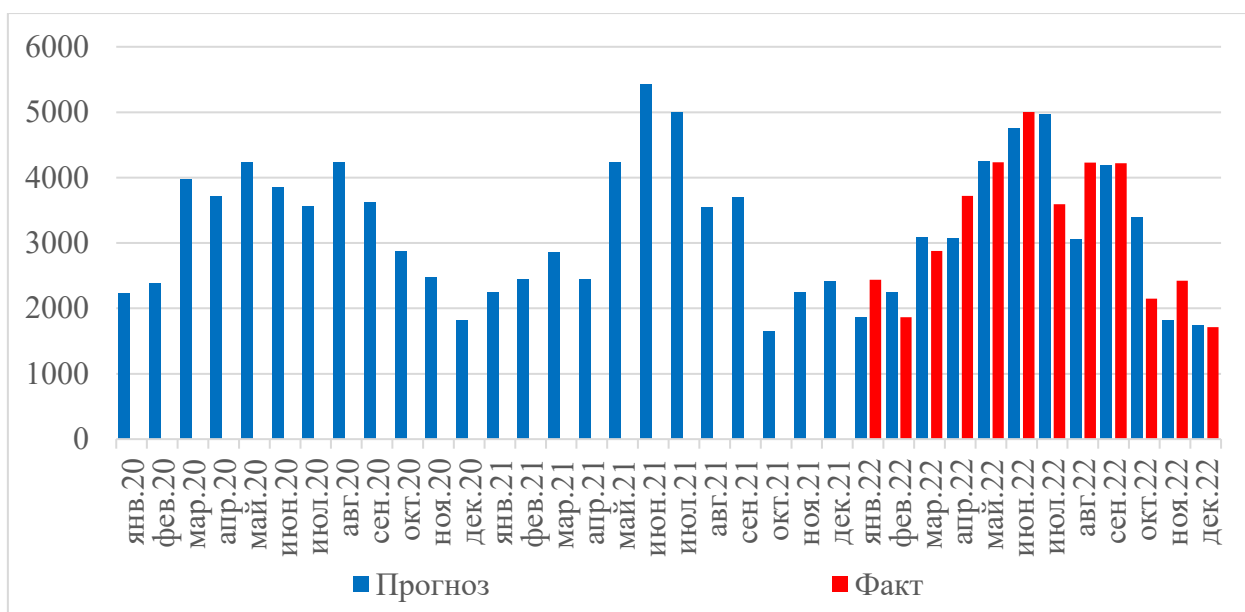
событий оказалось сдвинуто в более раннюю сторону на 5-6 дней, затем отклонение стало уменьшаться до 3 дней. Но так как третий клиент действовал по особой схеме, увеличивая максимальный объем запаса, когда запасы кончались быстро, и уменьшая максимальный объем, когда запасы заканчивались медленно, то в один момент из-за погрешности не произошло увеличение объема, и следующая покупка спрогнозировалась слишком рано. Отклонение объемов до того момента, когда увеличения/уменьшения выполнялись правильно, составило в среднем 2%.



Источник: составлено автором.

Рисунок 2.9 – Отклонения прогнозных дат будущих покупок от фактических, в днях

В случае необходимости построения прогноза в виде временного ряда определяем прогноз суммарных продаж по месяцам от всех трех клиентов, рисунок 2.10. Из-за того, что некоторые покупки сдвинуты и не попадают в соответствующий месяц, между фактическими значениями и прогнозными также наблюдаются отклонения. Однако эти отклонения уже не случайные, то, что не попало в текущий месяц, обязательно попадет в следующий. Если же нас интересует суммарный спрос за несколько месяцев, например за 3 месяца, то суммарный прогноз на 3 месяца практически не отличается от фактического, так как эти отклонения в основном компенсируются, таблица 2.3.



Источник: составлено автором.

Рисунок 2.10 – Прогноз продаж по месяцам, шт. в месяц

Таблица 2.3 – Суммарный прогноз на 3 месяца

с	по	Прогноз	Факт
январь 22	март 22	7181,891	7193,2098
февраль 22	апрель 22	8462,543	8400,6393
март 22	май 22	10832,76	10403,14
апрель 22	июнь 22	12956,87	12078,43
май 22	июль 22	12832,71	13980,807
июнь 22	август 22	12825,57	12790,237
июль 22	сентябрь 22	12040,73	12226,002
август 22	октябрь 22	10595,47	10647,887
сентябрь 22	ноябрь 22	8789,362	9412,377
октябрь 22	декабрь 22	6282,066	6964,5242

Источник: составлено автором.

Помимо прогноза будущих событий само восстановление функции $\hat{f}(t)$ скорости расхода запаса у наших клиентов может иметь отдельное большое практическое значение. Если просуммировать эту функцию по всем клиентам (а не суммировать продажи по месяцам), можно получить непрерывную суммарную скорость расхода выбранного товара всеми клиентами, которую можно использовать при планировании собственного производства или при управлении запасами, дополнительно анализировать другими средствами и смотреть на связь с изменениями глобальных мировых тенденций или внешних факторов.

**Пример на основе модели управления запасами № 2, одно событие
в 2–5 месяцев**

Рассмотрим похожий пример, но пусть в нем клиенты делают всего несколько покупок в год (достаточно увеличить максимальный запас в несколько раз), также пусть моделирование будет проводиться за 3 года. В результате будем иметь следующие данные, таблица 2.4, покупки возникают редко, между ними проходит около 2–5 месяцев.

Таблица 2.4 - Данные редких покупок

Клиент 1		Клиент 2		Клиент 3	
Дата t_i	Объем y_i	Дата t_i	Объем y_i	Дата t_i	Объем y_i
02.01.2020	3635,258	16.01.2020	2806,198	21.01.2020	2312,678
27.03.2020	3600,29	18.05.2020	2833,986	07.04.2020	2334,631
20.07.2020	3604,233	08.08.2020	2828,133	17.05.2020	4813,698
27.10.2020	3610,861	28.11.2020	2802,192	22.07.2020	4805,369
29.01.2021	3606,871	27.04.2021	2814,076	13.10.2020	4819,585
29.05.2021	3628,243	15.07.2021	2803,469	14.04.2021	2336,409
21.08.2021	3606,369	19.10.2021	2819,449	22.05.2021	4800,22
24.12.2021	3618,284	28.03.2022	2808,217	27.07.2021	4831,639
21.03.2022	3615,173	25.06.2022	2818,402	23.10.2021	4816,597
10.07.2022	3607,478	25.09.2022	2820,295	21.04.2022	2351,224
22.10.2022	3643,247	-	-	28.05.2022	4840,185
-	-	-	-	02.08.2022	4844,959
-	-	-	-	06.11.2022	2313,316

Источник: составлено автором.

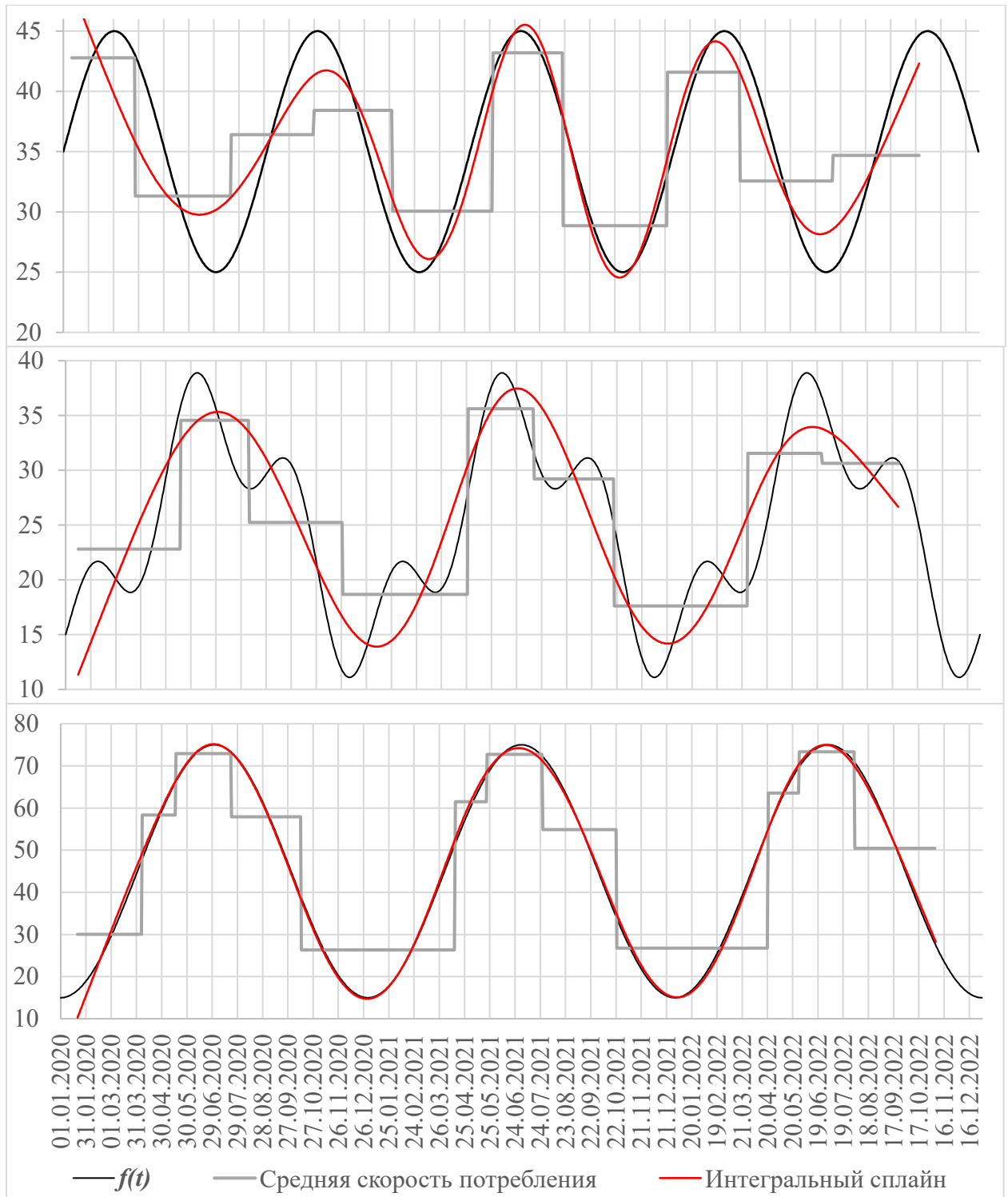
Если по этим данным построить временной ряд продаж по месяцам, по получится картина как на рисунке 2.11. Из такого временного ряда не просматривается никакой закономерности, представленные таким образом данные очень сложно анализировать.



Источник: составлено автором.

Рисунок 2.11. Редкие продажи, агрегированные по месяцам, шт. в месяц

После **шагов 1, 2 и 3** получаем восстановленные из данных редких событий функции скорости расхода продукции, рисунок 2.12. Конечно же из-за того, что у нас имелось намного меньше данных, точность восстановления снизилась, однако, закономерности в этих функциях бросаются в глаза сразу, можно сравнить рисунок 2.11 и новый рисунок 2.12.



Источник: составлено автором.

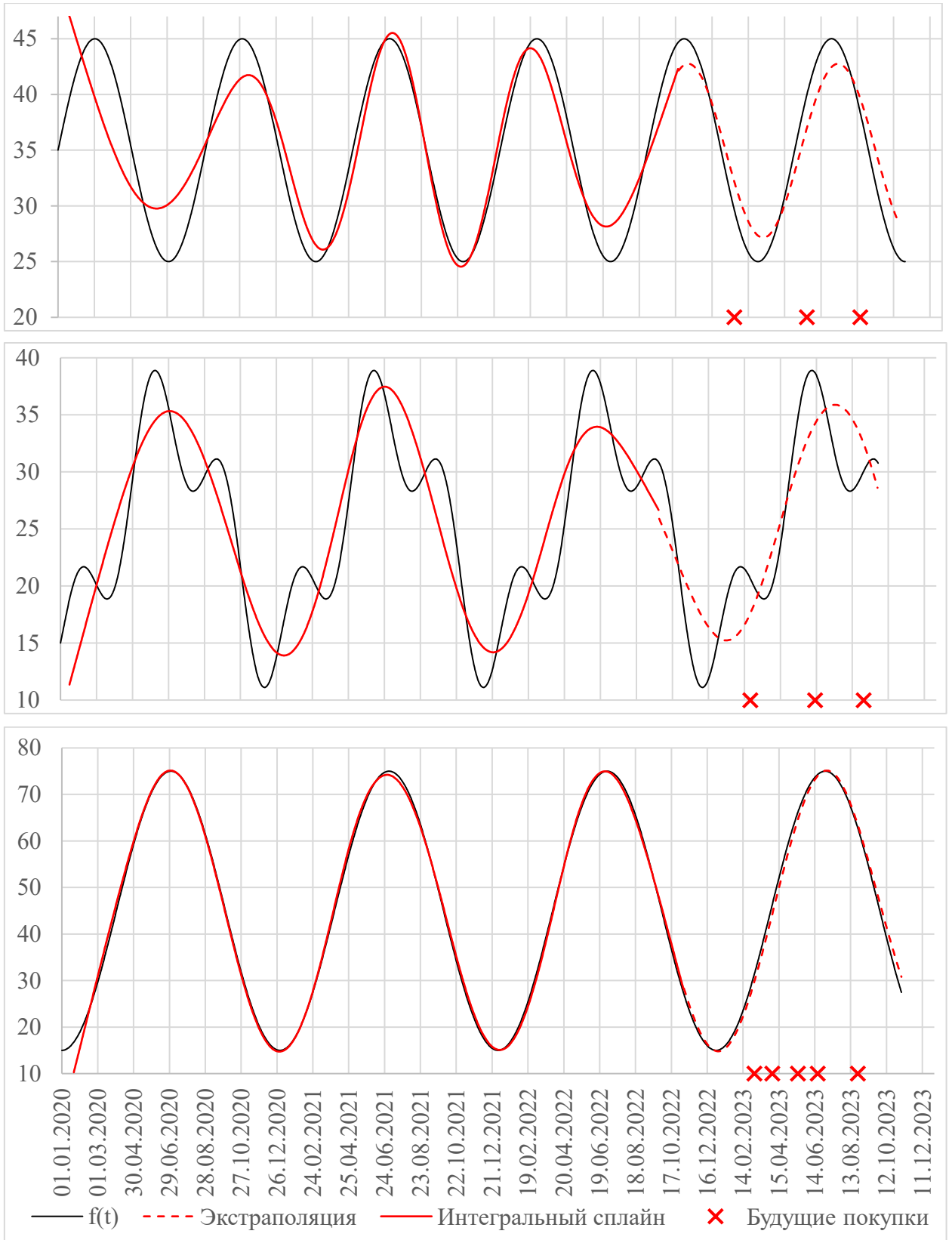
Рисунок 2.12 – Скорость расхода продукции первого, второго и третьего клиента соответственно, шт. в день

У первого клиента в самом начале восстановленная функция плохо совпадает с исходной. Такое произошло из-за нехватки информации (не было еще одной ступеньки слева, заставляющей функцию заворачивать вниз), вторая производная определяется меньше, чем следовало бы. Однако в центре

интервала восстановление произошло достаточно хорошо. На конце интервала снова не хватает информации, что функции надо начать поворачивать вниз. Если бы интервал наблюдения (время моделирования) был не 3 года, а больше, то центральный участок, где восстановление происходит хорошо, стал бы шире, но на новых концах интервала все равно могла наблюдаться та же картина, когда функция не знает куда стремиться. У второго клиента из-за того, что покупки происходят редко, определяется только основная гармоника и теряется гармоника с более высокой частотой. Такое происходит из-за того, что интеграл от быстро изменяющейся исходной функции как бы сглаживает эти быстрые изменения (данные содержат интегралы за широкие интервалы времени). Для третьего клиента, который менял свой уровень максимального запаса со временем, восстановление получилось достаточно хорошим, пусть даже покупок было мало. Такое произошло из-за того, что исходная функция изменялась достаточно медленно, и последовательность интегралов смогла объяснить ее изменения.

Шаг 4. Вновь воспользовавшись алгоритмом Куинна и Фернандеса [69; 70], проводим экстраполяцию восстановленных функций на год вперед (спустя последнее событие), рисунок 2.13. Вновь можно заметить, что из-за погрешности в определении параметров с ростом времени экстраполированные функции расходятся с истинными функциями $f(t)$, заложенными при моделировании.

Шаг 5. Запускаем процесс образования событий. Определить момент времени, когда закончится последняя покупка не составит труда, но, чтобы определить объем вновь сделанной покупки, требуется построить соответствующую модель.



Источник: составлено автором.

Рисунок 2.13 – Экстраполяция скорости расхода продукции первого, второго и третьего клиента соответственно, шт. в день

Для первых двух клиентов вновь предполагаем, что клиент каждый раз делает заказ до определенно постоянного максимума, который определяем как

$\frac{1}{n} \sum_{i=1}^n (y_i - \frac{1}{2} \hat{f}(t_i))$. Однако для третьего клиента, который меняет максимум своих запасов в зависимости от того, на сколько быстро закончился предыдущий запас продукции, требуется определить больше параметров. Из данных о покупках третьего клиента, таблица 2.4, можем увидеть, что повышение объема покупок произошло, когда время между покупками составило 40, 38 и 37 дней, максимальное среди них 40 (при моделировании время, при котором клиент увеличивал свой максимальный запас, задавалось равным 50 дней). Аналогично видим, что сокращение объема происходит, когда время между покупками составляет 183, 180 и 96 дней, минимальное среди которых 96 (при моделировании время, при котором клиент сокращал максимальный запас, задавалось равным 90 дней). Рассматривая объемы покупок без компенсации использования страховых запасов (определяемых как $y_i - \frac{1}{2} \hat{f}(t_i)$), выделяем две группы покупок со средним объемом 2310 и 4790. Тогда на основе уравнения $2(x + 2310) = x + 4790$ определяем критический уровень запасов $x = 170,5$, который определяет, когда клиент делает новый заказ (если уровень текущих запасов ниже критического, то делается новый заказ, в нашем случае покупка; при моделировании критический уровень задавался как 200). После определения этих параметров¹⁾ можем прогнозировать объемы следующих покупок и тем самым прогнозировать целую последовательность будущих событий.

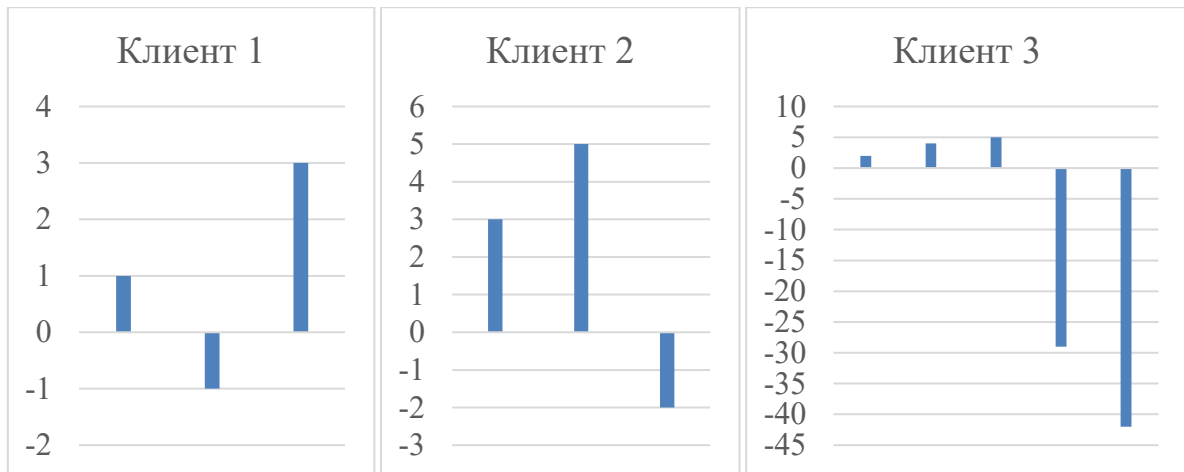
Наконец, запуская процесс образования событий, получаем прогноз будущих покупок на год вперед (относительно последней сделанной покупки), таблица 2.5, изображены крестиками на рисунке 2.13. Отклонения прогнозных дат от фактических показано на рисунке 2.14.

¹⁾ Определение параметров следует отнести к четвертому шагу, но они описаны в этом шаге для более последовательного изложения.

Таблица 2.5 - Прогноз будущих покупок трех клиентов

Клиент 1		Клиент 2		Клиент 3	
Дата t_i	Объем y_i	Дата t_i	Объем y_i	Дата t_i	Объем y_i
22.01.2023	3617,033	24.02.2023	2817,258452	05.03.2023	1093,569
22.05.2023	3612,909	12.06.2023	2832,31828	04.04.2023	2356,954
18.08.2023	3620,031	01.09.2023	2832,205922	17.05.2023	2371,955
-	-	-	-	19.06.2023	4800,605
-	-	-	-	25.08.2023	4811,806

Источник: составлено автором.



Источник: составлено автором.

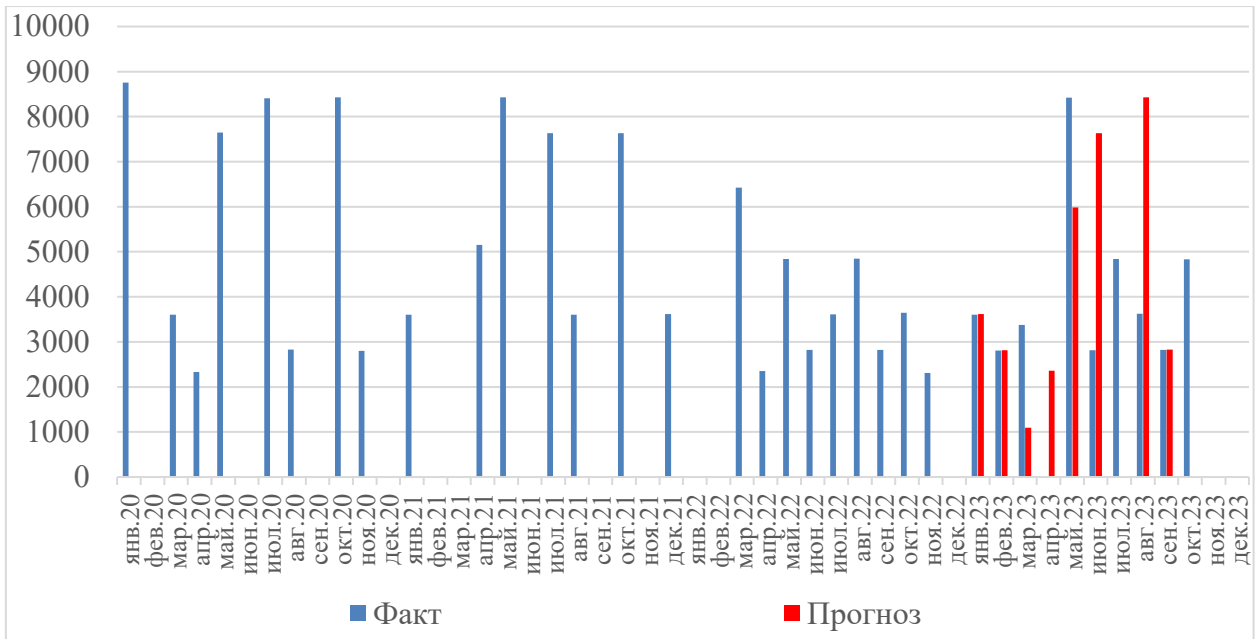
Рисунок 2.14 – Отклонения прогнозных дат будущих покупок от фактических, в днях.

Даты будущих покупок для первого и второго клиента определились достаточно хорошо несмотря на то, что восстановленная функция местами отличалась от исходной достаточно сильно. Так получилось из-за того, что покупки происходят редко, и за время между этими покупками превышение на одних участках и занижение на других участках успели компенсировать друг друга. Заметим, что отклонения составили 1-5 дней, при времени между событиями 3-4 месяца, это очень хороший результат. Отклонение прогнозного объема покупок от фактического у первого и второго клиента составило меньше 0,5%.

Для третьего клиента, который меняет свой максимальный запас в зависимости от того, как быстро закончилась предыдущая покупка, моменты времени только трех будущих покупок определились достаточно точно, а последующие с большой ошибкой. Объемы первых двух покупок определились также хорошо, отклонения 1,6% и 2%. Однако объем третьей

покупки определен неправильно, вместо объема покупки 4811,072 прогноз составил 2371,955. Так произошло из-за того, что время между третьей и второй покупкой составило 43 дня, когда параметр, отвечающий за время быстрой покупки, ранее определен как 40 дней, а при моделировании закладывался как 50 дней. В результате при моделировании (при получении исходных фактических покупок) произошло увеличение максимального запаса, а при запуске процесса с восстановленными параметрами (при получении прогнозных покупок) не произошло этого увеличения. Напомним, что исходных данных было очень мало, повышений и понижений объемов покупок встретилось лишь по 3 раза, поэтому из этих трех наблюдений восстановление параметров произошло не очень точно. Из-за того, что объем третьей покупки спрогнозировался в два раза меньше, даты четвертого и пятого события спрогнозировались с большим сдвигом.

В случае необходимости построения прогноза в виде временного ряда определяем прогноз суммарных продаж по месяцам от всех трех клиентов, рисунок 2.15. Прогноз на следующие два месяца хорошо совпадает с фактическим. В третьем месяце наблюдается отклонение, вызванное тем, что вторая фактическая покупка 4-того года третьего клиента должна была выпасть на 31.03.2023, а она спрогнозировалась на 04.04.2023 и попала в четвертый месяц, а не в третий. Не совпадение в мае вызвано тем, что третья прогнозная покупка третьего клиента не увеличилась в два раза. Далее расхождения становятся еще больше. Если же прогнозировать объем продаж на 3 месяца вперед, то часть расхождений, вызванных тем, что какие-то события промахиваются и не попадают в нужный месяц, компенсируются, однако расхождения, вызванные событиями, которые неправильно прогнозируются, могут не скомпенсироваться, таблица 2.6.



Источник: составлено автором.

Рисунок 2.15 – Прогноз продаж по месяцам на год вперед, шт. в месяц

Таблица 2.6 – Суммарный прогноз на 3 месяца

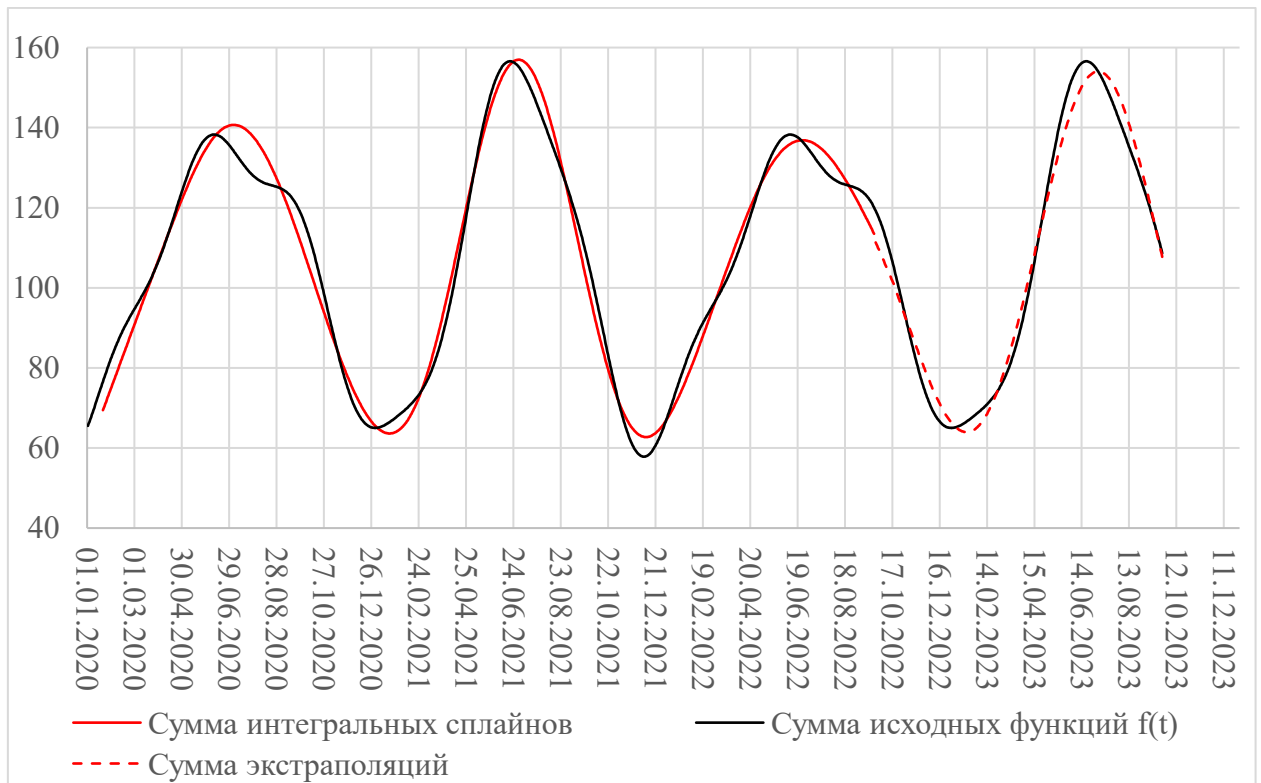
В штуках

с	по	Прогноз	Факт
янв.22	мар.22	9780,92	7527,86
фев.22	апр.22	6180,849	6267,781
мар.22	май.22	11801,46	9435,387
апр.22	июн.22	11241,26	15974,74
май.22	июл.22	16080,57	13617,79
июн.22	авг.22	11278,7	16064,76
июл.22	сен.22	11286,4	11264,04
авг.22	окт.22	11278,57	11264,04
сен.22	ноя.22	7655,845	2832,206
окт.22	дек.22	4831,48	0

Источник: составлено автором.

Однако, если нас интересуют суммарные продажи, возможно, нам будет интересно воспользоваться не суммой продаж по месяцам, а суммарной функцией скоростей потребления, рисунок 2.16. И это скорее будет предпочтительней, так как тут уже будет неважно, куда попало прогнозное наблюдение, на конец текущего месяца или уже в следующий, не надо будет строить модель для объемов покупок. С помощью такой суммарной функции можно строить долгосрочные планы производства или продаж. Однако, наибольшего экономического эффекта можно будет добиться только тогда,

когда будет осуществляться прогноз дискретных событий по каждому клиенту в отдельности (но потребуются определять больше параметров).



Источник: составлено автором.

Рисунок 2.16 – Суммарная скорость расхода продукции всех трех клиентов, шт. в день

Отметим, что за все время диссертационного исследования, аналогичные примеры на основе данных моделирования систем управления запасами строились несколько сотен, если не тысячу раз. Во всех этих примерах из данных успешно получалось восстановить заложенную исходную функцию скорости потребления.

Пример на реальных данных, поставка кваса

Теперь перейдем к примерам¹⁾ анализа и прогнозирования редких событий на реальных данных. Однако при работе с реальными данными у нас не будет возможности оценить точность восстановления исходных функций, так как никто нам свыше не сообщит какое было истинное значение функций. Вместе с тем мы ограничены имеющимися данными, у нас нет возможности

¹⁾ Результаты исследования получены совместно с исполнителями НИР – Головановой П.С. и Кострицей Т.А., которые на момент начала НИР являлись студентами автора данной диссертационной работы, в задачи которых входил поиск данных и первичная обработка.

проверить полученные прогнозы будущих событий, определить, как точно они сбылись. В итоге, данные примеры носят больше демонстрационный характер, показывая принципиальную возможность использования предлагаемого метода для анализа и прогнозирования редких событий в различных прикладных областях.

Имеются данные о поставках полуторалитровых бутылок кваса в универсам, таблица 2.7. Попробуем проанализировать и спрогнозировать будущие события (поставки).

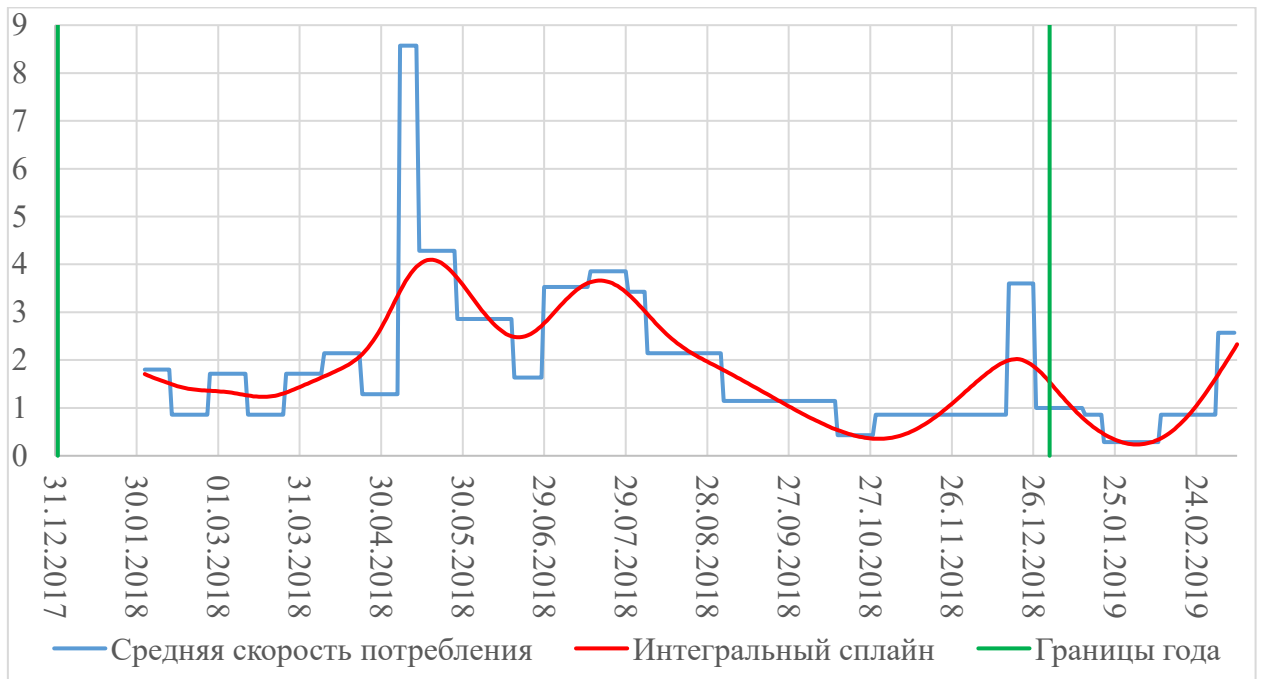
Таблица 2.7 – Данные о поставках бутылок кваса в универсам

Дата t_i	Объем y_i	Дата t_i	Объем y_i	Дата t_i	Объем y_i
02.02.2018	6	28.05.2018	60	06.11.2018	12
05.02.2018	12	18.06.2018	18	12.11.2018	6
12.02.2018	6	29.06.2018	60	10.12.2018	6
19.02.2018	6	16.07.2018	36	17.12.2018	36
26.02.2018	24	20.07.2018	18	27.12.2018	18
12.03.2018	12	30.07.2018	24	14.01.2019	6
26.03.2018	18	06.08.2018	30	21.01.2019	6
02.04.2018	6	20.08.2018	30	11.02.2019	18
09.04.2018	30	03.09.2018	36	04.03.2019	18
23.04.2018	18	10.09.2018	12	11.03.2019	6
07.05.2018	60	15.10.2018	6	-	-
14.05.2018	60	29.10.2018	18	-	-

Источник: составлено автором вместе с соавторами [61].

Предварительная подготовка данных (причесывание данных). Одним из нарушений основной предпосылки (подробнее будет в следующей главе) является то, что универсам преждевременно пополняет запас продукции, не дожидаясь окончания запасов, например, осуществляя две покупки (t_i, y_i) и (t_{i+1}, y_{i+1}) подряд или через очень малый период времени. Это приводит к тому, что за малый период времени будет предполагаться большая скорость расхода, что скорее всего неверно. Однако с этим очень просто бороться, достаточно объединить эти два события в одно $(t_i, y_i + y_{i+1})$, как будто сразу был сделан большой запас. Если же действительно за этот малый промежуток времени происходил очень большой расход продукции, то такое объединение соседних событий приводит лишь к усреднению расхода на этих двух интервалах, что не катастрофично. В этих данных можно увидеть, что между

первой и второй покупкой прошло только 3 дня, этот интервал кажется нам слишком коротким, объединим первую и вторую покупку в одну. Такие объединения покупок сделаем еще несколько раз, например, покупку 16.07.2018 объединяем с 20.07.2018 (не верим, что 36 бутылок продано за 4 дня). После этого восстанавливаем функцию скорости потребления бутылок кваса универсамом, рисунок 2.17. Построение такого графика является завершением третьего шага. Далее попробуем проанализировать график, построить модель и провести экстраполяцию.

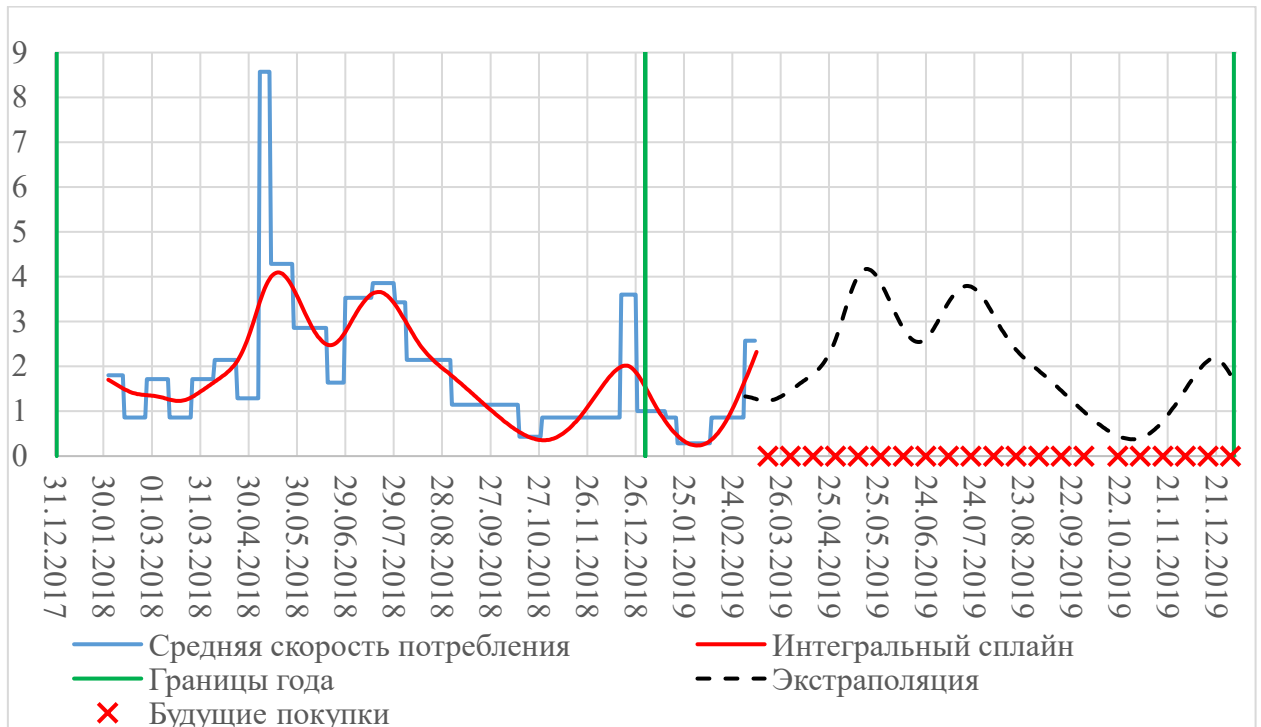


Источник: составлено автором вместе с соавторами [61].

Рисунок 2.17 – Скорость расхода кваса в универсаме, бутылок в день

По графику мы можем заметить, что наибольшее потребление кваса в универсаме происходит в период майских праздников и в середине лета, между которыми наблюдается небольшой спад. На протяжении всего лета потребление на высоком уровне, а осенью потребление минимально. Перед новогодними праздниками вновь заметен рост потребления. В результате мы видим, что скорость потребления кваса привязана к времени года, в жаркую погоду и к праздникам потребление возрастает (эти размышления ложатся в основу будущей модели). Логично предположить, что динамика потребления в будущем также будет привязана ко времени. Вместе с тем сделаем

оптимистичное предположение, что благосостояние населения повысится и спрос к концу следующего года вырастет на 10% (для того, чтобы показать возможность учитывать внешние факторы). Модель для экстраполяции значений получим с помощью переноса динамики прошлого года, которую дополнительно умножим на линейно возрастающий от 1 до 1.10 коэффициент, рисунок 2.18.



Источник: составлено автором вместе с соавторами [61].

Рисунок 2.18 - Прогноз скорости расхода кваса в универсаме, бутылок в день

Из исходных данных видим, что объем покупок кратен 6 и не превосходит 60, делаем вывод, что бутылки кваса поставляются в упаковках по 6 штук, причем не более 10 упаковок. Также видим, что интервалы между покупками почти всегда составляют 7, 14, 21, 28, 35, из них 7 и 14 встречается чаще всего. Можно предположить, что универсам проверяет запасы каждую неделю, а когда в момент проверки оказывается, что запаса не хватит до следующей проверки, универсам, ориентировочно прикидывая спрос, старается сделать двухнедельный запас, причем упаковками по 6 бутылей и не более 10 упаковок. Действуя по заданной схеме, получаем прогноз

последовательности будущих покупок, рисунок 2.18 (красные крестики на оси), таблица 2.8.

Таблица 2.8 – Прогноз будущих покупок кваса универсамом

Дата t_i	Объем y_i	Дата t_i	Объем y_i	Дата t_i	Объем y_i
18.03.2019	24	22.07.2019	48	02.12.2019	30
01.04.2019	24	05.08.2019	36	16.12.2019	30
15.04.2019	30	19.08.2019	30	30.12.2019	18
29.04.2019	48	02.09.2019	24	20.01.2020	6
13.05.2019	54	16.09.2019	18	10.02.2020	12
27.05.2019	48	30.09.2019	12	24.02.2020	24
10.06.2019	36	21.10.2019	6	09.03.2020	6
24.06.2019	42	04.11.2019	12	-	-
08.07.2019	54	18.11.2019	12	-	-

Источник: составлено автором.

Из-за того, что при планировании мы опираемся на прогноз спроса, а также предположили определенное поведение покупателя, такой прогноз покупок получился более ритмичным. Практически все покупки происходят через двухнедельные интервалы, только 3 покупки были спрогнозированы через трехнедельные интервалы. В исходных данных не было замечено строгой ритмичности, такое ощущение, что универсам выбирал объем покупок, не имея конкретного представления о будущем спросе, что было сделано у нас. С одной стороны, можно предложить (навязать) покупателю (универсаму) действовать в соответствии с нашим прогнозом его расхода продукции, что будет выгодно самому покупателю. С другой стороны, покупатель может действовать автономно и определять объем покупок, не опираясь ни на какие прогнозы помимо внутреннего чутья. В этом случае мы можем с уверенностью определить только дату самой первой следующей покупки как момент завершения запаса, сделанного последней покупкой, но не сможем определить последовательность будущих покупок (не зная объем будущей покупки нельзя будет сказать, когда запас продукции, пополненный этой покупкой, вновь закончится).

Такой проблемы не возникает при прогнозировании услуг, когда не нужно определять объем покупок.

Пример на реальных данных, сфера услуг, стрижка

Имеются данные о датах стрижки одного постоянного клиента, который непрерывно стрижется на протяжении 4 лет в одной и той же парикмахерской, располагающейся в г. Красногорске Московской области, таблица 2.9.

Таблица 2.9 – Данные стрижки

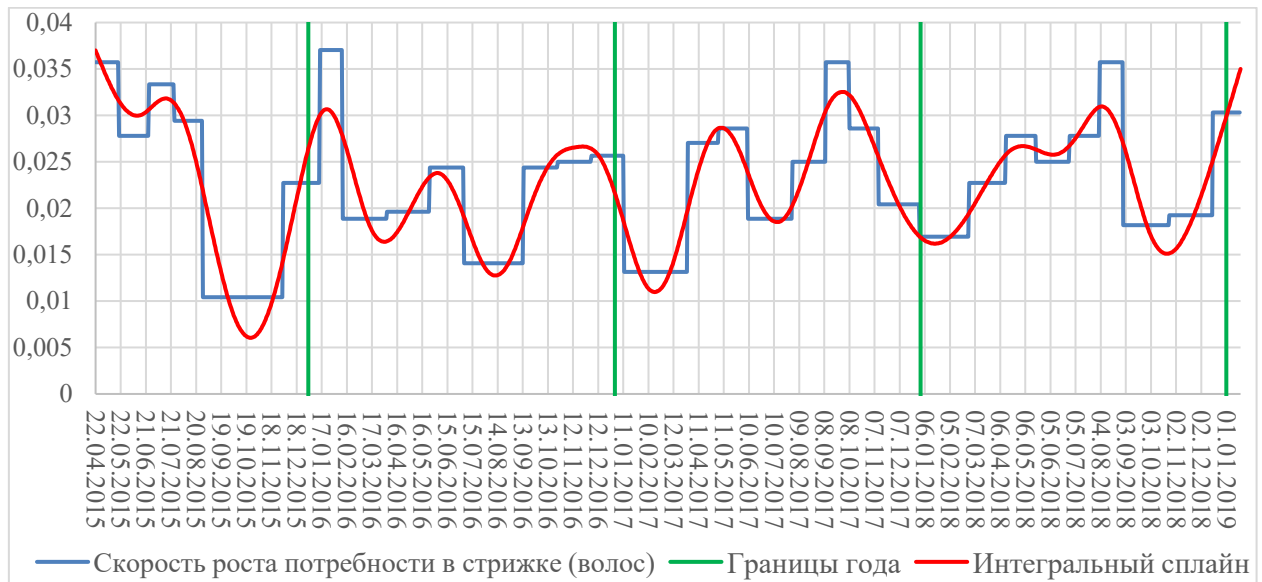
i	Дата t_i	i	Дата t_i	i	Дата t_i
1	22.04.2015	12	14.09.2016	23	31.12.2017
2	20.05.2015	13	25.10.2016	24	28.02.2018
3	25.06.2015	14	04.12.2016	25	13.04.2018
4	25.07.2015	15	12.01.2017	26	19.05.2018
5	28.08.2015	16	29.03.2017	27	28.06.2018
6	02.12.2015	17	05.05.2017	28	03.08.2018
7	15.01.2016	18	09.06.2017	29	31.08.2018
8	11.02.2016	19	01.08.2017	30	25.10.2018
9	04.04.2016	20	10.09.2017	31	16.12.2018
10	25.05.2016	21	08.10.2017	32	18.01.2019
11	05.07.2016	22	12.11.2017	-	-

Источник: составлено автором вместе с соавторами [62].

Шаг 1 автоматически выполнен, так как мы анализируем данные только одного клиента. Шаг 2, предполагаем, что процесс образования событий соответствует процессу накопления воздействия до определенного уровня, и моделируем его как емкость. В отличие от торговли, где величины u_i являются известными объемами покупок, при приобретении услуг величина u_i часто не измеряется. В этом случае в качестве величины u_i можно брать единицу, обозначающую сам факт приобретения услуги. Прогнозирование упрощается, так как не нужно определять будущие значения u_i , как было сделано при прогнозировании объемов покупок в торговле. Если же объем каждой приобретенной услуги по-прежнему выражен некоторым числом, то придется отдельно прогнозировать и его.

Шаг 3, при условии, что объем приобретенной услуги всегда равен единице $u_i = 1$, восстанавливаем функцию $f(t)$, рисунок 2.19, которую можно физически проинтерпретировать как относительную скорость роста волос. Относительная она от того, что выражена не в единицах длины за день,

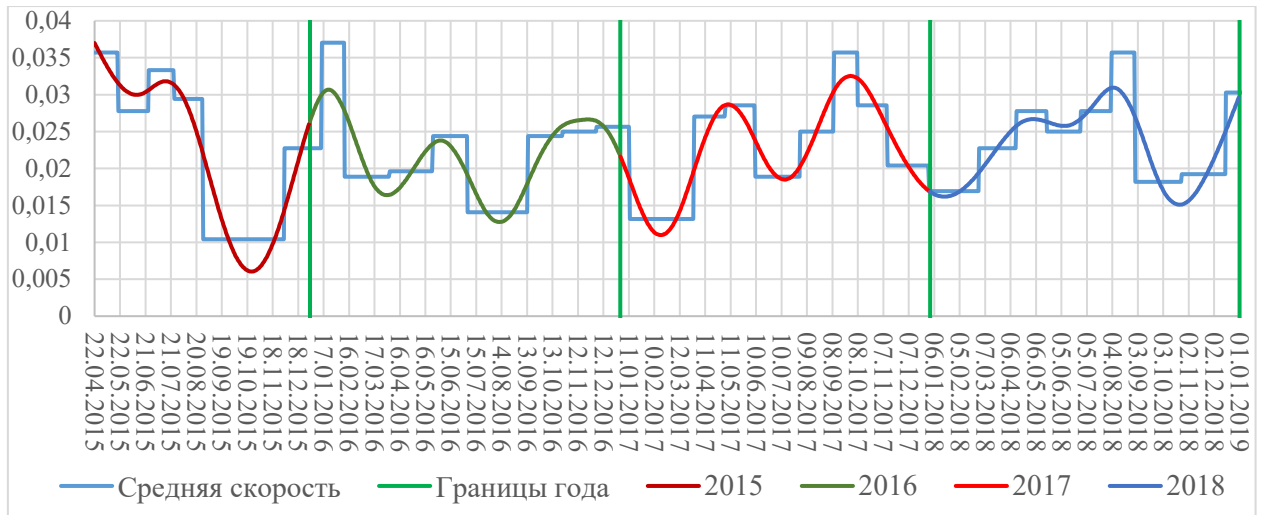
а в долях ко всей услуге. То есть значение точек этой функции показывает, насколько мы приближаемся каждый день, чтобы вновь приобрести услугу. Также можно дать еще одну интерпретацию: - как скорость роста желания человека подстричься.



Источник: составлено автором вместе с соавторами [62].

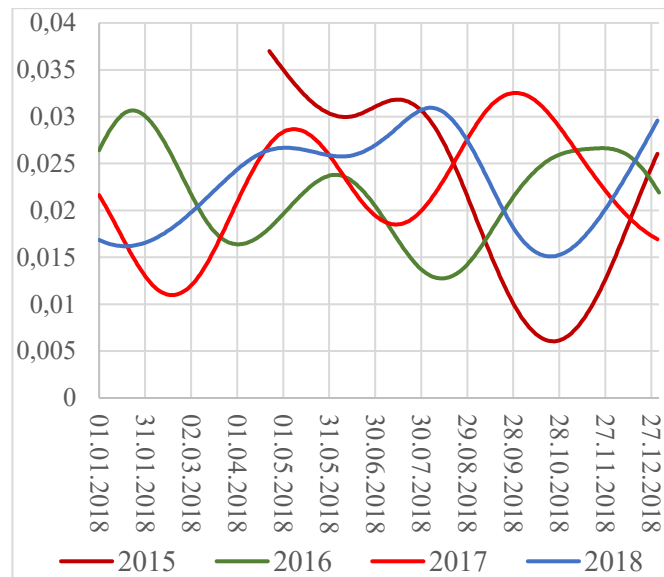
Рисунок 2.19 – Относительная скорость роста волос, долей в день

Далее переходим к этапу поиска закономерности и экстраполяции (Шаг 4). На этом графике можно заметить некоторую закономерность, в последних двух годах функция имеет два пика, в конце весны и в начале осени, а в середине лета заметен провал. Из открытых источников [72; 73] можно прочесть, что скорость роста волос зависит от температуры и времени года, чем выше температура, тем больше активность организма, и, как следствие, выше скорость роста волос. Изначально было желание предположить, что скорость роста волос должна жестко зависеть от времени года, однако из рисунка 2.19 этого не следует, вершины пиков сдвинуты по времени, расстояние между двумя этими пиками различное, имеется провал в летние дни. Это хорошо видно, если для каждого года мы перекрасим функцию в свой собственный цвет, рисунок 2.20, и наложим друг на друга, рисунок 2.21.



Источник: составлено автором.

Рисунок 2.20 – Перекрасили функцию скорости роста волос на каждом году в свой собственный цвет

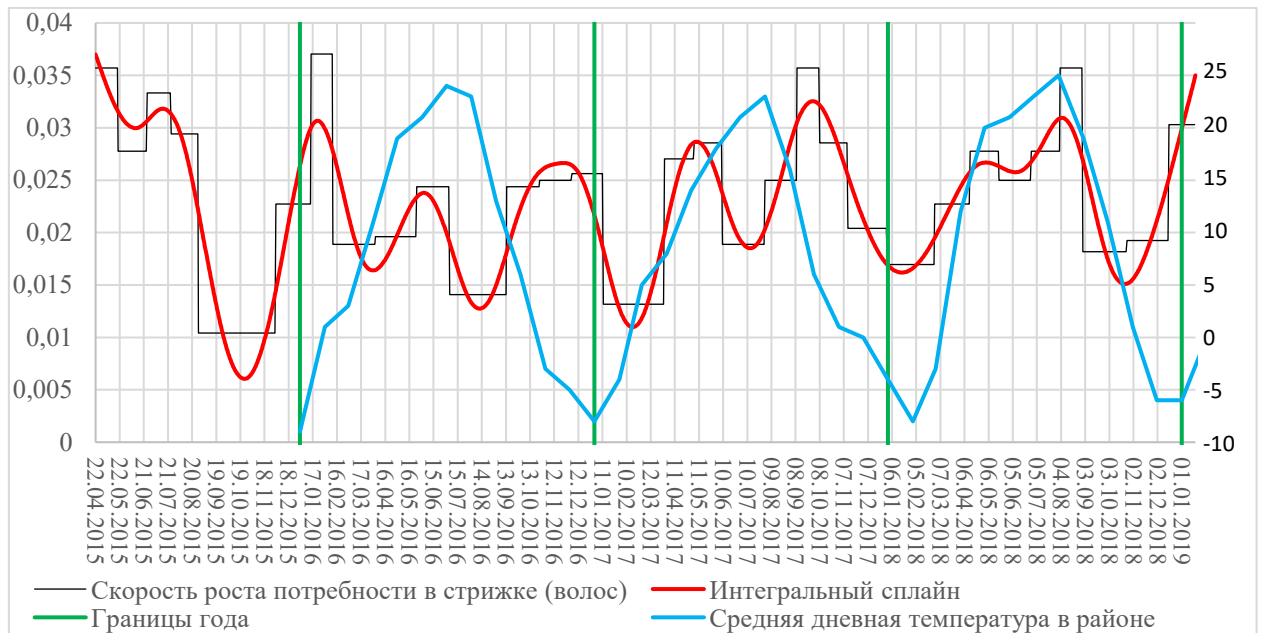


Источник: составлено автором.

Рисунок 2.21 - Сравнение функции скорости роста волос на разных годах

Пики в мае 2016 и 2017 года мало сдвинуты относительно друг друга, а вот пики в конце лета и начале осени 2016, 2017 и 2018 года сильно сдвинуты. Причем на разных годах расстояние между весенними и осенними пиками также сильно различается. Быть может, скорость роста волос привязана не к времени года, а именно к температуре, которая могла из-за погодных аномалий быть разной (в случае если было холодное лето). Однако после того, как удалось найти данные о средней дневной температуре в районе [74] и наложить график средней температуры на график полученной функции

скорости роста волос, рисунок 2.22, видно, что это утверждение также не выполняется (в 2017 году в максимально жаркие дни заметен провал).



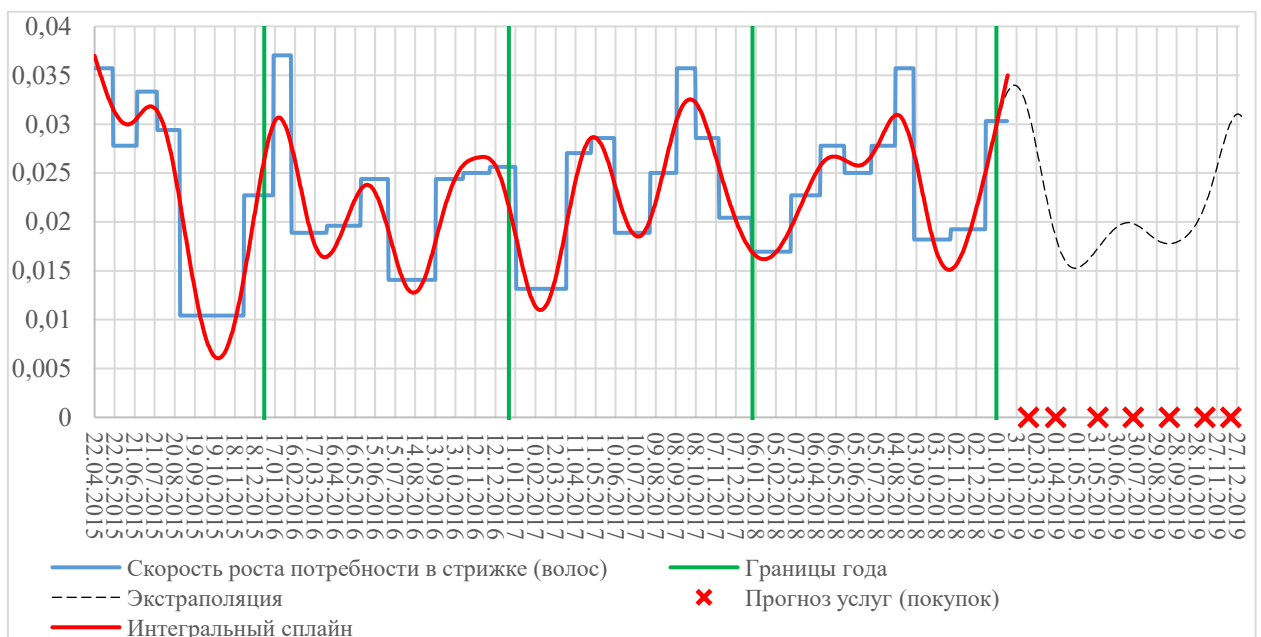
Источник: составлено автором.

Рисунок 2.22 – Сравнение со средней температурой

Клиент как будто наоборот в жаркие месяцы испытывал меньшую потребность в стрижке, больше похоже, что в переходные периоды весной и перед осенними похолоданиями скорость роста волос возрастала. Конечно, вполне возможно, что в жаркие летние месяцы клиент находился в отпуске и отдыхал по туристической путевке в другой стране, и, возможно, он очередной раз стригся в это время в местной парикмахерской, в результате чего в исходном наборе данных были пропущены некоторые наблюдения. Если бы такая информация была, то можно было бы скорректировать данные, назначив соответствующий объем услуги $y_i = 2$ (а не 1), т.е. за время от t_i до t_{i+1} произведено две стрижки. Однако такой информации у нас нет. К тому же в последний 2018 год в летние месяцы интервалы времени между стрижками отличаются не так сильно, чтобы можно было предположить, что пропущена еще одна услуга, но небольшой спад летом все равно заметен.

Таким образом, из-за недостатка информации остается сделать вывод, что скорость роста волос данного клиента не привязана ко времени года или температуре, как утверждается в открытых источниках, а скорость роста волос

больше зависит от внутреннего состояния человека, например, от изменения уровня гормонов и других факторов. Так как наблюдается некоторый колебательный процесс, можно предположить, что и эти факторы, от которых зависит скорость роста волос, имеют характер колебательного процесса. Можно предположить, что наблюдаемые изменения скорости роста волос имеют структуру, состоящую из периодически повторяющихся гармонических функций. Используя это предположение, с помощью того же алгоритма Куинна и Фернандеса (Quinn-Fernandes algorithm) [69; 70] можно попробовать представить функцию скорости роста волос в виде суммы ограниченного числа гармонических колебаний, после чего произвести экстраполяцию на год вперед, рисунок 2.23. Пунктирной линией изображен график экстраполированных значений функции скорости роста волос на год вперед (при модели в виде суммы гармонических функций). Дополнительно отмечу, что модели для экстраполяции можно строить различными способами, необходимы специальные исследования в каждой прикладной области (не берусь утверждать, что построенный график экстраполированных значений функции будет единственно верным, так как не являюсь специалистом в области физиологии роста волос).



Источник: составлено автором вместе с соавторами [62].

Рисунок 2.23 – Экстраполяция скорости роста волос и прогноз услуг, долей в день

Далее запускаем процесс накопления воздействия и получаем прогноз будущих событий (прогноз даты следующего обращения клиента за услугой) следующим образом. С самой последней даты приобретения клиентом услуги начинаем от нуля суммировать значения функции. Как только накопленное воздействие становится больше единицы, говорим, что клиент в эту дату вновь захочет обратиться за приобретением услуги, и в эту дату сбрасываем накопленное воздействие на ноль, и снова начинаем прибавлять значение функции, пока не получим дату очередной покупки (можно наоборот от единицы производить вычитание, как если бы происходил обычный процесс потребления из раздела систем управления запасами, но для покупателя услуги, а не для нас). Прогноз будущих стрижек нашего клиента изображен крестиками на оси времени, рисунок 2.23, и приведен в таблице 2.10.

Таблица 2.10 – Прогноз будущих моментов времени приобретения услуг

Прогноз услуг	
18.02.2019	17.09.2019
31.03.2019	09.11.2019
02.06.2019	18.12.2019
25.07.2019	-

Источник: составлено автором.

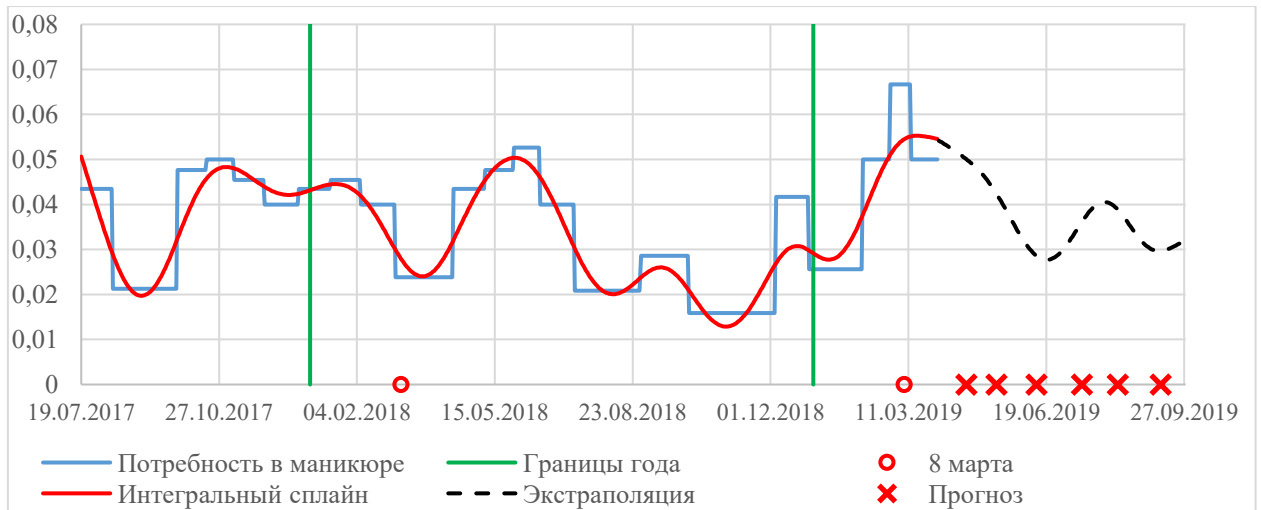
Пример на реальных данных, сфера услуг, маникюр

Имеется набор данных о маникюре постоянного клиента, который обращался в салон красоты на протяжении 2 лет, таблица 2.11. Используя описанный выше подход, также осуществляя поиск закономерности в виде гармонических функций, мы получим прогноз будущих обращений клиента за услугой, рисунок 2.24.

Таблица 2.11 – Данные о маникюре постоянного клиента

Клиент					
1	19.07.2017	9	07.02.2018	17	03.10.2018
2	11.08.2017	10	04.03.2018	18	05.12.2018
3	27.09.2017	11	15.04.2018	19	29.12.2018
4	18.10.2017	12	08.05.2018	20	06.02.2019
5	07.11.2017	13	29.05.2018	21	26.02.2019
6	29.11.2017	14	17.06.2018	22	13.03.2019
7	24.12.2017	15	12.07.2018	23	02.04.2019
8	16.01.2018	16	29.08.2018	-	-

Источник: составлено автором вместе с соавторами [62].



Источник: составлено автором вместе с соавторами [62].

Рисунок 2.24 – Потребность в маникюре и прогноз услуг, долей в день

Восстановленную из редких событий функцию можно проинтерпретировать как относительную скорость роста ногтей, опять же, выражена она не в миллиметрах, а в долях от всей услуги. По-другому можно проинтерпретировать эту функцию как скорость роста потребности в маникюре. У данного клиента можно заметить большую потребность в маникюре перед праздниками (8 марта и 1 января), а после окончания праздников потребность в маникюре падает, однако у других клиентов может быть совсем другая динамика потребности в услуге. Если получится определить закономерности в скорости роста потребности для всех клиентов, можно предсказывать будущие моменты обращения каждого клиента, определять общее число обращений за месяц, оптимизировать график работы мастеров этого салона красоты и т.д.

Опять повторюсь, что не берусь утверждать, что выбранный способ построения модели и экстраполяции восстановленной функции в примерах со стрижкой и маникюром является единственно правильным. Чтобы построить правильную модель для экстраполяции значений необходимо быть специалистом в соответствующей прикладной области, разбираться от чего может зависеть динамика наблюдаемого процесса. Я не работаю в парикмахерской, не являюсь аналитиком салона красоты. Представленные здесь примеры носят больше демонстративный характер, детали оставляю на

исследователей, которые могут заинтересоваться данным методом при проведении собственных исследований. Других математических методов, которые также подробно могли анализировать и предсказывать редкие события с точки зрения динамики процессов, образующих эти события, вообще нет (не встречались, хотя был проведен интенсивный поиск).

В приложении Б представлены дополнительные примеры анализа данных редких событий еще нескольких клиентов салона красоты, пример прогнозирования услуг связи, а также примеры из области финансов.

2.3 Формулирование условий, предпосылок, допущений

В данном параграфе мы сформулируем условия, которые должны соблюдаться для того, чтобы можно было корректно применять предложенный емкостный метод анализа и прогнозирования редких событий. Вначале мы сформулируем общие условия, которые должны выполняться для любой прикладной области, а после рассмотрим несколько примеров из некоторых прикладных областей и попробуем описать те допущения и предпосылки, которые должны для них соблюдаться.

Напомним, что в основе предложенного емкостного метода, когда мы анализируем и прогнозируем редкие события, мы предполагаем, что в источниках образования этих редких событий происходят процессы, схожие с опустошением/наполнением емкости. Причем величина совершенного события y_i есть интеграл от исходной неизвестной функции $f(t)$ скорости изменения уровня некоторой емкости за время от момента совершения этого события t_i до момента времени совершения следующего события t_{i+1} , то есть $y_i = \int_{t_i}^{t_{i+1}} f(t)dt$ (однако наблюдаемые величины содержат дополнительную погрешность).

Общие условия применения емкостного метода

Общими условиями применения емкостного метода будут являться следующие:

1) Прежде всего редкие события действительно должны образовываться с помощью процессов, схожих с процессами опустошения/наполнения емкости. Если такое не будет выполняться, то описанный механизм восстановления параметров процесса по наблюдаемым интегралам будет не подходить. Понадобятся другие механизмы исследования процессов. Некоторые такие механизмы будут предложены в пятой главе.

2) Процесс образования событий не должен меняться со временем, то есть не должно быть такого, чтобы часть времени события образуются по одному процессу, а часть - по совсем другому. Если такое не будет соблюдаться, то придется многократно устанавливать вид процесса и определять условия, при которых вид процесса меняется.

3) В процессе образования событий не должно возникать изменений в количестве параметров или изменение способов влияния этих параметров на ход формирования событий. Данное условие является частным случаем предыдущего условия, с тем отличием, что общий вид процесса остается неизменным, но меняются некоторые важные детали механизма формирования событий. Если же такие изменения происходят, то придется каким-то образом определять, что конкретно поменялось в механизме образования событий. Например, в процессах потребления и восполнения запаса не должно происходить изменение страхового уровня запасов товаров, желательно, чтобы страховой уровень запасов был постоянным (иначе в размере покупок будет откладываться разница изменения страхового запаса, что приведет к дополнительной погрешности). Если же страховой запас меняется, то придется восстанавливать механизм, по которому меняется этот страховой запас.

4) Закономерности изменений параметров не должны зависеть от ненаблюдаемых (или не интересующих нас) факторов, иначе восстановление закономерностей параметров должно происходить с учетом этих ненаблюдаемых (не интересующих нас) факторов. Иначе восстановление закономерностей может быть затруднительным или восстановленная

закономерность может содержать большую неопределенность (т.е. разброс), которую ничем кроме как случайностью назвать не получится (напоминаю, что наши старания были направлены на то, чтобы уменьшить неопределенность с помощью использования всей возможной информации), в итоге, определяемые закономерности будут содержать большую статистическую погрешность.

5) (Формальное необязательное условие) Исходная функция $f(t)$ скорости изменения уровня некоторой емкости (например запаса) должна быть определена на всем рассматриваемом промежутке времени и не должна быть импульсной. Иначе на самом деле вместо функции скорости будет восстанавливаться функция интенсивности¹⁾, которая будет показывать среднюю скорость изменения уровня за единицу времени. На самом деле данное ограничение редко выполняется на практике. При том же процессе потребления, например, при потреблении хлеба, потребление происходит дискретными порциями (мы непрерывно не высасываем содержимое буханки хлеба через трубочку). Данное условие внесится с той целью, чтобы исследователь, который воспользуется предложенным методом, понимал, что восстанавливаемые зависимости будут иметь непрерывный характер, когда на самом деле они могли быть дискретными. Однако существуют примеры, для которых данное ограничение действительно является важным. Например, в случае продажи билетов только перед очередным ежегодным чемпионатом метод восстановит функцию скорости продажи билетов как усредненную за годовой временной интервал. Другими словами, для редких событий, привязанных к определенным датам, такой метод не подходит (события не образуются процессом).

б) Значения y_i в анализируемых данных (t_i, y_i) должны иметь характер интегралов от функции $f(t)$ за период времени от t_i до t_{i+1}

¹⁾ Средняя скорость $f(t) = \lim_{\Delta t \rightarrow 0} E[X(t, t + \Delta t)]/\Delta t$, где $X(t, t + \Delta t)$ – изменение уровня за интервал времени от t до $t + \Delta t$, а $E(\cdot)$ – математическое ожидание случайной величины.

($y_i = \int_{t_i}^{t_{i+1}} f(t)dt$), т.е. отнесенных к началу интервала t_i . Если значение y_i относится к концу интервала, т.е. $y_i = \int_{t_{i-1}}^{t_i} f(t)dt$, то может произойти сдвиг в данных на одну позицию, что однако легко скорректировать сдвигом в обратную сторону, достаточно рассматривать данные (t_i, y_{i+1}) . Такое может происходить в следующем примере, например, когда у клиента выходят из строя некоторые однотипные электроприборы, и когда количество поломок превысит определенную величину, делается заказ на покупку соответствующего количества приборов. В этом случае наблюдаемое количество y_i показывает количество поломок начиная с предыдущего периода до текущего, а не с текущего до будущего (вот если бы наш клиент делал запас электроприборов, расходовал его, и делал очередной запас, когда этот запас закончится, то тогда ничего менять не надо). В случае адаптации предложенного метода к сдвинутым данным, мы получаем метод оценки нестационарной интенсивности потока событий по эмпирическим данным. Способов оценки нестационарной интенсивности по эмпирическим данным достаточно много, некоторые из них описаны в следующих работах [75; 76; 77; 78]. Однако очень редко эти методы могут применяться к исходной задаче, когда требуется определить скорость расхода запаса. Аналогично не всегда получается использовать метод определения скорости расхода запаса для определения интенсивности потока событий (интенсивность потока событий и скорость расхода запаса в этом контексте являются разными явлениями, относящимися к разным задачам).

7) Время следующего события t_{i+1} должно соответствовать моменту, когда полностью опустошилась/наполнилась емкость (опустошился запас или переполнилось терпение) без существенной задержки. То есть не должно быть периодичности проверки, когда момент наступления события может быть пропущен (перескочен) и только в момент проверки будет отмечено время следующего события t_{i+1} , что в результате вызовет задержку срабатывания события. Данная задержка при образовании события приведет к

дополнительной погрешности восстановления исходной функции $f(t)$. На самом деле в большинстве практических задач именно так и будет происходить, да и сама дискретизация времени по дням приводит к тому, что событие как бы возникает только в начале (или конце) дня, но никогда не в середине, первой четверти или второй трети этого дня (т.е. время округляют с точностью до дня). Данное условие введено для того, чтобы у исследователя, использующего данный метод, было понимание того, что точность фиксации моментов возникновения событий влияет на точность восстановления исходных зависимостей. Например, если рассматриваются события, про которые известно то, что погрешность фиксации дат (период проверки) составляет несколько недель, то точность восстановления исходной функции и прогноз будущих событий будут не очень точными.

На данный момент сформулировано 7 общих условий, однако вполне вероятно, что в некоторых определенных прикладных областях необходимо добавить дополнительные условия. Также возможно, что спустя время появится необходимость добавить больше условий, общих для любой прикладной области (возможно я пропустил и не обратил внимание на некоторые аспекты, которые достаточно важны для упоминания, надеюсь, что другие исследователи внесут соответствующие дополнения). Тем не менее введенные условия либо обязательны для применения предлагаемого метода, либо носят рекомендательный характер, и нарушение которых лишь может приводить к дополнительным погрешностям, однако сам метод еще остается работоспособным.

Далее рассмотрим несколько примеров из разных прикладных областей и дадим формулировки наших условий в терминах соответствующей прикладной области, чтобы можно было использовать емкостный метод для анализа и прогнозирования редких событий.

Пример предпосылок, розничная торговля

Рассмотрим пример¹⁾ розничной торговли, когда клиенты покупают продукцию у выбранного продавца. Метод же будет применяться на стороне продавца на основе имеющихся у него данных дискретных продаж своим клиентам. Предпосылки использования методы будут следующими:

1) Ненаблюдаемый запас каждого покупателя рассматривается как емкость, которая опустошается с разной скоростью, зависящей от времени. У каждого клиента своя скорость расхода запаса.

2) Как именно расходует свой запас конкретный клиент не рассматривается, предполагается лишь то, что этот расход должен описываться некоторой ненаблюдаемой функцией, которая имеет фиксированную закономерность. На самом деле клиент может не просто потреблять продукцию, а может реализовывать ее другим клиентам (в случае если наш клиент является промежуточным распространителем). Неважно, как именно происходит расход запаса клиента, лишь бы этот расход выражался некоторой закономерностью, которая в дальнейшем будет соблюдаться. Этим предположением мы ограничиваем факторы, от которых может зависеть восстанавливаемая функция для того, чтобы в дальнейшем успешно экстраполировать ее на будущее.

3) Благодаря предыдущему предположению мы можем ввести некоторые ограничения на поведение клиента: 1) считаем, что покупатели не избавляются от купленной продукции, т.е. запас убывает только в соответствии с потреблением; 2) запасов продукции у клиента всегда достаточно для удовлетворения потребности потребления, т.к. покупатели вовремя пополняют запасы (чтобы не было периодов времени, когда потребление задано, а продукция отсутствует). Эти предположения в целом соблюдаются, так как клиент не хочет лишней раз выбрасывать продукцию, за

¹⁾ Некоторые результаты исследования получены совместно с исполнителями НИР – Головановой П.С. и Кострицей Т.А., которые на момент начала НИР являлись студентами автора данной диссертационной работы, в задачи которых входил поиск данных и первичная обработка.

которую он заплатил, и так как клиенту всегда хочется, чтобы его потребности удовлетворялись. Эти условия нужны для того, чтобы сторонние факторы не влияли на то, как восстанавливается функция скорости расхода запаса. Если же сторонние факторы влияют, то это влияние должно подчиняться общей закономерности, восстанавливаемой из данных.

4) У каждого клиента может быть собственный максимальный запас, критический уровень и другие параметры, но эти параметры обязательно должны успешно восстанавливаться из имеющихся данных (например, как в параграфе 2.2). Конечно же восстановление параметров может быть неточным, и параметры могут обладать некоторой статистической погрешностью. Важна сама принципиальная возможность определения этих параметров. В то же время некоторые параметры могут быть непостоянными, однако должна быть возможность восстановить механизм изменения параметров из имеющихся данных.

5) Пополнение запасов клиентов (покупки) обязательно происходят с фиксацией момента времени (даты), объема покупки и некоторого идентификатора покупателя, позволяющего различать одного клиента от другого.

б) Покупатели преждевременно не пополняют запасы пока они не закончились (опустились ниже критического уровня). Например, не должно быть двух покупок подряд (например, когда за одну поездку вся продукция не влезла в кузов автомобиля, и покупатель вернулся для совершения второй покупки). Это условие необходимо для того, чтобы время расходования запасов продукции потребителя не фиксировалось преждевременно. Если в короткий промежуток покупатель дважды совершил покупки (t_i, y_i) и (t_{i+1}, y_{i+1}) , то получится так, что объем продукции y_i был израсходован за малое время $t_{i+1} - t_i$, что будет ошибочно (большой всплеск на этом периоде). В этом случае надо объединить обе покупки в одну и привязать к первому моменту времени $(t_i, y_i + y_{i+1})$.

7) Клиенты пополняют запасы только у нашего продавца, либо то, как они выбирают альтернативных продавцов, должно отображаться в закономерностях восстанавливаемой функции. Это действительно важное условие. Если оно не выполняется, то у нашего продавца будет только часть информации. Отсутствие информации сильно повлияет на восстановление исходной функции. Получается так, что собирать всю информацию возможно только от постоянных клиентов, которые не меняют своего поставщика продукции.

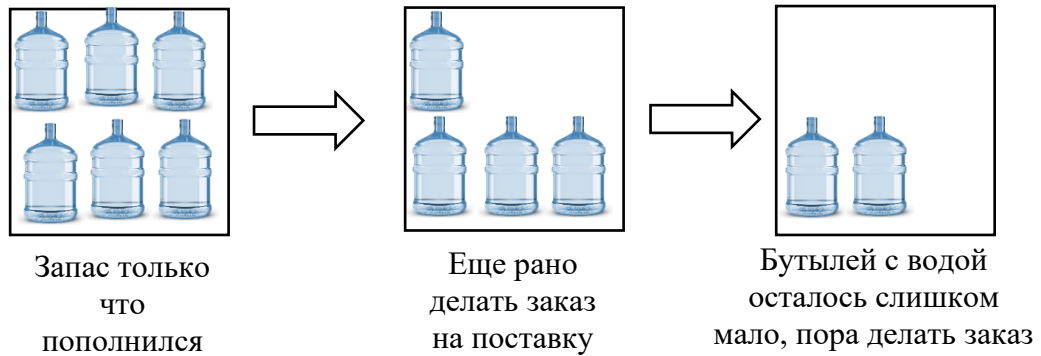
Однако, в то же время можно предположить, что различные продавцы (магазины) будут обмениваться информацией и передавать друг другу данные о покупках своих клиентов, чтобы эти данные можно было вновь соединить. Конечно же мало вероятно, что продавцы будут передавать информацию своим конкурентам. Полагаю, если цифровизация разовьется еще больше, когда все потоки информации от разных конкурирующих продавцов (сервисов) будут сливаться в один независимый центр прогнозирования, то в этом центре будет возможность анализировать действия каждого человека и предсказывать его будущие потребности. Это, конечно, звучит как в фантастике про антиутопию, что нас всех «посчитают» и о наших предпочтениях будет абсолютно все известно (даже когда мы никому явно ничего не сообщали), но на самом деле нынешний мир в одном шаге от этого (достаточно точкам обмениваться данными друг с другом, либо добровольно, либо из-за некоторого строгого регламента). Возможно, где-то это даже сейчас и происходит, особенно когда наблюдается тенденция, когда независимые магазины вытесняются крупными торговыми сетями, и вполне может получиться так, что в одном районе останется только одна торговая сеть, которая будет собирать информацию о своих клиентах в разных торговых точках. С одной стороны, это антиутопия, когда все наши действия будут полностью контролироваться, с другой стороны это есть развитие цифровой экономики, которая может дать огромные преимущества как производителям, так и потребителям, например, продавцы будут вовремя звонить и

напоминать, что пора бы пополнить запасы нашего любимого товара именно у них. Получается, как будто плановая экономика, но снизу-вверх.

Пример предпосылок, поставка бутилированной воды

Рассмотрим пример, в котором поставки осуществляются по заранее заключенному контракту, примером будет поставка бутилированной воды в крупную организацию. Данный пример интересен с той точки зрения, как изменяются предпосылки при изменении прикладной области.

Практически в каждом современном кабинете любой организации установлен кулер с питьевой водой. Бутыли с водой организация закупает у одной из множества компаний. Из-за того, что многие компании предоставляют скидку, если в момент приема новой партии бутылей сдается тара израсходованных бутылей, организации мотивированы заключать долгосрочный договор на поставку питьевой воды с одной компанией. Когда запас подходит к концу, организация сообщает необходимое количество бутылей воды и количество сдаваемых пустых. В момент замены бутылей t_i фургон компании приезжает, доставляет наполненные бутылки и забирает пустые. Можно предположить, что количество привезенных новых бутылей u_i примерно равно количеству сдаваемых пустых, но в общем случае это не обязательно. В простейшем случае процесс поставки бутилированной воды соответствует схеме, изображенной на рисунке 2.25. Используя емкостный метод, можно определить зависимость потребности в воде организации со временем, после чего спрогнозировать следующие моменты наступления событий, когда организация вновь будет делать заказ. Сама по себе выявленная закономерность может иметь особый интерес, она может показывать, как расход воды связан с временем года, температурой, или загруженностью сотрудников работой.



Источник: составлено автором вместе с соавторами [63].
 Рисунок 2.25 – Процесс формирования заказа на бутылки с водой

При рассмотрении этого примера можно выделить следующие условия для корректной работы метода.

1) Количество кулеров не должно меняться или, если оно меняется, то как минимум у всех страдающих от жажды людей должна быть возможность за приемлемое время дойти до другого работающего кулера. Если это условие не выполняется, то получается, что часть людей будет пользоваться другими источниками воды или страдать от жажды. В этом случае расход питьевой воды зависит не от жажды сотрудников, а от других факторов, например, от поломок кулеров. Тогда восстановленная зависимость будет показывать изменения в потреблении воды не от времени (или постоянных факторов), а от других случайных (или внешних) факторов, из-за которых выходят из строя кулеры. Это конечно тоже может иметь определенный интерес, но тогда при прогнозировании надо будет указать, какими будут эти самые дополнительные факторы.

2) Среднее количество сотрудников в организации желательно должно оставаться постоянным. Если это условие нарушается, то расход питьевой воды, который конечно же зависит от количества людей, будет также изменяться. В этом случае восстановленная функция будет содержать эти изменения, а не только зависеть от времени года, температуры или загруженности сотрудников. При экстраполяции необходимо, чтобы мы также имели представление, как будет меняться количество людей в организации.

3) Критический уровень запаса бутылей не должен меняться. Если это предположение не выполняется, то появляется дополнительная погрешность. Получается, что когда из запаса берется и откладывается в запас разное количество бутылей, то между событиями расходуется разное количество, изменение которого связано с другими факторами. Обратим внимание на то, что максимальный уровень запаса может меняться, и восстановление скорости расхода воды по-прежнему будет корректным (однако для прогноза последовательности будущих событий потребуется модель изменения максимального запаса).

Пример предпосылок, замена неисправных кондиционеров

Данный пример¹⁾ интересен тем, что исходный процесс образования событий связан со случайным образованием поломок в устройствах. Конечно же мы будем не в состоянии определить, когда произойдет следующая поломка, но мы сможем приблизительно определить, когда произойдет серия поломок, способствующих замене сразу многих поломок. Этот пример показывает, как можно с помощью предложенного метода анализировать потоки событий из теории случайных процессов.

Нормативный срок службы кондиционеров по ГОСТу 26963-86 от 10 лет, но все равно получается так, что некоторые кондиционеры выходят из строя раньше положенного срока. Поломки кондиционеров в большой компании можно моделировать как Пуассоновский поток, в котором единственным параметром будет являться интенсивность поломок λ , которую можно при отсутствии статистических данных выразить как $\lambda = N/\text{срок службы}$, где N общее количество установленных устройств. Однако очевидно, что количество поломок должно зависеть от условий эксплуатации, и в жаркую погоду, когда устройство работает намного чаще, интенсивность поломок должна быть значительно выше, чем во время

¹⁾ Некоторые результаты исследования получены совместно с исполнителями НИР – Головановой П.С. и Кострицей Т.А., которые на момент начала НИР были моими студентами, в задачи которых входил поиск данных и первичная обработка.

холодной погоды. Другими словами, интенсивность должна быть функцией от времени $\lambda(t)$. Однако, как правило, каждую поломку отдельно не фиксируют, то есть точно сказать, когда конкретное устройство вышло из строя, не получается. Допустим, что количество неисправных устройств определяется тогда, когда их количество превысит некоторую пороговую границу, когда уже проблему нельзя будет откладывать (условия работы становятся невыносимыми) или когда проходит запланированная инвентаризация. В этот момент компания делает заказ на установку новых кондиционеров, предполагается, что из-за одного неисправного кондиционера компания обращаться не будет. В результате что у компании, что у организации, занимающейся обслуживанием кондиционеров, имеется набор данных о датах t_i обращения за заменой некоторого количества y_i неисправного оборудования. Из этого набора данных можно определить с какой скоростью происходили поломки в разные периоды времени.

Обратим внимание в данном примере на то, что величина y_i , привязанная к моменту времени t_i , на самом деле показывает количество поломок за время от t_{i-1} до t_i (а не от t_i до t_{i+1} , как в основном предположении метода). Чтобы применить предлагаемый метод мы должны сдвинуть данные на одну позицию и анализировать данные (t_i, y_{i+1}) (чтобы не нужно было бы делать сдвиг должно выполняться условие, что величина y_i является количеством запасных устройств, а события должны образовываться тогда, когда этот запас заканчивается). После такого сдвига при применении метода мы сможем получить оценку интенсивности потока поломок устройств, данный метод будет предпочтительнее обычных методов сглаживания, таких как скользящее среднее или экспоненциальное сглаживание. После того как будет восстановлена эта интенсивность (или скорость образования поломок) можно выявить закономерности ее изменения от времени или других факторов. Из-за того, что процесс выхода оборудования из строя можно моделировать как Пуассоновский процесс, для дальнейшего анализа лучше использовать теорию случайных процессов.

Можно экстраполировать найденные зависимости на будущее и определить ожидаемое количество поломок к следующей проверке или определить ожидаемое время, когда поломок будет заданное количество.

Тем не менее для того, чтобы мы могли правильно экстраполировать найденные зависимости, необходимо выполнение следующих требований.

1) Площади компании-клиента не должны увеличиваться (и уменьшаться) в размере, в противном случае выявленная закономерность будет показывать не только рост количества неисправных кондиционеров, но также и рост новых площадей, для которых необходимо установить кондиционеры.

2) Самое первое обращение компании-клиента следует исключить из выборки, т.к. это обращение подразумевает установку всех кондиционеров (а не их замену).

3) Сломавшиеся кондиционеры желательно должны заменяться на такие же модели (либо незначительно усовершенствованные). В противном случае, если кондиционеры будут заменяться на принципиально новые модели, то выявленная зависимость будет показывать скорость поломок то старых, то новых устройств. В этом случае надо разделять поломки старых и новых моделей устройств, иначе произойдет перемешивание потоков событий.

4) Теплоизоляция фасада здания не должна меняться, иначе климатические условия в помещениях изменятся, и тогда настройки кондиционера будут подогнаны под новые значения (например, согласно новым настройкам, у всех кондиционеров изменится скорость обдува), а это значит, что его комплектующие будут изнашиваться по-другому. То есть выявленная закономерность будет показывать не только зависимость от времени, а зависимость от таких факторов, как изменение теплоизоляции.

5) Аналогично, окна и их количество в здании также не должны меняться (исключение: замена окон на такие же по коэффициенту сохранения тепла).

Приведенные в этих трех примерах условия являются лишь частью условий, которые удалось заметить, а не полным исчерпывающим перечнем. Возможно, опытный исследователь сможет заметить и другие условия, которые необходимы для корректного восстановления исходной зависимости, ее экстраполяции и прогнозирования будущих событий. На основе этих примеров хотелось показать, как могут меняться требования в зависимости от прикладной области. Исследователю в области анализа данных необходимо проявить особый талант, чтобы правильно разобрать какие факторы могут влиять на анализируемые зависимости.

Выводы по главе 2

1) Предложен авторский общий подход для анализа и прогнозирования редких событий в экономике, который опирается на информацию о процессе образования событий. Данный метод состоит из 5 шагов: 1 – разделение событий по источникам, где они образованы; 2 – выдвижение предположения о процессе/механизме образования событий в источнике; 3 – восстановление параметров процесса по выборке редких событий; 4 – экстраполяция значений параметров процесса на будущее; 5 – запуск самого процесса и получение прогноза будущих событий.

2) Разработан емкостный метод анализа и прогнозирования редких событий, который является частным случаем общего авторского подхода. В нем в качестве процесса образования событий рассмотрен процесс потребления или процесс накопления возмущений. Процесс образования событий в данном случае можно моделировать как процесс наполнения/опустошения емкости, что определило название метода как «ёмкостный метод». Рассмотрены примеры применения емкостного метода на данных моделирования систем управления запасами, а также на реальных данных в торговле и сфере услуг.

3) Сформулированы предпосылки, условия, допущения, которые должны соблюдаться для корректной работы емкостного метода. Описаны последствия возможного нарушения этих условий.

Глава 3

Восстановление функции по выборке событий

Напомним, что для анализа и прогнозирования редких событий, которые порождаются в результате процессов, схожих с процессами опустошения/наполнения емкости, требуется произвести восстановление функции скорости изменения уровня $f(t)$ этой емкости. Причем восстановление происходит по наблюдаемым значениям y_i , которые в соответствии с основным предположением (2.1) являются интегралами от исходной функции $f(t)$ за время от момента совершения этого события t_i до момента времени совершения следующего события t_{i+1} ($y_i = \int_{t_i}^{t_{i+1}} f(t)dt$). Вместе с тем значения y_i наблюдаются не достоверно, а с некоторой погрешностью ε_i , т.е. $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$. В результате перед нами стоит задача восстановления неизвестной функции $f(t)$ по последовательности интегралов y_i , которые наблюдаются с погрешностью (откуда берется погрешность и какие ее характеристики будет показано в главе 4). Материалы данной главы опубликованы автором в работах [79; 80; 81; 82; 83; 84].

3.1 Разработка оптимизационной задачи восстановления функции по интегралам

Восстановление неизвестной функции по интегралам обычно происходит с помощью сплайнов. Существует два подхода к построению интегральных сплайнов. Первый подход – это построение интегро-дифференциальных сплайнов (ИД-сплайны) $S(x)$ [85; 86; 87; 88]. Интегро-дифференциальные сплайны можно отнести к частному случаю коллокационных методов решения интегро-дифференциальных уравнений, [89; 90; 91]. Интегро-дифференциальные сплайны одновременно зависят от интегральных и дифференциальных условий. Дифференциальными условиями являются условия непрерывности производной в точках сочленения сплайнов. Интегральными условиями являются равенства

интегралов от сплайна наблюдаемым интегралам, $\int_{t_i}^{t_{i+1}} S(t)dt = y_i$, где $y_i = \int_{t_i}^{t_{i+1}} f(t)dt$ – наблюдаемые интегралы. Получение интегро-дифференциальных сплайнов происходит с помощью решения системы линейных уравнений, составленных из имеющихся равенств (условий непрерывности и интегральных условий). Однако добавление интегральных условий $\int_{t_i}^{t_{i+1}} S(t)dt = y_i$ (строгих равенств) в систему приводит к тому, что, по сути, получающийся сплайн получится интерполяционным сплайном (как обычный интерполяционный сплайн строго проходит через все точки наблюдений, так и в данном сплайне полученные интегралы будут строго равны наблюдаемым интегралам, которые на самом деле могут наблюдаться с погрешностью). Даже Карл де Бур [92, стр. 80] при описании похожей задачи упоминал данные условия как «интерполяционные» условия. В работах [93; 94] строится одномерный и двумерный сплайны по известной площади под кривой закона распределения, однако в указанной работе сплайн также является интерполяционным (формируется на основе строгого равенства), а не аппроксимирующим (где возможно отклонение интегралов восстанавливаемой функции от наблюдений).

Другой подход — это конструирование сглаживающих сплайнов. Считается (в англоязычной литературе), что основополагающей работой по приближенному восстановлению функций, является совместная работа Kimeldorf и Wahba [95]. На момент написания этой главы данная работа процитирована 797 раз в международных базах цитирования. Изучение всего списка цитирующих работ (основываясь на аннотациях) и ознакомление более чем с 100 полнотекстовых публикаций из этого списка, показало отсутствие готового решения той задачи, которая решается здесь (тем более в том виде, в котором она представлена тут). Научных публикаций других авторов и готовых решений для восстановления функции по интегралам, которые наблюдаются с погрешностью, обнаружить не удалось. Пришлось

разрабатывать метод самостоятельно. Для начала кратко рассмотрим, как устроены обычные сглаживающие сплайны (без интегралов).

Классические сглаживающие сплайны

Для начала рассмотрим классическую задачу аппроксимации с помощью кубического (сглаживающего) сплайна. В великолепно написанной работе Green & Silverman [96] представлено очень доходчивое объяснение всей необходимой теории. Сглаживание сплайном заключается в поиске такой функции $g(t)$, которая бы минимизировала сумму квадратов отклонения в точках t_i и добавочного штрафа на нелинейность с некоторым коэффициентом α , формула (3.1)

$$L(g) = \sum_{i=1}^n (y_i - g(t_i))^2 + \alpha \int_{t_1}^{t_n} (g''(x))^2 dx \rightarrow \min, \quad (3.1)$$

где n – количество наблюдений;

$\int_a^b (g''(x))^2 dx$ – штраф на нелинейность;

α – коэффициент сглаживания, также известный как коэффициент регуляризации Тихонова [97].

В классическом виде функция $g(t)$ представляет собой кубический сплайн, то есть сочленение кусочков из полиномов не выше третьей степени в точках t_i с условием непрерывности как самой функции, так и ее производной в точках сочленения. Причем на каждом участке функцию $g(t)$ записывают не как полином с четырьмя неизвестными коэффициентами, а выражают через только две неизвестные переменные: само значение функции в точке $g_i = g(t_i)$ и ее вторую производную в этой точке $\gamma_i = g''(t_i)$. Набор из этих значений полностью задает весь сплайн, значение сплайна $g(t)$ в произвольной точке t определяется по формуле (3.2)

$$g(t) = \frac{(t - t_i)g_{i+1} + (t_{i+1} - t)g_i}{t_{i+1} - t_i} - \frac{1}{6}(t - t_i)(t_{i+1} - t) \left\{ \left(1 + \frac{t - t_i}{t_{i+1} - t_i}\right) \gamma_{i+1} + \left(1 + \frac{t_{i+1} - t}{t_{i+1} - t_i}\right) \gamma_i \right\},$$

$$i: t_i \leq t \leq t_{i+1}. \quad (3.2)$$

Данное представление полиномов через значение сплайна g_i и через вторую производную γ_i называют *value-second derivative representation* [96, стр. 22–24]. Обращаем внимание на то, что точки сочленения t_i полиномов также являются и точками наблюдений.

Если взглянуть на структуру этого представления, видно, что сама функция $g(t)$ непрерывна в точках сочленения¹⁾. Если взять вторую производную, имеем

$$g''(t) = \frac{(t - t_i)\gamma_{i+1} + (t_{i+1} - t)\gamma_i}{t_{i+1} - t_i},$$

и в точках сочленения, что с левой стороны, что с правой получаем одинаковое значение²⁾. То есть условия непрерывности самой функции и ее второй производной выполняются автоматически. Однако условия непрерывности первой производной автоматически не выполняются, нужно это обеспечить во время расчета сплайна с помощью следующего условия. Если взять первую производную от (3.2) и для некоторого j подставить значения $t_j^+ = t_j + 0$ и $t_j^- = t_j - 0$, то получим:

$$g'(t_j^+) = \frac{g_{j+1} - g_j}{t_{j+1} - t_j} - \frac{1}{6}(t_{j+1} - t_j)\{\gamma_{j+1} + 2\gamma_j\},$$

$$g'(t_j^-) = \frac{g_j - g_{j-1}}{t_j - t_{j-1}} + \frac{1}{6}(t_j - t_{j-1})\{2\gamma_j + \gamma_{j-1}\}.$$

Записывая условие равенства первых производных с обеих сторон от точек сочленения, имеем равенство, которое должно выполняться во всех точках сочленения (всего имеем $n - 2$ точек сочленения, в результате имеем систему

¹⁾ Для некоторого j и $t = t_{j+1} - 0$ имеем $g(t) = g_{i+1} = g_{j+1}$, но для $t = t_{j+1} + 0$ имеем $g(t) = g_i = g_{j+1}$ (индекс i уже увеличился на единицу).

²⁾ Для некоторого j и $t = t_{j+1} - 0$ имеем $g''(t) = \gamma_{i+1} = \gamma_{j+1}$, но для $t = t_{j+1} + 0$ имеем $g''(t) = \gamma_i = \gamma_{j+1}$.

$n - 2$ уравнений). Перекидывая значения сплайна g_i в одну сторону, а значения вторых производных в другую, получают формулу (3.3)

$$\frac{g_{i+1} - g_i}{t_{i+1} - t_i} - \frac{g_i - g_{i-1}}{t_i - t_{i-1}} = \frac{1}{6}(t_{i+1} - t_i)\{\gamma_{i+1} + 2\gamma_i\} + \frac{1}{6}(t_i - t_{i-1})\{2\gamma_i + \gamma_{i-1}\},$$

$$i = 2, \dots, n - 1. \quad (3.3)$$

Данную систему уравнений компактно записывают с помощью матричного выражения, формула (3.4)

$$Q^T g = R\gamma, \quad (3.4)$$

где $g = (g_1, \dots, g_n)^T$ – столбец значений сплайна, $\gamma = (\gamma_2, \dots, \gamma_{n-1})^T$ – столбец вторых производных (вторая производная в натуральном сплайне в начальной и последней точке равна нулю, $\gamma_1 = \gamma_n = 0$), матрица Q размерностью $n \times (n - 2)$ и матрица R размерностью $(n - 2) \times (n - 2)$ являются трех-диагональными и заполняются следующим образом.

$$Q =$$

	2	3	...	$n - 1$
1	h_1^{-1}	0	...	0
2	$-h_1^{-1} - h_2^{-1}$	h_2^{-1}	...	0
3	h_2^{-1}	$-h_2^{-1} - h_3^{-1}$...	0
4	0	h_3^{-1}	...	0
...
$n - 2$		0	...	h_{n-2}^{-1}
$n - 1$	0	0	...	$-h_{n-2}^{-1} - h_{n-1}^{-1}$
n	0	0	...	h_{n-1}^{-1}

$$R =$$

	2	3	4	...	$n - 1$
2	$(h_1 + h_2)/3$	$h_2/6$	0	...	0
3	$h_2/6$	$(h_2 + h_3)/3$	$h_3/6$...	0
4	0	$h_3/6$	$(h_3 + h_4)/3$...	0
5	0	0	$h_4/6$...	0
...
$n - 2$	0	0	0	...	$h_{n-2}/6$
$n - 1$	0	0	0	...	$(h_{n-2} + h_{n-1})/3$

где $h_i = t_{i+1} - t_i$, $i = 1, \dots, n - 1$.

Благодаря условиям непрерывности $Q^T g = R\gamma$ можно выразить значения вторых производных γ через значения сплайна g или наоборот

$$\gamma = R^{-1}Q^T g. \quad (3.5)$$

Штраф на нелинейность $\int_{t_1}^{t_n} (g''(x))^2 dx$ упрощается [96, стр. 24–25] до операций с теми же матрицами, формула (3.6)

$$\int_{t_1}^{t_n} (g''(x))^2 dx = \gamma^T Q^T g = \gamma^T R \gamma = g^T [QR^{-1}Q^T]g = g^T K g, \quad (3.6)$$

где $K = QR^{-1}Q^T$ – симметричная матрица размерностью $n \times n$.

Оптимизационную задачу (3.1) представляют в матричном виде

$$\begin{aligned} L(g) &= (Y - g)^T(Y - g) + \alpha g^T K g = \\ &= Y^T Y - 2Y^T g + g^T(I + \alpha K)g \rightarrow \min, \end{aligned} \quad (3.7)$$

где $Y = (y_1, \dots, y_n)^T$ – столбец наблюдаемых значений. Данная оптимизационная задача зависит только от неизвестных значений g .

Для нахождения минимума выражения (3.7) частную производную по g приравнивают к нулю¹⁾, в результате получают формулу (3.8)

$$\begin{aligned} -2Y + 2(I + \alpha K)g &= 0, \\ (I + \alpha K)g &= Y, \\ g &= (I + \alpha K)^{-1}Y. \end{aligned} \quad (3.8)$$

Откуда, воспользовавшись системой уравнений из условий непрерывности можно найти $\gamma = R^{-1}Q^T g$. После чего сплайн полностью определен и его можно строить по найденным значениям g и γ с помощью выражения (3.2).

Это был классический сглаживающий сплайн (здесь не было моего вклада). Теперь перейдем к исходной задаче восстановления функции по интегралам с помощью уже интегральных сплайнов.

Восстановление функции по интегралам с помощью кубического интегрального сглаживающего сплайна

Возвращаясь к нашей задаче, у нас требуется найти функцию $\hat{f}(t)$, которая бы минимизировала сумму квадратов разницы между известным значением $y_i = \int_{t_i}^{t_{i+1}} f(t)$ и интегралом от найденной функции $\int_{t_i}^{t_{i+1}} \hat{f}(t)dt$

¹⁾ Формулы для взятия производной в матричной форме следующие $\frac{d(x^T b)}{dx} = b$; $\frac{d(bx)}{dx} = b^T$; $\frac{d(x^T A x)}{dx} = (A + A^T)x = 2Ax$, последнее справедливо, если матрица симметрична.

$$\sum_{i=1}^{n-1} \left(y_i - \int_{t_i}^{t_{i+1}} \hat{f}(t) dt \right)^2 + C \int_{t_1}^{t_n} (\hat{f}''(t))^2 dt \rightarrow \min. \quad (3.9)$$

Обращаю внимание на то, что сейчас у нас наблюдения y_i являются интегралами от исходной функции, а не просто значениями функции. В то же время можно заметить, что в сумме теперь $n - 1$ элементов. Готового решения такой задачи найти не удалось, поэтому решение было найдено самостоятельно. Во всех встречающихся решениях строился интерполяционный сплайн, а не сглаживающий. Причем все встретившиеся решения вытекали не из минимизации квадратов отклонения, а из решения систем уравнений с «интерполяционными» условиями; использование штрафа на нелинейность в таких задачах не встречалось.

Решение ищется в виде сплайна $\hat{f}(t) = g(t)$, где $g(t)$ – кубический сплайн, представленный в форме (3.2) через значения сплайна и вторые производные. Для решения нашей задачи будем использовать те же обозначения.

Оптимизационная задача, формула (3.10)

$$L(g) = \sum_{i=1}^{n-1} \left(y_i - \int_{t_i}^{t_{i+1}} g(t) dt \right)^2 + \alpha \int_{t_1}^{t_n} (g''(t))^2 dt \rightarrow \min. \quad (3.10)$$

Известные значения:

- $Y = (y_1, \dots, y_{n-1})^T$ – вектор столбец $n - 1$ значений наблюдений интегралов исходной функции;
- t_1, \dots, t_n – моменты наблюдений;
- α – параметр сглаживания¹⁾, отвечающий за то, с каким весом надо учитывать штраф на нелинейность (задается пользователем).

¹⁾ В литературе, посвященной сглаживающим сплайнам, слагаемое $\alpha \int_a^b (g''(x))^2 dx$ называют штрафом на нелинейность (шероховатость), коэффициент α называют просто сглаживающим параметром, однако в некоторых работах это второе слагаемое называют регуляризирующим слагаемым, а параметр α – регуляризирующим параметром [97]. В нашем случае параметр α также можно называть параметром регуляризации, так как в случае отсутствия этого параметра, когда $\alpha = 0$, наша задача будет некорректно поставленной, т.е. может существовать бесконечное множество решений, а получающаяся в результате аналитического решения матрица становится вырожденной, и вычисление решения становится невозможным. При очень малых значениях α решение получается нестабильным.

Искомые значения:

- $g = (g_1, \dots, g_n)^T$ – значения сплайна в точках t_1, \dots, t_n ;
- $\gamma = (\gamma_2, \dots, \gamma_{n-1})^T$ – значения вторых производных сплайна в точках t_2, \dots, t_{n-1} ($\gamma_1 = \gamma_n = 0$).

В оптимизационной задаче (3.10) по сравнению с задачей (3.1) изменения коснулись первой части (там появился интеграл), вторая часть (штраф на нелинейность) не поменялась. Так как решение по-прежнему ищется в виде сплайна $g(t)$ в форме (3.2), то условия непрерывности сплайна по-прежнему дают систему (3.4) $Q^T g = R\gamma$. Сконцентрируемся на первой части.

Найдем интеграл $\int_{t_i}^{t_{i+1}} g(t)dt$, формула (3.11)

$$\int_{t_i}^{t_{i+1}} g(t)dt = \int_{t_i}^{t_{i+1}} \frac{(t - t_i)g_{i+1} + (t_{i+1} - t)g_i}{t_{i+1} - t_i} dt - \int_{t_i}^{t_{i+1}} \frac{1}{6} (t - t_i)(t_{i+1} - t) \left\{ \left(1 + \frac{t - t_i}{t_{i+1} - t_i}\right) \gamma_{i+1} + \left(1 + \frac{t_{i+1} - t}{t_{i+1} - t_i}\right) \gamma_i \right\} dt. \quad (3.11)$$

Разобьем данное выражение на три части (во второй части два слагаемых, которые рассчитаем отдельно).

Первая часть. Вычисления тривиальны, формула (3.12)

$$\int_{t_i}^{t_{i+1}} \frac{(t - t_i)g_{i+1} + (t_{i+1} - t)g_i}{t_{i+1} - t_i} dt = \frac{1}{2} (t_{i+1} - t_i)g_{i+1} + \frac{1}{2} (t_{i+1} - t_i)g_i. \quad (3.12)$$

Вторая часть

$$\frac{\gamma_{i+1}}{6(t_{i+1} - t_i)} \int_{t_i}^{t_{i+1}} (t - t_i)(t_{i+1} - t)(t + t_{i+1} - 2t_i) dt.$$

Интегрирование таких длинных выражений может быть достаточно сложным занятием, если раскрыть скобки уже будет 12 слагаемых, а надо взять определенный интеграл, в результате будем иметь 24 слагаемых, после чего полученное выражение требуется упростить. Однако проще не раскрывать все скобки, а сгруппировать элементы определенным образом, тогда интегрирование дается намного проще. В приложении В представлен,

наверное, самый простой способ взятия этого интеграла. В итоге имеем, формула (3.13)

$$\begin{aligned} & \frac{\gamma_{i+1}}{6(t_{i+1} - t_i)} \int_{t_i}^{t_{i+1}} (t - t_i)(t_{i+1} - t)(t + t_{i+1} - 2t_i) dt \\ &= \frac{\gamma_{i+1}}{6(t_{i+1} - t_i)} \frac{(t_{i+1} - t_i)^4}{4} = \frac{\gamma_{i+1}}{24} (t_{i+1} - t_i)^3. \end{aligned} \quad (3.13)$$

Третья часть. Выполняя аналогичные преобразования, приложение Г, будем иметь, формула (3.14)

$$\begin{aligned} & \frac{\gamma_i}{6(t_{i+1} - t_i)} \int_{t_i}^{t_{i+1}} (t - t_i)(t_{i+1} - t)(2t_{i+1} - t - t_i) dt = \\ &= \frac{\gamma_i}{6(t_{i+1} - t_i)} \frac{(t_{i+1} - t_i)^4}{4} = \frac{\gamma_i}{24} (t_{i+1} - t_i)^3. \end{aligned} \quad (3.14)$$

Подставляя выражения (3.12), (3.13), (3.14) в (3.11), получаем (3.15)

$$\int_{t_i}^{t_{i+1}} g(t) dt = g_{i+1} \frac{h_i}{2} + g_i \frac{h_i}{2} - \gamma_{i+1} \frac{h_i^3}{24} - \gamma_i \frac{h_i^3}{24}, \quad (3.15)$$

где $h_i = t_{i+1} - t_i$, $i = 1, \dots, n - 1$.

Тогда для неизвестных $g = (g_1, \dots, g_n)^T$, $\gamma = (\gamma_2, \dots, \gamma_{n-1})^T$ оптимизационная задача (3.10) может быть записана в виде (тут $\gamma_1 = \gamma_n = 0$), формула (3.16)

$$\begin{aligned} L(g) &= \sum_{i=1}^{n-1} \left(y_i - \int_{t_i}^{t_{i+1}} g(t) dt \right)^2 + \alpha \int_{t_1}^{t_n} (g''(t))^2 dt = \\ & \sum_{i=1}^{n-1} \left(y_i - g_{i+1} \frac{h_i}{2} - g_i \frac{h_i}{2} + \gamma_{i+1} \frac{h_i^3}{24} + \gamma_i \frac{h_i^3}{24} \right)^2 + \alpha \int_{t_1}^{t_n} (g''(t))^2 dt = \\ &= (Y - Vg + P\gamma)^T (Y - Vg + P\gamma) + \alpha g^T K g, \end{aligned} \quad (3.16)$$

где матрица V размером $(n - 1) \times n$ и матрица P размером $(n - 1) \times (n - 2)$ есть матрицы коэффициентов при неизвестных g и γ

$$V = \frac{1}{2} \begin{pmatrix} h_1 & h_1 & 0 & \cdots & 0 & 0 \\ 0 & h_2 & h_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & h_{n-1} & h_{n-1} \end{pmatrix},$$

$$P = \frac{1}{24} \begin{pmatrix} h_1^3 & 0 & 0 & 0 & \dots & 0 & 0 \\ h_2^3 & h_2^3 & 0 & 0 & \dots & 0 & 0 \\ 0 & h_3^3 & h_3^3 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & h_{n-2}^3 & h_{n-2}^3 \\ 0 & 0 & 0 & 0 & \dots & 0 & h_{n-1}^3 \end{pmatrix}.$$

Далее благодаря тому, что условия непрерывности по-прежнему дают систему уравнений $Q^T g = R\gamma$, откуда имеем $\gamma = R^{-1}Q^T g$, выражение (3.16) перепишем так, чтобы была только одна неизвестная, формулы (3.17) и (3.18)

$$\begin{aligned} L(g) &= (Y - Vg + P\gamma)^T(Y - Vg + P\gamma) + \alpha g^T K g = \\ &= (Y - (V - PR^{-1}Q^T)g)^T(Y - (V - PR^{-1}Q^T)g) + \alpha g^T K g, \end{aligned} \quad (3.17)$$

$$L(g) = (Y - Cg)^T(Y - Cg) + \alpha g^T K g, \quad (3.18)$$

где $C = V - PR^{-1}Q^T$ – матрица размерностью $(n - 1) \times n$.

Для нахождения минимума выражения (3.18) преобразуем и приравняем производную по g к нулю, формулы (3.19) и (3.20)

$$L(g) = g^T(C^T C + \alpha K)g - 2Y^T C g + Y^T Y, \quad (3.19)$$

$$L'(g) = 2(C^T C + \alpha K)g - 2(Y^T C)^T = 0. \quad (3.20)$$

Откуда, наконец, находим g , формула (3.21) (после чего из условий непрерывности (3.4) находим $\gamma = R^{-1}Q^T g$)

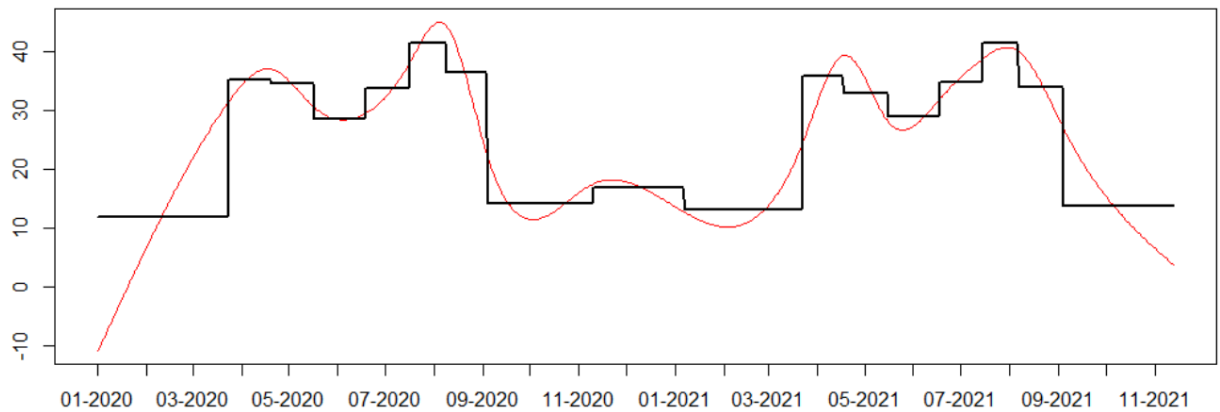
$$g = (C^T C + \alpha K)^{-1}C^T Y. \quad (3.21)$$

На этом задача решена, сплайн полностью определен, его можно рассчитать по найденным значениям g и γ с помощью выражения (3.2).

Недостатки из-за того, что узлы сплайна привязаны к точкам наблюдения

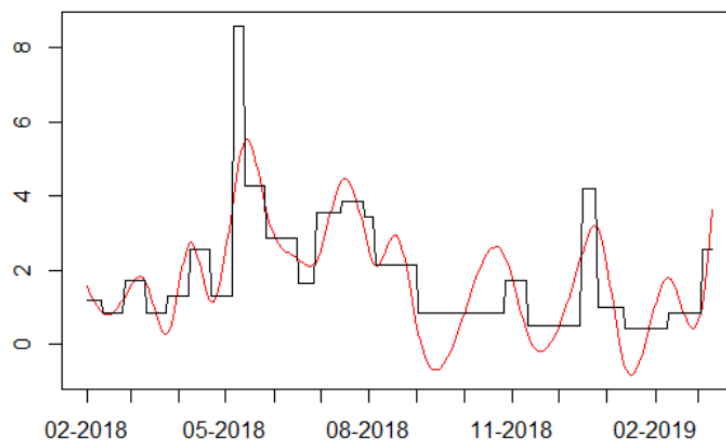
У данного сплайна моменты наблюдений t_1, \dots, t_n одновременно являются и узлами сплайна (точками сочленения полиномов), между которыми значения сплайна определяются как значения одного из полиномов (кусков сплайна) по формуле (3.2). В зависимости от того, как следуют наблюдения данный сплайн может обладать некоторыми недостатками.

1) Если расстояния между наблюдениями сильно отличаются, а такое может происходить, когда события идут то часто, то редко (например, когда потребитель изменяет величину максимального запаса), на некоторых участках заметна определенная асимметрия, рисунок 3.1 (поведение восстановленной функции достаточно интересно, она на узких интервалах изменяется быстро, а на широких интервалах изменения как будто замедляются). Такое поведение приводит иногда к негативным последствиям, рисунок 3.2.



Источник: составлено автором.

Рисунок 3.1 – Восстановление функции по интегралам, узлы сплайна совпадают с узлами наблюдений. Ступенчатая линия обозначает среднюю скорость $y_i/(t_{i+1} - t_i)$, площадь под ступенькой — это наблюдаемый интеграл y_i



Источник: составлено автором.

Рисунок 3.2 – Чередование узких и широких интервалов заставляет функцию изменяться более резко, из-за чего функция может опуститься ниже нуля

Когда интервалы следования сплайнов значительно изменяются (уже/шире), то функция имеет тенденцию то более резко возрастать/убывать, то более медленно, и при переключении интервалов это выглядит не всегда естественно. В этом случае следует увеличивать коэффициент сглаживания,

чтобы штраф на нелинейность был больше, и функция изменялась более плавно.

2) Вторая проблема (не такая значительная) заключается в том, что в случае, если наблюдений очень много, а наблюдения являются узлами сплайна, то размерность матриц может быть огромной, и трудоемкость вычислений может сильно возрасть (в нашем случае самих наблюдаемых интегралов не так много, чтобы эта проблема была серьезной).

Переход к сплайну с независимыми от точек наблюдения узлами

Для того, чтобы узлы сплайна были независимы от точек наблюдения, определим его в новых произвольных точках $s_1 < s_2 < \dots < s_m$, сам же сплайн определим в виде набора m функций $\beta_k(t)$, которые будут являться все теми же полиномами с формой (3.2), но определенными уже между новыми точками $k: s_k \leq t < s_{k+1}$. Сам сплайн $g(t)$ определяем следующим образом, формула (3.22)

$$g(t) = \sum_{k=1}^m \delta_k \beta_k(t), \quad (3.22)$$

где каждая такая функция $\beta_k(t)$ берется с некоторым коэффициентом δ_k , который является индикатором, принимающим значение 0 или 1 в зависимости от того, какая функция соответствует текущему моменту времени.

Данное представление сплайна напоминает базисный сплайн, когда сплайн определяется через сумму произвольных базисных сплайнов. У Бура [92] базисными сплайнами являются сплайны, составленные из кусочков полиномов, которые пересекаются областью определения (в одно мгновение t активно сразу несколько полиномов). В то же время сами кусочки полиномов никак не состыковываются. Непрерывность обеспечивается дополнительным условием, говорящим о том, что при переходе к новому узлу сумма активных полиномов (при пересечении узла один полином становится неактивным, а новый становится активным) совпадает. Так вот наше новое представление

можно назвать базисным сплайном, когда только один полином активен (где базисной функцией является полином в форме «*value-second derivative representation*» (3.2)), но в то же время непрерывность функции вытекает из самой формы полинома (3.2), а не из дополнительных условий. Однако, если бы мы выбрали узлы сплайна в точках наблюдения ($s_k = t_i, k = i = 1, \dots, n$), то наш сплайн превращается в ранее полученный сглаживающий интегральный сплайн.

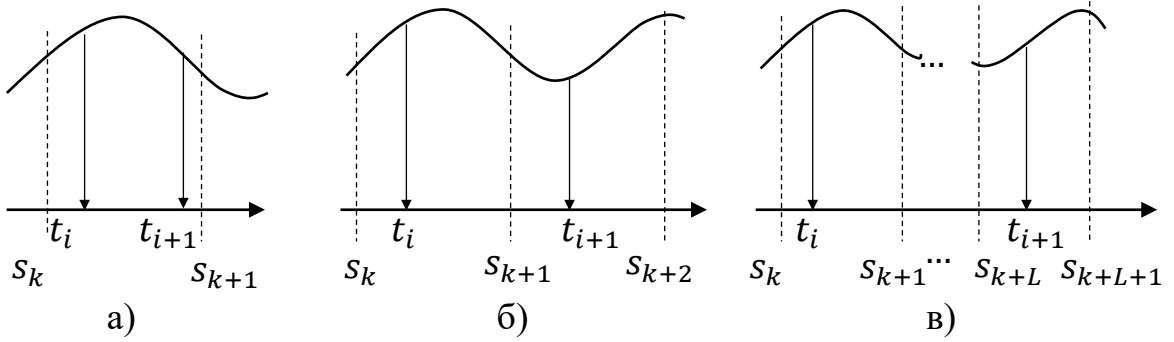
Дополнительно добавим возможность задавать вес каждого наблюдения w_i . Оптимизационная задача будет следующей, формула (3.23)

$$L_W(g) = \sum_{i=1}^{n-1} w_i \left\{ y_i - \int_{t_i}^{t_{i+1}} \sum_{k=1}^m \delta_k \beta_k(t) dt \right\}^2 + \alpha \int_{t_1}^{t_n} \left(\left(\sum_{k=1}^m \delta_k \beta_k(t) \right)'' \right)^2 dt \rightarrow \min . \quad (3.23)$$

Для решения этой задачи по-прежнему надо найти значения сплайна $g = (g_1, \dots, g_m)^T$ и его вторых производных $\gamma = (\gamma_2, \dots, \gamma_{m-1})^T$, но уже в новых точках $s_1 < s_2 < \dots < s_m$. Готового решение такой задачи найти не удалось, пришлось найти решение самостоятельно.

Штраф на нелинейность (вторая часть) по-прежнему будет выражаться как $\alpha g^T K g$, где $K = QR^{-1}Q^T$, но при этом размерность матриц Q и R будет зависеть не от n , а от m , и элементы будут зависеть от расстояния между новыми точками, где $h_k = s_{k+1} - s_k, k = 1, \dots, m - 1$. Сосредоточимся на первой части.

Рассчитаем интеграл $\int_{t_i}^{t_{i+1}} \sum_{k=1}^m \delta_k \beta_k(t) dt$, однако в зависимости от того, как выпадут точки наблюдений и как будут заданы новые точки сплайна, возможно несколько способов расчета. Два соседних наблюдения могут попасть в один интервал, рисунок 3.3а, могут попасть в два соседних интервала, рисунок 3.3б, а может получиться так, что два наблюдения будут находиться в L интервалах друг от друга, рисунок 3.3в.



Источник: составлено автором.

Рисунок 3.3 – Расположение соседних наблюдений на разных участках сплайна

Еще раз обращу внимание, что L – это количество интервалов между соседними наблюдениями, которое может быть равно 0, если оба наблюдения выпали на один интервал. Индекс k тут принимает значение номера интервала, в который попало i -тое наблюдение. Для того, чтобы получить универсальное выражение, подходящее для всех трех случаев, представим интеграл в следующем виде, формула (3.24)

$$\int_{t_i}^{t_{i+1}} \sum_{k=1}^m \delta_k \beta_k(t) dt = \sum_{l=0}^L \int_{s_{k+l}}^{s_{k+l+1}} \beta_{k+l}(t) dt - \int_{s_k}^{t_i} \beta_k(t) dt - \int_{t_{i+1}}^{s_{k+L+1}} \beta_{k+L}(t) dt, \quad (3.24)$$

$$L: s_{k+L} < t_{i+1} \leq s_{k+L+1},$$

$$k: s_k \leq t_i < s_{k+1}.$$

В правой стороне выражения (3.24) имеем три части. Первая часть выражения есть интеграл на всех L участках сплайна, причем, если $L = 0$, то в сумме будет одно слагаемое. Второе выражение есть интеграл от начала первого (k -того) полинома до текущего (i -того) наблюдения. Третье выражение есть интеграл от следующего наблюдения ($i + 1$ -того) до конца последнего интервала ($k + L$ -того), на который попало следующее наблюдение (k и L определяются в зависимости от того, куда попало текущее и следующее наблюдение).

Первая часть выражения (3.24) определяется по уже полученной ранее формуле (3.15), только лишь границами интервала стали новые точки s_k , формула (3.25)

$$\begin{aligned} & \sum_{l=0}^L \int_{s_{k+l}}^{s_{k+l+1}} \beta_{k+l}(t) dt = \\ & = \sum_{l=0}^L \left[\frac{h_{k+l}}{2} g_{k+l+1} + \frac{h_{k+l}}{2} g_{k+l} - \frac{h_{k+l}^3}{24} \gamma_{k+l+1} - \frac{h_{k+l}^3}{24} \gamma_{k+l} \right]. \end{aligned} \quad (3.25)$$

Находим вторую часть $\int_{s_k}^{t_i} \beta_k(t) dt$. После долгих преобразований, приложение Д, получаем относительно компактную формулу (3.26)

$$\begin{aligned} \int_{s_k}^{t_i} \beta_k(t) dt &= \frac{(h_k^{-i})^2}{2h_k} g_{k+1} + \frac{(h_k)^2 - (h_k^{+i})^2}{2h_k} g_k + \\ &+ \frac{\gamma_{k+1}}{24h_k} (h_k^{-i})^2 \left((h_k^{-i})^2 - 2(h_k)^2 \right) - \frac{\gamma_k}{24h_k} (h_k^{-i})^2 (h_k^{+i} + h_k)^2, \end{aligned} \quad (3.26)$$

где $h_k^{-i} = t_i - s_k$, $h_k^{+i} = s_{k+1} - t_i$, $h_k = s_{k+1} - s_k$.

Третью часть $\int_{t_{i+1}}^{s_{k+L+1}} \beta_{k+L}(t) dt$ находим аналогично приложению Д, формула (3.27)

$$\begin{aligned} & \int_{t_{i+1}}^{s_{k+L+1}} \beta_{k+L}(t) dt = \\ & = \frac{(h_{k+L})^2 - (h_{k+L}^{-(i+1)})^2}{2h_{k+L}} g_{k+L+1} + \frac{(h_{k+L}^{+(i+1)})^2}{2h_{k+L}} g_{k+L} - \\ & \quad - \frac{\gamma_{k+L+1}}{24h_{k+L}} (h_{k+L}^{+(i+1)})^2 (h_{k+L}^{-(i+1)} + h_{k+L})^2 + \\ & \quad + \frac{\gamma_{k+L}}{24h_{k+L}} (h_{k+L}^{+(i+1)})^2 \left((h_{k+L}^{+(i+1)})^2 - 2(h_{k+L})^2 \right), \end{aligned} \quad (3.27)$$

где $h_{k+L}^{-(i+1)} = t_{i+1} - s_{k+L}$, $h_{k+L}^{+(i+1)} = s_{k+L+1} - t_{i+1}$.

Подставляя найденные выражения (3.25),(3.26),(3.27) в (3.24), можем найти $\int_{t_i}^{t_{i+1}} \sum_{k=1}^m \delta_k \beta_k(t) dt$. Как и раньше форма этого выражения будет линейной по отношению к неизвестным g и γ , что позволяет использовать матричные операции. В результате оптимизационную задачу (3.23) для нахождения искомого сплайна интегралов можно записать в знакомом виде, формула (3.28)

$$L_W(g) = (Y - Vg + P\gamma)^T W(Y - Vg + P\gamma) + \alpha g^T K g \rightarrow \min, \quad (3.28)$$

где матрица V - матрица коэффициентов при неизвестных значениях g , P - матрица коэффициентов при неизвестных значениях второй производной γ (что значения, что вторая производная определяется уже в новых узлах сплайна $s_1 < s_2 < \dots < s_m$), W - диагональная матрица весов. Заполнение матриц V и P происходит только на основе наблюдений о моментах времени возникновения текущего и следующего события, в зависимости от того на какой интервал выпало это наблюдение.

Возможно будет удобно воспользоваться следующим представлением:

$$V = G^I - G^{II} - G^{III},$$

$$P = \Gamma^I - \Gamma^{II} - \Gamma^{III},$$

где матрицы G^I, G^{II}, G^{III} размерностью $(n-1) \times m$, матрица $\Gamma^I, \Gamma^{II}, \Gamma^{III}$ размерностью $(n-1) \times (m-2)$ (так как $\gamma_1 = \gamma_m = 0$ не участвуют). Элементы этих матриц заполняются следующим образом:

$$G_{i,k}^I = \frac{h_k}{2}, \quad s_k \leq t_i < s_{k+1},$$

$$G_{i,k+l}^I = (h_{k+l-1} + h_{k+l})/2, \quad l = 1, \dots, L: s_k \leq t_i, s_{k+L} < t_{i+1} \leq s_{k+L+1},$$

$$G_{i,k+L+1}^I = h_{k+L}/2, \quad L: s_{k+L} < t_{i+1} \leq s_{k+L+1},$$

$$G_{i,k}^{II} = h_k/2 - (h_k^{+i})^2/2h_k, \quad s_k \leq t_i < s_{k+1},$$

$$G_{i,k+1}^{II} = (h_k^{-i})^2/2h_k, \quad s_k \leq t_i < s_{k+1},$$

$$G_{i,k+L}^{III} = (h_{k+L}^{+(i+1)})^2/2h_{k+L}, \quad s_{k+L} < t_{i+1} \leq s_{k+L+1},$$

$$G_{i,k+L+1}^{III} = h_{k+L}/2 - (h_{k+L}^{-(i+1)})^2/2h_{k+L}, \quad s_{k+L} < t_{i+1} \leq s_{k+L+1},$$

$$\Gamma_{i,k}^I = h_k^3/24, \quad s_k \leq t_i < s_{k+1},$$

$$\Gamma_{i,k+l}^I = (h_{k+l-1}^3 + h_{k+l}^3)/24, \quad l = 1, \dots, L: s_k \leq t_i,$$

$$s_{k+L} < t_{i+1} \leq s_{k+L+1},$$

$$\Gamma_{i,k+L+1}^I = h_{k+L}^3/24, \quad L: s_{k+L} < t_{i+1} \leq s_{k+L+1},$$

$$\Gamma_{i,k}^{II} = (h_k^{-i})^2 (h_k^{+i} + h_k)^2/24h_k, \quad s_k \leq t_i < s_{k+1},$$

$$\Gamma_{i,k+1}^{II} = - (h_k^{-i})^2 \left((h_k^{-i})^2 - 2(h_k)^2 \right) / 24h_k, \quad s_k \leq t_i < s_{k+1},$$

$$\Gamma_{i,k+L}^{III} = - \left(h_{k+L}^{+(i+1)} \right)^2 \left(\left(h_{k+L}^{+(i+1)} \right)^2 - 2(h_{k+L})^2 \right) / 24h_{k+L},$$

$$s_{k+L} < t_{i+1} \leq s_{k+L+1},$$

$$\Gamma_{i,k+L+1}^{III} = \left(h_{k+L}^{+(i+1)} \right)^2 \left(h_{k+L}^{-(i+1)} + h_{k+L} \right)^2 / 24h_{k+L}, \quad s_{k+L} < t_{i+1} \leq s_{k+L+1}.$$

Подробнее о том, как заполняются элементы матриц V и P , будет в следующих двух параграфах.

После замены $C = V - PR^{-1}Q^T$ матрицей размерностью $(n-1) \times m$ оптимизационная задача (3.28) будет знакомого вида, формула (3.29)

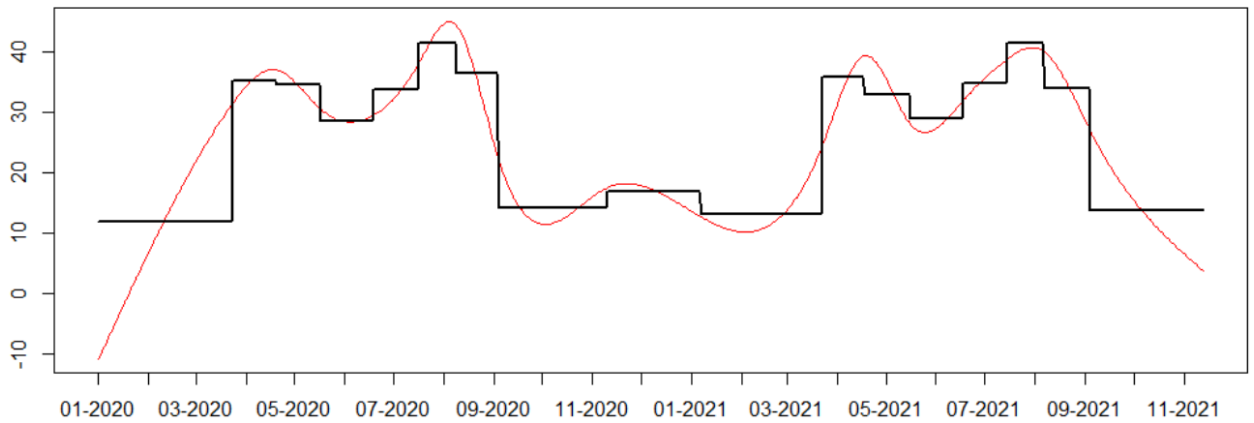
$$L_W(g) = (Y - Cg)^T W (Y - Cg) + \alpha g^T K g \rightarrow \min. \quad (3.29)$$

Приравнивая производную по g к нулю, получаем искомые значения g , а вслед за ними и $\gamma = R^{-1}Q^T g$, формула (3.30)

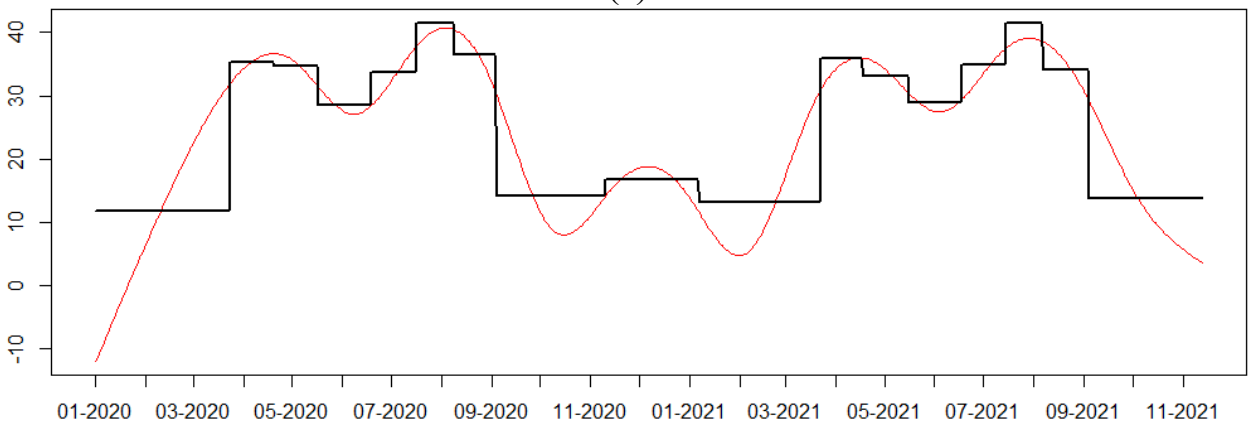
$$g = (C^T W C + \alpha K)^{-1} C^T W Y \quad (3.30)$$

Найденные значения g и γ определяют сплайн $g(t)$ в любой точке по формуле (3.2).

Для данного сплайна можно задавать произвольные узлы, не привязываясь к точкам наблюдений. Например, можно взять такое же количество узлов, что и наблюдений, но разместить их уже равномерно. Построенная функция будет обладать более естественным поведением, не будет недостатков, описанных выше, для сравнения показан рисунок 3.4.



(a)



(б)

Источник: составлено автором.

Рисунок 3.4 – Восстановление функции по интегралам: а) узлы сплайна привязаны к точкам наблюдения; б) узлы расположены равномерно

3.2 Выбор коэффициента сглаживания

Так как разница в сумме квадратов (3.9) считается между значениями интегралов, которые значительно превосходят значения самой функции, квадрат второй производной у которой достаточно мал, для какого-то видимого сглаживания погрешностей в наблюдаемых интегралах, параметр α , отвечающий за сглаживание, должен браться на достаточно большом уровне (в примере выше $\alpha = 10^5$).

Можно выбирать параметр сглаживания вручную (на глаз, чтобы визуально восстановленная функция хорошо соответствовала имеющимся данным), но существуют специальные процедуры автоматического расчета данного параметра на основе входных данных. Для этого используют либо

метод L -кривой [98; 99], либо метод кросс-валидации [100; 96, стр. 30]. Мною было проведено подробное исследование возможности применения этих подходов для нашей задачи восстановления функции по интегралам [84]. Оказалось, что Метод L -кривой не работает. Полученные формулы для расчета оценки кросс-валидации для случая интегрального сплайна следующие, формула (3.31) и (3.32)

$$CV(\alpha) = (n - 3)^{-1} \sum_{i=2}^{n-2} w_i \left(\frac{y_i - \sum_{j=1}^m C_{ij} g_j}{1 - \sum_{j=1}^m C_{ij} A_{ji}(\alpha)} \right)^2, \quad (3.31)$$

$$GCV(\alpha) = (n - 3)^{-1} \frac{\sum_{i=2}^{n-2} w_i (y_i - \sum_{j=1}^m C_{ij} g_j)^2}{\left(1 - (n - 3)^{-1} \sum_{i=2}^{n-2} \sum_{j=1}^m C_{ij} A_{ji}(\alpha) \right)^2}, \quad (3.32)$$

где $C = V - PR^{-1}Q^T$ та же матрица что и ранее;

$A(\alpha) = (C^T W C + \alpha K)^{-1} C^T W$ – матрица проекции (или hat - матрица);

w_i – вес наблюдения (элемент W_{ii});

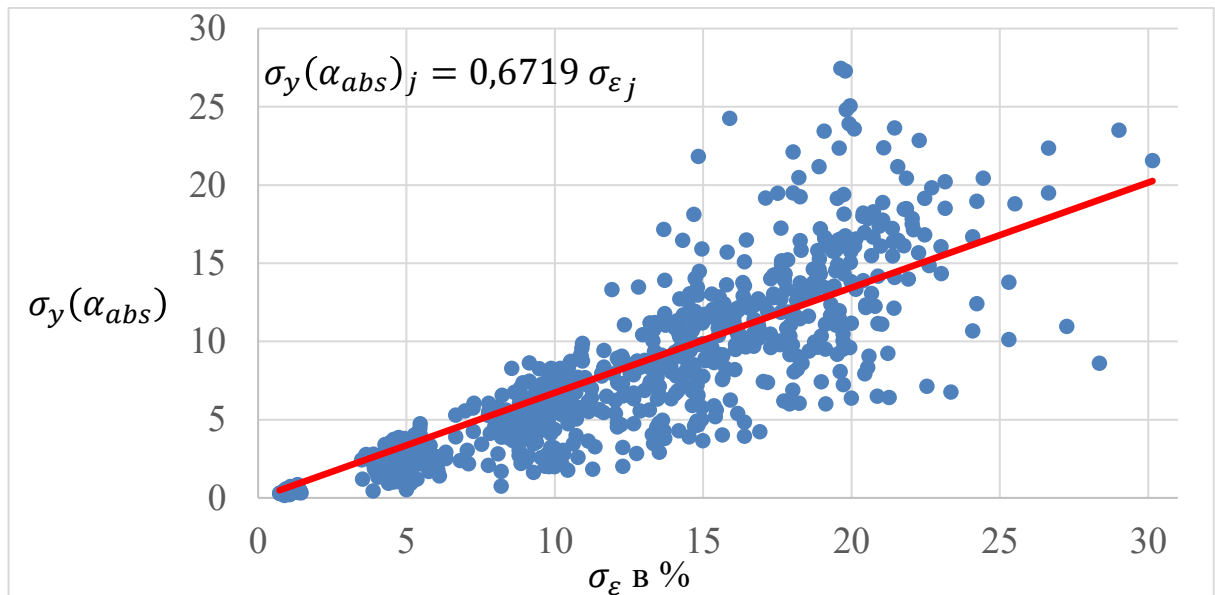
W – диагональная весовая матрица;

n – количество наблюдений.

Однако метод кросс-валидации работает только в половине случаев, в другой половине приводит либо к очень сильному занижению, либо завышению коэффициента сглаживания. Подробности выходят за рамки данной диссертационной работы, так как математика там достаточно сложная. Исследование выбора оптимального коэффициента сглаживания с помощью имитационного моделирования представлено в работе [84].

Другим подходом является принцип невязки Морозова (Morozov's discrepancy principle) [101; 102 стр. 72] или что то же самое рекомендация Рейнша [103], в которых предлагается брать коэффициент сглаживания α на таком уровне, чтобы он сглаживал наблюдения на величину, соответствующую уровню шума σ_ε в исходных данных. Однако в проведенном исследовании с помощью имитационного моделирования [84] оказалось, что наилучшее восстановление функции будет при таком коэффициенте сглаживания, который бы приглушал исходные наблюдения на

67,19% от уровня шума (после попытки скорректировать гетероскедастичность на $64,92 \approx 65 \%$), рисунок 3.5.



Источник: составлено автором.

Рисунок 3.5 – Диаграмма рассеивания и линейная модель для значения приглушения наблюдений¹⁾ при оптимальном коэффициенте сглаживания²⁾ $\sigma_y(\alpha_{abs})$ от величины фактического уровня шума в наблюдениях σ_ε

Например, если уровень шума исходных наблюдений σ_ε ожидается как 5%, то надо выбрать такой коэффициент сглаживания α , чтобы после сглаживания наблюдения приглушились на $0,65 \cdot 5 = 3,25\%$ (т.е. $\sigma_y(\alpha) = ((n-1)^{-1} \sum_{i=1}^{n-1} ((y_i - \bar{y}_i)/\bar{y}_i)^2)^{0,5} \times 100\% = 3,25\%$). В итоге можем сформулировать следующую рекомендацию. Если уровень шума в наблюдениях σ_ε известен, то следует выбирать такой коэффициент сглаживания α , который бы приглушал наблюдения на $\sim 65\%$ от этого уровня шума σ_ε , если же уровень шума неизвестен, его можно предположить на уровне 1%. С другой стороны, можно все же использовать метод кросс-валидации при условии, что будут реализованы специальные процедуры,

¹⁾ Уровень приглушения определяется как $\sigma_y(\alpha) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} \left(\frac{y_i - \bar{y}_i}{\bar{y}_i}\right)^2} \times 100\%$, где $\bar{y}_i = \sum_{j=1}^m C_{ij} g_j$ – прогнозное значение наблюдаемого интеграла.

²⁾ Оптимальный коэффициент сглаживания $\alpha_{abs} = \arg \min_{\alpha} R_{abs}(\alpha)$, где $R_{abs}(\alpha) = \frac{1}{t_n - t_1} \sum_{t=t_1}^{t_n} \left| \frac{f(t) - g(t, \alpha)}{f(t)} \right|$ – среднее относительное отклонение по модулю восстановленной функции $g(t, \alpha)$ от исходной $f(t)$ при заданном коэффициенте сглаживания α .

которые бы сообщали, что рассчитанный коэффициент сглаживания оказался слишком занижен или слишком завышен.

Однако это все касается автоматического способа выбора коэффициента сглаживания. Как видно из рисунка 3.5, вокруг красной линии (эти значения будут рекомендоваться методом) имеется значительный разброс. Как можно было видеть в исследовании [84], оценка коэффициента сглаживания часто ошибочна, и даже если она близка к истинному оптимальному коэффициенту сглаживания, она все равно с ним не совпадает. По моему мнению, нет ничего лучше, чем человеческий глаз опытного исследователя. Исследователь, глядя на ступенчатую функцию и на результат сглаживания, может понимать, насколько хорошо получилось аппроксимировать данные, стоит ли уменьшить или увеличить коэффициент сглаживания.

3.3 Алгоритм восстановления функции по интегралам

Подведем итоги предыдущего параграфа и реализуем полученный алгоритм на языке R [66], а также в MS Excel. Алгоритм восстановления функции по интегралам следующий.

По значениям $Y = (y_1, \dots, y_{n-1})^T$ (последнее значение y_n не используется) и $t = (t_1, \dots, t_n)^T$ найти значения $g = (g_1, \dots, g_m)^T$, $\gamma = (\gamma_2, \dots, \gamma_{m-1})^T$ (в начальной и последней точке вторая производная равна нулю $\gamma_1 = \gamma_m = 0$). Эти значения определяются с помощью полученных матричных выражений, формула (3.33)

$$\begin{aligned} C &= V - PR^{-1}Q^T, \\ K &= QR^{-1}Q^T, \\ g &= (C^TWC + \alpha K)^{-1}C^TWY, \\ \gamma &= R^{-1}Q^Tg, \end{aligned} \tag{3.33}$$

где основная сложность заключается в заполнении исходных матриц Q, R, V, P .

Задаем узлы сплайна $s = (s_1, \dots, s_m)$ так, чтобы они были упорядочены $s_1 < s_2 < \dots < s_m$. Можно брать количество узлов не

совпадающим с количеством наблюдений, например, $m = 3n$ в три раза больше, можно количество узлов задать в несколько раз меньше, например, $m = \lfloor n/3 \rfloor$ (округляя до целого), а можно задать количество узлов одинаковым с количеством наблюдений $m = n$.

Далее следует выбрать как располагать узлы сплайна, располагать можно совершенно произвольно. В самом простом и практичном случае удобно расположить узлы равномерно, $s_j = s_{j-1} + h$, $j = 2, \dots, m$, где $h = \frac{t_n - t_1}{m-1}$, а $s_1 = t_1$. В то же время, если надо расположить узлы сплайна так, чтобы они совпадали с узлами наблюдений (при условии $m = n$), достаточно присвоить $s = t$.

Матрица Q размерностью $m \times (m - 2)$ и **матрица R** размерностью $(m - 2) \times (m - 2)$ заполняются в зависимости только от узлов сплайна

$$Q = \begin{pmatrix} h_1^{-1} & 0 & \dots & 0 \\ -h_1^{-1} - h_2^{-1} & h_2^{-1} & \dots & 0 \\ h_2^{-1} & -h_2^{-1} - h_3^{-1} & \dots & 0 \\ 0 & h_3^{-1} & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & h_{m-2}^{-1} \\ 0 & 0 & \dots & -h_{m-2}^{-1} - h_{m-1}^{-1} \\ 0 & 0 & \dots & h_{m-1}^{-1} \end{pmatrix},$$

$$R = \begin{pmatrix} (h_1 + h_2)/3 & h_2/6 & 0 & \dots & 0 \\ h_2/6 & (h_2 + h_3)/3 & h_3/6 & \dots & 0 \\ 0 & h_3/6 & (h_3 + h_4)/3 & \dots & 0 \\ 0 & 0 & h_4/6 & \dots & 0 \\ \dots & \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \dots & h_{m-2}/6 \\ 0 & 0 & \dots & \dots & (h_{m-2} + h_{m-1})/3 \end{pmatrix},$$

где $h_k = s_{k+1} - s_k$, $k = 1, \dots, m - 1$ (если узлы выбраны равномерно, то для любого k имеем $h_k = h$).

Матрица V размерностью $(n - 1) \times m$ и **матрица P** размерностью $(n - 1) \times (m - 2)$ заполняются в зависимости от того, на какой интервал выпали наблюдения. Вначале для каждого наблюдения i определяется номер интервала k , на который попало это наблюдение. Далее определяется количество интервалов L , через сколько интервалов встречается следующее

наблюдение $i + 1$ (L может быть нулевым). Затем элементы i -той строки заполняются следующим образом (i – индекс наблюдения, k – номер интервала)

$$k: t_k \leq t_i < t_{k+1},$$

$$L: t_{k+L} < t_{i+1} \leq t_{k+L+1}.$$

Матрица V , заполнение i -той строки:

$$V_{i,k} = \frac{(h_k^{+i})^2}{2h_k},$$

$$V_{i,k+l} = \frac{h_{k+l-1} + h_{k+l}}{2}, \quad l = 1, \dots, L,$$

$$V_{i,k+1} = V_{i,k+1} - \frac{(h_k^{-i})^2}{2h_k},$$

$$V_{i,k+L} = V_{i,k+L} - \frac{(h_{k+L}^{+(i+1)})^2}{2h_{k+L}},$$

$$V_{i,k+L+1} = \frac{(h_{k+L}^{-(i+1)})^2}{2h_{k+L}},$$

где $h_k = s_{k+1} - s_k$, $h_k^{-i} = t_i - s_k$, $h_k^{+i} = s_{k+1} - t_i$, $h_{k+L}^{-(i+1)} = t_{i+1} - s_{k+L}$, $h_{k+L}^{+(i+1)} = s_{k+L+1} - t_{i+1}$.

Тут обратите внимание, что элементы $V_{i,k+1}, V_{i,k+L}$ после того, как получили значения в цикле по $l = 1, \dots, L$, еще раз изменяются (причем если $L = 0$, то изменяться они будут только один раз). Например, если расстояние между i -тым и $i + 1$ -вым наблюдением $L > 2$, то i -тый ряд будет выглядеть как в таблице 3.1.

Таблица 3.1 – Заполнение i -того ряда матрицы V при $L > 2$

$k - 1$	k	$k + 1$... $k + l$...	$k + L$	$k + L + 1$	$k + L + 2$
0	$\frac{(h_k^{+i})^2}{2h_k}$	$\frac{h_k + h_{k+1}}{2}$ $-\frac{(h_k^{-i})^2}{2h_k}$	$\frac{h_{k+l-1} + h_{k+l}}{2}$	$\frac{h_{k+L-1} + h_{k+L}}{2}$ $-\frac{(h_{k+L}^{+(i+1)})^2}{2h_{k+L}}$	$\frac{(h_{k+L}^{-(i+1)})^2}{2h_{k+L}}$	0

Источник: составлено автором.

Но если наблюдения попали на один интервал, $L = 0$, то ряд i будет содержать только 2 ненулевых значения, Таблица 3.2.

Таблица 3.2 – Заполнение i -того ряда матрицы V при $L = 0$

$k - 1$	k	$k + 1$	$k + 2$
0	$\frac{(h_k^{+i})^2}{2h_k} - \frac{(h_{k+L}^{+(i+1)})^2}{2h_{k+L}}$	$\frac{(h_{k+L}^{-(i+1)})^2}{2h_{k+L}} - \frac{(h_k^{-i})^2}{2h_k}$	0

Источник: составлено автором.

Матрица P , заполнение i -той строки:

$$P_{i,k} = \frac{h_k^3}{24} - \frac{(h_k^{-i})^2 (h_k^{+i} + h_k)^2}{24h_k},$$

$$P_{i,k+l} = \frac{h_{k+l-1}^3 + h_{k+l}^3}{24}, \quad l = 1, \dots, L,$$

$$P_{i,k+1} = P_{i,k+1} + \frac{(h_k^{-i})^2 ((h_k^{-i})^2 - 2h_k^2)}{24h_k},$$

$$P_{i,k+L} = P_{i,k+L} + \frac{(h_{k+L}^{+(i+1)})^2 ((h_{k+L}^{+(i+1)})^2 - 2(h_{k+L})^2)}{24h_{k+L}},$$

$$P_{i,k+L+1} = \frac{h_{k+L}^3}{24} - \frac{(h_{k+L}^{+(i+1)})^2 (h_{k+L}^{-(i+1)} + h_{k+L})^2}{24h_{k+L}},$$

где $h_k = s_{k+1} - s_k$, $h_k^{-i} = t_i - s_k$, $h_k^{+i} = s_{k+1} - t_i$, $h_{k+L}^{-(i+1)} = t_{i+1} - s_{k+L}$, $h_{k+L}^{+(i+1)} = s_{k+L+1} - t_{i+1}$ (такие же что для матрицы V).

Тут обратите внимание, что элементы $P_{i,k+1}, P_{i,k+L}$ после того, как получили значения в цикле по $l = 1, \dots, L$, еще раз изменяются (причем если $L = 0$, то изменяться они будут только один раз). Также при заполнении матрицы P надо учитывать, что ее вторая размерность $m - 2$, причем нумерация второго индекса начинается с 2 (как у матриц Q и R). Можно предложить поступить проще, заполнять матрицу P как матрицу размерностью $(n - 1) \times m$ ничего не меняя, после чего отбросить первый и последний столбец. Например, если расстояние между i -тым и $i + 1$ -вым наблюдением $L > 2$, то i -тый ряд будет выглядеть как в таблице 3.3.

Таблица 3.3 – Заполнение i -того ряда матрицы P при $L > 2$

$k - 1$	k	$k + 1$	$k + 2$
0	$\frac{h_k^3}{24} - \frac{(h_k^{-i})^2 (h_k^{+i} + h_k)^2}{24h_k}$	$\frac{h_k^3 + h_{k+1}^3}{24} + \frac{(h_k^{-i})^2 ((h_k^{-i})^2 - 2h_k^2)}{24h_k}$	$\frac{h_{k+1}^3 + h_{k+2}^3}{24}$
... $k + l$...	$k + L$	$k + L + 1$	$k + L + 2$
$\frac{h_{k+l-1}^3 + h_{k+l}^3}{24}$	$\frac{h_{k+L-1}^3 + h_{k+L}^3}{24} + \frac{(h_{k+L}^{+(i+1)})^2 ((h_{k+L}^{+(i+1)})^2 - 2(h_{k+L})^2)}{24h_{k+L}}$	$\frac{h_{k+L}^3}{24} - \frac{(h_{k+L}^{+(i+1)})^2 (h_{k+L}^{-(i+1)} + h_{k+L})^2}{24h_{k+L}}$	0

Источник: составлено автором.

Матрица W размерностью $(n - 1) \times (n - 1)$ является весовой матрицей, где по главной диагонали стоят веса каждого наблюдения. Если наблюдениям нет нужды задавать веса, то можно использовать единичную матрицу или пропустить эту операцию с умножением на матрицу W в формуле (3.30).

Когда все матрицы заполнены, дальше остается воспользоваться формулами (3.31) и затем построить саму функцию по формуле (3.2). Далее рассмотрим программную реализацию.

3.4 Реализация на R

Реализуем описанный алгоритм в виде функции `int_spline`, которая будет принимать набор аргументов и возвращать координаты построенного сплайна. Аргументами будут:

- t – массив n координат наблюдений t_i ;
- Y – массив $n - 1$ значений наблюдений y_i ;
- s – массив узлов сплайна s_k , по умолчанию совпадает с массивом t ;
- m – количество узлов сплайна, по умолчанию равен `length(s)`;
- W – весовая матрица, по умолчанию единичная матрица;
- $alpha$ – параметр сглаживания, по умолчанию 10^5 ;

– x – массив координат, в которых будут рассчитываться значения сконструированного сплайна и которые затем возвратятся, по умолчанию x принимает все значения от t_1 до t_n с шагом 1.

Первым делом определяем длину набора данных.

```
n=length(t)
```

Предусмотрим возможность автоматически задавать узлы сплайна с одинаковым расстоянием между узлами, когда число узлов задано отличным от числа наблюдений (в случае если s не задается, то оно по умолчанию выбирается как t и его длина будет равна n).

```
if (m != length(s))
  s=seq(t[1],t[n],length=m)
```

Для заполнения матриц понадобится знать расстояние между узлами сплайна.

```
h=array(0,dim=m-1)
h [1:(m-1)]=s[2:m]-s[1:(m-1)]
```

Наконец переходим к заполнению матриц.

Матрица Q .

```
Q=matrix(0,nrow=m,ncol=m-2)
for (i in 1:(m-2))
{
  Q[i,i]=1/h[i]
  Q[i+1,i]=-1/h[i]-1/h[i+1]
  Q[i+2,i]=1/h[i+1]
}
```

Матрица R .

```
R=matrix (0,nrow=m-2,ncol=m-2)
for (i in 1:(m-2))
{
  R[i,i]=1/3*(h[i]+h[i+1])
  if (i<m-2)
  {
    R[i+1,i]=1/6*h[i+1]
    R[i,i+1]=1/6*h[i+1]
  }
}
```

Расчет матрицы K .

```
inv_R=solve(R) # обратная матрица
t_Q=t(Q)       # транспонирование
```

K=Q %%% inv_R %%% t_Q

Заполнение матрицы V .

```
V=matrix(0,nrow=n-1,ncol=m)
k=1
while((s[k]<=t[1])&(s[k+1]<t[1])) # находим k
  k=k+1
for (i in 1:(n-1))
{
  for (L in 0:(m-k-1)) # находим L
    if (t[i+1]<s[k+L+1])
      break;
  l=1;
  V[i,k]=(s[k+1]-t[i])^2/h[k]/2
  while (l<=L)
  {
    V[i,k+1]=(h[k+1-1]+h[k+1])/2
    l=l+1;
  }
  V[i,k+1]=V[i,k+1]-(t[i]-s[k])^2/h[k]/2
  V[i,k+L]=V[i,k+L]-(s[k+L+1]-t[i+1])^2/h[k+L]/2
  V[i,k+L+1]=(t[i+1]-s[k+L])^2/h[k+L]/2
  k=k+L
}
```

Заполнение матрицы P .

```
P=matrix(0,nrow=n-1,ncol=m)
k=1
while((s[k]<=t[1])&(s[k+1]<t[1])) # находим k
  k=k+1
for (i in 1:(n-1))
{
  for (L in 0:(m-k-1)) # находим L
    if (t[i+1]<s[k+L+1])
      break;
  l=1;
  P[i,k]=h[k]^3/24-(t[i]-s[k])^2*(s[k+1]-t[i]+h[k])^2
  /h[k]/24
  while (l<=L)
  {
    P[i,k+1]=(h[k+1-1]^3+h[k+1]^3)/24
    l=l+1;
  }
  P[i,k+1]=P[i,k+1]+(t[i]-s[k])^2*((t[i]-s[k])^2-
  2*h[k]^2)/h[k]/24
  P[i,k+L]=P[i,k+L]+(s[k+L+1]-t[i+1])^2*((s[k+L+1]-
  t[i+1])^2-2*h[k+L]^2)/h[k+L]/24
  P[i,k+L+1]=h[k+L]^3/24-(s[k+L+1]-t[i+1])^2*(t[i+1]-
  s[k+L]+h[k+L])^2/h[k+L]/24
}
```

```

    k=k+L
}
P=P[1:(n-1),2:(m-1)] #отбрасываем столбец с номером 1 и m

```

Расчет матрицы C .

```
C=V-P %**% inv_R %**% t_Q
```

Расчет значений g и γ .

```

t_C=t(C)
A=t_C %**% W %**% C + alpha*K
g=solve(A, t_C %**% W %**% Y )
gamma=inv_R %**% t_Q %**% g

```

Все значения посчитаны, осталось рассчитать и вернуть значения сплайна. Сначала, чтобы было проще работать с индексами, вектор γ дополняем нулями с обеих сторон. Затем рассчитываем значения y в координатах x (которые по умолчанию задаются как все точки от t_1 до t_n), и, наконец, возвращаем координаты y .

```

g2=c(0,gamma,0)
y=rep(0,length(x))          #изначально заполняем нулями
k=1                         #начинаем с первого интервала
for (j in (1:length(x)))
{
  while (x[j]<t[n] & x[j]>s[k]+h[k]) #если вышли на следующий
    k=k+1;                          #интервал, то сдвигаем интервал
  y[j]=((x[j]-s[k])*g[k+1]+(s[k+1]-x[j])*g[k])/h[k]-
    1/6*(x[j]-s[k])*(s[k+1]-x[j])*
    (g2[k+1]*(1+(x[j]-s[k])/h[k]) + g2[k]*(1+(s[k+1]-x[j])/h[k]) )
}
return(y)

```

3.5 Пример применения алгоритма на языке R

Пусть у нас имеется набор данных о датах t_i и объемах y_i купленной продукции некоторым покупателем, таблица 3.4 (данные получены в результате моделирования систем управления запасами). Требуется восстановить скорость расхода продукции $f(t)$ у покупателя.

Таблица 3.4 – Данные редких событий

Дата	Объем	Дата	Объем	Дата	Объем
20.01.2020	5081	07.01.2021	5017	26.08.2021	5032
24.03.2020	10035	11.03.2021	5043	27.10.2021	5061
21.06.2020	5105	05.05.2021	5013	26.12.2021	5086
26.07.2020	5135	12.06.2021	5106	-	-
07.09.2020	10090	16.07.2021	5029	-	-

Источник: составлено автором.

Пусть данные находятся в файле .csv и записаны в две колонки с заголовками. Считывание данных осуществляем следующим образом.

```
library(lubridate)          #библиотека для работы с датами
filename="F:/DIR/f_in.csv";
MyData=read.csv(file=filename,header=TRUE,sep=";",
                stringsAsFactors = FALSE, dec=",")
t=as.numeric(dmy(MyData[[1]])) #строку преобразуем в дату
                                #а затем дату в число

n=length(t)
Y=MyData[[2]]                 #берем значения Y из второго столбца
Y=Y[1:(n-1)]                 #последнее значение не используется
```

Простейший вызов функции для расчета нашего интегрального сплайна требует только два аргумента.

```
y=int_spline(t,Y)
```

При данном вызове функции узлы сплайна определяются в тех же точках что и точки наблюдений. По умолчанию значения сплайна рассчитываются во всех точках от первого до последнего наблюдения с шагом 1. Однако координата x не возвращается, для построения графика координату x надо рассчитать самостоятельно.

Вывод графика сплайна вместе с графиком средних значений можно осуществить следующим образом.

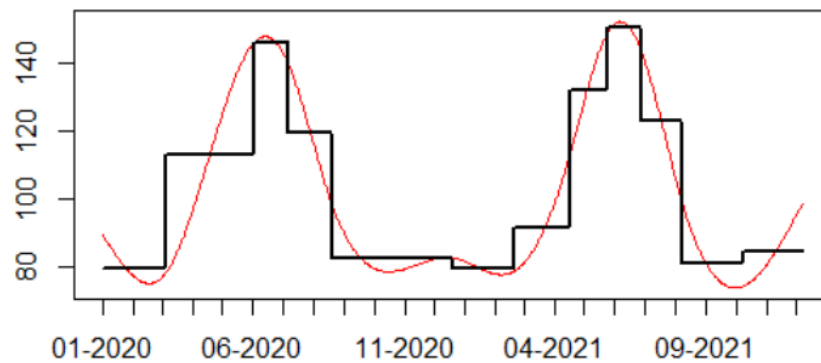
```
x=seq(t[1],t[n],1)
x2=NULL
y2=NULL
for (i in 1:(n-1))
{
  x2=c(x2,t[i],t[i+1])
  y2=c(y2,Y[i]/(t[i+1]-t[i]),Y[i]/(t[i+1]-t[i]))
}
plot(x,y,col="red",type="l",lwd="1",lty=1,xaxt="n",
```

```

ylim=range(c(y,y2)),xlim=range(c(x,x2)))
axis.Date(1,at=seq(min(dmy(MyData[[1]]),max(dmy(MyData[[1]]
))), by="months"),format="%m-%Y " )
lines(x2,y2,col="black",type="l",lwd="2",lty=1)

```

Восстановленная по интегралам функция изображена на рисунке 3.6. Алгоритм пытается подогнать кривую таким образом, чтобы площадь под кривой приближалась к площади под ступенькой с учетом штрафа на нелинейность. На рисунке 3.6 узлы сплайна и наблюдения, которые идут через разные интервалы времени, совпадают. Когда узлы сплайна привязаны к наблюдениям, а наблюдения идут то часто, то редко, то могут быть негативные последствия, которые были упомянуты в прошлом параграфе.



Ступенчатая линия обозначает среднюю скорость $y_i/(t_{i+1} - t_i)$, площадь под каждой ступенькой представляет собой значение интеграла y_i .

Источник: составлено автором.

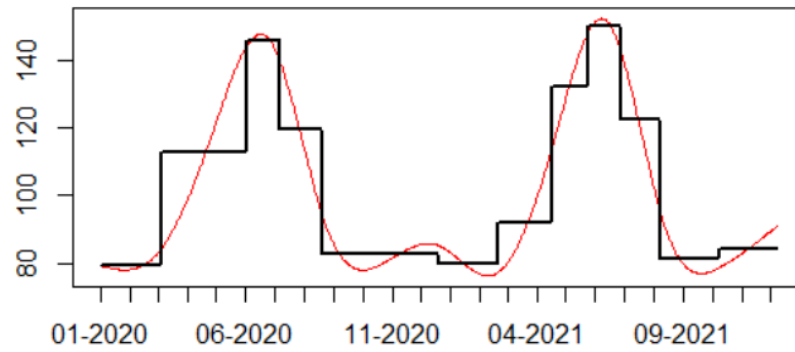
Рисунок 3.6 – Восстановленная функция, простейший вызов функции

Чтобы создать сплайн с равномерно следующими узлами сначала заполняем массив узлов s , а при вызове функции передаем этот массив как третий аргумент. Результатом будет сплайн, в котором узлы уже не привязаны к точкам наблюдения, рисунок 3.7. Обратите внимание, что количество узлов не поменялось, поменялось лишь их расположение.

```

s=seq(t[1],t[n],length=n)
y=int_spline(t,Y,s=s)

```

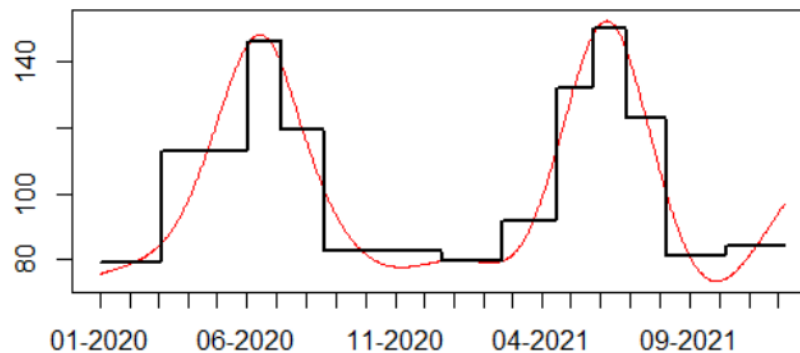


Источник: составлено автором.

Рисунок 3.7 – Сплайн с равномерно следующими узлами, $m = n$

Однако, чтобы задать сплайн с равномерно следующими узлами, достаточно указать количество узлов отличным от количества наблюдений, например, взять в три раза больше узлов чем наблюдений, рисунок 3.8. Количество узлов m не сильно влияет на вид восстанавливаемой функции, главное, чтобы узлов было достаточно, чтобы можно было описать все основные изменения этой функции.

`y=int_spline(t,Y,m=n*3)`



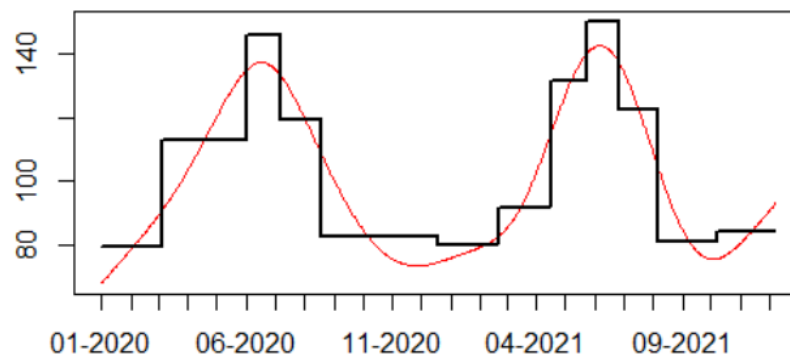
Источник: составлено автором.

Рисунок 3.8 – Сплайн с равномерно следующими узлами, $m = 3n$

Наибольшее влияние оказывает коэффициент сглаживания α . Если коэффициент сглаживания α близок к бесконечности, то функция вырождается в прямую линию. Если α близко к нулю, то сглаживания практически не происходит. Если же $\alpha = 0$, то задача является некорректно поставленной, существует бесконечное множество решений, матрица $A = C^TWC + \alpha K$ будет вырожденной и от нее обратная матрица не возьмется. Так как определенные интегралы от функции принимают большие значения, а

вторая производная как правило небольшие значения, то сглаживание начинает быть заметным, когда коэффициент сглаживания принимает большие значения. Например, для того же набора данных визуальных изменений практически не наблюдается, что при $\alpha = 10^{-1}$, что при $\alpha = 10^5$. Начиная со значений около $\alpha = 10^6$ и выше, сглаживание начинает сильно изменять вид восстанавливаемой функции. На рисунке 3.9 изображен результат для $\alpha = 10^7$, для того, чтобы указать коэффициент сглаживания достаточно вызвать функцию следующим образом (если параметр не указан, как в предыдущих случаях, то коэффициент сглаживания берется на уровне $\alpha = 10^5$, по-хорошему надо определять оптимальный коэффициент сглаживания, как было сказано в параграфе 3.2, но это следует делать отдельной функцией, подробное обсуждение этого вопроса выходит за рамки данной работы).

```
y=int_spline(t,Y,m=n*3,alpha=10^7)
```



Источник: составлено автором.

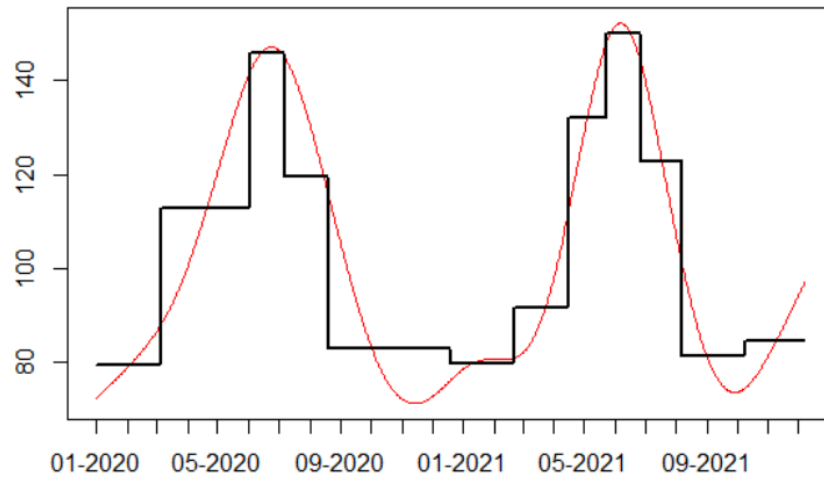
Рисунок 3.9 – Сплайн с равномерно следующими узлами, $m = 3n$, $\alpha = 10^7$

Также мы можем задавать веса конкретным наблюдениям, если мы укажем вес наблюдения 0, то такое наблюдение не будет учитываться при восстановлении функции. По умолчанию все веса заданы единичными. Если мы укажем вес наблюдения больше 1, то этому наблюдению будет уделено больше внимания. Вес наблюдений задается с помощью диагональной матрицы W . На рисунке 3.10 представлен результат восстановления функции, когда 4-тое наблюдение проигнорировано, а 6-тому наблюдению уделяется в 100 раз больше внимание, вызов происходит следующим образом.

```

W=diag(n-1)
W[4,4]=0
W[6,6]=100
y=int_spline(t,Y,m=n*3,W=W)

```



Источник: составлено автором.

Рисунок 3.10 – Восстановление функции по интегралам, вес 4-того наблюдения (ступеньки) задан как 0, вес 6-го наблюдения как 100

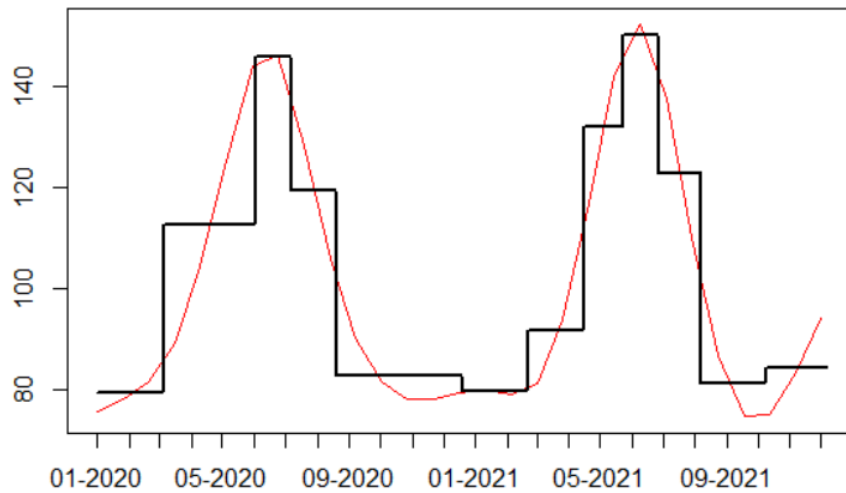
Ранее мы говорили, что функция `int_spline` рассчитывает и возвращает значения сплайна во всех точках от самого первого момента наблюдения t_1 до самого последнего t_n с шагом 1. Чтобы рассчитать сплайн в других координатах, достаточно задать новый вектор x . Например расчет сплайна в каждой 25 точке происходит следующим образом, результат представлен на рисунке 3.11.

```

x=seq(t[1],t[n],25)
y=int_spline(t,Y,m=n*3,x=x)

```

С помощью данного программного инструментария можно очень гибко производить восстановление функции по интегралам.



Источник: составлено автором.

Рисунок 3.11 – Расчет сплайна в каждой 25 точке (через каждые 25 дней)

3.6 Реализация в MS Excel

До сих пор Excel остается очень популярным для анализа и визуального представления данных. Рассмотрим, как реализовать полученный алгоритм восстановления функции по интегралам в Excel (ограничимся только сплайнами с одинаковым расстоянием между узлами).

Первым делом, рисунок 3.12, вставляем данные редких событий (пусть будет всего 25 наблюдений) в столбцы *A* и *B*, оставляем минимум 3 столбца свободными, чтобы в дальнейшем там рассчитать вспомогательные величины. Подсчитываем разницу по времени между первым и последним событием. Задаем количество узлов сплайна, пусть их будет 51. Пусть узлы у нас будут расположены на равномерной сетке с шагом $h_k = h$, рассчитываем величину шага. Далее формируем сами узлы s_k , всего 51 узел.

	A	B	C	D	E	F	G	H	I	J	K
1	date	Value					Общее время T=			679	=A26-A2
2	01.01.2018	1355,076					Количество узлов сплайна m=			51	
3	07.02.2018	1357,701					Шаг h=	13,58	1/h=	0,073637703	=1/H3
4	22.03.2018	1321,186					Коэффициент сглаживания alpha=			1000000	
5	18.04.2018	1298,131					Новые узлы сплайна				=J1/(J2-1)
6	10.05.2018	1284,903					k	S _k			=A2
7	02.06.2018	1313,847					1	01.01.2018			=H7+\$H\$3
8	28.06.2018	1337,623					2	14.01.2018			
9	23.07.2018	1334,214					3	28.01.2018			
10	13.08.2018	1277,61					4	43141,74			
11	03.09.2018	1332,674					5	24.02.2018			
12	27.09.2018	1322,015					6	09.03.2018			
				
24	06.09.2019	1293,613									
25	03.10.2019	1329,85					49	14.10.2019			
26	11.11.2019	1353,229					50	28.10.2019			
27							51	11.11.2019			

Источник: составлено автором.
Рисунок 3.12 – Ввод начальных данных

Далее рассчитываем величины $h_k^{-i} = t_i - s_k$ и $h_k^{+i} = s_{k+1} - t_i$. Для этого вначале определяем на какой интервал попало каждое наблюдение и рассчитываем индекс этого интервала. Индекс последнего наблюдения задаем как номер последнего узла. Рассчитываем h_k^{-i} и h_k^{+i} , рисунок 3.13.

	A	B	C	D	E	F	G	H
1	date	Value	Индекс	h_k^{-i}	h_k^{+i}		Общее время T=	
2	01.01.2018	1355,076	=ПОИСК	0	13,58		Количество узлов сплайна n	
3	07.02.2018	1357,701	=ПОИСКПОЗ(ВПР(A2;\$H\$7:\$H\$57;1);\$H\$7:\$H\$57;0)	3	9,84	3,74	Шаг h=	13,58
4			=A2-ИНДЕКС(\$H\$7:\$H\$57;C2)	6	12,1	1,48	Коэффициент сглаживания	
5			=ИНДЕКС(\$H\$7:\$H\$57;C2+1)-A2		11,94	1,64	Новые узлы сплайна	
6	10.05.2018	1284,903		49	6,79	5,8	k	S _k
7	02.06.2018	1313,847		12	2,02	10,96	1	01.01.2018
8	28.06.2018	1337,623		14	1,46	12,12	2	14.01.2018
			
25	03.10.2019	1329,85		48	1,74	11,84		
26	11.11.2019	1353,229		51			48	01.10.2019
27							49	14.10.2019
							50	28.10.2019
							51	11.11.2019

Источник: составлено автором.
Рисунок 3.13 – Определение интервала и значений h_k^{-i}, h_k^{+i}

Теперь переходим к заполнению матриц. Каждую матрицу лучше задавать на отдельном листе, чтобы ее можно было проще масштабировать.

Для матрицы Q размерностью $m \times (m - 2)$ пронумеровываем индекс строки и столбца, затем в первую ячейку матрицы вводим формулу “=ЕСЛИ(\$A3=B\$2;-2*\$C\$1;ЕСЛИ(\$A3+1=B\$2;\$C\$1;ЕСЛИ(\$A3=B\$2+1;\$C\$1;0)))”. Эту формулу растягиваем на все 51 строк и столбцов, но позже, при обращении к матрице Q , первый и последний столбец брать не будем, рисунок 3.14.

	A	B	C	D	E	F	AAA	AX	AY	AZ
1	Матрица	$h^{-1} =$	0,073638	=Лист1!\$J\$3						
2	Q	1	2	3	4	5		49	50	51
3	1	=ЕСЛИ(\$A3=B\$2;-2*\$C\$1;ЕСЛИ(\$A3+1=B\$2;\$C\$1;ЕСЛИ(\$A3=B\$2+1;\$C\$1;0)))	0,073638	0	0	0		0	0	0
4	2	0,073638	-0,14728	0,073638	0	0		0	0	0
5	3	0	0,073638	-0,14728	0,073638	0		0	0	0
6	4	0	0	0,073638	-0,14728	0,073638		0	0	0
7	5	0	=ЕСЛИ(\$A3=B\$2;-2*\$C\$1;ЕСЛИ(\$A3+1=B\$2;\$C\$1;ЕСЛИ(\$A3=B\$2+1;\$C\$1;0)))	0,14728	0,073638	0		0	0	0
8	6	0	0	0	0	0		0	0	0
50	48	0	0	0	0	0		0,073638	0	0
51	49	0	0	0	0	0		-0,14728	0,073638	0
52	50	0	0	0	0	0		0,073638	-0,14728	0,073638
53	51	0	0	0	0	0		0	0,073638	-0,14728
54										

Источник: составлено автором.
Рисунок 3.14 – Заполнение матрицы Q

Для матрицы R размерностью $(m - 2) \times (m - 2)$ в первой ячейке используем формулу “=ЕСЛИ(\$A3=B\$2;2*\$C\$1/3;ЕСЛИ(\$A3+1=B\$2;\$C\$1/6;ЕСЛИ(\$A3=B\$2+1;\$C\$1/6;0)))”. Опять же протягиваем формулу на все столбцы и строки, но позже не будем использовать первые и последние столбцы и строки (отбросим их), рисунок 3.15.

Матрицу $K = QR^{-1}Q^T$ размерностью $m \times m$ заполняем через матричные операции Excel “=МУМНОЖ(МУМНОЖ(Q!C3:AY53; МОБР('R!C4:AY52)); ТРАНСП(Q!C3:AY53))”, указывая диапазоны ячеек матриц Q и R (не забываем, что для использования матричных операций в Excel надо применять формулу ко всему диапазону, нажимая комбинацию клавиш Ctrl+Shift+Enter, а не протягивать формулы), рисунок 3.16.

	A	B	C	D	E	AAA	AX	AY	AZ	
1	Матрица	h=	13,58	=Лист1!H3						
2	R	1	2	3	4		49	50	51	
3	1	9,053333	2,263333	0	0		0	0	0	
4	2	2,263333	9,053333	2,263333	0		0	0	0	
5	3	0	2,263333	9,053333	2,263333		0	0	0	
6	4	0	0	2,263333	9,053333		0	0	0	
7	5	0	0	0	2,263333		0	0	0	
48	48									
49	47									
50	48	=ЕСЛИ(\$A3=B\$2;2*\$C\$1/3; ЕСЛИ(\$A3+1=B\$2;\$C\$1/6; ЕСЛИ(\$A3=B\$2+1;\$C\$1/6;0)))						2,263333	0	0
51	49						9,053333	2,263333	0	
52	50	0	0	0	0		2,263333	9,053333	2,263333	
53	51	0	0	0	0		0	2,263333	9,053333	

Источник: составлено автором.
Рисунок 3.15 – Заполнение матрицы R

	A	B	C	AAA	AY	AZ
1	Матрица	$K=QR^{-1}Q^T$				
2		1	2		50	51
3	1	=МУМНОЖ	0,00146		-1,3E-30	2,1E-31
4	2	-0,00146	0			-30
52	50	-1,3E-30	7			146
53	51	2,1E-31	-1,3E-30		-0,00146	0,000642

Источник: составлено автором.
Рисунок 3.16 – Расчет матрицы K

Для того, чтобы заполнить матрицу V размерностью $(n - 1) \times m$, рисунок 3.17, надо в начале скопировать с первого листа значения столбцов Индекс h_k^- и h_k^+ . Лучше для этого сослаться на первую ячейку и протянуть ее на все три столбца и на все строки кроме последней. Затем в первую ячейку матрицы ввести большую формулу:

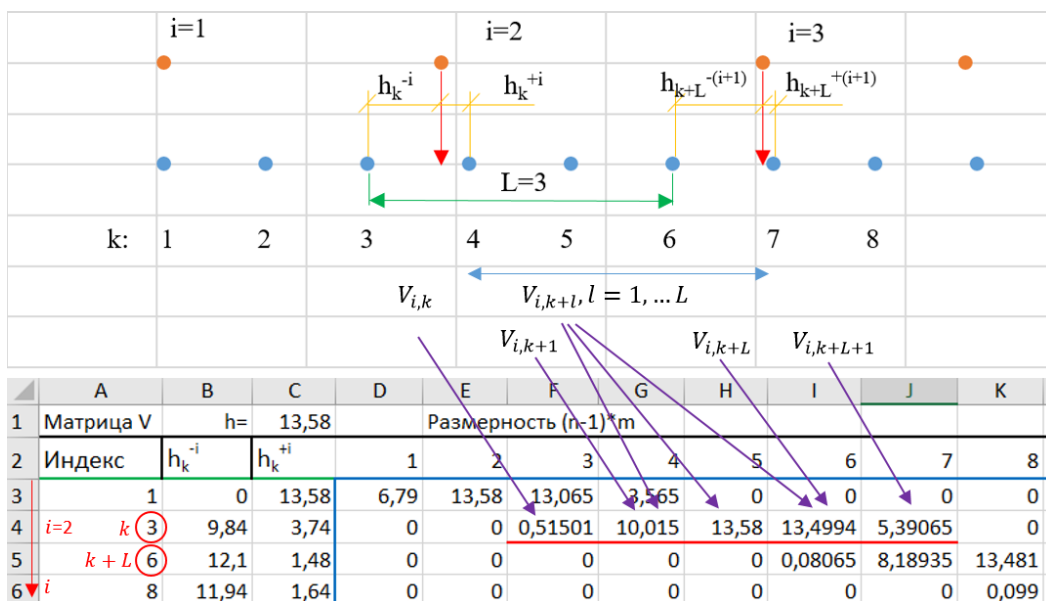
$$\begin{aligned}
 &=ЕСЛИ(\$A3=D\$2;(\$C3^2)/(2*\$C\$1);0) \\
 &+ЕСЛИ((D\$2>\$A3)*(D\$2<=\$A4);\$C\$1;0) \\
 &+ЕСЛИ(D\$2=(\$A4+1);\$B4^2/(2*\$C\$1);0) \\
 &-ЕСЛИ(D\$2=\$A3+1;\$B3^2/(2*\$C\$1);0) \\
 &-ЕСЛИ(D\$2=\$A4;\$C4^2/(2*\$C\$1);0)
 \end{aligned}$$

Эту формулу надо растянуть на все ячейки матрицы. Однако из-за особенностей Excel, последняя ячейка определяется неправильно, чтобы это скорректировать, достаточно в этой последней ячейке все выражение взять в скобки и поделить на 2.

	A	B	C	D	ЕСЛИ(лог_выражение; [значение_если_истина]; [значение_если_ложь])										VA	VB
1	Матрица V	h=	13,58		=Лист1!ИЗ											
2	Индекс	h_k^{-i}	h_k^{+i}	1	2	3	4	5	6	7	8			50	51	
3	1	0	13,58	=ЕСЛИ(\$	13,58	13,065	3,565	0	0	0	0			0	0	
4	=Лист1!С2 протянуть на 3 столбца			0	0	0,51501	10,015	13,58	13,4994	5,39065	0			0	0	
5			7,4	0	0	0	0	0	0,08065	8,18935	13,481			0	0	
6			5,4	0	0	0	0	0	0	0	0,099			0	0	
7	10	6,78	6,8	0	=ЕСЛИ(\$A3=D\$2;(\$C3^2)/(2*\$C\$1);0) +ЕСЛИ((D\$2>\$A3)*(D\$2<=\$A4);\$C\$1;0) +ЕСЛИ(D\$2=(\$A4+1);\$B4^2/(2*\$C\$1);0) -ЕСЛИ(D\$2=\$A3+1;\$B3^2/(2*\$C\$1);0) -ЕСЛИ(D\$2=\$A4;\$C4^2/(2*\$C\$1);0) Протянуть на всю матрицу										0	0
20	30	3,56	1,78	0										0	0	
21	35	3,56	1,78	0										0	0	
22	41	4,8	8,78	0										0	0	
23	43	0,64	12,94	0										0	0	
24	44	8,06	5,52	0										0	0	
25	46	1,9	11,68	0										0	0	
26	48	1,74	11,84	0	0	0	0	0	0	0	0			13,58	6,79	
27	51															

Источник: составлено автором.
Рисунок 3.17 - Расчет матрицы V

Более подробно взглянем на то, как заполняется матрица V, рисунок 3.18. Рассмотрим второе наблюдение $i = 2$. Это наблюдение попадает на третий интервал $k = 3$. Следующее наблюдение $i + 1 = 3$ попадает на шестой интервал $k + L = 6$. Для одного этого наблюдения $i = 2$ надо заполнить элементы $V_{i,k}$ с $k = 3$ по $k + L + 1 = 7$, итого 5 элементов.



Источник: составлено автором.
Рисунок 3.18 – Пояснение, заполнение элементов матрицы V

Для заполнения матрицы P размерностью $(n - 1) \times (m - 2)$, надо скопировать столбцы Индекс, h_k^{-i} и h_k^{+i} , затем в первую ячейку матрицы ввести большую формулу

$$\begin{aligned} &= \text{ЕСЛИ}(\$A3=D\$2; \$C\$1^3/24 - \$B3^2 * (\$C3 + \$C\$1)^2 / \$C\$1/24; 0) \\ &+ \text{ЕСЛИ}((D\$2 > \$A3) * (D\$2 \leq \$A4); \$C\$1^3/12; 0) \\ &+ \text{ЕСЛИ}(D\$2 = \$A3 + 1; \$B3^2 * (\$B3^2 - 2 * \$C\$1^2) / \$C\$1/24; 0) \\ &+ \text{ЕСЛИ}(D\$2 = \$A4; \$C4^2 * (\$C4^2 - 2 * \$C\$1^2) / \$C\$1/24; 0) \\ &+ \text{ЕСЛИ}(D\$2 = \$A4 + 1; \$C\$1^3/24 - \$C4^2 * (\$B4 + \$C\$1)^2 / \$C\$1/24; 0) \end{aligned}$$

Эту формулу также растягиваем на все столбцы и строки, при этом первый и последний столбец использоваться не будет, рисунок 3.19.

	A	B	C	D	E	F	G	H	I	J	K	AAA	AZ	BA	BB	
1	Матрица P	h=	13,58	=Лист1!H3												
2	Индекс	h_k^{-i}	h_k^{+i}	1	2	3	4	5	6	7	8			49	50	51
3	1	0	13,58	104,35	208,7	193,47	80,81	0	0	0	0			0	0	0
4	3	0,84	3,74	0	0	15,229	127,9	208,7	206,2	99,92	0			0	0	0
5			1,48	0	0	0	0	0	2,464	108,8	205,7			0	0	0
6			1,64	0												0
7			6,8	0												0
23	43	0,64	12,94	0												0
24	44	8,06	5,52	0												0
25	46	1,9	11,68	0										3,398	0	0
26	48	1,74	11,84	0										205,3	208,7	208,7
27	51															

Источник: составлено автором.

Рисунок 3.19 – Расчет матрицы P

Все исходные матрицы введены, далее необходимо провести расчеты с этими матрицами. В Excel нет операции вычитания или сложения матриц, поэтому приходится делать больше промежуточных операций, рисунок 3.20.

Скриншоты Excel-таблиц, иллюстрирующие промежуточные расчеты:

- 1. Матрица $PR^{-1}Q^T$ размерностью 24 на 51. Формула: $=МУМНОЖ(МУМНОЖ(Р!Е3:ВA26; МОБР('R'!C4:AY52)); ТРАНСП(Q!C3:AY53))$
- 2. Матрица $C=V-PR^{-1}Q^T$ размерностью 24 на 51. Формула: $=V!D3-PRQ!B3$
- 3. Матрица C^T размерностью 51 на 51. Формула: $=МУМНОЖ(ТРАНСП('C'!B3:AZ26); 'C'!B3:AZ26)$
- 4. Матрица C^T+aK размерностью 51 на 51. Формула: $=СТC!B3+Лист1!$J$4*K!B3$

Источник: составлено автором.
Рисунок 3.20 – Промежуточные расчеты

После завершения промежуточных расчетов возвращаемся на первый лист и рассчитываем искомый столбец значений сплайна g и вторых производных γ , рисунок 3.21.

	A	B	C	D	E	F	G	H	I	J	
1	date	Value	Индекс	h_k^{-1}	h_k^{+1}		Общее время T=			679	
2	01.01.2018	1355,1	1	0	13,58		Количество узлов сплайна m=			51	
3	07.02.2018	1357,7	3	9,84	3,74		Шаг h=	13,58	1/h=	0,073637703	
4	22.03.2018	1321,2	6	12,1	1,48		Коэффициент сглаживания alpha=			1000000	
5	18.04.2018	1298,1	8	11,94	1,64		Новые узлы сплайна	Значение сплайна	Вторая производная		
6	10.05.2018	1284,9	10	6,78	6,8		k	S_k	$g=(C^T+aK)^{-1}C^T \gamma$	$\gamma=R^{-1}Q^T g$	
7	02.06.2018	1313,8	12	2,62	10,96		1	01.01.2018	$=МУМНОЖ(МУМНОЖ($	0	
8	28.06.2018	1337,6	14	1,46	12,12		2	14.01.2018	37,51711206	0,001591728	
9	23.07.2018	1334,2	15	12,88	0,7		3	28.01.2018	33,05203472	0,007613283	
10	13.08.2018	1277,6	17	6,72	6,86		4	43141,74	30,1210213	0,017866014	
11	03.09.2018	1332,7	19	0,56	13,02		5	24.02.2018	30,30719688	0,022340617	
12	27.09.2018	1322	20	10,98	2,6		6	09.03.2018	34,22313741	0,017665985	
13	01.11.2018	1336,2	23	5,24	8,34		7	09.03.2018	34,22313741	0,017665985	
14	08.12.2018	1340	26	1,5	12,08		8	09.03.2018	34,22313741	0,003450983	
15	09.01.2019	1335,9	28	6,34	7,24		9	09.03.2018	34,22313741	-0,010892905	
16	21.02.2019	1361,8	31	8,6	4,98		10	09.03.2018	34,22313741	-0,019308576	
17	29.03.2019	1306,7	34	3,86	9,72		11	09.03.2018	34,22313741	23244	
18	22.04.2019	1293	36	0,7	12,88		12	09.03.2018	34,22313741	41466	
19	14.05.2019	1299	37	9,12	4,46		13	09.03.2018	34,22313741	25744	
20	07.06.2019	1311,4	39	5,96	7,62		14	12.06.2018	36,04323732	0,013004681	
21	03.07.2019	1321,5	41	4,8	8,78		15	26.06.2018	52,18208308	0,014382577	
22	26.07.2019	1290,2	43	0,64	12,94		16	10.07.2018	54,74523436	0,008405046	
23	16.08.2019	1322	44	8,06	5,52		17	23.07.2018	58,60572	-0,005793899	
24	06.09.2019	1293,6	46	1,9	11,68		18	06.08.2018	61,49283311	-0,01689819	
25	03.10.2019	1329,9	48	1,74	11,84		19	19.08.2018	61,62110317	-0,016372485	
26	11.11.2019	1353,2	51				20	02.09.2018	58,72556769	-0,015991589	
								16.09.2018	53,01184315	-0,011351132	
										...	
								49	14.10.2019	37,45890347	0,000988843
								50	28.10.2019	31,32136638	0,000235457
								51	11.11.2019	25,24317053	0

Источник: составлено автором.
Рисунок 3.21 – Расчет значений g и вторых производных γ сплайна

Построение графика сплайна $g(t)$ в произвольный момент времени t происходит через известные значения $g_k, g_{k+1}, \gamma_k, \gamma_{k+1}$, где $k: s_k \leq t \leq s_{k+1}$. Способ построения графика показан на рисунке 3.22.

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Общее время T=			679					=ИНДЕКС(\$H\$7:\$H\$57;M7)					
2	Количество узлов спл:			51					=ИНДЕКС(\$H\$7:\$H\$57;M7+1)					
3	h=	13,58	1/h=	0,07364					=ИНДЕКС(\$I\$7:\$I\$57;M7)					
4	alpha=			1000000					=ИНДЕКС(\$I\$7:\$I\$57;M7+1)					
5	Новые узлы сплайн	Значение сплайн	Вторая производная											Сплайн
6	k	s_k	$g=(C^T C + \alpha \gamma = R^{-1} Q^T g$		t	Индекс k	t_k	t_{k+1}	g_k	g_{k+1}	γ_k	γ_{k+1}		$g(t)$
7	1	01.01.2018	42,4119	0	01.01.2018	1	01.01.2018	14.01.2018	42,41	37,52	0	0,00159	0	42,41189
8	2	14.01.2018	37,5171	0,00159	14.01.2018	1	01.01.2018	14.01.2018	42,41	37,52	0	0,00159	0	42,04786
9	3	28.01.2018	33,052	0,0076	28.01.2018	1	01.01.2018	14.01.2018	42,41	37,52	0	0,00159	0	41,68396
10	4	10.02.2018	30,12	0,0137	10.02.2018	1	01.01.2018	14.01.2018	42,41	37,52	0	0,00159	0	41,32029
11	5	24.02.2018	28,185	0,01787	24.02.2018	1	01.01.2018	14.01.2018	42,41	37,52	0	0,00159	0	40,9697
12	6	09.03.2018	26,9412	0,02234	09.03.2018	1	01.01.2018	14.01.2018	42,41	37,52	0	0,00159	0	40,62412
13	7	23.03.2018	25,78185	0,02787	23.03.2018	1	01.01.2018	14.01.2018	42,41	37,52	0	0,00159	0	40,28185
14	8	06.04.2018	24,78029	0,0345	06.04.2018	1	01.01.2018	14.01.2018	42,41	37,52	0	0,00159	0	39,9412
...
56	50	28.10.2019	31,3214	0,00024	19.02.2018	4	10.02.2018	24.02.2018	30,12	30,31	0,01787	0,02234	0	29,78901
57	51	11.11.2019	25,2432	0	20.02.2018	4	10.02.2018	24.02.2018	30,12	30,31	0,01787	0,02234	0	29,84045
...
685					10.11.2019	50	28.10.2019	11.11.2019	31,32	25,24	0,00024	0	0	25,69022
686					11.11.2019	50	28.10.2019	11.11.2019	31,32	25,24	0,00024	0	0	25,24317

Источник: составлено автором.
Рисунок 3.22 – Построение графика сплайна

Значения столбца $g(t)$ (столбец T на первом листе) и будут являться искомыми значениями восстановленной функции по интегралам, которую уже потом можно вывести в виде диаграммы, анализировать и экстраполировать любыми другими методами.

Выводы по главе 3

1) Разработан математический метод восстановления параметров процессов наполнения/опустошения емкости по выборке редких событий, который базируется на сплайновой коллокации, восстанавливающий неизвестную функцию по последовательности интегралов. Метод базируется на кубических сплайнах в форме представления через значения и значения второй производной (value-second derivative representation) с использованием штрафа на нелинейность (шероховатость, roughness). Особенностью также

является то, что узлы сплайна можно выбирать независимо от точек наблюдения.

2) Представлены результаты исследования по выбору оптимального параметра сглаживания для данного математического метода. Делается вывод, что метод L -кривой не подходит, а метод кросс-валидации не работает в половине случаев. Дается рекомендация использовать принцип невязки Морозова, причем надо брать такой коэффициент сглаживания, который бы приглушал исходные наблюдения на 65% от предполагаемого уровня шума.

3) Разработан программный инструментарий на языке R, реализующий соответствующий математический метод сплайновой коллокации восстановления функции по последовательности интегралов. Представлен код на языке R и продемонстрированы примеры по использованию данного метода.

4) Представлена реализация соответствующего метода сплайновой коллокации с помощью матричных операций в Excel. Подробно и наглядно представлено как заполняется каждая матрица, представлены формулы, которые нужно вводить в ячейки, чтобы рассчитывались значения этих матриц.

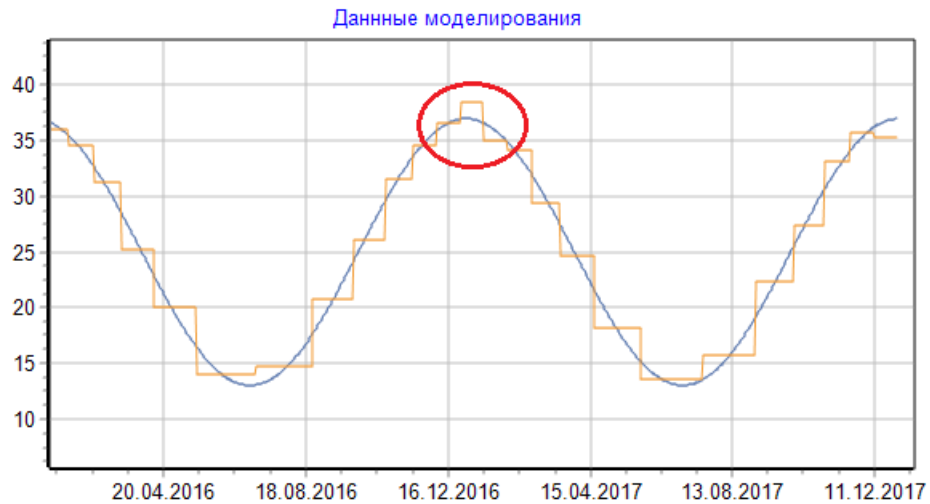
Глава 4

Оценка точности ёмкостного метода

Данная глава будет посвящена вопросам точности. Напомним, что основное предположение ёмкостного метода заключается в том, что наблюдаемая величина y_i есть интеграл от искомой функции $f(t)$ за время от момента совершения этого события t_i до момента времени совершения следующего события t_{i+1} , $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$. Мы разберем, откуда возникает эта погрешность ε_i в наблюдаемых интегралах y_i . Посмотрим, как влияет эта погрешность на точность восстановления исходной функции $f(t)$. Произведем сравнение с существующими методами, предназначенными для анализа и прогнозирования будущих событий. Поговорим о точности прогноза будущих событий. Заметим, что для того, чтобы определять погрешности, нам потребуется эталон, с которым сравнивать. Поэтому исследования базируются на данных имитационного моделирования, когда мы можем априори задать исходную функцию потребления. В то же время результаты моделирования объяснены аналитическими расчетами и численными примерами. Материалы данной главы опубликованы в [104; 105; 106; 107; 108; 109; 110; 111; 112].

4.1 Погрешность наблюдений, связанная с дискретизацией времени

Иногда во время моделирования процессов потребления, когда мы получаем даты t_i и объемы покупок y_i , можно наблюдать следующую картину, рисунок 4.1. Здесь построенная по данным редких событий (покупок) ступенчатая функция, показывающая среднюю скорость расхода продукции (построенная как $\bar{f}(t) = y_i / (t_{i+1} - t_i)$ на каждом интервале $t_i \leq t < t_{i+1}$), на некоторых участках оказывается либо завышенной, либо заниженной (восстановленная функция $\bar{f}(t)$ лежит выше или ниже исходной функции $f(t)$). Как такое может быть, если $\bar{f}(t)$ усредняет исходную функцию $f(t)$ на интервале между двумя покупками?



Гладкая линия – исходная функция потребления $f(t)$, ступенчатая линия – средняя скорость потребления ($\bar{f}(t) = y_i / (t_{i+1} - t_i)$), построенная по наблюдаемым покупкам

Источник: составлено автором.

Рисунок 4.1 – Аномальное усреднение

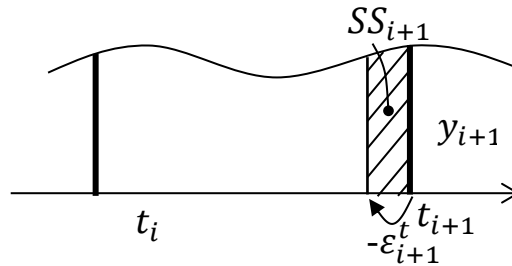
На самом деле это происходит из-за того, что при моделировании (да и в реальности тоже) время покупок измеряется не непрерывно (до миллисекунд), а дискретно. Например, при моделировании процесса потребления момент совершения покупки t_i мы определяем текущим временем, которое при моделировании измеряется дискретно по дням (в моделях потребления или управления запасами чаще всего время изменяется по дням). То есть при моделировании время изменяется целыми сутками, нет никаких промежуточных дробных значений (в реальности как правило также ограничиваются датой покупки). Рассмотрим это более подробно (на всякий случай стоит оговорить, что моменты времени t_i будут обозначать конец i -тых суток).

Рассмотрим процесс образования покупки¹⁾. Когда в момент t_{i+1} запас, сделанный предыдущей i -той покупкой, подходит к концу и опускается ниже критического уровня, а опускается он ниже критического уровня из-за дискретности изменения времени в модели (так как спрос измеряется по дням), то часть продукции берется из запасов в объеме SS_{i+1} . Если бы время было непрерывным, то событие завершения запаса должно было возникнуть чуть

¹⁾ Условимся, что заказ движется мгновенно, нет необходимости содержать дополнительные страховые запасы из-за времени движения заказа. Страховой запас используется для компенсации возможного дефицита.

раньше на величину времени ε_{i+1}^t , рисунок 4.2. Получается, что наблюдаемое время t_{i+1} (дата покупки) оказывается смещенным в большую сторону из-за дискретности исчисления времени. Начиная с момента времени $t_{i+1} - \varepsilon_{i+1}^t$, используются страховые запасы SS_{i+1} до момента проверки, поэтому в истории продаж мы увидим только время t_{i+1} (так как у нас время изменяется дискретно). В итоге объем используемых страховых запасов, формула (4.1)

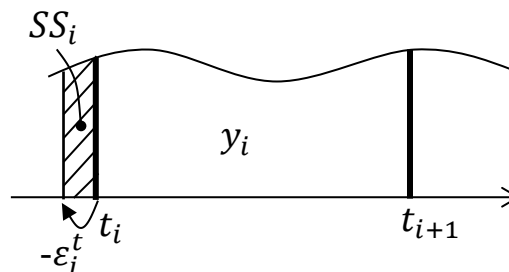
$$SS_{i+1} = \int_{t_{i+1} - \varepsilon_{i+1}^t}^{t_{i+1}} f(t) dt . \quad (4.1)$$



Источник: составлено автором.

Рисунок 4.2 – Запасы должны были закончиться чуть ранее

Теперь рассмотрим момент t_i , когда была сделана покупка y_i из-за того, что запас, пополненный предыдущей покупкой ($i - 1$) закончился. В момент t_i также происходит использование части страховых запасов SS_i . Наблюдаемое время t_i (дата покупки) также получается смещенным в большую сторону, завершение запаса должно было возникнуть чуть раньше на величину времени ε_i^t , рисунок 4.3.



Источник: составлено автором.

Рисунок 4.3 – Для обеих покупок запасы должны были закончиться чуть раньше

Затем, когда определяется объем покупки, то этот объем покупки должен восполнить эти страховые запасы SS_i (после покупки часть отложится в страховые запасы). Поэтому наблюдаемый объем покупки $Q_i = y_i + SS_i$

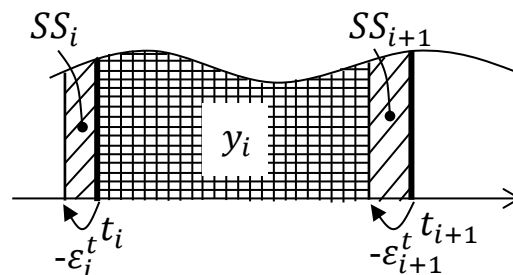
становится завышенным, но до следующей покупки (к моменту времени t_{i+1}) часть SS_i не расходуется, так эта часть продукции откладывается для восполнения страхового запаса (ранее мы говорили, что исходными данными были даты t_i и объемы покупок y_i , но теперь мы говорим, что на самом деле наблюдаемыми покупками являются завышенные объемы Q_i , когда y_i остаются объемами потребления).

Далее, как говорилось выше, момент следующей покупки также смещен $t_{i+1} - \varepsilon_{i+1}^t$, количество товара y_i должно было израсходоваться к этому моменту времени $t_{i+1} - \varepsilon_{i+1}^t$ ($y_i = \int_{t_i}^{t_{i+1} - \varepsilon_{i+1}^t} f(t) dt$). Однако фиксируется лишь момент t_{i+1} . Поэтому к этому моменту времени t_{i+1} будет израсходован объем продукции $y_i + SS_{i+1}$, так как за время от $t_{i+1} - \varepsilon_{i+1}^t$ до t_{i+1} дополнительно расходуется объем SS_{i+1} .

В итоге имеем, наблюдаемый объем покупки $y_i + SS_i$, а реально расходуемое до следующего события количество продукции $y_i + SS_{i+1}$. Обратите внимание на индексы. В итоге количество проданной продукции некоторому клиенту не совпадает с количеством продукции, которое расходует этот клиент на наблюдаемом интервале времени от t_i до t_{i+1}

$$y_i + SS_i \neq y_i + SS_{i+1}. \quad (4.2)$$

То есть мы наблюдаем превышение на SS_i , которое откладывается в запасы и не расходуется, но не наблюдаем SS_{i+1} , которое берется из запасов, так как продукция расходуется до момента t_{i+1} , а не $t_{i+1} - \varepsilon_{i+1}^t$, рисунок 4.4. Причем SS_i и SS_{i+1} изменяются независимо, одна величина может отклониться в большую сторону, когда другая в меньшую.



Источник: составлено автором.

Рисунок 4.4 – Несовпадение наблюдаемого объема $y_i + SS_i$ и расходуемого $y_i + SS_{i+1}$

Определим пределы, в которых изменится страховой запас SS_i . Основной запас продукции мог закончиться как сразу при наступлении нового дня (очередного дискретного момента времени), так и в самом конце дня

$$0 = \int_{t_i-0}^{t_i} f(t)dt < SS_i = \int_{t_i-\varepsilon_i^t}^{t_i} f(t)dt \leq \int_{t_i-1}^{t_i} f(t)dt = f(t_i),$$

$$0 < SS_i \leq f(t_i). \quad (4.3)$$

Есть основание полагать, что величина SS_i распределена по равномерному закону в интервале $(0, f(t_i)]$. Среднее значение используемого страхового запаса тогда будет равно, формула (4.4)

$$\overline{SS}_i \cong f(t_i)/2 \quad (4.4)$$

Несовпадение наблюдаемого и расходуемого объема продукции на интервале от t_i до t_{i+1} может показывать превышение на $f(t_i)$, когда $SS_i = f(t_i)$ и $SS_{i+1} = 0$, или занижение на $f(t_{i+1})$, когда $SS_i = 0$ и $SS_{i+1} = f(t_{i+1})$, формула (4.5)

$$\begin{aligned} \varepsilon_i^Q &= Q_i - \int_{t_i}^{t_{i+1}} f(t)dt = y_i + SS_i - \int_{t_i}^{t_{i+1}} f(t)dt = \\ &= \int_{t_i}^{t_{i+1}-\varepsilon_{i+1}} f(t)dt + SS_i - \int_{t_i}^{t_{i+1}} f(t)dt = \\ &= SS_i - \int_{t_{i+1}-\varepsilon_{i+1}^t}^{t_{i+1}} f(t)dt = SS_i - SS_{i+1} \in (-f(t_{i+1}), f(t_i)). \end{aligned} \quad (4.5)$$

Величина этого несовпадения ε_i^Q , так как является результатом взаимодействия двух равномерных величин, будет распределена по трапецеидальному закону от $-f(t_{i+1})$ до $f(t_i)$. Среднее несовпадение наблюдаемых и реально расходуемых объемов продукции тогда будет

$$\overline{\varepsilon}_i^Q = \frac{f(t_i) - f(t_{i+1})}{2}. \quad (4.6)$$

Если на участках, где исходная функция $f(t)$ возрастает, это несовпадение будет в среднем отрицательным, а там где убывает, в среднем положительным. Обращаю внимание на то, что если в точках t_i и t_{i+1} значения функции совпадают, то разброс величины ε_i^Q происходит

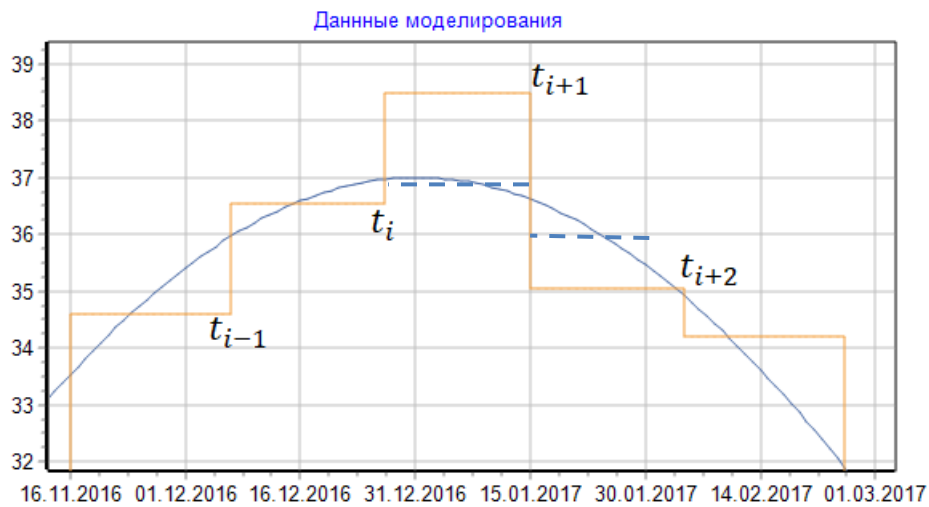
относительно нуля. В других случаях, когда функция $f(t)$ меняется медленно, а покупки происходят относительно часто, и значения функции изменяются не сильно, можно считать, что разброс происходит приблизительно по треугольному закону вокруг нуля.

Как раз из-за величины этого несовпадения ε_i^Q на Рисунке 4.1 появляется характерная аномальная картина усреднения. Отклонение среднего значения $\bar{f}(t)$ из-за дискретности моментов времени будет

$$\varepsilon_{\bar{f}(t)} = \frac{\varepsilon_i^Q}{t_{i+1} - t_i} = \frac{SS_i - SS_{i+1}}{t_{i+1} - t_i} \in \left(-\frac{f(t_{i+1})}{t_{i+1} - t_i}, \frac{f(t_i)}{t_{i+1} - t_i} \right). \quad (4.7)$$

Численный пример

Наши аналитические расчеты подтверждаются численными расчетами. Рассмотрим следующие результаты моделирования, рисунок 4.5, таблица 4.1, и применим к ним полученные формулы для определения погрешности.



Источник: составлено автором.

Рисунок 4.5 – Численный пример, аномальное усреднение

Таблица 4.1 – Численный пример, данные моделирования

Индекс	Дата t_i	Покупка Q_i	SS_i	$\int_{t_i}^{t_{i+1}} f(t)dt$	$\bar{f}(t)$
$i - 1$	06.12.2016	730,84	30,84	731,12	36,54
i	26.12.2016	731,12	31,12	701,25	38,48
$i + 1$	14.01.2017	701,25	1,25	718,43	35,06
$i + 2$	03.02.2017	718,43	18,43	705,14	34,21

Источник: составлено автором.

Примечание - Так как при моделировании клиент пополняет запасы до постоянного максимума запасов, то здесь видно, что объем покупки совпадает с объемом

израсходованной продукции на предыдущем периоде. Однако если клиент изменяет максимальный запас или просто выбирает произвольный объем покупки, то такого наблюдаться не будет. Может сложиться ложное впечатление, что надо произвести сдвиг в данных для большей точности, но это будет ошибкой.

В результате моделирования имеем даты t_i и объемы покупок Q_i . В то же время при моделировании было задано, что потребитель закупал продукцию до своего максимума в 700 единиц. Тогда мы можем определить значения используемых страховых запасов SS_i . Также мы можем посчитать исходное суммарное потребление между двумя покупками $\int_{t_i}^{t_{i+1}} f(t) dt$.

Далее для наблюдения i рассчитываем наблюдаемую среднюю скорость $\bar{f}(t_i)$ и исходную среднюю скорость, после чего определяем ошибку

$$\bar{f}(t_i) = \frac{731,12}{19} = 38,48,$$

$$\frac{\int_{t_i}^{t_{i+1}} f(t) dt}{t_{i+1} - t_i} = \frac{701,25}{19} = 36,91,$$

$$\varepsilon_{\bar{f}(t)} = 38,48 - 36,91 = 1,57.$$

В то же время воспользовавшись полученной нами формулой (4.7), получаем значение

$$\varepsilon_{\bar{f}(t)} = \frac{SS_i - SS_{i+1}}{t_{i+1} - t_i} = \frac{31,12 - 1,25}{19} = 1,57,$$

которое полностью совпадает с наблюдаемым отклонением на участке от t_i до t_{i+1} (наблюдаемое среднее значение $\bar{f}(t)$ превышает истинную среднюю скорость расходования продукции на 1.57), на рисунке 4.5 это тоже заметно, все точки ступенчатой функции лежат выше оригинальной функции $f(t)$.

Аналогично для наблюдения $i + 1$ имеем

$$\bar{f}(t_{i+1}) = \frac{701,25}{20} = 35,06,$$

$$\frac{\int_{t_{i+1}}^{t_{i+2}} f(t) dt}{t_{i+2} - t_{i+1}} = \frac{718,43}{20} = 35,92,$$

$$\varepsilon_{\bar{f}(t)} = 35,06 - 35,92 = -0,86.$$

И воспользовавшись формулой (4.7) через SS_{i+1} и SS_{i+2} , получаем точно такое же значение

$$\varepsilon_{\bar{f}(t)} = \frac{SS_{i+1} - SS_{i+2}}{t_{i+2} - t_{i+1}} = \frac{1,25 - 18,43}{20} = -0,86,$$

которое совпадает с наблюдаемым отклонением на участке от t_{i+1} до t_{i+2} (наблюдаемое среднее значение $\bar{f}(t)$ не достаёт до реальной средней скорости расходования продукции на 0,86), на рисунке 4.5 это тоже заметно, все точки ступенчатой функции лежат ниже оригинальной функции $f(t)$.

Глядя на то, что в таблице 4.1 объёмы покупок совпадают с объёмом потребления за прошлый период, напрашивается алгоритм, который позволит скорректировать ошибку ε_i^0 и $\varepsilon_{\bar{f}(t)}$, а именно можно вместо y_i подставлять y_{i+1} при восстановлении функции $\hat{f}(t)$ и расчете средней скорости $\bar{f}(t)$. Однако такое возможно только при условии, что мы точно знаем, что потребитель не меняет объём покупаемой продукции и всегда придерживается некоторого постоянного уровня максимальных запасов. Но эта информация нам не доступна, более того клиент может менять свой объём покупок произвольным образом, и исходная постановка метода будет продолжать работать. Но если мы произведем сдвиг в данных, а покупатель решит уменьшить или увеличить объём покупки, например в k раз, то на том же наблюдаемом периоде от t_i до t_{i+1} мы получим, что расходуется изменённый в k раз запас продукции, что неверно.

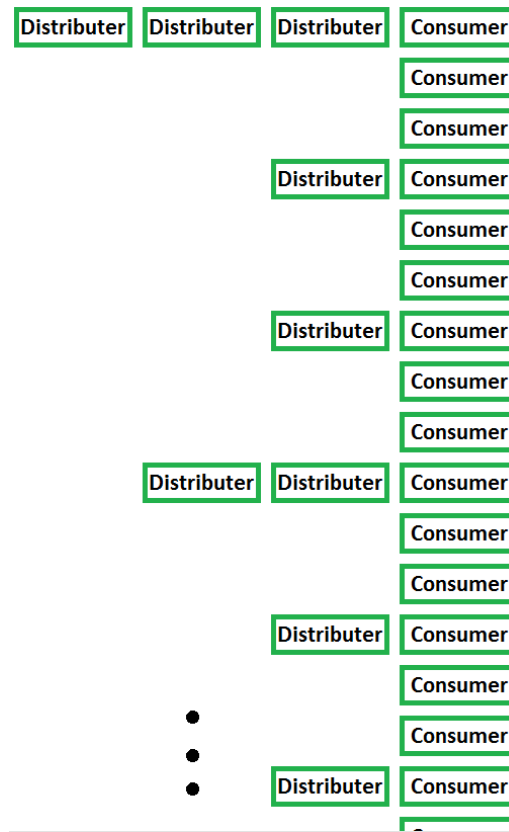
В этом параграфе в результате наших расчетов (возвращаясь к основному предположению $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$) мы определили, что погрешность ε_i , связанная с дискретизацией времени, соответствует погрешности ε_i^0 , которая связана с использованием страховых запасов в дни совершения события, причем эта погрешность распределена по трапецеидальному закону или приблизительно можно считать по треугольному закону относительно нуля.

4.2 Исследование вопросов точности метода в зависимости от удаленности от конечного потребителя с помощью моделирования

В этом параграфе рассмотрим процессы, когда потребление происходит через цепочку промежуточных распространителей. Причем спрос формируется снизу вверх, то есть потребление происходит только конечными потребителями. Когда у конечных потребителей заканчивается продукция, они делают заказ у распространителя, когда у распространителя заканчивается продукция, те в свою очередь делают заказ у вышестоящих распространителей. В логистике такая картина называется реактивным способом управления запасами [65], когда клиенты вытягивают запасы у своих поставщиков по мере необходимости. Будем моделировать такой процесс и фиксировать даты и объемы сделанных каждым участником цепи реализации продукции покупок. Затем по полученным данным будем производить восстановление исходных функций и смотреть как изменяется точность от удаленности от конечного потребителя. Результаты моделирования потом объясним с помощью математического анализа.

Стоит оговорить, что всевозможных схем цепи распространителей может быть большое множество. Схема цепочки поставок может иметь произвольную древовидную (иерархическую) структуру. Количество распространителей, которое товару необходимо последовательно пройти от корневого распространителя (производителя) до конечного потребителя назовем **количеством уровней**. Количество дочерних распространителей или конечных потребителей у выбранного распространителя назовем **количеством потребителей**. Чтобы сконцентрироваться на цели исследования - определение влияния удаленности от конечного потребителя на точность восстановления исходной скорости потребления продукции $f(t)$, для корректного сравнения результатов экспериментов осуществим моделирование для единообразных схем. Такими схемами могут быть такие, когда количество потребителей у всех распространителей одинаково. Например, схема, в которой у каждого распространителя по 3 покупателя,

рисунок 4.6. Меняя количество уровней, будем измерять точность восстановления исходной зависимости.



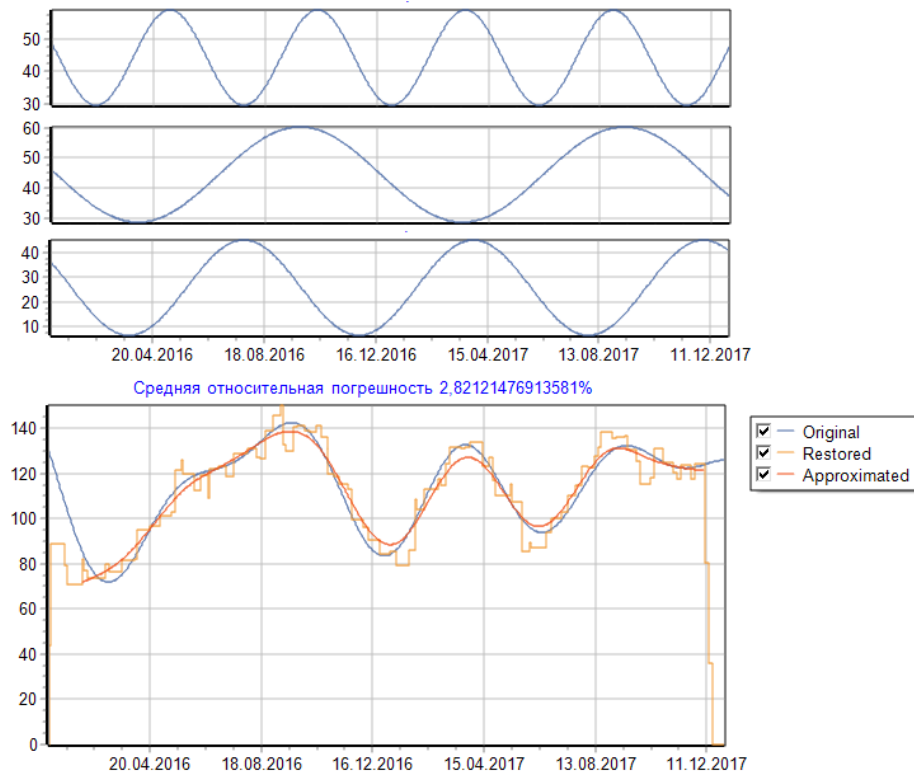
Источник: составлено автором.

Рисунок 4.6 – Схема цепочки распространителей с тремя покупателями

Каждый конечный потребитель может обладать собственным максимальным запасом продукции, критическим уровнем и функцией потребления. Для упрощения пусть на протяжении всего эксперимента максимальный уровень запасов потребителей остается постоянным. В данном эксперименте потребители не меняют поставщиков продукции, хотя на самом деле это не обязательно. Начальный уровень запасов потребителя устанавливается в начале эксперимента, как доля от максимального запаса, либо вручную, либо генерируется случайным образом.

Только у конечных потребителей есть функция скорости расхода продукции $f(t)$, которая показывает, как расходуется их запас продукции со временем. У распространителей такой функции нет, их запас расходуется только по мере того, как у них делают заказ их потребители. Так как нас интересует вопрос точности восстановления исходной скорости расхода

продукции $f(t)$ в произвольной точке цепочки распространителей, нам нужна исходная функция для сравнения. Причем, так как выбранный узел цепочки распространителей содержит свое множество потребителей, а те в свою очередь также содержат дочерних потребителей и т.д., эталонную функцию для осуществления сравнения будем получать суммированием функций скорости потребления всех конечных потребителей, имеющих отношение к осуществлению потребления продукции в выбранном узле. Например, для узла, у которого имеется три дочерних покупателя с разными функциями скорости потребления, сравнение может выглядеть следующим образом, рисунок 4.7. Функции потребления конечных потребителей будем задавать гармоническими функциями с различной частотой, фазой и амплитудой.



Источник: составлено автором.

Рисунок 4.7 – Сравнение восстановленной функции с исходной

Чтобы на выводы о зависимости точности восстановления от позиции в цепочке распространителей ничего кроме самой позиции в этой цепочке не влияло, необходимо обеспечить, чтобы частота покупок товара в разных узлах цепи распространителей была одинаковой. Для этого надо предусмотреть, чтобы максимального запаса продукции каждого распространителя хватало

приблизительно на такое же время, как и запасов дочерних потребителей, не забывая о том, что самих дочерних потребителей может быть множество. Для этого можно сделать так, чтобы максимальный запас распространителя равнялся сумме максимальных запасов его дочерних потребителей. Тогда получится, если у покупателей запасы заканчиваются, например, раз в 10 дней, а самих покупателей, например всего 5, то и у распространителя запасы заканчиваться будут приблизительно те же каждые 10 дней при условии, что его максимальный запас может обеспечить товара для этих 5 покупателей.

Говоря о восстановлении исходной функции скорости потребления $f(t)$ для распространителя, будем понимать способность восстанавливать скорость, с которой расходовалась продукция у выбранного распространителя, на основе данных о покупках этого же распространителя у поставщика высшего уровня. В то же время для поставщика высшего уровня не составит большого труда посчитать суммарную скорость потребления дочерних покупателей. Однако большей информативностью обладает персональная восстановленная функция скорости расхода продукции, так как по ней вышестоящий распространитель может предсказывать даты, когда у его покупателей заканчивается запас продукции.

Теперь опишем процесс одного прогона модели. Для схемы с тремя уровнями распространителей по три покупателя, как на рисунке 4.6, в результате моделирования имеем даты и объемы покупок, совершенных покупателями у своих поставщиков. Для самого близкого к конечным потребителям распространителя (распространителя третьего уровня) информация может иметь такой вид, как в таблице 4.2. Для чистоты эксперимента не будем округлять объем заказа до целого (так как скорость потребления задана вещественными функциями, то количество израсходованной продукции будет не целым числом). Моделирование проводится в течение двух лет. Также не будем использовать случайные числа, скорость потребления строго соответствует своим заданным функциям. Случайные числа будут использоваться для формирования схемы нового

прогона модели, для определения параметров самих функций потребления, максимальных запасов продукции, начальных запасов продукции и др.

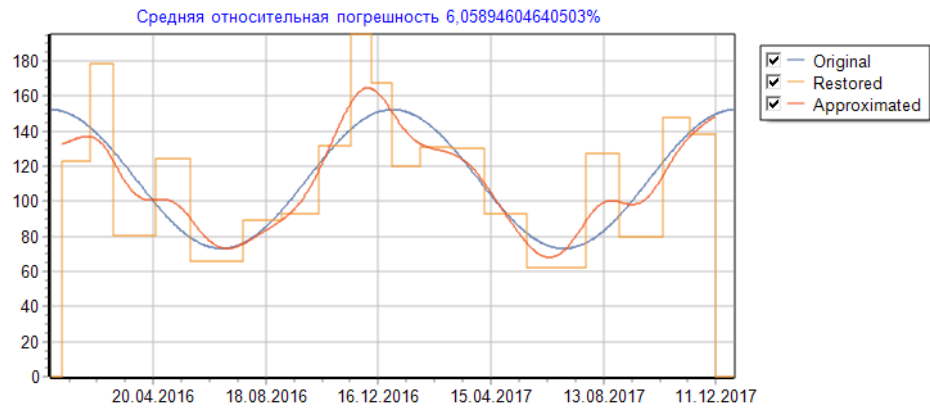
Таблица 4.2 – Данные о продажах у распространителя третьего уровня

Покупатель 1		Покупатель 2		Покупатель 3	
дата	объем	дата	объем	дата	объем
11.01.2016	1185,24	11.01.2016	617,92	13.01.2016	1189,56
30.01.2016	1192,56	24.01.2016	653,34	02.02.2016	1240,30
19.02.2016	1209,91	06.02.2016	631,99	23.02.2016	1237,13
...
16.12.2017	1184,76	25.12.2017	657,97	07.11.2017	1185,83

Источник: составлено автором.

При моделировании будем обращать внимание на количество посредников между наблюдаемым распространителем и конечными потребителями.

Итак, начинаем с распространителей с одним посредником. По данным продаж восстанавливаем исходную функцию потребления, после чего считаем относительную погрешность. Относительную погрешность измеряем от восстановленной функции. Картина может получиться как на рисунках 4.8 - 4.10. Повторяем тоже самое 20 раз для аналогичной схемы, когда у распространителя имеется 1 посредник, но с другими параметрами, т.е. выполняем дополнительные прогоны при фиксированном расстоянии до конечных потребителей. В итоге получаем 20 значений среднего относительного отклонения на двухлетнем временном промежутке. Теперь делаем тоже самое для распространителя, у которого встречается 2 и 3 уровня посредников до конечного потребителя. Полученные результаты занесем в таблицу 4.3, значения отклонений округлены до четвертого знака после запятой. В таблице 4.4 представлены средние значения и квадратичные отклонения по всем 20 прогонам.



Источник: составлено автором.

Рисунок 4.8 – Восстановление исходной функции для распространителя с одним посредником



Источник: составлено автором.

Рисунок 4.9 – Восстановление исходной функции для распространителя с двумя посредниками до конечного потребителя



Источник: составлено автором.

Рисунок 4.10 – Восстановление исходной функции для распространителя с тремя посредниками до конечного потребителя

Таблица 4.3 – Данные 20 прогонов с 1–3 уровнями посредников до конечных потребителей, по 3 покупателя у распространителей

Номер прогона	Среднее относительное отклонение в процентах		
	1 посредник	2 посредника	3 посредника
1	6,0589	8,1665	7,7042
2	4,4492	7,0122	8,5183
3	3,8206	12,1525	8,9015
4	9,6343	6,9123	10,9366
5	5,3771	11,1905	9,6241
6	4,3904	8,2253	10,9213
7	5,7136	9,2977	8,6745
8	6,018	8,3653	8,1735
9	8,688	7,0874	7,0018
10	12,2289	8,7743	4,1247
11	9,7904	8,2171	9,5669
12	6,846	12,9519	6,3553
13	3,3225	11,818	7,9162
14	7,6846	8,6079	6,6991
15	3,1397	6,8984	9,7203
16	4,9831	8,3273	10,0206
17	7,616	4,0429	7,1258
18	10,428	7,8877	7,1145
19	7,3258	4,9267	10,2031
20	4,1924	8,0084	8,8516

Источник: составлено автором.

Таблица 4.4 – Результаты моделирования для схемы распространителей с 3 покупателями у каждого

В процентах

-	1 посредник	2 посредника	3 посредника
Среднее значение относительного отклонения по 20 прогонам, в процентах	6,5854	8,3305	8,3606
Корень выборочной дисперсии относительного отклонения по 20 прогонам, в процентах	2,5562	2,2392	1,6748

Источник: составлено автором.

Глядя на эти результаты, можно сказать, что погрешность увеличивается с удалением от конечного потребителя. Однако при такой дисперсии этот вывод не внушает доверия. Возможно, это увеличение погрешности получилось в результате каких-либо других факторов, а не от изменения позиции в цепочке распространителей.

Повторим эксперимент для схемы, когда у каждого распространителя по 5 покупателей. Результаты 20 прогонов приведены в таблице 4.5 и 4.6.

Таблица 4.5 – Данные 20 прогонов с 1-3 уровнями посредников до конечных потребителей, по 3 покупателя у распространителей

Номер прогона	Среднее относительное отклонение в процентах		
	1 посредник	2 посредника	3 посредника
1	2,8298	5,1647	4,9741
2	3,6861	4,3132	4,4719
3	5,5754	5,6762	7,2097
4	4,4646	6,2413	5,6578
5	3,3521	4,8821	5,6241
6	4,0997	5,7721	7,6642
7	5,9071	6,5638	5,2471
8	4,821	3,9836	6,726
9	2,9903	6,8282	7,3417
10	3,5561	6,1996	4,5723
11	3,7494	6,4211	5,6519
12	4,3705	6,7689	4,1944
13	3,8393	5,4128	6,9989
14	4,9498	4,2735	5,8537
15	2,797	9,586	6,4262
16	3,3409	6,7915	5,1041
17	4,4591	3,269	9,0653
18	5,6656	4,1067	6,4863
19	5,9078	5,5455	5,5274
20	5,4166	5,2883	5,9748

Источник: составлено автором.

Таблица 4.6 - Результаты моделирования для схемы распространителей с 5 покупателями у каждого

В процентах

-	1 посредник	2 посредника	3 посредника
Среднее значение относительного отклонения по 20 прогонам, в процентах	4,2889	5,6544	6,0386
Корень выборочной дисперсии относительного отклонения по 20 прогонам, в процентах	1,0284	1,3967	1,2145

Источник: составлено автором.

Эти результаты более убедительно говорят о том, что погрешность увеличивается с удалением от конечного потребителя. Тем не менее с такой дисперсией сохраняется вероятность того, что такая картина получилась в результате других неизвестных факторов. Примечательно, что рост средних значений погрешности в этих 6 экспериментах сохраняет общую тенденцию. Не было такого, чтобы в одном из экспериментов наблюдалась противоположная картина. Однако в Таблице 4.4 для экспериментов с 2 и 3

посредниками разница составляет всего лишь несколько сотых, что сохраняет некую степень недоверия.

В этих двух сериях экспериментов мы восстанавливали по данным продаж, имеющихся у выбранного распространителя, исходную скорость потребления продукции для его покупателей. Полученные графики показывали, как выбранный распространитель может видеть скорость расхода продукции своего покупателя. Это дает возможность сделать прогноз о дате, когда у его покупателя будет подходить запас продукции к концу. Это дает дополнительные преимущества при планировании собственного плана пополнения запасов или производства, а также дает определенные маркетинговые преимущества, способствующие стабильному сбыту продукции.

Если выбранный распространитель будет суммировать все определенные функции расхода продукции своих покупателей, то он получит некую агрегированную функцию, которая будет показывать с какой приблизительной скоростью у него самого расходуется продукция. Но тут стоит сказать, что эта информация будет иметь меньший практический интерес, так как покупатели делают свои покупки дискретно. Однако по этой функции можно будет прикинуть общую тенденцию, происходящую на рынке, и строить долгосрочные планы.

В этом исследовании мы проводили по 20 прогонов каждого эксперимента. Обращаю внимание, что схема цепочки распространителей представляет собой дерево, где у каждого распространителя есть несколько дочерних распространителей, у тех в свою очередь есть свои покупатели и т.д. В общей сложности количество участников в цепочке распространителей будет равно количеству покупателей у распространителя, возведенное в степень количества уровней. Так для схемы, состоящей из 3 уровней распространителей по 3 покупателя у каждого, составит 27 конечных потребителей, 3+9 промежуточных распространителей. Для схемы из 4 уровней составит 81 конечных потребителей, 3+9+27 промежуточных

распространителей. Если же рассматривать схему, в которой у каждого распространителя по 5 покупателей, то для 3 уровней будем иметь дело с 125 конечными потребителями, 5+25 промежуточных распространителя. Для 4 уровней - 625 конечных потребителя, 5+25+125 промежуточных распространителя. Каждое моделирование проводится в течение 2 лет с фиксацией данных продаж. По данным продаж восстанавливаются графики исходных функций скорости потребления. В общей сложности выходят приличные затраты времени и памяти для проведения каждого прогона. Для больших схем моделирование выполняется достаточно долго. Большое время также занимает просмотр и анализ результатов каждого прогона. Часть времени занимала генерация схемы и генерация параметров, видов функций и начальных значений. В результате по мере проведения экспериментов было решено ограничиться 20 прогонами (для увеличения точности результатов необходимо предварительно значительно оптимизировать процесс моделирования).

Для полноты картины приведем результаты моделирования цепочки с 10 уровнями промежуточных распространителей. Так как количество участников растет экспоненциально от количества уровней в цепочке распространителей, используем схему, где у каждого распространителя всего по 2 дочерних покупателя. Получается модель с $2^{10}=1024$ конечными потребителями продукции и большим числом промежуточных распространителей. Средняя относительная погрешность за 2 года представлена в таблице 12 (так как всего по 2 покупателя у каждого распространителя, заметно уменьшение точности).

Таблица 4.7 – Средняя относительная погрешность, 10 уровней посредников до конечных потребителей (схема с 2 покупателями у каждого распространителя), погрешность

В процентах

Прогоны	Количество посредников										
	0	1	2	3	4	5	6	7	8	9	10
1	5,81	4,77	14,55	10,94	11,11	12,83	10,85	11,78	13,82	10,81	12,40
2	0,85	3,40	13,83	12,25	12,99	13,34	10,48	9,75	11,47	12,04	12,18
3	2,32	7,92	6,86	8,86	6,60	4,86	6,84	7,79	7,65	13,48	13,41

Источник: составлено автором.

По 3 прогонам видим, что точность падает, но падение как будто замедляется.

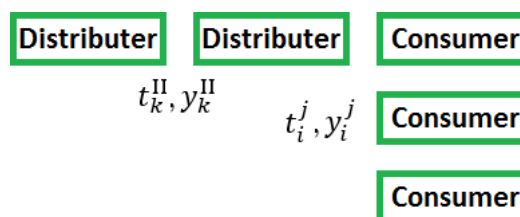
В результате всех экспериментов мы видим, что точность падает с удалением от конечного потребителя, но падает как будто с замедлением (при большом удалении как будто стабилизируется). Восстановление исходной функции скорости потребления по-прежнему происходит достаточно хорошо. Далее попробуем с математической точки зрения объяснить, почему при удалении от конечного потребителя падает точность восстановленной функции и почему это падение замедляется.

4.3 Математическое обоснование падения точности при удалении от конечного потребителя

Требуется определить и объяснить погрешность восстановления исходной функции скорости потребления продукции покупателей в зависимости от позиции в цепочке распространителей. Вначале рассмотрим схему, когда имеется лишь один уровень промежуточных распространителей до конечного потребителя.

Один уровень промежуточных распространителей

Итак, пусть расход продукции j -того конечного потребителя задан функцией $f^j(t)$. Когда запас продукции подходит к концу, конечные потребители совершают покупки у распространителя. В свою очередь у распространителя заканчивается запас продукции, и он также пополняет запас, совершая покупки у распространителя, находящегося на уровень выше. Пусть есть 3 покупателя (посредника), которые покупают продукцию у одного распространителя (верхнего уровня), схема цепочки распространения продукции будет выглядеть как на рисунке 4.11.



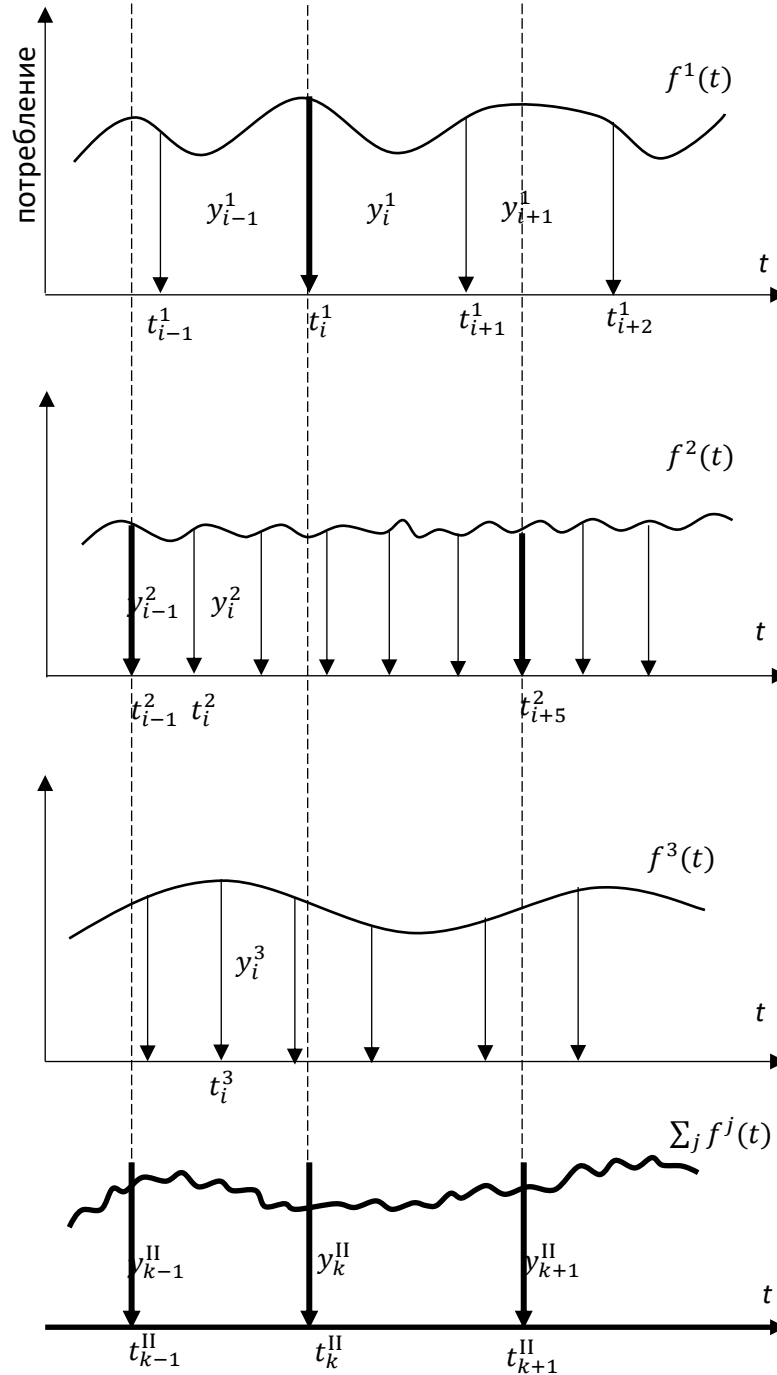
Источник: составлено автором.

Рисунок 4.11 – Один уровень посредников до конечного потребителя

Имеются данные t_i^j , y_i^j времени и объемов покупок конечных потребителей у распространителя, эти данные никак не связаны друг с другом (независимы). Рассмотрим, как производит покупки распространитель, который реализует продукцию этим конечным потребителям. Этот распространитель совершает покупки у вышестоящего распространителя. Пусть первый раз запас продукции распространителя заканчивается в момент t_{k-1}^{II} , когда к нему пришел второй потребитель в момент времени t_{i-1}^2 , рисунок 4.12. Затем потребители совершали другие покупки, и пусть очередной раз запас закончился в момент времени t_k^{II} , когда пришел первый покупатель в момент времени t_i^1 . И так далее. Также имеются данные об объеме совершенной покупки y_k^{II} в моменты времени t_k^{II} . Попробуем определить погрешность определения суммарной скорости потребления продукции $\sum_j f^j(t)$ по отметкам y_k^{II} , t_k^{II} . Нас не интересует уровень запасов посредников и самого распространителя как функция от времени, поэтому привычных пилообразных графиков из моделей управления запасами мы строить не будем, иначе это только усложнит восприятие. На рисунке 4.12 стрелки показывают покупки конечных потребителей, когда у тех заканчивается продукция, жирными стрелками показаны покупки, которые приводят к тому что у распространителя (в самом низу рисунка) заканчивается продукция, и он пополняет ее, делая покупки y_k^{II} , t_k^{II} уже у своего выше стоящего распространителя. Каждая стрелка обозначает покупку, причем объем покупки по основному предположению соответствует интегралу скорости расхода продукции от времени покупки до времени следующей покупки.

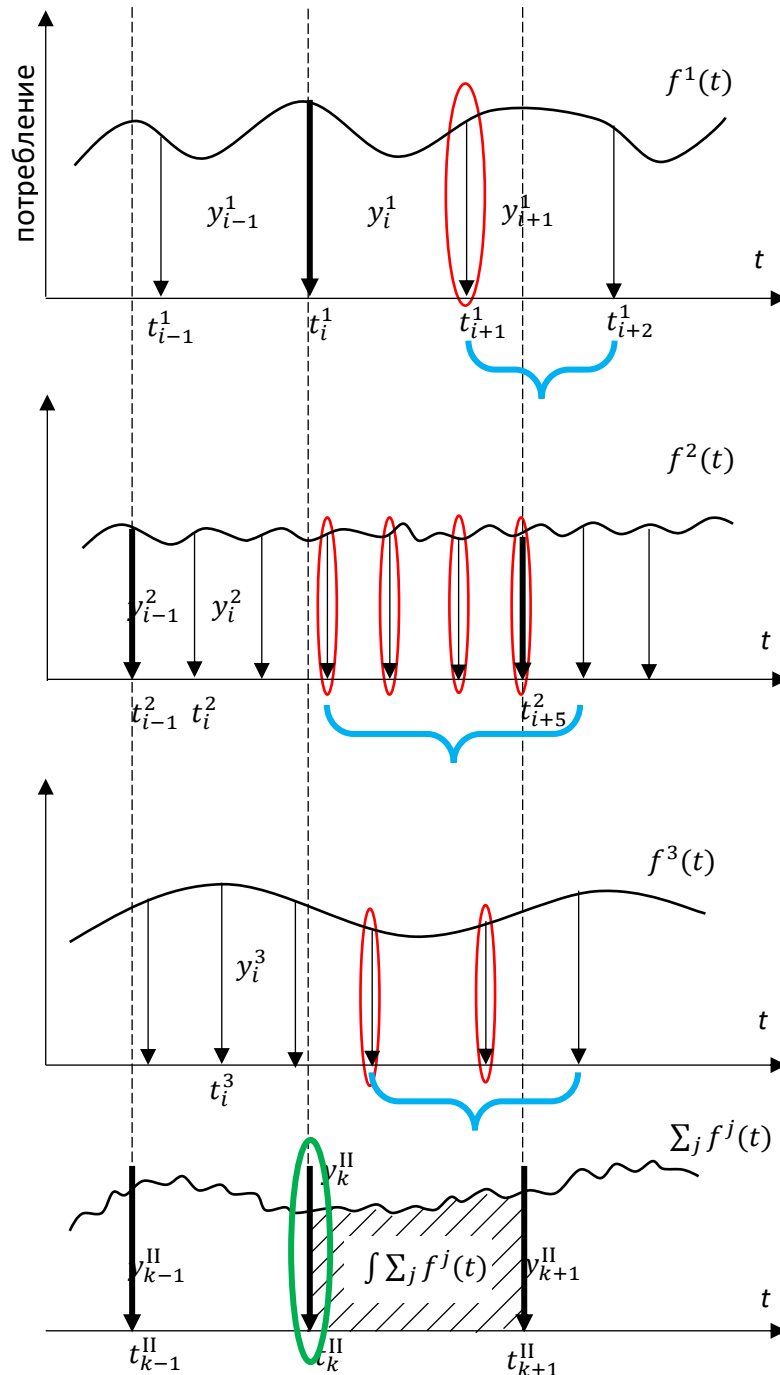
Объема сделанной покупки y_k^{II} хватает на все покупки, совершенные потребителями в период с t_k^{II} до t_{k+1}^{II} , рисунок 4.13. На рисунке овалами обведены покупки y_i^j , на которые хватает запаса y_k^{II} распространителя верхнего уровня, сделанного в момент времени t_k^{II} . Причем покупка y_i^1 не включается, так как она расходует запасы y_{k-1}^{II} , сделанные в предшествующий период времени t_{k-1}^{II} (пока не рассматриваем необходимость использования

страховых запасов). В то же время покупка y_{i+5}^2 , сделанная в момент времени $t_{i+5}^2 = t_{k+1}^{\text{II}}$, включается, так как именно она исчерпывает запасы распространителя (пока без использования страховых запасов) и заставляет его делать следующую покупку y_{k+1}^{II} для восстановления своих запасов.



Источник: составлено автором.

Рисунок. 4.12 – Пополнение продукции с одним уровнем промежуточных посредников



Источник: составлено автором.

Рисунок 4.13 – Запаса y_k^{II} хватает на обозначенные покупки

Для начала будем предполагать, что объема продукции y_k^{II} хватает на все покупки без необходимости использования страхового запаса, а потом усложним и введем страховой запас. Если объема запаса y_k^{II} точно хватает на все покупки, включая последнюю (пока без использования страхового запаса), то можно записать, формула (4.8)

$$y_k^{\Pi} = \sum_j \sum_{i: t_k^{\Pi} < t_i^j \leq t_{k+1}^{\Pi}} y_i^j. \quad (4.8)$$

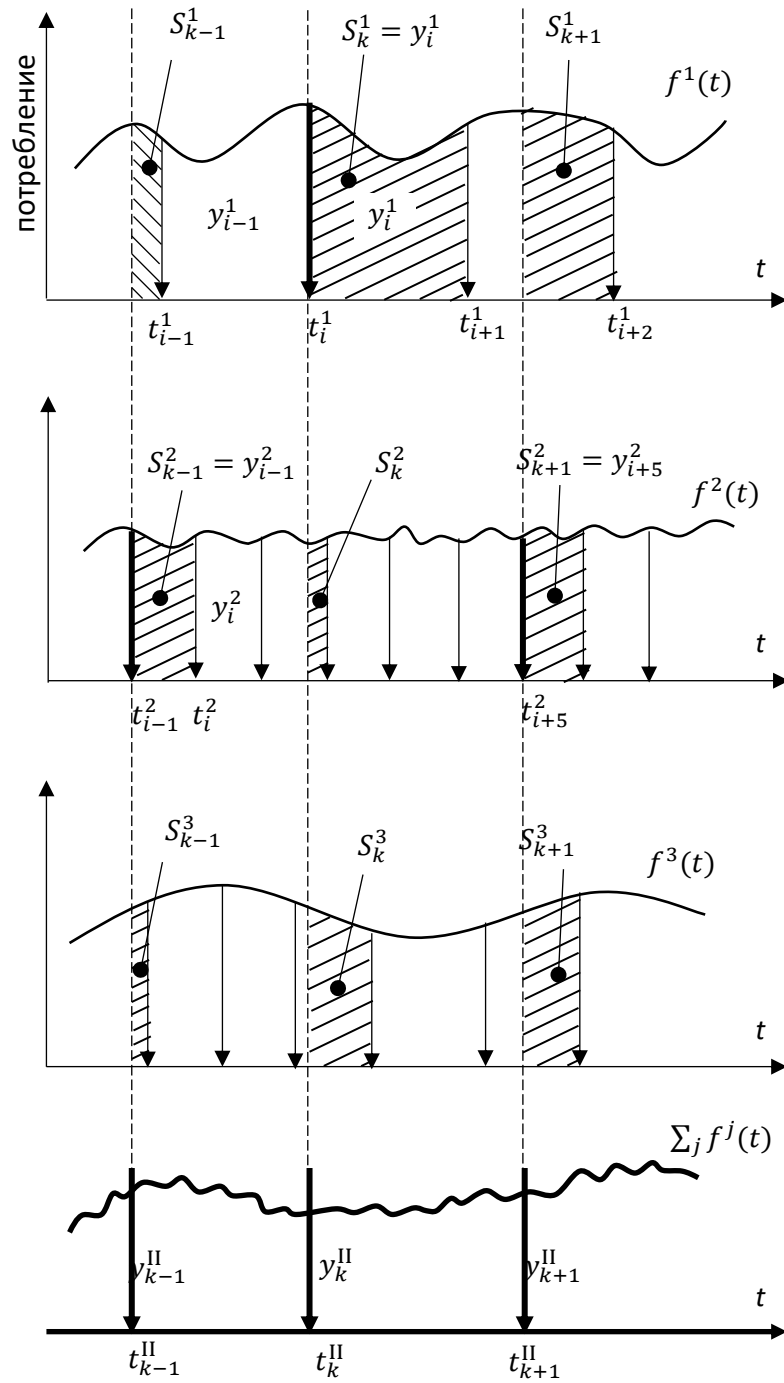
Сравним наблюдаемый объем покупки y_k^{Π} с расходуемым $\int_{t_k^{\Pi}}^{t_{k+1}^{\Pi}} \sum_j f^j(t) dt$ объемом потребления за время от t_k^{Π} до t_{k+1}^{Π} . Еще раз вспомним основное предположение, что объем совершенной покупки y_i^j есть интеграл от функции $f^j(t)$ скорости потребления за время от момента совершения этой покупки t_i^j до момента времени совершения следующей покупки t_{i+1}^j . Так как $\int_a^b f^j(t) dt + \int_b^c f^j(t) dt = \int_a^c f^j(t) dt$, то, формула (4.9)

$$y_k^{\Pi} = \sum_j \sum_{i: t_k^{\Pi} < t_i^j \leq t_{k+1}^{\Pi}} \int_{t_i^j}^{t_{i+1}^j} f^j(t) dt = \sum_j \int_{t_k^{\Pi}}^{t_{\max(i)+1}^j} f^j(t) dt. \quad (4.9)$$

Фигурными скобками на рисунке 4.13 указаны временные промежутки, за которые происходит расходование продукции y_k^{Π} .

В итоге видим, рисунок 4.14, что наблюдаемый из данных продаж объем покупки y_k^{Π} занижен на $\sum_j S_k^j$ и превышен на $\sum_j S_{k+1}^j$ по сравнению с ненаблюдаемым расходуемым объемом продукции $\int_{t_k^{\Pi}}^{t_{k+1}^{\Pi}} \sum_j f^j(t) dt$ на интервале от t_k^{Π} до t_{k+1}^{Π} , где величина S_k^j показывает значение объема потребляемой продукции, относящегося к предшествующей покупке, формула (4.10)

$$S_k^j = \int_{t_k^{\Pi}}^{t_{\min(i)}^j, t_i^j > t_k^{\Pi}} f^j(t) dt. \quad (4.10)$$



Источник: составлено автором.

Рисунок 4.14 – Объем y_k^{II} занижен на $\sum_j S_k^j$, и превышен на $\sum_j S_{k+1}^j$

Величина несоответствия ΔQ_S между y_k^{II} и $\int_{t_k^{II}}^{t_{k+1}^{II}} \sum_j f^j(t) dt$ будет,

формула (4.11)

$$\Delta Q_S = \sum_j S_{k+1}^j - \sum_j S_k^j = \sum_j (S_{k+1}^j - S_k^j). \quad (4.11)$$

Оценим величины S_k^j . Можно увидеть, формула (4.12)

$$0 < S_k^j \leq y_{\min(i)-1}^j = y_{\text{предш.}k}^j \quad (4.12)$$

$$i: t_k^{\text{II}} < t_i^j$$

Величина $y_{\text{предш.}k}^j$ показывает объем предшествующей покупки покупателя j относительно момента времени t_k^{II} , причем, если запас распространителя заканчивается как раз из-за покупки в момент $t_i^j = t_k^{\text{II}}$, то $S_k^j = y_{\text{предш.}k}^j = y_i^j$. Есть основания полагать, что величина S_k^j распределена в интервале $(0, y_{\text{предш.}k}^j]$ по равномерному закону. Тогда среднее значение можно грубо посчитать как, формула (4.13)

$$\overline{S_k^j} \cong \frac{y_{\text{предш.}k}^j}{2}. \quad (4.13)$$

Величина несоответствия ΔQ_S между y_k^{II} и $\int_{t_k^{\text{II}}}^{t_{k+1}^{\text{II}}} \sum_j f^j(t) dt$ лежит в диапазоне, формула (4.14)

$$\Delta Q_S \in \left(- \sum_j y_{\text{предш.}k}^j, \sum_j y_{\text{предш.}k+1}^j \right). \quad (4.14)$$

Стоит сказать, что это только в самом худшем из худших случаев возможно, что ΔQ_S будет близко к границам этого интервала, ведь в начале по каждому покупателю j величины S_{k+1}^j и S_k^j могут компенсировать друг друга, а потом, когда складываются по j могут еще раз компенсировать друг друга. Если S_k^j равномерные случайные величины, то $\Delta Q_S = \sum_j (S_{k+1}^j - S_k^j)$ приближается к нормальному распределению (есть основание полагать, что линейная корреляция между S_{k+1}^j и S_k^j отсутствует).

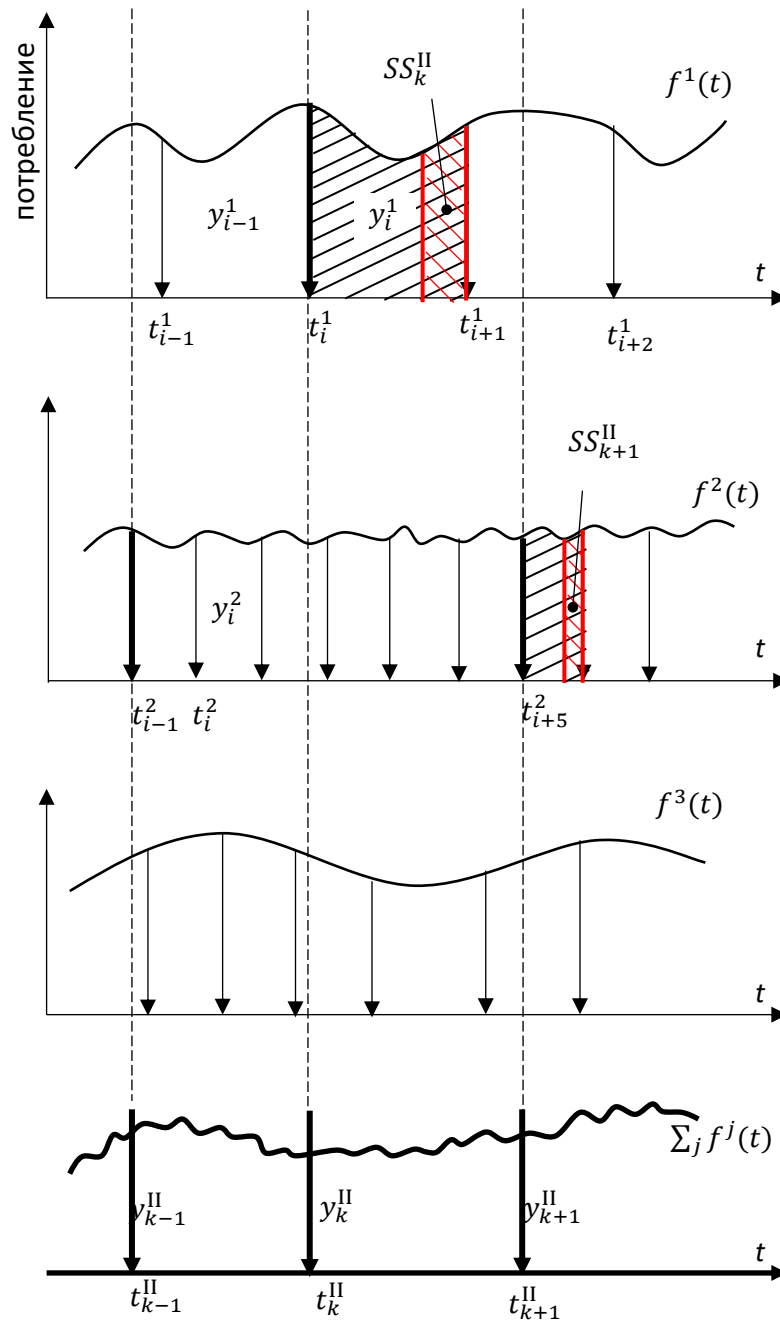
Стоит сказать, что в случае, когда покупатели не меняют свой максимальный запас продукции, т.е. когда величина покупок y_i^j постоянна (пока не считаем восполнение страхового запаса), то никакой дополнительной систематической ошибки возникать не будет, так как среднее значение несоответствия будет равно нулю, формула (4.15)

$$\overline{\Delta Q_S} = \sum_j (\overline{S_{k+1}^j} - \overline{S_k^j}) = \frac{1}{2} \sum_j (y_{\text{предш.}k+1}^j - y_{\text{предш.}k}^j) = 0. \quad (4.15)$$

Но разброс относительно нуля этого несоответствия все равно останется, причем приблизительно по нормальному закону распределения.

Тут стоит сказать несколько замечаний относительно расчета таких характеристик как $S_k^j = \int_{t_k^{\text{II}}}^{t_{\text{min}}^j(i), t_i^j > t_k^{\text{II}}} f^j(t) dt$. Во-первых, эту характеристику можно наблюдать только при моделировании, ведь истинных значений функции расхода продукции $f^j(t)$ и запасов продукции покупателей знать невозможно. Во-вторых, при моделировании надо учесть также дискретность исчисления времени, то есть на самом деле запас должен закончиться чуть раньше на время $\varepsilon_i^{t,j}$ от полной единицы времени, когда происходит проверка и выявляется необходимость пополнения запаса (моделирование проводится по дням, а не миллисекундам). В этом случае каждый покупатель использует собственные страховые запасы $SS_i^j = \int_{t_i^j - \varepsilon_i^{t,j}}^{t_i^j} f^j(t) dt$, которые приводят к тому, что величина S_k^j также должна считаться за интервал времени на $\varepsilon_i^{t,j}$ короче, так как SS_i^j включаются в наблюдаемую покупку y_i^j . Подробнее об этом было рассказано в предыдущем параграфе, картина будет похожа на ту, что описана ниже.

Теперь проанализируем случай использования страховых запасов, когда запасов u_k^{II} распространителя верхнего уровня не хватает точь-в-точь на все покупки и приходится использовать страховые запасы. В моменты t_k^{II} и t_{k+1}^{II} когда приходили покупатель 1 и 2, был превышен критический уровень запаса, и запасы брались из страховых запасов, рисунок 4.15.



Источник: составлено автором.

Рисунок 4.15 – Использование страховых запасов

После того как первый покупатель совершил покупку y_i^1 в момент времени $t_i^1 = t_k^{II}$ наш распространитель выдал часть товара из запаса, затем запас закончился (превышен критический уровень), часть товара SS_k^{II} он взял из страхового запаса. После этой сделки, так как критический уровень запаса был превышен, распространитель совершил покупку $y_k^{II} + SS_k^{II}$ (или сделал заказ) у своего поставщика, из которой часть продукции опять отложилась в

страховой запас SS_k^{II} . Но при этом дальше расходовалась только продукция y_k^{II} , то есть наблюдаем превышение на SS_k^{II} .

Когда в момент времени t_{k+1}^{II} приходит второй покупатель и совершает покупку в момент $t_{i+5}^2 = t_{k+1}^{\text{II}}$, эта покупка приводит к израсходованию запаса, часть продукции берется из запасов, а часть SS_{k+1}^{II} берется из страховых запасов. Общий расход продукции за наблюдаемый интервал времени от t_k^{II} до t_{k+1}^{II} составит $y_k^{\text{II}} + SS_{k+1}^{\text{II}}$.

Имеем дело с несовпадением наблюдаемого сделанного запаса продукции с расходуемым, формула (4.16)

$$y_k^{\text{II}} + SS_k^{\text{II}} \neq y_k^{\text{II}} + SS_{k+1}^{\text{II}}. \quad (4.16)$$

Можно показать, что SS_k^{II} лежит в интервале, формула (4.17)

$$0 < SS_k^{\text{II}} \leq \max_j y_{\text{предш.}k}^j. \quad (4.17)$$

Здесь берем максимум по j , так как покупка любого из конечных потребителей может привести к завершению запаса нашего распространителя и заставит использовать его свои страховые запасы. Есть основания предполагать, что величина SS_k^{II} распределена приблизительно равномерно по интервалу $(0; \max_j y_{\text{предш.}k}^j]$.

Среднее значение можно приблизительно посчитать как, формула (4.18)

$$\overline{SS_k^{\text{II}}} \cong \frac{\sum_{j=1}^n y_{\text{предш.}k}^j}{2n}, \quad (4.18)$$

где n – количество покупателей (в рассматриваемом примере 3).

Несовпадение наблюдаемого и расходуемого объема продукции на интервале времени от t_k^{II} до t_{k+1}^{II} из-за необходимости использовать страховые запасы будет, формула (4.19)

$$\Delta Q^{SS} = SS_{k+1}^{\text{II}} - SS_k^{\text{II}} \in \left(-\max_j y_{\text{предш.}k}^j, \max_j y_{\text{предш.}k+1}^j \right). \quad (4.19)$$

Аналогично предыдущему случаю можно сказать, что если покупатель не склонен изменять объем заказа продукции, то систематической ошибки возникать не будет, разброс будет относительно нуля $\overline{\Delta Q^{SS}} = 0$. Причем если

величина SS_k^{II} распределена равномерно, то так как в образовании несовпадения ΔQ^{SS} участвуют только две величины, распределение погрешности из-за использования страховых запасов будет подчинено трапециевидальному распределению, которое можно приближенно считать треугольным распределением относительно нуля.

В итоге имеем два вида погрешности:

- 1) ΔQ^S - из-за того, что часть покупок идет в зачет другого периода времени;
- 2) ΔQ^{SS} - из-за того, что приходится использовать страховые запасы.

Причем эти два вида погрешности обязательно вычитаются друг из друга, так как объем покупки y_k^{II} занижен на S_k^j и завышен на SS_k^{II} . Это приводит к тому, что общая погрешность ΔQ будет, формула (4.20)

$$\Delta Q = \Delta Q^S - \Delta Q^{SS}. \quad (4.20)$$

Численный пример.

Рассмотрим конкретный числовой пример, где один распространитель продает продукцию трем другим более мелким распространителям, которые работают напрямую с конечными покупателями. На рисунке 4.16 сплошной гладкой линией изображена суммарная исходная функция расхода продукции $\sum_j f^j(t)$, ступенчатой функцией изображена средняя скорость $y_k^{\text{II}} / (t_{k+1}^{\text{II}} - t_k^{\text{II}})$, пунктирными горизонтальными линиями указана средняя истинная скорость расхода продукции.



Источник: составлено автором.

Рисунок 4.16 – Пример расчета погрешности с одним уровнем промежуточных распространителей

У распространителя имеется информация только о времени и объеме покупок, таблица 4.8, столбцы 2 и 3. Общее исходное потребление за время между покупками собрано во время моделирования и представлено в столбце 4. Во время моделирования также можно узнать величину используемых страховых запасов, столбец 5, а также величину оставшейся части покупок, которая относится к предыдущему периоду, столбец 6.

Таблица 4.8 – Результаты численного примера

Индекс	Дата t_k^{II}	Покупка y_k^{II}	$\int_{t_k^{\text{II}}}^{t_{k+1}^{\text{II}}} \sum_j f^j(t) dt$	SS_k^{II}	$\sum_j S_k^j$
$k - 1$	13.10.2016	4336,58	4811,05	736,59	1784,22
k	26.11.2016	4168,35	3506,35	568,35	1141,51
$k + 1$	24.12.2016	3701,08	3696,03	101,08	1336,24
$k + 2$	22.01.2017	4123,84	4470,85	523,84	1764,05
$k + 3$	01.03.2017	-	-	762,38	1655,59

Источник: составлено автором.

При моделировании наблюдается определенная сложность со сбором данных S_k^j , так как потребители тоже используют свои страховые запасы. Для этого надо с момента времени последней покупки t_k^{II} суммировать $f^j(t)$ до момента времени $t_{\min(i)}^j$, $t_i^j > t_k^{\text{II}}$ (следующей покупки покупателя j). После этого от этой суммы надо отнять величину страховых запасов, которые использовал покупатель j перед тем, как совершил покупку $y_{\min(i)}^j$, $t_i^j > t_k^{\text{II}}$ для обслуживания уже своих покупателей (в нашем случае конечных потребителей, потребление которых задается функцией $f^j(t)$).

Для интервала $k - 1$ имеем несоответствие ΔQ в объемах купленной и израсходованной продукции, а также разницу средних значений

$$\Delta Q = y_{k-1}^{\text{II}} - \int_{t_{k-1}^{\text{II}}}^{t_k^{\text{II}}} \sum_j f^j(t) dt = 4336,58 - 4811,05 = -474,47,$$

$$\frac{y_{k-1}^{\text{II}}}{t_k^{\text{II}} - t_{k-1}^{\text{II}}} - \frac{\int_{t_{k-1}^{\text{II}}}^{t_k^{\text{II}}} \sum_j f^j(t) dt}{t_k^{\text{II}} - t_{k-1}^{\text{II}}} = \frac{4336,58}{44} - \frac{4811,05}{44} = 98,56 - 109,34 =$$

$$= -10,78.$$

Наблюдаем занижение на 10,78, на графике также видим соответствующую разницу (ступенчатая функция лежит ниже пунктирной, которая усредняет оригинальную функцию).

Воспользуемся формулами (4.11), (4.19) и (4.20)

$$\Delta Q^S = \sum_j (S_k^j - S_{k-1}^j) = 1141,51 - 1784,22 = -642,71,$$

$$\Delta Q^{SS} = SS_k^{\text{II}} - SS_{k-1}^{\text{II}} = 568,35 - 736,59 = -168,24,$$

$$\Delta Q = \Delta Q^S - \Delta Q^{SS} = -642,71 - (-168,24) = -474,47,$$

$$\Delta Q / (t_k^{\text{II}} - t_{k-1}^{\text{II}}) = -474,47 / 44 = 10,78.$$

Рассчитанное по формулам несоответствие ΔQ в объемах купленной и израсходованной продукции совпадает со значением, полученным при моделировании. Разница средних значений объясняется этим несовпадением

$$\Delta Q / (t_k^{\text{II}} - t_{k-1}^{\text{II}}) = -474,47 / 44 = 10,78.$$

Проделаем аналогичные расчеты для интервала k .

$$\Delta Q = y_k^{\text{II}} - \int_{t_k^{\text{II}}}^{t_{k+1}^{\text{II}}} \sum_j f^j(t) dt = 4168,35 - 3506,35 = 662,$$

$$\frac{y_k^{\text{II}}}{t_{k+1}^{\text{II}} - t_k^{\text{II}}} - \frac{\int_{t_k^{\text{II}}}^{t_{k+1}^{\text{II}}} \sum_j f^j(t) dt}{t_{k+1}^{\text{II}} - t_k^{\text{II}}} = \frac{4168,35}{28} - \frac{3506,35}{28} = 148,87 - 125,23 = 23,64.$$

Наблюдаем превышение на 23,64, на графике также видим соответствующую разницу (ступенчатая функция лежит выше пунктирной, которая усредняет оригинальную функцию).

Воспользуемся формулами (4.11), (4.19) и (4.20):

$$\Delta Q^S = \sum_j (S_{k+1}^j - S_k^j) = 1336,24 - 1141,51 = 194,73,$$

$$\Delta Q^{SS} = SS_{k+1}^{\text{II}} - SS_k^{\text{II}} = 101,08 - 568,35 = -467,27,$$

$$\Delta Q = \Delta Q^S - \Delta Q^{SS} = 194,73 - (-467,27) = 662.$$

Что также полностью совпадает с ранее полученными значениями. Разница средних значений объясняется этим несовпадением

$$\Delta Q / (t_k^{\text{II}} - t_{k-1}^{\text{II}}) = 662/28 = 23,64.$$

Для остальных интервалов $k + 1, k + 2$ значения расчетов ΔQ , полученных по разнице между покупкой и интегралом, и по формулам, также будут полностью совпадать, можете в этом убедиться самостоятельно.

Этот пример показывает справедливость полученных нами формул, поясняет откуда берется разница между восстановленной и оригинальной суммарной функцией скорости потребления. Однако мы еще не ответили на главный вопрос: почему и как с удалением распространителя от конечного потребителя растет относительная погрешность.

Два уровня промежуточных распространителей

Проведем похожие рассуждения относительно распространителя, у которого два уровня посредников до конечного потребителя. Сравним наблюдаемый объем покупки y_L^{III} с $\int_{t_L^{\text{III}}}^{t_{L+1}^{\text{III}}} \sum_r \sum_j f^{r,j}(t) dt$ – суммарным объемом потребления за время от t_L^{III} до t_{L+1}^{III} (r – количество покупателей у распространителя).

На рисунке 4.17 в самом низу изображены покупки распространителя y_L^{III} и y_{L+1}^{III} в моменты времени t_L^{III} и t_{L+1}^{III} . Причем объема покупки хватает на все продажи своим клиентам вплоть до момента времени t_{L+1}^{III} . Однако каждый клиент также реализует свою продукцию через других посредников. Потребителей наших посредников изобразим выше и в стороне. Для удобства восприятия рассмотрим случай, когда только продажи посреднику с номером $r = 1$ приводят к завершению запаса нашего распространителя.

Если объема запаса точно хватает на все покупки, включая последнюю (пока без использования страхового запаса), то можно записать

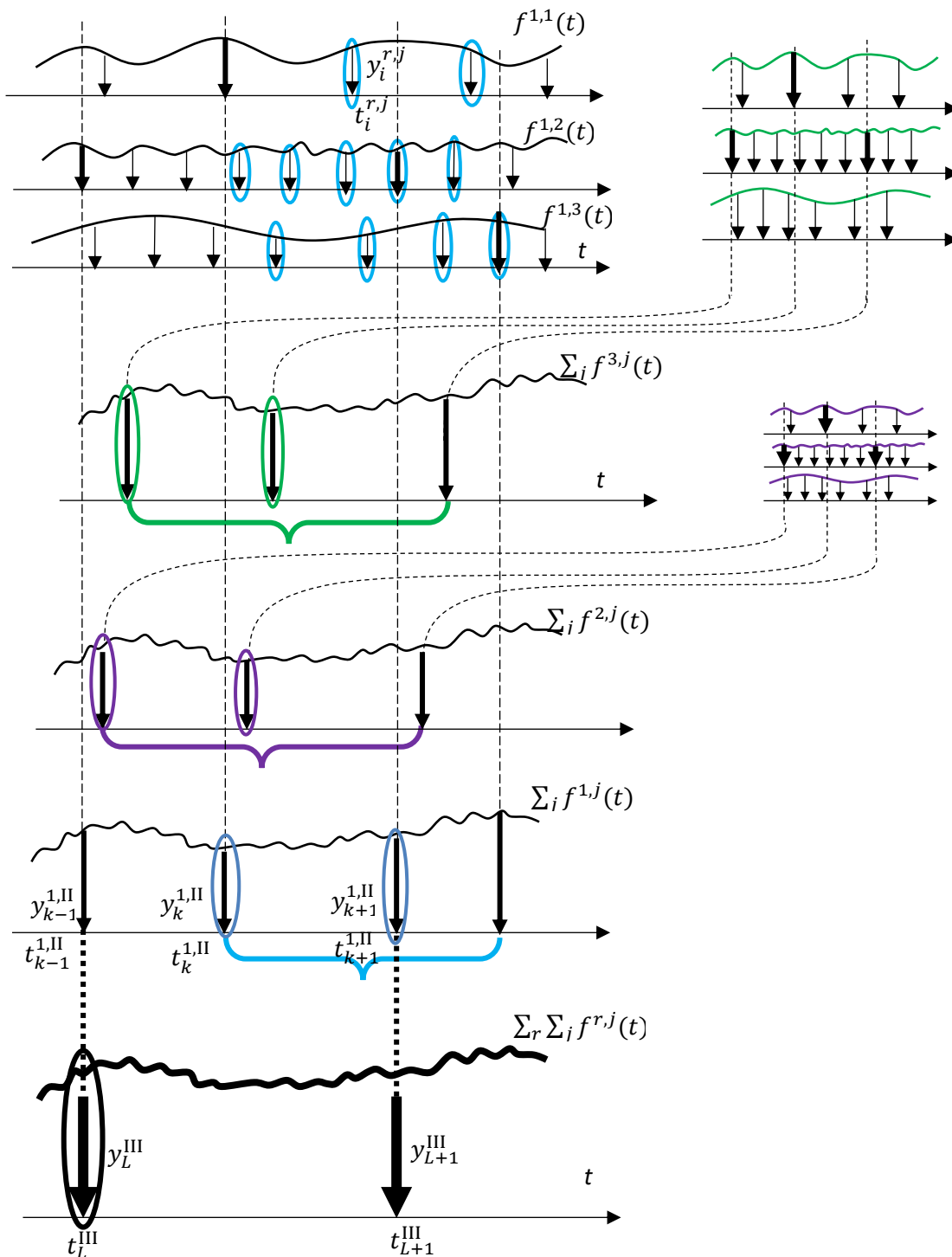
$$y_L^{\text{III}} = \sum_r \sum_{k: t_L^{\text{III}} < t_k^{r,\text{II}} \leq t_{L+1}^{\text{III}}} y_k^{r,\text{II}} = \sum_r \sum_{k: t_L^{\text{III}} < t_k^{r,\text{II}} \leq t_{L+1}^{\text{III}}} \sum_j \sum_{i: t_k^{r,\text{II}} < t_i^{r,j} \leq t_{k+1}^{r,\text{II}}} y_i^{r,j}. \quad (4.21)$$

На рисунке 4.17 обведены овалами покупки $y_k^{r,\text{II}}$ и $y_i^{r,j}$, которые содержатся в большой покупке y_L^{III} (на которые хватает запаса y_L^{III}).

Фигурными стрелками обозначены периоды времени, за которые расходуется этот запас продукции

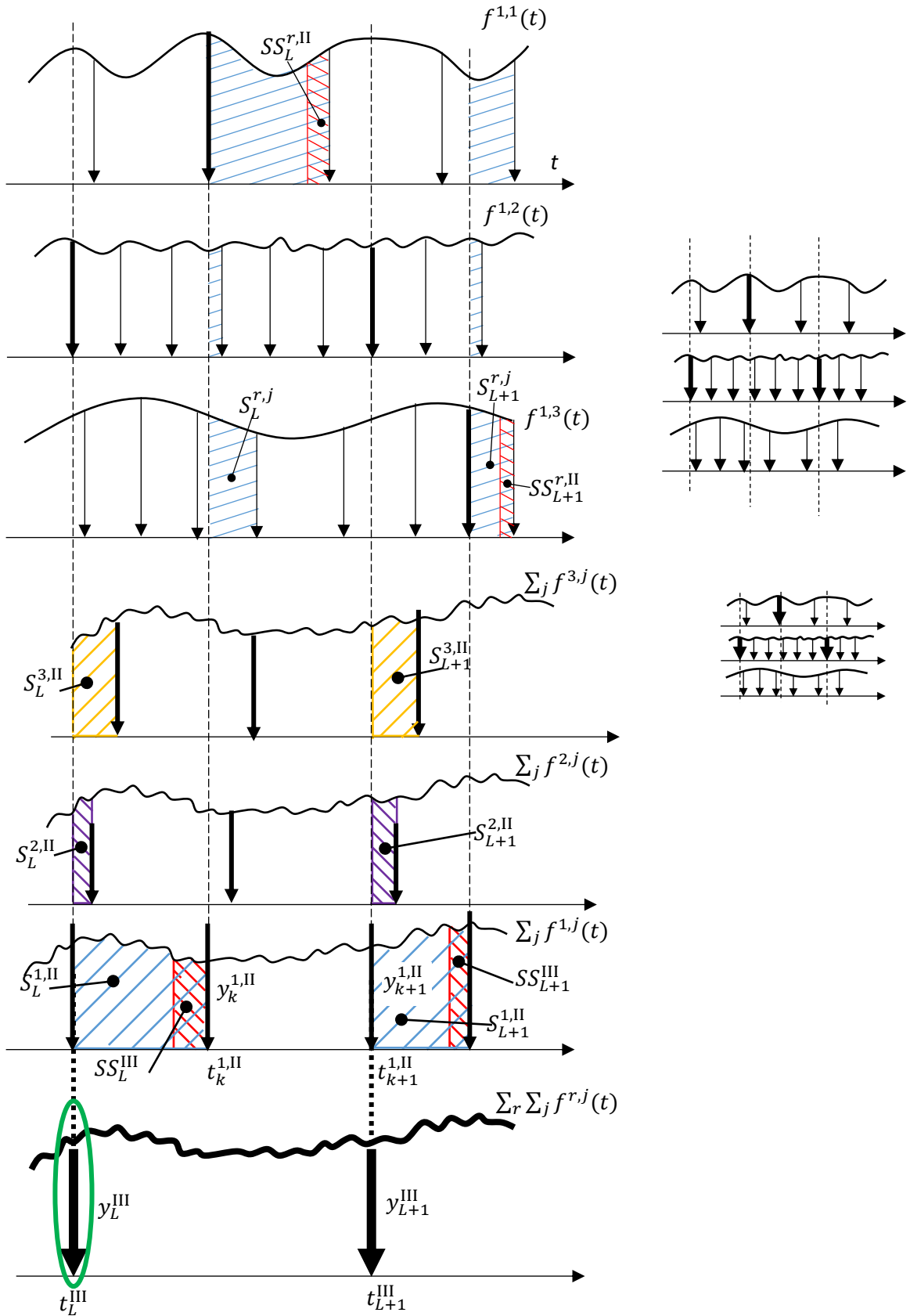
$$y_L^{III} = \sum_r \sum_j \int_{t_{\min(i)}^{r,j}}^{t_{\max(i)+1}^{r,j}} f^{r,j}(t) dt. \quad (4.22)$$

$i: t_k^{r,II} < t_i^{r,j} \leq t_{k+1}^{r,II}$
 $k: t_L^{III} < t_k^{r,II} \leq t_{L+1}^{III}$



Источник: составлено автором.

Рисунок 4.17 – Два уровня посредников до конечного потребителя



Источник: составлено автором.

Рисунок 4.18 – Несоответствие наблюдаемого и расходуемого объема продукции

Опять видим, что объем покупки y_L^{III} занижен на $\sum_r S_L^{r,\text{II}}$ и превышен на $\sum_r S_{L+1}^{r,\text{II}}$,

где

$$S_L^{r,\text{II}} = \int_{t_L^{\text{III}}}^{t_{\min(k)}^{r,\text{II}}, t_k^{r,\text{II}} > t_L^{\text{III}}} \sum_j f^{r,j}(t). \quad (4.23)$$

Но на предыдущем уровне также наблюдается занижение на $\sum_j S_L^{r,j}$ и превышение на $\sum_j S_{L+1}^{r,j}$, аналогично рассмотренному примеру с одним промежуточным распространителем.

Величина несоответствия из-за того, что часть покупок идет в зачет других периодов будет, формула (4.24)

$$\Delta Q_S = \sum_r S_{L+1}^{r,\text{II}} - \sum_r S_L^{r,\text{II}} + \sum_r \sum_j (S_{L+1}^{r,j} - S_L^{r,j}). \quad (4.24)$$

Помимо этого, возникает та же ситуация с использованием страховых запасов, наблюдается завышение на SS_L^{III} и занижение на SS_{L+1}^{III} . Но на предыдущем уровне также возникает потребность использовать страховые запасы, поэтому наблюдается завышение на $SS_L^{r,\text{II}}$ и занижение на $SS_{L+1}^{r,\text{II}}$.

Величина несоответствия в связи с использованием страховых запасов

$$\Delta Q^{SS} = SS_{L+1}^{\text{III}} - SS_L^{\text{III}} + \sum_r (SS_{L+1}^{r,\text{II}} - SS_L^{r,\text{II}}). \quad (4.25)$$

Эти два вида погрешности вычитаются друг из друга, формула (4.26)

$$\Delta Q = \Delta Q^S - \Delta Q^{SS}. \quad (4.26)$$

Величины $S_L^{r,\text{II}}$ и SS_L^{III} имеют такой же интервал значений, относительно предшествующих покупок, как и в случае с одним промежуточным распространителем, формулы (4.27) и (4.28)

$$0 < S_L^{r,\text{II}} \leq y_{\text{предш.}L}^{r,\text{II}}, \quad (4.27)$$

$$0 < SS_L^{\text{III}} \leq \max_r y_{\text{предш.}L}^{r,\text{II}}. \quad (4.28)$$

А величины $S_L^{r,j}$ и $SS_L^{r,\text{II}}$ привязаны к покупкам на предыдущем (ближе к потребителям) уровне и имеют то же самое значение, как и в случае с одним промежуточным распространителем.

Справедливость полученных формул можно подтвердить численным примером, однако он не нагляден, его трудно изобразить одним рисунком, и он занимает много объема. Если есть желание, можете самостоятельно представить ход моделирования с помощью таблиц Excel (это займет более 400 строк и 30 столбцов) или какого-нибудь языка программирования. При расчете величин $S_L^{r,II}$ во время моделирования как сумму значений функции $f^{r,j}(t)$ надо делать корректировку на величину используемых страховых запасов (величина покупки минус максимальный запас), так как они расходуются в текущем периоде, а не предыдущем. С помощью численного примера убеждаемся, что полученные формулы для величины несоответствия наблюдаемой и расходуемой продукции верны.

Теперь нам достаточно абсолютно всего, чтобы можно было объяснить рост погрешности при удалении от конечного потребителя.

Определение скорости роста погрешности с удалением от конечного потребителя

Как было показано выше абсолютное значение погрешности ΔQ определения исходной зависимости зависит от объема совершаемых покупок $u_{\text{предш.}L}^{r,II}$. При отдалении от конечного потребителя объемы покупок растут. Причем если бы в погрешности не участвовали покупки от более низких уровней, то относительная погрешность оставалась бы одинаковой. Но у нас в итоговую погрешность включаются погрешности с предшествующих уровней в цепочке распространения, поэтому погрешность растет.

Стоит заметить, что влияние значения объема покупок предшествующих распространителей на относительную погрешность падает с экспоненциальной скоростью при удалении от конечного потребителя. Это происходит из-за того, что объем покупок распространителей верхнего уровня всегда в несколько раз больше, чем предшествующих распространителей. Вклад предшествующих распространителей в погрешность при расчете

относительной погрешности надо будет делить на значение, зависящее от объемов покупок на текущем уровне.

Рассмотрим еще раз величину несоответствия ΔQ_S из-за того, что часть покупок относилась к другому временному интервалу:

$$\mathbf{1 \text{ уровень посредников: } \Delta Q_S = \sum_j (S_{k+1}^j - S_k^j),}$$

$$\mathbf{2 \text{ уровня посредников: } \Delta Q_S = \sum_r (S_{L+1}^{r,II} - S_L^{r,II}) + \sum_r \sum_j (S_{L+1}^{r,j} - S_L^{r,j}).}$$

Для сокращения записи обозначим разницу в скобках как $dS^j, dS^{r,II}, dS^{r,j}$. Величина несоответствия будет:

$$\mathbf{1 \text{ уровень посредников: } \Delta Q_S = \sum_j dS^j,}$$

$$\mathbf{2 \text{ уровня посредников: } \Delta Q_S = \sum_r dS^{r,II} + \sum_r \sum_j dS^{r,j}.}$$

Если повторить все рассуждения для 3 уровней посредников, то можно показать, что величина несоответствия будет:

$$\mathbf{3 \text{ уровня посредников: } \Delta Q_S = \sum_m dS^{m,III} + \sum_m \sum_r dS^{m,r,II} + \sum_m \sum_r \sum_j dS^{m,r,j}.}$$

Теперь рассмотрим величину несоответствия из-за использования страховых запасов ΔQ_{SS} :

$$\mathbf{1 \text{ уровень посредников: } \Delta Q_{SS} = (SS_{k+1}^{II} - SS_k^{II}),}$$

$$\mathbf{2 \text{ уровня посредников: } \Delta Q_{SS} = (SS_{L+1}^{III} - SS_L^{III}) + \sum_r (SS_{L+1}^{r,II} - SS_L^{r,II}).}$$

Также для сокращения записи сделаем замену:

$$\mathbf{1 \text{ уровень посредников: } \Delta Q_{SS} = dSS^{II},}$$

$$\mathbf{2 \text{ уровня посредников: } \Delta Q_{SS} = dSS^{III} + \sum_r dSS^{r,II}.}$$

Повторяя все рассуждения для трех уровней посредников, можно показать, что величина несоответствия будет:

$$\mathbf{3 \text{ уровня посредников: } \Delta Q_{SS} = dSS^{IV} + \sum_m dSS^{m,III} + \sum_m \sum_r dSS^{m,r,II}.}$$

Величины превышений/занижения наблюдаемой и реально расходуемой продукции зависят от величин предшествующих покупок: S_k^j распределена в интервале $(0, y_{\text{предш.}k}^j]$; SS_k^{II} распределена в интервале $(0, \max_j y_{\text{предш.}k}^j]$. Значит разницы dS^j , dSS^{II} также пропорциональны предыдущим покупкам.

Зафиксируем все условия, чтобы определить влияние только позиции в цепочке распространения на величину ошибки, а все остальные факторы не вносили своего влияния. Для этого предположим, что на каждом уровне посредников, у каждого распространителя одинаковое количество покупателей $\max t = \max r = \max j = n$. Также предположим, что запасов распространителей хватает на одинаковое количество покупок P более мелких распространителей. То есть величина покупки распространителя $y_{\text{предш.}L}^{r,II}$ всегда относится к величине покупки $y_{\text{предш.}k}^j$ (дочерних распространителей) как некоторая постоянная величина P . Тогда величина $dS^{r,II}$ относится к dS^j так же как P , и $dS^{m,III}$ относится к $dS^{r,II}$ как P . Аналогично остальные величины dSS^{IV} , dSS^{III} , dSS^{II} также относятся последовательно друг к другу как P .

Величина несоответствия ΔQ_S в зависимости от количества уровней посредников будет выглядеть следующим образом:

$$\begin{aligned} \mathbf{1 \text{ уровень:}} \quad \Delta Q_S &= \sum_j dS^j = n \cdot dS, \\ \mathbf{2 \text{ уровня:}} \quad \Delta Q_S &= \sum_r dS^{r,II} + \sum_r \sum_j dS^{r,j} = n \cdot P \cdot dS + n^2 dS, \\ \mathbf{3 \text{ уровня:}} \quad \Delta Q_S &= \sum_m dS^{m,III} + \sum_m \sum_r dS^{m,r,II} + \sum_m \sum_r \sum_j dS^{m,r,j} = \\ &= n \cdot P^2 \cdot dS + n^2 \cdot P \cdot dS + n^3 \cdot dS. \end{aligned}$$

Для величины несоответствия ΔQ_{SS} в связи с использованием страхового запаса аналогично получаем следующее:

$$\begin{aligned} \mathbf{1 \text{ уровень посредников:}} \quad \Delta Q_{SS} &= dSS^{II} = dSS, \\ \mathbf{2 \text{ уровня посредников:}} \quad \Delta Q_{SS} &= dSS^{III} + \sum_r dSS^{r,II} = P \cdot dSS + n \cdot dSS, \\ \mathbf{3 \text{ уровня посредников:}} \quad \Delta Q_{SS} &= dSS^{IV} + \sum_m dSS^{m,III} + \sum_m \sum_r dSS^{m,r,II} = \\ &= P^2 \cdot dSS + n \cdot P \cdot dSS + n^2 \cdot dSS. \end{aligned}$$

Общее несоответствие $\Delta Q = \Delta Q_S - \Delta Q_{SS}$ тогда будет:

$$\begin{aligned} \mathbf{1 \text{ уровень:}} \quad \Delta Q &= n \cdot dS - dSS, \\ \mathbf{2 \text{ уровня:}} \quad \Delta Q &= n \cdot P \cdot dS + n^2 dS - P \cdot dSS - n \cdot dSS = \\ &= P \cdot (n \cdot dS - dSS) + n \cdot (n \cdot dS - dSS) = (n \cdot dS - dSS)(P + n), \end{aligned}$$

3 уровня:

$$\begin{aligned}\Delta Q &= n \cdot P^2 \cdot dS + n^2 \cdot P \cdot dS + n^3 \cdot dS - P^2 \cdot dSS + n \cdot P \cdot dSS + n^2 \cdot dSS = \\ &= P^2(n \cdot dS - dSS) + nP(n \cdot dS - dSS) + n^2(n \cdot dS - dSS) = \\ &= (n \cdot dS - dSS)(P^2 + nP + n^2).\end{aligned}$$

Так как нас интересует относительная погрешность, то надо делить величину несоответствия на объем покупки рассматриваемого распространителя. То есть для 1 уровня посредников величину ΔQ , которая пропорциональна объему покупок y_i^j на предшествующем уровне, надо делить на объем покупки y_k^{II} , которое как было условлено будет в P раз больше.

1 уровень посредников

$$\frac{\Delta Q}{y_k^{\text{II}}} = \frac{(n \cdot dS - dSS)}{y_k^{\text{II}}} = \frac{(n \cdot dS - dSS)}{P \cdot y_i^j}, \quad (4.29)$$

2 уровня посредников

$$\begin{aligned}\frac{\Delta Q}{y_L^{\text{III}}} &= \frac{(n \cdot dS - dSS)(P + n)}{y_L^{\text{III}}} = \frac{(n \cdot dS - dSS)(P + n)}{P \cdot y_k^{\text{II}}} = \\ &= \frac{(n \cdot dS - dSS)(P + n)}{P^2 \cdot y_i^j} = \frac{(n \cdot dS - dSS)}{P \cdot y_i^j} \left(1 + \frac{n}{P}\right),\end{aligned} \quad (4.30)$$

3 уровня посредников

$$\begin{aligned}\frac{\Delta Q}{y^{\text{IV}}} &= \frac{(n \cdot dS - dSS)(P^2 + nP + n^2)}{y_L^{\text{IV}}} = \frac{(n \cdot dS - dSS)(P^2 + nP + n^2)}{P^3 \cdot y_i^j} = \\ &= \frac{(n \cdot dS - dSS)}{P \cdot y_i^j} \left(1 + \frac{n}{P} + \left(\frac{n}{P}\right)^2\right).\end{aligned} \quad (4.31)$$

Если продолжить эту последовательность дальше, аналогично повторяя все рассуждения, то получим следующее, формула (4.32)

N уровней посредников

$$\frac{\Delta Q}{y^{[N+1]}} = \frac{(n \cdot dS - dSS)}{P \cdot y_i^j} \left(1 + \frac{n}{P} + \left(\frac{n}{P}\right)^2 + \dots + \left(\frac{n}{P}\right)^{N-1}\right). \quad (4.32)$$

Выражение в скобках является суммой геометрически убывающей прогрессии при условии, что величина $\left(\frac{n}{P}\right)$ меньше 1. Эта величина будет меньше единицы, так как P показывает на сколько покупок хватает запаса продукции распространителя, а n показывает количество покупателей у распространителя. Только в случае, если $n \geq P$ выражение в скобках не будет убывающей прогрессией. Это возможно, если объема покупок распространителя не хватает, чтобы удовлетворить лишь по одной покупке своих покупателей, то есть объем покупок покупателей превосходит доли $1/n$ от объема покупки распространителя.

Расчет скорости роста дисперсии погрешности с удалением от конечного потребителя

При увеличении количества уровней посредников увеличивается общее количество задействованных покупателей, и увеличивается общее количество случайных (неопределенных) величин. Если дисперсия будет расти как сумма убывающей геометрической прогрессии, то относительное среднее квадратичное отклонение должно расти как корень из этой дисперсии, то есть еще медленнее (ошибки могут компенсировать друг друга). Однако большая часть вклада в дисперсию осуществляется первыми $2(n+1)$ случайными величинами (по 2 величины на значения dS и dSS), вклад следующих $2(n^2+n)$ случайных величин падает в P раз, а очередных $2(n^3+n^2)$ падает уже в P^2 раз, и так далее. Закон распределения относительной погрешности не будет соответствовать нормальному распределению (только приближенно), так как в нормальном распределении предполагается, что вклад всех случайных величин одинаков. Закон распределения относительной погрешности будет характеризоваться в основном суммой первых $2(n+1)$ равномерных случайных величин (распределенных от 0 до $\frac{1}{P}$, так как каждая величина лежит в диапазоне от 0 до $y_{\text{предш.к}}^j$, а затем она делится на y_k^{II} , которое в P раз больше).

Попробуем определить дисперсию относительного отклонения $\frac{\Delta Q}{y^{[N+1]}}$. Дисперсия базовых равномерных величин равна $1/12$. Тогда суммарная дисперсия первых $2(n+1)$ величин будет $\frac{2(n+1)}{12P^2}$ (дисперсия суммы/разницы независимых случайных величин равна сумме дисперсий, умножение случайной величины на константу приводит к увеличению дисперсии в константу в квадрате раз). Следующие $2(n^2+n)$ случайных величин увеличивают дисперсию на $\frac{2(n^2+n)}{12P^4}$, а очередные $2(n^3+n^2)$ случайные величины увеличивают дисперсию на $\frac{2(n^3+n^2)}{12P^6}$, и так далее.

В результате дисперсия относительного отклонения в связи с наличием N уровней посредников будет, формула (4.33)

$$\begin{aligned} D \left[\frac{\Delta Q}{y^{[N+1]}} \right] &= \frac{2(n+1)}{12P^2} + \frac{2(n^2+n)}{12P^4} + \frac{2(n^3+n^2)}{12P^6} + \dots + \frac{2(n+1)n^{N-1}}{12P^{2N}} = \\ &= \frac{n+1}{6P^2} \left(1 + \frac{n^1}{P^2} + \frac{n^2}{(P^2)^2} + \dots + \frac{n^{N-1}}{(P^2)^{N-1}} \right). \end{aligned} \quad (4.33)$$

Выражение в скобках опять является суммой геометрически убывающей прогрессии. Сумма первых N слагаемых будет $\frac{1-q^N}{1-q}$, где $q = \frac{n}{P^2}$ - знаменатель прогрессии. В результате можем получить значение дисперсии для относительного отклонения, формула (4.34)

$$D \left[\frac{\Delta Q}{y^{[N+1]}} \right] = \frac{(n+1) \left(1 - \left(\frac{n}{P^2} \right)^N \right)}{6P^2 \left(1 - \frac{n}{P^2} \right)}. \quad (4.34)$$

Среднее квадратичное отклонение получаем как корень из дисперсии:

$$\sigma \left[\frac{\Delta Q}{y^{[N+1]}} \right] = \sqrt{\frac{(n+1) \left(1 - \left(\frac{n}{P^2} \right)^N \right)}{6P^2 \left(1 - \frac{n}{P^2} \right)}}. \quad (4.35)$$

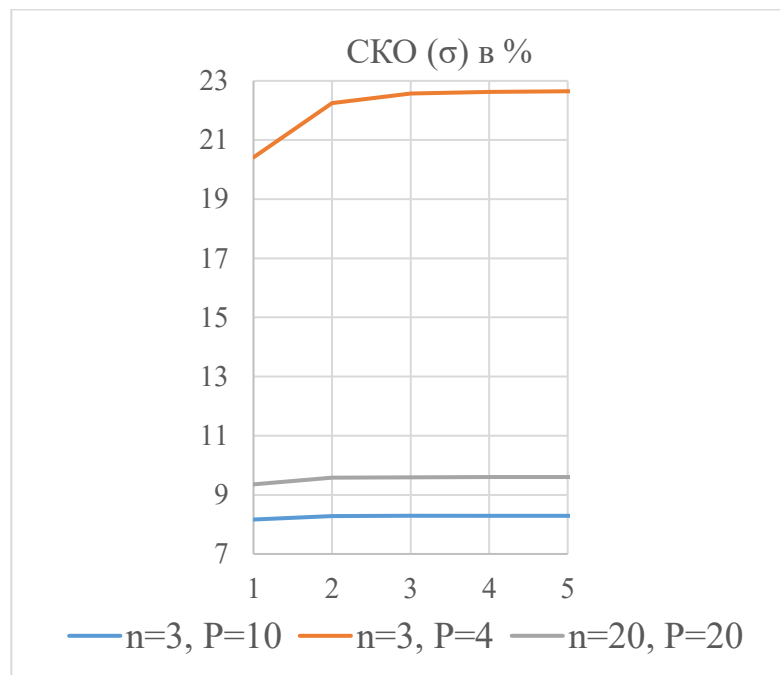
Для рассмотренного абстрактного примера, когда у распространителя по 3 покупателя ($n = 3$ более мелких распространителей), если объем покупок распространителей при удалении от конечного потребителя растет как $P = 10$,

то можем посчитать дисперсию и среднее квадратичное отклонение для относительного отклонения, таблица 4.9, столбец 1. График зависимости относительной погрешности от удаленности от конечных потребителей представлен на рисунке 4.19. Как видно из таблицы и графика, относительная погрешность растет незначительно, уже при удаленности на 3 позиции погрешность изменяется лишь в четвертом знаке.

Таблица 4.9 – Расчет дисперсии и среднего квадратичного отклонения относительной ошибки при удалении от конечного потребителя

N	n = 3, P = 10		n = 3, P = 4		n = 20, P = 20	
	D	σ в процентах	D	σ в процентах	D	σ в процентах
1	0,006666667	8.164966	0,041666667	20.412415	0,00875	9.3541435
2	0,006866667	8.286535	0,049479167	22.243910	0,0091875	9.5851448
3	0,006872667	8.290155	0,05094401	22.570780	0,009209375	9.5965489
4	0,006872847	8.290263	0,051218669	22.631542	0,009210469	9.5971187
5	0,006872852	8.290267	0,051270167	22.642917	0,009210523	9.5971472

Источник: составлено автором.



Источник: составлено автором.

Рисунок 4.19 – Относительная погрешность в зависимости от удаленности от конечных потребителей (длины цепочки посредников), в процентах

Подведем итоги. Получены формулы (4.20) и (4.26) для расчета дополнительной ошибки восстановления исходной функции в зависимости от позиции в цепочке распространителей. Получены значения интервалов, в

которых лежит ошибка. Ошибка имеет нулевое математическое ожидание при неизменности исходных факторов. На численном примере показана справедливость формул. Определено, что относительная погрешность при удалении от конечного потребителя (без влияния других факторов) будет расти приблизительно также, как растет сумма геометрически убывающей прогрессии. Однако так как элементы суммы являются случайными величинами, дисперсия и среднее квадратичное отклонение растет еще медленнее. Получены формулы (4.33–4.35) для их расчета в идеальных условиях, когда никакие другие факторы не оказывают никакого воздействия. Показано, что при удалении от конечного потребителя дисперсия и среднее квадратичное отклонение для относительной ошибки растет очень медленно. Однако даже в худшем случае, когда все случайные величины накладываются и ошибка увеличивается, а не компенсируется, погрешность все равно не превысит суммы геометрической убывающей прогрессии.

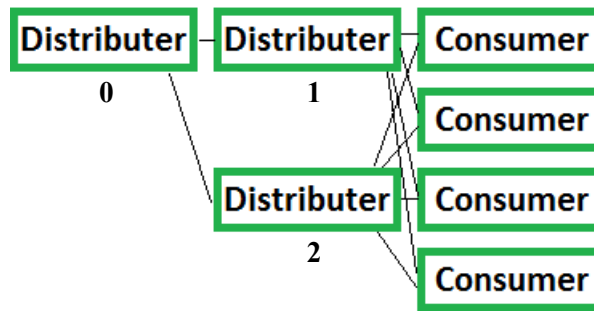
В этом параграфе в результате наших расчетов (возвращаясь к основному предположению $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$) мы определили, что погрешность ε_i , связанная с удаленностью от конечного потребителя, образуется из-за несоответствия наблюдаемой и реально расходуемой продукции $\Delta Q = \Delta Q^S - \Delta Q^{SS}$. Это несоответствие связано с тем, что часть покупок относится к другому временному интервалу, а также с использованием страховых запасов в дни совершения события. Эту погрешность можно приближенно считать нормальной случайной величиной с нулевым математическим ожиданием (т.к. имеется сумма большого числа случайных величин, разбросанных у нуля). Дисперсия этой ошибки при удаленности от конечного потребителя растет, но растет как геометрически убывающая прогрессия. Среднее квадратичное отклонение, как корень из этой дисперсии, растет еще медленнее. Получены формулы для расчета дисперсии и среднего квадратичного отклонения этой ошибки.

4.4 Изучение влияния конкуренции или потери части данных на точность восстановления исходной функции

Рассмотрим, как влияет потеря части данных на точность восстановления исходной функции скорости опустошения/наполнения емкости $f(t)$ (скорости потребления). Под потерей данных понимается случай, например, когда клиент решает совершить покупки не у того же самого распространителя, а у другого. На самом деле это приводит лишь к тому, что в выборке данных о продажах у распространителя будет отсутствовать часть данных. Эта картина очень легко моделируется. Например, при моделировании процесса расхода продукции можно с указанной вероятностью осуществлять покупку у одного из множества поставщиков. В реальности это действительно происходит из-за конкуренции, когда клиенты меняют свое предпочтение между поставщиками. То есть можно говорить, что вследствие конкуренции происходит потеря части данных в истории покупок клиентов у распространителей. Если же брать лишь одного клиента, то моделировать потерю данных еще проще, достаточно из выборки с заданной вероятностью исключить каждую порцию данных о покупке.

Влияние потери данных на поведение восстановленной функции

Прежде чем говорить о точности, рассмотрим, например, следующую простую схему, рисунок 4.20. Четыре конечных потребителя могут пополнять свои запасы у одного из двух вышестоящих распространителей (распространители, отмеченные номерами 1 и 2), которые в свое время могут пополнять свой запас продукции только у одного вышестоящего распространителя (распространитель, отмеченный номером 0). Пусть вероятность выбора конечными потребителями одного из распространителей будет равняться 50%. Также пусть исходная функция скорости расхода продукции у самого первого потребителя отличается частотой от скорости расхода продукции остальных трех (это нужно, чтобы лучше было видно, к чему приводит переключение потребителей).

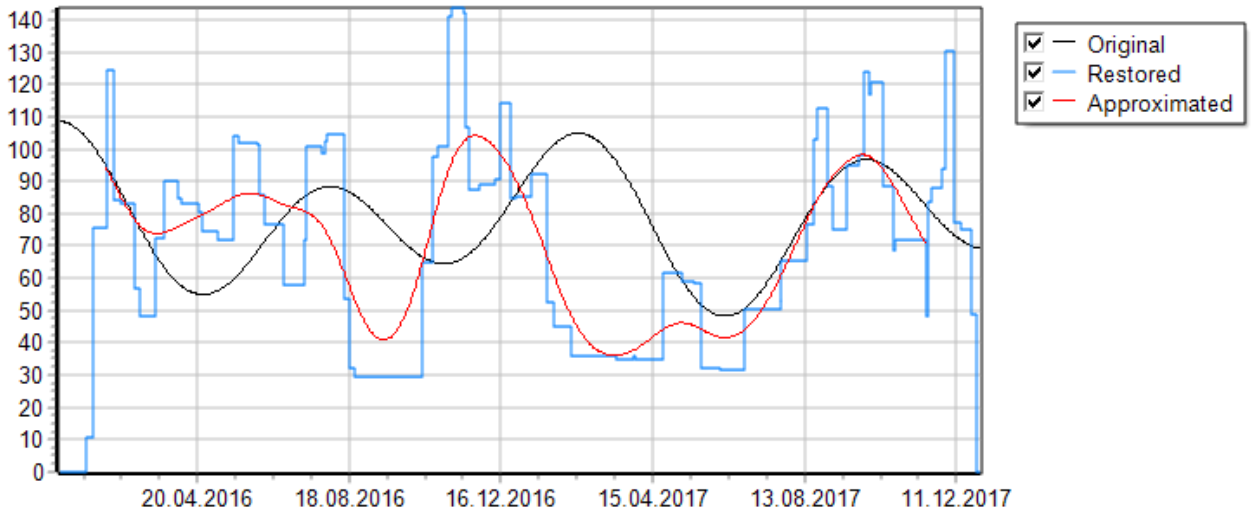


Источник: составлено автором.

Рисунок 4.20 – Схема цепочки распространителей

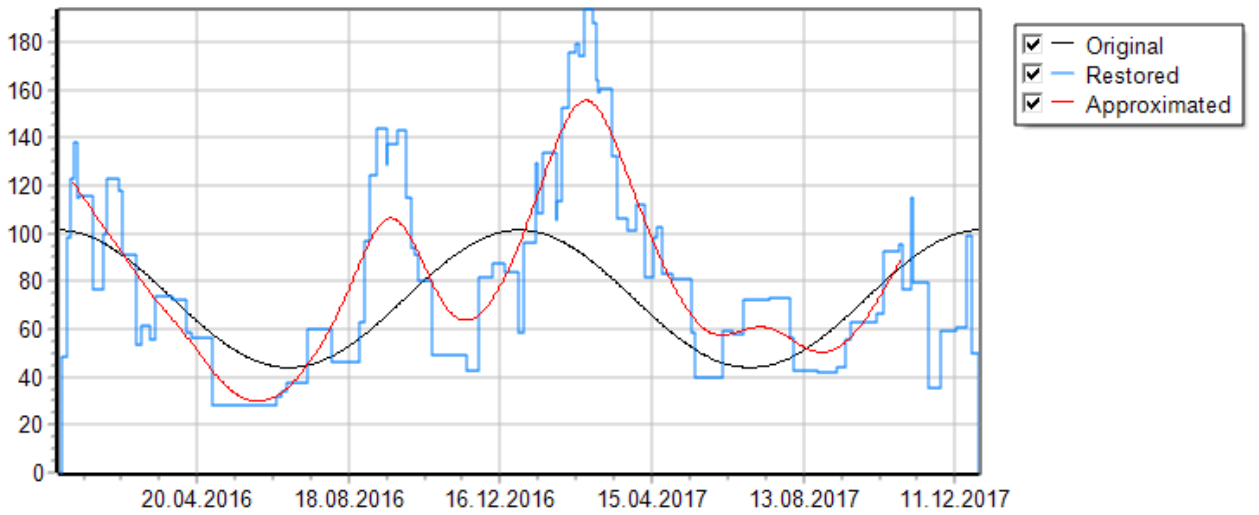
На рисунках 4.21 и 4.22 ступенчатой линией изображены восстановленные функции скорости расхода продукции у распространителей 1 и 2. Ступенчатая функция показывает среднюю скорость расхода продукции, восстановленную из данных. Гладкая черная линия показывает исходную (суммарную) скорость расхода продукции у закрепленных за этим распространителем покупателей (суммарный расход продукции двумя потребителями). Визуально можно заметить, что там, где восстановленная функция ниже исходной на одном рисунке, на другом рисунке она выше, и наоборот, там, где восстановленная ступенчатая функция выше, на другом рисунке она ниже. Такое происходит из-за того, что на некотором участке потребление некоторых клиентов переключилось с одного распространителя на другого.

Теперь, если взглянуть на рисунок 4.23, где ступенчатой функцией показана восстановленная скорость расхода вышестоящего распространителя, отмеченного номером 0, можно увидеть, что восстановленная функция достаточно хорошо соответствует исходной скорости расхода продукции (сплошные линии - суммарная скорость расхода четырех потребителей, красная – восстановленная, черная - оригинальная). Неточность вызвана в основном удаленностью от конечного потребителя (появляется 1 уровень посредников).



Источник: составлено автором.

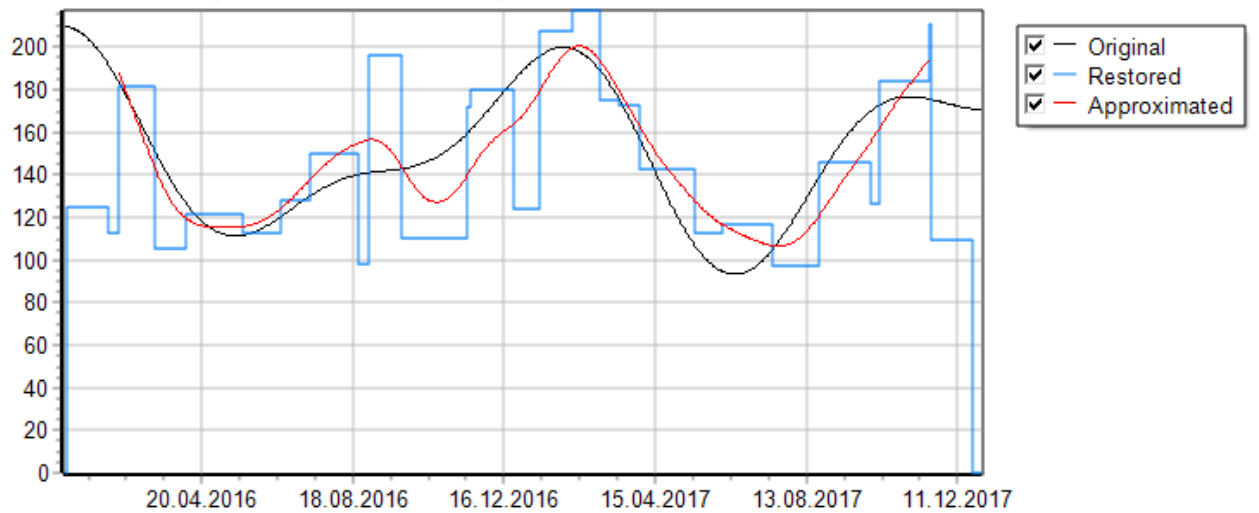
Рисунок 4.21 – Восстановленная скорость расхода продукции у распространителя 1



Источник: составлено автором.

Рисунок 4.22 – Восстановленная скорость расхода продукции у распространителя 2

Средняя относительная погрешность 7,97724666515%



Источник: составлено автором

Рисунок 4.23 – Восстановленная скорость расхода продукции у распространителя 0

В итоге можно сказать, что когда потребители переключаются между распространителями (продавцами) случайным образом, то в эти моменты восстановленная функция скорости расхода также переключается между этими распространителями. Восстановление функции из данных событий корректно показывает эти перемены, причем заметно, что когда у одного распространителя скорость потребления клиентами падает, то у другого она в это время возрастает, и наоборот. В то же время, пусть даже потребители переключаются между распространителями, если у вышестоящего распространителя (через один или несколько уровней посредников) его покупатели не переключаются на других вышестоящих распространителей, то есть все потребление происходит у нашего вышестоящего распространителя, то восстановленная функция суммарной скорости корректно восстанавливает исходную скорость потребления всех потребителей. Это дает возможность прогнозирования событий на уровне этого вышестоящего распространителя.

Определение точности восстановления исходной функции при потере части данных и сравнение с классическим подходом

Теперь перейдем к вопросу точности восстановления исходной функции при потере части данных (например, из-за конкуренции). Будем определять, как падает точность при увеличении вероятности потерять часть данных (вероятности переключения покупателя на конкурента). При этом с целью сравнения этого падения точности в нашем методе с другим методом будем осуществлять аналогичные действия в классическом подходе, когда из данных строится временной ряд, например, продаж по месяцам (где будем определять скорость расхода продукции как отношение объема продаж в месяце к количеству дней в этом месяце). То есть будем сравнивать наш метод (сглаживание интегральным сплайном производить не будем¹⁾), будем рассматривать только ступенчатую функцию средних скоростей) и

¹⁾ Иначе для справедливости надо также сглаживать ступенчатую функцию, получающуюся в классическом подходе.

классический, где временной ряд превратим в ступенчатую функцию скорости расхода продукции в каждый месяц.

Будем определять среднее значение для относительной ошибки, посчитанной для обоих методов, формула (4.36)

$$R = \frac{1}{n_t} \sum_{t=t_{\text{старт}}}^{t_{\text{конец}}} \frac{|\bar{f}(t) - f(t)|}{f(t)}. \quad (4.36)$$

где $\bar{f}(t)$ – средняя (не сглаженная) скорость расхода продукции;

$f(t)$ – исходная функция скорости расхода продукции;

$t_{\text{старт}}$ и $t_{\text{конец}}$ – начальный и конечный момент времени, для которого возможно сравнение исходной и восстановленной функции;

n_t – количество таких сравнений.

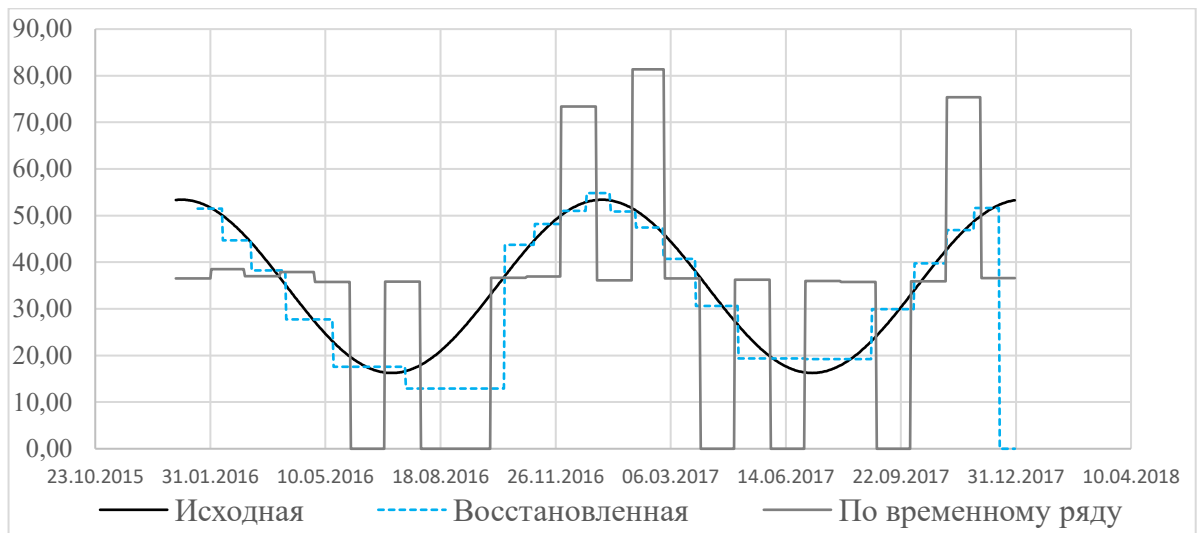
Дополнительно будем измерять среднюю квадратичную относительную ошибку для обоих методов, формула (4.37)

$$\sigma = \sqrt{\frac{1}{n_t} \sum_{t=t_{\text{старт}}}^{t_{\text{конец}}} \left(\frac{\bar{f}(t) - f(t)}{f(t)} \right)^2}. \quad (4.37)$$

Для получения результатов создадим модель, в которой в каждом отдельном прогоне случайным образом задается исходная функция расхода продукции $f(t)$ (средний уровень, амплитуда, частота и фаза задаются случайными числами из некоторого промежутка), случайным образом задается начальный запас, максимальный запас. Затем моделируем процесс расхода продукции (как в управлении запасами) за 2 года. Будем варьировать вероятность P того, что клиент обратится к другому поставщику и данные потеряются (информация о покупке пропадет). Разыгрывание вероятности также происходит случайным образом, то есть при 10% вероятности потери данных из выборки, состоящей из 20 наблюдений, может потеряться как 2, так и 1, 0 или же 3 и 4 наблюдения, а может и все 20 наблюдений (количество потерянных наблюдений соответствует биномиальному распределению). В то

же время есть вероятность того, что потери будут идти подряд (нет строгого чередования).

Пример моделирования показан на рисунке 4.24. В этом примере вероятность потери данных была 10%. Из 22 наблюдений было потеряно лишь одно, потерянное наблюдение приходилось на 9-тый месяц. Пунктирная ступенчатая линия показывает среднюю скорость, полученную с помощью емкостного метода. Сплошная ступенчатая линия показывает временной ряд (площадь ступеньки есть продажи за месяц), преобразованный в скорость расхода продукции. Гладкая сплошная линия показывает истинную скорость расхода продукции, которую требовалось восстановить из данных о продажах. В этом прогоне средняя относительная погрешность восстановленной функции составила 11,9% для емкостного метода и 58,4% для классического, средняя квадратичная относительная погрешность составила 18,7% и 70,1% соответственно. В таблице 4.10 приведены результаты 20 прогонов для вероятности потери данных 10%.



Источник: составлено автором.

Рисунок 4.24 – Пример моделирования для оценки погрешности

Таблица 4.10 – Относительная погрешность при 10% вероятности потери данных

Номер прогона	Средняя относительная ошибка в процентах		Разница	Средняя квадратичная относительная ошибка в процентах		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
1	15,1538	35,6751	20,5213	25,7686	40,5931	14,8245
2	19,2385	42,6782	23,4397	31,4277	55,4565	24,0287
3	10,4954	23,9023	13,4069	20,6358	28,7282	08,0924
4	20,1894	41,3341	21,1447	32,1758	53,1899	21,0141
5	25,6441	55,8185	30,1744	34,7822	87,6615	52,8793
6	16,5491	36,9385	20,3894	27,5394	49,4287	21,8893
7	11,0051	22,9302	11,9251	19,8246	30,479	10,6544
8	20,6452	70,037	49,3918	29,0427	101,1585	72,1158
9	10,1343	24,5582	14,424	19,2464	31,128	11,8816
10	14,3658	38,2592	23,8934	26,9225	48,9228	22,0003
11	13,921	48,0131	34,0921	22,801	63,1846	40,3836
12	6,8549	46,397	39,5421	9,1229	60,8571	51,7342
13	20,5882	28,2138	07,6256	32,345	38,6496	06,3046
14	6,7678	26,0652	19,2974	12,4631	33,1063	20,6432
15	13,3678	19,7823	06,4146	24,2151	27,0365	02,8215
16	10,7269	24,8431	14,1162	20,0973	34,3362	14,2389
17	15,9477	38,1086	22,1609	26,0582	49,6658	23,6076
18	15,7876	46,3486	30,561	25,6549	58,6784	33,0235
19	14,9275	58,5897	43,6622	23,3633	71,5303	48,167
20	19,536	69,6921	50,1562	25,8494	94,0599	68,2105

Источник: составлено автором.

Классический подход с формированием временного ряда полностью проигрывает в каждом эксперименте. Относительная погрешность емкостного метода (для получения средней скорости) в среднем на 24.8% меньше (на 28.4 для средней квадратичной относительной погрешности).

Проведем аналогичное моделирование с вероятностью потери данных 30%. Результаты представлены в таблице 4.11.

Таблица 4.11 – Относительная погрешность при 30% вероятности потери данных

Номер прогона	Средняя относительная ошибка в процентах		Разница	Средняя квадратичная относительная ошибка в процентах		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
1	40,661	49,478	8,817	45,467	60,349	14,882
2	38,775	39,536	0,761	48,271	49,835	1,564
3	29,129	59,977	30,848	36,811	72,629	35,818
4	26,551	35,727	9,176	39,613	46,353	6,740
5	27,419	32,571	5,152	38,257	46,821	8,564
6	24,549	35,926	11,378	36,321	43,810	7,489
7	26,360	30,149	3,790	37,601	37,165	-0,435
8	33,776	39,033	5,257	44,006	52,257	8,252
9	33,975	37,804	3,829	44,284	44,110	-0,174
10	38,246	54,343	16,096	45,526	64,534	19,008
11	27,181	50,703	23,522	37,889	61,949	24,059
12	16,978	54,400	37,422	26,519	66,288	39,769
13	30,821	41,901	11,081	40,159	54,406	14,248
14	47,786	60,916	13,130	58,851	72,022	13,171
15	37,405	49,773	12,368	47,676	60,656	12,980
16	32,871	46,228	13,357	42,424	57,839	15,415
17	36,227	37,889	1,662	45,674	45,668	-0,006
18	38,379	72,399	34,019	46,106	86,715	40,609
19	13,814	67,873	54,059	19,269	84,973	65,704
20	31,854	36,204	4,351	39,892	42,649	2,757

Источник: составлено автором.

При вероятности потери данных 30% ошибка восстановления исходной функции уже большая, тем не менее в каждом эксперименте средняя относительная ошибка емкостного метода меньше (в среднем на 15%), средняя квадратичная относительная ошибка в основном меньше (в среднем на 16,5%), правда кое-где ошибка сравнялась.

Наконец проведем моделирование, в котором вероятность потери данных будет 50%. Результаты представлены в таблице 4.12.

Таблица 4.12 – Относительная погрешность при 50% вероятности потери данных

Номер прогона	Средняя относительная ошибка в процентах		Разница а	Средняя квадратичная относительная ошибка в процентах		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
1	47,746	56,215	8,468	27,933	44,038	16,104
2	35,898	71,731	35,833	44,054	81,189	37,135
3	53,394	74,495	21,102	60,303	82,048	21,745
4	35,768	68,061	32,293	43,550	77,317	33,767
5	52,566	84,793	32,227	62,241	93,685	31,444
6	51,347	63,975	12,628	57,779	75,714	17,935
7	44,194	61,661	17,467	51,086	72,101	21,015
8	61,588	62,980	1,392	65,735	74,196	8,461
9	51,375	70,069	18,695	56,987	81,231	24,244
10	46,771	73,436	26,664	56,859	79,695	22,836
11	47,270	66,255	18,985	55,215	72,996	17,781
12	39,825	45,107	5,282	46,423	57,808	11,384
13	45,203	66,544	21,341	54,167	73,718	19,551
14	59,472	77,924	18,452	64,622	86,228	21,606
15	54,556	79,175	24,620	61,530	86,874	25,344
16	48,729	70,110	21,381	55,809	77,713	21,904
17	52,266	62,035	9,769	62,942	74,023	11,081
18	53,202	62,206	9,004	61,232	72,977	11,746
19	34,200	78,717	44,517	41,817	91,231	49,414
20	57,245	84,039	26,794	64,176	94,453	30,277

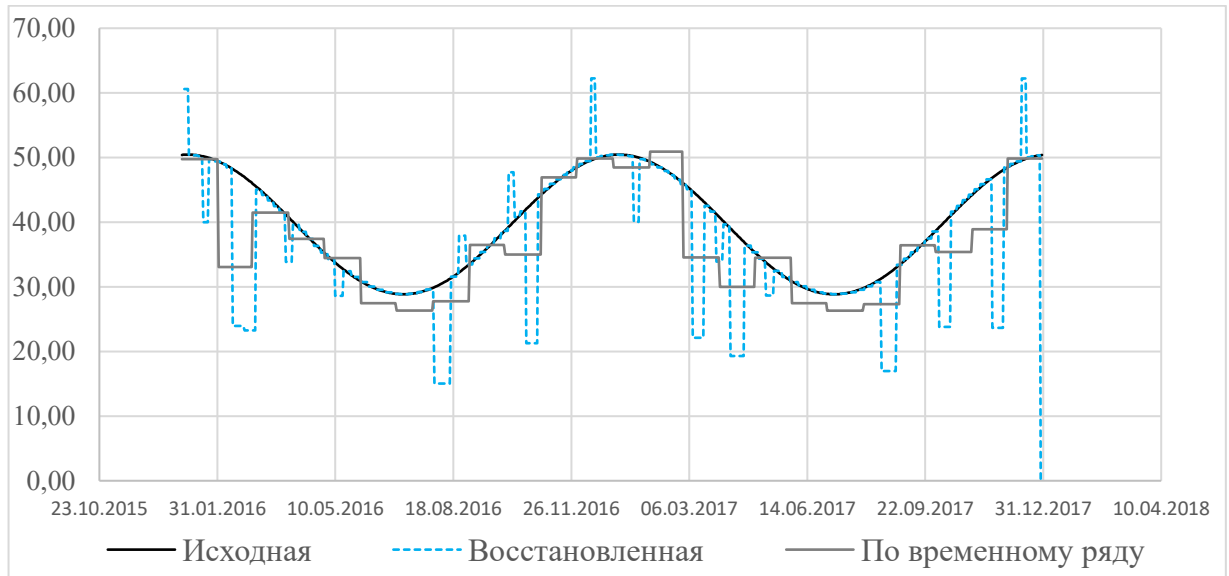
Источник: составлено автором.

Погрешность стала совсем огромной, однако опять в каждом эксперименте средняя относительная ошибка емкостного метода оказалась меньше (в среднем на 20,3%), средняя квадратичная относительная ошибка также меньше (в среднем на 22,7%), правда кое-где ошибка сравнялась.

Во всех трех сериях экспериментов наглядно видно, что емкостный метод обладает меньшей погрешностью при потерях данных, следовательно, при анализе редких событий (продаж) по возможности надо всегда использовать его, а не строить временные ряды с нулями продаж за выбранный период.

Однако, что, если мы имеем дело не с редкими продажами, а с частыми, например, когда каждый покупатель делает много покупок за месяц. Как в этом случае будет изменяться точность при потере части данных. Временной ряд не будет содержать нулей, каждое пропущенное наблюдение лишь

уменьшает величину расхода за месяц. Пример моделирования для вероятности потери данных 10% изображен на рисунке 4.25. Результаты моделирования потери данных с вероятностью 10% при частых продажах показаны в таблице 4.13.



Источник: составлено автором.

Рисунок 4.25 - Пример моделирования для оценки погрешности при частых продажах

Таблица 4.13 – Относительная погрешность при 10% вероятности потери данных при частых продажах

Номер прогона	Средняя относительная ошибка в процентах		Разница	Средняя квадратичная относительная ошибка в процентах		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
1	11,252	11,424	0,172	21,540	16,222	-5,318
2	14,576	16,303	1,727	25,636	20,293	-5,343
3	9,628	11,772	2,144	18,070	16,242	-1,828
4	13,941	22,024	8,083	20,583	29,383	8,801
5	8,484	9,130	0,646	18,776	11,951	-6,825
6	11,813	12,338	0,526	22,886	16,999	-5,887
7	16,154	19,831	3,676	24,162	26,511	2,350
8	11,650	14,152	2,502	22,791	18,475	-4,316
9	13,174	14,780	1,605	25,119	21,557	-3,562
10	12,732	13,664	0,932	23,989	17,287	-6,702

Источник: составлено автором.

В каждом эксперименте средняя относительная погрешность емкостного метода по-прежнему меньше. Однако средняя квадратичная относительная погрешность в большинстве случаев наоборот больше. Это связано с тем, что из-за потери данных при частых продажах емкостный метод

показывает занижение на узком интервале времени, которое достаточно значительно по величине. Возведение в квадрат этой разницы вносит большой вклад в оценку погрешности даже при условии, что эти спады кратковременны. Стоит заметить, что любое сглаживание способно решить эту проблему.

В результате этих экспериментов мы определили, как влияет потеря данных на точность восстановления исходной функции. Также мы определили, что емкостный метод анализа редких событий в случае потери части данных теряет меньше в точности восстановления исходной зависимости, чем классический метод, базирующийся на построении временных рядов. В каждом проведенном эксперименте средняя относительная погрешность емкостного метода была меньше, чем у классического. В случае анализа частых событий, когда временной ряд не содержит нулевых значений, по-прежнему лучше использовать емкостный метод, так как он способен дать много информации о протекающих в источниках процессах.

Для объяснения полученных результатов было проведено обоснование с помощью математического анализа, приложение Е. Однако для этого обоснования при сравнении двух методов пришлось принять несколько допущений, которые делают соответствующее обоснование недостаточно строгим. Тем не менее результат достаточно интересен. Получается, что даже при частых событиях, если количество событий за каждый интервал времени не больше 2, то восстановление исходной функции $f(t)$ при потере части данных будет более точным у емкостного метода. Только при частоте событий за интервал 3 и более точность классического метода сравнивается и начинает превосходить точность емкостного метода. Однако, редкие события характерны тем, что частота этих событий значительно меньше 1. Поэтому в условиях редких событий, даже при потере части данных, надо все равно отказываться от стандартных временных рядов и переходить на емкостный метод анализа редких событий.

4.5 Изучение влияния неопределенности внутри самого процесса, протекающего в источнике, на точность метода

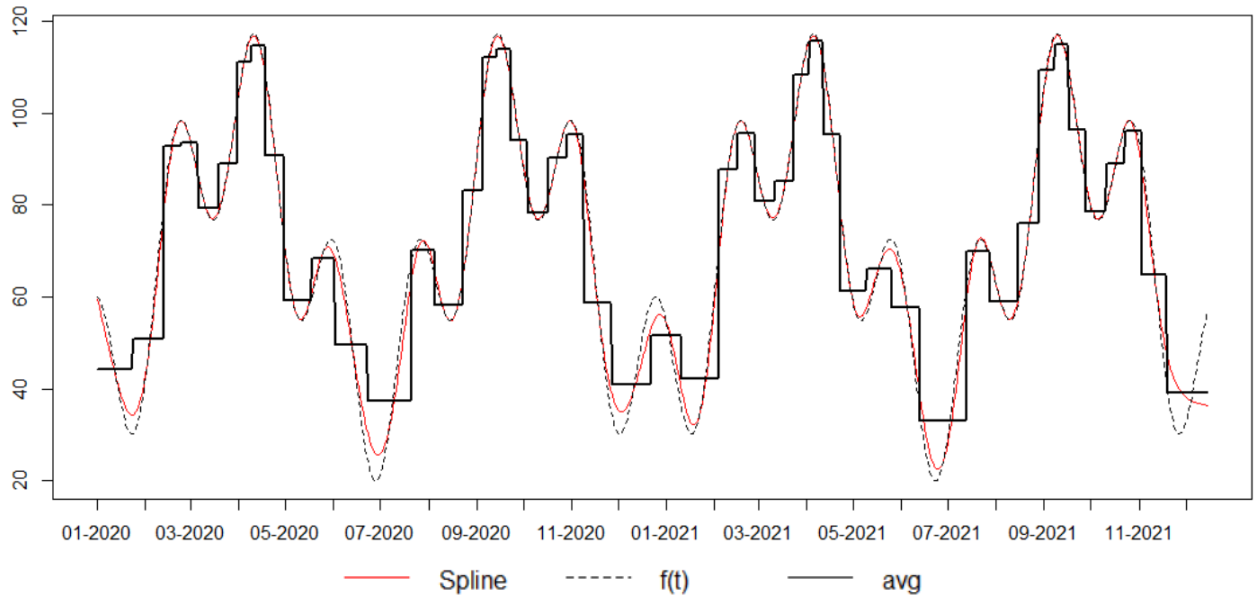
Производить исследование будем с помощью имитационного моделирования (как и ранее), в результате получаем моменты времени t_i и объемы пополнения запаса y_i . Помимо этого, также получим точные значения интегралов от исходной функции за время между моментами пополнения запаса $I_i = \int_{t_i}^{t_{i+1}} f(t) dt$. Эти интегралы не будут совпадать с объемами пополнения запаса, так как часть от этого объема будет компенсировать использование страховых запасов (см. параграф 4.1 и 4.3). По данным моделирования будем восстанавливать исходную функцию $\hat{f}(t)$ из наблюдаемых объемов пополнения запаса с помощью интегрального сплайна, описанного в параграфе 3 (параметр сглаживания α определим один раз и для разных экспериментов будем использовать одно и то же значение). Для определения точности восстановления исходной функции $f(t)$ будем находить среднюю относительную погрешность восстановления в процентах

$$R = \frac{1}{T} \sum_{t=1}^T \frac{|\hat{f}(t) - f(t)|}{f(t)} \times 100\%. \quad (4.38)$$

Вместе с тем будем определять $\sigma_{\text{факт}}^{\text{набл}}$ фактическое СКО относительного отклонения наблюдаемых объемов пополнения запасов от рассчитанных аналитически интегралов $\frac{y_i - I_i}{I_i}$ и попробуем показать связь между относительной погрешностью восстановления функции и этим разбросом в наблюдениях.

Обратим внимание, что даже если мы восстанавливаем функцию по истинным интегралам I_i от функции, то все равно мы имеем ненулевую погрешность, рисунок 4.26. Такое происходит из-за того, что на концах интервала недостаточно информации, сплайн не знает куда стремиться. Восстановление функции по истинным интегралам, показанное на рисунке 4.26, происходит со средней относительной погрешностью 3.7%. Сплайн

строится таким образом, чтобы площадь под графиком приближалась к площади под каждой ступенькой, где ступенчатая линия показывает среднее значение функции на интервале $I_i/(t_{i+1} - t_i)$. Как можно заметить, наибольшая погрешность наблюдается на правом конце интервала. Этот результат показателен, даже при идеальных наблюдениях восстановление не идеально, это надо учесть при дальнейшем анализе.



Ступенчатая линия avg - средние значения $I_i/(t_{i+1} - t_i)$, гладкая сплошная линия Spline - восстановленная функция, гладкая пунктирная линия f(t) – исходная функция.

Источник: составлено автором.

Рисунок 4.26 – Восстановление функции по истинным значениям интегралов I_i

Теперь перейдем к результатам экспериментов. В начале определим точность не от неопределенности внутри процесса, а от погрешности в самих наблюдениях.

Эксперимент 1, погрешность в самих наблюдениях

Моделируя процесс потребления продукции, получаем даты t_i и объемы y_i пополнения запасов, также определяем истинные значения интегралов I_i исходной функции между этими датами. Но восстанавливать значение функции будем по истинным интегралам, взятым с некоторой дополнительной относительной погрешностью $y_i = I_i \cdot (1 + \varepsilon_i)$. То есть наблюдениями будут значения, которые случайным образом будут чуть меньше или чуть больше истинных значений интегралов на случайное

количество процентов. Величина ε_i будет генерироваться по нормальному закону распределения с нулевым математическим ожиданием и с заданным средним квадратичным отклонением в 1%, 5%, 10%, 20% и 30%. Так как случайные величины с нормальным законом распределения могут отклоняться на несколько сигм в большую или меньшую сторону, то отдельные наблюдения могут достаточно серьезно отклоняться от истинных значений. Для каждого установленного значения СКО наблюдений проводим 20 независимых прогонов (реализаций случайных чисел), таблица 4.14.

Таблица 4.14 – Среднее относительное отклонение восстановленной функции от фактического СКО наблюдений

В процентах

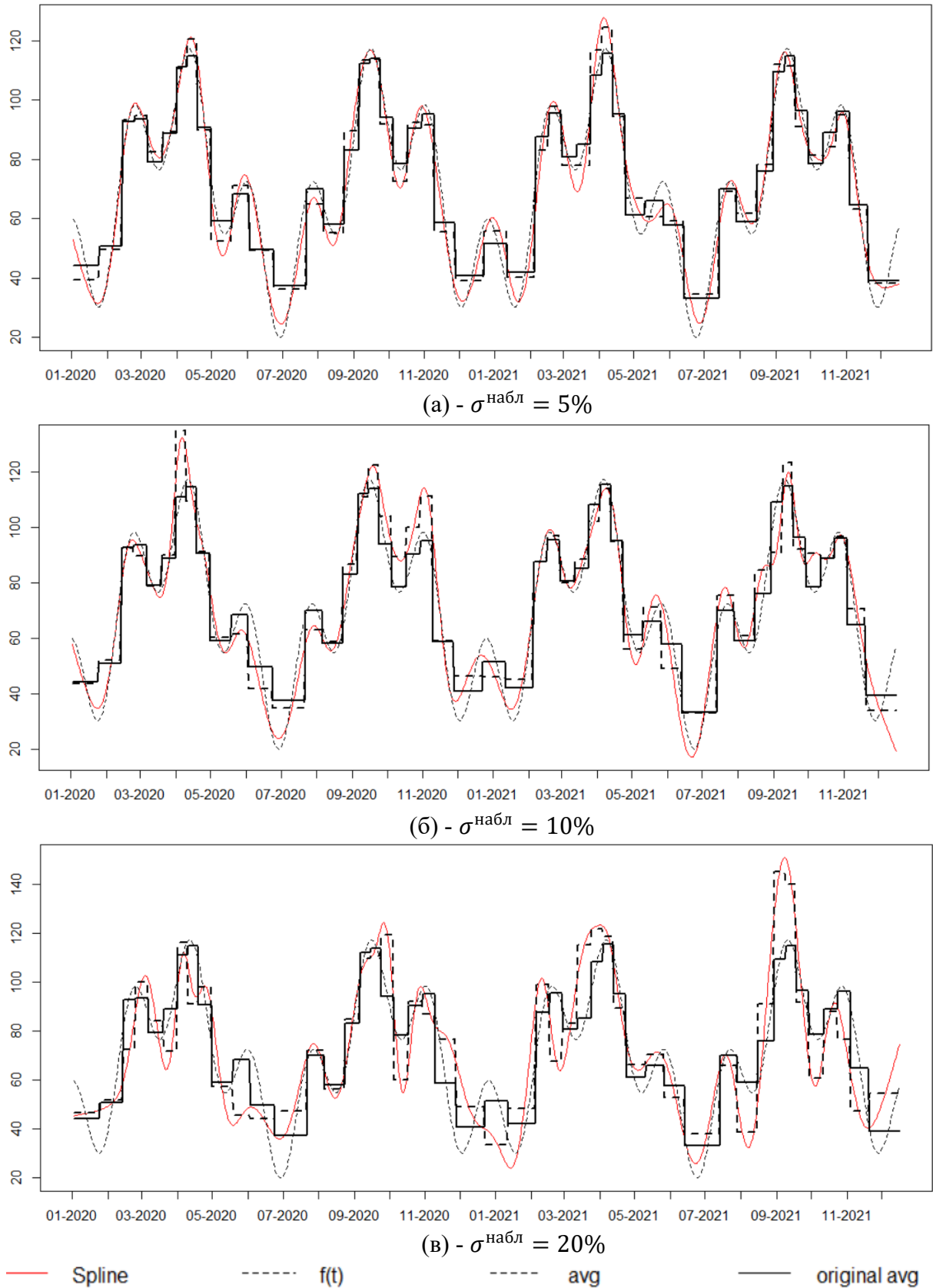
Прогоны	$\sigma^{\text{набл}} = 1\%$		$\sigma^{\text{набл}} = 5\%$		$\sigma^{\text{набл}} = 10\%$		$\sigma^{\text{набл}} = 20\%$		$\sigma^{\text{набл}} = 30\%$	
	R	$\sigma_{\text{факт}}^{\text{набл}}$	R	$\sigma_{\text{факт}}^{\text{набл}}$	R	$\sigma_{\text{факт}}^{\text{набл}}$	R	$\sigma_{\text{факт}}^{\text{набл}}$	R	$\sigma_{\text{факт}}^{\text{набл}}$
1	3,959	0,958	6,788	5,051	9,910	9,437	19,964	18,946	26,672	27,420
2	4,282	0,936	7,292	4,972	11,840	9,922	22,931	21,148	36,855	37,383
3	3,968	0,978	6,116	4,590	11,003	11,306	17,438	15,139	29,734	30,575
4	3,979	0,877	7,987	5,017	9,052	8,725	20,876	20,793	34,634	30,828
5	3,870	1,074	6,057	5,060	13,540	11,600	19,484	19,038	28,170	27,926
6	4,239	1,061	7,017	4,652	9,875	9,633	18,517	17,500	33,216	32,552
7	3,802	1,081	7,856	5,904	9,880	9,433	20,394	20,829	28,924	28,519
8	4,056	1,073	6,654	4,874	11,503	10,762	22,970	23,316	29,102	25,802
9	3,877	0,861	6,525	4,414	15,092	11,953	17,169	18,459	24,398	25,092
10	3,783	1,088	6,205	4,262	11,434	10,155	16,815	18,471	27,825	28,338
11	4,202	1,076	6,790	5,015	12,730	9,918	19,897	20,233	36,034	32,030
12	3,974	0,837	6,257	4,362	10,701	9,209	21,802	21,902	31,457	30,687
13	4,019	1,045	7,721	5,047	12,022	11,077	18,806	18,909	33,146	28,528
14	4,072	1,090	6,143	5,274	9,479	9,319	25,098	22,724	32,491	28,472
15	3,932	0,823	6,505	5,545	11,839	11,022	22,588	22,041	31,982	33,311
16	4,138	1,193	6,540	5,259	9,072	10,639	19,975	22,463	28,013	31,434
17	4,138	1,161	6,961	5,902	11,646	10,207	22,453	23,913	25,215	26,945
18	4,108	1,039	6,381	4,785	11,355	9,814	21,334	22,429	35,981	33,053
19	4,039	0,999	5,687	4,479	11,729	9,837	21,003	18,220	27,879	31,475
20	4,251	0,951	5,778	4,288	10,776	10,804	18,083	17,357	28,471	26,623

Источник: составлено автором.

При погрешности наблюдений в 1% средняя относительная погрешность восстановления исходной функции колеблется от 3,8 до 4,2%, но не забываем, что даже при нулевой погрешности наблюдений погрешность восстановления функции была 3,669414%, поэтому увеличение погрешности

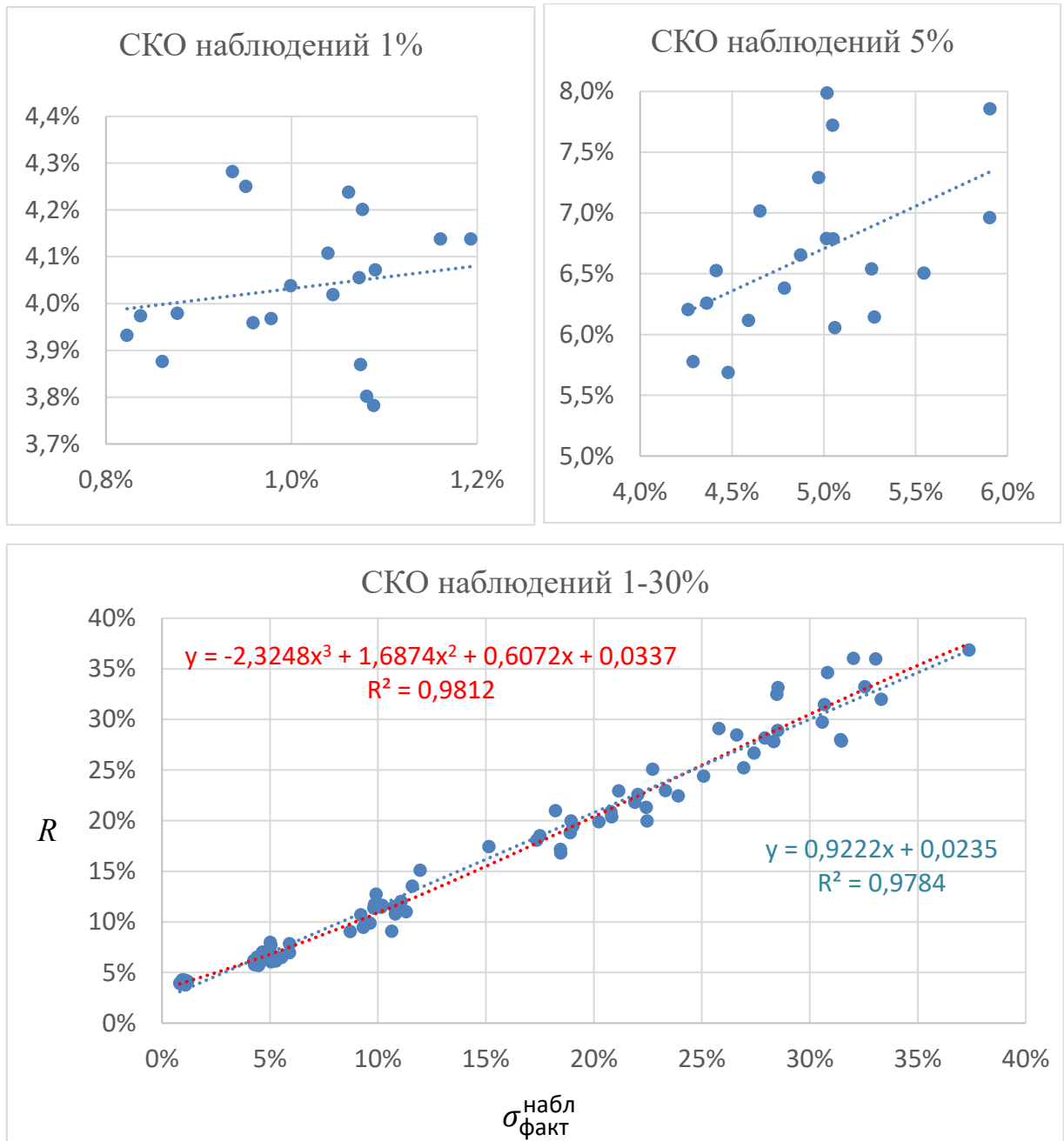
наблюдений с 0 до 1% незначительное. Также надо обратить внимание, что фактическое СКО наблюдений $\sigma_{\text{факт}}^{\text{набл}}$ не полностью объясняет погрешность восстановления исходной функции R (если несколько наблюдений подряд отклоняются в одну и ту же сторону, на этом участке восстановленная функция становится сильно смещена, оценка СКО отклонений не содержит информации о том, как эти отклонения шли). На рисунке 4.27 показаны результаты некоторых прогонов восстановления функции по интегралам при заданных относительных погрешностях наблюдений $\sigma^{\text{набл}}$. На диаграмме рассеивания на рисунок 4.28 заметно, что при сравнительно одних и тех же значениях $\sigma_{\text{факт}}^{\text{набл}}$ точность восстановления исходной функции R может значительно отличаться. Визуально заметен практически линейный рост R от $\sigma_{\text{факт}}^{\text{набл}}$, но полиномиальная модель подходит чуть лучше.

При большом разбросе наблюдений от истинных значений, например, когда СКО относительной погрешности составляет 30%, появляются ложные дополнительные пики функции, точность восстановления функции неудовлетворительна, рисунок 4.29а. В этом случае можно увеличить сглаживающий коэффициент α и убрать высокочастотные пики, рисунок 4.29б. Однако, конечно же, при этом теряются и высокочастотные составляющие исходной функции.



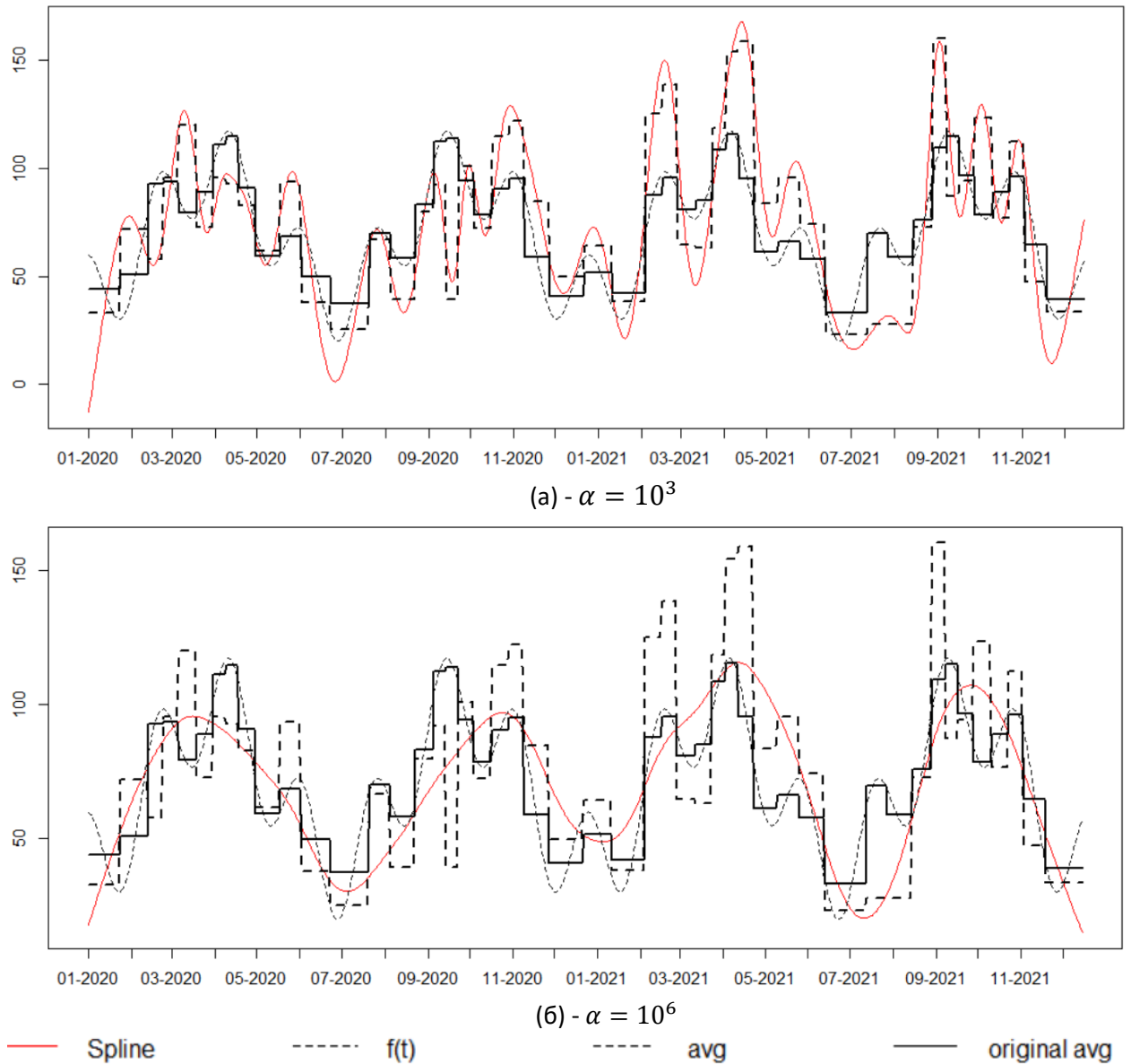
Источник: составлено автором.

Рисунок 4.27 – Восстановление функции по наблюдениям, образованным от истинных значений интегралов с дополнительной относительной погрешностью $y_i = I_i \cdot (1 + \varepsilon_i)$:
а – 5%; б – 10%; в – 20%



Источник: составлено автором.

Рисунок 4.28 – Диаграмма рассеивания погрешности R восстановления исходной функции от фактического СКО наблюдений $\sigma_{\text{факт}}^{\text{набл}}$



Источник: составлено автором.

Рисунок 4.29 – Восстановление функции по интегралам при сильно зашумленных данных $\sigma^{\text{набл}} = 30\%$, удаление шумов увеличением коэффициента сглаживания:

а) $\alpha = 10^3$; б) $\alpha = 10^6$

Эксперимент 2, погрешность в процессе образования событий во время проверки запасов

В данном эксперименте мы проверим как влияет неопределенность внутри самого процесса образования событий на точность восстановления исходной функции. Напоминаю, что процессом образования событий является процесс потребления, когда из запасов ежедневно вычитается некоторый объем, заданный определенной функцией. Происходит проверка уровня запасов, если он опустился ниже критического уровня, то происходит

мгновенное пополнение запаса. Предположим, что неопределенность кроется в процессе проверки уровня запасов, когда потребитель на глаз определяет достаточно ли еще запасов или нет. Предположим, что во время такой проверки существует разброс, соответствующий нормальному распределению с нулевым математическим ожиданием и с определенным средним квадратичным отклонением, заданным относительно критического уровня в процентах. То есть пусть проверка необходимости пополнения запаса происходит по следующему условию (здесь величина ε_i имеет нормальное распределение с нулевым математическим ожиданием и с заданным средним квадратичным отклонением $\sigma^{\text{кр.ур.}}$, которое будем изменять во время эксперимента)

$$\text{Stock} \leq \text{CriticalStorage} \cdot (1 + \varepsilon_i).$$

Стоит обратить внимание на то, что проверка уровня запасов во время моделирования происходит каждый день. Более вероятно, что условие будет выполняться чуть раньше, так как достаточно, чтобы лишь в один из дней отклонение от критического уровня произошло в меньшую сторону на достаточную величину. Для того, чтобы пополнение запаса произошло позже, необходимо, чтобы ни в один день до этого (спустя предыдущее пополнение) не было таких отклонений, вероятность чего намного меньше. Однако пусть объем пополнения запаса после такой проверки будет происходить как обычно без учета этого случайного отклонения, т.е. запас пополняется до максимума на момент этой проверки. В этом случае объемы пополнения запасов будут меньше, так как израсходовалось меньше запасов. Получается, что пополнение запасов будет происходить в среднем чуть раньше, но и объемы пополнения в среднем будут чуть меньше. Тем не менее по-прежнему будет возникать погрешность в наблюдениях $\sigma_{\text{факт}}^{\text{набл}}$, так как наблюдаемый объем пополнения запаса будет не совпадать с реально расходуемым количеством продукции. Причем величина этого несовпадения будет пропорциональна разбросу при сравнении критического уровня $\sigma^{\text{кр.ур.}}$ (наблюдаемый объем

пополнения запаса хоть и связан с объемом расходуемой продукцией, но он определяет на сколько хватит продукции до следующего события, а не до текущего: тем самым наблюдение и интеграл от функции с точки зрения расходуемой продукции относятся к разным интервалам времени). После того, как все нюансы проведения эксперимента обговорены, приведем его результаты, таблица 4.15.

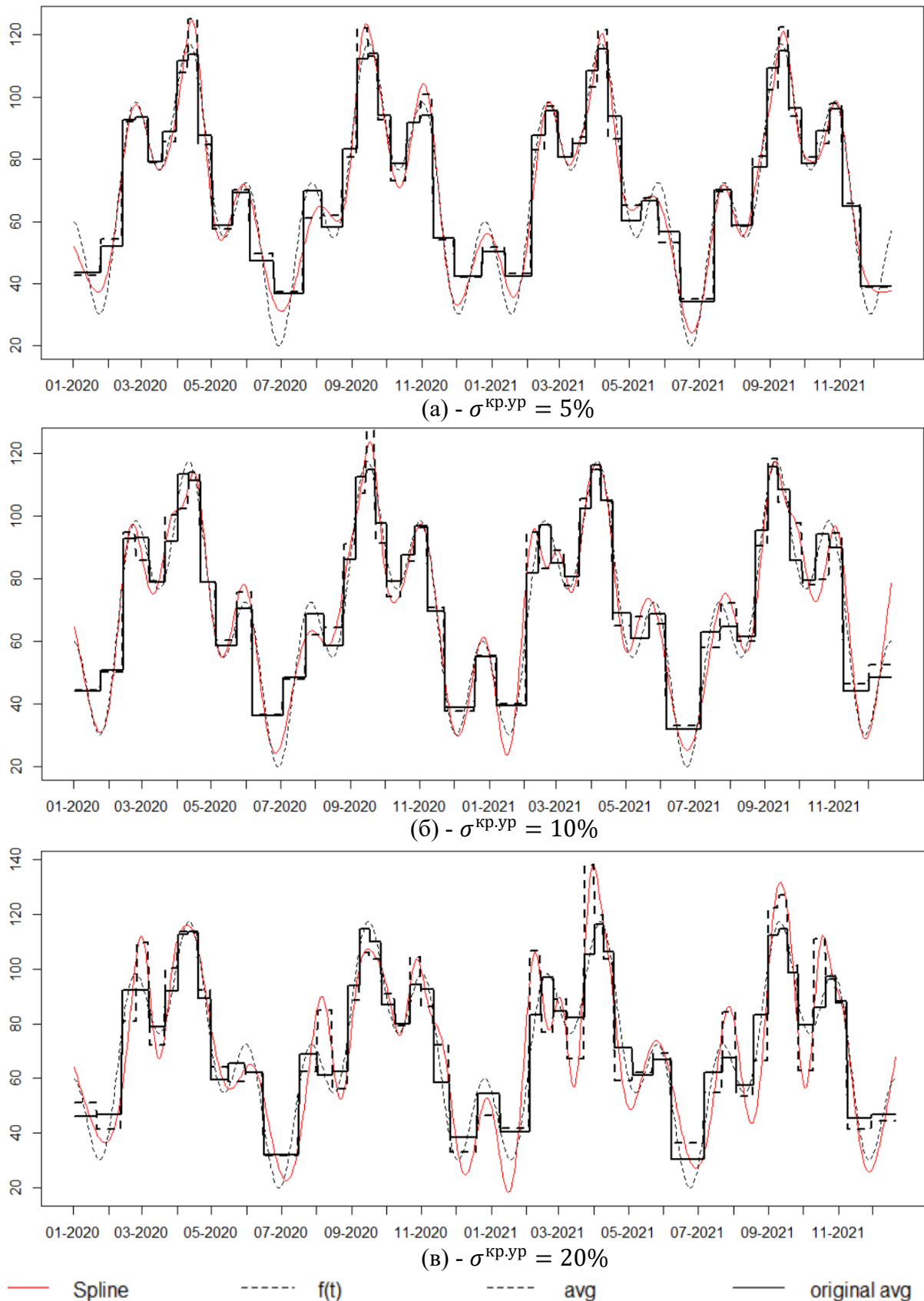
Таблица 4.15 – Среднее относительное отклонение R восстановленной функции и фактическое СКО наблюдений от разброса при проверке уровня запасов, в процентах

Прогнозы	$\sigma^{\text{кр.ур}} = 1\%$		$\sigma^{\text{кр.ур}} = 5\%$		$\sigma^{\text{кр.ур}} = 10\%$		$\sigma^{\text{кр.ур}} = 20\%$		$\sigma^{\text{кр.ур}} = 30\%$	
	R	$\sigma_{\text{факт}}^{\text{набл}}$	R	$\sigma_{\text{факт}}^{\text{набл}}$	R	$\sigma_{\text{факт}}^{\text{набл}}$	R	$\sigma_{\text{факт}}^{\text{набл}}$	R	$\sigma_{\text{факт}}^{\text{набл}}$
1	5,119	3,328	7,013	4,656	7,147	6,724	14,697	14,174	18,513	17,680
2	4,760	3,150	6,816	4,629	6,799	6,819	13,656	12,829	12,919	13,582
3	4,597	2,891	5,710	4,823	7,067	6,010	15,285	13,025	14,493	18,531
4	5,651	3,333	5,458	3,909	7,082	7,200	12,395	10,708	10,700	14,327
5	4,951	3,177	5,697	4,477	7,769	7,526	12,782	12,548	15,001	15,634
6	4,344	3,070	4,016	3,674	8,069	6,693	11,344	11,790	22,514	23,348
7	4,822	2,996	5,964	4,884	8,738	6,683	10,317	9,388	16,715	16,533
8	5,145	3,041	7,191	4,819	10,765	8,350	10,708	9,872	14,745	15,777
9	5,021	3,070	7,413	5,384	7,519	6,931	11,564	12,579	17,315	18,146
10	4,162	3,282	6,216	3,844	7,388	5,768	12,022	10,240	15,773	15,853
11	5,021	3,070	5,723	4,758	9,086	8,520	11,418	10,640	14,162	17,326
12	4,042	2,890	5,336	4,885	8,574	7,041	10,760	11,293	21,328	20,465
13	5,001	3,282	5,392	4,182	7,157	6,474	11,010	11,285	13,156	18,410
14	5,005	3,142	5,659	4,677	7,433	6,857	12,183	12,236	18,570	18,755
15	4,659	3,544	6,027	5,647	7,344	6,209	14,498	15,300	14,670	15,075
16	4,329	2,950	7,044	5,382	9,920	7,101	11,117	13,105	18,975	21,918
17	5,431	3,623	5,163	4,355	9,011	6,987	12,381	11,508	21,095	21,712
18	5,200	3,378	6,145	4,381	10,019	8,974	10,794	10,135	23,465	23,128
19	5,316	3,702	7,537	4,939	8,426	7,030	15,109	14,028	19,108	20,920
20	5,475	3,345	5,808	4,123	7,980	8,346	12,120	12,623	26,281	29,147

Источник: составлено автором.

Из результатов проведения экспериментов видно, что увеличение разброса погрешности при проверке уровня запасов приводит к тому, что увеличивается среднее квадратичное отклонение самих наблюдений, при этом оно растет медленнее. То, что рост погрешности происходит медленнее, объяснить можно тем, что критический уровень запасов был меньше, чем средний объем пополнения запаса (который зависит от максимального

уровня), поэтому отклонение от критического уровня вносит меньше погрешности, чем отклонения самих объемов пополнения запаса. Восстановление функции из наблюдений даже при разбросе в 20% при проверке уровня запасов происходит достаточно хорошо, рисунок 4.30. Тем не менее можно констатировать, что неопределенность при проверке уровня запасов всего лишь влияет на погрешность самих наблюдений. Именно погрешность наблюдений влияет на погрешность восстановления исходной функции. Из диаграммы рассеивания на рисунке 4.31 видна очень похожая линейная зависимость, однако коэффициент наклона меньше (был 0,9222, стал 0,821, причем доверительные интервалы параметра наклона не пересекаются даже для доверительной вероятности 99%). Утверждать, что погрешность растет медленнее, чем в предыдущем эксперименте, не получится, потому что СКО наблюдений не говорит о том, как отклонения шли во время самого прогона. Так два подряд отклонения в одну сторону сильнее заставят функцию отклониться, нежели если отклонения будут чередоваться то в одну сторону, то другую, а грубая оценка среднего квадратичного отклонения не способна этого отобразить. Весьма вероятно, что при других прогонах, которые будут обладать той же оценкой СКО наблюдений, погрешность восстановления функции будет значительно отличаться.

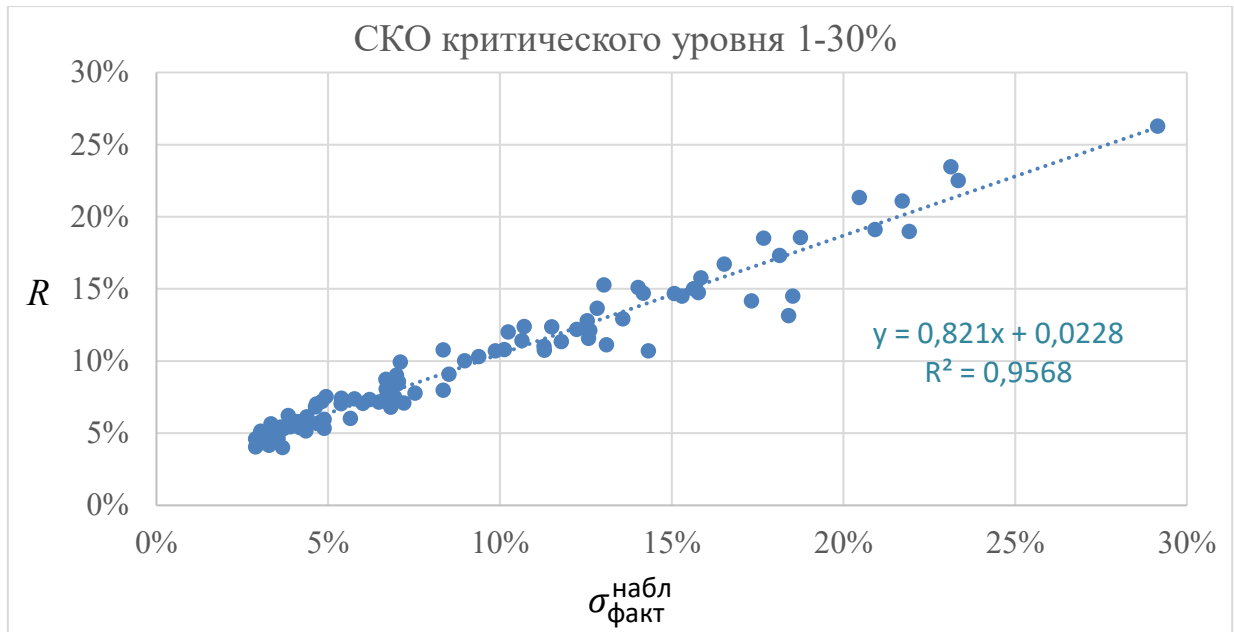


Источник: составлено автором.

Рисунок 4.30 – Восстановление функции по наблюдениям, образованным в процессе потребления с погрешностью при сравнении уровня запасов

$$\text{Stock} \leq \text{CriticalStorage} \cdot (1 + \varepsilon_i):$$

а – 5%, б – 10%; в – 20%



Источник: составлено автором.

Рисунок 4.31 – Диаграмма рассеивания погрешности R восстановления исходной функции от фактического СКО наблюдений $\sigma_{\text{факт}}^{\text{набл}}$ при разбросе во время проверки уровня запасов

Эксперимент 3, погрешность в дате события

В предыдущем эксперименте разброс при проверке критического уровня приводил к сдвигу даты события как правило в одну сторону (событие возникало раньше). В этом эксперименте мы проверим как влияет разброс в дате возникновения события на точность восстановления исходной функции. Под разбросом даты возникновения события можно представить такую картину, когда потребитель идет пополнять запасы чуть позже или чуть раньше положенного срока. Одной из причин такого может быть по-прежнему то, что потребитель не может точно оценить свои запасы. Другой причиной может быть то, что потребитель занят в определенный день и планирует пополнить запасы либо заранее, либо чуть позже, когда освободится.

Для проведения эксперимента внесем следующие изменения в модель. Вначале определяем дату t_i завершения запаса также, как и в первом эксперименте при условии, что уровень запасов опустился ниже критического уровня (без разброса). Затем от данной даты откладываем разброс, выраженный через относительное значение интервала времени от предыдущего события, после чего определяем новую дату события:

$$\hat{t}_i = t_i + \Delta t_i \cdot \varepsilon_i,$$

$$\Delta t_i = t_i - t_{i-1}.$$

Случайная величина ε_i по-прежнему будет соответствовать нормальному закону с нулевым математическим ожиданием и с заданным средним квадратичным отклонением $\sigma^{\Delta t_i}$. Однако из-за того, что интервалы времени между событиями Δt_i неодинаковые, то разброс $\Delta t_i \cdot \varepsilon_i$ на самом деле соответствует нестационарному Гауссовскому процессу. Так как моделирование происходит по дням, то пусть разброс происходит в целых днях, а не в дробных, то есть будем округлять разброс до ближайшего целого. Так как интервалы между событиями не такие большие (колеблются от 9 до 28 дней), отклонение в 1% с последующим округлением ничего не дает. Начинать проводить эксперименты сразу будем с $\sigma^{\Delta t_i} = 5\%$. При этом мы будем определять фактическое относительное среднее квадратичное отклонение даты события $\sigma_{\text{факт}}^{\Delta t_i}$ в процентах от длительности интервала, которое рассчитаем уже после всех округлений.

В определенный таким образом момент времени наступления события \hat{t}_i наш потребитель будет пополнять запасы, причем пополнять будет те запасы, которые он израсходует к этому моменту \hat{t}_i (то есть к сдвинутому), а не к первоначальному (несдвинутому) t_i . Таким образом, если потребитель чуть раньше решит пополнять запасы, то и объем пополнения запасов будет меньше, а если он решит позже пополнять запасы, то объем пополнения запасов будет больше. Расчет следующего события будет происходить именно с этого нового момента времени \hat{t}_i (сдвинутого), то есть при моделировании придется откатить модельное время немного назад, если разброс времени события был отрицательный, и продвинуть немного вперед, если разброс был положительным. Хоть объемы пополнения запаса будут меняться из-за сдвига, все равно будет наблюдаться разброс в самих наблюдениях по сравнению с истинными интегралами от исходной функции (истинные интегралы определяются также до сдвинутой даты \hat{t}_i). В каждом прогоне будем как и раньше определять фактическое среднее квадратичное отклонение самих

наблюдений $\sigma_{\text{факт}}^{\text{набл}}$ в процентах. Результаты моделирования представлены в таблице 4.16, а результаты некоторых прогонов изображены на рисунке 4.32.

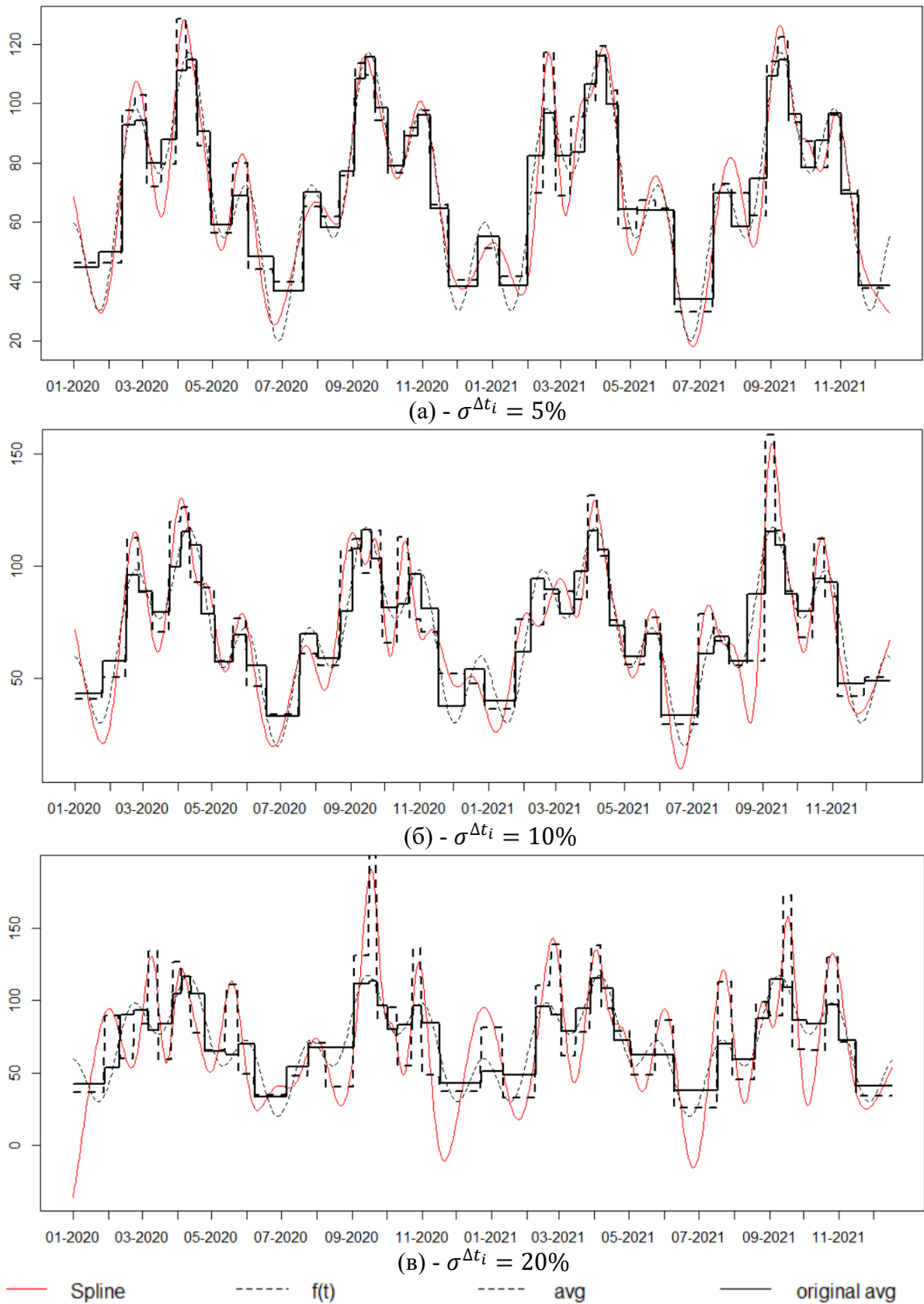
Из результатов эксперимента видно, что разброс относительно даты события вызывает достаточно значительный разброс в самих наблюдениях. Так при $\sigma_{\text{факт}}^{\Delta t_i} = 10,958\%$ разброс в наблюдениях составил 16,868%, что в свою очередь привело к погрешности восстановления функции $R = 16,640\%$. Объяснить это можно тем, что несовпадение наблюдаемых объемов пополнения запаса и реально расходуемых объемов зависит от изменения интервала с обоих концов, то есть обладает удвоенной дисперсией, что должно приводить к тому, что разброс наблюдений должен быть в среднем на 1.41 больше (из-за того, что исходная функция нестационарная, разброс может быть еще больше).

Еще хочется обратить внимание на то, что сам по себе разброс даты событий иногда приводит к тому, что событие возникает абсурдно рано. Так, например, при разбросе в $\sigma^{\Delta t_i} = 20\%$, если отклонение произойдет на 3 сигмы, произойдет сдвиг аж на 60% от интервала времени, то есть от предыдущего события прошло лишь 40% от положенного времени. Получается, что очередное событие возникло в 2.5 раза раньше. Такое значительное отклонение будет тяжело назвать случайным, его легко спутать с каким-то структурным фактором. Дальнейшее увеличение разброса даты событий, например, $\sigma^{\Delta t_i} = 30\%$ или 40% вообще может привести к тому, что очередное событие появится через нулевой период времени или должно появиться до появления текущего (в прошлом). Большие отклонения в датах события маловероятны, на практике такое трудно представить.

Таблица 4.16 – Среднее относительное отклонение R восстановленной функции, фактическое СКО наблюдений и фактическое СКО даты событий от разброса в дате образования события, в процентах

Прогнозы	$\sigma^{\Delta t_i} = 5\%$			$\sigma^{\Delta t_i} = 10\%$			$\sigma^{\Delta t_i} = 20\%$		
	R	$\sigma_{\text{набл}}^{\text{факт}}$	$\sigma_{\text{факт}}^{\Delta t_i}$	R	$\sigma_{\text{набл}}^{\text{факт}}$	$\sigma_{\text{факт}}^{\Delta t_i}$	R	$\sigma_{\text{набл}}^{\text{факт}}$	$\sigma_{\text{факт}}^{\Delta t_i}$
1	9,853	9,141	5,052	16,640	16,868	10,958	37,770	35,168	21,582
2	9,126	7,807	5,201	13,224	11,810	9,148	27,999	30,582	26,842
3	10,005	8,517	5,367	24,391	21,434	11,343	27,815	30,156	23,606
4	9,965	8,765	5,501	11,024	11,373	8,927	20,885	26,269	25,831
5	7,582	6,600	5,499	13,917	11,356	8,708	35,196	32,752	21,257
6	8,870	8,169	6,025	14,191	13,167	8,342	31,770	30,089	20,668
7	10,763	8,781	5,561	14,027	14,455	11,014	32,355	34,865	23,430
8	10,255	8,751	5,510	16,057	16,328	10,957	29,163	35,353	28,833
9	8,535	7,689	5,299	16,992	18,160	10,351	36,468	35,337	29,040
10	10,041	8,956	5,987	16,472	16,408	10,357	20,734	24,238	18,118
11	11,086	9,107	5,465	19,672	16,569	9,745	31,335	27,144	17,385
12	11,124	7,541	3,994	16,727	14,809	10,175	33,706	32,106	22,201
13	12,396	9,692	5,386	13,665	12,372	8,469	22,821	23,309	16,438
14	8,396	7,555	5,111	16,178	15,666	10,980	30,813	37,947	28,022
15	9,722	6,794	5,373	14,807	13,972	8,648	35,335	28,307	19,815
16	9,092	8,904	6,516	15,220	16,387	11,979	25,431	28,997	22,877
17	10,021	9,311	5,587	19,917	17,846	13,230	24,071	23,455	17,580
18	10,372	9,748	6,104	19,292	20,349	11,942	31,596	27,755	19,910
19	12,656	13,668	7,212	13,052	11,982	8,731	33,720	37,302	22,554
20	10,056	9,371	5,274	16,158	15,017	9,408	29,485	26,171	17,962

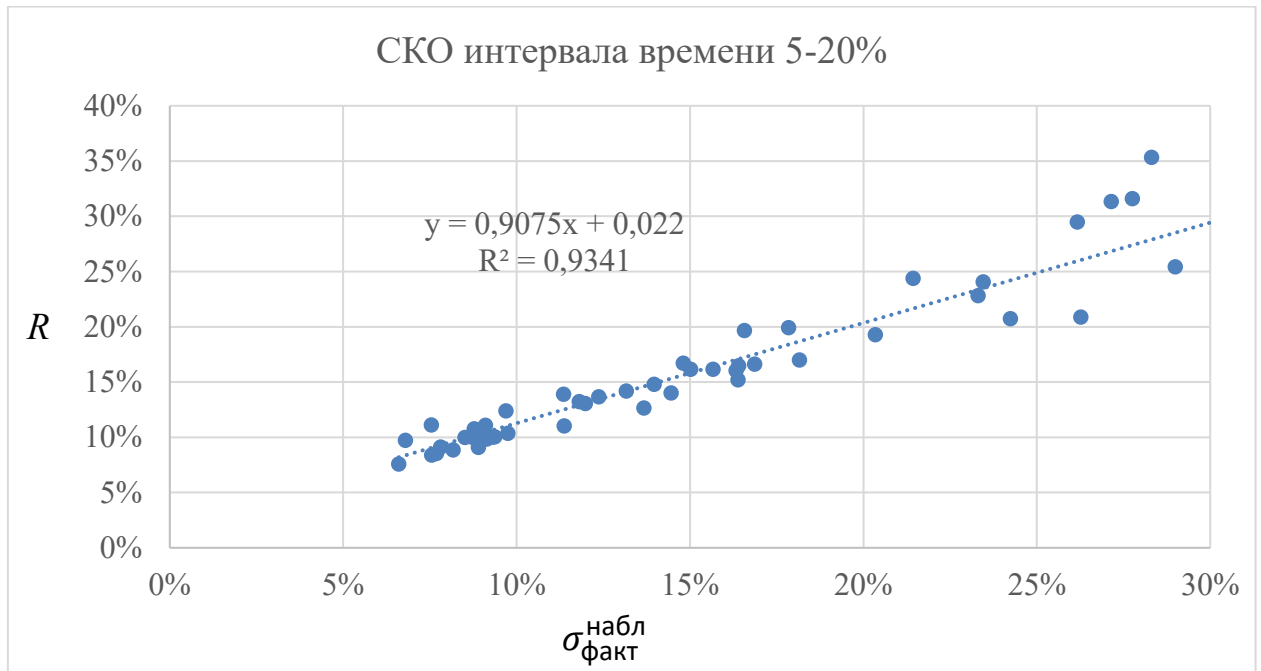
Источник: составлено автором.



Источник: составлено автором.

Рисунок 4.32 – Восстановление функции по наблюдениям, образованным в процессе потребления при разбросе в дате образования события с СКО интервала времени $\sigma^{\Delta t_i}$: а – 5%, б – 10%; в – 20%

На диаграмме рассеивания, изображенной на рисунке 4.33, видна линейная зависимость погрешности восстановления функции от фактической погрешности наблюдений, коэффициент наклона получился равен 0,907 (в первом эксперименте был 0,9222, во втором 0,821). Доверительный интервал для коэффициента наклона в этом эксперименте пересекается с доверительными интервалами в предыдущих двух экспериментах.



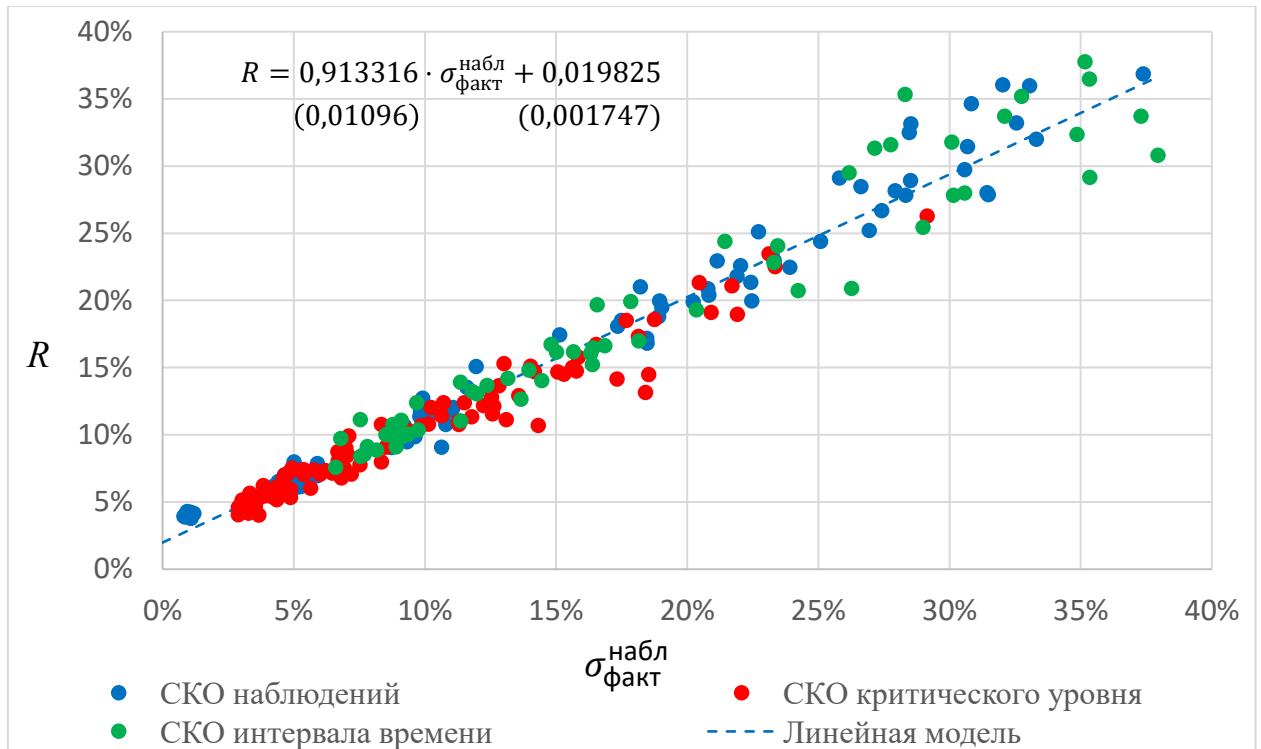
Источник: составлено автором.

Рисунок 4.33 – Диаграмма рассеивания погрешности R восстановления исходной функции от фактического СКО наблюдений $\sigma_{\text{факт}}^{\text{набл}}$ при разбросе даты события

На рисунке 4.34 представлена диаграмма рассеивания для данных всех трех экспериментов, объединенных вместе. Результаты каждого эксперимента обозначены своим цветом. Наблюдается линейная зависимость относительной погрешности восстановления исходной функции R от фактического среднеквадратичного отклонения наблюдений $\sigma_{\text{факт}}^{\text{набл}}$. Получившаяся линейная зависимость имеет уравнение, формула (4.39)

$$R = 0,913316 \cdot \sigma_{\text{факт}}^{\text{набл}} + 0,019825 \quad (4.39)$$

(0,010959) (0,001747)



Источник: составлено автором.

Рисунок 4.34 – Диаграмма рассеивания погрешности R восстановления исходной функции от фактического СКО наблюдений $\sigma_{\text{факт}}^{\text{набл}}$ для всех трех экспериментов

По результатам трех экспериментов мы видим, что погрешность внутри самого процесса образования событий приводит к появлению погрешности в наблюдениях. Дальнейший анализ влияния неопределенности внутри процессов образования событий можно свести к анализу погрешности самих наблюдений. Точность восстановления исходной функции по последовательности интегралов зависит линейно от этой погрешности в наблюдениях. При погрешности наблюдений $\sigma_{\text{факт}}^{\text{набл}}$ около 5% восстановление функции происходит очень хорошо, восстановленная функция повторяет все изгибы исходной функции, экстремумы приблизительно находятся на соответствующем месте, амплитуда отклонений не очень большая. При погрешности $\sigma_{\text{факт}}^{\text{набл}}$ около 10% по-прежнему восстановление происходит достаточно хорошо, однако амплитуда отклонения восстановленной функции становится заметно больше, также экстремумы могут быть немного смещены. При погрешности $\sigma_{\text{факт}}^{\text{набл}}$ в 20 и 30% начинают появляться ложные экстремумы там, где их не должно быть. То есть на восстановленную функцию как будто

накладывается дополнительный шум, из случайных отклонений. Данный шум можно попробовать сгладить с помощью использования более высокого значения коэффициента сглаживания в процессе восстановления функции по интегралам. Однако при удалении высокочастотных шумов также теряются высокочастотные колебания самой исходной функции. Этот подход можно использовать при сильно зашумленных данных, когда интерес представляет восстановление по интегралам хотя бы основных гармоник исходной функции.

4.6 Сравнение емкостного метода с другими методами

Прежде чем приступить к сравнению емкостного метода с другими методами, надо понять, что другие методы применяются вообще-то для разных задач, да и входные данные могут сильно отличаться. Для того чтобы произвести сравнение, придется либо емкостный метод адаптировать под задачи, которые решаются другими методами, либо другие методы адаптировать под задачи, решаемые емкостным методом. В то же время входные данные придется преобразовывать соответствующим образом, то есть из отдельных событий придется вновь строить временной ряд, причем временной ряд можно строить как по месяцам, так и по дням, что повлияет на результаты. Будем использовать за основу пример из параграфа 2.2.2, будем к входным данным из этого примера применять другие методы.

Сравнение с методом Кростона

Метод Кростона анализирует данные в виде временных рядов, разделяя данные на две выборки: выборку ненулевых значений и выборку длительности интервалом между ненулевыми значениями. С помощью метода Кростона можно определить среднее время до ненулевого спроса и среднее значение ненулевого спроса (усреднение с помощью обычного экспоненциального сглаживания), благодаря чему рассчитывается ожидаемый спрос за единицу времени. Также с помощью метода Кростона определяются страховые запасы.

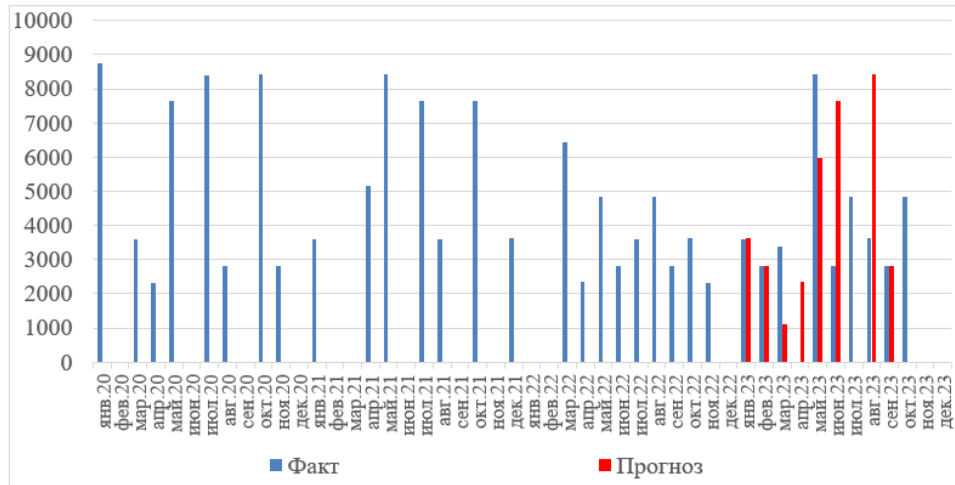
Обратим внимание на то, что метод Кростона, дает прогноз только на один месяц, а чтобы получить прогноз на несколько месяцев, мы вынуждены

добавлять в выборку данных фактические значения будущего периода (как в адаптивных методах). В емкостном методе мы сразу можем получить прогноз на несколько месяцев вперед, в примерах ниже мы не увеличиваем обучающую выборку (т.е. не как адаптивный, если бы в емкостном методе также делать прогноз только на одно событие вперед и каждый раз добавлять фактические данные в обучающую выборку, то результат будет еще лучше).

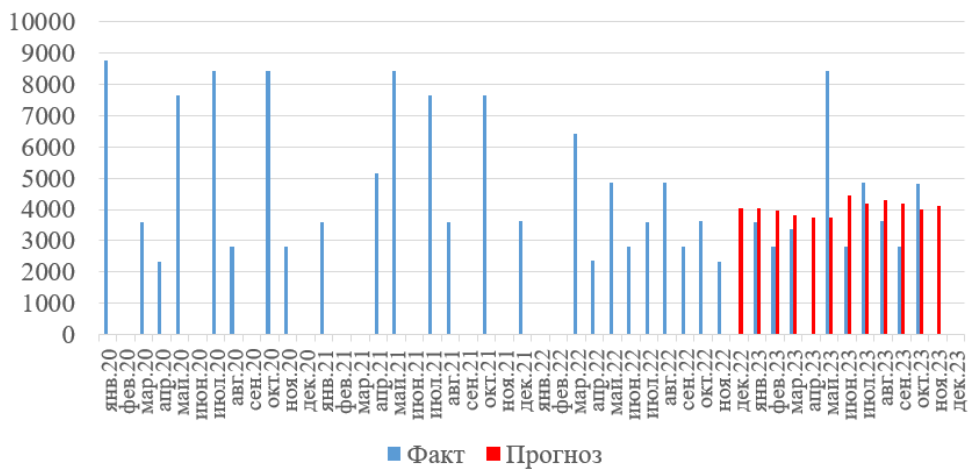
Прогноз спроса на месяц с помощью емкостного метода изображен на рисунке 4.35(а), а на рисунке 4.35 (б) с помощью метода Кростона, на вход которого подавался временной ряд, построенный с помощью агрегирования событий по месяцам, если же агрегировать события по дням, то прогноз получается, как на рисунке 4.35 (в).

Ожидаемый спрос за единицу времени, определенный с помощью емкостного метода, показан на рисунке 4.36 (а), а с помощью метода Кростона на рисунке 4.36(б) и 4.36(в).

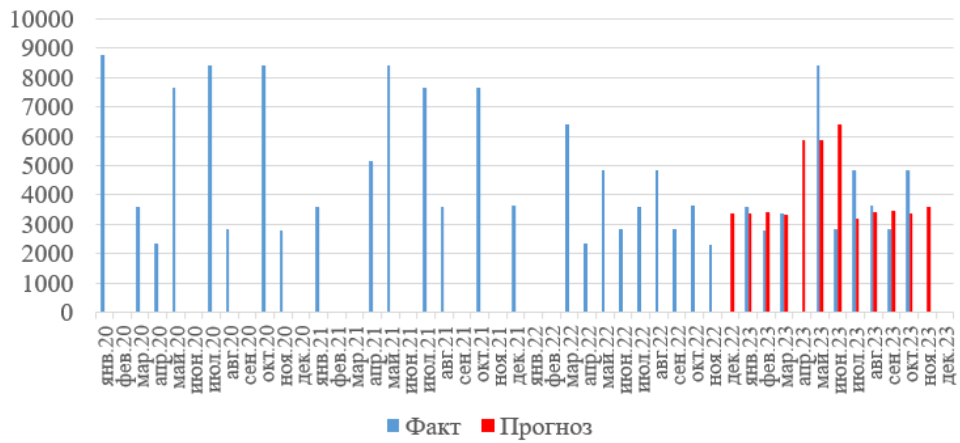
Для сравнения определяемых страховых запасов придется адаптировать емкостный метод, так как ранее мы им только получали прогноз будущих событий. Для этого в емкостном методе, зная (предполагая) погрешность восстановления функции, можем заранее запланировать пополнение запаса со сдвигом от конкретной прогнозной даты (например, если точность восстановления функции скорости потребления составляла 5%, то заложим сдвиг в 15% от интервала между событиями). Метод Кростона планирует практически одинаковый уровень запаса, даже когда он используется как адаптивный метод. Картина динамики запасов, изменяемых дискретными покупками, показана на рисунке 4.37.



а) Емкостный метод



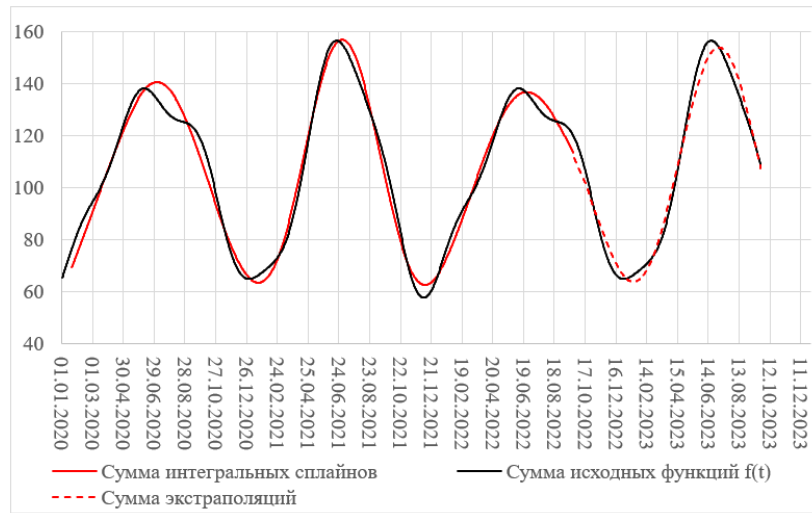
б) Метод Кростона (агрегирование по месяцам)



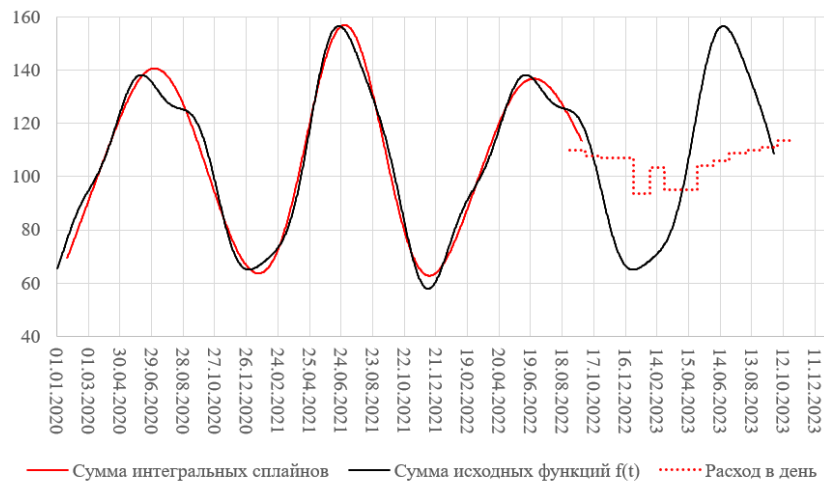
в) Метод Кростона (агрегирование по дням)

Источник: составлено автором.

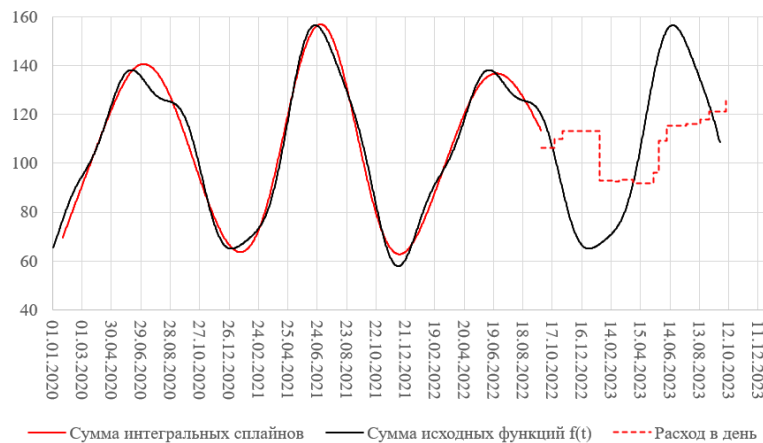
Рисунок 4.35 – Сравнение прогноза на месяц с методом Кростона



а) Емкостный метод



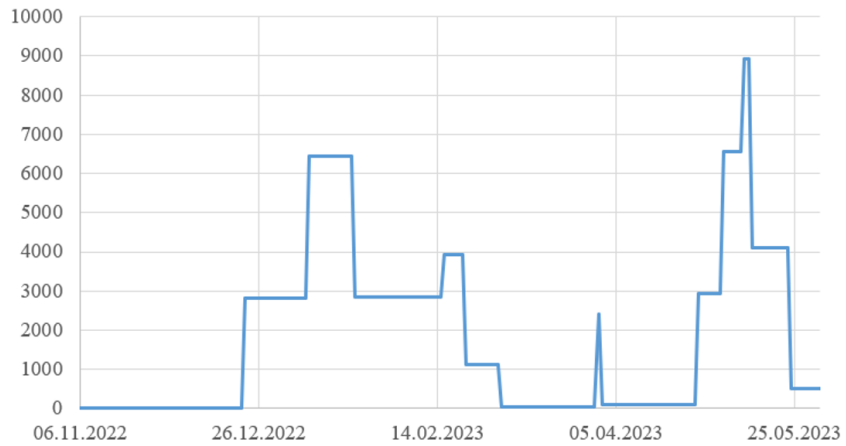
б) Метод Кростона (только пунктирная линия, агрегирование событий по месяцам)



в) Метод Кростона (только пунктирная линия, агрегирование событий по дням)

Источник: составлено автором.

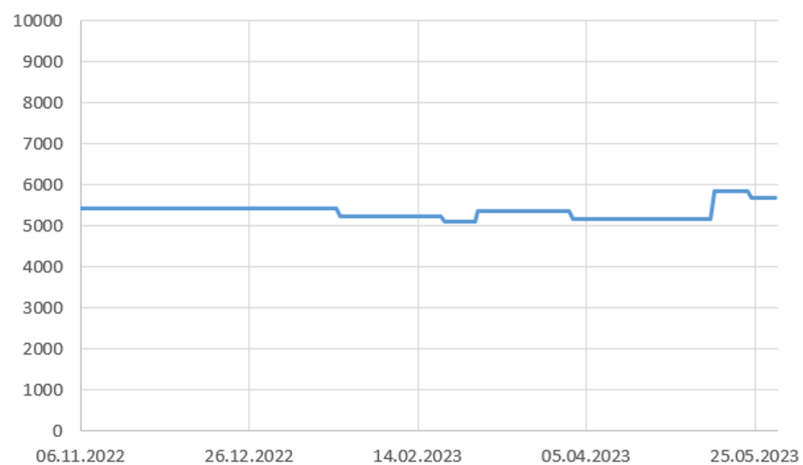
Рисунок 4.36 – Потребление в день, сравнение с методом Кростона



а) Емкостный метод



б) Метод Кростона (агрегирование событий по месяцам)



в) Метод Кростона (агрегирование событий по дням)

Источник: составлено автором.

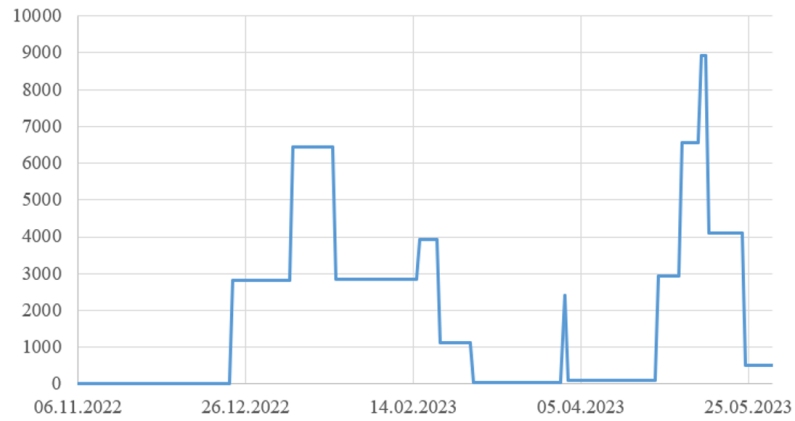
Рисунок 4.37 – Динамика запасов у нашего продавца, который подготавливает запасы в соответствии с указанным методом

Как видим емкостный метод дает более точный прогноз спроса на первые несколько месяцев, но прогноз ухудшается при увеличении временной

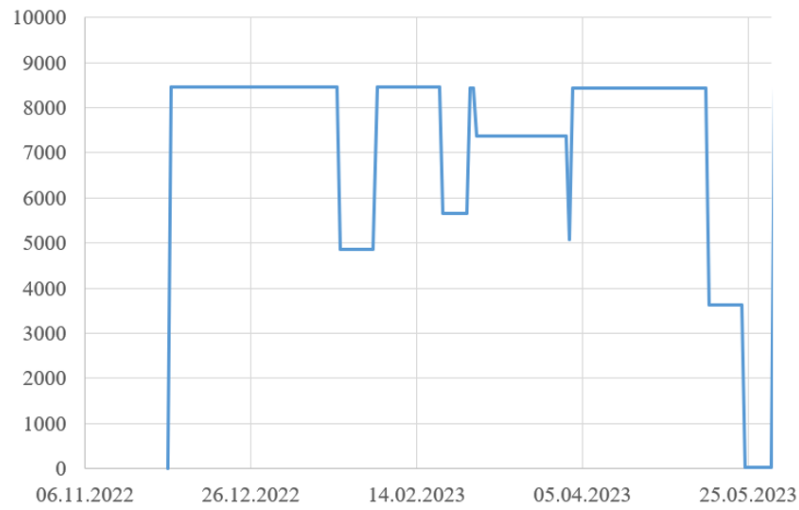
перспективы. Если же прогноз пересчитывать каждый раз, как это было сделано в методе Кростона, то его точность будет хорошей и в последних месяцах. Прогноз потребления за единицу времени всегда оказался точнее, а динамика запасов за полгода показала, что почти всегда хранится меньше запасов чем в методе Кростона.

Сравнение с методом Виллемейна

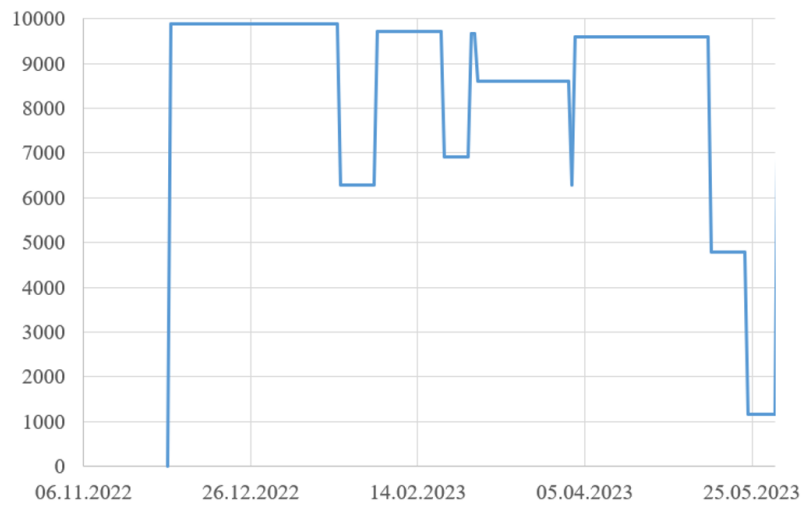
В методе Виллемейна определяется только объем заказа. Входными данными является опять временной ряд (агрегируем события либо по месяцам, либо по дням). Моделируя динамику изменения запасов, получаем следующую картину, рисунок 4.38. Уровень запасов при использовании емкостного метода всегда оказывается ниже, чем в методе Виллемейна, а агрегирование событий по дням в методе Виллемейна только ухудшает картину.



а) Емкостный метод



б) Метод Виллимейна (агрегирование по месяцам)



в) Метод Виллимейна (агрегирование по дням)

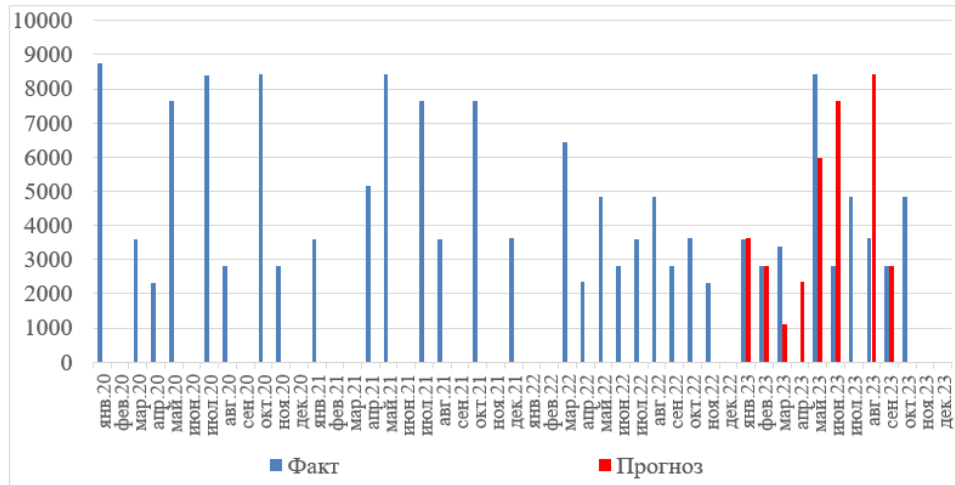
Источник: составлено автором.

Рисунок 4.38 – Динамика запасов у нашего продавца, который подготавливает запасы в соответствии с указанным методом

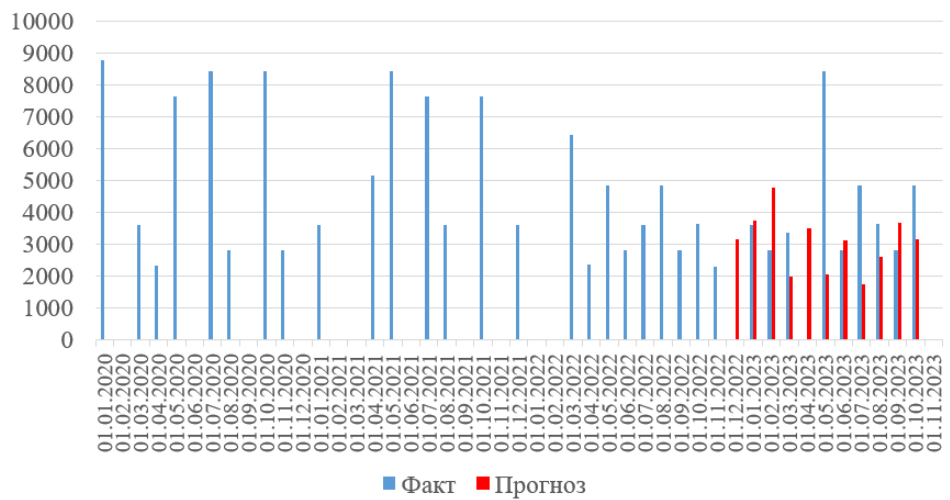
Сравнение с методом ближайших соседей

Иногда используются методы классификации, которые по определенным наблюдаемым признакам могут предсказать будущее значение. В логистике при анализе прерывистого спроса иногда используют метод ближайших соседей, когда вектором признаков выступает предшествующая последовательность значений фиксированной длины самого временного ряда. Тогда из всей имеющейся выборки (из всего временного ряда) находятся такие «соседи», которые наиболее близки к нашему вектору признаков, критерием выступает эвклидово расстояние. Следующее после каждого найденного соседа значение временного ряда выступает как некоторый прогноз, но если отбираются несколько соседей, то прогнозные значения усредняются в соответствии с весами, обратно пропорциональными расстоянию до вектора признаков. Причем временной ряд также можно формировать по дням, в этом случае будет огромное количество нулевых значений. Причем метод ближайших соседей используем как адаптивный (добавляя в обучающую выборку новые данные).

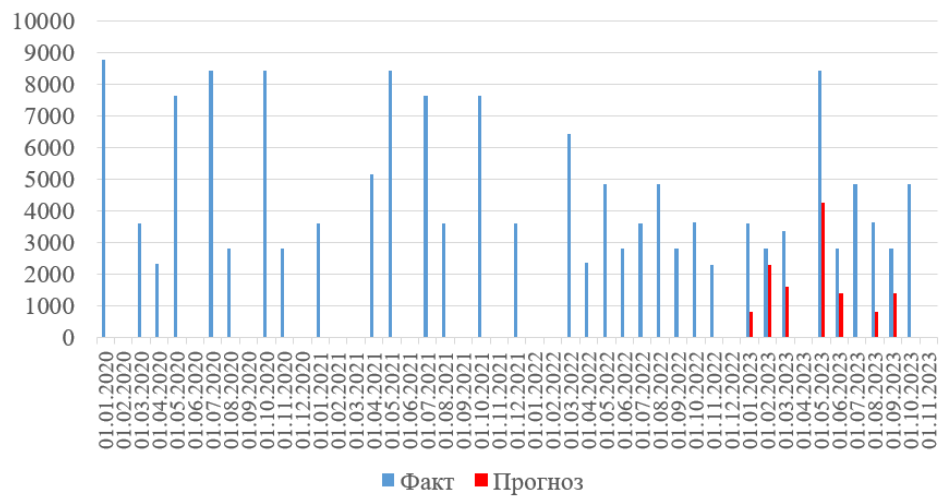
Сравнение прогноза спроса на месяц, полученного с помощью емкостного метода и с помощью метода ближайших соседей, показано на рисунке 4.39. Если находятся несколько соседей, то смешивание прогноза будет приводить к его занижению. Обращаем внимание на то, что в методе ближайших соседей надо выбрать длину вектора признаков (ширину скользящего окна), результаты ниже соответствуют самому лучшему выбору этой длины вектора, дающего наименьшую ошибку прогноза (из всех опытов отобран самый лучший результат). Уже в первый месяц, где спрос должен отсутствовать, метод ближайших соседей прогнозирует ненулевой спрос.



а) Емкостный метод



б) Метод ближайших соседей (агрегирование по месяцам)



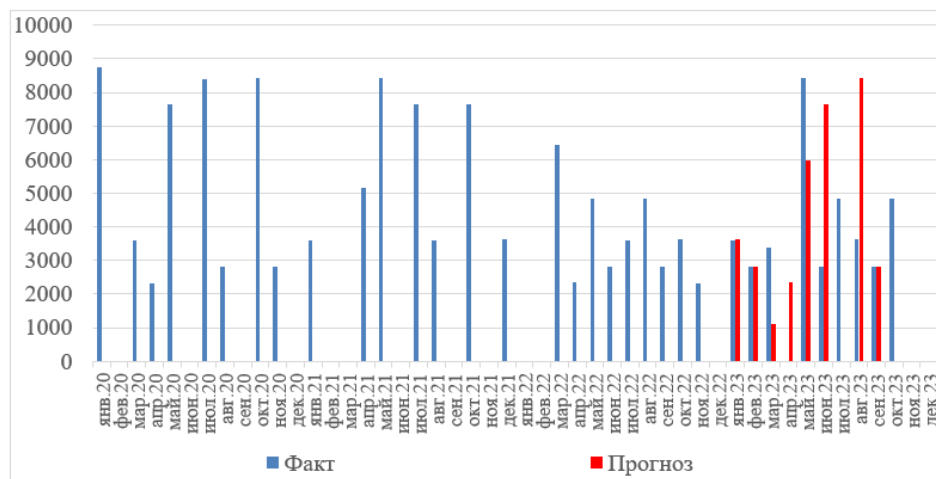
в) Метод ближайших соседей (агрегирование по дням)

Источник: составлено автором.

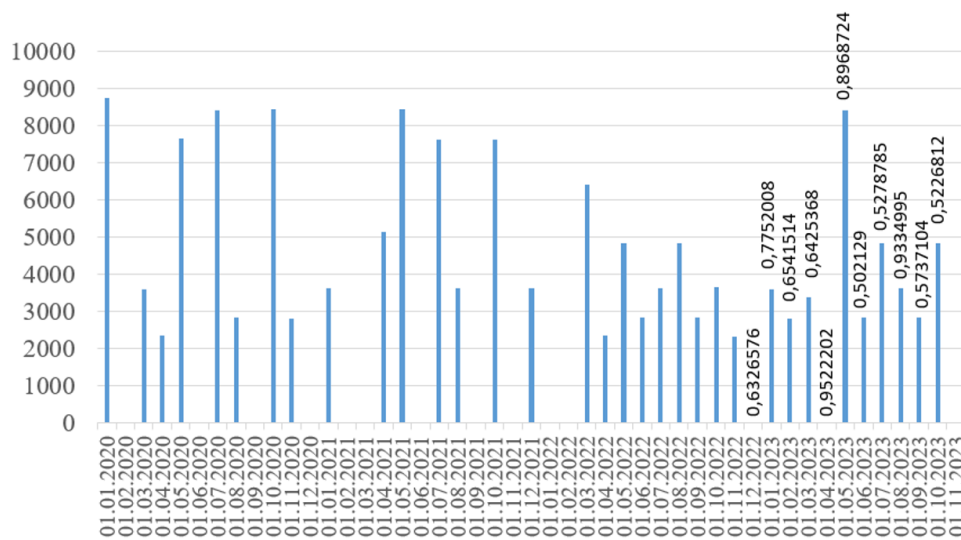
Рисунок 4.39 – Прогноз на месяц, сравнение с методом ближайших соседей

Сравнение с логистической регрессией

Используя методы классификации, такие как метод логистической регрессии, можно получить оценку вероятности возникновения будущего ненулевого спроса (но не его значение). На рисунке 4.40 показано сравнение емкостного метода с попыткой предсказать возникновение спроса в следующий месяц с помощью логистической регрессии (длина вектора признаков подобрана наилучшим способом).



а) Емкостный метод



б) Логистическая регрессия (агрегирование по месяцам)

Источник: составлено автором.

Рисунок 4.40 – Прогноз на месяц (вероятность ненулевого спроса в следующий месяц)

Как видно, в интервалы, где спроса не было, логистическая регрессия прогнозирует вероятность ненулевого спроса с большой вероятностью. Если же использовать агрегирование событий по дням, когда имеется огромное

количество нулевых значений, то логистическая регрессия не справляется (вероятность ненулевого события всюду прогнозируется как 0.5).

4.7 Методика проверки адекватности прогнозных моделей, построенных с помощью емкостного метода

Идея проверки адекватности для моделей прогнозирования будущих событий, построенных с помощью емкостного метода, будет, как и в эконометрических методах, заключаться в проверке того, что прогнозируемое событие попадет в соответствующий доверительный интервал, построенный для этого прогнозируемого события. В исследовании [84], посвященном вопросу выбора оптимального коэффициента сглаживания (вынесено за рамки данной работы, результаты которого кратко описаны в параграфе 3.2), с помощью имитационного моделирования (приблизительно на 1000 моделях) определялось, какой коэффициент сглаживания α минимизирует абсолютную $R_{abs}(\alpha)$ и квадратичную среднюю относительную погрешность $R_{sqr}(\alpha)$ восстановления исходной функции $f(t)$ в виде сплайна $g(t, \alpha)$, формулы (4.40) и (4.41)

$$R_{abs}(\alpha) = \frac{1}{t_n - t_1} \sum_{t=t_1}^{t_n} \left| \frac{f(t) - g(t, \alpha)}{f(t)} \right|, \quad (4.40)$$

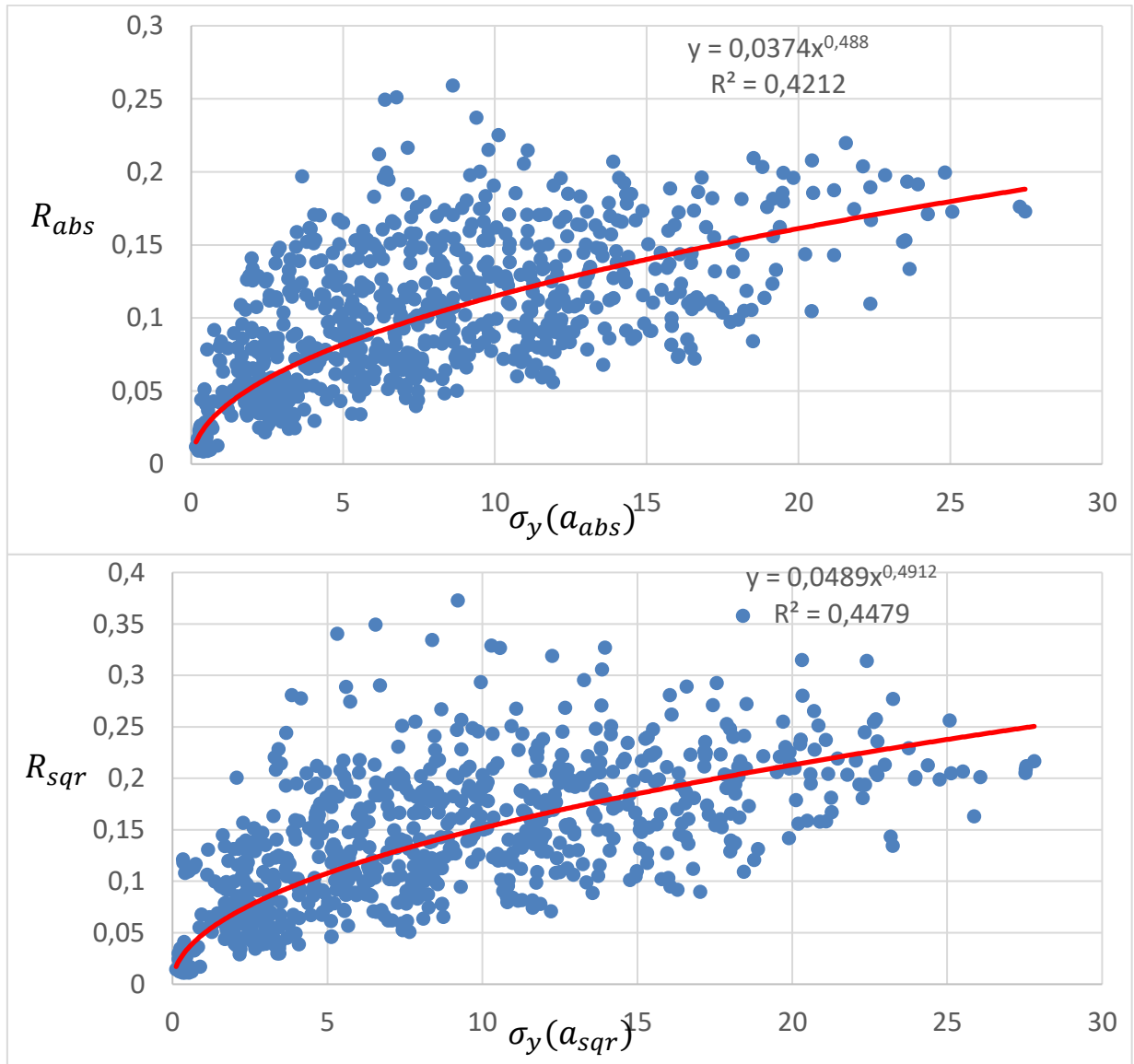
$$R_{sqr}(\alpha) = \sqrt{\frac{1}{t_n - t_1} \sum_{t=t_1}^{t_n} \left(\frac{f(t) - g(t, \alpha)}{f(t)} \right)^2}. \quad (4.41)$$

При этом были определены оптимальные коэффициенты сглаживания $a_{abs} = \arg \min_{\alpha} R_{abs}(\alpha)$ и $a_{sqr} = \arg \min_{\alpha} R_{sqr}(\alpha)$. Для этих значений также определялось насколько приглушаются (в процентах %) исходные наблюдения y_i , были определены $\sigma_y(a_{abs})$ и $\sigma_y(a_{sqr})$ по формуле (4.42)

$$\sigma_y(\alpha) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} \left(\frac{y_i - \bar{y}_i}{\bar{y}_i} \right)^2} \times 100\%. \quad (4.42)$$

То исследование предполагало сравнение значений коэффициентов сглаживания a_{CV} и a_L , рассчитанных с помощью кросс-валидации и метода L -кривой, с этими оптимальными значениями a_{abs} и a_{sqr} , а также

определения для принципа невязки Морозова насколько приглушаются наблюдения. Эти результаты исследования можно еще раз использовать, но уже чтобы построить модель зависимости погрешностей восстановления исходной функции при оптимальных коэффициентах сглаживания $R_{abs}(a_{abs})$ или $R_{sqr}(a_{sqr})$ от уровня приглушения наблюдений $\sigma_y(a_{abs})$ или $\sigma_y(a_{sqr})$. Модели получились следующими, рисунок 4.41.



Источник: составлено автором.

Рисунок 4.41 – Зависимость погрешности восстановления исходной функции от уровня приглушения наблюдений при оптимальных коэффициентах сглаживания

Оценить величину приглушения данных наблюдения σ_y можно по формуле (4.43)

$$\sigma_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} w_i \left(\frac{y_i - \sum_{j=1}^m C_{ij} g_j}{\sum_{j=1}^m C_{ij} g_j} \right)^2}, \quad (4.43)$$

где матрица C такая же как определено в параграфе 3.1 (здесь $\sum_{j=1}^m C_{ij} g_j$ показывает ожидаемое значение интеграла от функции \bar{y}_i);

g_j – рассчитанные значения сплайна в узлах сплайна;

w_i – вес наблюдений, используемый при восстановлении функции;

n – количество наблюдений;

y_i – сами наблюдения.

Если предположить, что во время восстановления функции используется оптимальный коэффициент сглаживания a_{abs} или a_{sqr} , то по уровню приглушения наблюдений σ_y можно приблизительно оценить погрешность восстановления функции \hat{R}_{abs} или \hat{R}_{sqr} .

После восстановления функции следует этап экстраполяции. Если предположить, что на этом этапе закономерности найдены верно и экстраполяция правильно осуществляет отображение этой закономерности на будущий период без какой-либо существенной дополнительной погрешности, то тогда с этой же погрешностью будет происходить опустошение/наполнение емкости. Прогноз (оценка момента времени) следующего события \hat{t}_{n+1} определяется как

$$\hat{t}_{n+1} = t_n + \widehat{\Delta t}, \quad (4.44)$$

где $\widehat{\Delta t}$ – оценка времени до следующего события. Эта оценка $\widehat{\Delta t}$ связана с интегралом от восстановленной функции, которая имеет погрешность \hat{R}_{abs} или \hat{R}_{sqr} . Тогда сам момент возникновения события \hat{t}_{n+1} будет также определяться с этой погрешностью \hat{R}_{abs} или \hat{R}_{sqr} (предполагая, что восстановленная функция скорости опустошения/наполнения емкости на интервале времени от текущего до следующего события сдвинута в одну и ту же сторону, имеем $D(\hat{t}_{n+1}) = \widehat{\Delta t} \cdot D(\widehat{\Delta t})$, где $D(\widehat{\Delta t})$ относительная).

Для построения доверительного интервала для момента будущего события \hat{t}_{n+1} надо найти дисперсию $D(\widehat{\Delta t})$, которая связана с погрешностью \hat{R}_{abs} или \hat{R}_{sqr} . Однако у оценок этой погрешности есть свой разброс, поэтому расчет будет немного усложняться. В качестве примера проведем такие рассуждения. Пусть $X \sim \text{norm}(m_1, s_1)$, причем $s_1 \sim \text{norm}(m_2, s_2)$, $m_1, m_2, s_2 = \text{const}$, требуется построить доверительный интервал для X . Для этого представим

$$X = m_1 + s_1 \cdot u_1 = m_1 + (m_2 + s_2 \cdot u_2) \cdot u_1,$$

где u_1, u_2 соответствуют $\text{norm}(0,1)$. Тогда определяем дисперсию $D(X)$

$$\begin{aligned} D(X) &= D(m_1) + D[(m_2 + s_2 \cdot u_2) \cdot u_1] = \\ &= D(m_1) + [E(m_2 + s_2 \cdot u_2)]^2 \cdot D(u_1) + [E(u_1)]^2 \cdot D(m_2 + s_2 \cdot u_2) \\ &\quad + D(m_2 + s_2 \cdot u_2) \cdot D(u_1) = \\ &= D(m_1) + m_2^2 \cdot 1 + 0^2 \cdot D(m_2 + s_2 \cdot u_2) + D(m_2 + s_2 \cdot u_2) \cdot 1 = \\ &= D(m_1) + m_2^2 + D(m_2) + [E(s_2)]^2 D(u_2) + [E(u_2)]^2 D(s_2) + D(s_2) D(u_2) = \\ &= D(m_1) + m_2^2 + D(m_2) + s_2^2 \cdot 1 + 0^2 \cdot 0 + 0 \cdot 1 = \\ &= D(m_1) + m_2^2 + D(m_2) + s_2^2. \end{aligned}$$

Тут было использовано свойство дисперсии произведения двух независимых случайных величин ($D(XY) = [E(X)]^2 D(Y) + [E(Y)]^2 D(X) + D(X)D(Y)$), также специально оставлено $D(m_1)$ и $D(m_2)$ так как на практике мы имеем оценки мат. ожиданий, дисперсии которых также учитываются при построении доверительного интервала. Вернемся к дисперсии интервала времени до следующего события $D(\widehat{\Delta t})$, в силу вышесказанного

$$D(\widehat{\Delta t}) = D(E(\widehat{\Delta t})) + \hat{R}^2 + D(E(\hat{R})) + D(\hat{R}), \quad (4.45)$$

где мы вынуждены предположить $D(E(\widehat{\Delta t})) = 0$ (для упрощения расчетов, иначе надо считать дисперсию параметров самого сплайна), \hat{R} – один из способов расчета погрешности, а вот $D(E(\hat{R})) + D(\hat{R})$ удобнее посчитать как одну величину.

Зависимости R_{abs} и R_{sqr} лучше всего описываются степенной моделью, для которой после логарифмирования получаются следующие линеаризованные модели

$$\text{Ln}(\hat{R}_{abs}) = \begin{matrix} -3,2872 + 0,48804 \cdot \text{Ln}(\sigma_y) + \varepsilon \\ (0,02899) \quad (0,01478) \quad (0,41878) \end{matrix}, \quad (4.46)$$

$$\text{Ln}(\hat{R}_{sqr}) = \begin{matrix} -3,0175 + 0,49116 \cdot \text{Ln}(\sigma_y) + \varepsilon \\ (0,02948) \quad (0,01414) \quad (0,41482) \end{matrix}, \quad (4.47)$$

$$\hat{R}_{abs} = 0,03736 \cdot \sigma_y^{0,48804}, \quad (4.48)$$

$$\hat{R}_{sqr} = 0,04892 \cdot \sigma_y^{0,49116}. \quad (4.49)$$

Для дальнейших рассуждений потребуется определить дисперсию

$$D(\text{Ln}(\hat{R}_{abs})) = 0,02899^2 + 0,01478^2 \cdot [\text{Ln}(\sigma_y)]^2 + 0,41878^2, \quad (4.50)$$

$$D(\text{Ln}(\hat{R}_{sqr})) = 0,02948^2 + 0,01414^2 \cdot [\text{Ln}(\sigma_y)]^2 + 0,41482^2, \quad (4.51)$$

$$D(\hat{R}_{abs}) = D(e^{\text{Ln}(\hat{R}_{abs})}) = (e^{\text{Ln}(\hat{R}_{abs})})^2 \cdot D(\text{Ln}(\hat{R}_{abs})) = (\hat{R}_{abs})^2 \cdot D(\text{Ln}(\hat{R}_{abs})), \quad (4.52)$$

$$D(\hat{R}_{sqr}) = D(e^{\text{Ln}(\hat{R}_{sqr})}) = (e^{\text{Ln}(\hat{R}_{sqr})})^2 \cdot D(\text{Ln}(\hat{R}_{sqr})) = (\hat{R}_{sqr})^2 \cdot D(\text{Ln}(\hat{R}_{sqr})). \quad (4.53)$$

В последних двух выражениях использовалось правило расчета дисперсии для сложной функции $D(Y(x_1, \dots, x_n)) = \left(\frac{\partial Y}{\partial x_1}\right)^2 D(x_1) + \dots + \left(\frac{\partial Y}{\partial x_n}\right)^2 D(x_n)$. Причем эти два выражения показывают суммарный разброс математического ожидания вместе с дисперсией ($D(E(\hat{R})) + D(\hat{R})$ в выражении 4.44).

В итоге дисперсия $D(\widehat{\Delta t})$ будет следующей, формула (4.54)

$$\begin{aligned} D(\widehat{\Delta t}) &= (\hat{R}_{sqr})^2 + (\hat{R}_{sqr})^2 \cdot D(\text{Ln}(\hat{R}_{sqr})) = (\hat{R}_{sqr})^2 \left(1 + D(\text{Ln}(\hat{R}_{sqr}))\right) = \\ &= (0,04892 \cdot \sigma_y^{0,49116})^2 \left(1 + 0,02948^2 + 0,01414^2 \cdot [\text{Ln}(\sigma_y)]^2 + 0,41482^2\right) = \\ &= (0,04892 \cdot \sigma_y^{0,49116})^2 \left(1,17294 + 0,01414^2 \cdot [\text{Ln}(\sigma_y)]^2\right). \end{aligned} \quad (4.54)$$

Так, если уровень приглушения наблюдений σ_y составит 5%, то расчет $\sqrt{D(\widehat{\Delta t})}$ дает 0,1683 или 11,7% (при $\sigma_y = 10\%$ имеем $\sqrt{D(\widehat{\Delta t})} = 16,42\%$).

Для построения доверительного интервала откладываем $\widehat{\Delta t} \sqrt{D(\widehat{\Delta t})}$ от ожидаемого момента времени наступления следующего события \hat{t}_{n+1} ,

распределение считаем нормальным (слишком много участвует факторов). Например, доверительный интервал в 95% будет следующим

$$t_{n+1} \in \left[\hat{t}_{n+1} - 2\widehat{\Delta t} \cdot D(\widehat{\Delta t})^{0.5}; \hat{t}_{n+1} + 2\widehat{\Delta t} D(\widehat{\Delta t})^{0.5} \right]. \quad (4.55)$$

Так при уровне приглушения наблюдений σ_y равном 5%, доверительный интервал составляет $\pm 23,4\%$ от интервала между событиями (при $\sigma_y = 10\%$ получим $\pm 32,84\%$), откладываемый от прогнозного момента времени \hat{t}_{n+1} . В примере №2 из раздела 2.2 первые прогнозируемые события попадают в соответствующий доверительный интервал. Для примеров на основе реальных данных осуществляем ретроспективный прогноз для последнего события. В примере «поставка кваса» прогнозная дата последнего события совпадает с фактической (т.к. проверка запаса осуществляется раз в 7 дней). В «парикмахерская» прогнозная дата последнего события отличается на 2 дня (при интервале между событиями в 33 дня). Таким образом события попадают в соответствующие доверительные интервалы.

Это был доверительный интервал для момента времени появления будущего события t_{n+1} , но мы также могли делать прогноз и для величины самого события y_{n+1} . В примерах из главы 2 величину y_{n+1} мы определяли совсем по другим принципам, в основном эмпирически на основе здравого смысла. Например, в том же примере №2 из раздела 2.2 для первого и второго клиента объем покупки вычислялся на основе имеющихся покупок, а для третьего клиента на основе дополнительных расчетов о том, когда клиент пожелает увеличить максимальный запас, а когда уменьшить. В каждом таком случае можно также своим способом определить величину погрешности и построить доверительный интервал для значения y_{n+1} . В примерах же из сферы услуг величина y_i всегда равна единице и разброса нет, значит и строить доверительный интервал не нужно.

Заметим, что если будущее событие t_{n+1} не попадает в рассчитанный доверительный интервал (4.55), то это говорит о том, что на одном из этапов была допущена ошибка. Ошибка могла быть допущена как на этапе

экстраполяции (выбрана неправильная модель), так и на этапе восстановления функции (неправильно подобран коэффициент сглаживания или было недостаточно данных), и конечно же на этапе, когда выдвигалось предположение о самом характере процесса образования событий, действующего в источнике событий.

Выводы по главе 4

1) Показано как появляется погрешность наблюдений, связанная с тем, что время измеряется дискретно (например, по дням). Получены формулы для расчета соответствующей погрешности. На численном примере подтверждена справедливость полученных формул.

2) Для процесса потребления показана какая появляется дополнительная погрешность в наблюдениях, если само потребление осуществляется конечными потребителями через последовательность промежуточных распространителей, а данные собираются на одном из уровней этой цепочки распространения. Результаты исследования показывают, что погрешность растет с замедлением как геометрически убывающая прогрессия. Исследование проведено как с помощью имитационного моделирования, так и с помощью математического анализа. Получены формулы для значения дополнительной погрешности и ее дисперсии в зависимости от удаленности от конечного потребителя, количества покупателей у каждого промежуточного распространителя и количества покупок, на сколько хватает запасов каждого распространителя. Все формулы подтверждены численным примером.

3) Определено как растет погрешность восстановления функции от потери части наблюдений. Показано, что погрешность при использовании емкостного метода растет медленнее чем у классического подхода, когда по наблюдениям строится временной ряд. Исследование проведено с помощью имитационного моделирования. С помощью математического анализа удалось показать, что до тех пор, пока за один период времени встречается меньше 2 событий, погрешность емкостного метода будет меньше.

4) Определена зависимость возникающей погрешности восстановления исходной функции при увеличении погрешности исходных наблюдений, при появлении неопределенности в самом процессе образования событий и при появлении погрешности в дате образования события. Для всех этих зависимостей оценена линейная модель.

5) Проведено сравнение емкостного метода анализа и прогнозирования редких событий с другими существующими методами, используемыми в экономике. Результаты показали, что с помощью емкостного метода можно точнее прогнозировать будущие события, что, например, в торговле позволит держать запасы на более низком уровне.

6) Разработана методика для проверки адекватности получающихся моделей прогнозирования редких событий, который опирается на построение доверительного интервала для момента времени следующего события. Получены формулы для расчета соответствующего доверительного интервала, которые основываются на величине приглушения данных наблюдения на этапе восстановления параметров процесса образования событий.

Глава 5

Обобщение идеи емкостного метода

В этом разделе методика исследования редких событий будет обобщена на произвольные процессы, не обязательно на процессы потребления. Материалы исследования опубликованы в [113; 114; 115; 116; 117; 118].

5.1 Использование емкостного метода для анализа исторических, социальных и политических процессов

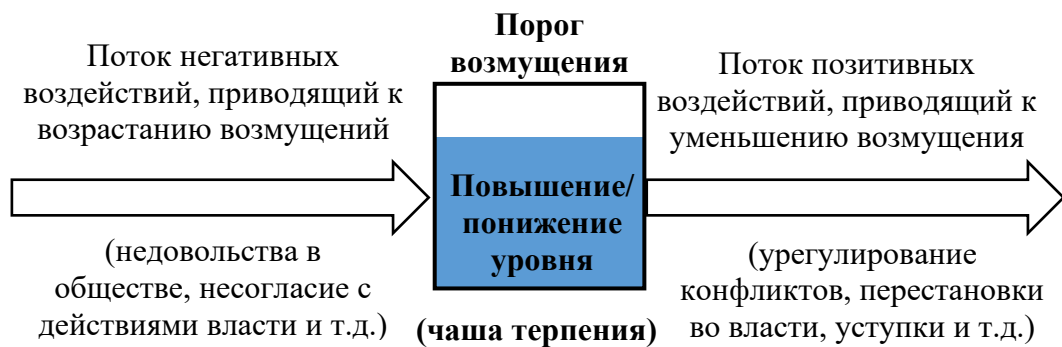
В этом параграфе¹⁾ все еще используется предположение, что процесс образования событий можно моделировать как процесс наполнения/опустошения некоторой емкости.

Использование емкостного метода для анализа «цветных революций»

Будем считать, что наступление редкого события происходит вследствие накопления некоторого потока воздействий (на общество). Этими потоками воздействий могут быть: сообщения в средствах массовой информации и социальных сетях, несущие негативный окрас действующей власти (так называемого «режима»); изменения в экономическом благополучии групп граждан; разногласия и конфликты между слоями населения; политическая деятельность противоположных групп; неудачные высказывания и слабость политиков со стороны действующей власти; и т. д. В то же время существует поток положительных воздействий, которые могут снижать негативные настроения. Например, сообщения о нахождении компромисса между правительством и протестующими, отмена лишней бюрократизации, перестановки во власти, частичное выполнение условий оппозиции, различные уступки, и т.д..

¹⁾ Некоторые результаты исследования получены совместно с исполнителями НИР – Головановой П.С. и Кострицей Т.А., которые на момент начала НИР являлись студентами автора диссертационной работы, в задачи которых входил поиск данных и первичная обработка.

Наступление самого дискретного события будем считать происходит при накоплении негативного возмущения и достижения порога возмущения (переполнения чаши терпения), вследствие дисбаланса негативного и положительного потока воздействий. Причем само возникновение этого события приводит к выплескиванию этого возмущения, что в свою очередь так или иначе приводит к сбрасыванию этого возмущения, и процесс накопления возмущения происходит по новой. На рисунке 5.1 изображена модель процесса образования таких событий.



Источник: составлено автором вместе с соавторами [113].
Рисунок 5.1 – Модель процесса образования социальных событий

Заметим, что данный механизм образования событий является лишь упрощенным отображением социально-исторических процессов, приводящих к возникновению событий, данная модель не может полностью заменить существующие социальные и исторические исследования.

Чтобы использовать емкостный метод, требуется, чтобы события образовывались одним процессом (у них должен быть общий источник). Поэтому для дальнейших рассуждений необходимо ввести предположение – *события появляются вследствие функционирования одного и того же процесса друг за другом.* Это предположение, конечно, дискуссионное, оно требует, чтобы подготовка к следующей цветной революции (вспышке, а не завершению) начиналось от предыдущей вспышки. Для выполнения этого предположения необходимо, чтобы силы, которые разжигают революции по всему миру (в данном случае только на постсоветском пространстве), не могли распределять свои усилия на несколько стран одновременно. В то же время, в

исходных данных о цветных революциях, таблица 5.1, присутствуют несколько событий, которые происходят в один и тот же год. Такие события мы вынуждены объединить в одно событие (как будто подготовка к этим двум событиям началась одновременно).

Другим вопросом является то, что выбрать за величину u_i . В качестве этой величины можно выбрать: сам факт возникновения события, количество участников протестов (в открытых источниках приводится лишь приблизительная оценка), оценку результата протестов для разных сторон. Причем при объединении нескольких событий в одно, величины надо складывать (для факта и количества) или брать среднее значение (для результата).

Таблица 5.1 – Цветные революции на постсоветском пространстве

Даты t_i	Название (популярное), присвоенное событию	Варианты u_i			
		Факт	Кол-во участников (приблизительно)	Результат {поражение/ничья/победа}	
				{1,2,3}	{3,2,1}
02.11.2003	Революция роз в Грузии	1	100000	3	1
22.11.2004	Оранжевая революция в Украине	1	250000	3	1
22.03.2005	Тюльпановая революция в Киргизии	1	50000	3	1
19.03.2006	попытка Васильковой революции в Беларуси	1	30000	1	3
20.02.2008	Протесты в Армении	1	30000	1	3
06.04.2009	Протесты в Молдавии	1	7000	1	3
06.04.2010	Вторая дынная революция в Киргизии	1	300000	3	1
04.12.2011	Выступления на Болотной площади в России	1	170000	1	3
21.11.2013	Коричневая революция в Украине	1	50000	3	1
17.06.2015	Революция розеток (электромайдан) в Армении	1	20000	1	3
06.09.2015	Хризантемовая революция в Молдавии	1	50000	3	1
17.06.2016	Попытка цветной революции (захват заложников) в Армении	1	20000	1	3
11.04.2018	Бархатная революция (любви) в Армении	1	265000	3	1
09.08.2020	Массовые протесты в Белоруссии	1	1000000	1	3

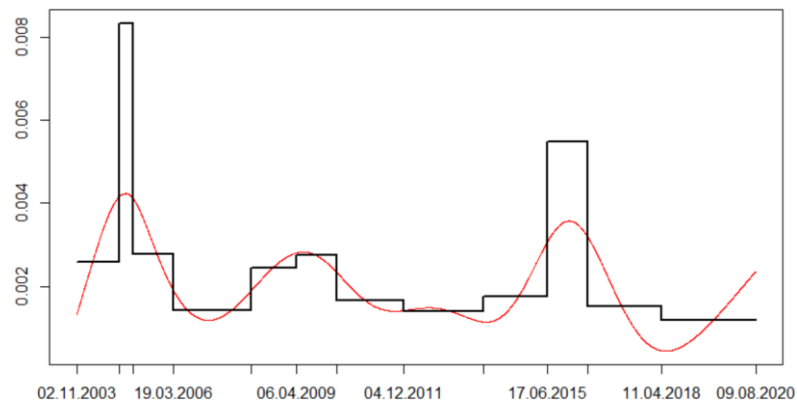
Источник: [119-138].

При применении емкостного метода становится возможным восстановить значение только результирующего потока (разницы между потоком положительных и негативных возмущений)

$$f(t) = \text{входящий поток} - \text{исходящий поток.}$$

Однако отдельные составляющие, т.е. отдельно входящий и отдельно исходящий поток мы восстановить, увы, не сможем. Восстановленная функция может иметь разную интерпретацию в зависимости от того, что было выбрано за величину y_i .

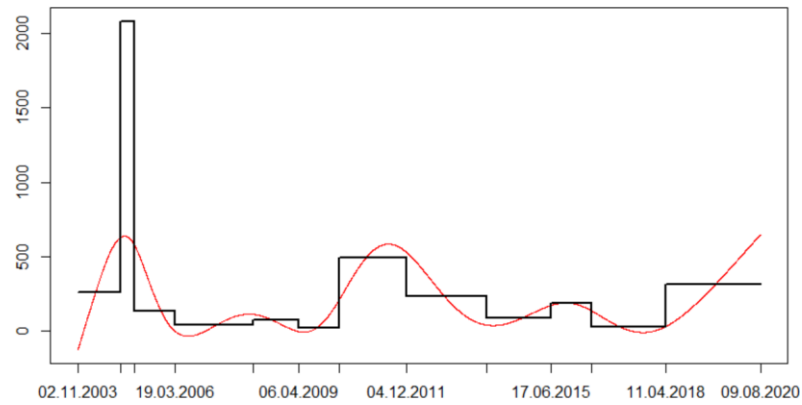
1) *По факту.* Восстановленную функцию можно интерпретировать как относительную скорость роста недовольства в обществе или относительная скорость подготовки к подрыву общественного порядка и попытке смены власти (доля в день), рисунок 5.2.



Источник: составлено автором вместе с соавторами [113].

Рисунок 5.2 – Восстановление результирующего потока по факту возникновения события (относительная скорость нарастания недовольства в обществе, относительная скорость подготовки к попытке смены власти), доля в день

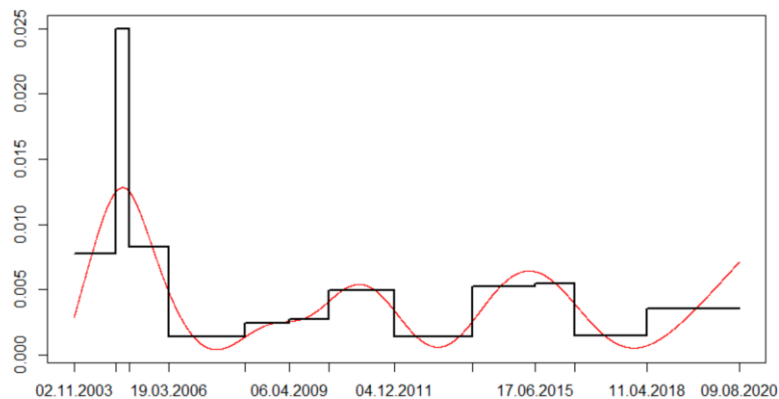
2) *По количеству участников протеста.* Восстановленную функцию можно интерпретировать как скорость роста количества подготовленных протестующих в обществе, рисунок 5.3.



Источник: составлено автором вместе с соавторами [113].

Рисунок 5.3 – Восстановление результирующего потока по количеству участников конфликта (скорость подготовки протестующих), кол-во/день

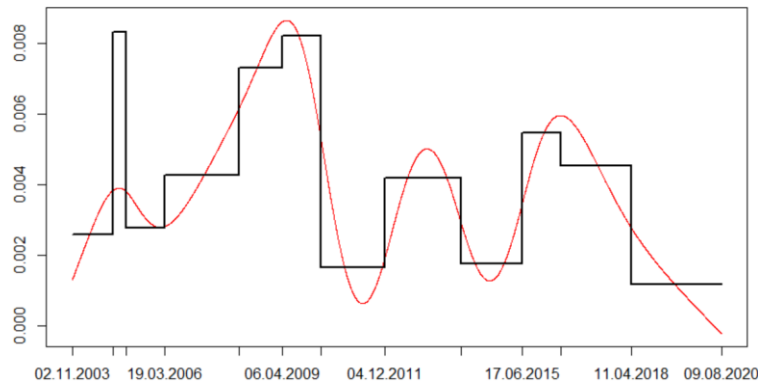
3) *По результату (победа протестующих)*. Восстановленную функцию можно интерпретировать как скорость подготовки к протестам, но уже для достижения выполнения условий протестующих, или как интенсивность усилий для достижения целей, рисунок 5.4.



Источник: составлено автором вместе с соавторами [113].

Рисунок 5.4 – Восстановление результирующего потока по результату: подавление властью – 1, ничья – 2, победа протестующих – 3 (интенсивность усилий для достижения целей), доля победы/день

4) *По результату (победа действующей власти)*. Восстановленную функцию можно интерпретировать как интенсивность усилий действующей власти, направленных на подавление возникающих недовольств, рисунок 5.5.



Источник: составлено автором вместе с соавторами [113].

Рисунок 5.5 – Восстановление результирующего потока по результату, победа протестантов 1, ничья 2, подавление протестов 3 (интенсивность усилий государства для подавления восстания), доля победы/день

На данном этапе исследования отказываемся от попытки экстраполяции восстановленных закономерностей и прогноза будущих событий, так как для самого восстановления мы были вынуждены принять достаточно неоднозначные дискуссионные условия, строить модель для экстраполяции в этом случае будет безответственно. Однако сама возможность исследовать события, сам подход может заинтересовать историков и социологов. Восстанавливаемые зависимости могут иметь интересную интерпретацию.

Использование емкостного метода для анализа исторических событий на примере Русско-Турецких войн

Аналогично, как в параграфе 5.1.1 механизм образования событий (войн) можно попробовать представить, как переполнение некоторой чаши терпения, которая наполняется под воздействием определенного потока возмущения. Таким потоком могут быть различные разногласия, недопонимания, нежелание идти на компромисс, различные мелкие стычки и конфликты, неприязнь к другой культуре, религии и обычаям, и так далее. В то же время может существовать поток позитивных воздействий, приводящий к уменьшению возмущения. Таким потоком может служить дипломатия, переговоры и урегулирование споров, уступки и компромисс, позитивное влияние культуры, взаимный обмен и торговля, и тому подобное.

Так же, как и в предыдущем случае, в качестве величины y_i можно взять разные значения, например, сам факт, длительность войны, результат, когда победе уделяется большее внимание, и результат, когда поражению уделяется большее внимание, таблица 5.2.

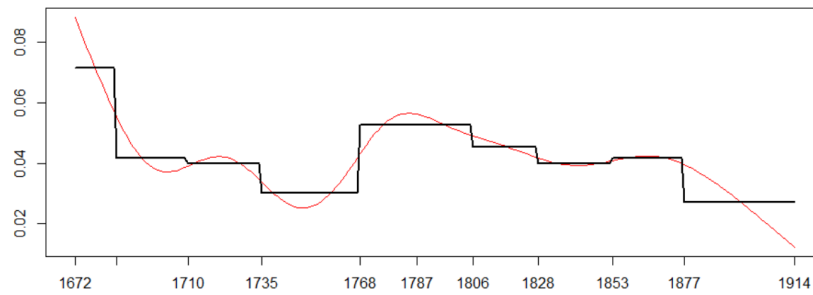
Таблица 5.2 – Данные Русско-Турецких войн

Годы	Название	Начало t_i	Варианты y_i			
			Факт	Длитель- ность, лет	Результат {поражение/ ничья/победа}	
					{1,2,3}	{3,2,1}
1568-1570	Русско-турецкая война	1568	1	3	3	1
1672-1681	Русско-турецкая война	1672	1	10	2	2
1686-1700	Русско-турецкая война	1686	1	15	3	1
1710-1713	Русско-турецкая война	1710	1	4	1	3
1735-1739	Русско-турецкая война	1735	1	5	2	2
1768-1774	Русско-турецкая война	1768	1	7	3	1
1787-1791	Русско-турецкая война	1787	1	5	3	1
1806-1812	Русско-турецкая война	1806	1	7	3	1
1828-1829	Русско-турецкая война	1828	1	2	3	1
1853-1856	Крымская война	1853	1	4	1	3
1877-1878	Русско-турецкая война	1877	1	2	3	1
1914-1918	Кавказский фронт Первой мировой войны	1914	1	5	1	3

Источник: [139].

С помощью емкостного метода можно восстановить динамику результирующего потока возмущений. Этому результирующему потоку можно дать разную интерпретацию в зависимости от того, какие величины выбирались в качестве величин y_i .

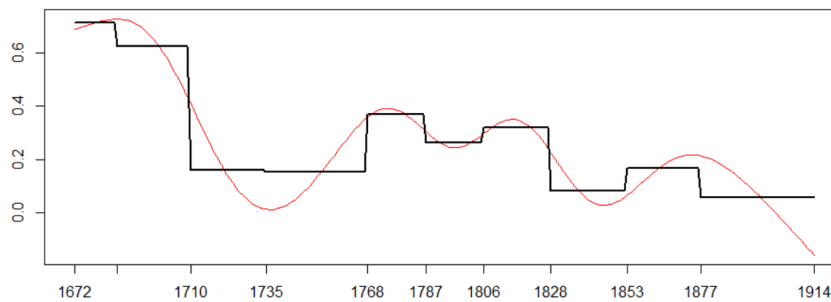
1) *По факту*. Восстановленную функцию можно интерпретировать как относительную скорость подготовки к очередной войне (доля в день) или как скорость нарастания уровня непонимания между двумя сторонами, выраженную в долях к такому уровню непонимания, который приводит к войне, рисунок 5.6.



Источник: составлено автором вместе с соавторами [114].

Рисунок 5.6 – Восстановление результирующего потока по факту возникновения события (относительная скорость подготовки к очередной войне, скорость нарастания непонимания), доля в год

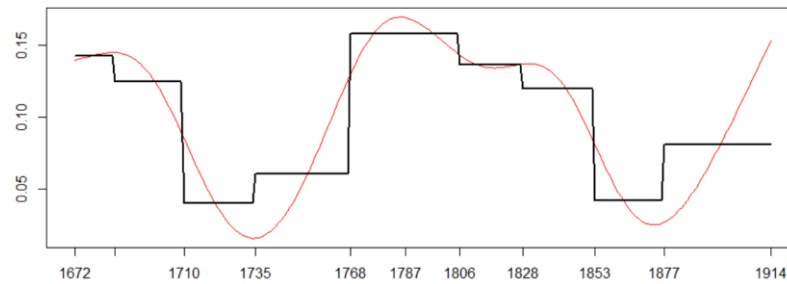
2) *По длительности.* Восстановленную функцию можно интерпретировать как усталость от затяжного конфликта, выраженную как отношение военного времени к одному году жизни, или как долю времени, которую приходилось тратить на подготовку к одному году войны, рисунок 5.7.



Источник: составлено автором вместе с соавторами [114].

Рисунок 5.7 – Восстановление результирующего потока по длительности конфликта (усталость от затяжного конфликта, доля времени на подготовку к войне)

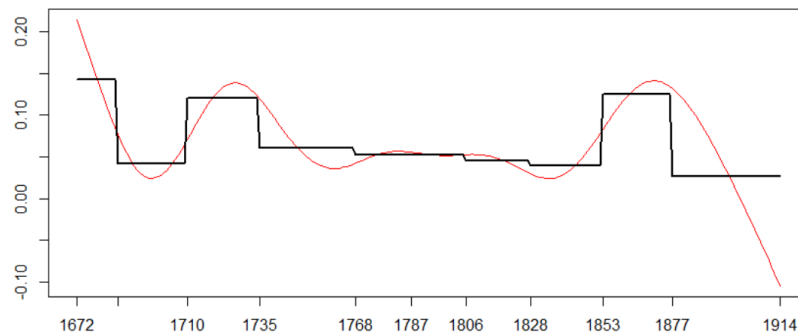
3) *По результату (победа).* Восстановленную функцию можно интерпретировать как некоторый эмоциональный окрас заданному периоду времени, как меняется настроение в результате военных действий или как скорость (интенсивность) подготовки к военным действиям, но уже для достижения победы, рисунок 5.8.



Источник: составлено автором вместе с соавторами [114].

Рисунок 5.8 – Восстановление результирующего потока по результату, поражение 1, ничья 2, победа 3 (настроение от успехов военных действий, интенсивность усилий для победы)

4) *По результату (поражение)*. Восстановленную функцию можно интерпретировать как интенсивность подрывной деятельности некоторых сил, приводящей к поражению, или как интенсивность подготовки противника к войне для своей победы, рисунок 5.9.



Источник: составлено автором вместе с соавторами [114].

Рисунок 5.9 – Восстановление результирующего потока по результату, поражение 3, ничья 2, победа 1 (интенсивность подрывной деятельности некоторых сил, интенсивность подготовки противника к войне для своей победы)

В некоторых местах восстановленная функция принимает отрицательные значения, однако ранее мы говорили, что мы восстанавливаем результирующий поток. Так в самом последнем случае отрицательные значения могут показывать либо борьбу с подрывными силами, либо некоторые контрдействия, препятствующие противнику подготовке к победе в очередной войне. С другой стороны, это может противоречить физическому смыслу, в этом случае надо вносить соответствующие ограничения в сам математический метод восстановления функции.

В данной части исследования мы не берем на себя ответственность экстраполировать найденные зависимости, так как это должны делать историки, разбирающиеся в сложных государственных отношениях, союзах

между странами и образующихся блоков в разные исторические эпохи. Однако сам подход для анализа исторических событий может быть интересен соответствующим специалистам.

5.2 Восстановление функции по разным функционалам для анализа и прогнозирования редких событий, вызванных произвольными процессами

Во всех предыдущих разделах восстанавливалась функция скорости наполнения/опустошения емкости по данным редких событий, которые рассматривались как интегралы от ненаблюдаемой функции. Однако для анализа других видов процессов образования событий может потребоваться другая информация. В данном параграфе будет получен метод восстановления функции одновременно по разным видам функционалов. Пусть события будут нести информацию о таких значениях, как значения самой функции в определенный момент времени, значения первой и второй производной в некоторые моменты времени, значения определенных интегралов на определенных периодах времени:

- $y_i = f(t_i) + \varepsilon_i, i = 1, \dots, n_f;$
- $y'_j = f'(t_j) + \varepsilon'_j, j = 1, \dots, n_{df};$
- $y''_l = f''(t_l) + \varepsilon''_l, l = 1, \dots, n_{d^2f};$
- $Y_u = \int_{t_u^a}^{t_u^b} f(t)dt + \varepsilon_u^{int}, u = 1, \dots, n_{int},$

где t_i, t_j, t_l - моменты времени наблюдений за значениями неизвестной функции $f(t)$, ее первой и второй производной;

t_u^a и t_u^b - нижний и верхний диапазон интегрирования у соответствующего наблюдения за интегралом;

$\varepsilon_i, \varepsilon'_j, \varepsilon''_l, \varepsilon_u^{int}$ - погрешности наблюдений значений, первой производной, второй производной и интегралов соответственно (с нулевым математическим ожиданием, дисперсии могут быть разными);

n_f , n_{df} , n_{d^2f} , n_{int} – количество наблюдений значений, первых производных, вторых производных, интегралов искомой функции, соответственно.

Причем эти разные по характеру данные могут быть доступны как все вместе, так и может быть доступно что-то одно, в то же время объемы выборок входных данных разных характеристик могут не совпадать (или равны нулю, если таких данных нет). Опять же готового решения найти не удалось, особенно в том виде, в котором оно получено здесь. Описанное здесь решение опирается на работу Green и Siverman [96].

Разработка метода восстановления функции по разным функционалам

Для восстановления функции по многим функционалам будем минимизировать сразу несколько сумм квадратов отклонения и штрафа на нелинейность. Так как величины, в которых измеряются значения, производные и интегралы, могут сильно отличаться друг от друга, добавим соответствующие коэффициенты, чтобы можно было увеличивать вес каждой группе наблюдений. Дополнительно добавим возможность изменять веса каждого отдельного наблюдения. В результате оптимизационная задача примет следующий вид, формула (5.1)

$$S(f) = \sum_{i=1}^{n_f} w_i^f (y_i - f(t_i))^2 + \mu \sum_{j=1}^{n_{df}} w_j^{df} (y_j' - f'(t_j))^2 + \nu \sum_{l=1}^{n_{d^2f}} w_l^{d^2f} (y_l'' - f''(t_l))^2 + \psi \sum_{u=1}^{n_{int}} w_u^{int} \left(Y_u - \int_{t_u^a}^{t_u^b} f(t) dt \right)^2 + \alpha \int_{t_{start}}^{t_{end}} (f''(t))^2 dt, \quad (5.1)$$

где w_i^f , w_j^{df} , $w_l^{d^2f}$, w_u^{int} - индивидуальные веса соответствующих групп наблюдений;

μ (мю) - вес всей группы наблюдений первых производных;

ν (ню) - вес всей группы наблюдений вторых производных;

ψ (пси) - вес всей группы наблюдений интегралов;

последнее слагаемое - штраф на нелинейность (шероховатость);

α - коэффициент сглаживания (регуляризации);

t_{start} и t_{end} соответственно границы, в которых происходит восстановление функции.

Заметим, что отсутствует весовой коэффициент для всей группы наблюдений обычных значений, то есть он предполагается равным единице, а все остальные коэффициенты тогда показывают вес по сравнению с этой первой группой наблюдений.

Неизвестную функцию $f(t)$ восстанавливаем в виде кубического сплайна $g(t)$, который имеет такое же представление через значение и вторую производную (value-second derivative representation), как и в главе 3. Но для удобства его можно более компактно записать в следующем виде

$$g(t) = \frac{h_k^-}{h_k} g_{k+1} + \frac{h_k^+}{h_k} g_k - \frac{h_k^- h_k^+ (h_k + h_k^-)}{6h_k} \gamma_{k+1} - \frac{h_k^- h_k^+ (h_k + h_k^+)}{6h_k} \gamma_k, \quad (5.2)$$

$$k: s_k \leq t_i < s_{k+1};$$

где s_k – узлы сплайна;

$h_k = s_{k+1} - s_k$, $h_k^- = t - s_k$, $h_k^+ = s_{k+1} - t$ – есть шаг между узлами или между узлом и наблюдением;

$g_k = g(s_k)$ – значение сплайна в узле s_k ;

$\gamma_k = g''(s_k)$ – вторая производная в узле s_k .

Эта запись эквивалента записи из главы 3, но более компактна.

Для решения оптимизационной задачи (5.1) надо вычислить соответствующие функционалы, подставляя выражение для сплайна (5.2).

Запись соответствующих функционалов будет следующей

$$g(t_i) = \frac{h_k^{-i}}{h_k} g_{k+1} + \frac{h_k^{+i}}{h_k} g_k - \frac{h_k^{-i} h_k^{+i} (h_k + h_k^{-i})}{6h_k} \gamma_{k+1} - \frac{h_k^{-i} h_k^{+i} (h_k + h_k^{+i})}{6h_k} \gamma_k, \quad (5.3)$$

$$k: s_k \leq t_i < s_{k+1},$$

$$g'(t_j) = \frac{g_{k+1}}{h_k} - \frac{g_k}{h_k} - \left(\frac{h_k}{6} - \frac{(h_k^{-j})^2}{2h_k} \right) \gamma_{k+1} + \left(\frac{h_k}{6} - \frac{(h_k^{+j})^2}{2h_k} \right) \gamma_k, \quad (5.4)$$

$$k: s_k \leq t_j < s_{k+1},$$

$$g''(t_l) = \frac{h_k^{-l}}{h_k} \gamma_{k+1} + \frac{h_k^{+l}}{h_k} \gamma_k, \quad (5.5)$$

$$k: s_k \leq t_l < s_{k+1},$$

$$\int_{t_u^a}^{t_u^b} g(t) dt = \sum_{l=0}^L \int_{s_{k+l}}^{s_{k+l+1}} g(t) dt - \int_{s_k}^{t_u^a} g(t) dt - \int_{t_u^b}^{s_{k+L+1}} g(t) dt = \quad (5.6)$$

$$L: s_{k+L} < t_u^b \leq s_{k+L+1}, \\ k: s_k \leq t_u^a < s_{k+1},$$

$$= \sum_{l=0}^L \left[\frac{h_{k+l}}{2} g_{k+l+1} + \frac{h_{k+l}}{2} g_{k+l} - \frac{h_{k+l}^3}{24} \gamma_{k+l+1} - \frac{h_{k+l}^3}{24} \gamma_{k+l} \right] \\ - \frac{(h_k^-)^2}{2h_k} g_{k+1} - \frac{h_k^2 - (h_k^+)^2}{2h_k} g_k - \frac{(h_k^-)^2 ((h_k^-)^2 - 2h_k^2)}{24h_k} \gamma_{k+1} + \frac{(h_k^-)^2 (h_k^+ + h_k)^2}{24h_k} \gamma_k \\ - \frac{h_{k+L}^2 - (h_{k+L}^-)^2}{2h_{k+L}} g_{k+L+1} - \frac{(h_{k+L}^+)^2}{2h_{k+L}} g_{k+L} \\ + \frac{(h_{k+L}^+)^2 (h_{k+L}^- + h_{k+L})^2}{24h_{k+L}} \gamma_{k+L+1} - \frac{(h_{k+L}^+)^2 ((h_{k+L}^+)^2 - 2h_{k+L}^2)}{24h_{k+L}} \gamma_{k+L}, \quad (5.7)$$

$$L: s_{k+L} < t_u^b \leq s_{k+L+1}, \\ k: s_k \leq t_u^a < s_{k+1}.$$

Здесь величины h_k^{-i} , h_k^{-j} , h_k^{-l} – показывают расстояние от узла до наблюдения, h_k^{+i} , h_k^{+j} , h_k^{+l} – показывают расстояние от наблюдения до следующего узла, $h_k^{-a} = t_u^a - s_k$ – показывает расстояние от узла до нижнего предела интеграла, $h_k^{+a} = s_{k+1} - t_u^a$ – расстояние от нижнего предела интеграла до следующего узла, $h_{k+L}^{-b} = t_u^b - s_{k+L}$ – расстояние от узла до верхнего предела интеграла, $h_{k+L}^{+b} = s_{k+L+1} - t_u^b$ – расстояние от верхнего предела интеграла до следующего узла. Во всех этих выражениях в начале определяется на какой интервал k выпало наблюдение. В самом последнем выражении для интеграла необходимо в начале определить на какой интервал k выпал нижний предел интегрирования t_u^a и на какой интервал $k + L$ выпал верхний предел интегрирования t_u^b , где L – количество интервалов между ними (L может быть равно 0, если оба выпали на один интервал).

Все эти выражения имеют линейную форму относительно неизвестных параметров сплайна g_k и γ_k . Поэтому оптимизационную задачу (5.1) можно выразить в следующем матричном виде, формула (5.8)

$$\begin{aligned}
S(g) = & (Y_f - V_f g + P_f \gamma)^T W_f (Y_f - V_f g + P_f \gamma) + \\
& + \mu (Y_{df} - V_{df} g + P_{df} \gamma)^T W_{df} (Y_{df} - V_{df} g + P_{df} \gamma) + \\
& + \nu (Y_{d^2f} - 0g + P_{d^2f} \gamma)^T W_{d^2f} (Y_{d^2f} - 0g + P_{d^2f} \gamma) + \\
& + \psi (Y_{int} - V_{int} g + P_{int} \gamma)^T W_{int} (Y_{int} - V_{int} g + P_{int} \gamma) + \\
& + \alpha g^T K g \rightarrow \min,
\end{aligned} \tag{5.8}$$

где $Y_f, Y_{df}, Y_{d^2f}, Y_{int}$ - столбцы наблюдений;

V_f, V_{df}, V_{int} - матрицы коэффициентов при неизвестных g_k ;

$P_f, P_{df}, P_{d^2f}, P_{int}$ - матрицы коэффициентов при неизвестных γ_k ;

$W_f, W_{df}, W_{d^2f}, W_{int}$ - диагональные матрицы индивидуальных весов наблюдений.

Матрицы заполняются следующим образом:

V_f размерностью $n_f \times m$, ее каждая i -тая строка выглядит как

$$\begin{array}{cccccccc}
1 & \dots & k-1 & k & k+1 & k+2 & \dots & m \\
0 & \dots & 0 & h_k^{+i}/h_k & h_k^{-i}/h_k & 0 & \dots & 0
\end{array}$$

P_f размерностью $n_f \times (m-2)$, ее каждая i -тая строка выглядит как

$$\begin{array}{cccccccc}
2 & \dots & k-1 & k & k+1 & k+2 & \dots & m-1 \\
0 & \dots & 0 & h_k^{-i} h_k^{+i} (h_k + h_k^{+i}) / 6h_k & h_k^{-i} h_k^{+i} (h_k + h_k^{-i}) / 6h_k & 0 & \dots & 0
\end{array}$$

V_{df} размерностью $n_{df} \times m$, ее каждая j -тая строка выглядит как

$$\begin{array}{cccccccc}
1 & \dots & k-1 & k & k+1 & k+2 & \dots & m \\
0 & \dots & 0 & -1/h_k & 1/h_k & 0 & \dots & 0
\end{array}$$

P_{df} размерностью $n_{df} \times (m-2)$, ее каждая j -тая строка выглядит как

$$\begin{array}{cccccccc}
2 & \dots & k-1 & k & k+1 & k+2 & \dots & m-1 \\
0 & \dots & 0 & -h_k/6 + (h_k^{+j})^2 / 2h_k & h_k/6 + (h_k^{-j})^2 / 2h_k & 0 & \dots & 0
\end{array}$$

P_{d^2f} размерностью $n_{d^2f} \times (m-2)$, ее каждая l -тая строка выглядит как

$$\begin{array}{cccccccc}
2 & \dots & k-1 & k & k+1 & k+2 & \dots & m-1 \\
0 & \dots & 0 & -h_k^{+l}/h_k & -h_k^{-l}/h_k & 0 & \dots & 0
\end{array}$$

V_{int} размерностью $n_{int} \times m$, ее каждая u -тая строка заполняется следующим образом (как в главе 3):

$$\begin{aligned}
V_{u,k} = \frac{(h_k^{+a})^2}{2h_k}; \quad V_{u,k+l} = \frac{h_{k+l-1} + h_{k+l}}{2}, \quad l = 1, \dots, L; \quad V_{u,k+1} = V_{u,k+1} - \frac{(h_k^{-a})^2}{2h_k}; \\
V_{u,k+L} = V_{u,k+L} - \frac{(h_{k+L}^{+b})^2}{2h_{k+L}}; \quad V_{u,k+L+1} = \frac{(h_{k+L}^{-b})^2}{2h_{k+L}}.
\end{aligned} \tag{5.9}$$

P_{int} размерностью $n_{int} \times (m - 2)$, ее каждая u -тая строка заполняется следующим образом (как в главе 3):

$$P_{u,k} = \frac{h_k^3}{24} - \frac{(h_k^{-a})^2(h_k^{+a} + h_k)^2}{24h_k}; P_{u,k+l} = \frac{h_{k+l-1}^3 + h_{k+l}^3}{24}, l = 1, \dots, L;$$

$$P_{u,k+1} = P_{u,k+1} + \frac{(h_k^{-a})^2((h_k^{-a})^2 - 2h_k^2)}{24h_k}; P_{u,k+L} = P_{u,k+L} + \frac{(h_{k+L}^{+b})^2((h_{k+L}^{+b})^2 - 2h_{k+L}^2)}{24h_{k+L}}; \quad (5.10)$$

$$P_{u,k+L+1} = \frac{h_{k+L}^3}{24} - \frac{(h_{k+L}^{+b})^2(h_{k+L}^{-b} + h_{k+L})^2}{24h_{k+L}}.$$

Заметим, что в зависимости от L некоторые элементы матриц V и P могут изменяться дважды.

Благодаря условиям непрерывности первой производной, позволяющим выразить $\gamma = R^{-1}Q^T g$, оптимизационную задачу (5.8) можно записать более компактно только через одну неизвестную g , формула (5.11)

$$S(g) = (Y_f - C_f g)^T W_f (Y_f - C_f g) +$$

$$\mu (Y_{df} - C_{df} g)^T W_{df} (Y_{df} - C_{df} g) +$$

$$\mu (Y_{df} - C_{df} g)^T W_{df} (Y_{df} - C_{df} g) + \quad (5.11)$$

$$\nu (Y_{d^2f} - C_{d^2f} g)^T W_{d^2f} (Y_{d^2f} - C_{d^2f} g) +$$

$$\psi (Y_{int} - C_{int} g)^T W_{int} (Y_{int} - C_{int} g) +$$

$$\alpha g^T K g \rightarrow \min ,$$

где $C_f = V_f - P_f R^{-1} Q^T$, $C_{df} = V_{df} - P_{df} R^{-1} Q^T$, $C_{d^2f} = 0 - P_{d^2f} R^{-1} Q^T$,
 $C_{int} = V_{int} - P_{int} R^{-1} Q^T$.

Приравнивая производную от $S(g)$ по g к нулю, находим выражение для искомым параметров g , формула (5.12)

$$g = \left(C_f^T W_f C_f + \mu C_{df}^T W_{df} C_{df} + \nu C_{d^2f}^T W_{d^2f} C_{d^2f} + \psi C_{int}^T W_{int} C_{int} + \alpha K \right)^{-1} \times \quad (5.12)$$

$$\left(C_f^T W_f Y_f + \mu C_{df}^T W_{df} Y_{df} + \nu C_{d^2f}^T W_{d^2f} Y_{d^2f} + \psi C_{int}^T W_{int} Y_{int} \right).$$

Зная g , рассчитывается $\gamma = R^{-1}Q^T g$, после чего уже можно строить сплайн $g(t)$ в произвольной точке t по выражению (5.2).

Выбор параметра сглаживания α

Процедуры автоматического выбора параметра сглаживания для задачи восстановления по разным функционалам найти не удалось. Пришлось

самостоятельно получить модифицированные формулы для расчета оценки кросс-валидации

$$CV(\alpha) = n_f^{-1} \sum_{i=1}^{n_f} w_i^f (y_i - g^{(-i)}(t_i, \alpha))^2 + n_{df}^{-1} \mu \sum_{j=1}^{n_{df}} w_j^{df} (y_j' - g'^{(-j)}(t_j, \alpha))^2 + \\ + n_{d^2f}^{-1} \nu \sum_{l=1}^{n_{d^2f}} w_l^{d^2f} (y_l'' - g''^{(-l)}(t_l, \alpha))^2 + n_{int}^{-1} \psi \sum_{\{u=1\}}^{n_{int}} w_u^{int} \left(Y_u - \int_{t_u^a}^{t_u^b} g^{(-u)}(t, \alpha) dt \right)^2 = \quad (5.13)$$

$$= n_f^{-1} \sum_{i=1}^{n_f} w_i^f \left(\frac{y_i - g(t_i, \alpha)}{1 - \sum_{k=1}^m C_{ik}^f A_{ki}^f(\alpha)} \right)^2 + n_{df}^{-1} \sum_{j=1}^{n_{df}} w_j^{df} \left(\frac{y_j' - g'(t_j, \alpha)}{1 - \sum_{k=1}^m C_{jk}^{df} A_{kj}^{df}(\alpha)} \right)^2 + \\ + n_{d^2f}^{-1} \sum_{l=1}^{n_{d^2f}} w_l^{d^2f} \left(\frac{y_l'' - g''(t_l, \alpha)}{1 - \sum_{k=1}^m C_{lk}^{d^2f} A_{kl}^{d^2f}(\alpha)} \right)^2 + n_{int}^{-1} \sum_{\{u=1\}}^{n_{int}} w_u^{int} \left(\frac{Y_u - \int_{t_u^a}^{t_u^b} g(t, \alpha) dt}{1 - \sum_{k=1}^m C_{uk}^{int} A_{ku}^{int}(\alpha)} \right)^2 = \quad (5.14)$$

$$= n_f^{-1} \sum_{\{i=1\}}^{n_f} w_i^f \left(\frac{y_i - \sum_{k=1}^m C_{ik}^f g_k}{1 - \sum_{k=1}^m C_{ik}^f A_{ki}^f(\alpha)} \right)^2 + n_{df}^{-1} \sum_{j=1}^{n_{df}} w_j^{df} \left(\frac{y_j' - \sum_{k=1}^m C_{jk}^{df} g_k}{1 - \sum_{k=1}^m C_{jk}^{df} A_{kj}^{df}(\alpha)} \right)^2 + \\ + n_{d^2f}^{-1} \sum_{\{l=1\}}^{n_{d^2f}} w_l^{d^2f} \left(\frac{y_l'' - \sum_{k=1}^m C_{lk}^{d^2f} g_k}{1 - \sum_{k=1}^m C_{lk}^{d^2f} A_{kl}^{d^2f}(\alpha)} \right)^2 + n_{int}^{-1} \sum_{u=1}^{n_{int}} w_u^{int} \left(\frac{Y_u - \sum_{k=1}^m C_{uk}^{int} g_k}{1 - \sum_{k=1}^m C_{uk}^{int} A_{ku}^{int}(\alpha)} \right)^2. \quad (5.15)$$

Здесь верхние индексы $(-i)$, $(-j)$, $(-l)$, $(-u)$ означают, что сплайн $g(t, \alpha)$ оценен по данным без этого конкретного наблюдения. Матрицы $C^f, C^{df}, C^{d^2f}, C^{int}$ такие же как и в формулах 5.11, 5.12. Матрицы $A^f(\alpha) = A(\alpha)C_f^T W_f$, $A^{df}(\alpha) = A(\alpha)\mu C_{df}^T W_{df}$, $A^{d^2f}(\alpha) = A(\alpha)\nu C_{d^2f}^T W_{d^2f}$, $A^{int}(\alpha) = A(\alpha)\psi C_{int}^T W_{int}$, где $A(\alpha) = (C_f^T W_f C_f + \mu C_{df}^T W_{df} C_{df} + \nu C_{d^2f}^T W_{d^2f} C_{d^2f} + \psi C_{int}^T W_{int} C_{int} + \alpha K)^{-1}$. Минимизация $CV(\alpha)$ любым известным способом относительно α дает искомое значение параметра сглаживания α .

Заметим, что с помощью метода кросс-валидации подбирается такой параметр алгоритма, который делает процедуру оценки функции эффективной с точки зрения предсказания будущих наблюдений, то есть прогноз обладает наименьшей дисперсией, но в качестве новых прогнозных значений по очереди берется одно из уже известных наблюдений, а функция оценивается без этого конкретного наблюдения. Кросс валидация хорошо работает на обычных сплайнах. Однако у нас необычный случай, у нас происходит восстановление также по первым/вторым производным и интегралам. В итоге оценка кросс-валидации показывает не обычную дисперсию наблюдений, а

дисперсию наблюдений как за значениями, так и за производными и интегралами (происходит смешивание разных дисперсий, но при самом построении сплайна мы также смешивали квадраты ошибок разных наблюдений). Но проблема в том, что при исключении некоторых наблюдений может катастрофически сильно поменяться вид восстановленной функции. В описанном ниже примере будет очевидно, что исключение любого наблюдения приведет к тому, что функция определится очень неточно, при этом рассчитывается ошибка прогноза именно для исключенного наблюдения. Когда каждое наблюдение несет крайне необходимую информацию для восстановления функции, исключение наблюдения из выборки приведет к очень и очень большим ошибкам. В этом случае кросс-валидация не подходит. Но если же у нас очень много наблюдений, одни наблюдения лишь статистически дополняют информацию других наблюдений, и исключение одного из наблюдений не приводит к значительным изменениям восстановленной функции, то метод кросс-валидации может быть хорошим решением для определения параметра сглаживания.

Пример анализа редких событий с использованием метода восстановления функции по разным функционалам

Пусть есть некоторая неподконтрольная нам организация, от которой в определенные дискретные моменты времени поступают сигналы, несущие преимущественно качественную (а не количественную) информацию. Причем в каждом сигнале эта качественная информация может относиться к разным характеристикам (не обязательно, что каждый сигнал сообщает обо всех характеристиках). Для примера представьте такую картину, когда от некоторой компании поступают следующие сигналы:

- a) Мы начали тонуть – надо срочно что-то делать!
- b) Мы все еще сильно тонем!
- c) Кажется, снижение начало замедляться.
- d) Мы начали всплывать.
- e) Хорошо идем, вот бы всегда так.

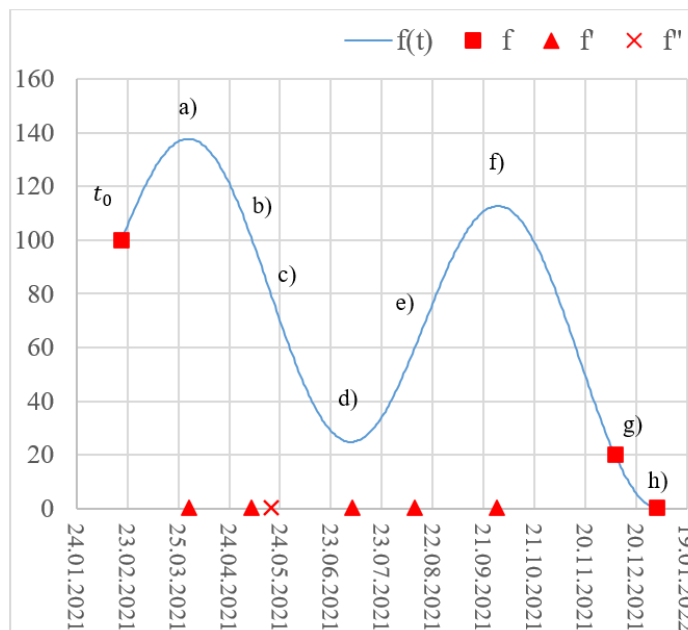
- f) О нет, мы вновь тонем!
- g) Мы прошли отметку невозврата...
- h) Мы на дне.

Можно догадаться, что события образованы внутри такой организации в результате наблюдения за некоторым внутренним показателем, который изменяется динамически со временем (возможно вследствие каких-то управляющих воздействий). Образование событий связано с операцией сравнения этого показателя с некоторыми критическими значениями. Если воспользоваться предложенной методикой, восстановить динамику ненаблюдаемого напрямую внутреннего показателя организации, построить модель изменения этой динамики, провести экстраполяцию и вновь запустить процесс образования событий, то возможно будет предсказать будущие события (и возможно даже управляющие воздействия).

Что же нам известно из подобного вида сигналов, что поможет восстановить динамику изменения ненаблюдаемого напрямую показателя. Оказывается, из имеющихся сигналов можно извлечь следующие данные: моменты времени как моменты получения сообщений (с возможной поправкой на задержку в получении); значения первой производной или положение точек экстремума; значения второй производной или положение точек перегиба; значения самого изменяющегося показателя в некоторых точках. Для большей полноты также предположим, что могут наблюдаться интегралы от функции изменения того же ненаблюдаемого показателя (отсутствует в описанном выше примере, но можно предположить, что если интересующий нас показатель указывает на финансовые средства на счетах, то начисленные проценты по этим счетам за некоторые периоды времени будут этими интегралами). Следует отметить, что все наблюдения могут быть неточными, причем неточными могут быть как сами моменты наблюдения, так и значения в этих наблюдениях. Однако погрешность в моментах наблюдения так или иначе может быть сведена к погрешности в значениях этих же наблюдений (например, если истинное положение экстремума чуть сдвинуто

от момента наблюдения, то в этот момент наблюдения значение производной будет смещено относительно нуля).

Обозначим неизвестную искомую функцию скрытой динамики как $f(t)$, которую в дальнейшем будем восстанавливать по имеющимся данным в виде сплайна. Для простоты пусть функция $f(t)$ будет безразмерной, и в начальный момент времени t_0 значение этой функции будет принято за исходное значение, а все остальные значения выражаются в процентах от этого исходного (то есть полагаем, что $f(t_0) = 100$). Далее, для примера, пусть исходная неизвестная функция ведет себя так, как показано на рисунке 5.10.



Функция $f(t)$ безразмерная, показывает количество % от начального уровня (значение в начальный момент t_0 берется как исходное 100%), данные о значении функции отмечены на самом графике, данные о производных отложены на горизонтальной оси и обозначают их позицию.

Источник: составлено автором.

Рисунок 5.10 – Пример исходной функции $f(t)$ и исходных данных в виде событий

Изначально было сказано, что мы имеем дело с преимущественно качественными данными, но этим качественным данным не составит труда дать количественную оценку, причем достаточно дать приближенную оценку (или же значения могли быть изначально приближенно известны). Так в событиях a, c, d, f говорим, что наблюдаются нулевые значения первых или вторых производных. Для событий b и e предположим, что нам приближенно

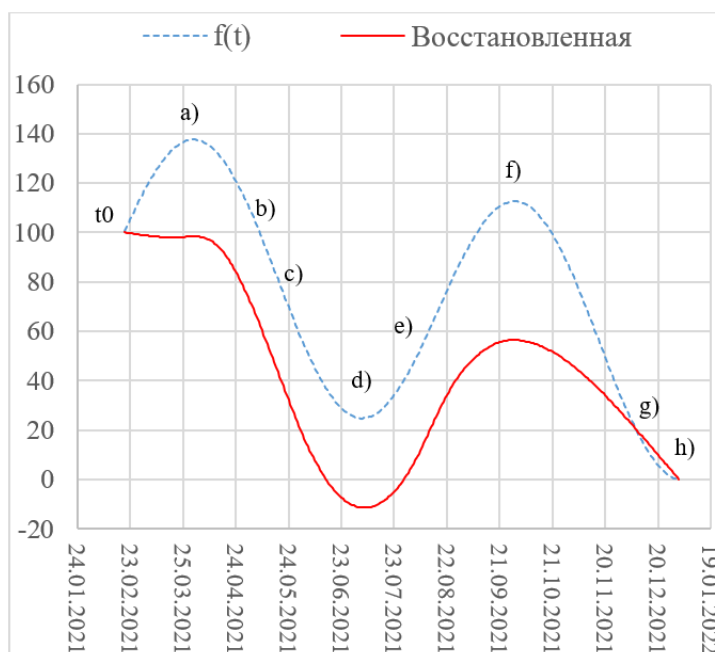
известен тангенс угла наклона, для события g предположим (или знаем), что уровень невозврата начинается на 20%. Мы давали приближенную оценку значениям, а моменты времени событий наблюдаются и известны. В результате имеющиеся данные могут быть следующими, таблица 5.3.

Таблица 5.3 – Имеющиеся приблизительные данные

t_f	y_f	t_{df}	y_{df}	t_{d^2f}	y_{d^2f}
20.02.2021	100	31.03.2021	0	19.05.2021	0
08.12.2021	20	07.05.2021	-1.75	-	-
01.01.2022	0	06.07.2021	0	-	-
-	-	11.08.2021	1.55	-	-
-	-	29.09.2021	0	-	-

Источник: составлено автором.

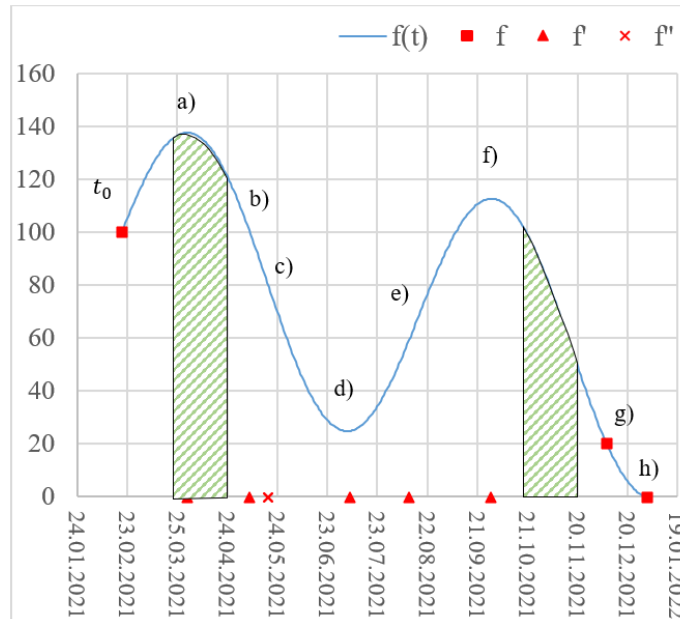
Как можно заметить, объем выборок может сильно отличаться, для второй производной имеем только одно наблюдение. Реализуя описанный метод восстановления функции по функционалам из имеющихся данных (все коэффициенты μ, ν, ψ, α равны 1), получим следующий результат, рисунок 5.11. Задача оказалась плохо обусловленной, было недостаточно информации для хорошего восстановления функции (ничто не указывало на то, что в самом начале функция должна была возрастать).



Источник: составлено автором.

Рисунок 5.11 – Плохо обусловленная задача, недостаточно информации, чтобы точнее восстановить функцию

Если же добавить больше информации, например, в виде интегралов искомой функции, рисунок 5.12, таблица 5.4, то восстановить исходную функцию получается намного точнее, рисунок 5.13.



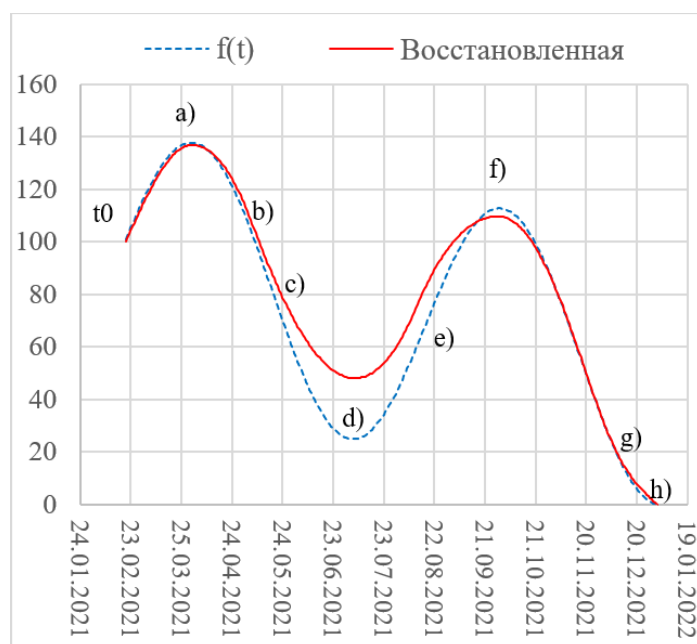
Источник: составлено автором.

Рисунок 5.12 – Исходная функция $f(t)$ и исходные данные в виде событий с добавлением информации об интегралах функции

Таблица 5.4 – Данные с добавлением информации об интегралах

t_f	y_f	t_{df}	y_{df}	t_{d^2f}	y_{d^2f}	t_{int}^a	t_{int}^b	Y_{int}
20.02.2021	100	31.03.2021	0	19.05.2021	0	25.03.2021	24.04.2021	4000
08.12.2021	20	07.05.2021	-1.75	-	-	21.10.2021	20.11.2021	2282
01.01.2022	0	06.07.2021	0	-	-	-	-	-
-	-	11.08.2021	1.55	-	-	-	-	-
-	-	29.09.2021	0	-	-	-	-	-

Источник: составлено автором.



Источник: составлено автором.

Рисунок 5.13 – Восстановление функции по данным с добавлением информации об интегралах

Конечно же восстановление все еще не идеальное, но это связано с нехваткой самой информации, а не с недостатком математического метода. Если добавить больше наблюдений, то можно добиться очень хорошего восстановления. Заметим, что с помощью настройки весов как самих наблюдений, так и групп наблюдений, можно очень гибко настраивать какой информации стоит больше уделять внимания, а какой меньше.

Дальнейшее прогнозирование будущих событий подчиняется все той же схеме (в данном примере мы можем вовсе и не ждать никаких будущих событий). Программная реализация метода представлена в работе [115].

5.3 Разработка универсального метода исследования произвольных процессов

В данном параграфе будет представлен универсальный метод восстановления параметров процесса образования редких событий, причем этот процесс образования событий может быть описан уже произвольным алгоритмом.

Описание процесса в виде алгоритмической модели

Сделаем единственное ограничение на вид процесса образования событий – *процесс не должен быть случайными* (в том смысле, в котором предполагает теория случайных процессов, где определяются законы распределения для интервалов между событиями), то есть в котором для образования событий используются случайные числа. Нас интересуют такие процессы, в которых события образуются как бы детерминировано, по некоторым «физическим» законам. В то же время можно допустить, что внутри процесса присутствует неопределенность, но тогда нас будет интересовать только динамика математического ожидания (усреднение по реализациям) всех нестационарных параметров, то есть опять параметры могут быть описаны аналитически в виде функциональных зависимостей. Однако при самом образовании событий мы будем допускать, что наблюдения за этими событиями могут фиксироваться с некоторой дополнительной погрешностью. В итоге от нас будет требоваться определить нестационарные параметры процесса по зашумленным данным.

В этой части исследования мы предполагаем, что процесс может быть представлен произвольным алгоритмом, а не математическим выражением. Как в эконометрике или анализе данных исследователь формирует математическую модель, так и здесь исследователь будет формировать модель, но не математическую, а алгоритмическую. Ключевым элементом в такой модели процесса является образование дискретных событий в результате некоторых операций сравнения. При формировании событий помимо момента времени t_i может возвращаться информация об этом событии y_i (например, в процессах потребления этой информацией будет тот же объем покупок). Для упрощения предположим, что y_i будут скалярными величинами, хотя не составит труда обобщить описанный ниже алгоритм на случай, когда каждое событие может возвращать вектор из значений. Также допустим, что сами события могут образовываться в разной части процесса в результате разных проверок.

Процесс может состоять как из переменных, так и использовать некоторые параметры, которые будут изменяться со временем в зависимости от некоторых внешних условий и независимо от того, что происходит в самом процессе. Параметры могут быть как стационарными, так и нестационарными. Эти параметры не наблюдаемы, их нельзя назвать экзогенными переменными (переменные процесса тоже не наблюдаемы, наблюдаются только события). Их можно сравнить с коэффициентами эконометрических моделей (a_0, a_1, \dots), но допустить, что они будут динамическими (то есть $a_0(t), a_1(t), \dots$).

В качестве примера процесса рассмотрим процесс потребления, как в моделях управления запасами (с пополнением запаса до максимума), но с учетом того, что этот процесс использует параметры. Параметрами будут максимальный запас и спрос, критический уровень запаса выбирается равным 0, так как он оказывается как бы «спутан» с или «мультиколлинеарен» максимальному запасу, можно одновременно изменять оба из них на одинаковое число, и результат не изменится¹⁾. Начальное и конечное время задается извне модели (не являются ни параметрами, ни переменными). Алгоритмическая модель процесса будет следующей:

- 1) Максимальный запас = P[1];
- 2) Спрос = P[2];
- 3) Критический запас = 0;
- 4) Время = Начальное время;
- 5) Запас = Максимальный запас;
- 6) Пока (Время \leq Конечное время)
- 7) {
- 8) Запас = Запас – Спрос;
- 9) Если (Запас \leq Критический запас), то
- 10) {
- 11) Создать событие (Время, Максимальный запас – Запас);
- 12) Запас = Максимальный запас;
- 13) }

¹⁾ Тогда под параметром «Максимальный запас» понимается некоторый размах от неизвестного критического уровня до некоторого истинного максимального запаса.

- 14) Продвинуть время и обновить параметры;
- 15) Максимальный запас = $P[1]$;
- 16) Спрос = $P[2]$;
- 17) }

Здесь $P[1]$ и $P[2]$ обозначают искомые параметры процесса, а «Запас» и «Время» как внутренние переменные. Функция «Создать событие» добавляет в выборку событие с заданным временем и значением (объемом покупки/заказа), равным тому, сколько не хватает до максимума. Процедура «Продвинуть время и обновить параметры» продвигает время на один шаг, пусть в данном примере на день, обновление параметров изменяет их значение в соответствии с временем модели. В этом примере логично предположить, что первый параметр $P[1]$ будет стационарным, в этом случае строку 15 можно не выполнять.

Структура параметров процесса и ее представление

Напомню, что динамические параметры процесса как бы задают медленно меняющиеся условия, в которых функционирует процесс, а статические параметры задают неизменяющиеся условия (могут различаться в разных источниках событий, например, у каждого клиента свой собственный максимальный запас). Будем описывать такие параметры в виде непрерывных гладких функций. Может появиться вопрос, а почему именно гладких и непрерывных, неужели параметры не могут изменяться скачками? Напомню, что в самом процессе задается алгоритм изменения внутренних переменных, внутренние переменные могут изменяться скачками, но в соответствии с алгоритмом. Параметры же изменяются сами по себе со временем. Можно предположить, что такие изменения происходят непрерывно¹⁾. Если допустить возможность скачкообразных изменений параметров в произвольный момент времени, то такие параметры не получится восстановить, а наша цель

¹⁾ Корректнее сказать, что на определенном уровне абстракции можно принять эти изменения непрерывными, на определенном масштабе дискретность теряется/сливается/размывается, и ее агрегированное воздействие можно описать непрерывной функцией. Например, на низком уровне можно рассматривать движение молекул, а на более высоком, потоки тепла.

восстановить эти неизвестные параметры процесса. Кстати, параметры в виде непрерывных гладких функций точно восстановить не получится, а только приближенно, и то налагая определенные ограничения на вид функций.

Вариантов представления функций может быть большое множество. Но обычно для больших периодов времени используют сплайны. Чаще всего используют кубические сплайны. То есть форма функции ограничивается полиномами третьей степени, которые состыковываются в точках сочленения (узлах сплайна). В итоге здесь мы также будем определять параметры процесса в виде кубических сплайнов. В главе 3 мы математически восстанавливали такие сплайны по редким событиям (для процессов, схожих с опустошением/наполнением емкости). Тогда сплайн (полином) $g(t)$ между двумя узлами s_k и s_{k+1} задавался через значения сплайна $g(s_k)$, $g(s_{k+1})$ и значение его второй производной в узлах сплайна $g''(s_k)$, $g''(s_{k+1})$. Но для нашего случая, когда мы моделируем функционирование процесса, получается, что придется заранее задать значение $g(s_{k+1})$ и вторую производную $g''(s_{k+1})$ параметра для будущего момента времени (в следующем узле s_{k+1}), что будет не удобно.

Чтобы при определении сплайна не приходилось «залезать в будущее», лучше определить его только через значение $g(s_k)$ в предыдущем узле, и через значение его третьей производной $g'''(s_k)$ в этом же предыдущем узле. Но как вы понимаете, значение в предыдущем узле само также определяется через значение и третью производную в предпредыдущем узле и так далее. Как вы уже могли догадаться, наш сплайн будет немного похож на дифференциальное уравнение. Он будет определяться через начальные значения в стартовом узле, а на всех других узлах будет меняться только его третья производная. Замечу, что такая форма определения сплайна эквивалента представлению через значения и вторые производные (value-second derivative representation), которая в свою очередь эквивалента классическому представлению через 4 параметра при степенях переменной (я провел все соответствующие

аналитические расчеты, чтобы убедиться в этом). Теперь опишем это более формально.

Задавать динамические параметры процесса будем в виде кубических сплайнов $g(t)$, представленных через изменяющуюся скачками третью производную $g'''(s_k)$ в каждом узле s_k . Наш сплайн будет немного похож на дифференциальное уравнение. В итоге, чтобы задать значение изменяющегося со временем параметра процесса в виде сплайна $g(t)$, требуется:

1) задать узлы сплайна $s_1 < s_2 < \dots < s_m$ (точки сочленения полиномов, количество узлов m и их местоположение выбирается исследователем априори, самый распространенный способ выбрать узлы, совпадающими с наблюдениями, или распределить их равномерно на всем интервале наблюдений);

2) задать начальное значение, а также первую и вторую производную в самом первом узле $g(s_1), g'(s_1), g''(s_1)$;

3) задать значения третьей производной $g'''(s_k)$ во всех предшествующих моменту времени t узлах, $\forall k: s_k < t$.

Благодаря такой форме необходимо определять только те значения сплайна, которые предшествовали интересующему нас периоду времени t (не «залезаем в будущее»). Заметим, что в последнем узле s_m значение третьей производной можно не задавать, так как после последнего узла мы не определяем значение сплайна.

Значения сплайна в произвольный момент времени t определяется через значения и производные, соответствующие предшествующему узлу s_k (из исходной формы сплайна последовательными преобразованиями выводится форма, соответствующая ряду Тейлора, что было ожидаемо). Для времени t и предшествующего этому времени узла s_k выполняем:

$$g''(t) = g''(s_k) + (t - s_k)g'''(s_k), \quad (5.16)$$

$$g'(t) = g'(s_k) + (t - s_k)g''(s_k) + \frac{1}{2}(t - s_k)^2 g'''(s_k), \quad (5.17)$$

$$g(t) = g(s_k) + (t - s_k)g'(s_k) + \frac{1}{2}(t - s_k)^2 g''(s_k) + \frac{1}{6}(t - s_k)^3 g'''(s_k). \quad (5.18)$$

Заметим, что данная форма представления сплайна также гарантирует непрерывность вплоть до второй производной, только третья производная меняется скачками в узлах сплайна (третья производная константа между узлами). Также заметим, что в примере процесса образования событий мы вводили функцию «Продвинуть время и обновить параметры», поэтому обновление параметров удобнее реализовать в ней. Причем если время продвигается всего на единицу (например, по дням), то все вычисления еще раз значительно упрощаются, достаточно обновить значения не от предыдущего узла, а от предыдущего момента времени:

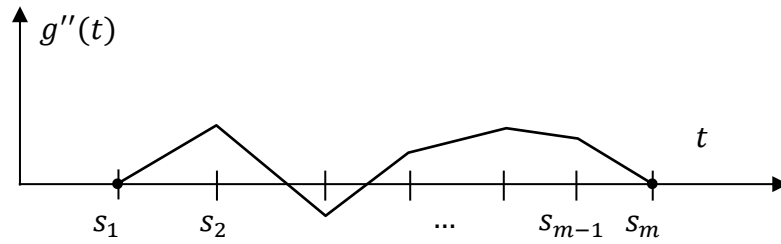
$$g(t) = g(t) + g'(t) + \frac{1}{2}g''(t) + \frac{1}{6}g'''(s_k), \quad (5.19)$$

$$g'(t) = g'(t) + g''(t) + \frac{1}{2}g'''(s_k), \quad (5.20)$$

$$g''(t) = g''(t) + g'''(s_k). \quad (5.21)$$

Дополнительно заметим, что практически во всех работах, посвященных восстановлению неизвестных функций сплайнами, используются натуральные сплайны, которые в самом начале и в самом конце превращаются в прямую линию. То есть вторая производная на концах сплайна обнуляется, в самом начале она берется равной нулю, и к самому окончанию она также должна прийти к нулю. Если мы введем такое условие, то вторую производную в начальном узле можно будет не подбирать. Также можно будет не подбирать значения третьей производной в предпоследнем узле s_{m-1} , так как это значение можно выразить через ранее определенные значения. Если $h_k = s_{k+1} - s_k$ - расстояние между узлами, $g''(s_1) = 0$, то $g''(s_m) = \sum_{k=1}^{m-1} h_k g'''(s_k) = 0$, пояснения на рисунке 5.14, откуда

$$g'''(s_{m-1}) = -\frac{\sum_{k=1}^{m-2} h_k g'''(s_k)}{h_{m-1}}. \quad (5.22)$$



Источник: составлено автором.

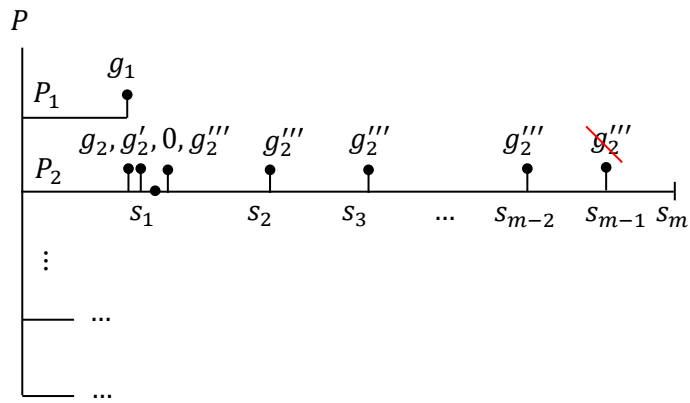
Рисунок 5.14 – Поведение второй производной $g''(t)$ в кубическом натуральном сплайне, третья производная изменяется скачками, вторая кусочно-линейно, вторая производная начинается в нуле и заканчивается в нуле

Следующий момент. Ранее было сказано, что наблюдения могут быть с погрешностью (данные зашумлены). Для борьбы с шумом к функции потерь добавляют штраф на чрезмерную гладкость (шероховатость, roughness). Для кубического сплайна штраф на шероховатость рассчитывается как $\int_{s_1}^{s_m} (g''(t))^2 dt$. В нашей задаче также потребуется рассчитывать этот штраф. Его можно рассчитывать как во время самого функционирования процесса, суммируя квадрат второй производной, так и после завершения процесса. Пропуская промежуточные вычисления (надо интегрировать квадрат $g''(t) = g''(s_k) + (t - s_k)g'''(s_k)$), скажем, что штраф на шероховатость для нашего сплайна можно определить как

$$\int_{s_1}^{s_m} (g''(t))^2 dt = \sum_{k=1}^{m-1} \frac{(g''(s_k) + g'''(s_k)h_k)^3 - (g''(s_k))^3}{3g'''(s_k)}, \quad (5.23)$$

где $g''(s_k) = g'''(s_1)h_1 + \dots + g'''(s_{k-1})h_{k-1}$.

Далее возвращаясь к процессу образования событий, вспоминаем, что параметров может быть несколько. И для каждого динамического параметра необходимо предусмотреть все вышеописанное. Для простоты удобнее всем динамическим параметрам задавать узлы сплайна одинаковым способом. В итоге для модели процесса необходимо предусмотреть следующую схему значений для описания параметров процесса, рисунок 5.15, тут первый параметр статический, а второй динамический.



Источник: составлено автором.

Рисунок 5.15 – Пример схемы значений для описания параметров процесса формирования событий, P_1 – первый параметр (статический), P_2 - второй параметр (динамический)

Для описанного выше примера процесса в алгоритмической форме может использоваться именно такая схема параметров, но только с 2 параметрами P_1, P_2 .

Так как наша задача восстановить параметры событий, то теперь искомыми значениями будут являться значения сплайнов и производных в самом первом узле и значения третьих производных в каждом последующем узле. Определять их будем с помощью соотнесения имеющейся выборки редких событий с выборкой событий, полученной в результате функционирования нашей модели процесса. Для этого задаем функцию потерь и оптимизируем ее численными методами

Функция потерь

В результате функционирования модели процесса с динамически изменяющимися параметрами будут образовываться события. Эти события мы будем сравнивать с имеющейся выборкой редких событий. Обозначим моменты появления событий как t_i и t'_i для свершившихся событий (наблюдений) и для событий, полученных в результате функционирования модели процесса (моделирования). Заметим, что время возникновения событий в экономике как правило измеряется (фиксируется) дискретно с шагом в один день, редко когда время событий фиксируют с точностью до минут/наносекунд. Далее, каждое событие несет некоторую информацию/признак/воздействие (например, объем покупки). Ранее мы

говорили, что ограничимся случаем, когда каждое событие характеризуется только одним скалярным значением, а не целым вектором значений (множеством признаков). Обозначим тогда эти признаки как y_i и y'_i соответственно для исходных наблюдений и для событий, полученных в результате моделирования. Введем функцию потерь, которая будет показывать различия в двух выборках: исходной и сгенерированной. Обычно практически во всех методах определяют сумму квадратов отклонений, реже встречается сумма абсолютных значений (модулей) отклонений. Мы также будем рассчитывать квадраты отклонений. Но так как у нас по каждому событию есть две характеристики, то у нас будет несколько сумм. В итоге можем составить следующую функцию потерь, формула (5.24)

$$S = \sum_{i=2}^n (t_i - t'_i)^2 + \mu \sum_{i=2}^n (y_i - y'_i)^2, \quad (5.24)$$

где нумерация начинается со второго события, так как первое событие служит отправной точкой, с которой начинает функционировать модель процесса; μ – весовой коэффициент, показывающий какую долю внимания мы уделяем отклонениям для моментов времени, а какую для признаков; n – размер выборки. Так как величины t_i и y_i измеряются в разных единицах измерения (например, время может измеряться в днях, и отклонения будут незначительными, а признаки измеряться в миллионах), то логично перейти к безразмерным относительным величинам, формула (5.25)

$$S = \sum_{i=2}^n \left(\frac{t_i - t'_i}{t_i - t_{i-1}} \right)^2 + \mu \sum_{i=2}^n \left(\frac{y_i - y'_i}{y_i} \right)^2, \quad (5.25)$$

где для времени берется отношение отклонения к интервалу времени между возникновением событий, а для значений отношение к наблюдаемому значению (однако в начале мы говорили, что наблюдения фиксируются с погрешностью, поэтому можно делить на y'_i , т.е. будет определяться относительное отклонение наблюдения от прогнозного значения).

Далее. Для того, чтобы восстановить параметры процесса можно было единственным способом, мы вводим штраф на гладкость (шероховатость) для функции каждого параметра $\int_{s_1}^{s_m} (g''(t))^2 dt$, как он рассчитывается было сказано выше. Если этот штраф не вводить, то можно будет бесконечным числом способов задать каждую такую функцию, и каждый такой способ будет правильный (обеспечивать минимум). Поэтому для каждого динамического параметра мы вводим соответствующий штраф с соответствующим весовым коэффициентом, формула (5.26)

$$S = \sum_{i=2}^n \left(\frac{t_i - t'_i}{t_i - t_{i-1}} \right)^2 + \mu \sum_{i=2}^n \left(\frac{y_i - y'_i}{y_i} \right)^2 + \sum_{z=1}^{N_{\text{дин.пар.}}} \alpha_z \int_{s_1}^{s_m} (g''_z(t))^2 dt. \quad (5.26)$$

Здесь z – номер одного из динамических параметров, $N_{\text{дин.пар.}}$ – количество динамических параметров, $g''_z(t)$ – вторая производная от функции (сплайна) динамического параметра, α_z – коэффициент, с которым учитывается данный штраф в общей функции потерь (также известен как параметр сглаживания или коэффициент альфа регуляризации Тихонова). Если у нас в модели процесса только один динамический параметр, то будет участвовать только один такой штраф, а если все параметры статические, то такого штрафа вовсе не будет. Заметим, что данная задача может также интерпретироваться как многокритериальная, когда каждому критерию назначается свой вес. Также обратим внимание на то, что этот штраф интегрирует абсолютные значения квадрата второй производной. Получается немного странная ситуация, когда при сравнении событий используются относительные отклонения, а штраф дает абсолютные значения. Значение для параметра сглаживания в этом случае надо выбирать намного меньше, например 10^{-6} .

Надо обратить внимание еще на один важный момент. Дело в том, что в ходе дальнейшей оптимизации, когда будут в процесс подставляться различные значения параметров, в случае неудачного выбора параметров может быть получена выборка событий, которая по размеру отличается от

ожидаемого количества событий. Например, в выборку вообще не было добавлено ни одного события, или когда мы ожидали получить 3 события, а получили только 2. Забегая вперед, скажем, что мы будем подбирать параметры процесса не сразу во всех узлах, а сначала подбирать его в нескольких начальных узлах, чтобы получить несколько первых событий. Так вот, если выбранные значения параметров не дают нам нужного количества событий, мы можем сразу сказать, что функция потерь обращается в бесконечность (но процесс образования событий нужно моделировать чуть дольше, чем самое последнее наблюдение, чтобы дать шанс сформироваться нужному событию, об этом чуть ниже).

Оптимизация

Для того чтобы определить параметры процесса, мы минимизируем функцию потерь. Процесс запускается с разными значениями параметров, получается выборка событий. Эта выборка сравнивается с имеющейся исходной выборкой наблюдений, в результате такого сравнения получаем числовую оценку. Взять производную от целевой функции не получится, так как ее значение определяется в результате функционирования модели процесса, которая в общем случае может быть произвольным алгоритмом. Аналитические методы и методы, предполагающие аналитическое вычисление значений производных, не годятся. Можно численно определять градиент и идти в направлении этого градиента. Целевая функция каждый раз возвращает скалярное значение, можно по нескольким точкам определить градиент. Но, к сожалению, у нас во многих точках целевая функция будет не существовать (возвращает бесконечность), может быть не гладкой, расчет градиента будет затруднительным.

Для этого случая используют хорошо зарекомендовавший себя алгоритм Нелдера-Мида [140], который действительно является универсальным. В этом методе выбирается на одну точку больше, чем количество переменных, тем самым формируется симплекс. Вершины этого симплекса оцениваются, после чего сортируются. Далее самая плохая вершина

зеркально отражается относительно центра противоположащей грани. Для ускорения схождения используется дополнительное растяжение или сжатие выбранного направления, в направлении которого происходит зеркалирование. Тратить время на более подробное объяснение не будем, подробнее смотрите в литературе, на той же Википедии есть очень доходчивое и наглядное описание. Алгоритм Нелдера-Мида способен двигаться в направлении лучшего значения, даже если среди выбранных вершин симплекса есть вершины, в которых расчет целевой функции невозможен. На одной итерации в большинстве случаев происходит только одно оценивание целевой функции.

Существенной проблемой является наличие множества локальных оптимумов. Существует несколько подходов. Самый простой из них это оптимизация на сетке, когда оптимизация запускается многократно из разных начальных точек, после чего из обнаруженных локальных экстремумов выбирается наилучший. Другой подход – это генетические алгоритмы [141; 142]. В них также надо некоторым образом задать начальную популяцию (множество точек), вероятность отбора особей из популяции при минимизации обратно пропорциональна значению целевой функции, потомки наследуют часть хромосом (параметров) от родителей, дополнительно предусмотрена процедура случайной мутации отдельных генов. Еще один современный подход – это разбросанный поиск (scatter search, поиск по разбросу)¹⁾ [143; 144]. В нем основная идея заключается в том, что используется некоторая мера разброса между точками, например, эвклидово расстояние, которая используется совместно со значениями целевой функции при добавлении этих точек в популяцию. В итоге популяция для дальнейшего поиска представлена точками, которые будут одновременно и обладать хорошим значением целевой функции и при этом хорошо распределены по пространству. И тот и другой алгоритм использует некоторую рандомизацию

¹⁾ На данный момент разбросанный поиск используется в оптимизаторе OptQuest, который встроен во многие современные средства имитационного моделирования, включая Anylogic.

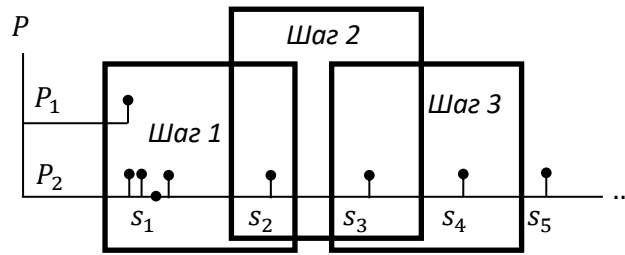
при определении следующих точек. В этих методах, как и в обычном методе оптимизации на сетке, опять происходит локальный поиск в разных стартовых точках, но эти стартовые точки выбираются алгоритмом, а не задаются исследователем. Лично мое мнение, что намного честнее и понятнее для себя использовать обычный поиск на сетке, более того, исходные данные и некоторые представления о модели процесса могут подсказать какой шаг сетки следует выбирать по каждой переменной.

В итоге для определения параметров процесса будет использоваться оптимизация на сетке с помощью метода Нелдера-Мида. Причем, если при такой оптимизации мы покидаем заданную ячейку, то алгоритм прекращается, и происходит начинается уже из следующей ячейки. Возможно, такой подход будет не самым эффективным с точки зрения производительности, и другие исследователи предпочтут другие методы оптимизации, но такой подход справляется со своей задачей.

Важным является выбор шага сетки. Причем для каждого параметра можно задавать свой собственный шаг. В этот момент желательно использовать дополнительную информацию о процессе образования событий. Например, если у нас используется процесс потребления, такой, как описано выше, то для первого параметра (максимального запаса) логично использовать такую сетку, которая бы покрывала значения, равные наблюдаемым объемам покупок (с некоторым запасом). То есть если объемы покупок измеряются в районе 1000, то бессмысленно перебирать значения максимально запаса в области меньше 100, или больше 10000. Аналогичный подход следует применить и для остальных параметров. Но заметим, что для динамических параметров следует подбирать стартовое значение, стартовую скорость, а также третью производную в каждом узле. Естественно, для разных величин надо выбирать разную сетку, для меньших величин определять меньший шаг. Также отметим, что значения производных могут быть отрицательны, а вот начальные значения только положительны (операция вычитания или суммирования задается в самой модели процесса).

В начальном узле по каждому динамическому параметру надо определить 3 значения, а во всех последующих узлах 1 значение. Если каждое значение перебирается на сетке, то получается задача огромной размерности, количество комбинаций значений параметров будет расти экспоненциально. Мы не будем подбирать значения параметров сразу в каждом узле, мы будем подбирать значения для такого количества узлов, чтобы получить всего несколько событий, а не все сразу. Идея заключается в том, чтобы вначале подобрать параметры так, чтобы первые события, образовавшиеся процессом, хорошо совпадали с первыми событиями из исходной выборки. На первом шаге мы подбираем параметры процесса на N событий вперед, для всех таких параметров запускаем оптимизацию на сетке. На следующем шаге мы как бы сдвигаем скользящее окно на одну позицию, на одно событие, и начинаем от него подбирать параметры для N событий. Например, если $N = 2$, то на первом шаге подбираются параметры, отвечающие за образование первого и второго события¹⁾, на втором шаге мы подбираем параметры, отвечающие за образование второго и третьего, подбираем их опять на сетке. При этом в целевой функции учитываем не вошедшие в скользящее окно события, а параметры, отвечающие за образование не вошедших событий, также используем в оптимизации (не на сетке), предполагаем, что их определенные на предыдущем шаге значения лежат в нужном нам локальном оптимуме. На рисунке 5.16 изображено какие параметры перебираются на сетке на каждом новом шаге для случая $N = 2$ (при условии, что узлы сетки s_k совпадают с наблюдениями).

¹⁾ Т.к. время первого события является отправной точкой, то первое образованное событие на самом деле будет вторым событием.



Источник: составлено автором.

Рисунок 5.16 – Подбор параметров на $N = 2$ событий вперед, на первом шаге полностью перебираются на сетке комбинации 5 параметров, на втором - комбинации 2 параметров, на третьем - тоже комбинации 2 параметров

Обратим внимание, что на первом шаге на вход оптимизатора подается вектор из 5 значений (у параметра P_2 третье значение всегда 0), в качестве начальных точек берется полная комбинация всех возможных значений параметров. На втором шаге на вход оптимизатора подается 6 значений, но количество комбинаций ограничено только комбинациями последних двух значений (в узлах s_2 и s_3 берется полная комбинация возможных значений, но на вход оптимизатора каждый раз подаются все значения, включая значения в узле s_1). На третьем шаге на вход оптимизатора подается 7 значений, но комбинации определяются перебором на сетке только значений в узлах s_3 и s_4 . Можно сформулировать следующее предположение, которое тут используется:

Параметры процесса, соответствующие событию, зависят незначительно от событий, расположенных в более чем N позициях от текущего события.

То есть ранее запущенный алгоритм оптимизации нашел такой экстремум, который при добавлении новых событий и соответствующих им значений параметров может по-прежнему оказаться глобальным экстремумом при оптимизации на сетке новых добавленных значений. Это предположение может быть верно, так как восстанавливаемый динамический параметр (как функция) на одном участке может практически не зависеть от поведения этого же параметра на другом сильно удаленном участке. В то же время подавать на вход оптимизатора ранее найденные параметры для их последующей корректировки необходимо, иначе возможен тот самый «эффект бабочки», и

малые расхождения в самом начале приведут к катастрофическим последствиям в конце.

Желательно выбирать N (ширину окна событий) как можно больше, в идеале делать полный перебор всех комбинаций значений параметров для всей выборки сразу, но это очень сильно влияет на объем вычислений. Нужно найти некоторый баланс между временем выполнения всех расчетов и точностью определяемых параметров. Время выполнения самого первого шага может быть самым большим, так как на нем определяется больше всего параметров, время второго шага может быть наименьшим, далее с ростом номера шага время будет линейно увеличиваться, т.к. процесс будет запускаться на время пока не образуется нужное количество событий, плюс количество самих параметров, подаваемых на оптимизатор будет линейно возрастать. Тем не менее, с использованием описанного предположения получается обойти проблему не полиномиального роста объема вычислений.

Пример определения параметров процесса, заданного в алгоритмической форме

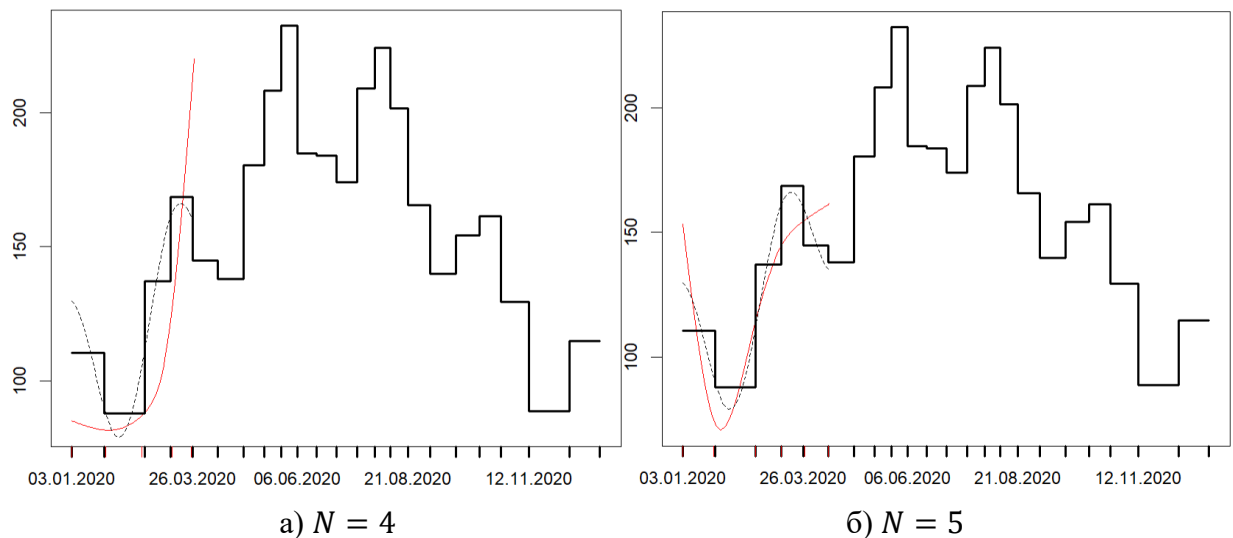
Рассмотрим работу метода на примере процесса потребления, описанного выше с помощью алгоритмической модели. В качестве начальных данных возьмем данные, образованные этим же процессом с некоторыми предустановленными параметрами (предполагаем их неизвестными). Нашей целью будет восстановить эти значения параметров, располагая только выборкой событий, таблица 5.5.

Таблица 5.5 – Данные редких событий

t_i	y_i	t_i	y_i	t_i	y_i	t_i	y_i
03.01.2020	2429	30.04.2020	2526	17.07.2020	2508	09.10.2020	2418
25.01.2020	2461	14.05.2020	2500	29.07.2020	2466	24.10.2020	2457
22.02.2020	2465	26.05.2020	2556	09.08.2020	2418	12.11.2020	2478
11.03.2020	2528	06.06.2020	2402	21.08.2020	2484	10.12.2020	2409
26.03.2020	2462	19.06.2020	2574	05.09.2020	2517	31.12.2020	2478
12.04.2020	2481	03.07.2020	2437	23.09.2020	2467	-	-

Источник: составлено автором.

В качестве целевой функции используем функцию потерь с относительными отклонениями (5.26). Пусть для нас определение даты событий t_i имеет большее значение, чем характеристика этого события y_i (объем покупки), возьмем $\mu = 0.1$ (вес суммы квадратов относительных отклонений y_i). Для штрафа на нелинейность возьмем коэффициент сглаживания $\alpha = 10^{-6}$. Прежде чем запускать алгоритм, надо определиться на сколько событий вперед N следует подбирать параметры. Если подбирать параметры на 4 события вперед, то первый шаг получается таким, как на рисунке 5.17(a), а если на 5 событий вперед, как на рисунке 5.18(б), таблица 5.6 и 5.7 показывает числовые результаты первого шага.



Гладкая сплошная линия показывает восстановленное значение скорости потребления $g_2(t)$ (шт./день), пунктирная линия показывает искомую функцию потребления, заложенную в модель, ступенчатая линия показывает среднее потребление $y_i/(t_{i+1} - t_i)$, площадь под каждой ступенькой будет объемом покупки y_i .

Источник: составлено автором.

Рисунок 5.17 – Первый шаг алгоритма определения параметров

Таблица 5.6 – Значения параметров после первого шага алгоритма

Параметр	Значение	
	$N = 4$	$N = 5$
Максимальный запас (истинное значение 2400)	$g_1 = 1992,141$	$g_1 = 2371,628$
Начальная скорость потребления	$g_2(s_1) = 85,084$	$g_2(s_1) = 153,222$
Первая производная скорости потребления	$\dot{g}_2(s_1) = -0,21103$	$\dot{g}_2(s_1) = -4,8323$
Третья производная скорости потребления в узле s_1	$\ddot{g}_2(s_1) = -6,5058 \cdot 10^{-4}$	$\ddot{g}_2(s_1) = 1,4825 \cdot 10^{-2}$
Третья производная скорости потребления в узле s_2	$\ddot{g}_2(s_2) = -6,5419 \cdot 10^{-4}$	$\ddot{g}_2(s_2) = -1,4202 \cdot 10^{-2}$
Третья производная скорости потребления в узле s_3	$\ddot{g}_2(s_3) = 1,8618 \cdot 10^{-2}$	$\ddot{g}_2(s_3) = 5,6984 \cdot 10^{-5}$
Третья производная скорости потребления в узле s_4	$\ddot{g}_2(s_4)$ выражается через предыдущие значения	$\ddot{g}_2(s_4) = 4,0556 \cdot 10^{-3}$
Третья производная скорости потребления в узле s_5	-	$\ddot{g}_2(s_5)$ выражается через предыдущие значения

Источник: составлено автором.

Таблица 5.7 – Полученные события после первого шага алгоритма

Модель	Дата					
	03.01.2020	25.01.2020	22.02.2020	11.03.2020	26.03.2020	12.04.2020
Исходные	03.01.2020	25.01.2020	22.02.2020	11.03.2020	26.03.2020	12.04.2020
$N = 4$	-	26.01.2020	20.02.2020	12.03.2020	25.03.2020	-
$N = 5$	-	24.01.2020	22.02.2020	11.03.2020	27.03.2020	12.04.2020
Модель	Значения (объем покупки)					
Исходные	2429	2461	2465	2528	2462	2481
$N = 4$	-	1992,141	2079,106	2116,739	2169,692	-
$N = 5$	-	2386,468	2485,890	2378,407	2413,642	2532,705

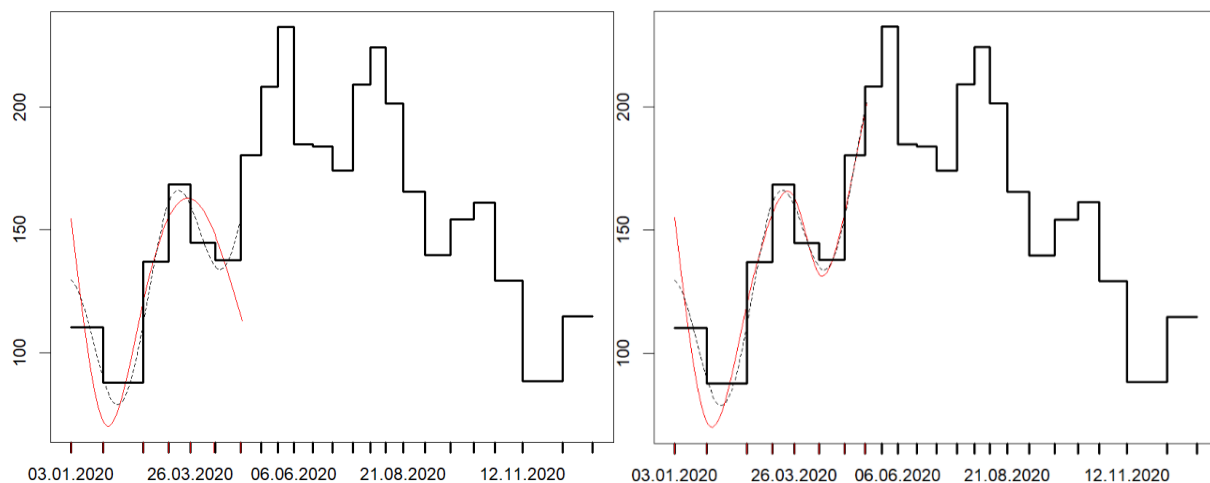
Источник: составлено автором.

Так как алгоритм опирается только на значения целевой функции, рассчитываемые по отклонениям полученных событий от выборки наблюдений, то в первом случае ($N = 4$) алгоритм подобрал максимальный запас равным 1992, что позволило с заниженным потреблением сформировать события, расположенные близко к имеющимся наблюдениям. Из-за неправильно подобранного значения максимального запаса объемы покупок в этом случае оказались значительно ниже. Однако так как мы измеряем квадрат относительных отклонений, а не абсолютных, да еще используем вес $\mu = 0.1$, это оказывает незначительное влияние на функцию потерь. То есть по первым 4 событиям, глядя только на функцию потерь, не получилось

корректно подобрать значения параметров. Ситуация немного улучшится, если продолжить выполнять алгоритм на следующих шагах (с тем же значением $N = 4$), но когда скользящее окно сдвинется на одно событие, параметры в первом узле могут оказаться в локальном минимуме, а не глобальном (они перестают перебираться на сетке). Картина заметно улучшается, если определять параметры процесса сразу на $N = 5$ событий вперед. Пятое событие вносит нужную информацию, позволяющую точнее определить начальные параметры и значения производных в узлах. Далее результаты идут для значения $N = 5$.

На втором шаге сдвигается скользящее окно событий вправо, и для $N = 5$ событий осуществляется подбор параметров, не забывая о значениях параметров, определенных для самого первого сформированного события. Значения параметров, определенные для первого события, подставляются в оптимизатор как есть, а для значений параметров, необходимых для образования следующих 5 событий, берется множество возможных комбинаций значений, оптимизация запускается для каждой такой комбинации. Это позволяет кардинально пересмотреть ранее полученные значения параметров, попавших в скользящее окно (со второго по пятое событие). Значения параметров для первого образованного события также изменяются при оптимизации, но уже только относительно полученных на первом шаге значений. Так значение максимального запаса изменилось с ранее определенного 2371.628 на 2410.712. На третьем и последующем шаге скользящее окно вновь сдвигается на одну позицию, и попавшие в это скользящее окно значения перебираются на сетке, а не попавшие лишь корректируются. Алгоритм останавливается после 18 шага, когда будут подобраны все значения параметров, необходимые для формирования всех 22 событий. Результат работы алгоритма после шага 2 и 3 изображен на рисунке 5.18, а за все 18 шагов на рисунке 5.19. Таблица 5.8 показывает события, образованные в результате функционирования процесса, с подобранными параметрами. Только в 1 из 22 событий дата события отличается на 1 день.

Значения событий (объемы покупок) отличаются в большинстве событий меньше 1%, только в последних событиях на 2-3%.

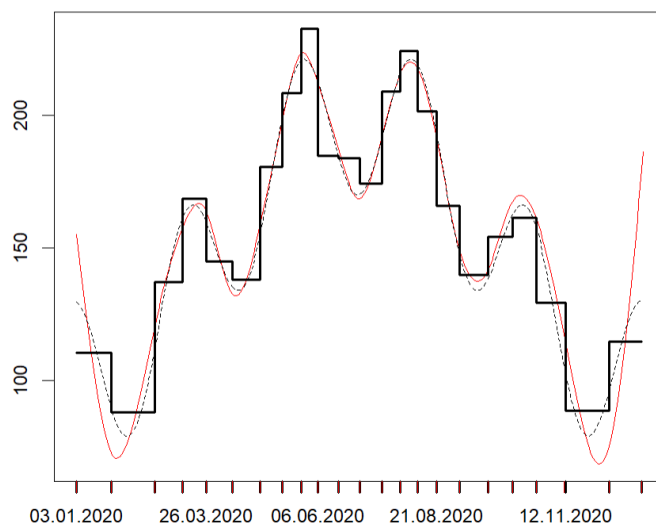


Шаг 2

Шаг 3

Источник: составлено автором.

Рисунок 5.18 – Результат работы алгоритма после 2 и 3 шага.



Источник: составлено автором.

Рисунок 5.19 – Результат работы алгоритма после всех 18 шагов

Таблица 5.8 – Полученные события после 18 шага алгоритма

-	Дата					
Исходные	Полное совпадение в 21 из 22 событий, только для события с датой 12.11.2020, процесс определил событие на день раньше (11.11.2020)					
Процесс						
-	Значения (объем покупки)					
Исходные	2429	2461	2465	2528	2462	2481
Процесс	-	2468.564	2472.071	2547.122	2464.828	2494.400
Исходные	2526	2500	2556	2402	2574	2437
Процесс	2545,614	2511,836	2566,986	2411,377	2585,020	2445,382
Исходные	2508	2466	2418	2484	2517	2467
Процесс	2498,655	2470,175	2406,726	2462,082	2515,078	2523,396
Исходные	2418	2457	2478	2409	2478	-
Процесс	2489,236	2509,046	2522,387	2406,723	2575,255	-

Источник: составлено автором.

В результате применения описанного метода получилось очень хорошо определить параметры процесса, который был задан алгоритмической моделью, один параметр алгоритма был статический, а второй динамическим. Напомню, что мы определяем не собственную скорость расхода продукции, а ненаблюдаемую функцию скорости расхода продукции неподконтрольного нам клиента. Ранее в примерах главы 2 математическим методом также получалось восстановить скорость потребления, а остальные параметры процесса получалось определить эмпирическими методами, но сейчас мы определили все параметры процесса в рамках одного метода.

Стоит обратить внимание на то, что благодаря использованию скользящего окна событий, у нас получилось разбить одну большую оптимизационную задачу огромной размерности на небольшое количество задач значительно меньшей размерности. Если бы мы подбирали значения параметров абсолютно для всех событий из выборки, то трудоемкость задачи была бы астрономической, мы не смогли бы за приемлемое время осуществить необходимые вычисления.

Программная реализация на языке R опубликована в [117].

5.4 Разработка метода исследования неизвестных событий

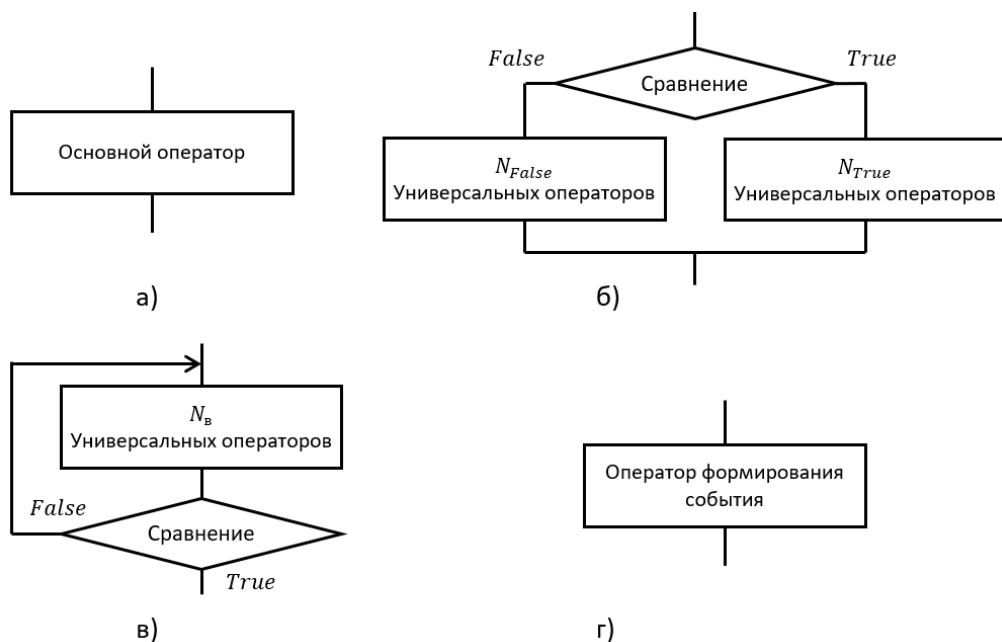
В предыдущем параграфе мы восстанавливали параметры процесса образования событий, заданного исследователем в виде алгоритмической модели. Можно пойти еще дальше и дополнительно предположить, что сам процесс является совсем неизвестным, и требуется подобрать не только параметры процесса, но и сам процесс. То есть если в предыдущем параграфе мы решали оптимизационную задачу $S(P) \rightarrow \min$, то теперь предстоит подобрать не только параметры P , но и саму модель процесса образования событий *Program*, формула (5.27)

$$S(\text{Program}, P) \rightarrow \min. \quad (5.27)$$

Основная идея заключается в том, что, введя ограничения на количество операторов, количество внутренних переменных и параметров, мы будем перебирать возможные схемы операторов модели процесса *Program*.

После чего для каждой такой модели происходит восстановление параметров процесса P . Среди всех моделей $Program$ будет определяться лучшая.

Механизм перебора алгоритмов модели процесса $Program$ образования событий следующий. Основным элементом будет являться универсальный оператор, который в свою очередь может оказаться одним из 4 видов: основной оператор, оператор ветвления, оператор цикла и оператор формирования события. Причем внутри оператора ветвления и цикла могут располагаться другие универсальные операторы, которые опять могут быть 4 видов, рисунок 5.20.



а) основной оператор; б) оператор ветвления; в) оператор цикла; г) оператор формирования события.

Источник: составлено автором.

Рисунок 5.20 – Структура универсального оператора

Основной оператор лишь выполняет одну из арифметических операций (+, -, *, /, =, где = означает отсутствие операции как таковой) над внутренними переменными и параметрами процесса и сохраняет результат в одну из внутренних переменных. Внутри оператора ветвления может быть N_{true} и N_{false} универсальных операторов, а внутри оператора цикла $N_{в}$ универсальных операторов, которые могут быть как обычными основными операторами, так и вновь операторами ветвления или операторами цикла. При этом в них осуществляется операция сравнения. Ключевым элементом

процесса является оператор формирования события. Этот оператор добавляет запись о дискретном событии в выборку дискретных событий. Оператор формирования события очень похож на обычный основной оператор, в нем выполняется одна из арифметических операций (+, -, *, /) над внутренними переменными и параметрами, но вместо сохранения результата во внутренние переменные этот результат используется для формирования события (возвращается момент времени t_i и значение/результат события y_i).

Задав ограничение на количество основных операторов N_{opers} и количество сравнений N_{if} , которые используются в ветвлениях и циклах, рекурсивным образом перебираются все возможные схемы процесса. Далее для каждой фиксированной схемы, задав ограничения на количество внутренних переменных N_x и параметров N_p , перебираются все возможные комбинации арифметических операций и операндов.

Отдельно надо сказать пару слов про оператор формирования события, предполагается, что событие может формироваться только внутри оператора ветвления. То есть внутри цикла не может быть оператора формирования события, только если внутри цикла будет другой оператор ветвления, и только в нем оператор формирования события. В то же время можно допустить, что события формируются в нескольких местах алгоритма, тогда надо задать ограничение на количество операторов формирования событий N_{Events} . Получающиеся алгоритмы, перебирающие все возможные комбинации схем процесса и комбинации операций внутри каждой схемы, являются достаточно сложными для составления рекурсивными алгоритмами, однако сами алгоритмы оказываются достаточно быстродействующими.

Реализация на языке R и подробное описание алгоритма автоматического перебора модели процесса образования событий *Program* опубликованы в работе [118] (из-за ограничения объема нет возможности привести подробное описание в данной диссертационной работе).

В результате появляется возможность автоматически перебирать всевозможные процессы образования событий, заданные алгоритмическими

моделями. Количество схем процесса получается не таким большим, но количество комбинаций операторов в каждой схеме большое. Например, при ограничении в 3 основных оператора и 1 сравнение – всего 30 комбинаций схем (причем без циклов, так как одно сравнение нужно для генерации события). При 3 основных операторах и 2 сравнениях - уже 450 комбинаций схем. При 4 основных операторах и 1 сравнении – всего 70 комбинаций схем, при 4 основных операторах и 2 сравнениях – уже 1638 комбинаций схем. При 5 основных операторах и 1 сравнении – всего 140 комбинаций схем, при 5 основных операторах и 2 сравнениях – уже 4761 комбинаций схем. Это лишь количество комбинаций схем. Но для каждой схемы надо перебрать комбинации операторов. Количество комбинаций операторов зависит от того, из каких операций составлена эта схема, а также от заданного количества внутренних переменных и количества параметров процесса. Причем это количество будет очень большим. Например, для схемы из 3 основных операторов и 1 сравнения, если количество переменных $N_x = 2$ и количество параметров тоже $N_p = 2$, получается уже 10.892.880 комбинаций операторов/операндов. При большем количестве внутренних переменных и параметров количество комбинаций становится астрономическим, но все же конечным числом. Поэтому остается гипотетическая возможность перебрать все возможные комбинации. К тому же исследование каждой комбинации схемы и операторов поддается распараллеливанию.

Есть некоторые соображения, как можно уменьшить количество комбинаций операторов внутри каждой схемы. Большая часть различающихся комбинаций может не приводить к изменению генерируемых событий. Например, если процесс образования событий задан 3 операторами, но только 2 оператора участвуют в изменении переменных, отвечающих за момент времени или возвращаемое значение событий, а третий оператор делает бесполезную ни на что не влияющую операцию, то этот оператор можно изменять многими способами (лишь бы они тоже не влияли на получаемые процессом события). Тем самым количество комбинаций растет, но, по сути,

процесс образования событий остается одним и тем же. К сожалению, этот вопрос еще не исследован, но есть ощущение, что можно многократно уменьшить количество перебираемых комбинаций, если пропускать комбинации операций, которые совпадают с ранее исследованными.

На само исследование каждой модели процесса может требоваться разное время. Так если модель процесса не позволяет создать ни одно событие, то оптимизация прекращается сразу же. Также процесс определения параметров завершается раньше, если погрешность превосходит определенную заданную величину. Например, иногда (при неудачно выбранных моделях процесса) за 5 минут получается исследовать ~730 комбинаций операторов (~2.5 комбинаций в секунду), но если модель процесса выбрана удачно, то поиск значений параметров этого процесса занимает около 10 секунд (в случае, если параметры статичные, если параметры динамические, то время может быть намного больше). Конечно же это все еще очень большое время, если требуется проверить все возможные модели процессов образования событий. Но как было сказано выше, все вычисления можно легко распараллелить (на данный момент распараллеливание вычислений происходит только при оптимизации во время подбора параметров). Тем самым существует принципиальная возможность исследовать совершенно неизвестные события с помощью подбора алгоритмической модели процесса образования событий. Заметим, что при таком подходе мы не строим модель процесса, опираясь на некоторую известную информацию, а алгоритм за нас сам подберет наилучшую модель процесса, после чего мы можем заняться интерпретацией этой модели. Однако такой подход требует дальнейших исследований. В одиночку боюсь не получится справиться, нужно подключение широкого круга исследователей.

Выводы по главе 5

1) Произведено обобщение емкостного метода для анализа исторических, социальных и политических событий. Рассмотрены примеры цветных революций и русско-турецких войн. В этом случае емкостью является

«чаша терпения», которая наполняется или опустошается под воздействием потоков возмущений, и при превышении порога возмущения образуется событие.

2) Разработан математический метод восстановления функции сразу по разным функционалам, таким как значение функции, ее первой и второй производной, а также значениям определенного интеграла, который основан на сплайновой коллокации со штрафной функцией. Данный метод может использоваться для восстановления параметров процессов образования событий, которые несут преимущественно качественную информацию. Для данного математического метода восстановления функции по разным функционалам получена формула для расчета оценки-кросс валидации, минимизация которой может дать оптимальный параметр сглаживания.

3) Разработан универсальный метод анализа и прогнозирования редких событий, в котором процесс образования событий может быть задан произвольным алгоритмом, у которого могут быть динамические параметры. Восстановление динамических параметров такого процесса осуществляется с помощью методов оптимизации черного ящика с использованием специального приема, который ограничивает размерность получаемой задачи.

4) Сформулирован метод исследования совершенно неизвестных событий на основе ранее разработанного универсального метода анализа и прогнозирования редких событий. В данном методе предполагается подбирать не только параметры процесса, но также и сам процесс образования событий в виде алгоритма, составленного из определенных операторов.

Заключение

В данной диссертационной работе предложен новый подход для анализа и прогнозирования редких событий в экономике, основная идея которого заключается в рассмотрении событий с точки зрения процессов, происходящих в источниках этих событий, в результате функционирования которых как раз образуются эти события. Метод состоит из 5 шагов: 1) разделение событий в зависимости от источников, в которых они образованы;

2) выдвижение предположения о процессе образования событий в этом источнике; 3) восстановление (регрессия) параметров процессов по выборке редких событий; 4) экстраполяция значений параметров процесса на будущее; 5) запуск самого процесса с установленными значениями параметров и получение прогноза будущих событий.

В качестве частного случая процесса образования событий в данной работе преимущественно рассматриваются процессы, которые можно моделировать как процессы опустошения/наполнения некоторой емкости. В результате данного исследования был разработан «ёмкостный» метод анализа и прогнозирования редких событий в экономике. В качестве примеров рассматриваются примеры процесса потребления, когда событиями являются покупки у неподконтрольных нам клиентов, например, как в моделях управления запасами. В этом случае получается прогнозировать будущие покупки этих клиентов. Хочется обратить внимание на то, что клиенты могут даже не догадываться, что их очередной раз «посчитали» и для них могут быть составлены соответствующие модели прогнозирования будущих покупок. Это позволяет применять определенные маркетинговые меры, чтобы клиент совершал покупки у «правильных» продавцов. Также можно более аккуратно планировать уровень запасов, подготавливая нужный объем к заданной дате будущих покупок клиентов (не договариваясь с ним). Стоит заметить, что никакие другие методы анализа и прогнозирования редких событий не способны дать прогноз конкретной даты будущих покупок и их объемов, а лишь вероятности возникновения события за интервал времени или ожидаемого количества событий за интервал времени. Помимо процессов потребления также можно рассматривать процессы накопления возмущения до некоторого критического уровня, после которого происходит образование события. Рассматриваются примеры из сферы услуг, например, стрижка волос, маникюр и др.

Разработан математический метод восстановления параметров данных процессов образования событий, который основан на сплайновой коллокации.

В данном методе неизвестная функция восстанавливается в виде кубического сплайна по значениям наблюдаемого с погрешностью функционала, которым является определенный интеграл от функции. В пятой главе данный метод обобщается, в результате функционалами также могут выступать сами значения, значения первых и вторых производных от функции, все функционалы также могут наблюдаться с погрешностью. На сплайн налагается штраф на гладкость (шероховатость), который используется для регуляризации параметров. Для этого метода проведено исследование по выбору оптимального коэффициента сглаживания с помощью метода L -кривой, кросс-валидации и невязки Морозова, даны рекомендации по выбору параметра сглаживания. Для данного метода разработан программный инструментарий на языке R. Так как данный метод используется для восстановления параметров процесса образования событий, то восстанавливаемая функция сама по себе может иметь интересную интерпретацию и представлять теоретический интерес. Например, для сферы услуг парикмахерской можно получить зависимость скорости роста волос от времени, что должно быть интересно в соответствующих областях исследования физиологии человека. В других областях восстановление параметров процессов образования событий может иметь свое собственное практическое и теоретическое значение.

Проведено исследование различных вопросов точности. Выявлены дополнительные причины возникновения погрешности в самих наблюдениях. Получены формулы для расчета этой погрешности. Исследована точность восстановления исходной зависимости от погрешности в наблюдениях, в зависимости от неопределенности внутри самого процесса формирования событий. Получен вывод, что все неопределенности внутри процесса образования событий можно свести к погрешности наблюдений. Построена зависимость погрешности восстановления исходной функции от погрешности наблюдений. Благодаря проведенным исследованиям вопросов точности восстановления исходной функции был предложен подход проверки

адекватности получающихся моделей прогнозирования на основе построения доверительного интервала для момента времени прогнозируемого события. Рассчитаны формулы для построения соответствующего доверительного интервала.

Самым многообещающим результатом работы является обобщение полученных идей на произвольные процессы образования событий, которые можно задавать произвольными алгоритмами. Разработан метод определения динамически изменяющихся параметров у таких алгоритмических моделей образования событий. Данный подход открывает новое направление, когда исследователи также, как исследователи в эконометрике составляют структурные модели, будут формировать алгоритмические модели образования редких событий, восстанавливать их параметры в виде динамических функций, экстраполировать их на будущее и давать прогноз будущим событиям. Под это направление исследований можно сформировать новое направление обучения студентов. Вместе с тем появляется возможность исследования совершенно неизвестных событий с помощью алгоритмических моделей образования событий, где сами эти модели перебираются и выискивается наилучшая, а подобрав алгоритмическую модель можно в дальнейшем давать интерпретацию получающимся закономерностям.

В целом поставленные в диссертационном исследовании задачи были решены полностью, а исследование получилось комплексным. Открывшееся новое направление исследований конечно же является источником новых вопросов и задач, которые не представляется возможным рассмотреть в рамках одной работы и над которыми предстоит работать многим исследователям.

Список литературы

1. Hasofer, A.M. Exact and invariant second-moment code format / A.M. Hasofer, L. Niels // *Journal of Engineering Mechanics*. – 1974. – № 1. Volume 100. – P. 111–121. – ISSN 0733-9399.
2. Kiureghian, A. Second-order reliability approximations / A. Kiureghian, H. Lin, S.J. Hwang // *Journal of Engineering Mechanics*. – 1987. – № 8. Volume 113. – P. 1208–1225. – ISSN 0733-9399.
3. Schuëller, G.I. A critical appraisal of reliability estimation procedures for high dimensions / G.I. Schuëller, J. Pradlwarter, P.S. Koutsourela // *Probabilistic Engineering Mechanics*. – 2004. – № 4. Volume 19. – P. 463-474. – ISSN 0266-8920.
4. Angelis, M. Advanced line sampling for efficient robust reliability analysis / M. Angelis, E. Patelli, M. Beer // *Structural Safety*. – 2015. – № 1. Volume 52. – P. 170–182. – ISSN 0167-4730.
5. Виноградов, В.В. История слов: около 1500 слов и выражений и более 5000 слов, с ними связанных / В. В. Виноградов; Российская академия наук. Отделение литературы и языков. Научный совет "Русского языка". Институт русского языка им. В. В. Виноградова. – Москва, 1999. – 1138 с. – ISBN 5-88744-033-3.
6. Толковый словарь Даля : [сайт] / В.И. Даль 2008-2022. – Текст : электронный. – URL: <http://slovardalja.net/> (дата обращения: 19.01.2021).
7. Талев, Н.Н. Черный лебедь. Под знаком непредсказуемости / Н. Талев ; перевод с английского В. Сонькина [и др.]. – Москва : КоЛибри, 2010. – 525 с. : ил. – ISBN 978-5-389-00573-0.
8. Шеннон, К. Э. Работы по теории информации и кибернетике : сборник статей ; перевод с английского / С предисловием А. Н. Колмогорова; под редакцией Р. Л. Добрушина и О. Б. Лупанова. – Москва : Издательство иностранной литературы, 1963. – 829 с. – ISBN отсутствует.
9. Carreno, A. Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework / A. Carreno, I. Inza, J.A. Lozano // *Artificial Intelligence Review*. – 2020. – Volume 53. – P. 3575-3594. – ISSN 0269-2821. – DOI 10.1007/s10462-019-09771-y.
10. Halim, S.Z. Time-trend analysis of offshore fire incidents using nonhomogeneous Poisson process through Bayesian inference / S.Z. Halim, N. Quaddus, H. Pasman // *Process Safety and Environmental Protection*. – 2021. – № 147. – P. 421-429. – ISSN 1744-3598. – DOI 10.1016/j.psep.2020.09.049.
11. Makhduomi, H. A Comparative Study of First-Order Reliability Method-Based Steepest Descent Search Directions for Reliability Analysis of Steel Structures / H. Makhduomi, B. Keshtegar, M. Shahraki // *Advances in Civil*

Engineering. – 2017. – Volume 2017. – P. 1-10. – ISSN 1687-8086. DOI 10.1155/2017/8643801.

12. Kiureghian, A.D. Efficient algorithm for second-order reliability analysis/ A.D. Kiureghian, M.D. Stefano // Journal of Engineering Mechanics. – 1991. – № 12. Volume 117. – P. 2904–2923. – ISSN 1943-7889. – DOI 10.1061/(ASCE)0733-9399(1991)117:12(2904).

13. Song, J. Active learning line sampling for rare event analysis / J. Song, P. Wei, M. Valbenito, M. Beer // Mechanical Systems and Signal Processing. – 2021. – Volume 147. – P. 107113. – ISSN 0888-3270. – DOI 10.1016/j.ymssp.2020.107113.

14. Papaioannou, I. Combination line sampling for structural reliability analysis / I. Papaioannou, D. Straub // Structural Safety. – 2021. – № 3. Volume 88. – P. 102025. – ISSN 0167-4730. – DOI 10.1016/j.strusafe.2020.102025.

15. Воробьева, А.В. Модели оценки и управления рисками ипотечного кредитования: специальность 08.00.13 «Математические и инструментальные методы экономики»: диссертация на соискание ученой степени кандидата экономических наук / Воробьева Анна Владимировна; РЭУ имени Г.В. Плеханова. – Москва, 2018. – 162 с. – Библиогр.: с. 149-158.

16. Обзор ключевых показателей деятельности страховщиков (за 2017) : / Банк России. – 2018. – Текст : электронный. – URL: http://www.cbr.ru/collection/collection/file/24028/review_insure_17q4.pdf (дата обращения: 25.11.2022).

17. Показатели состояния безопасности дорожного движения : [сайт] / Госавтоинспекция. – 2019. – Текст : электронный. – URL: <http://stat.gibdd.ru/> (дата обращения: 11.01.2019).

18. Единая государственная информационная система учета результатов научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения (ЕГИСУ НИОКТР) : [сайт] / ФГАНУ ЦИТиС. – Текст : электронный. – URL: <https://rosrid.ru/> (дата обращения: 11.01.2019).

19. Kaya, G. O. Intermittent demand forecasting: a guideline for method selection / G. O. Kaya, M. Sahin, O. Demirel // Sadhana - Academy Proceedings in Engineering Sciences. – 2020. – № 1. Volume 45. – P. 45-51. – ISSN 0973-7677. – DOI 10.1007/s12046-020-1285-8.

20. Pince, C. Intermittent demand forecasting for spare parts: A Critical review / C. Pince, L. Turrini, J. Meissner // Omega (United Kingdom). – 2021. – Volume 105. – P. 102513. – ISSN 0305-0483. – DOI 10.1016/j.omega.2021.102513.

21. Croston, J.D. Forecasting and stock control for intermittent demands / J. D. Croston // Journal of the Operational Research Society. – 1972. – № 3. Volume 23. – P. 289-303. – ISSN 1476-9360. – DOI 10.1057/jors.1972.50.

22. Qingzheng, X. A Review of Croston's method for intermittent demand forecasting / X. Qingzheng, W. Na, S. Heping // Proceedings - 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2012. – 2012. – P. 1468-1472. – DOI 10.1109/FSKD.2012.6234258.

23. Efron, B. Bootstrap methods: another look at jackknife / B. Efron // The Annals of Statistics. – 1979. – № 1. Volume 7. – P. 1-26. – ISSN 0090-5364.

24. Efron, B. A leisurely look at the bootstrap, the jackknife, and cross-validation / B. Efron, G. Gong // American Statistician. – 1983. – № 1. Volume 37. – P. 36-48. – ISSN 1537-2731.

25. Efron, B. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy / B. Efron, R. Tibshirani // Statistical Science. – 1986. – № 1. Volume 1. – P. 54-75. – ISSN 0883-4237. – DOI 10.1214/SS/1177013815.

26. Efron, B. An Introduction to the Bootstrap / B. Efron, R.J. Tibshirani. – New York : Chapman & Hall, 1993. – 456 p. – ISBN 0-412-04231-2.

27. Willemain, T.R. A new approach to forecasting intermittent demand for service parts inventories / T.R. Willemain, C.N. Smart, H.F. Schwarz // International Journal of Forecasting. – 2004. – № 3. Volume 20. – P. 375-387. – ISSN 0169-2070. – DOI 10.1016/S0169-2070(03)00013-X

28. Willemain, T.R. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method / T.R. Willemain, C.N. Smart, J.H. Shockor, P.A. DeSautels // International Journal of Forecasting. – 1994. – № 10. – P. 529-538. – ISSN 0169-2070. – DOI 10.1016/0169-2070(94)90021-3.

29. Cover, T. Nearest neighbor pattern classification/ T. Cover, P. Hart // IEEE Transactions on Information Theory. – 1967. – № 1. Volume 13. – P. 21-27. – ISSN 1557-9654.

30. Fix, E. Discriminatory Analysis. Nonparametric Discrimination / E. Fix, J.L.Jr. Hodges, University of California; Berkley // USAF School of Aviation Medicine; Randolph Field; Texas; Project 21-49-004; Report 4; Contact AF41(128)-31. – California : Berkley, 1951. – ISBN отсутствует.

31. Pekalska, E. Who invented the nearest neighbor rule? / E. Pekalska, B. Duin, (Robert P.W.) // 37 Steps. Understanding, teaching and developing tools for pattern recognition : [сайт]. – 2014. – Текст : электронный. – URL: <https://37steps.com/4370/nn-rule-invention/> (дата обращения: 25.11.2022).

32. Pelillo, M. Alhazen and the nearest neighbor rule / M. Pelillo // Pattern Recognition Letters. – 2014. – Volume 38. – P. 34-37. – ISSN 0167-8655.

33. Lindberg, D.C. Theories of Vision from Al-Kindi to Kepler / D.C. Lindberg – Chicago : The University of Chicago Press, 1976. – 324. – ISBN 0226482340, 9780226482347.

34. Singh, S. Forecasting using Fuzzy Nearest Neighbor Method / S. Singh // Proc. 6th International Conference on Fuzzy Theory and Technology, Fourth Joint Conference on Information Sciences (JCIS'98) (23-28 October, 1998). – North Carolina, 1998. – Volume 1. – P. 80-83. – ISBN отсутствует.
35. Hosmer, D.W. Applied logistic regression / D.W. Hosmer, S. Lemeshow – Chichester : Wiley, 2000. – 2nd ed. – ISBN 9781118548394.
36. Goodfellow, I. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – MIT Press, 2016. – ISBN 0262035618. – Текст : электронный. – URL: <http://www.deeplearningbook.org> (дата обращения: 25.11.2022).
37. Hastie, T. The Elements of Statistical Learning. Data Mining, Inference, and Prediction / T. Hastie, R. Tibshirani, J. Friedman. – 2nd ed. – New York : Springer-Verlag, 2009. – ISBN 978-0-387-84857-0. – DOI 10.1007/978-0-387-84858-7.
38. Kashnitsky, Y. Открытый курс машинного обучения. Тема 4. Линейные модели классификации и регрессии / Y. Kashnitsky // Хабр : [сайт]. – 2017. – Текст : электронный. – URL: <https://habr.com/ru/company/ods/blog/323890/> (дата обращения: 25.11.2022).
39. Cramer, J.S. The origins of logistic regression / J.S. Cramer // Tinbergen Institute Working Paper. – 2002. – № 119/4. – Текст : электронный. – URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=360300 (дата обращения: 25.11.2022). – DOI10.2139/ssrn.360300.
40. Berkson, J. Application of the Logistic Function to Bio-assay / J. Berkson // Journal of the American Statistical Association. – 1944. – № 227. Volume 9. – P. 357-365. – ISSN 0162-1459.
41. Bliss C.I. The method of probits / C.I. Bliss // Science. – 1934. – № 2037. Volume 79. – P. 38-39. – ISSN 0036-8075. – DOI10.1126/science.79.2037.38.
42. Bliss, C. I. The method of probits – A correction / C.I. Bliss // Science. – 1934. – № 2053. Volume 79. – P. 409–410. – ISSN 0036-8075. – DOI 10.1126/science.79.2053.409.
43. Vannucci, M. Classification of unbalanced datasets and detection of rare events in industry: Issues and solutions / M. Vannucci, V. Colla // Communications in Computer and Information Science. – 2016. – № 629. – P. 337-351. – ISSN 1865-0929.
44. Choe, W. Detection of rare events by neural networks / W. Choe, O.K. Ersoy, M. Blina // Intelligent Engineering Systems Through Artificial Neural Networks. – 1998. – pp. 5-10. – ISSN отсутствует.
45. Kaiser, L. Learning to remember rare events / L. Kaiser, A. Roy, O. Nachum, S. Bengio // Conference Track Proceedings : 5th International

Conference on Learning Representations (ICLR) 2017 (France, Toulon, April 24-26, 2017). – France : Toulon, 2017. – ISBN отсутствует.

46. Вентцель, Е. С. Теория случайных процессов и ее инженерные приложения / Е. С. Вентцель, Л. А. Овчаров : учебное пособие для вузов. – 2-е издание стереотипное. – Москва : Высшая школа, 2000. – 383 с.: ил. – ISBN 5-06-003831-9

47. Лукинский, В. Методы управления запасами: расчет показателей запаса для товарных групп, относящихся к редким событиям (часть I) / В. Лукинский, Д. Замалетдинова // Логистика. – 2015. – № 1. Том 98. – С. 28-33. – ISSN 2219-7222.

48. Лукинский, В. Методы управления запасами: расчет показателей запаса для товарных групп, относящихся к редким событиям (часть II) / В. Лукинский, Д. Замалетдинова // Логистика. – 2015. – № 2. Том 99. – С. 24-27. – ISSN 2219-7222.

49. Вожжов, А.П. Формирование страховых запасов предприятия при пуассоновском характере поступающих и выдаваемых потоков / А.П. Вожжов, О.В. Луняков, С.П. Вожжов // Экономика и управление: теория и практика. – 2015. – № 1. Том 1. – С. 30-35. – ISSN 2412-8376

50. Henderson, D.J. Applied Nonparametric Econometrics / D.J. Henderson, C.F. Parmeter. – New York : Cambridge University Press, 2015. – 367 p. – ISBN 978-1-107-01025-3.

51. Scott, D.W. Multivariate Density Estimation: Theory, Practice, and Visualization. – 2nd ed. – New Jersey : Wiley, 2015. – 384 p. – ISBN 978-0-471-69755-8.

52. Li, Q. Nonparametric Econometrics: Theory and Practice / Q. Li, J.S. Racine. – Princeton University Press, 2007. – 768 p. – ISBN 978-0-691-12161-1.

53. Syntetos, A.A. Forecasting of compound Erlang demand / A.A. Syntetos, M.Z. Babai, S. Luo // Journal of the Operational Research Society. – 2015. – № 12. Volume 66. – P. 2061-2074. – ISBN 0160-5682.

54. Shale, E.A. Forecasting for intermittent demand: the estimation of an unbiased average / E.A. Shale, J.E. Boylan, F.R. Johnston // Journal of the Operational Research Society. – 2006. – № 57. – P. 588-592. – ISBN 0160-5682.

55. Иванько, Р.С. Краткосрочное прогнозирование нестационарного спроса в оптовой торговле: специальность 08.00.12 «Бухгалтерский учет, статистика»: диссертация на соискание ученой степени кандидата экономических наук / Иванько Роман Сергеевич ; МЭСИ. – Москва, 2005. – 143 с. – Библиогр.: с. 116-125.

56. Кораблев, Ю.А. Альтернативное определение спроса / Ю.А. Кораблев // Россия 2030 глазами молодых ученых : материалы

III Всероссийской научной конференции (Москва, 26 апреля 2012 г.). – Москва. 2012. – С. 137-140. – ISBN 978-5-91290-192-8.

57. Кораблев, Ю.А. Емкостный метод определения функции скорости потребления / Ю.А. Кораблев // Экономика и менеджмент систем управления. – 2015. – № 1.1. Том 15. – С. 140-150. – ISSN 2223-0432.

58. Кораблев, Ю.А. Емкостный метод анализа редких событий в экономике / Ю.А. Кораблев // Экономика: теория и практика. – 2016. – № 4 (44). – С. 59-64. – ISSN 2224-042X.

59. Кораблев, Ю.А. Емкостный метод анализа редких продаж в Excel / Ю.А. Кораблев // Экономика и управление: проблемы, решения. – 2017. – № 6. Том 3 (66). – С. 224-229. – ISSN 2227-3891.

60. Кораблев, Ю.А. Анализ и прогнозирование редких событий в экономике / Ю.А. Кораблев // Современная математика и концепции инновационного математического образования : труды VI Международной научно-практической конференции (23 мая 2019). – Москва: Издательский дом МФО, 2019. – С. 123-132. – ISSN 2412-9895.

61. Кораблев, Ю.А. Емкостный метод анализа редких событий в торговле различными товарами / Ю.А. Кораблев, П.С. Голованова, Т.А. Кострица // Бизнес. Образование. Право. Вестник Волгоградского института бизнеса. – 2019. – № 3 (48). – С. 121-131. – ISSN 1990-536X.

62. Кораблев, Ю.А. Емкостный метод анализа редких событий в сфере услуг / Ю.А. Кораблев, П.С. Голованова, Т.А. Кострица // Экономическая наука современной России. – 2020. – № 3 (90). – С. 132-142. – ISSN 1609-1442.

63. Кораблев, Ю.А. Емкостный метод анализа редких событий, допущения при использовании в сфере услуг / Ю.А. Кораблев, П.С. Голованова, Т.А. Кострица // Современная экономика: проблемы и решения. – 2019. – № 11 (119). – С. 22-32. – ISSN 2078-9017.

64. Кораблев, Ю.А. Обоснование емкостного метода определения спроса / Ю.А. Кораблев // Экономика, статистика и информатика. Вестник УМО. – 2015. – № 5. – С. 100-104. – ISSN 1994-7844.

65. Бауэрсок, Д.Д. Логистика: интегрированная цепь поставок / Д.Д. Бауэрсок, Д.Д. Клосс. – 2-е издание. – Москва : ЗАО "Олимп-Бизнес", 2008. – 640 с.: ил. с. – ISBN 978-5-9693-0124-5.

66. R Core Team. R: A Language and Environment for Statistical Computing : Vienna, Austria: 2020. – Текст : электронный. – URL: <https://www.r-project.org/> (дата обращения: 16.11.2020).

67. ALGLIB. Least squares fitting (linear/nonlinear) / Cross-platform numerical analysis and data processing library ALGLIB : [сайт]. – Текст : электронный. – URL: https://www.alglib.net/interpolation/least_squares.php (дата обращения: 16.11.2020).

68. diakin. Практическое применение преобразования Фурье для анализа сигналов. Введение для начинающих. // Хабр : [сайт]. –2015. – Текст : электронный. – URL: <https://habr.com/ru/post/269991/> (дата обращения: 17.11.2020).
69. Quinn, B.G. A Fast Efficient Technique for the Estimation of Frequency / B.G. Quinn, J.M. Fernandes // *Biometrika*. – 1991. – № 3. Volume 78. – P. 489-497. – ISSN 0006-3444.
70. Quinn, B.G. The Estimation and Tracking of Frequency / B.G. Quinn, E.J. Hannan. – Cambridge: Cambridge University Press, 2001. – 278 pp. – ISBN 0-521-80446-9.
71. Vladimir. Экстраполяция цен методом Фурье - индикатор для MetaTrader 5/Vladimir, gpwr//MQL5 community. – 2010. – Текст : электронный. – URL: <https://www.mql5.com/ru/code/130> (дата обращения: 17.11.2020).
72. С какой скоростью растут волосы на голове // Здоровье женских и мужских волос : [сайт]. – Текст : электронный. – URL: <https://volosylady.ru/uhod/skorost-rosta-volos-na-golove> (дата обращения: 24.11.2020).
73. Скорость роста волос: что влияет и как ускорить // Эксперт по волосам : [сайт]. – Текст : электронный. – URL: <https://expertpovolosam.com/uhod/rost/skorost> (дата обращения: 24.11.2020).
74. Погода в Красногорске // Погодный сервис World Weather : [сайт]. – Текст : электронный. – URL: <https://world-weather.ru/pogoda/russia/krasnogorsk/> (дата обращения: 24.11.2020).
75. Беккерман, Е.Н. Алгоритм оценивания числа состояний и значений интенсивностей МС-потока событий / Е.Н. Беккерман, С.С. Катаева // *Вестник ТГУ*. – 2009. – № 3 (8). – С. 19-26. – ISSN 1998-8605.
76. Водинчар, Г.М. Свойства одного алгоритма оценки параметров нестационарных Пуассоновских процессов / Г.М. Водинчар // *Вестник КРАУНЦ. Физ.-мат. науки*. – 2012. – № 1 (4). – С. 11-17. – ISSN 2079-6641.
77. Резник, А.Л. Об оценке интенсивности потоков при регистрации отсчетов / А.Л. Резник // *Автометрия*. – 1980. – № 3. – С. 123-126. – ISSN 0320-7102.
78. Суворова, А.В. Построение доверительных интервалов оценок интенсивности рискованного поведения на основе неравенства Чебышева / А.В. Суворова, А.Е. Пащенко, Т.В. Тулупьева, А.Л. Тулупьев // *Труды СПИИРАН*. – 2009. – № 10. – С. 96-109. – ISSN 2078-9181.
79. Кораблев, Ю.А. Метод восстановления функции по интегралам для анализа и прогнозирования редких событий в экономике / Ю.А. Кораблев // *Экономика и математические методы*. – 2020. – № 3. Том 56. – С. 113-124. – ISSN 0424-7388.

80. Кораблев, Ю.А. Программный продукт, реализующий метод восстановления функции по последовательности интегралов, которые наблюдаются с погрешностью, основанный на интегральных сплайнах со штрафной функцией : Свидетельство о государственной регистрации программы для ЭВМ. – № 2020617973, 15.07.2020. – Заявка № 2020617018 от 06.07.2020.

81. Кораблев, Ю.А. Емкостный метод анализа редких событий в экономике, интегральный сплайн в Excel / Ю.А. Кораблев // Научный журнал «Мягкие измерения и вычисления». – 2020. – № 10. Том 35. – С. 32-45. – ISSN 2618-9976.

82. Кораблев, Ю.А. Кросс-валидация для интегрального сплайна, восстанавливающего скорость потребления по редким событиям / Ю.А. Кораблев // Системный анализ в экономике – 2020 : сборник трудов VI Международной научно-практической конференции-биеннале ; под общей редакцией Г.Б. Клейнера, С.Е. Щепетовой. – Москва : Издательский дом «Наука», 2021. – С. 410-413. – ISBN 978-5-6046256-0-6.

83. Korablev, Yu.A. Restoration of function by integrals with cubic integral smoothing spline in R = Восстановление функции по интегралам с помощью кубического интегрального сглаживающего сплайна в R / Yu. A. Korablev // ACM Transactions on Mathematical Software. – 2022. – № 2. Volume 48. – P. 1-17. – ISSN 0098-3500. – DOI 10.1145/3519384.

84. Korablev, Yu.A. Restoration of the product consumption rate with integral cubic smoothing spline, study of the best smoothing parameter choice = Восстановление скорости потребления продукции с помощью интегрального кубического сглаживающего сплайна, исследование выбора лучшего параметра сглаживания / Yu. A. Korablev // Acta Applicandae Mathematicae. – 2022. – № 180 (8). – P. 1-30. – ISSN 1572-9036.

85. Киреев, В.И. Интегральный метод приближения функций алгебраическими многочленами и биквадратными сплайнами / В.И. Киреев // Вестник Московского авиационного института. – 1994. – № 1. Том 1. – С. 48–58. – ISSN 0869-6101.

86. Киреев, В.И. Полиномиальные интегродифференциальные одномерные и двумерные сплайны / В.И. Киреев, Т.К. Бирюкова // Вычислительные технологии. – 1998. – № 3. Том 3. – С. 19–34. – ISSN 1560-7534.

87. Киреев, В.И. Интегродифференциальный метод обработки информации и его применение в численном анализе/ В.И. Киреев, Т.К. Бирюкова. – Москва : ИПИ РАН, 2014. – 267 с. – ISBN 978-5-91993-040-2.

88. Бирюкова, Т.К. Методы численного дифференцирования и восстановления сеточных функций по интегралам, основанные на

интегродифференциальных сплайнах / Т.К. Бирюкова, В.И. Киреев, М.М. Гершкович // Системы компьютерной математики и их приложения : материалы XVII Международной научной конференции. – Смоленск, 2016. – С. 106–112.

89. Dagnino, C. Spline quasi-interpolating projectors for the solution of nonlinear integral equations / C. Dagnino, A. Dallefrate, S. Remogna // *J. Comput. Appl. Math.* – 2019. – Volume 354. – P. 360–372. – ISSN 0377-0427.

90. Самойлова, Э.Н. Сплайновые приближения решения сингулярного интегродифференциального уравнения / Э.Н. Самойлова // *Известия высших учебных заведений. Математика.* – 2001. – № 11 (474). – С. 35-45. – ISSN 0021-3446.

91. Zemlyanova, A.Y. A new B-spline collocation method for singular integro-differential equations of higher orders / A.Y. Zemlyanova, A. Machina // *J. Comput. Appl. Math.* – 2020. – Volume 380. – P. 112949. – ISSN 0377-0427.

92. Boor, C.D. *A Practical Guide to Splines* / C.D. Boor. – revised ed. – Berlin: Springer Verlag, 2001. – 366 p. – ISBN 0-387-95366-3.

93. Федорова, О.П. Об одном подходе к приближению функции сплайнами / О.П. Федорова // *Вестник Томского государственного университета. Математика и механика.* – 2008. – № 2 (3). – С. 61–66. – ISSN 1998-8621.

94. Федорова, О.П. Метод построения сплайна, сохраняющего интеграл функции двух переменных по области ее задания / О.П. Федорова // *Научный альманах.* – 2016. – № 1–3 (15). – С. 31–35. – eISSN 2411-7609.

95. Kimeldorf, G. Some results on Tchebychefian spline functions / G. Kimeldorf, G. Wahba // *Journal of mathematical analysis and applications.* – 1971. – Volume 33. – P. 82-95. – ISSN 0022-247X. – DOI 10.1016/0022-247X(71)90184-3

96. Green, P.J. *Nonparametric regression and generalized linear models. A roughness penalty approach* / P.J. Green, B.W. Silverman. – New York : Chapman & Hall, 1994. – 194 p. – ISBN 0-412-30040-0.

97. Тихонов, А.Н. *Методы решения некорректных задач* / А.Н. Тихонов, В.Я. Арсенин. – 2-е издание. – Москва : Наука, 1979. – 285 с. – ISBN отсутствует.

98. Hansen, P. C. Analysis of Discrete Ill-Posed Problems by Means of the L-Curve / P.C. Hansen // *SIAM Rev.* – 1992. – № 4. Volume 34. – P. 561–580. – ISSN 0036-1445.

99. Hansen, P.C. The L-Curve and its Use in the Numerical Treatment of Inverse Problems / P.C. Hansen // In book: *Computational Inverse Problems in Electrocardiology*; edited by: P.R. Johnston, Griffith University, Australia – Southampton : WIT Press, 2001. – ISBN 978-1-85312-614-7. – P. 119–142.

100. Craven, P. Smoothing noisy data with spline functions - Estimating the correct degree of smoothing by the method of generalized cross-validation / P. Craven, G. Wahba // *Numerische Mathematik*. – 1978. – № 31. Volume 4. – P. 377-403. ISSN 0029-599X.

101. Morozov, A. Methods for solving incorrectly posed problems / A. Morozov // New York : Springer-Verlag, 1984. – 257 p. – ISBN: 978-0-387-96059-3.

102. Mueller, J.J. Linear and Nonlinear Inverse Problems with Practical Applications / J.J. Mueller, S. Siltanen. – Philadelphia : SIAM, 2012. – 338 p. – ISBN 978-1-61197-233-7.

103. Reinsch, Ch. H. Smoothing by spline function / Ch. H. Reinsch // *Numerische Mathematik*. – 1967. – № 10. – P. 177–183. – ISSN 0029-599X.

104. Кораблев, Ю.А. Исследование точности определения скорости потребления продукции с помощью емкостного метода при прохождении товара через цепочку распространителей / Ю.А. Кораблев // *Современные проблемы моделирования социально-экономических систем : материалы VIII международной научно-практической конференции (1-10 апреля 2016 г.)*. – Бердянск : Ткачук А.В., 2016. – С. 17-21. – ISBN 978-617-7291-55-7.

105. Кораблев, Ю.А. Разбор причин и оценка погрешности аномальных картин в емкостном методе анализа редких событий / Ю.А. Кораблев // *Экономика и управление: проблемы, решения*. – 2017. – № 8. Том 6 (68). – С. 8-12. – ISSN 2227-3891.

106. Кораблев, Ю.А. Информационная экономика: этапы развития, методы управления, модели (Пункт 3.6. Влияние позиции в цепочке распространителей на точность емкостного метода) : монография / Ю.А. Кораблев ; под редакцией д.э.н, профессора В.С. Пономаренко, д.э.н., профессора Т.С. Клебановой. – Харьков : ВШЭМ – ХНЭУ имени С. Кузнеца, 2018. – С. 502-523. – 668 с. – ISBN 978-80-89654-45-1.

107. Кораблев, Ю.А. Исследование точности емкостного метода от позиции в цепочке распространителей / Ю.А. Кораблев // *Экономика и управление: проблемы, решения*. – 2018. – № 5. Том 7 (77). – С. 106-121. – ISSN 2227-3891.

108. Кораблев, Ю.А. Влияние позиции в цепочке распространителей на точность емкостного метода / Ю.А. Кораблев // *Научный журнал «Мягкие измерения и вычисления»*. – 2018. – № 10. Том 11. – С. 47-62. – ISSN 2618-9976.

109. Кораблев, Ю.А. Дисперсия емкостного метода от позиции в цепочке распространителей / Ю.А. Кораблев // *Системный анализ в экономике – 2018 : сборник трудов V Международной научно-практической конференции – биеннале ; под общей редакцией Г.Б. Клейнера,*

С.Е. Щепетовой. – Москва : Прометей, 2018. – С. 197-200. – ISBN 978-5-907100-80-0.

110. Кораблев, Ю.А. Погрешность емкостного метода анализа редких событий, удаленность от конечного потребителя / Ю.А. Кораблев // Известия Кабардино-Балкарского научного центра РАН. – 2019. – № 3 (89). – С. 48-77. – ISSN 1991-6639.

111. Кораблев, Ю.А. Емкостный метод анализа редких событий, оценка погрешности вследствие конкуренции или потери данных / Ю.А. Кораблев // Современная экономика: проблемы и решения. – 2019. – № 10 (118). – С. 18–31. – ISSN 2078-9017.

112. Кораблев, Ю.А. Исследование точности емкостного метода анализа редких событий от неопределенности внутри процесса образования событий (Часть 1) / Ю.А. Кораблев // Известия Кабардино-Балкарского научного центра РАН. – 2020. – № 5 (97). – С. 49-67. – ISSN 1991-6639.

113. Кораблев, Ю.А. Использование емкостного метода для анализа социальных событий / Ю.А. Кораблев, П.С. Голованова, Т.А. Кострица // Научный журнал «Мягкие измерения и вычисления». – 2021. – № 6. Том 43. – С. 81-92. – ISSN 2618-9976.

114. Кораблев, Ю.А. Использование емкостного метода для анализа исторических событий / Ю.А. Кораблев, П.С. Голованова, Т.А. Кострица // KANT. – 2021. – № 1 (38). – С. 27-32. – ISSN 2222-243X.

115. Кораблев, Ю.А. Об одном алгоритме восстановления функции по разным функционалам для прогнозирования редких событий в экономике / Ю.А. Кораблев // Финансы: теория и практика / Finance: Theory and Practice. – 2022. – № 3. Том 26. – С. 196-225. – ISSN 2587-5671.

116. Кораблев, Ю.А. Определение параметров процесса образования редких событий в экономике для их последующего прогнозирования / Ю.А. Кораблев // Экономика и математические методы. – 2022. – № 2. Том 58. – С. 80-91. – ISSN 0424-7388.

117. Кораблев, Ю.А. Восстановление параметров процесса образования событий в экономике, заданного алгоритмической моделью / Ю.А. Кораблев // Известия Кабардино-Балкарского научного центра РАН. – 2022. – № 4 (108). – С. 96–114. – ISSN 1991-6639.

118. Кораблев, Ю.А. Алгоритм автоматического перебора произвольных моделей процесса образования редких событий в экономике / Ю.А. Кораблев // Научный журнал «Мягкие измерения и вычисления». – 2022. – № 9. Том 58. – С. 62-88. – ISSN 2618-9976.

119. Цветные революции // Русский эксперт — патриотическая энциклопедия о России и мире : [сайт]. – Текст : электронный. – URL: https://ruxpert.ru/Цветные_революции (дата обращения: 01.08.2021).

120. Искандарян, А. Кавказ-2003. Выборный год. Ежегодник КИСМИ / А. Искандарян – Ереван : Кавказский институт СМИ, 2005. – 228 с. – ISBN 99930-78-73-5

121. Верч, Д. Грузия после "революции роз" / Д. Верч // Кавказ и глобализация. – 2006. – № 1. Том 1. – С. 62-77. – ISSN 1817-7100.

122. Гахокидзе, Э. Грузия: политические партии до и после "революции роз" / Э. Гахокидзе // Центральная Азия и Кавказ. – 2005. – № 2 (38). – С. 66-73. – ISSN 1403-7068.

123. Котляревский, Ю.Л. Оранжевая революция. Глазами консультанта / Ю.Л. Котляревский – Ростов-на-Дону : Феникс, 2005. – 123 с. – ISBN 5-222-06685-1.

124. Затулина, К.Ф. На фоне оранжевой революции. Украина между Востоком и Западом: вчера. Сегодня. Завтра / Институт стран СНГ (Институт диаспоры и интеграции ; [под редакцией К.Ф. Затулина]. – Москва : Институт стран СНГ, 2005. – 238 с. – ISBN 5-7853-0477-5.

125. Наумова, А.Ю. «Цветные революции» на постсоветском пространстве / А.Ю. Наумова, В. Авдеев, А. Наумов. – Санкт-Петербург : Алетейя, 2013. – 162 с. – ISBN 978-5-91419-896-8.

126. Князев, А. Государственный переворот 25 марта 2005 г. в Киргизии / А. Князев – 3-е издание – Бишкек : Общественный Фонд А. Князева, 2007. – 272 с. – ISBN 5-8355-1455-7.

127. Костенко, Д. Партизанская республика. Акции протеста в Минске 19-25 марта 2006 года / Д. Костенко, Фокс, А. Кашко [и др.]; под редакцией Е. Михайловской. – Москва : РОО «Центр «Панорама», 2006. – 219 с. – ISBN 5-94420-023-5.

128. Республика Беларусь накануне президентских выборов 2006 г. Серия информационно-аналитических записок / Исследовательская группа ЦИРКОН. – Москва : ЦИРКОН, 2006. – 69 с. – ISBN отсутствует. – URL: <http://www.zircon.ru/upload/File/russian/publication/1/060303.pdf> (дата обращения: 28.11.2022)

129. Democracy on rocky ground: Armenia's disputed 2008 presidential election, post-election violence, and the one-sided pursuit of accountability / Human Rights Watch. – New York, HRW, 2009 – 65 p. – ISBN 978-1-56432-444-3.

130. 7 апреля — 29 июля 2009 года: государственный переворот в Молдове // интернет-портал AVA : [сайт]. – 2012. – Текст : электронный. – URL: <https://ava.md/2012/04/07/7-aprelya-29-iyulya-2009-goda-gosudarstvennyy/> (дата обращения: 01.08.2021).

131. Чернявский, С. Киргизская революция 2010 года: причины и перспективы постреволюционного развития Киргизии / С. Чернявский //

Центральная Азия и Кавказ. – 2010. – № 2. Том 13. – С. 44-53. – ISSN 1403-7068.

132. Акции протеста в России после выборов 4 декабря 2011 года. // РИА Новости : [сайт]. – 2013. – Текст : электронный. – URL: <https://ria.ru/20130506/933060828.html> (дата обращения: 01.08.2021).

133. Протестное движение в России 2011-2012 годов: проблема субъектности / Д.Г. Зайцев, В.Е. Карастелев // Государство и общество в пространстве власти и политических коммуникаций. Политическая наука. Ежегодник 2013 ; под редакцией Соловьева А.И. – Москва : РОССПЭН, 2013. – С. 231-266. – ISBN 978-5-8243-1773-2.

134. Пихорович, В.Д. Украина между Западом и Востоком: Война на Донбассе / В.Д. Пихорович. – Москва : URSS ЛЕНАНД, 2015. – 201 с. – ISBN 978-5-9710-1582-6.

135. Итоги 2014: Украина пережила самый тяжелый год за время независимости // РИА Новости : [сайт]. – 2014. – Текст : электронный. – URL: <https://ria.ru/20141231/1040986726.html> (дата обращения: 01.08.2021).

136. Протесты в Армении (2015) // Википедия. Свободная энциклопедия : [сайт]. – Текст : электронный. – URL: [https://ru.wikipedia.org/wiki/Протесты_в_Армении_\(2015\)](https://ru.wikipedia.org/wiki/Протесты_в_Армении_(2015)) (дата обращения: 02.08.2021).

137. Захват заложников в Ереване (2016). // Википедия. Свободная энциклопедия : [сайт]. – Текст : электронный. – URL: [https://ru.wikipedia.org/wiki/Захват_заложников_в_Ереване_\(2016\)](https://ru.wikipedia.org/wiki/Захват_заложников_в_Ереване_(2016)) (дата обращения: 02.08.2021).

138. Протесты в Белоруссии (2020—2021) // Википедия. Свободная энциклопедия : [сайт]. – Текст : электронный. – URL: [https://ru.wikipedia.org/wiki/Протесты_в_Белоруссии_\(2020—2021\)](https://ru.wikipedia.org/wiki/Протесты_в_Белоруссии_(2020—2021)) (дата обращения: 02.08.2021).

139. Русско-турецкие войны // Википедия. Свободная энциклопедия : [сайт]. – Текст : электронный. – URL: https://ru.wikipedia.org/wiki/Русско-турецкие_войны (дата обращения: 14.09.2022).

140. Nelder, J.A. A Simplex Method for Function Minimization / J. A. Nelder, R. Mead // The Computer Journal. – 1965. – № 4. Volume 7. – P. 308-313. – ISSN 0010-4620/ – DOI <https://doi.org/10.1093/comjnl/7.4.308>

141. Goldberg, D.E. Genetic algorithms in search, optimization, and machine learning / D.E. Goldberg. – Addison-Wesley, 1989. – 432 pp. – ISBN 0-201-15767-5.

142. Саймон, Д. Алгоритмы эволюционной оптимизации / Д. Саймон; перевод с английского А.В. Логунова. – Москва : ДМК Пресс, 2020. – 1002 с. – ISBN 978-5-97060-707-7.

143. Laguna, M. Scatter Search – Methodology and Implementations in C / M. Laguna, R. Marti. – Boston : Kluwer Academic Publishers, 2003. – ISBN 978-1-4020-7376-2.

144. Laguna, M. Scatter Search / M. Laguna, R. Marti // Metaheuristic Procedures for Training Neural Networks ; ed. E. Alba, R. Marti. – New York : Springer, 2006.– ISBN 978-0-387-33415-8. – P. 139-152.

145. Кораблев, Ю.А. Восстановление потребности в заемных средствах с помощью емкостного метода анализа редких событий / Ю.А. Кораблев // Научный журнал «Мягкие измерения и вычисления». – 2019. – № 6. Том 19. – С. 22-37. – ISSN 2618-9976.

146. Большой энциклопедический словарь / главный редактор А. М. Прохоров. – Москва : Советская энциклопедия; Санкт-Петербург : Фонд «Ленингр. галерея», 2002. – 1628 с. – ISBN 5-85270-015-0.

147. Nadaraya, E.A. On Estimating Regression / E.A. Nadaraya // Теор. Вероятност. i Применен. – 1964. – № 1 (9). – 157-159. – ISSN 0040-361X; Theory of Probability and its Applications. – 1964. – № 1 (9). – P. 141–142. – ISSN 0040-585X.

148. Watson, G.S. Smooth Regression Analysis / G.S. Watson // Sankhyā: The Indian Journal of Statistics, Series A. – 1964. – № 4. Volume 26.– P. 359–372. – ISSN 0581-572x.

Приложение А

(информационное)

Алгоритм Куина и Фернандеса на языке R

Адаптированная под язык R версия алгоритма Куина и Фернандеса (Quinn-Fernandes algorithm) для разложения на гармоники и экстраполяции значений ряда.

Исходная версия алгоритма для системы MetaTrader 5 выложена пользователем Vladivir (за авторством некоторого gpwr).

```
s="f_in.csv";
MyData <- read.csv(file=s, header=TRUE, sep=";", dec=",", stringsAsFactors = FALSE)
#dim(MyData)
#x=MyData[[1]]
y=MyData[[2]] #when data have 2 columns

#+-----+
#| Fourier_Interpolator_of_Price.mq5          Copyright 2010, gpwr |
#|      Extrapolation by trigonometric (multitone) model"      |
#+-----+

Npast=length(y) ; # Past values, to which trigonometric series is fitted
Nfut =365; # Predicted future values
Nharm =10; # Harmonics in model
FreqTOL =0.000001; # Tolerance of frequency calculations

#---- Initialization
N=max(Npast,Nfut+1);
av=sum(y[1:Npast])/Npast;
xm=rep(av,N);
ym=rep(av,Nfut);

layout(matrix(c(1,2), 2, 1, byrow = TRUE))
plot(1:(Npast+Nfut),c(y[1:Npast],rep(av,Nfut)), type='l')

#+-----+
#| Quinn and Fernandes algorithm for finding frequency      |
#+-----+
Freq=function(x,n)
{
  z=rep(0,n);#
  alpha=0.0;#
  beta=2.0;#
  z[1]=x[1]-xm[1];
  while(abs(alpha-beta)>FreqTOL)
  {
    alpha=beta;
    z[2]=x[2]-xm[2]+alpha*z[1];
```

```

num=z[1]*z[2]; #
den=z[1]*z[1]; #
for(i in 3:n)
{
  z[i]=x[i]-xm[i]+alpha*z[i-1]-z[i-2];
  num=num+z[i-1]*(z[i]+z[i-2]);
  den=den+z[i-1]*z[i-1];
}
beta=num/den;
}
w=acos(beta/2.0);
r=TrigFit(x,n,w);

return(c(w,r))
}

#+-----+
#| Least-squares fitting of trigonometric series |
#+-----+
TrigFit=function(x,n,w)
{
  Sc =0.0;
  Ss =0.0;
  Scc=0.0;
  Sss=0.0;
  Scs=0.0;
  Sx =0.0;
  Sxc=0.0;
  Sxs=0.0;
  for(i in 1:n)
  {
    c=cos(w*i);
    s=sin(w*i);
    dx=x[i]-xm[i];
    Sc=Sc+c;
    Ss=Ss+s;
    Scc=Scc+c*c;
    Sss=Sss+s*s;
    Scs=Scs+c*s;
    Sx=Sx+dx;
    Sxc=Sxc+dx*c;
    Sxs=Sxs+dx*s;
  }
  Sc=Sc/n;
  Ss=Ss/n;
  Scc=Scc/n;
  Sss=Sss/n;
  Scs=Scs/n;
  Sx=Sx/n;
  Sxc=Sxc/n;
  Sxs=Sxs/n;
}

```

```

if(w==0.0)
{
  m=Sx;
  a=0.0;
  b=0.0;
}
else
{
  # calculating a, b, and m
  den=(Scs-Sc*Ss)^2-(Scx-Sc*Sc)*(Sss-Ss*Ss);
  a=((Sxs-Sx*Ss)*(Scs-Sc*Ss)-(Scx-Sx*Sc)*(Sss-Ss*Ss))/den;
  b=((Scx-Sx*Sc)*(Scs-Sc*Ss)-(Sxs-Sx*Ss)*(Scx-Sc*Sc))/den;
  m=Sx-a*Sc-b*Ss;
}
return(c(m,a,b))
}

#--- fit trigonometric model and calculate predictions
for(harm in 1:Nharm)
{
  r=Freq(y,Npast);
  w=r[1];
  m=r[2];
  a=r[3];
  b=r[4];
  for(i in 0:(N-1))
  {
    xm[i+1]=xm[i+1]+m+a*cos(w*i)+b*sin(w*i);
  }
  for(j in 0:(Nfut-1))
  ym[j+1]=ym[j+1]+m+a*cos(w*(N+j))+b*sin(w*(N+j));
}
plot(1:(length(xm)+length(ym)),c(xm,ym), type='l')
MyWriteData=data.frame(y=c(xm,ym))
s2="f_out.csv";
write.csv(MyWriteData, file = s2,dec="," ,row.names=FALSE)

```

Приложение Б

(информационное)

Дополнительные примеры применения метода

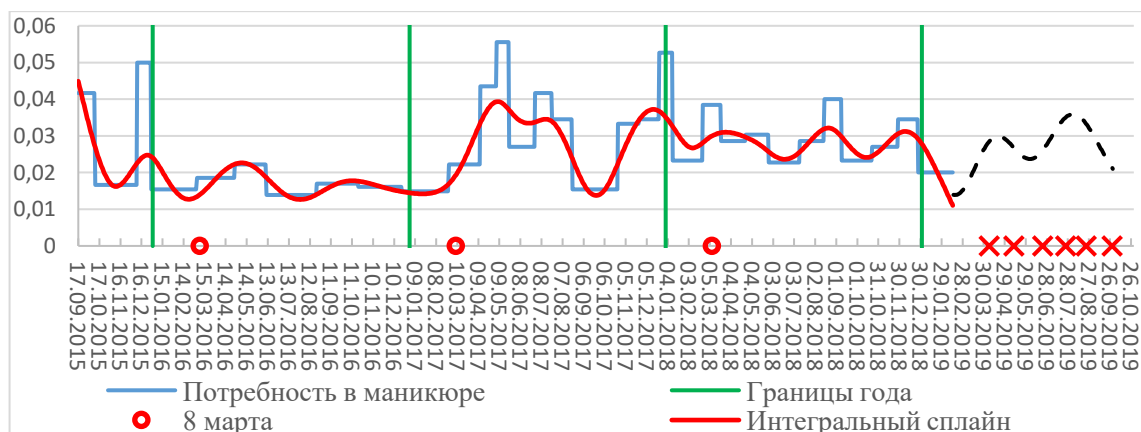
Маникюр 2

Имеется набор данных о маникюре нескольких клиентов, таблица Б.1. Первый клиент на протяжении более 3 лет ходил в один и тот же салон красоты. Из набора данных надо исключить моменты, когда клиент обращается в салон два раза подряд через крайне небольшой период времени (предполагая, что клиент не успел получить нужные услуги за одно обращение 25.10.2017, поэтому он пришел второй раз 27.10.2017). Используя описанный предложенный метод, получим следующий результат, рисунок Б.1 и Б.3.

Таблица Б.1 – Данные о маникюре первого и второго клиента

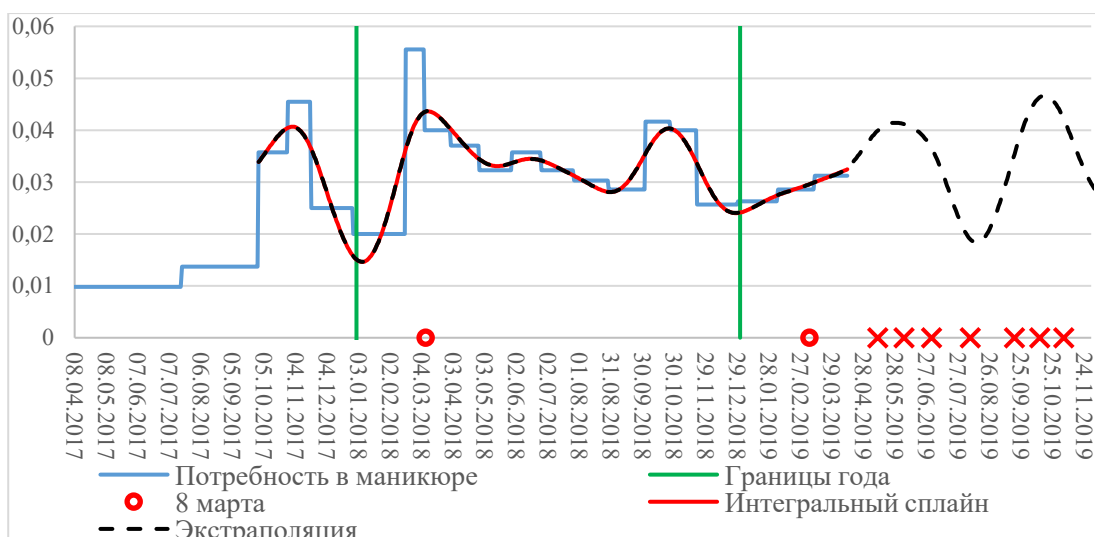
Клиент 1					
i	Дата	i	Дата	i	Дата
1	17.09.2015	13	05.05.2017	25	21.03.2018
2	11.10.2015	14	23.05.2017	26	25.04.2018
3	10.12.2015	15	29.06.2017	27	28.05.2018
4	30.12.2015	16	23.07.2017	28	11.07.2018
5	04.03.2016	17	21.08.2017	29	15.08.2018
6	27.04.2016	18	25.10.2017	30	09.09.2018
7	11.06.2016	19	27.10.2017	31	18.10.2018
8	22.08.2016	20	24.11.2017	32	22.10.2018
9	20.10.2016	21	23.12.2017	33	28.11.2018
10	21.12.2016	22	11.01.2018	34	05.12.2018
11	26.02.2017	23	23.02.2018	35	27.12.2018
12	12.04.2017	24	05.03.2018	36	15.02.2019
Клиент 2					
1	08.04.2017	8	07.03.2018	15	03.10.2018
2	19.07.2017	9	01.04.2018	16	27.10.2018
3	30.09.2017	10	28.04.2018	17	21.11.2018
4	28.10.2017	11	29.05.2018	18	30.12.2018
5	19.11.2017	12	26.06.2018	19	06.02.2019
6	29.12.2017	13	27.07.2018	20	13.03.2019
7	17.02.2018	14	29.08.2018	21	14.04.2019

Источник: составлено автором вместе с соавторами [62].



Источник: составлено автором вместе с соавторами [62].

Рисунок Б.1 – Анализ и прогноз услуг первого клиента, долей в день



Источник: составлено автором.

Рисунок Б.2 – Анализ и прогноз услуг второго клиента, долей в день

У первого клиента можно подметить, что наибольшая потребность в маникюре возникает перед Новым годом, также небольшой скачок наблюдается перед 8 Марта. Второй клиент первые полгода нерегулярно обращался в салон, не используем эти данные для построения модели и экстраполяции. У второго клиента весной вблизи 8 Марта происходит рост, после чего незначительный спад. С октября по середину ноября наблюдается рост, возможно в это время у клиента какой-нибудь праздник, способствующий желанию выглядеть красиво.

Услуги связи

Имеются данные за 4 года о пополнении баланса сотовой связи абонента, представленные в таблице Б.2. Пользователь услуги связи подключен к тарифу, в котором отсутствует ежемесячная или любая другая абонентская плата. Расходование средств происходит только за счет звонков, sms и mss сообщений.

Таблица Б.2 – Данные пополнения баланса

i	Дата	Пополнение	i	Дата	Пополнение	i	Дата	Пополнение
1	17.09.2016	300	4	29.05.2017	300	7	28.08.2018	300
2	19.11.2016	300	5	07.01.2018	300	8	26.11.2018	400
3	02.04.2017	300	6	03.06.2018	300	9	16.04.2019	300

Источник: составлено автором вместе с соавторами [62].

Используя предложенный метод, восстановив и определив закономерность, можно спрогнозировать следующую дату пополнения баланса абонентом. Можно заметить резкий скачок в апреле 2017 года, после которого идет достаточно продолжительный спад. В 2018 году происходит рост к лету и незначительное убывание к зиме. Однако не обязательно

строить модель расхода средств полностью только на прошлых данных. Можно предположить тенденцию поведения искомой функции на данных из других источников. Например, можно предположить, что стоимость всех услуг со временем всегда растет, или абонент с развитием цифровых технологий и различных сервисов будет более активно использовать средства связи, рисунок Б.3. Эти дополнительные предположения можно комбинировать с закономерностями, найденными из данных.



Источник: составлено автором вместе с соавторами [62].

Рисунок Б.3 – Расход средств на связь и прогнозирование пополнение баланса, руб. в день

Потребность в заемных средствах

Московская область. В распоряжении системы Thomson Reuters ¹⁾ имеется история только с 2002 по 2008 год. Восстановленная потребность в заемных средствах (без сглаживания) представлена на рисунке Б.4 (опубликовано в [145]).



Источник: составлено автором.

Рисунок Б.4 – Расход земных средств Московской области, млрд. руб.

¹⁾ Thomson Reuters Eikon. (2019). [MOSCOW CITY; Debt & Credit; Thomson Reuters Eikon]. Retrieved 14-Feb-2019.

Новосибирск. В распоряжении системы Thomson Reuters имелась история только с 2003 по 2017 год. Восстановленная потребность в заемных средствах (без сглаживания) представлена на рисунке Б.5.



Источник: составлено автором.

Рисунок Б.5 – Расход заемных средств г. Новосибирск, млрд. руб.

Заметно снижение активности в период кризиса 2008 года. Однако снижение могло быть вызвано отсутствием возможности взять заемные средства в этот период.

Приложение В
(информационное)
Расчет интеграла 1

Расчет интеграла $\int_a^b (t-a)(b-t)(t+b-2a)dt$.

Представляем подынтегральное выражение в следующем виде:

$$(t-a)(-t+b)(2t-2a-t+b).$$

Группируем таким образом

$$2(t-a)^2(-t+b) + (t-a)(-t+b)^2.$$

Выражение $2(t-a)^2(-t+b)$ представляем как

$$2(t-a)^2(-t+a-a+b) = -2(t-a)^3 - 2(t-a)^2(b-a).$$

Выражение $(t-a)(-t+b)^2$ представляем как

$$(-1)^2(t-b)^2(t-b+b-a) = (t-b)^3 + (t-b)^2(b-a).$$

Теперь берем определенный интеграл от a до b от всех четырех слагаемых:

$$\begin{aligned} \int_a^b -2(t-a)^3 dt &= -\frac{2}{4}(t-a)^4 \Big|_a^b = -\frac{1}{2}(b-a)^4 \\ \int_a^b -2(t-a)^2(b-a) dt &= (b-a) \frac{2}{3} (t-a)^3 \Big|_a^b = \frac{2}{3} (b-a)^4 \\ \int_a^b (t-b)^3 dt &= \frac{1}{4}(t-b)^4 \Big|_a^b = -\frac{1}{4}(a-b)^4 = -\frac{1}{4}(b-a)^4 \\ \int_a^b (t-b)^2(b-a) dt &= (b-a) \frac{1}{3} (t-b)^3 \Big|_a^b = -\frac{1}{3}(b-a)(a-b)^3 = \\ &= \frac{1}{3}(b-a)^4. \end{aligned}$$

В итоге имеем

$$\begin{aligned} \int_a^b (t-a)(b-t)(t+b-2a) dt &= (b-a)^4 \left(-\frac{1}{2} + \frac{2}{3} - \frac{1}{4} + \frac{1}{3} \right) \\ &= (b-a)^4 \left(\frac{-6+8-3+4}{12} \right) = \frac{1}{4}(b-a)^4. \end{aligned}$$

Приложение Г
(информационное)
Расчет интеграла 2

Расчет интеграла $\int_{t_i}^{t_{i+1}} (t-a)(b-t)(2b-t-a)dt$

Представляем подынтегральное выражение в следующем виде

$$(t-a)(-t+b)(2b-2t+t-a).$$

Группируем таким образом

$$(t-a)2(-t+b)^2 + (t-a)^2(-t+b).$$

Выражение $2(t-a)(-t+b)^2$ представляем как

$$2(t-b+b-a)(t-b)^2 = 2(t-b)^3 + 2(t-b)^2(b-a).$$

Выражение $(t-a)^2(-t+b)$ представляем как

$$\begin{aligned} (t-a)^2(-1)(t-a+a-b) &= -(t-a)^3 - (t-a)^2(a-b) = \\ &= -(t-a)^3 + (t-a)^2(b-a). \end{aligned}$$

Теперь берем определенный интеграл от a до b от всех четырех слагаемых:

$$\begin{aligned} \int_a^b 2(t-b)^3 dt &= \frac{2}{4}(t-b)^4 \Big|_a^b = -\frac{1}{2}(a-b)^4 = -\frac{1}{2}(b-a)^4, \\ \int_a^b 2(t-b)^2(b-a) dt &= (b-a) \frac{2}{3} (t-b)^3 \Big|_a^b = -\frac{2}{3} (b-a)(a-b)^3 = \\ &= \frac{2}{3} (b-a)^4, \\ \int_a^b -(t-a)^3 dt &= -\frac{1}{4}(t-a)^4 \Big|_a^b = -\frac{1}{4}(b-a)^4, \\ \int_a^b (t-a)^2(b-a) dt &= (b-a) \frac{1}{3}(t-a)^3 \Big|_a^b = \frac{1}{3}(b-a)^4. \end{aligned}$$

В итоге имеем

$$\begin{aligned} \int_a^b (t-a)(b-t)(2b-t-a) dt &= (b-a)^4 \left(-\frac{1}{2} + \frac{2}{3} - \frac{1}{4} + \frac{1}{3} \right) = \\ &= (b-a)^4 \left(\frac{-6+8-3+4}{12} \right) = \frac{1}{4}(b-a)^4. \end{aligned}$$

Приложение Д
(информационное)
Расчет интеграла 3

Расчет интеграла

$$\int_a^c g(t) dt = \int_a^c \frac{(t-a)g_{k+1} + (b-t)g_k}{b-a} dt$$

$$- \int_a^c \frac{1}{6}(t-a)(b-t) \left\{ \left(1 + \frac{t-a}{b-a}\right) \gamma_{k+1} + \left(1 + \frac{b-t}{b-a}\right) \gamma_k \right\} dt.$$

Первая часть тривиальна

$$\int_a^c \frac{(t-a)g_{k+1} + (b-t)g_k}{b-a} dt = \frac{\frac{(t-a)^2}{2} \Big|_a^c g_{k+1} - \frac{(b-t)^2}{2} \Big|_a^c g_k}{b-a} =$$

$$= \frac{(c-a)^2}{2(b-a)} g_{k+1} - \frac{(b-c)^2 - (b-a)^2}{2(b-a)} g_k.$$

Вторая часть

$$\frac{\gamma_{k+1}}{6(b-a)} \int_a^c (t-a)(b-t)(t+b-2a) dt.$$

В начале упростим выражение $(t-a)(b-t)(t+b-2a)$:

$$\begin{aligned} (t-a)(b-t)(t+b-2a) &= (bt - t^2 - ab + at)(t+b-2a) = \\ &= bt^2 - t^3 - abt + at^2 + b^2t - t^2b - ab^2 + atb - 2abt + 2at^2 + 2a^2b - 2a^2t = \\ &= -t^3 + 3at^2 + (b^2 - 2ab - 2a^2)t - ab^2 + 2a^2b = \\ &= -t^3 + 3at^2 - 3a^2t + 3a^2t + a^3 - a^3 + (b^2 - 2ab - 2a^2)t - ab^2 + 2a^2b = \\ &= -(t-a)^3 + 3a^2t - a^3 + (b^2 - 2ab - 2a^2)t - ab^2 + 2a^2b = \\ &= -(t-a)^3 + (b^2 - 2ab + a^2)t - ab^2 + 2a^2b - a^3 = \\ &= -(t-a)^3 + (b-a)^2t - a(b^2 - 2ab + a^2) = \\ &= -(t-a)^3 + (b-a)^2t - a(b-a)^2 = \\ &= -(t-a)^3 + (b-a)^2(t-a). \end{aligned}$$

Далее берем определенный интеграл

$$\int_a^c (-(t-a)^3 + (b-a)^2(t-a)) dt =$$

$$= -\frac{(t-a)^4}{4} \Big|_a^c + (b-a)^2 \frac{(t-a)^2}{2} \Big|_a^c =$$

$$= -\frac{(c-a)^4}{4} + (b-a)^2 \frac{(c-a)^2}{2} =$$

$$= -\frac{(c-a)^2}{4} ((c-a)^2 - 2(b-a)^2).$$

В итоге вторая часть будет

$$\begin{aligned} & \frac{\gamma_{k+1}}{6(b-a)} \int_a^c (t-a)(b-t)(t+b-2a) dt = \\ & = -\frac{\gamma_{k+1}(c-a)^2}{24(b-a)} ((c-a)^2 - 2(b-a)^2). \end{aligned}$$

Переходим к третьей части

$$\frac{\gamma_k}{6(b-a)} \int_a^c (t-a)(b-t)(-t+2b-a) dt.$$

В начале упростим выражение $(t-a)(b-t)(-t+2b-a)$:

$$\begin{aligned} (t-a)(b-t)(-t+2b-a) &= (bt-t^2-ab+at)(-t+2b-a) = \\ &= -bt^2+t^3+abt-at^2+2b^2t-2bt^2-2ab^2+2abt-abt+at^2+a^2b-a^2t = \\ &= -bt^2+t^3+2b^2t-2bt^2-2ab^2+2abt+a^2b-a^2t = \\ &= t^3-3bt^2+t(2b^2+2ab-a^2)-2ab^2+a^2b = \\ &= t^3-3bt^2+3b^2t-3b^2t-b^3+b^3+t(2b^2+2ab-a^2)-2ab^2+a^2b = \\ &= t^3-3bt^2+3b^2t-b^3+t(-b^2+2ab-a^2)+b^3-2ab^2+a^2b = \\ &= (t-b)^3-t(b-a)^2+b(b^2-2ab+a^2) = \\ &= (t-b)^3-t(b-a)^2+b(b-a)^2 = \\ &= (t-b)^3-(b-a)^2(t-b). \end{aligned}$$

Далее берем определенный интеграл

$$\begin{aligned} & \int_a^c ((t-b)^3-(b-a)^2(t-b)) dt = \\ &= \frac{(t-b)^4}{4} \Big|_a^c - (b-a)^2 \frac{(t-b)^2}{2} \Big|_a^c = \\ &= \frac{(c-b)^4}{4} - \frac{(a-b)^4}{4} - (b-a)^2 \frac{(c-b)^2}{2} + (b-a)^2 \frac{(a-b)^2}{2} = \\ &= \frac{(c-b)^4}{4} + \frac{(a-b)^4}{4} - \frac{2(b-a)^2(c-b)^2}{4} = \\ &= \frac{1}{4} ((c-b)^2 - (b-a)^2)^2 = \\ &= \frac{1}{4} (c^2 - 2cb + b^2 - b^2 + 2ab - a^2)^2 = \\ &= \frac{1}{4} (c^2 - 2cb + 2ab - a^2)^2 = \\ &= \frac{1}{4} (c^2 - a^2 - 2b(c-a))^2 = \\ &= \frac{1}{4} ((c-a)(c+a) - 2b(c-a))^2 = \end{aligned}$$

$$= \frac{1}{4}(c-a)^2(c+a-2b)^2.$$

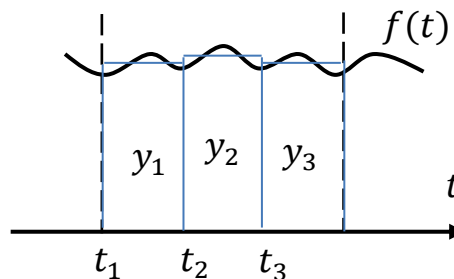
В итоге третья часть будет

$$\begin{aligned} \frac{\gamma_k}{6(b-a)} \int_a^c (t-a)(b-t)(-t+2b-a) dt &= \\ &= \frac{\gamma_k(c-a)^2}{24(b-a)} (c+a-2b)^2 = \\ &= \frac{\gamma_k(c-a)^2}{24(b-a)} (c-b+a-b)^2. \end{aligned}$$

Приложение Е
(информационное)

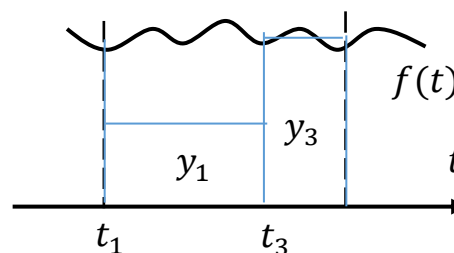
Математическое обоснование падения точности от потери части данных

Рассмотрим случай частых событий, когда на каждый интервал времени τ будет приходиться несколько событий, и попробуем определить среднеквадратичную относительную погрешность восстановления функции (без сглаживания). Из объяснения результатов относительно частых событий получится вывод, относящийся к редким событиям. Для упрощения представим картину, когда на интервал τ выпадает три события (покупки), причем первая выпадает в самом начале интервала, для следующего интервала первое событие также выпадает в самом начале, рисунок Е.1. Величина события y_i есть суммарное воздействие оригинальной функции $f(t)$ за время между моментами текущего события t_i и следующего события t_{i+1} (площадь под кривой $y_i = \int_{t_i}^{t_{i+1}} f(t)dt$), другими словами запас продукции заканчивается к моменту следующего пополнения запаса. В простейшем случае можно определить среднюю скорость $\tilde{F}(t)$ расхода запаса за время между двумя событиями $y_i/(t_{i+1} - t_i)$, которая будет являться горизонтальной линией, усредняющей оригинальную функцию.



Источник: составлено автором.
Рисунок Е.1 – На интервал попало 3 события

Теперь представим, что потерялось одно из наблюдений, например, пусть потерялись данные второго события. Тогда визуально получается, что величина первого события y_1 приходится на период от t_1 до t_3 , рисунок Е.2. Средний уровень определится как $\tilde{F}(t) = y_1/(t_3 - t_1)$, будет наблюдаться занижение.



Источник: составлено автором.
Рисунок Е.2 – Второе наблюдение потерялось

Для упрощения рассуждений предположим, что среди значений y_i нет выделяющихся по величине, то есть различия между ними не существенны $y_1 \approx y_2 \approx y_3 \approx y$. Введем обозначение η – среднее количество событий за интервал τ , это позволит рассматривать произвольное количество событий за интервал, а не только когда их 3.

Тогда на участке, где происходит занижение, имеем

$$\tilde{F}(t) = \frac{\int_{t_i}^{t_i + \frac{\tau}{\eta}} f(t) dt}{2\tau/\eta} = \frac{y_1}{2\tau/\eta} \approx \frac{y}{2\tau/\eta}.$$

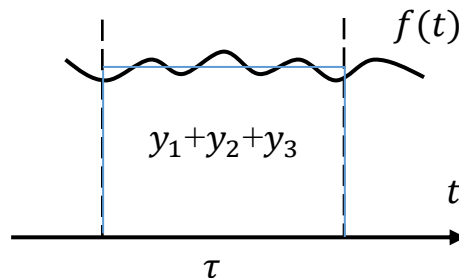
На участке, где не было занижения имеем

$$\tilde{F}(t) = \frac{\int_{t_i + \frac{2\tau}{\eta}}^{t_i + \frac{3\tau}{\eta}} f(t) dt}{\tau/\eta} = \frac{y_3}{\tau/\eta} \approx \frac{y}{\tau/\eta}.$$

Определим суммарную квадратичную относительную погрешность на всем участке τ :

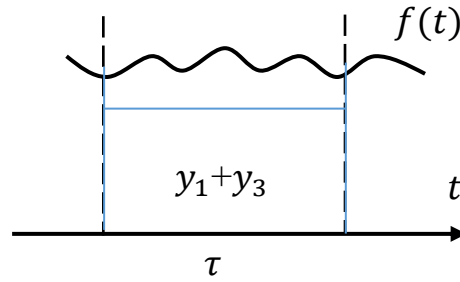
$$\begin{aligned} & \int_{t_i}^{t_i + \frac{2\tau}{\eta}} \left(\frac{\frac{\eta y}{2\tau} - f(t)}{f(t)} \right)^2 dt + \sum_{k=2}^{\eta-1} \int_{t_i + \frac{k\tau}{\eta}}^{t_i + \frac{(k+1)\tau}{\eta}} \left(\frac{\frac{\eta y}{\tau} - f(t)}{f(t)} \right)^2 dt = \\ & = \int_{t_i}^{t_i + \frac{2\tau}{\eta}} \left(\frac{\eta y}{2\tau f(t)} \right)^2 dt - \int_{t_i}^{t_i + \frac{2\tau}{\eta}} \frac{2\eta y}{2\tau f(t)} dt + \int_{t_i}^{t_i + \frac{2\tau}{\eta}} 1 dt + \\ & + \sum_{k=2}^{\eta-1} \left(\int_{t_i + \frac{k\tau}{\eta}}^{t_i + \frac{(k+1)\tau}{\eta}} \left(\frac{\eta y}{\tau f(t)} \right)^2 dt - \int_{t_i + \frac{k\tau}{\eta}}^{t_i + \frac{(k+1)\tau}{\eta}} \frac{2\eta y}{\tau f(t)} dt + \int_{t_i + \frac{k\tau}{\eta}}^{t_i + \frac{(k+1)\tau}{\eta}} 1 dt \right). \end{aligned}$$

Теперь рассмотрим классический подход, когда формируется временной ряд как суммарная величина всех событий, попавших на рассматриваемый интервал τ . Значение временного ряда на этом интервале будет $y_1 + y_2 + y_3$. Для сравнения методов надо оценить возможность восстановления исходной функции $f(t)$. Если пытаемся восстановить скорость расхода запаса на интервале, то в простейшем случае можем определить среднюю скорость $\bar{F}(t) = (y_1 + y_2 + y_3)/\tau$, которая будет горизонтальной линией, усредняющей исходную функцию $f(t)$ на всем периоде, рисунок Е.3.



Источник: составлено автором.
Рисунок Е.3 – На интервал попало 3 события

При потере второго наблюдения, на интервал τ попадет два события, средний уровень $\bar{F}(t) = (y_1 + y_3)/\tau$ будет занижен, рисунок Е.4.



Источник: составлено автором.

Рисунок Е.4 – Второе наблюдение потерялось

Используя обозначение η – среднее количество событий за интервал τ , запишем

$$\bar{F}(t) = \frac{\int_{t_i}^{t_i+\tau} f(t) dt - \int_{t_i+\frac{\tau}{\eta}}^{t_i+\frac{2\tau}{\eta}} f(t) dt}{\tau} \approx \frac{y(\eta - 1)}{\tau}.$$

Опять найдем суммарную квадратичную относительную погрешность на всем участке τ :

$$\begin{aligned} & \int_{t_i}^{t_i+\tau} \left(\frac{y(\eta - 1)}{\tau} - f(t) \right)^2 dt = \\ & = \int_{t_i}^{t_i+\tau} \left(\frac{y(\eta - 1)}{\tau f(t)} \right)^2 dt - \int_{t_i}^{t_i+\tau} \frac{2y(\eta - 1)}{\tau f(t)} dt + \int_{t_i}^{t_i+\tau} 1 dt. \end{aligned}$$

Теперь надо сравнить ранее полученное выражение с этим. В первом и втором выражении есть $\int_{t_i}^{t_i+\frac{2\tau}{\eta}} 1 dt + \sum_{k=2}^{\eta-1} \int_{t_i+\frac{k\tau}{\eta}}^{t_i+\frac{(k+1)\tau}{\eta}} 1 dt = \int_{t_i}^{t_i+\tau} 1 dt$, поэтому осталось сравнить

$$\int_{t_i}^{t_i+\frac{2\tau}{\eta}} \left(\frac{\eta y}{2\tau f(t)} \right)^2 dt - \int_{t_i}^{t_i+\frac{2\tau}{\eta}} \frac{2\eta y}{2\tau f(t)} dt + \sum_{k=2}^{\eta-1} \left(\int_{t_i+\frac{k\tau}{\eta}}^{t_i+\frac{(k+1)\tau}{\eta}} \left(\frac{\eta y}{\tau f(t)} \right)^2 dt - \int_{t_i+\frac{k\tau}{\eta}}^{t_i+\frac{(k+1)\tau}{\eta}} \frac{2\eta y}{\tau f(t)} dt \right)$$

и

$$\int_{t_i}^{t_i+\tau} \left(\frac{y(\eta - 1)}{\tau f(t)} \right)^2 dt - \int_{t_i}^{t_i+\tau} \frac{2y(\eta - 1)}{\tau f(t)} dt.$$

Рассмотрим первое выражение, его можно представить в виде

$$\begin{aligned} & \left(\frac{\eta y}{2\tau} \right)^2 \int_{t_i}^{t_i+\frac{2\tau}{\eta}} \left(\frac{1}{f(t)} \right)^2 dt + \left(\frac{\eta y}{\tau} \right)^2 \sum_{k=2}^{\eta-1} \int_{t_i+\frac{k\tau}{\eta}}^{t_i+\frac{(k+1)\tau}{\eta}} \left(\frac{1}{f(t)} \right)^2 dt \\ & - \frac{\eta y}{\tau} \int_{t_i}^{t_i+\frac{2\tau}{\eta}} \frac{1}{f(t)} dt - \frac{2\eta y}{\tau} \sum_{k=2}^{\eta-1} \int_{t_i+\frac{k\tau}{\eta}}^{t_i+\frac{(k+1)\tau}{\eta}} \frac{1}{f(t)} dt. \end{aligned}$$

Множитель перед интегралом без суммы в 4 и в 2 раза меньше, чем множитель перед интегралом под суммой. Подынтегральные выражения совпадают, сумму интегралов можно заменить интегрированием по всему периоду τ , объединив интегралы:

$$-\frac{3}{4}\left(\frac{\eta y}{\tau}\right)^2 \int_{t_i}^{t_i+\frac{2\tau}{\eta}} \frac{1}{f^2(t)} dt + \left(\frac{\eta y}{\tau}\right)^2 \int_{t_i}^{t_i+\tau} \frac{1}{f^2(t)} dt \\ + \frac{\eta y}{\tau} \int_{t_i}^{t_i+\frac{2\tau}{\eta}} \frac{1}{f(t)} dt - \frac{2\eta y}{\tau} \int_{t_i}^{t_i+\tau} \frac{1}{f(t)} dt.$$

Рассмотрим второе выражение и раскроем скобки:

$$\left(\frac{y(\eta-1)}{\tau}\right)^2 \int_{t_i}^{t_i+\tau} \frac{1}{f^2(t)} dt - \frac{2y(\eta-1)}{\tau} \int_{t_i}^{t_i+\tau} \frac{1}{f(t)} dt = \\ = \frac{y^2\eta^2 - 2y^2\eta + y^2}{\tau^2} \int_{t_i}^{t_i+\tau} \frac{1}{f^2(t)} dt - \frac{2y\eta - 2y}{\tau} \int_{t_i}^{t_i+\tau} \frac{1}{f(t)} dt.$$

Часть слагаемых совпадает, тогда осталось сравнить

$$-\frac{3}{4}\left(\frac{\eta y}{\tau}\right)^2 \int_{t_i}^{t_i+\frac{2\tau}{\eta}} \frac{1}{f^2(t)} dt + \frac{\eta y}{\tau} \int_{t_i}^{t_i+\frac{2\tau}{\eta}} \frac{1}{f(t)} dt$$

и

$$\frac{-2y^2\eta + y^2}{\tau^2} \int_{t_i}^{t_i+\tau} \frac{1}{f^2(t)} dt + \frac{2y}{\tau} \int_{t_i}^{t_i+\tau} \frac{1}{f(t)} dt.$$

Сравнению мешают интегралы за разные периоды от одной и той же подынтегральной функции. В первом выражении площадь под функцией за время $\frac{2\tau}{\eta}$, а во втором выражении за весь интервал времени τ . Можно попробовать сравнить пропорции между этими площадями. Если площадь под функцией от события к событию приблизительно одинаковая, то тогда два интеграла относятся друг к другу также как ширина, за которую происходит интегрирование, то есть как $\frac{2\tau}{\eta}$ к τ . Тогда далее необходимо сравнить следующие выражения, где интегралы заменены на пропорции между ними и на некоторые множители K_1, K_2 :

$$-\frac{3}{4}\left(\frac{\eta y}{\tau}\right)^2 \frac{2\tau}{\eta} K_1 + \frac{\eta y}{\tau} \frac{2\tau}{\eta} K_2 = \frac{-1.5\eta y^2}{\tau} K_1 + 2y K_2$$

и

$$\frac{-2\eta y^2 + y^2}{\tau^2} \tau K_1 + \frac{2y}{\tau} \tau K_2.$$

Вторые слагаемые совпадают. Первые слагаемые в числителе отличаются друг от друга на $-0.5y^2\eta + y^2 = -y^2(0.5\eta - 1)$. Тогда получаем, если среднее количество событий на интервале времени $\eta > 2$, то разница отрицательна, иначе положительна. То есть получается, что даже при частых событиях, если количество событий за каждый интервал времени не больше 2, то восстановление исходной функции $f(t)$ будет более

точным у емкостного метода. При редких событиях частота η много меньше единицы, так что для редких событий это еще более справедливо. Кроме того, при редких событиях возникают дополнительные эффекты, когда события не попадают на интервал времени, что при формировании временных рядов приводит к нулевым значениям, и, следовательно, к большой погрешности классического подхода. Построение временных рядов из редких событий заведомо проигрышная затея.