

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

АВТОРЕФЕРАТ  
диссертации на соискание ученой степени  
доктора экономических наук

Научный консультант

Бывшев Виктор Алексеевич,  
доктор технических наук, профессор

Москва – 2023

Диссертация представлена к публичному рассмотрению и защите в порядке, установленном ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации» в соответствии с представленным правом самостоятельно присуждать учёные степени кандидата наук, учёные степени доктора наук согласно положениям пункта 3.1 статьи 4 Федерального закона от 23 августа 1996 г. №127-ФЗ «О науке и государственной научно-технической политике».

Публичное рассмотрение и защита диссертации состоится 20 сентября 2023 г. в 15:00 часов на заседании диссертационного совета Финансового университета Д 505.001.111 по адресу: Москва, Ленинградский проспект, д. 51, корп. 1, аудитория 1001.

С диссертацией можно ознакомиться в диссертационном зале Библиотечно-информационного комплекса ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации» по адресу: 125167, Москва, Ленинградский проспект, д. 49/2, комн. 100 и на официальном сайте Финансового университета в информационно-телекоммуникационной сети «Интернет» по адресу: [www.fa.ru](http://www.fa.ru)

Персональный состав диссертационного совета:

председатель – Соловьев В.И., д.э.н., доцент;  
заместитель председателя – Коровин Д.И., д.э.н., доцент;  
ученый секретарь – Золотова Т.В., д.физ.-мат.н., доцент;

члены диссертационного совета:

Абдикеев Н.М., д.техн.н., профессор;  
Афанасьев А.А., д.э.н., доцент;  
Васильева Е.В., д.э.н., доцент;  
Владова А.Ю., д.техн.н., доцент;  
Гатауллин Т.М., д.э.н., профессор;  
Судаков В.А., д.техн.н., доцент;  
Трегуб И.В., д.э.н., профессор.

Автореферат диссертации разослан 31 марта 2023 г.

Учёный секретарь диссертационного совета  
Финансового университета Д 505.001.111

Т.В. Золотова

## I Общая характеристика работы

**Актуальность темы исследования.** Способность анализировать и предсказывать будущие события всегда имело и будет иметь большое значение для любой области деятельности. Знание о грядущем событии позволит грамотно к нему подготовиться, тем самым извлечь определенную выгоду или уменьшить возможные убытки. В современной экономике это особенно актуально. На торгах информация о грядущем спаде или скачке цен позволяет некоторым игрокам 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 хотя 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. Исследование существующих методов показало, что их 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 методов, применяемых в экономике, является то, что все полученные оценки параметров статистических 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 данном исследовании 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. 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 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., Беляева Ю.К., Соловьева А.Д., Ушакова И.А., Кокса Д., Барлоу Р., Прошана 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) и второго порядка (SORM), выборкой по линиям (Line sampling) 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 Т. 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 С. 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, схожих с наполнением/опустошением некоторых материальных или нематериальных емкостей. 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).

6) 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 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: 1) аналитически (использование математического анализа, теории вероятностей, математической статистики, линейной алгебры); 2) экспериментально (при помощи моделей); 3) практикой. 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: на III Всероссийской научной конференции «Россия 2030 глазами молодых ученых» (Москва,

ИНИОН РАН, 26 апреля 2012 г.); на VI Международной научно-практической Интернет-конференции «Современные проблемы моделирования социально-экономических систем» (г. Харьков, Харьковский национальный экономический университет, 3–12 апреля 2014 г.); на VII Международной научно-практической Интернет-конференции «Современные проблемы моделирования социально-экономических систем» (г. Харьков, Харьковский национальный экономический университет, 2–10 апреля 2015 г.); на ежегодном Международном круглом столе «Системная экономика, социально-экономическая кибернетика, мягкие измерения в экономике – 2017» (Москва, Финансовый университет, 8 июня 2017 г.); на Ежегодном международном круглом столе «Системная экономика, социально-экономическая кибернетика, мягкие измерения в экономике – 2018» (Москва, Финансовый университет, 6 июня 2018 г.); на Всероссийском межвузовском круглом столе «Мягкие измерения в научной и учебной деятельности» (Москва, Финансовый университет, 17 сентября 2018 г.); на V Международной научно-практической конференции-биеннале «Системный анализ в экономике» - 2018 (Москва, Финансовый университет, 21-23 ноября 2018 г.); на Международной научно-практической конференции «Анализ данных, принятие решений и Финансовые технологии», посвященной 100-летию Финансового университета (Москва, Финансовый университет, 23 мая 2019 г.); на Ежегодной международной конференции «Системная экономика, социально-экономическая кибернетика, мягкие измерения в экономике – 2019» (Москва, Финансовый университет, 7 июня 2019 г.); на Международной научно-практической конференции «Системная экономика, социально-экономическая кибернетика и мягкие измерения в экономике» (Москва, Финансовый университет, 20 мая 2020 г.); на VII Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 25 июня 2020 г.); на VI Международной научно-практической конференции-биеннале «Системный анализ в экономике – 2020» (Москва, Финансовый университет, 9–11 декабря 2020 г.); на VIII Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 28 мая 2021 г.); на Международной научно-практической конференции «Системная экономика, социально-экономическая кибернетика и мягкие измерения в экономике» (Москва, Финансовый университет, 24–25 июня 2021 г.); на Научном семинаре в ЦЭМИ – «Прикладная статистика и моделирование реальных процессов» (Москва, ЦЭМИ, 8 декабря 2021 г.); на Научном семинаре на секции «Эконометрика» в Департаменте математики (Москва, Финансовый университет, 26 января 2022 г.); на IX Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 9 июня 2022 г.); на Международной научно-практической конференции «Системная экономика,

социально-экономическая кибернетика и мягкие измерения в экономике» (Москва, Финансовый университет, 1 июля 2022 г.).

Результаты исследования использованы при выполнении научно-исследовательской работы (далее - НИР) по одноименной теме проекта исследований гранта Российского Фонда Фундаментальных Исследований «Емкостный метод анализа редких событий в экономике» в 2019–2021 г., № 19-010-00154.

Результаты исследования 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 инструментальный 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 диссертации используются кафедрой «Системный анализ в экономике» Финансового университета в преподавании учебных дисциплин: «Экономико-математическое моделирование логистики», «Математические модели и методы в логистике».

Апробация и внедрение результатов исследования подтверждены соответствующими документами.

**Публикации.** Результаты диссертационного исследования опубликованы в 30 научных работах общим объемом 41,51 п.л. (авторский объем 38,57 п.л.), в том числе в 20 статьях общим объемом 26,96 п.л. (авторский объем – 24,02 п.л.), опубликованных в рецензируемых научных изданиях, определенных ВАК при Минобрнауки России, в 3 статьях авторским объемом 6,65 п.л., входящих в цитатно-аналитическую базу RSCI, в 1 статье авторским объемом 2,1 п.л., опубликованной в издании, входящем в цитатно-аналитическую базу «Scopus» (Q1) и в 1 статье

авторским объемом 2,0 п.л., опубликованной в издании, входящем в цитатно-аналитическую базу «Scopus» (Q2).

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

## **II Основное содержание работы**

В **первой главе** диссертации делается обзор существующих научных работ, в которых используется термин «редкое событие», выявляется список задач и делается обзор существующих методов для работы с редкими событиями.

Во-первых, под термином «событие» в данной работе понимается более общее понятие, как «происшествие, что сбылось», нежели более узкое понятие из теории вероятностей как «результат некоторого опыта, испытания». Во-вторых, под «редкостью» понимается либо насколько количество одних событий превосходит количество других событий, либо насколько вероятности возникновения этих событий за промежуток времени отличаются друг от друга. В то же время обращается внимание на то, что вопрос о частоте или редкости относится ко второму закону диалектики, и само деление может не иметь четкой грани. Различия определяются от контекста. В некоторых работах вместо термина «редкие события» используют термин «аномалии», «выбросы», «новизна», «инцидент», но при этом подразумевают то же самое. Также под редкими событиями иногда понимают более значимые, более серьезные события по сравнению с другими событиями (сравнение опять может не иметь четкой грани), либо события, которые несут больше информации.

Понятие «редкие события» 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 наибольшее значение для экономики составляет выявление закономерностей и прогнозирование появления событий, что является не простой задачей.

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 анализируются лишь появление или непоявление событий, но не способ появления событий. Вместе с тем получающиеся модели оказываются статичными, хотя образование событий может происходить в динамически изменяющихся условиях (обучающая выборка рассматривается как статичная картина, оцененная модель будет иметь фиксированные параметры, а не динамические).

Обращается внимание на то, что в существующих методах никак не используется информация о способах появления самих этих событий, то есть о процессах образования событий. Делается вывод о необходимости обратить внимание на внутренние процессы, приводящие к образованию событий.

Также замечается, что 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).

**Во второй главе** непосредственно разрабатывается новый метод анализа и прогнозирования редких событий в экономике.

1) Разработан новый общий подход для анализа и прогнозирования редких событий в экономике, который опирается на информацию о процессе образования событий. Данный метод состоит из 5 шагов:

- Шаг 1. Разделение событий по источникам, где они образованы (необходим отказ от агрегирования или смешивания событий, информация не должна теряться).

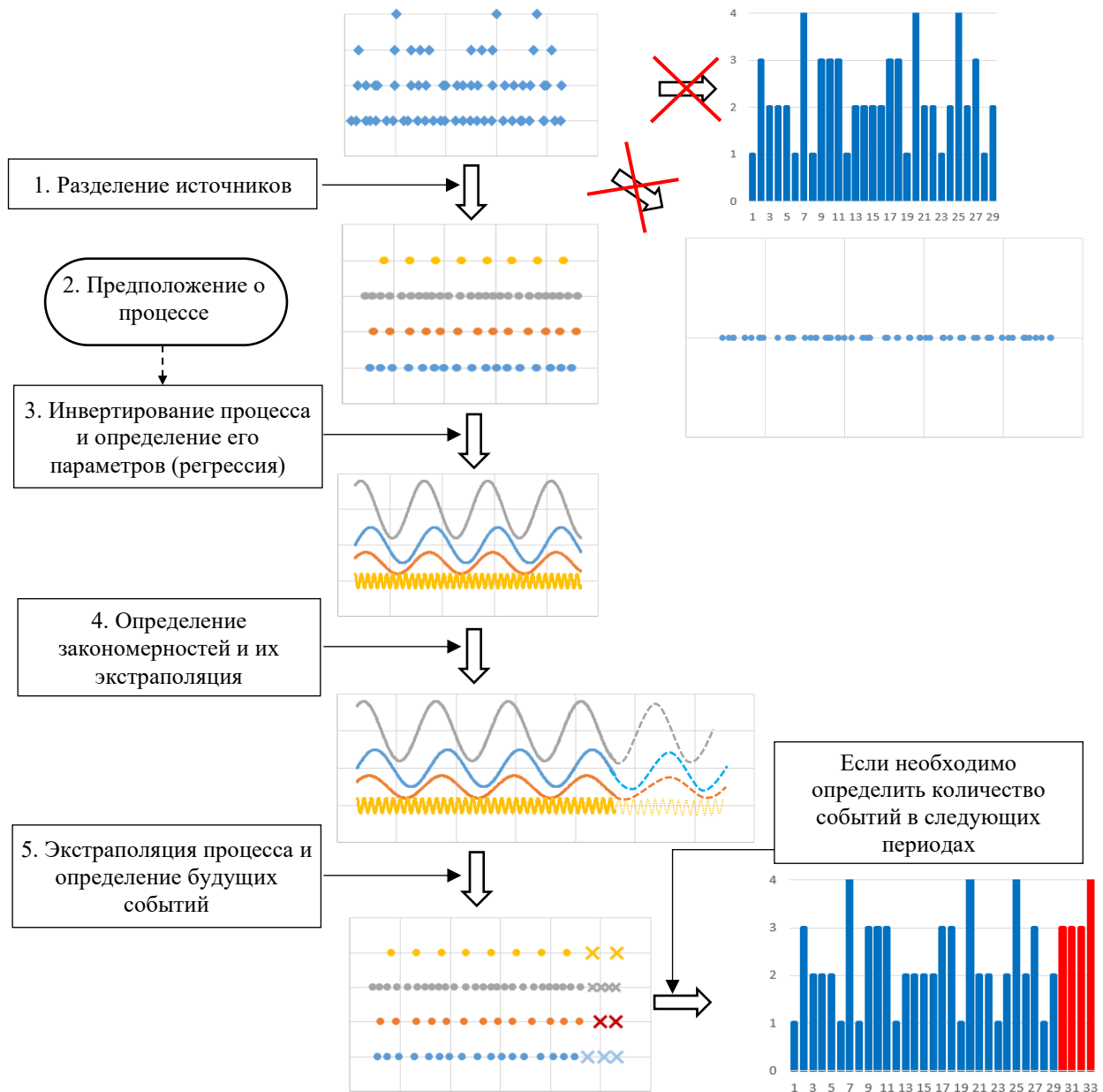
- Шаг 2. Выдвижение предположения о процессе/механизме образования событий в источнике (в каждом источнике образования событий может быть свой собственный уникальный процесс образования событий).

- Шаг 3. Восстановление параметров процесса по выборке редких событий (регрессия параметров процесса, причем параметры могут быть динамическими).

- Шаг 4. Экстраполяция значений параметров процесса на будущее (на этом этапе исследователь свободен в выборе метода для поиска закономерностей изменения параметров со временем или в зависимости от внешних факторов).

- Шаг 5. Запуск самого процесса с установленными значениями параметров и получение прогноза будущих событий (для каждого источника).

Если в дальнейшем потребуется получить прогноз будущих событий в виде временного ряда, не составит труда агрегировать полученные события за выбранные интервалы времени. Схема этого общего подхода изображена на рисунке 1.



Источник: составлено автором.

Рисунок 1 - Схема анализа и прогнозирования редких событий

2) Разработан метод анализа и прогнозирования редких событий в экономике, процесс образования которых схож с процессами наполнения/опустошения некоторой емкости (что определило название метода как «ёмкостный метод»). Данный метод является частным случаем общего метода, описанного выше. Предполагается, что модель процесса образования событий следующая: некоторая емкость опустошается (или наполняется) с нестационарной скоростью (интенсивностью), само образование событий происходит в тот момент, когда эта емкость полностью опустошится (или полностью наполнится), в момент образования события эта емкость вновь наполняется (или опустошается). Данный метод подходит для процессов потребления, когда истощение запаса заставляет совершать новые покупки для пополнения запаса (как в моделях управления запасами), или процессов накопления некоторого возмущения, когда достижение критического уровня возмущения приводит к определенному событию, в

результате которого возмущение уменьшается или сбрасывается. Последнее справедливо для сферы услуг, например услуг парикмахерской.

Примечание - В системах управления запасами моделируется расход собственных запасов или запасов подконтрольных пунктов распространения, причем прогноз спроса уже задан в каждом подконтрольном пункте распространения. В предложенном емкостном методе в данном примере происходит определение скорости расхода продукции у неподконтрольных нам клиентов, после чего прогнозируются дискретные моменты их покупок. Мы не управляем ничьими запасами, не считаем издержки и не оптимизируем точку заказа, а прогнозируем будущие покупки неподконтрольных клиентов.

Самой важной характеристикой для дальнейшего прогноза события является функция  $f(t)$  скорости (интенсивности) наполнения/опустошения, например расхода запаса (накопления воздействия). Достаточно знать только лишь эту функцию  $f(t)$  для того, чтобы затем уметь моделировать процесс возникновения следующего событий (отталкиваясь от объема последней совершенной покупки). Примером такой функции может являться: функция спроса как функция времени, которая «вытягивает» товар из запасов на складе промежуточного распространителя, который пополняет свои запасы (совершая покупки у производителя); индивидуальная скорость потребления продукции конечного потребителя, который закупается у нас (у владельца магазина); интенсивность покупок неизвестных покупателей у выбранного покупателя (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. Например, можно предположить, что объем будущей покупки будет соответствовать тому, сколько не хватает до максимального уровня запасов, который также может меняться по определенному правилу. Если же мы имеем дело со сферой услуг, то объем приобретаемой услуги можно предположить всегда равен константе (например, 1, как сам факт приобретения услуги), в этом случае никакие дополнительные параметры вводить не надо. Это был второй шаг предложенной схемы анализа и прогнозирования редких событий (первый шаг выполняется автоматически, если в нашем примере данные в выборке относятся только к одному клиенту).

На третьем шаге необходимо восстановить параметры процесса. Для восстановления функции  $f(t)$  (скорости изменения запаса или накопления воздействия) инвертируем процесс потребления продукции, получаем обратную задачу к задаче управления запасами. По имеющимся данным редких событий  $(t_i, y_i)$  - моментах времени и величинах воздействия события (например, объемы покупок) определяется скорость изменения уровня емкости  $f(t)$ . 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  $t_{i+1}$  (наблюдения  $y_i$  могут содержать погрешность  $\varepsilon_i$ ), формула (1)

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

Например, следующее  $(i + 1)$ -е событие происходит в тот момент времени, когда запас, сделанный во время текущего  $i$ -того события, закончился. С использованием этого предположения задача восстановления функции  $f(t)$  представляет из себя задачу восстановления неизвестной функции  $\hat{f}(t)$  по последовательности интегралов, которую можно сформулировать в виде задачи минимизации суммы квадратов отклонений, формула (2)

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

где второе слагаемое – штраф на нелинейность;

$\alpha$  – параметр сглаживания;

$w_i$  – вес наблюдения;

$n$  – номер последнего события.

Решение этой задачи получено в виде кубического интегрального сглаживающего (аппроксимирующего) сплайна. Решение данной задачи и программная реализация вынесены в третью главу диссертации. Определение других параметров, например, таких как максимальный уровень запасов, можно осуществить после восстановления функции  $f(t)$  непосредственно из известных данных  $(t_i, y_i)$  более простыми методами.

Четвертый шаг заключается в экстраполяции параметров процесса на будущее. На этом этапе предполагается полная свобода и полная ответственность исследователя, можно использовать любые известные математические или статистические средства.

На пятом этапе моделируется процесс формирования событий. Для этого в том же примере реализуется процесс потребления так же, как в моделях управления запасами.

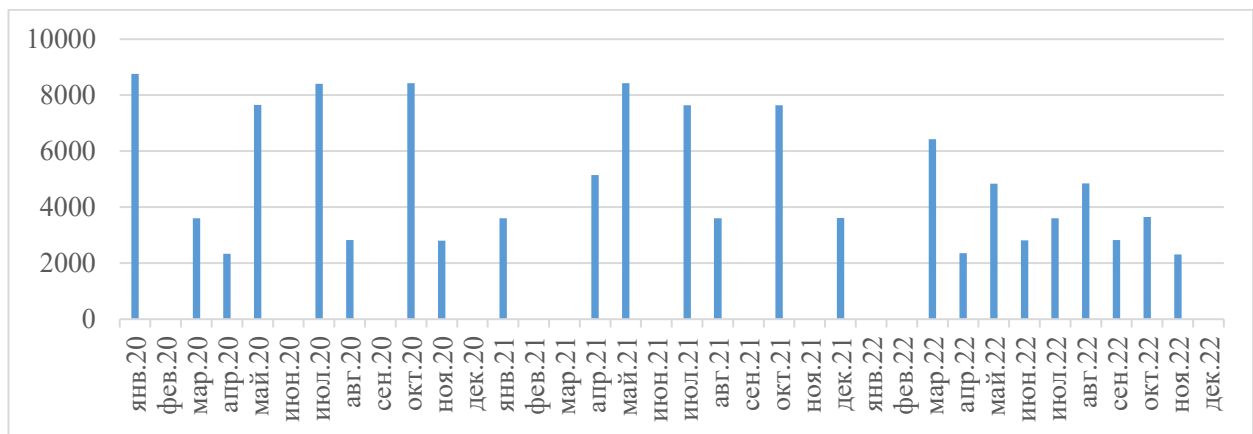
Пример 1. Целесообразно показать работоспособность метода на данных, получающихся в результате модели управления запасами, когда можно задать нужные параметры, а потом по сформированным событиям восстановить эти же параметры и предсказать новые события. Соответствующие входные данные представлены в таблице 1. Особенно стоит отметить, что если агрегировать эти данные по месяцам для получения временного ряда, то получается картина как на рисунке 2, из которой уже очень тяжело найти какие-либо зависимости. Если же не смешивать данные, то удастся восстановить скорость расхода запаса и другие параметры для каждого клиента, после чего экстраполировать все параметры на будущее (в данном случае для экстраполяции подходит алгоритм Куинна-Фернандеса) и запустить сам процесс для получения прогноза будущих событий. Результат применения метода изображен на рисунке 3.



Таблица 1 – Пример входных данных (на основе модели управления запасами)

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

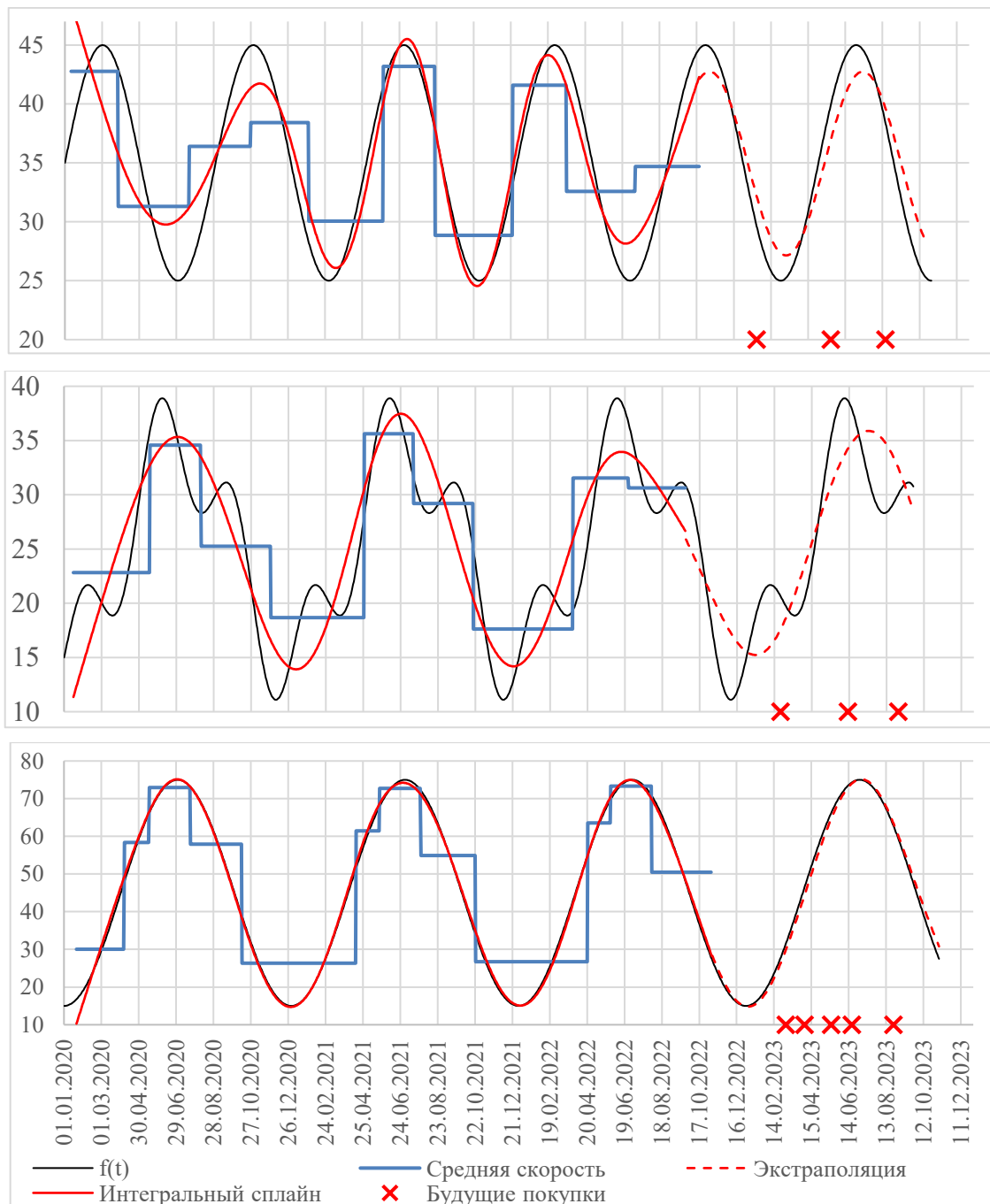
Источник: составлено автором.



Источник: составлено автором.

Рисунок 2 – Данные, агрегированные по месяцам, шт. в месяц

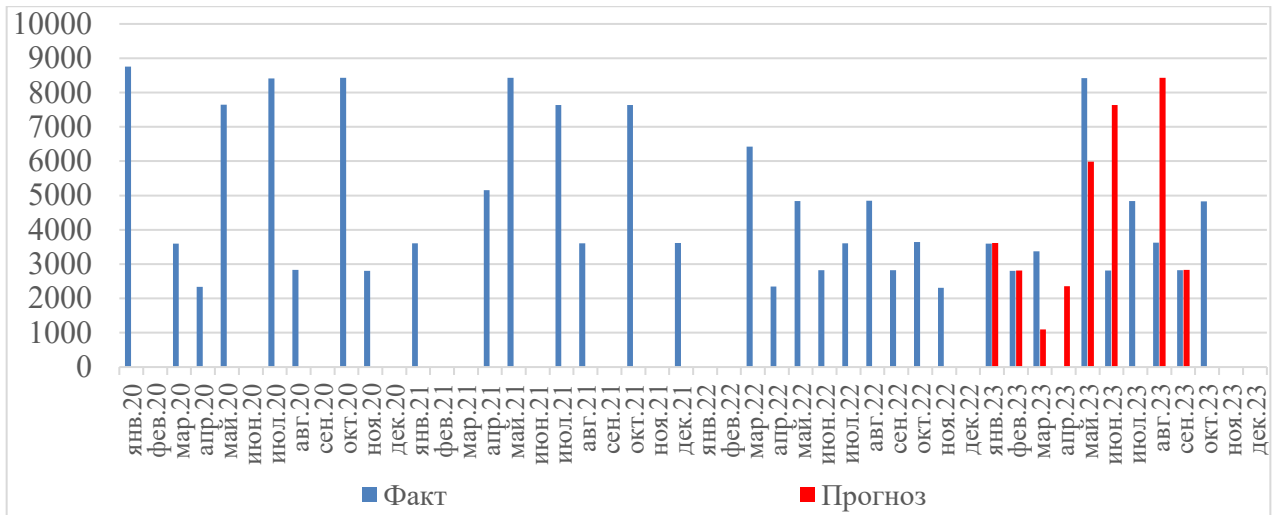
В случае необходимости построения прогноза в виде временного ряда определяем прогноз суммарных продаж по месяцам от всех трех клиентов, рисунок 4. Прогноз на следующие два месяца практически совпадает с фактическим, для более далекой временной перспективы наблюдаются расхождения (для третьего месяца погрешность момента времени в несколько дней привела к тому, что событие одного из клиентов попало в следующий месяц). Однако, возможно, будет предпочтительнее воспользоваться не суммой дискретных покупок по месяцам, а суммарной функцией скоростей потребления, рисунок 5, так как уже будет неважно куда попало прогнозное событие, на конец текущего месяца или уже в следующий, не надо будет строить модель для объемов покупок.



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

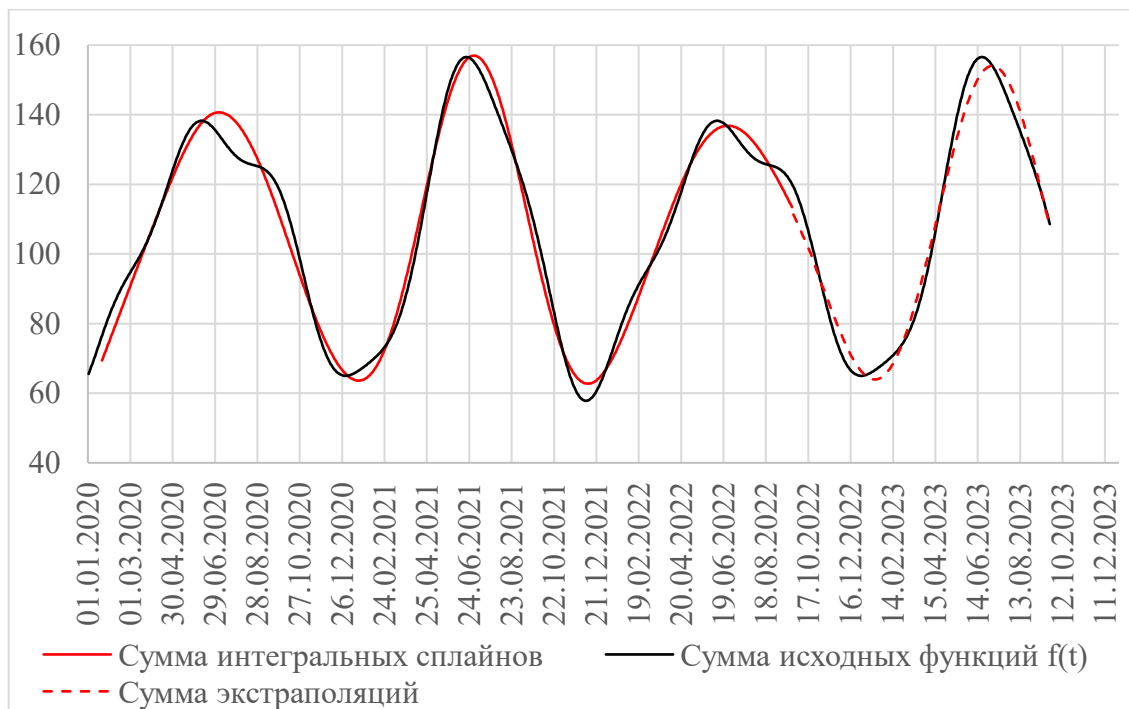
Источник: составлено автором.

Рисунок 3 – Восстановление скорости расхода продукции, экстраполяция и прогноз будущих покупок первого, второго и третьего клиента соответственно, шт. в день



Источник: составлено автором.

Рисунок 4 – Прогноз будущих покупок по месяцам на год вперед, шт. в месяц



Источник: составлено автором.

Рисунок 5 – Суммарная скорость расхода продукции всех трех клиентов, шт. в день

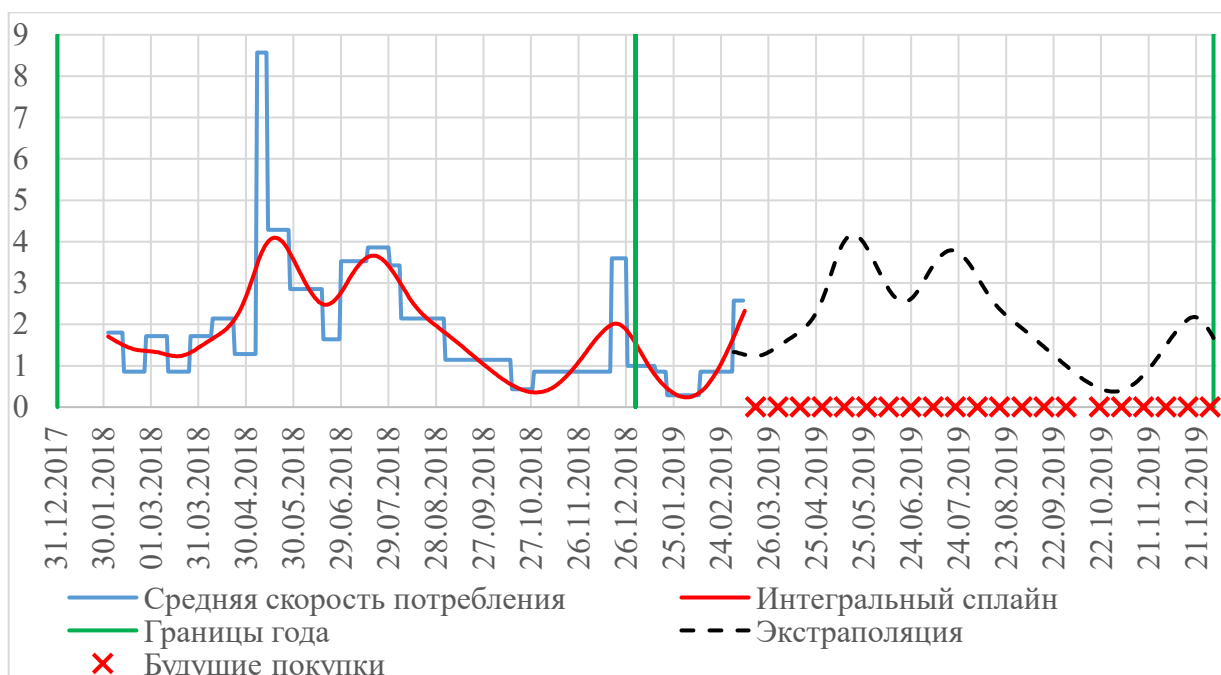
Пример 2. Имеются данные о поставках полторалитровых бутылок кваса в универсам, таблица 2. Для устранения нарушения основной предпосылки несколько событий объединяются в одно  $(t_i, y_i + y_{i+1})$ , если универсам преждевременно пополняет запасы (время между событиями неправдоподобно мало). После восстановления функции скорости потребления бутылок кваса универсамом, рисунок 6, замечаем, что скорость потребления кваса привязана к времени года, в жаркую погоду и к праздникам потребление возрастает (эти размышления ложатся в основу для экстраполяции), тогда экстраполяция осуществляется с помощью переноса динамики прошлого года.

Таблица 2 – Данные о поставках бутылок кваса в универсамах

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.

Из исходных данных видим, что интервалы между пополнениями запаса кратны 7, причем наиболее часто пополнение происходит через 7 или 14 дней. Можно сделать предположение, что универсам старается сделать двухнедельный запас упаковками по 6 бутылок. Действуя по заданной схеме, получается сделать прогноз серии будущих покупок (крестики на горизонтальной оси, рисунок 6). Однако, универсам может действовать независимо от нашего прогноза, мы можем с относительной уверенностью определить только дату самой первой будущей покупки универсама, как момент завершения запаса, сделанного последней покупкой. С другой стороны, появляется возможность самим навязывать данному универсаму объем покупок, чтобы его запасов ему хватало на 2 недели.

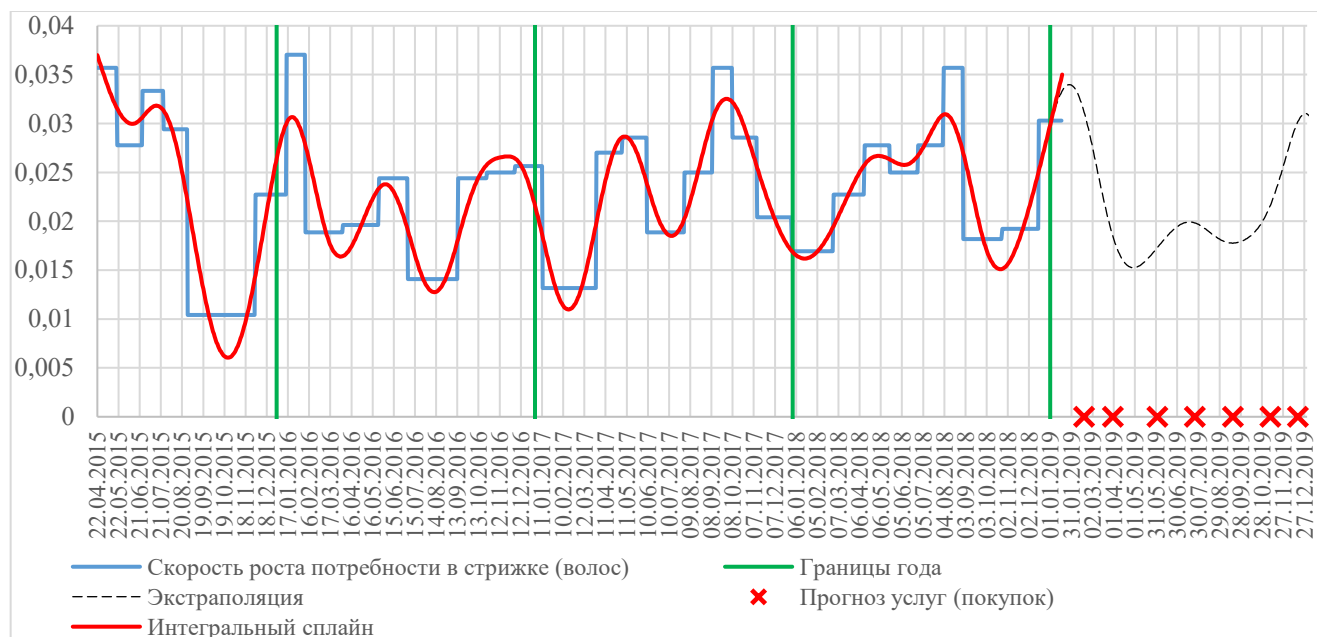


Источник: составлено автором вместе с соавторами <sup>1)</sup>.

Рисунок 6 – Прогноз скорости расхода кваса в универсаме (бутылок в день) и прогноз событий

<sup>1)</sup> Кораблев, Ю.А. Емкостный метод анализа редких событий в торговле различными товарами / Ю.А. Кораблев, П.С. Голованова, Т.А. Кострица // Бизнес. Образование. Право. – 2019. – № 3 (48). – С. 121-131. – ISSN 1990-536X.

Пример 3. Прогноз услуг. Аналогично можно поступать при прогнозировании услуг. Например, по данным о датах (объем приобретенной услуги всегда равен единице  $y_i = 1$ ) стрижки одного постоянного клиента в парикмахерской на протяжении 4 лет можно восстановить некоторую относительную скорость роста волос (относительно ко всей услуге) или скорость роста желания подстричься. Экстраполировав полученную зависимость на будущее любым известным способом (в данном примере опять использовался алгоритм Куинна-Фернандеса), запустив сам процесс образования событий, можно получить прогноз будущих услуг. Результат изображен на рисунке 7.



Источник: составлено автором вместе с соавторами<sup>1)</sup>.

Рисунок 7 – Восстановление и экстраполяция относительной скорости роста волос (относительно ко всей услуге), доли в день, и прогноз приобретения услуги

Сами восстанавливаемые закономерности из данных редких событий могут представлять особенный интерес для специалистов соответствующей прикладной области (восстановленная функция может быть интересна специалистам по физиологии роста волос).

3) Сформулированы предпосылки, условия, допущения, которые должны соблюдаться для корректной работы емкостного метода.

- Прежде всего редкие события действительно должны образовываться с помощью процессов, схожих с процессами опустошения/наполнения емкости.

- Процесс образования событий не должен меняться со временем, то есть не должно быть такого, чтобы часть времени события образуются по одному процессу, а часть совсем по-другому.

<sup>1)</sup> 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.

- В процессе образования событий не должно возникать изменений в количестве параметров или изменение способов влияния этих параметров на ход формирования событий.

- В изменении значений параметров со временем должна быть закономерность, позволяющая экстраполировать эти значения параметров на будущее.

- Исходная функция  $f(t)$  скорости изменения уровня некоторой емкости (запаса) должна быть определена на всем рассматриваемом промежутке времени и не должна быть импульсной (например, она должна быть непрерывной).

- 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  $t_{i+1}$  ( $y_i = \int_{t_i}^{t_{i+1}} f(t)dt$ ), 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 легко скорректировать сдвигом в обратную сторону, достаточно рассматривать данные  $(t_i, y_{i+1})$ .

- Время следующего события  $t_{i+1}$  должно соответствовать моменту, когда полностью опустошилась/наполнилась емкость без существенной задержки или заблаговременного пополнения (эта задержка воспринимается как погрешность).

В зависимости от каждой конкретной прикладной области могут появляться дополнительные условия/предпосылки. Последствием нарушения этих условий/предпосылок будет возрастающая погрешность при восстановлении функции и прогнозе будущих событий, либо же неспособность применить данный метод вовсе.

**В третьей главе** получены следующие результаты.

1) Разработан математический метод восстановления параметров процессов опустошения/наполнения емкости по выборке редких событий, который базируется на сплайновой коллокации, восстанавливающий неизвестную функцию по последовательности интегралов.

Решение получается в виде сплайна  $\hat{f}(t) = g(t)$ , где  $g(t)$  – кубический сплайн. Оптимизационная задача имеет следующий вид, формула (3)

$$L_W(g) = \sum_{i=1}^{n-1} w_i \left( y_i - \int_{t_i}^{t_{i+1}} g(t)dt \right)^2 + \alpha \int_{t_1}^{t_n} (g''(t))^2 dt \rightarrow \min, \quad (3)$$

где  $\int_{t_1}^{t_n} (g''(t))^2 dt$  - штраф на нелинейность;

$\alpha$  – параметр сглаживания;

$w_i$  – вес  $i$ -того наблюдения;

$n$  – количество наблюдений.

Сплайн  $g(t)$  имеет представление через значения и значения второй производной в узлах (value-second derivative representation), имеющий следующий вид, формула (4)

$$g(t) = \frac{(t - s_k)g_{k+1} + (s_{k+1} - t)g_k}{s_{k+1} - s_k} - \frac{1}{6}(t - s_k)(s_{k+1} - t) \left\{ \left(1 + \frac{t - s_k}{s_{k+1} - s_k}\right) \gamma_{k+1} + \left(1 + \frac{s_{k+1} - t}{s_{k+1} - s_k}\right) \gamma_k \right\},$$

$$k: s_k \leq t \leq s_{k+1}, \quad (4)$$

где  $s_1 < s_2 < \dots < s_m$  – узлы сплайна (могут не совпадать с точками наблюдений, количество узлов  $m$  может отличаться от количества наблюдений  $n$ );

$g_k = g(s_k)$  – значение сплайна в узле  $s_k$ ;

$\gamma_k = g''(s_k)$  – значение второй производной в узле  $s_k$ .

Для идентификации сплайна необходимо найти  $g = (g_1, \dots, g_m)^T$  и  $\gamma = (\gamma_2, \dots, \gamma_{m-1})^T$  (из условий натурального сплайна  $\gamma_1 = \gamma_m = 0$ ) из оптимизационной задачи (3).

Если узлы совпадают с наблюдениями  $s_k = t_i$ , то интеграл от сплайна рассчитывается по формуле (5)

$$\int_{t_i}^{t_{i+1}} g(t) dt = g_{k+1} \frac{h_k}{2} + g_k \frac{h_k}{2} - \gamma_{k+1} \frac{h_k^3}{24} - \gamma_k \frac{h_k^3}{24}, \quad (5)$$

где  $h_k = s_{k+1} - s_k$ ,  $k = 1, \dots, m - 1$ .

Но если узлы  $s_k$  выбираются независимо от наблюдений, то расчет усложняется.

Интеграл на интервале от  $t_i$  до  $t_{i+1}$  разбивается на 3 части, формула (6)

$$\int_{t_i}^{t_{i+1}} g(t) dt = \sum_{l=0}^L \int_{s_{k+l}}^{s_{k+l+1}} g(t) dt - \int_{s_k}^{t_i} g(t) dt - \int_{t_{i+1}}^{s_{k+L+1}} g(t) dt,$$

$$L: s_{k+L} < t_{i+1} \leq s_{k+L+1}, \quad (6)$$

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

где  $L$  – это количество интервалов между соседними наблюдениями, которое может быть равно 0, если оба наблюдения выпали на один интервал, индекс  $k$  принимает значение номера интервала, на который выпало  $i$ -ое наблюдение.

Первая, вторая и третья часть рассчитываются следующим образом, формулы (7)–(9)

$$\sum_{l=0}^L \int_{s_{k+l}}^{s_{k+l+1}} g(t) dt = \sum_{l=0}^L \left[ \frac{h_{k+l}}{2} g_{k+l+1} + \frac{h_{k+l}}{2} g_{k+l} - \frac{h_{k+l}^3}{24} \gamma_{k+l+1} - \frac{h_{k+l}^3}{24} \gamma_{k+l} \right], \quad (7)$$

$$\int_{s_k}^{t_i} g(t) dt = \frac{(h_k^-)^2}{2h_k} g_{k+1} + \frac{(h_k)^2 - (h_k^+)^2}{2h_k} g_k + \frac{\gamma_{k+1}}{24h_k} (h_k^-)^2 \left( (h_k^-)^2 - 2(h_k)^2 \right) - \frac{\gamma_k}{24h_k} (h_k^-)^2 (h_k^+ + h_k)^2, \quad (8)$$

$$\int_{t_{i+1}}^{s_{k+L+1}} g(t)dt = \frac{(h_{k+L})^2 - (h_{k+L}^{-(i+1)})^2}{2h_{k+L}} g_{k+L+1} + \frac{(h_{k+L}^{+(i+1)})^2}{2h_{k+L}} g_{k+L} -$$

$$- \frac{\gamma_{k+L+1}}{24h_{k+L}} (h_{k+L}^{+(i+1)})^2 (h_{k+L}^{-(i+1)} + h_{k+L})^2 + \frac{\gamma_{k+L}}{24h_{k+L}} (h_{k+L}^{+(i+1)})^2 \left( (h_{k+L}^{+(i+1)})^2 - 2(h_{k+L})^2 \right), \quad (9)$$

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

Подставляя (7)–(9) в (6), получаем выражение для  $\int_{t_i}^{t_{i+1}} g(t)dt$ .

Форма выражения (5) или выражения (6) будет иметь линейный вид по отношению к неизвестным  $g = (g_1, \dots, g_m)^T$  и  $\gamma = (\gamma_2, \dots, \gamma_{m-1})^T$ . В результате чего оптимизационную задачу (3) можно записать в виде матричного выражения (10)

$$L_W(g) = (Y - Vg + P\gamma)^T W (Y - Vg + P\gamma) + \alpha g^T K g \rightarrow \min, \quad (10)$$

где  $Y = (y_1, \dots, y_{n-1})^T$  – вектор столбец значений  $n - 1$  наблюдения;

$V$  – матрица коэффициентов при неизвестных  $g$  размерностью  $(n - 1) \times m$ ;

$P$  – матрица коэффициентов при неизвестных  $\gamma$  размерностью  $(n - 1) \times (m - 2)$ ;

$W$  – диагональная матрица весов наблюдений размерностью  $(n - 1) \times (n - 1)$ ;

$g^T K g$  – штраф на нелинейность ( $\int_{t_1}^{t_n} (g''(t))^2 dt$ );

$\alpha$  – коэффициент сглаживания.

Матрица  $V$  и  $P$  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). Заполнение  $i$ -той строки матрицы  $V$  и  $P$  происходит по формулам (11) и (12) соответственно

$$V_{i,k} = (h_k^+)^2 / 2h_k;$$

$$V_{i,k+l} = (h_{k+l-1} + h_{k+l}) / 2, \quad l = 1, \dots, L;$$

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

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

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

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

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

$$P_{i,k+1} = P_{i,k+1} + (h_k^-)^2 \left( (h_k^-)^2 - 2h_k^2 \right) / 24h_k; \quad (12)$$

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

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



Элементы  $V_{i,k+1}$ ,  $V_{i,k+L}$  и  $P_{i,k+1}$ ,  $P_{i,k+L}$  после того, как получили значения в цикле по  $l = 1, \dots, L$ , еще раз изменяются. Например, если расстояние между  $i$ -ым и  $i + 1$ -ым наблюдением  $L > 2$ , то  $i$ -ая строка матрицы  $V$  и  $P$  будет выглядеть как в таблице 3 и 4 соответственно.

Таблица 3 – Заполнение  $i$ -той строки матрицы  $V$  при  $L > 2$

$k - 1$	$k$	$k + 1$	$\dots k + l \dots$	$k + L$	$k + L + 1$	$k + L + 2$
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.

Таблица 4 – Заполнение  $i$ -той строки матрицы  $P$  при  $L > 2$

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

Источник: составлено автором.

Матрица  $K = QR^{-1}Q^T$  симметричная матрица размерностью  $m \times m$ , которая строится из матриц  $Q$  и  $R$  (матрицы  $Q$  и  $R$  образуются из условий непрерывности первой производной в узлах сплайна  $g'(s_k - 0) = g'(s_k + 0)$ , группируя в левой части слагаемые с  $g_k$ , а в правой с  $\gamma_k$ , система имеет  $m - 2$  уравнения, компактная запись условий непрерывности  $Q^T g = R\gamma$ ).

Матрица  $Q$  размерностью  $m \times (m - 2)$  и матрица  $R$  размерностью  $(m - 2) \times (m - 2)$  заполняются в зависимости только от расположения узлов сплайна ( $h_k = s_{k+1} - s_k$ ), рисунок 8.

Благодаря условиям непрерывности  $Q^T g = R\gamma$ , оптимизационную задачу (10) можно записать только от одной неизвестной  $g$ , подставив  $\gamma = R^{-1}Q^T g$  и заменив  $C = V - PR^{-1}Q^T$ ,

$$L_W(g) = (Y - Cg)^T W(Y - Cg) + \alpha g^T K g \rightarrow \min. \quad (13)$$

Минимизируя (13), получаем искомые значения  $g$  (вслед за ними и  $\gamma = R^{-1}Q^T g$ ) по формуле (14)

$$g = (C^T W C + \alpha K)^{-1} C^T W Y. \quad (14)$$

Найденные значения  $g$  и  $\gamma$  определяют сплайн  $g(t)$  в любой точке по формуле (4).

	2	3	...	$m-1$
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
...	...	...	...	...
$m-2$		0	...	$h_{m-2}^{-1}$
$m-1$	0	0	...	$-h_{m-2}^{-1} - h_{m-1}^{-1}$
$m$	0	0	...	$h_{m-1}^{-1}$

$Q =$

	2	3	4	...	$m-1$
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
...	...	...	...	...	...
$m-2$	0	0	0	...	$h_{m-2}/6$
$m-1$	0	0	0	...	$(h_{m-2} + h_{m-1})/3$

$R =$

Источник: составлено автором.  
Рисунок 8 – Заполнение матриц  $Q$  и  $R$

Особенностью данного сплайна является возможность задавать произвольные узлы сплайна, не привязываясь к точкам наблюдений.

2) Представлены результаты исследования по выбору оптимального параметра сглаживания  $\alpha$  для данного математического метода.

Оказалось, что Метод  $L$ -кривой не работает. Формулы для расчета оценки кросс-валидации  $CV(\alpha)$  и обобщенной кросс-валидации  $GCV(\alpha)$  следующие, формулы (15) и (16)

$$CV(\alpha) = (n-3)^{-1} \sum_{i=2}^{n-2} w_i \left( \frac{y_i - \sum_{j=1}^m C_{ij} g_j}{1 - \sum_{j=1}^m C_{ij} A_{ji}(\alpha)} \right)^2, \quad (15)$$

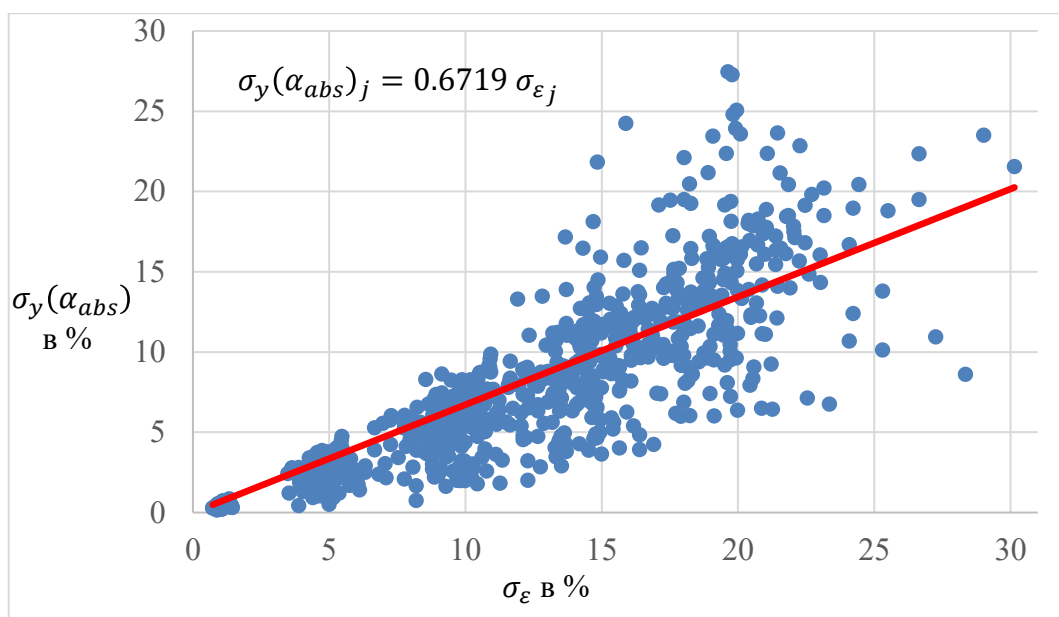
$$GCV(\alpha) = (n-3)^{-1} \frac{\sum_{i=2}^{n-2} w_i (y_i - \sum_{j=1}^m C_{ij} g_j)^2}{(1 - (n-3)^{-1} \sum_{i=2}^{n-2} \sum_{j=1}^m C_{ij} A_{ji}(\alpha))^2}, \quad (16)$$

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

$A(\alpha) = (C^T W C + \alpha K) C^T W$  – матрица проекции (*hat* - матрица).

Однако, метод кросс-валидации работает только в половине случаев, в другой половине приводит либо к очень сильному занижению, либо завышению коэффициента сглаживания  $\alpha$ .

Другим подходом является принцип невязки Морозова, в котором предлагается брать коэффициент сглаживания  $\alpha$  на таком уровне, чтобы он сглаживал наблюдения на величину, соответствующую уровню шума в исходных данных. Однако в проведенном с помощью имитационного моделирования исследовании оказалось, что наилучшее восстановление функции будет при таком коэффициенте сглаживания, который приглушал бы исходные наблюдения на 67,19% от уровня шума (после попытки скорректировать гетероскедастичность на 64,92%), рисунок 9.



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  $\bar{y}_i = \sum_{j=1}^m C_{ij} g_j$  – прогнозируемое значение наблюдаемого интеграла. 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  $\alpha$ .

Источник: составлено автором.

Рисунок 9 – Диаграмма рассеивания и линейная модель для значения приглушения наблюдений при оптимальном коэффициенте сглаживании  $\sigma_y(\alpha_{abs})$  от величины фактического уровня шума в наблюдениях  $\sigma_{\epsilon}$

В итоге сформулирована рекомендация. Если уровень шума в наблюдениях  $\sigma_{\epsilon}$  известен, то следует выбирать такой коэффициент сглаживания  $\alpha$ , который бы приглушал наблюдения на  $\sim 65\%$  от этого уровня шума  $\sigma_{\epsilon}$ , 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, которые бы сообщали, что определенный коэффициент сглаживания оказался слишком занижен или слишком завышен.

3) Разработан программный инструментарий на языке R, реализующий соответствующий математический метод сплайновой коллокации восстановления функции по последовательности интегралов.

Данный математический метод реализован в виде функции `int_spline`, по соответствующей программной реализации 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».

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. Например, расчет матрицы  $V$  изображен на рисунке 10, все расчеты могут занимать более 10 листов.

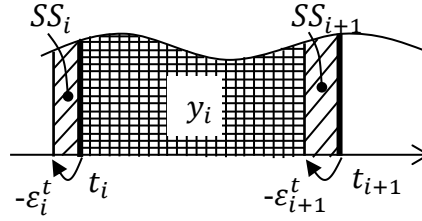
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	WA	WB
1	Матрица V	h=	13,58		=Лист1!H3																			
2	Индекс	$h_k^{-1}$	$h_k^{+1}$	1	2	3	4	5	6	7	8												50	51
3	1	0	13,58	=ЕСЛИ(\$	13,58	13,065	3,565	0	0	0	0												0	0
4				74	0	0,51501	10,015	13,58	13,4994	5,39065	0												0	0
5				48	0	0	0	0	0,08065	8,18935	13,481												0	0
6				64	0	0	0	0	0	0	0,099												0	0
7	10	6,78	6,8	0																		0	0	
8	30	3,22	3,22	0																		0	0	
9	32	3,22	3,22	0																		0	0	
10	33	3,22	3,22	0																		0	0	
11	34	3,22	3,22	0																		0	0	
12	35	3,22	3,22	0																		0	0	
13	36	3,22	3,22	0																		0	0	
14	37	3,22	3,22	0																		0	0	
15	38	3,22	3,22	0																		0	0	
16	39	3,22	3,22	0																		0	0	
17	40	3,22	3,22	0																		0	0	
18	41	4,8	8,78	0																		0	0	
19	42	4,8	8,78	0																		0	0	
20	43	0,64	12,94	0																		0	0	
21	44	8,06	5,52	0																		0	0	
22	45	8,06	5,52	0																		0	0	
23	46	1,9	11,68	0																		0	0	
24	47	1,9	11,68	0																		0	0	
25	48	1,74	11,84	0																		13,58	6,79	
26	49	1,74	11,84	0																				
27	51																							

Источник: составлено автором.  
Рисунок 10 - Расчет матрицы  $V$  в Excel

**В четвертой главе** изучаются вопросы точности и происходит сравнение емкостного метода с другими существующими методами.

1) Показано как появляется погрешность наблюдений, связанная с тем, что время измеряется дискретно (например, по дням).

В процессе опустошения/наполнения емкости возникает дополнительная погрешность наблюдений, связанная с тем, что время изменяется дискретно. Событие не может образоваться (зафиксироваться) именно в тот момент, когда уровень емкости опустился/превысил критический уровень, а из-за дискретности времени (например, по дням) образуется чуть позже. То есть следующее событие должно было быть датировано моментом времени  $t_{i+1} - \varepsilon_{i+1}^t$ , где  $\varepsilon_{i+1}^t$  погрешность из-за округления времени, но датируется моментом  $t_{i+1}$ . Получается, что за период от  $t_{i+1} - \varepsilon_{i+1}^t$  до  $t_{i+1}$  происходит перерасход на величину  $SS_{i+1}$ , которая, например, в моделях управления запасами может показывать количество продукции, которая берется из страховых запасов. Аналогично происходит со всеми событиями, на рисунке 11 проиллюстрирована эта картина. Но проблема не в том, что используются страховые запасы, а в том, что фактически расходуемый объем до очередного события будет не совпадать с наблюдаемым объемом, так как тот откладывается для компенсации этого перерасхода  $SS_{i+1}$ .



Источник: составлено автором.

Рисунок 11 – Несовпадение наблюдаемого объема  $y_i + SS_i$  и расходуемого  $y_i + SS_{i+1}$

Величина несовпадения наблюдаемого и расходуемого объема  $\epsilon_i^Q$  за наблюдаемый период от  $t_i$  до  $t_{i+1}$  есть разница величин  $SS_i$  и  $SS_{i+1}$ , формула (17), которые распределены равномерно от 0 до  $f(t_i)$  и  $f(t_{i+1})$  соответственно

$$\epsilon_i^Q = SS_i - SS_{i+1} \in (-f(t_{i+1}), f(t_i)). \quad (17)$$

Величина несовпадения  $\epsilon_i^Q$  будет распределена по трапецидальному закону, причем среднее значение не обязательно равно нулю (ноль, только если  $f(t)$  стационарна или когда  $f(t_i) = f(t_{i+1})$  для всех наблюдений). На участках, где  $f(t)$  возрастает, она будет в среднем отрицательна, а где убывает - в среднем положительна.

Данная величина  $\epsilon_i^Q$  приводит к дополнительной погрешности наблюдений, которая прибавляется к (или содержится в) погрешности наблюдений  $\epsilon_i$  (в наблюдениях  $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \epsilon_i$ ). Соответствующие расчеты подтверждены на численном примере.

2) Для процесса потребления показана какая появляется дополнительная погрешность в наблюдениях, если само потребление осуществляется конечными потребителями, удаленными на заданное количество позиций (посредников) от точки сбора данных (в цепочке распространителей).

Получено выражение для расчета дополнительной относительной погрешности  $\epsilon_i^Q/y_i$ , возникающей, когда конечный потребитель (не подконтрольный нам клиент) удален на несколько промежуточных распространителей от нашего продавца (то есть появляются посредники), где величина  $\epsilon_i^Q$  показывает несовпадение наблюдаемого и фактически расходуемого объема продукции (всеми конечными потребителями).

Величина  $\epsilon_i^Q/y_i$  будет иметь приблизительно нормальное распределение (так как участвует слишком много случайных величин) с нулевым математическим ожиданием. Дисперсия такой величины будет расти как сумма убывающей геометрической прогрессии, а среднее квадратичная погрешность будет следующей, формула (18)

$$\sigma_{\frac{\epsilon_i^Q}{y_i}}(N) = \sqrt{\frac{n+1}{6P^2} \left( 1 + \frac{n^1}{P^2} + \frac{n^2}{(P^2)^2} + \dots + \frac{n^{N-1}}{(P^2)^{N-1}} \right)} = \sqrt{\frac{(n+1) \left( 1 - \left( \frac{n}{P^2} \right)^N \right)}{6P^2 \left( 1 - \frac{n}{P^2} \right)}}, \quad (18)$$

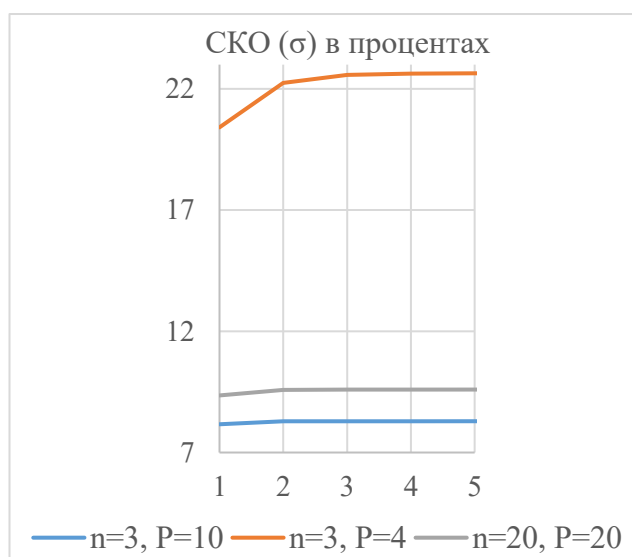
где  $N$  – количество посредников (промежуточных распространителей);  
 $n$  – количество покупателей на каждом уровне распространителей;  
 $P$  – во сколько раз увеличивается запас вышестоящего распространителя.

Пример расчета дисперсии и среднего квадратичного отклонение для относительной погрешности  $\varepsilon_i^0/y_i$  представлен в таблице 5, график  $\sigma_{\varepsilon^0/y_i}(N)$  представлен на рисунке 12. Как видно из таблицы и графика, относительная погрешность растет незначительно, уже при удаленности на 3 позиции погрешность изменяется лишь в четвертом знаке.

Таблица 5 – Расчет дисперсии и среднего квадратичного отклонения относительной ошибки при удалении от конечного потребителя  $N$

$N$	$n = 3, P = 10$		$n = 3, P = 4$		$n = 20, P = 20$	
	$D$	$\sigma$ в процентах	$D$	$\sigma$ в процентах	$D$	$\sigma$ 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 12 – Относительная погрешность в зависимости от удаленности от конечных потребителей (длины цепочки посредников), в процентах

Данная относительная погрешность  $\varepsilon_i^0/y_i$  опять приводит к дополнительной погрешности наблюдений, которая прибавляется к (или уже содержится в) погрешности наблюдений  $\varepsilon_i$  (в наблюдениях за  $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$ ). Соответствующие расчеты также подтверждены на численном примере.

3) Определено, как растет погрешность восстановления функции от потери части наблюдений и показано, что погрешность при использовании емкостного метода растет медленнее, чем у классического подхода, когда по наблюдениям строится временной ряд.

С помощью имитационного моделирования и серии экспериментов показано, что в случае потери части данных (10%, 30% и 50%) относительная погрешность  $R$  и  $\sigma$  восстановления исходной функции значительно меньше у емкостного метода, чем если бы использовался классический подход, когда события агрегировались бы во временной ряд и из него находилась скорость расхода продукции. Погрешности  $R$  и  $\sigma$  рассчитывались по формулам (19) и (20)

$$R = \frac{1}{n_t} \sum_{t=t_{\text{старт}}}^{t_{\text{конец}}} \frac{|\bar{f}(t) - f(t)|}{f(t)}, \quad (19)$$

$$\sigma = \sqrt{\frac{1}{n_t} \sum_{t=t_{\text{старт}}}^{t_{\text{конец}}} \left( \frac{\bar{f}(t) - f(t)}{f(t)} \right)^2}, \quad (20)$$

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.

Так, например, при потере 10% данных средняя по 20 прогонам погрешность  $R$  была 15,09% против 39,90%, при потере 30% данных погрешность  $R$  была 31,63% против 46,64%, а при потере 50% данных наблюдений средняя по 20 прогонам погрешность  $R$  составила 48,63% у емкостного метода против 68,97% у классического, основанного на построении временного ряда.

Также с помощью математического анализа удалось показать, 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 (отказываться от агрегирования по интервалам и начинать рассматривать каждое событие с точки зрения процесса образования событий).

4) Определено влияние неопределенности внутри самого процесса образования событий на точность восстановления скорости опустошения/наполнения емкости (скорости потребления продукции).

Проведено 3 серии экспериментов. В первой серии экспериментов определялась погрешность  $R$  восстановления исходной функции  $f(t)$  скорости опустошения/наполнения емкости (например, скорость потребления продукции) в зависимости от уровня шума в самих

наблюдениях (уровня шума  $\sigma^{\text{набл}}$  у погрешности  $\varepsilon_i$  в наблюдениях  $y_i$ ). Во второй серии экспериментов определялась та же погрешность  $R$  в зависимости от неопределенности внутри процесса образования событий, когда в модели управления запасами при проверке собственных запасов клиент оценивал их с погрешностью  $\varepsilon_i$  (в зависимости от уровня шума  $\sigma^{\text{кр.ур}}$  погрешности  $\varepsilon_i$  при проверке условия  $Stock \leq CriticalStorage \times (1 + \varepsilon_i)$ , отвечающего за образование события). В третьей группе экспериментов определялась зависимость погрешности восстановления исходной функции  $R$  от неопределенности выбора момента времени возникновения события, например, когда клиент задерживает либо опережает время пополнения запаса (в зависимости от уровня шума  $\sigma^{\Delta t_i}$  погрешности  $\varepsilon_i$  при определении даты события  $\hat{t}_i = t_i + \Delta t_i \cdot \varepsilon_i$ , где  $\Delta t_i = t_i - t_{i-1}$ ). Во всех этих экспериментах под погрешностью  $R$  понимается средняя относительная погрешность восстановления исходной функции (в процентах), формула (21)

$$R = \frac{1}{T} \sum_{t=1}^T \frac{|\hat{f}(t) - f(t)|}{f(t)} \times 100\%, \quad (21)$$

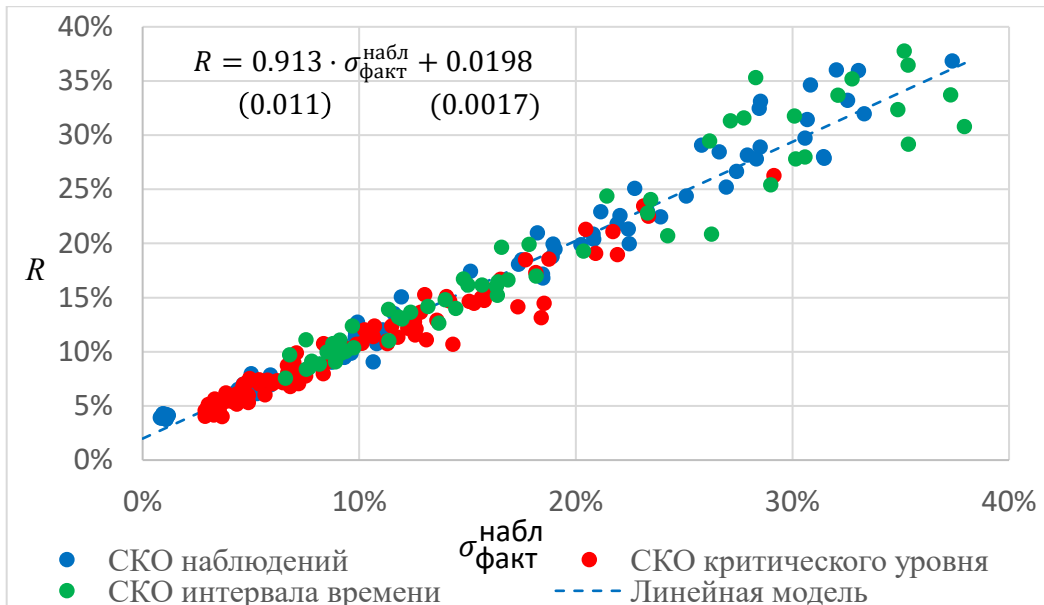
где  $f(t)$  – исходная функция;

$\hat{f}(t)$  – восстановленная с помощью интегрального сплайна функция;

$T$  – общее количество сравнений (дней в примере модели управления запасами).

По результатам исследования во всех трех сериях экспериментов показано, что все неопределенности внутри процесса можно свести к неопределенности непосредственно самих наблюдений  $y_i$ . Во время экспериментов рассчитывалась какая получается фактическая относительная погрешность  $(y_i - I_i)/I_i$  у самого наблюдения  $y_i$  по сравнению с истинным значением интеграла  $I_i = \int_{t_i}^{t_{i+1}} f(t) dt$ , после чего определялся уровень этой фактической погрешности  $\sigma_{\text{факт}}^{\text{набл}}$ . Определено влияние уровня неопределенности  $\sigma^{\text{кр.ур}}$  и  $\sigma^{\Delta t_i}$  на неопределенность наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$ . Также по результатам всех трех серий экспериментов определена зависимость погрешности восстановления исходной функции  $R$  от фактического уровня шума в наблюдениях  $\sigma_{\text{факт}}^{\text{набл}}$ . Полученная зависимость оказалась линейной, диаграмма рассеивания и построенная линейная модель изображены на рисунке 13.



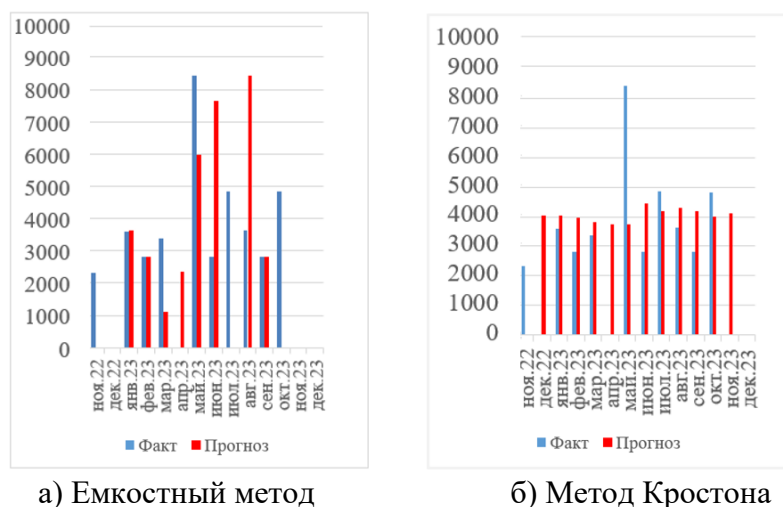


Источник: составлено автором.

Рисунок 13 – Диаграмма рассеивания погрешности  $R$  восстановления исходной функции от фактического СКО наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$  для всех трех экспериментов

5) Проведено сравнение емкостного метода анализа и прогнозирования редких событий с другими существующими методами, использующимися в экономике.

Проведено сравнение емкостного метода с методом Кростона и Виллимейна, которые используются в логистике при прогнозировании прерывистого (редкого) спроса. Показано, что с помощью емкостного метода (агрегировав события по месяцам) можно получить более точный прогноз спроса на следующие несколько месяцев (на далекую перспективу точность падает), рисунок 14, точнее определить скорость потребления в день, рисунок 15, а также на основе емкостного метода можно спланировать запасы для каждого конкретного события (к будущей покупке каждого клиента), что позволит держать запасы на более низком уровне (возможность меньше хранить запасов освобождает финансовые средства для других целей), рисунок 16.

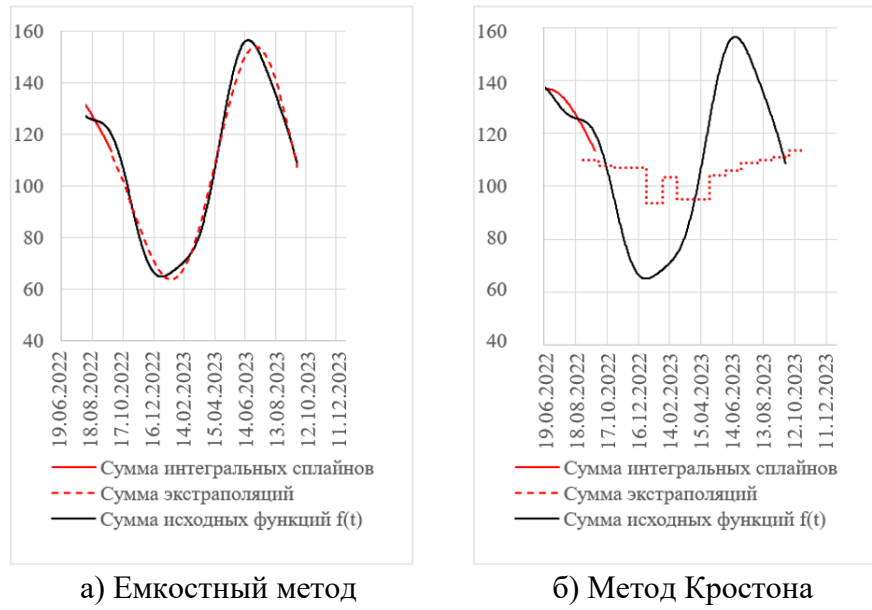


а) Емкостный метод

б) Метод Кростона

Источник: составлено автором.

Рисунок 14 – Сравнение прогноза спроса на месяц с методом Кростона



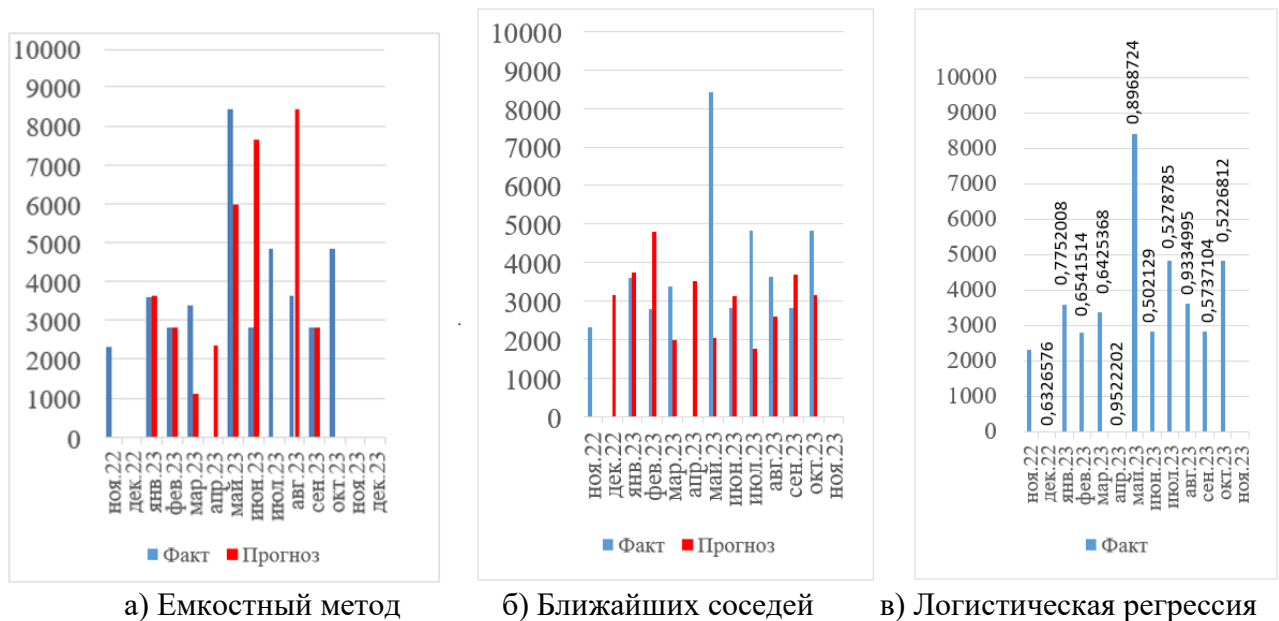
Источник: составлено автором.

Рисунок 15 – Потребление в день, сравнение с методом Кростона, шт. в день



Источник: составлено автором.

Рисунок 16 – Динамика запасов у нашего продавца, который подготавливает запасы в соответствии с указанным методом, штук



Источник: составлено автором.

Рисунок 17 – Прогноз на месяц, шт. (вероятность ненулевого спроса в следующем месяце)

Проведено сравнение емкостного метода с методом ближайших соседей и логистической регрессией. Емкостный метод прогнозирует события в следующем месяце лучше, чем оба эти метода (тут логистическая регрессия определяет только вероятность ненулевого спроса), рисунок 17.

б) Разработана методика проверки адекватности получающихся моделей прогнозирования редких событий, которая опирается на построение доверительного интервала для момента времени наступления следующего события.

В исследовании получена зависимость погрешностей восстановления исходной функции  $R_{abs}(a_{abs})$  или  $R_{sqr}(a_{sqr})$  при оптимальных коэффициентах сглаживания  $a_{abs} = \arg \min_{\alpha} R_{abs}(\alpha)$  и  $a_{sqr} = \arg \min_{\alpha} R_{sqr}(\alpha)$  от уровня приглушения наблюдений  $\sigma_y(a_{abs})$  или  $\sigma_y(a_{sqr})$ , рассчитываемых по формулам (22)-(24)

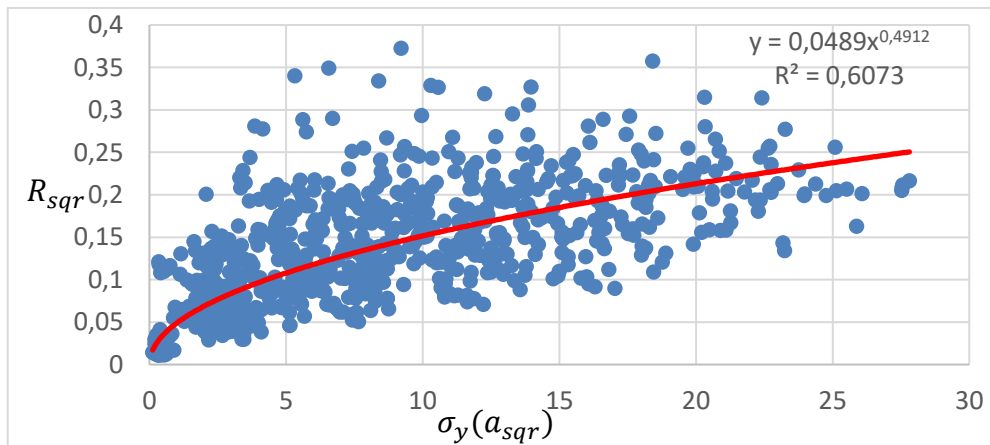
$$R_{abs}(\alpha) = \frac{1}{t_n - t_1} \sum_{t=t_1}^{t_n} \left| \frac{f(t) - g(t, \alpha)}{f(t)} \right|, \quad (22)$$

$$R_{sqr}(\alpha) = \sqrt{\frac{1}{t_n - t_1} \sum_{t=t_1}^{t_n} \left( \frac{f(t) - g(t, \alpha)}{f(t)} \right)^2}, \quad (23)$$

$$\sigma_y(\alpha) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} \left( \frac{y_i - \bar{y}_i}{\bar{y}_i} \right)^2} \times 100\%, \quad (24)$$

где  $f(t)$  – исходная функция;  
 $g(t, \alpha)$  – оценка сплайна при выбранном коэффициенте сглаживания;  
 $\bar{y}_i$  – оценка интеграла от восстановленной функции.

Зависимость оказалась степенной, рисунок 18.



Источник: составлено автором.

Рисунок 18 – Зависимость погрешности восстановления исходной функции  $R_{sqr}$  от уровня приглушения наблюдений  $\sigma_y$  при оптимальном коэффициенте сглаживания  $a_{sqr}$

Получены формулы для расчета доверительного интервала момента времени следующего события, которые основываются на величине приглушения данных наблюдения  $\sigma_y$  на этапе восстановления параметров процесса образования событий. Дисперсия интервала времени до следующего события  $D(\widehat{\Delta t})$  определяется в зависимости от известного значения величины приглушения наблюдений  $\sigma_y$  следующим образом, формула (25)

$$\begin{aligned} D(\widehat{\Delta t}) &= (\widehat{R}_{sqr})^2 + (\widehat{R}_{sqr})^2 \times D(\text{Ln}(\widehat{R}_{sqr})) = (\widehat{R}_{sqr})^2 \times \left(1 + D(\text{Ln}(\widehat{R}_{sqr}))\right) = \\ &= (0,04892 \times \sigma_y^{0,49116})^2 \left(1 + 0,02948^2 + 0,01414^2 \times [\text{Ln}(\sigma_y)]^2 + 0,41482^2\right) = \\ &= (0,04892 \times \sigma_y^{0,49116})^2 \left(1,17294 + 0,01414^2 \times [\text{Ln}(\sigma_y)]^2\right). \end{aligned} \quad (25)$$

Так, если уровень приглушения наблюдений  $\sigma_y$  составит 5%, то расчет  $\sqrt{D(\widehat{\Delta t})}$  дает 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  $\hat{t}_{n+1}$  (где  $\widehat{\Delta t} = \hat{t}_{n+1} - t_n$  - оценка времени до следующего события), распределение считаем нормальным (слишком много участвует факторов). Например, доверительный интервал в 95% будет следующим, формула (26)

$$t_{n+1} \in \left[ \hat{t}_{n+1} - 2\widehat{\Delta t} \times D(\widehat{\Delta t})^{0,5}; \hat{t}_{n+1} + 2\widehat{\Delta t} \times D(\widehat{\Delta t})^{0,5} \right]. \quad (26)$$

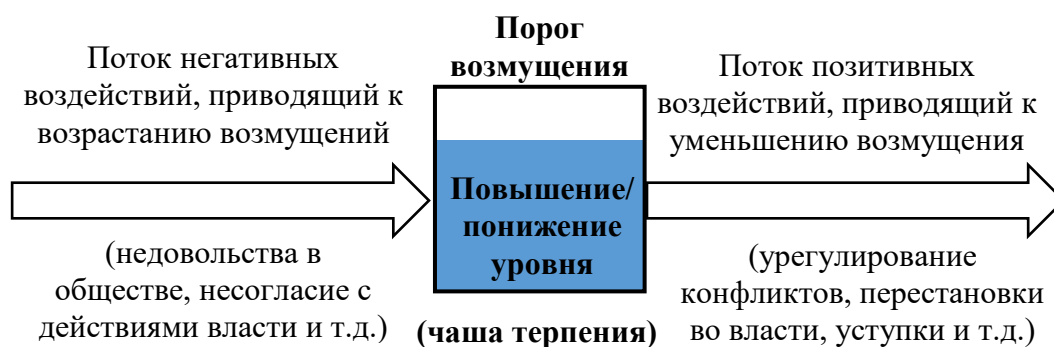
Так при уровне приглушения наблюдений  $\sigma_y$  равным 5%, доверительный интервал составляет  $\pm 23,4\%$  (при  $\sigma_y = 10\%$  получим  $\pm 32,84\%$ ) от интервала  $\widehat{\Delta t}$ , откладываемый от момента  $\hat{t}_{n+1}$ . В рассмотренных во второй главе примерах ретроспективный прогноз для последнего события попадает в соответствующий доверительный интервал относительно фактического события.

Если будущее событие  $t_{n+1}$  не попадает в рассчитанный доверительный интервал, 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.

**В пятой главе** происходит обобщение идеи емкостного метода на произвольные процессы.

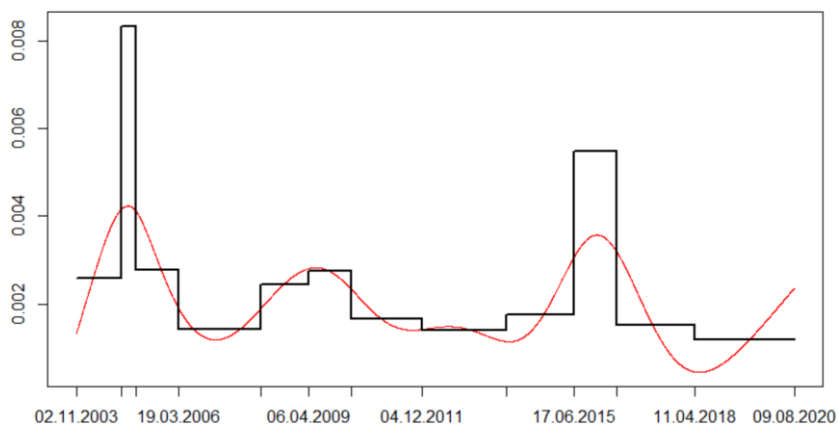
1) Произведено обобщение емкостного метода для анализа исторических, социальных и политических событий на примере «цветных революций» и русско-турецких войн. Модель процесса образования дискретного события основана на накоплении негативного возмущения и достижения порога возмущения (переполнения чаши терпения) вследствие дисбаланса негативного и положительного потока воздействий. Пример модели образования социальных

событий изображен на рисунке 19. С использованием некоторых допущений (события образованы одним процессом) можно восстановить результирующий поток (разницы между потоком положительных и негативных возмущений)  $f(t)$ . В зависимости от того, что выбирается за величину  $y_i$  (сам факт, количество участников, длительность, результат), восстановленный результирующий поток может иметь разную интерпретацию, например, если использовать сам факт появления события ( $y_i = 1$  для всех наблюдений), то восстановленная зависимость, рисунок 20, может иметь интерпретацию как относительная скорость роста недовольства или относительная скорость подготовки к попытке смены власти (относительная, так как выражена в долях ко всему событию).



Источник: составлено автором вместе с соавторами<sup>1)</sup>.

Рисунок 19 – Модель процесса образования социальных событий



Источник: составлено автором вместе с соавторами<sup>2)</sup>.

Рисунок 20 – Восстановление результирующего потока по факту возникновения события (относительная скорость нарастания недовольства в обществе, скорость подготовки к попытке смены власти), доля в день

Данный подход к анализу социальных и исторических процессов, конечно же, сильно упрощает сложные процессы (высокий уровень абстракции), происходящие в реальности, однако он все равно может быть интересен специалистам (историкам, социологам), занимающимся

<sup>1)</sup> 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.

<sup>2)</sup> там же

изучением соответствующих событий, им может пригодиться математический метод для обоснования выдвинутых гипотез.

2) Для восстановления параметров сложных процессов образования событий разработан математический метод восстановления функции по разным функционалам.

Образующиеся процессами события могут нести информацию о значениях некоторой функции, ее первой и второй производной, а также информацию о значениях определенного интеграла.

Входными данными могут являться:

$$y_i = f(t_i) + \varepsilon_i, \quad i = 1, \dots, n_f,$$

$$y'_j = f'(t_j) + \varepsilon'_j, \quad j = 1, \dots, n_{df},$$

$$y''_l = f''(t_l) + \varepsilon''_l, \quad l = 1, \dots, n_{d^2f},$$

$$Y_u = \int_{t_u^a}^{t_u^b} f(t)dt + \varepsilon_u^{int}, \quad u = 1, \dots, n_{int},$$

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 у значений, первой производной, второй производной и интегралов соответственно (с нулевым математическим ожиданием, дисперсии могут быть разными);

$n_f, n_{df}, n_{d^2f}, n_{int}$  – соответственно количество наблюдений значений, первых производных, вторых производных, интегралов искомой функции (могут не совпадать, могут быть равны нулю, если таких данных нет).

Решение основывается на сплайновой коллокации, где так же, как и в третьей главе используется кубический сплайн со штрафной функцией в представлении через значение и вторую производную (value-second derivative representation). Оптимизационная задача имеет следующий вид, формула (27)

$$S(f) = \sum_{i=1}^{n_f} w_i^f (y_i - f(t_i))^2 + \mu \sum_{j=1}^{n_{df}} w_j^{df} (y'_j - f'(t_j))^2 + \nu \sum_{l=1}^{n_{d^2f}} w_l^{d^2f} (y''_l - f''(t_l))^2 + \psi \sum_{u=1}^{n_{int}} w_u^{int} \left( Y_u - \int_{t_u^a}^{t_u^b} f(t)dt \right)^2 + \alpha \int_{t_{start}}^{t_{end}} (f''(t))^2 dt, \quad (27)$$

где  $w_i^f, w_j^{df}, w_l^{d^2f}, w_u^{int}$  – индивидуальные веса соответствующих групп наблюдений;

$\mu$  (мю) – вес всей группы наблюдений первых производных;

$\nu$  (ню) – вес всей группы наблюдений вторых производных;

$\psi$  (пси) – вес всей группы наблюдений интегралов (весовой коэффициент для обычных значений отсутствует, он предполагается равным 1, остальные коэффициенты показывают вес по сравнению с этой первой группой наблюдений);

последнее слагаемое - штраф на нелинейность (шероховатость);

$\alpha$  – коэффициент сглаживания (регуляризации);

$t_{start}$  и  $t_{end}$  – границы, в которых происходит восстановление функции.

Неизвестную функцию  $f(t)$  восстанавливаем в виде кубического сплайна  $g(t)$ , который имеет тот же вид, что в третьей главе, формула (4). Вычисление соответствующих функционалов будет следующим, формулы (28)-(31)

$$g(t_i) = \frac{h_k^{-i}}{h_k} g_{k+1} + \frac{h_k^{+i}}{h_k} g_k - \frac{h_k^{-i} h_k^{+i} (h_k + h_k^{-i})}{6h_k} \gamma_{k+1} - \frac{h_k^{-i} h_k^{+i} (h_k + h_k^{+i})}{6h_k} \gamma_k, \quad (28)$$

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

$$g'(t_j) = \frac{g_{k+1}}{h_k} - \frac{g_k}{h_k} - \left( \frac{h_k}{6} - \frac{(h_k^{-j})^2}{2h_k} \right) \gamma_{k+1} + \left( \frac{h_k}{6} - \frac{(h_k^{+j})^2}{2h_k} \right) \gamma_k, \quad (29)$$

$k: s_k \leq t_j < s_{k+1};$

$$g''(t_l) = \frac{h_k^{-l}}{h_k} \gamma_{k+1} + \frac{h_k^{+l}}{h_k} \gamma_k, \quad (30)$$

$k: s_k \leq t_l < s_{k+1};$

$$\int_{t_u^a}^{t_u^b} g(t) dt = \sum_{l=0}^L \left[ \frac{h_{k+l}}{2} g_{k+l+1} + \frac{h_{k+l}}{2} g_{k+l} - \frac{h_{k+l}^3}{24} \gamma_{k+l+1} - \frac{h_{k+l}^3}{24} \gamma_{k+l} \right] -$$

$$- \frac{(h_k^{-a})^2}{2h_k} g_{k+1} - \frac{h_k^2 - (h_k^{+a})^2}{2h_k} g_k - \frac{(h_k^{-a})^2 ((h_k^{-a})^2 - 2h_k^2)}{24h_k} \gamma_{k+1} + \frac{(h_k^{-a})^2 (h_k^{+a} + h_k)^2}{24h_k} \gamma_k -$$

$$- \frac{h_{k+L}^2 - (h_{k+L}^{-b})^2}{2h_{k+L}} g_{k+L+1} - \frac{(h_{k+L}^{+b})^2}{2h_{k+L}} g_{k+L} +$$

$$+ \frac{(h_{k+L}^{+b})^2 (h_{k+L}^{-b} + h_{k+L})^2}{24h_{k+L}} \gamma_{k+L+1} - \frac{(h_{k+L}^{+b})^2 ((h_{k+L}^{+b})^2 - 2h_{k+L}^2)}{24h_{k+L}} \gamma_{k+L}, \quad (31)$$

$L: s_{k+L} < t_u^b \leq s_{k+L+1},$   
 $k: s_k \leq t_u^a < s_{k+1},$

где  $h_k^{-i}, h_k^{-j}, h_k^{-l}$  – расстояние от узла до наблюдения;

$h_k^{+i}, h_k^{+j}, h_k^{+l}$  – расстояние от наблюдения до следующего узла;

$h_k^{-a} = t_u^a - s_k$  – расстояние от узла до нижнего предела интегрирования;

$h_k^{+a} = s_{k+1} - t_u^a$  – расстояние от нижнего предела интегрирования до следующего узла;

$h_{k+L}^{-b} = t_u^b - s_{k+L}$  – расстояние от узла до верхнего предела интегрирования;

$h_{k+L}^{+b} = s_{k+L+1} - t_u^b$  – расстояние от верхнего предела интегрирования до следующего узла.

Во всех этих выражениях в начале определяется на какой интервал  $k$  выпало наблюдение. В самом последнем выражении для интеграла необходимо 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  $\gamma_k$ . Поэтому оптимизационную задачу можно выразить в матричном виде (32)

$$\begin{aligned} S(g) = & (Y_f - V_f g + P_f \gamma)^T W_f (Y_f - V_f g + P_f \gamma) + \\ & + \mu (Y_{df} - V_{df} g + P_{df} \gamma)^T W_{df} (Y_{df} - V_{df} g + P_{df} \gamma) + \\ & + \nu (Y_{d^2f} - 0g + P_{d^2f} \gamma)^T W_{d^2f} (Y_{d^2f} - 0g + P_{d^2f} \gamma) + \\ & + \psi (Y_{int} - V_{int} g + P_{int} \gamma)^T W_{int} (Y_{int} - V_{int} g + P_{int} \gamma) + \\ & + \alpha g^T K g \rightarrow \min, \end{aligned} \quad (32)$$

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  $\gamma_k$ ;

$W_f, W_{df}, W_{d^2f}, W_{int}$  - диагональные матрицы индивидуальных весов наблюдений.

Матрицы заполняются следующим образом (в зависимости от  $L$  некоторые элементы матриц  $V$  и  $P$  могут изменяться дважды).

$V_f$  размерностью  $n_f \times m$ , ее каждая  $i$ -тая строка выглядит как

$$\begin{array}{cccccccc} 1 & \dots & k-1 & k & k+1 & k+2 & \dots & m \\ \hline 0 & \dots & 0 & h_k^{+i}/h_k & h_k^{-i}/h_k & 0 & \dots & 0 \end{array}$$

$P_f$  размерностью  $n_f \times (m - 2)$ , ее каждая  $i$ -тая строка выглядит как

$$\begin{array}{cccccccc} 2 & \dots & k-1 & & k & & k+1 & & k+2 & \dots & m-1 \\ \hline 0 & \dots & 0 & h_k^{-i} h_k^{+i} (h_k + h_k^{+i}) / 6h_k & h_k^{-i} h_k^{+i} (h_k + h_k^{-i}) / 6h_k & & 0 & \dots & 0 \end{array}$$

$V_{df}$  размерностью  $n_{df} \times m$ , ее каждая  $j$ -тая строка выглядит как

$$\begin{array}{cccccccc} 1 & \dots & k-1 & k & k+1 & k+2 & \dots & m \\ \hline 0 & \dots & & 0 & -1/h_k & 1/h_k & & 0 & \dots & & 0 \end{array}$$

$P_{df}$  размерностью  $n_{df} \times (m - 2)$ , ее каждая  $j$ -тая строка выглядит как

$$\begin{array}{cccccccc} 2 & \dots & k-1 & & k & & k+1 & & k+2 & \dots & m-1 \\ \hline 0 & \dots & 0 & -h_k/6 + (h_k^{+j})^2 / 2h_k & h_k/6 + (h_k^{-j})^2 / 2h_k & & 0 & \dots & 0 \end{array}$$

$P_{d^2f}$  размерностью  $n_{d^2f} \times (m - 2)$ , ее каждая  $l$ -тая строка выглядит как

$$\begin{array}{cccccccc} 2 & \dots & k-1 & k & k+1 & k+2 & \dots & m-1 \\ \hline 0 & \dots & 0 & -h_k^{+l}/h_k & -h_k^{-l}/h_k & 0 & \dots & 0 \end{array}$$

$V_{int}$  размерностью  $n_{int} \times m$ , ее каждая  $u$ -тая строка заполняется по формулам (33)-(34)



$$V_{u,k} = \frac{(h_k^{+a})^2}{2h_k}; V_{u,k+l} = \frac{h_{k+l-1} + h_{k+l}}{2}, l = 1, \dots, L; V_{u,k+1} = V_{u,k+1} - \frac{(h_k^{-a})^2}{2h_k};$$

$$V_{u,k+L} = V_{u,k+L} - \frac{(h_{k+L}^{+b})^2}{2h_{k+L}}; V_{u,k+L+1} = \frac{(h_{k+L}^{-b})^2}{2h_{k+L}}. \quad (33)$$

$P_{int}$  размерностью  $n_{int} \times (m - 2)$ , ее каждая  $u$ -тая строка заполняется по формулам (34):

$$P_{u,k} = \frac{h_k^3}{24} - \frac{(h_k^{-a})^2(h_k^{+a} + h_k)^2}{24h_k}; P_{u,k+l} = \frac{h_{k+l-1}^3 + h_{k+l}^3}{24}, l = 1, \dots, L;$$

$$P_{u,k+1} = P_{u,k+1} + \frac{(h_k^{-a})^2((h_k^{-a})^2 - 2h_k^2)}{24h_k}; P_{u,k+L} = P_{u,k+L} + \frac{(h_{k+L}^{+b})^2((h_{k+L}^{+b})^2 - 2h_{k+L}^2)}{24h_{k+L}}; \quad (34)$$

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

Решение оптимизационной задачи (32) будет следующим, формула (35)

$$g = \left( C_f^T W_f C_f + \mu C_{df}^T W_{df} C_{df} + \nu C_{d^2f}^T W_{d^2f} C_{d^2f} + \psi C_{int}^T W_{int} C_{int} + \alpha K \right)^{-1} \times$$

$$\times \left( C_f^T W_f Y_f + \mu C_{df}^T W_{df} Y_{df} + \nu C_{d^2f}^T W_{d^2f} Y_{d^2f} + \psi C_{int}^T W_{int} Y_{int} \right) \quad (35)$$

где  $C_f = V_f - P_f R^{-1} Q^T$  ;  $C_{df} = V_{df} - P_{df} R^{-1} Q^T$  ;  $C_{d^2f} = 0 - P_{d^2f} R^{-1} Q^T$  ;  
 $C_{int} = V_{int} - P_{int} R^{-1} Q^T$ .

Сплайн  $g(t)$  строится по выражению (4) по значениям  $g$  и  $\gamma = R^{-1} Q^T g$ .

Для выбора параметра сглаживания  $\alpha$  можно воспользоваться процедурой кросс-валидации, для чего требуется найти минимум выражения  $CV(\alpha)$  относительно  $\alpha$ , формула (36)

$$CV(\alpha) = n_f^{-1} \sum_{\{i=1\}}^{n_f} w_i^f \left( \frac{y_i - \sum_{k=1}^m C_{ik}^f g_k}{1 - \sum_{k=1}^m C_{ik}^f A_{ki}^f(\alpha)} \right)^2 + n_{df}^{-1} \sum_{\{j=1\}}^{n_{df}} w_j^{df} \left( \frac{y_j' - \sum_{k=1}^m C_{jk}^{df} g_k}{1 - \sum_{k=1}^m C_{jk}^{df} A_{kj}^{df}(\alpha)} \right)^2 +$$

$$+ n_{d^2f}^{-1} \sum_{\{l=1\}}^{n_{d^2f}} w_l^{d^2f} \left( \frac{y_l'' - \sum_{k=1}^m C_{lk}^{d^2f} g_k}{1 - \sum_{k=1}^m C_{lk}^{d^2f} A_{kl}^{d^2f}(\alpha)} \right)^2 + n_{int}^{-1} \sum_{\{u=1\}}^{n_{int}} w_u^{int} \left( \frac{Y_u - \sum_{k=1}^m C_{uk}^{int} g_k}{1 - \sum_{k=1}^m C_{uk}^{int} A_{ku}^{int}(\alpha)} \right)^2, \quad (36)$$

где  $C^f, C^{df}, C^{d^2f}, C^{int}$  - такие же матрицы, как и в формуле (33);  $A^f(\alpha) = A(\alpha) C_f^T W_f$  ;  
 $A^{df}(\alpha) = A(\alpha) \mu C_{df}^T W_{df}$  ;  $A^{d^2f}(\alpha) = A(\alpha) \nu C_{d^2f}^T W_{d^2f}$  ;  $A^{int}(\alpha) = A(\alpha) \psi C_{int}^T W_{int}$  ;  
 $A(\alpha) = \left( C_f^T W_f C_f + \mu C_{df}^T W_{df} C_{df} + \nu C_{d^2f}^T W_{d^2f} C_{d^2f} + \psi C_{int}^T W_{int} C_{int} + \alpha K \right)^{-1}$ .

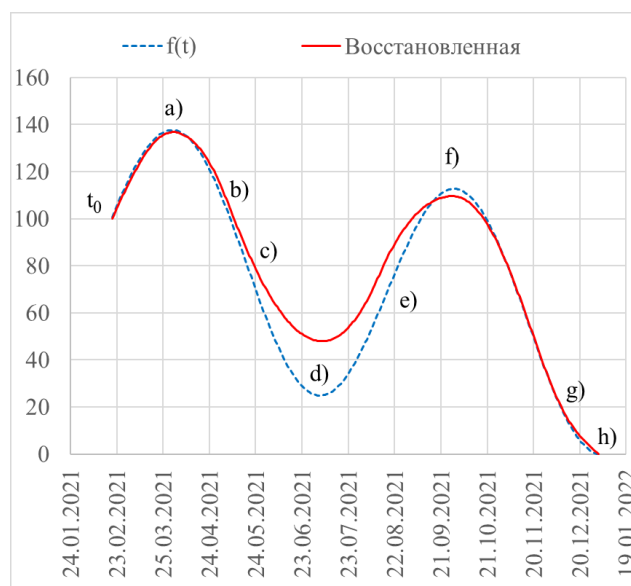
Пример. Пусть от некоторой неподконтрольной организации поступают сообщения следующего вида: а) мы начали тонуть – надо срочно что-то делать; б) мы все еще сильно тонем; в) кажется, снижение начало замедляться; г) мы начали всплывать; д) хорошо идем, вот бы всегда так; е) мы вновь тонем; ж) мы прошли отметку невозврата; з) мы на дне. Предполагаем, что события образованы операцией сравнения некоторого внутреннего показателя с определенными критическими значениями. Предполагаем, что искомая функция скрытой динамики  $f(t)$  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

(  $f(t_0) = 100$  ). Даем приблизительную количественную оценку поступающим событиям, таблица 6, по которым получается восстановить ненаблюдаемую скрытую динамику, рисунок 21.

Таблица 6 – Пример, входные данные

$t_f$	$y_f$	$t_{df}$	$y_{df}$	$t_{d^2f}$	$y_{d^2f}$	$t_{int}^a$	$t_{int}^b$	$Y_{int}$
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 21 – Восстановленная по разным функционалам скрытая динамика

3) Разработан универсальный метод анализа и прогнозирования редких событий, в котором процесс образования событий может быть задан произвольным алгоритмом с динамическими параметрами.

Процесс образования событий может быть задан алгоритмической моделью с неизвестными параметрами  $P$ , подлежащими определению. Причем параметры могут быть динамическими (как например, если бы в эконометрике вместо параметров  $a_0, a_1$  были бы параметры  $a_0(t), a_1(t)$ ). Параметры задаются вновь кубическим сплайном, но уже через значения третьей производной в каждом узле (плюс начальные условия). В итоге, чтобы задать значение изменяющегося со временем параметра процесса в виде сплайна  $g(t)$  требуется задать:

- узлы сплайна  $s_1 < s_2 < \dots < s_m$  (количество узлов  $m$  и их местоположение выбирается исследователем, например можно выбрать узлы, совпадающими с наблюдениями, или распределить их равномерно на всем интервале наблюдений);

- начальное значение, а также первую и вторую производную в самом первом узле  $g(s_1), g'(s_1)$  (вторая производная  $g''(s_1) = 0$  у натурального сплайна);

- значения третьей производной  $g'''(s_k)$  во всех предшествующих моменту времени  $t$  узлах,  $\forall k: s_k < t$ .

При продвижении времени значения динамических параметров определяются по формулам разложения в ряд Тейлора. Однако, что статические, что динамические параметры, неизвестны, они подлежат определению. Поэтому для динамических параметров требуется определить начальные условия и значения третьей производной в каждом узле. Если использовать условие натурального сплайна, у которого вторая производная в самом начале и в самом конце равна нулю ( $g''(s_1) = g''(s_m) = \sum_{k=1}^{m-1} h_k g'''(s_k) = 0$ ), то определять надо меньше значений (определять  $g'''(s_{m-1})$  не надо, ее можно выразить через значения в других узлах  $g'''(s_{m-1}) = -\sum_{k=1}^{m-2} h_k g'''(s_k) / h_{m-1}$ ). Штраф на нелинейность каждого динамического параметра рассчитывается по формуле (37)

$$\int_{s_1}^{s_m} (g''(t))^2 dt = \sum_{k=1}^{m-1} \frac{(g''(s_k) + g'''(s_k)h_k)^3 - (g''(s_k))^3}{3g'''(s_k)}, \quad (37)$$

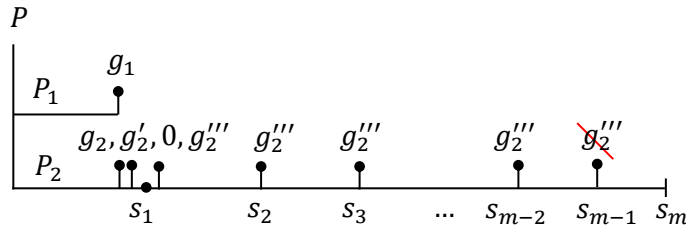
где  $g''(s_k) = g'''(s_1)h_1 + \dots + g'''(s_{k-1})h_{k-1}$ .

Примером такой алгоритмической модели процесса образования событий может быть следующий алгоритм (как в модели управления запасами):

- 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) }
- 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  $t_i$  и значением  $y_i$  (объемом покупки/заказа), равным тому, сколько не хватает до максимума. Процедура «Продвинуть время и обновить параметры» продвигает время на один шаг (день). В этом примере логично предположить, что первый параметр  $P[1]$  будет стационарным, в этом случае строку 15 можно не выполнять. Для определения параметров нужно найти значение стационарного параметра и значения третьей производной в каждом узле плюс начальные условия в первом узле второго параметра. Схема подлежащих определению значений параметров представлена на рисунке 22.



Источник: составлено автором.

Рисунок 22 – Пример схемы значений для описания параметров процесса формирования событий,  $P_1$  – первый параметр (статический),  $P_2$  – второй параметр (динамический)

Значения параметров  $P$  определяются с помощью соотнесения имеющейся выборки редких событий с выборкой событий, полученной в результате функционирования нашей модели процесса. Для этого задаем следующую функцию потерь (относительных отклонений), формула (38)

$$S(P) = \sum_{i=2}^n \left( \frac{t_i - t'_i}{t_i - t_{i-1}} \right)^2 + \mu \sum_{i=2}^n \left( \frac{y_i - y'_i}{y_i} \right)^2 + \sum_{z=1}^{N_{\text{дин.пар.}}} \alpha_z \int_{s_1}^{s_m} (g''_z(t))^2 dt, \quad (38)$$

где  $t_i$  – моменты времени появления событий из имеющейся выборки;

$t'_i$  – моменты времени появления событий, полученные в результате функционирования модели процесса (моделирования);

$y_i$  – информация/признак/воздействие, которую несет событие из имеющейся выборки;

$y'_i$  – информация/признак/воздействие, которую несет событие, полученное в результате функционирования модели процесса (моделирования);

$\int_{s_1}^{s_m} (g''_z(t))^2 dt$  – штраф на нелинейность (шероховатость) параметра с номером  $z$ ;

$N_{\text{дин.пар.}}$  – количество динамических параметров;

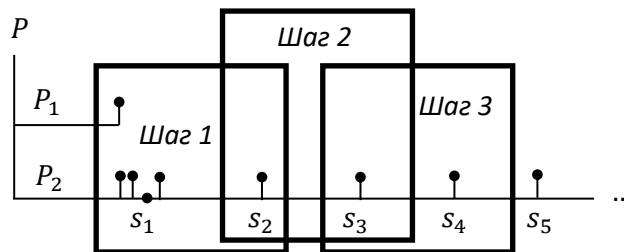
$\alpha_z$  – коэффициент, с которым учитывается данный штраф в общей функции потерь;

$\mu$  – весовой коэффициент, показывающий какую долю внимания мы уделяем отклонениям для моментов времени, а какую - для признаков.

Если при функционировании процесса получается недостаточное количество событий, то функция потерь возвращает  $\infty$ , однако, надо дать шанс процессу сгенерировать эти события, для чего длительность моделирования надо увеличивать.

Оптимизация функции потерь  $S(P)$  происходит численными методами. Так как функция потерь часто возвращает нечисловые значения, то хорошо подходит алгоритм Нелдера-Мида. Так как функция потерь может содержать много локальных минимумов, оптимизацию следует осуществлять на сетке (выбор шага сетки для каждого параметра может быть подсказан значениями данных, например, если объемы покупок измеряются в районе 1000, то бессмысленно перебирать значения максимального запаса в области много меньше или больше 1000).

Чтобы уменьшить размерность получаемой оптимизационной задачи, используется следующий прием, рисунок 23. В начале надо 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 события, на втором шаге мы подбираем параметры, отвечающие за образование второго и третьего, подбираем их опять на сетке.



Подбор значений параметров для процесса, у которого первый параметр статический, а второй динамический. На первом шаге полностью перебираются на сетке комбинации 5 параметров (у статического параметра надо подобрать одно значение, у динамического параметра надо подбирать  $g(s_1)$ ,  $g'(s_1)$  и  $g'''(s_1)$  (вторая производная  $g''(s_1) = 0$ ) в первом узле и  $g'''(s_2)$  во втором узле), на втором шаге перебираются комбинации 2 параметров  $g'''(s_2)$  и  $g'''(s_3)$ , на третьем тоже 2 параметров  $g'''(s_3)$  и  $g'''(s_4)$ .

Источник: составлено автором.

Рисунок 23 – Подбор параметров на  $N = 2$  события вперед

При этом параметры, отвечающие за образование не вошедших в скользящее окно событий, также используем в оптимизации, но уже не на сетке. На рисунке 23 на втором шаге на вход оптимизатора подаются все шесть значений, но первые четыре (не вошедшие в окно) подаются каждый раз как рассчитанные на первом шаге значения, когда два значения (вошедшие в окно) подаются на вход с разными значениями (перебор их на сетке). При оптимизации могут поменяться все шесть значений на втором шаге, все семь значений на третьем шаге и так далее. (просто начальные комбинации, откуда стартует оптимизация, зависят только от последних значений, но во время самой оптимизации определяются все значения). предполагается, что найденные на прошлом шаге не вошедшие в скользящее окно значения параметров находятся как

бы в глобальном оптимуме, а не в локальном. То есть используется предположение «Параметры процесса, соответствующие событию, зависят незначительно от событий, расположенных в более чем  $N$  позициях от текущего события». Предпочтительнее выбирать ширину окна событий  $N$  как можно больше, но это очень сильно влияет на объем вычислений. Нужно найти некоторый баланс между производительностью и точностью определяемых параметров. В результате используемого приема удастся обойти проблему неполиномиального роста объема вычислений (правда за счет возможной потери точности).

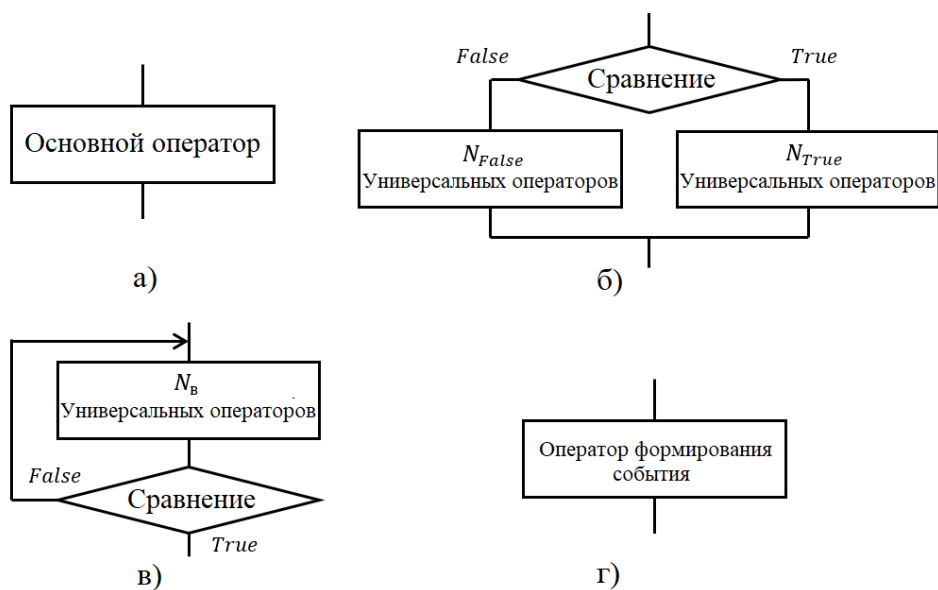
4) Разработан метод исследования совершенно неизвестных событий на основе ранее разработанного универсального метода анализа и прогнозирования редких событий.

В данном методе предполагается подбирать не только параметры  $P$  процесса, но также модель *Program* самого процесс образования событий в виде алгоритма. Функция потерь рассчитывается так же, как в формуле (38), с тем отличием, что меняется сам процесс *Program* образования событий, формула (39)

$$S(Program, P) \rightarrow \min. \quad (39)$$

Ограничивая количество операторов, количество внутренних переменных и параметров, перебираем возможные схемы алгоритма модели процесса *Program*. После чего для каждой такой модели происходит восстановление параметров процесса  $P$ .

Механизм перебора алгоритмов модели процесса *Program* образования событий следующий. Основным элементом является универсальный оператор, который в свою очередь может оказаться одним из 4 видов: основной оператор, оператор ветвления, оператор цикла и оператор формирования события, как на рисунке 24. При этом внутри этого универсального оператора могут содержаться другие универсальные операторы.



Возможно 4 вида: а) основной оператор; б) оператор ветвления; в) оператор цикла; г) оператор формирования события

Источник: составлено автором.

Рисунок 24 – 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  $N_B$  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  $(t_i, y_i)$ .

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 (событие может формироваться только внутри оператора ветвления). 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 операторов и 2 сравнениях уже 450 комбинаций схем; при 4 основных операторов и 1 сравнении – всего 70 комбинаций схем; при 4 основных операторов и 2 сравнениях – уже 1638 комбинаций схем; при 5 основных операторов и 1 сравнении – всего 140 комбинаций схем; при 5 основных операторов и 2 сравнениях – уже 4761 комбинаций схем.

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. 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 получается 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. Можно уменьшить количество перебираемых комбинаций, пропуская такие, которые не вносят изменения в ранее исследованные процессы образования событий. Тем самым существует принципиальная возможность исследовать совершенно неизвестные события с помощью подбора алгоритмической модели процесса образования событий. Однако такой подход требует дальнейших исследований, нужно подключение широкого круга исследователей.

### III Заключение

В данном исследовании разработан новый метод анализа и прогнозирования редких событий в экономике, основная идея которого заключается в рассмотрении событий с точки

зрения процессов, в результате функционирования которых как раз образуются эти события. Для процессов, которые можно моделировать как процессы опустошения/наполнения некоторой емкости, был разработан «емкостный» метод. В качестве примеров рассматриваются примеры процесса потребления, когда событиями являются покупки неподконтрольных нам клиентов, например, как в моделях управления запасами. В этом случае получается прогнозировать будущие покупки таких 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 их будущих покупок. Это позволяет применять определенные маркетинговые меры, например, напомнить клиенту о потребности пополнить его запасы. 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. Искомая функция восстанавливается в виде кубического сплайна по наблюдаемым с погрешностью значениям функционалов, которыми являются определенные интегралы 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  $R$ . Восстановленная динамика изменения параметров процесса сама по себе может иметь интересную интерпретацию и представлять отдельный теоретический интерес.

Проведено исследование различных вопросов точности. Выявлены дополнительные причины возникновения погрешности в самих наблюдениях. Построена зависимость погрешности восстановления параметров процесса в зависимости от этой погрешности в наблюдениях. Предложен подход проверки адекватности получающихся моделей прогнозирования на основе



построения доверительного интервала для момента времени прогнозируемого события. Рассчитаны формулы для построения соответствующего доверительного интервала.

Разработан универсальный метод анализа и прогнозирования редких событий, в котором процесс образования событий может быть задан произвольным алгоритмом с динамическими параметрами. Данный подход открывает новые возможности, исследователи будут формировать алгоритмические модели образования редких событий (аналог составления структурных моделей в эконометрике), 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, когда сами модели процессов подбираются автоматически.

Поставленные в диссертационном исследовании задачи были решены полностью, а исследование получилось комплексным. Открывшееся новое направление исследований может являться источником новых вопросов и задач, решение которых возможно потребует привлечение более широкого круга исследователей.

#### **IV Список работ, опубликованных по теме диссертации**

*Публикации в рецензируемых научных изданиях,  
определенных ВАК при Минобрнауки России:*

1. 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.
2. 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.
3. 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.
4. Кораблев, Ю.А. Емкостный метод анализа редких продаж в Excel / Ю.А. Кораблев // Научно-практический, теоретический журнал «Экономика и управление: проблемы, решения». – 2017. – № 6. Том 3 (66). – С. 224-229. – ISSN 2227–3891.
5. Кораблев, Ю.А. Разбор причин и оценка погрешности аномальных картин в емкостном методе анализа редких событий / Ю.А. Кораблев // Научно-практический, теоретический журнал «Экономика и управление: проблемы, решения». – 2017. – № 8. Том 6 (68). – С. 8-12. – ISSN 2227-3891.

6. Кораблев, Ю.А. Исследование точности емкостного метода от позиции в цепочке распространителей / Ю.А. Кораблев // Научно-практический, теоретический журнал «Экономика и управление: проблемы, решения». – 2018. – № 5. Том 7 (77). – С. 106-121. – ISSN 2227-3891.
7. 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.
8. Кораблев, Ю.А. Емкостный метод анализа редких событий в торговле различными товарами / Ю.А. Кораблев, П.С. Голованова, Т.А. Кострица // Бизнес. Образование. Право. – 2019. – № 3 (48). – С. 121-131. – ISSN 1990-536X.
9. 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.
10. 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.
11. 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.
12. 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. (RSCI).
13. Кораблев, Ю.А. Моделирование и восстановление предпочтений покупателей между двумя альтернативными товарами с помощью емкостного метода анализа редких событий в экономике (часть 1; часть 2) / Ю.А. Кораблев // Известия Кабардино-Балкарского научного центра РАН. – 2020. – № 3 (95); 4 (96). – С. 74-91; 78-88. – ISSN 1991-6639.
14. 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.
15. 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.
16. 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.

17. 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.

18. 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. (RSCI).

19. 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.

20. 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.

*Публикации в других научных изданиях:*

21. Кораблев, Ю.А. Альтернативное определение спроса / Ю.А. Кораблев // Россия 2030 глазами молодых ученых : материалы III Всероссийской научной конференции. – Москва : Научный эксперт, 2012. – С. 137-140. – ISBN 978-5-91290-192-8.

22. 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 В.С. Пономаренко, Т.С. Клебановой. – Харьков : ВШЭМ – ХНЭУ имени С. Кузнеца, 2018. – С. 502-523. – 668 с. – 150 экз. – ISBN 978-80-89654-45-1.

23. 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.

24. 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.

25. 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. – Москва : Издательский дом МФО, 2019. – С. 123-132. – ISSN 2412–9895.

26. 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.

27. Кораблев, Ю.А. Cross-validation for an Integral Spline Restoring the Consumption Rate of Rare Events = 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.

28. 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. (RSCI).

29. Кораблев, Ю.А. Algorithm 1023: Restoration of Function by Integrals with Cubic Integral Smoothing Spline in  $R$  = Восстановление функции по интегралам с помощью кубического интегрального сглаживающего сплайна в  $R$  / Ю.А. Кораблев // ACM Transactions on Mathematical Software. – 2022. – № 2. Volume 48. – ISSN 0098-3500. – Текст : электронный. – DOI 10.1145/3519384. – URL: <https://dl.acm.org/doi/10.1145/3519384> (дата обращения: 17.10.2022) (Scopus, Q1).

30. Кораблев, Ю.А. 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 / Ю.А. Кораблев // Acta Applicandae Mathematicae. – 2022. – № 180 (8). – ISSN 1572-9036. – Текст : электронный. – DOI 10.1007/s10440-022-00509-7. – URL: <https://link.springer.com/article/10.1007/s10440-022-00509-7> (дата обращения: 17.10.2022) (Scopus, Q2).