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

На правах рукописи

Кораблев Юрий Александрович

ПРОГНОЗИРОВАНИЕ РЕДКИХ ЭКОНОМИЧЕСКИХ СОБЫТИЙ С ПОМОЩЬЮ ВОССТАНОВЛЕНИЯ МЕХАНИЗМОВ ИХ ОБРАЗОВАНИЯ

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
	Судаков Владимир Анатольевич, доктор технических наук, доцент

Москва – 2026

Диссертация представлена к публичному рассмотрению и защите в порядке, установленном ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации» в соответствии с предоставленным правом самостоятельно присуждать ученые степени кандидата наук, ученые 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 состоятся 7 октября 2026 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.

Персональный состав диссертационного совета:

председатель – Михайлова С.С., 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;
Васильева Е.В., д.э.н., доцент;
Владова А.Ю., д.техн.н., доцент;
Иванюк В. А., д.э.н., доцент;
Соловьев В.И., д.э.н., профессор;
Судаков В.А., д.техн.н., доцент;
Трегуб И.В., д.э.н., профессор.

Автореферат диссертации разослан 26 мая 2026 г.

Ученый секретарь диссертационного совета
Финансового университета Д 505.001.111 Т.В. Золотова

I Общая характеристика работы

Актуальность темы исследования. Прогнозирование событий дает возможность к ним подготовиться, позволяет извлекать выгоду или уменьшать убытки, определяет успех или спасает от неудач. Прогноз событий востребован как в государственном управлении, так и в управлении частным бизнесом. Своевременный прогноз может позволить правительству заранее подготавливать меры поддержки, центральному банку – заранее корректировать монетарную политику. Умение прогнозировать события позволит бизнесу предвидеть изменения спроса, оптимизировать издержки на хранение запасов и логистику, избегать дефицита и затоваривания, оптимизировать работу своего персонала, управлять как подконтрольными, так и неподконтрольными клиентами, инвестировать в перспективные проекты и избегать рискованные инвестиции. Сложно переоценить способность предсказывать будущие события. В особенности это касается редких экономических событий.

Первое, с чем ассоциируются «редкие события», – это финансовые кризисы, дефолт того или иного банка, землетрясения, выигрыш в лотерее, прорывное научное открытие, внезапный скачок продаж или цен и так далее. То есть что-то экстраординарное по своей величине и значению. Однако что для одного может являться значимым, может таковым не являться для другого. Не величина играет главную роль, а внезапность. В этом понимании «редкие события» обозначают моменты времени, когда нарушаются ожидания людей. Это те моменты, когда мысленные или математические модели разбиваются об объективную реальность. Такие редкие события Нассим Николас Талеб называет «черными лебедями». Однако Талеб на протяжении всей своей знаменитой работы говорит о невозможности прогнозирования момента времени, когда ваша модель сломается.

Ожидания нарушаются лишь по нашему ошибочному представлению. Нужно лучше изучать происходящие в мире процессы и разбираться в причинах возникновения событий. «Черные лебеди», которые привлекли внимание, заставляют строить модели процессов, в результате чего события перестают быть «черными лебедями». Мы занимаемся прогнозированием тех событий, модели образования которых можно построить. Талеб называет их «серыми лебедями». Если модели дают ошибочный прогноз, то это будет лишь сигналом о необходимости скорректировать модель.

В этой работе под редкими событиями понимается не нарушение ожиданий. Редкие события относятся все к тем же дискретным событиями, на которые по той или иной причине обращается внимание и которые могли бы вовсе не обладать смыслом редких или значимых событий. Но в результате моделирования и предсказания событий некоторые из них могут получиться такими, которые будут обладать тем самым уникальным свойством, которое позволило бы назвать такие события редкими или значимыми. Предсказание редких событий базируется на способности предсказывать обычные события. С этого момента можно вообще опустить само понятие «редкие», ведь, занимаясь предсказанием событий, в том числе будем получать и редкие. Вместо значимости под редкостью понимается «не частота», подчеркивая дискретный характер данных.

В области моделирования редких событий одинаковые понятия часто обозначаются разными терминами, а под одним и тем же названием могут скрываться принципиально разные концепции и решаемые задачи. Среди названий можно встретить такие, как аномалии, новизна, выбросы, инциденты. Среди задач, использующих термин «редкие события», встречаются такие, как задача обнаружения событий во временных рядах, классификация по наблюдаемым признакам, определение вероятности возникновения события в заданном периоде времени, определение вероятности выхода за границы допустимых условий и др. Исследование существующих методов показало, что большинство из них основано на выявлении статистических связей между наблюдаемыми признаками и интервалами времени до следующего события или самим фактом появления события на интервале времени. Но когда в наблюдаемых признаках не содержится информация, способная объяснить интервалы времени до следующего события, или таких признаков вовсе нет, то такие статистические методы могут работать неудовлетворительно. Другие методы ограничиваются лишь определением закона распределения для интервалов времени между событиями, что не способно объяснить, почему получающиеся интервалы оказываются то короче, то длиннее. Модели, основанные на применении машинного обучения, в особенности нейросетевые модели или модели, использующие градиентный бустинг, применяющиеся в прогнозировании событий, имеют характер черного ящика. Это создает проблемы, связанные с объяснимостью и пониманием устройства самих этих моделей, что делает результат прогнозирования неубедительным и ненадежным. Анализ множества работ показал, что задача прогнозирования редких событий изучена недостаточно подробно. Отсутствует комплексная методология прогнозирования редких экономических событий. Существующим методам необходим большой объем обучающей выборки, что недоступно при анализе редких событий.

Отсутствие комплексной методологии прогнозирования редких экономических событий составляет научную проблему исследования. Задача прогнозирования редких экономических событий может решаться в рамках новой методологии, в которой предлагается по небольшой зашумленной выборке событий восстанавливать параметры алгоритмических моделей механизмов их образования. Такой подход по точности является конкурентным передовым существующим методам машинного обучения, зачастую превосходя их. Разрабатываемая методология закладывает основы новому научному направлению, которое назовем «эвентометрика».

Степень разработанности темы исследования. Вопросами математического моделирования экономических процессов занимались Кантарович Л.В., Юдин Д.Б., Данциг Г., Беллман Р., Леонтьев В.В., Нейман Дж., Моргенштерн О., Слуцкий Е.Е., Фельдман Г.А., Немчинов В.С., Поспелов И.Г., Моисеев Н.Н., Иванилов Ю.П., Макаров В.Л., Бахтизин А.Р. и др.

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., Ширяев А.Н. и др. Однако чисто вероятностное моделирование не способно объяснить отклонения в большую или меньшую сторону от своего ожидаемого значения, что делает прогноз событий менее точным и обоснованным. Прогнозирование интервалов времени до следующего события на основе наблюдаемых признаков возможно осуществить проверенными регрессионными методами, основы которых заложили Лежандр А.М., Гаусс К.Ф., Адраин Р., Нейман Дж., Эйткен А.С., Фишер И., Фриш Р., Тинберген Я., Шумпетер Й., Андерсон О., Айвазян С.А., Бывшев В.А., Бабешко Л.О. и многие другие.

Появление событий часто прогнозируют методами бинарной или множественной классификации, где по наблюдаемым сопутствующим признакам определяют класс того или иного наблюдения. Для этого используются хорошо известные методы, которые последнее время выделяют в группу методов машинного обучения, такие как логистическая и пробит-регрессия, метод ближайших соседей, метод опорных векторов, деревья решений, градиентный бустинг, нейронные сети и другие методы, основы которых были заложены в работах Берксона Дж., Блисса Ч., Вапника В.Н., Червоненкиса А., Босера Б., Бравермана Э., Cramer J.S., Cortes C., Fix E., Cover T., Hart P., Bentley J.L., Omohundro S.M., Singh S., Morgan J.N., Sonquist J.A., Belson W., Breiman L., Friedman J., Olshen R., Stone C.J., Landwehr N., Hall M., Frank E., Kearns M., Valiant L.G., Schapire R.E, Freund Y., Rosenblatt F., Choe. W., Ersoy O.K., McCullagh P., Nelder J., Hosmer D.W., Lemeshow S., Goodfellow I., Hastie T., Tibshirani R., Каширина И.Л., Демченко М.В., Судакова В.А., Соловьева В.И. и др. Редкие события зачастую связаны с проблемой несбалансированности выборки, решение которой представлено в работах Vannucci M., Colla V. и др. Методы классификации имеют очень широкое применение в работах отечественных и зарубежных исследователей: для определения устойчивости (дефолта/банкротств) компаний или банков – Гусятникова П.В., Могилат А.Н., Туктарова П.А., Заиченко Е.М., Рыгина В.Е., Биджояна Д.С., Kumar K., Tan C., Jardin P., Johnsen T. и многих других (где дефолты/банкротства определяются как редкие события); в информационной безопасности для определения атак или эксплойтов – Лифанова К.А, Зайцева К.С.; в прогнозировании ненулевого спроса – Пивкина К.С.; при определении потенциальных покупателей – Заказчиковой Н.А.; во многих других областях экономики.

Очень часто редкие события рассматривают с точки зрения потоков событий или 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., Вентцель Е.С., Овчарова Л.А. и др. В экономике теория потоков событий находит очень широкое применение: при прогнозировании спроса и моделировании собственных объемов страховых запасов – Истомина А.А., Ян Л.А., Замалетдинова Д.А., Syntetos A.A., Babai M.Z., Luo S., Shale E.A.; 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. В теории экстремальных событий также оперируют термином «редкие события», хотя в ней не делается предсказание событий. События моделируются пуассоновскими потоками, а ущерб от событий – с помощью распределений с тяжелыми хвостами. В этой области работали Embrechts P., Klüppelberg C., Mikosch T., Paul E. и др.

Прогнозирование событий перекликается с задачами создания систем раннего предупреждения, например кризисов, которые развивали зарубежные ученые Berg A., Pattillo C. и др. Современные системы раннего предупреждения, основанные на методах машинного обучения, рассматривали в своих работах Xi L., Yining W., Qianqian F., Qinyun L., Tian W. и др. Также редкие события встречаются в работах по теории случайного блуждания, таких как случайные процессы Леви и теории больших прыжков. Но в этой теории не предсказываются отдельные события, а определяются законы распределения для суммарного движения. В этом направлении работали Wiener N., Lévy P., Burioni R., Vezzani A., Holl M., Barkai E.

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., Efron B., Willemain T.R., Smart C.N., Hua Z.S., Kaya G.O., Pince C. 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., Проскан Ф. и др.

Полученные в этой диссертационной работе результаты опираются на методы непараметрической регрессии, такие как ядерные функции, теория приближения функций сплайнами, сплайновая коллокация, интегро-дифференциальные сплайны, методы решения некорректно поставленных задач, определение оптимального параметра сглаживания, основы которых заложены в работах Nadaraya E.A., Watson G.S., Henderson D.J., Бубнова И.Г., Whittaker E.T., Schoenberg I.J., Curry H.B., Reinsch Ch.H., Ahlberg J.H., Nilson E.N., Walsh J.L., Boor C.D., Schumaker L.L., Kimeldorf G., Craven P., Wahba G., Ramsay J.O., Green P.J., Silverman B.W., Тихонова А.Н., Арсенина В.Я., Wright I.W., Morozov A., Hansen P.C., Киреева В.И., Бирюковой Т.К., Самойловой Э.Н., Федоровой О.П., Zemlyanova A.Y., Machina A., Dagnino C. Обеспечение положительности восстанавливаемой функции основывается на работах Elfving T., Andersson L., Vlachkova K., Greiner H., Ramsay J.O., Nagahara M., Green P.J., Silverman B.W., Chan V., Tsui K.W., Wei Y., Смоляка С.А. Решение задач квадратичного программирования опирается на работы Лагранжа Ж.Л., Лемке К.Э., Katta G.M., Wright M.H., Писарука Н.Н.

Цель исследования – разработка методологии прогнозирования экономических событий, в том числе редких, с помощью восстановления механизмов их образования для повышения точности и обоснованности прогноза.

Задачи исследования:

– Анализ и адаптация существующих методов для прогнозирования экономических событий (методы классификации, регрессионные методы, методы машинного обучения, методы теории случайных процессов и др.), анализ границ их применимости и недостатков.

– Разработка методологии прогнозирования экономических событий, основанной на составлении алгоритмических моделей и восстановлении параметров механизмов их образования.

– Анализ применимости существующих математических методов для восстановления параметров механизмов образования событий.

– Разработка метода восстановления параметров механизма образования событий, основанного на методах монотонной сплайновой коллокации.

– Разработка способов экстраполяции восстановленных параметров на будущее, чтобы механизм образования событий функционировал в условиях, соответствующих будущему периоду времени.

– Разработка метода подбора гиперпараметров для применяемых алгоритмов восстановления и экстраполяции параметров механизма образования событий.

– Разработка способов проверки адекватности построенных алгоритмических моделей прогнозирования событий.

– Разработка метода восстановления параметров механизмов образования событий, заданных в алгоритмической форме, на основе численной оптимизации.

– Разработка методов автоматического подбора самих алгоритмических моделей образования событий.

– Сравнение нового подхода прогнозирования событий с существующими методами на практических примерах.

– Реализация разработанных математических методов и алгоритмов с использованием современных информационных технологий на современных языках программирования.

Объект исследования – экономические события, в том числе редкие, и методы их прогнозирования.

Предмет исследования – механизмы образования экономических событий, в том числе редких, и методы восстановления их параметров.

Область исследования. Диссертация подготовлена в соответствии с п. 11. «Компьютерные методы и программы моделирования экономических процессов», п. 14. «Эконометрические и статистические методы анализа данных, формирования и тестирования гипотез в экономических исследованиях. Эконометрическое и экономико-статистическое моделирование» Паспорта научной специальности: 5.2.2. Математические, статистические и инструментальные методы в экономике (экономические науки).

Научная новизна исследования заключается в разработке новой методологии прогнозирования экономических событий, основанной на восстановлении механизмов их образования. В отличие от существующих статистических методов в новом подходе строятся алгоритмические модели механизмов образования событий. По выборке событий восстанавливаются параметры моделей. Разработаны аналитические средства восстановления параметров, основанные на монотонной сплайновой коллокации, а также средства, основанные на численных методах. Восстановленные параметры экстраполируются на будущее и используются при моделировании самих механизмов для получения прогноза будущих событий.

Положения, выносимые на защиту:

1) Произведена адаптация существующих методов анализа данных и машинного обучения, таких как методы классификации, регрессии и методы на основе потоков событий, для задачи прогнозирования событий. Адаптация заключается в способах: построения матриц признаков из последовательности событий как для обучения, так и для прогнозирования; формирования объясняемой переменной; интерпретации результата как прогноза событий. Рассмотрены недостатки и границы применимости существующих методов, даны рекомендации по их применению. Полученные результаты конкретизируют существующие подходы и методы прогнозирования в области экономических событий (С. 40–101).

2) Разработан новый подход к прогнозированию событий, в котором строятся алгоритмические модели механизмов образования событий. Под механизмом понимается детерминированная последовательность операторов над переменными и параметрами.

Переменные инициализируются и изменяются внутри механизма, параметры изменяются вовне механизма. Параметры могут быть как статичными, так и динамичными. По выборке событий происходит восстановление этих параметров аналитическими методами, в случае если модель механизма это позволяет, или численными методами. После экстраполяции параметров прогноз будущих событий получается в результате моделирования механизмов образования событий. Полученные результаты являются принципиально новыми положениями по отношению к существующим методам прогнозирования экономических событий (С. 104–177).

3) Разработан новый метод восстановления динамических параметров по выборке событий, основанный на монотонной сплайновой коллокации восстановления функции по разным функционалам, таким как сами значения искомой функции, ее первые и вторые производные, определенные интегралы на заданных, возможно, пересекающихся интервалах. Данный метод обеспечивает неотрицательность восстанавливаемой функции, что востребовано содержанием многих экономических задач. Ранее аналитическими методами параметры механизмов образования событий не восстанавливались. Полученные результаты дополняют и расширяют комплекс существующих статистических и математических методов на область прогнозирования экономических событий (С. 138–168).

4) Предложены способы экстраполяции параметров механизма образования событий: с помощью переноса тенденций с прошлого периода времени, с помощью регрессионных методов и с помощью разложения на сумму ограниченного количества гармонических функций. Данные практические предложения позволяют моделировать механизм образования событий в условиях, соответствующих будущему периоду времени, а не прошлому, тем самым повышая адекватность получаемых прогнозов (С. 169–173; 227–237).

5) Разработан способ определения оптимальных гиперпараметров для применяемых методов и алгоритмов восстановления и экстраполяции параметров механизма образования событий. В отличие от популярных методов анализа данных и машинного обучения, где гиперпараметры определяются только поиском на сетке, в разработанном способе гиперпараметры определяются на основе комбинирования поиска на сетке и локальной оптимизации в границах ячейки сетки алгоритмом Нелдера-Мида. Функция потерь минимизирует погрешность прогноза нескольких последних событий. Составлены принципиально новые метрики оценки качества прогноза событий на основе средней и среднеквадратичной, абсолютной или относительной погрешности прогноза их дат наступления и их характеристик. Подобных метрик для дискретных событий ранее не было (С. 175–178; 232–246).

6) Разработаны способы построения доверительных интервалов для проверки адекватности получающихся моделей прогнозирования экономических событий: а) на основе

оценки дисперсии параметров сплайна; б) на основе имитационного моделирования; г) на основе моделирования механизма образования событий с параметрами, отклоненными от их среднего значения. При промахе фактического события мимо рассчитанного доверительного интервала принимается гипотеза о неадекватности модели. Положения являются новыми, так как ранее доверительные интервалы для прогноза событий таким образом не рассчитывались (С. 178–186).

7) Разработан новый метод восстановления параметров модели механизма образования событий, основанный на численной оптимизации. Динамические параметры задаются кубическим сплайном, который кодируется через начальные значения и значения третьей производной в узлах сплайна. Функция потерь минимизирует квадрат относительных отклонений (моментов времени и значений) событий, полученных в результате моделирования, от фактических событий, плюс штраф на нелинейность динамических параметров. Используется специальный прием, ограничивающий объем вычислений. Полученные положения дают возможность восстанавливать параметры сложных алгоритмических моделей (С. 191–206).

8) Предложен метод подбора самих алгоритмических моделей механизмов образования событий, основанный либо на полном переборе с возможностью пропуска повторяющихся комбинаций операторов, либо на генетическом программировании. В этом способе функция потерь минимизируется не только по параметрам механизма, но и по структурной форме модели механизма образования событий. Благодаря полученным положениям появляется принципиальная возможность исследовать совершенно неизвестные события с помощью подбора алгоритмической модели механизма образования событий. Отличительной особенностью является алгоритм перебора схем операторов, а также их операндов (С. 208–215).

9) Разработан программный инструментарий на языке R и Python, реализующий комплекс методов, необходимых для прогнозирования будущих экономических событий. Создана программа на языке Python, реализующая метод восстановления функции по разным функционалам с помощью монотонной сплайновой коллокации. Разработана на языке Python и выложена в репозитарий PyPi библиотека eventometrics, которая реализует основные методы и алгоритмы прогнозирования экономических событий с помощью восстановления механизмов их образования. Библиотека имеет привычный для аналитиков данных интерфейс, реализует эффективные вычисления, ключевые алгоритмы оптимизированы с помощью Cython, используются параллельные вычисления для поиска оптимальных гиперпараметров. Библиотека отличается высоким быстродействием. Запрограммирован на языке R способ восстановления параметров механизма образования событий с помощью численных методов. На языке R реализован способ автоматического перебора самих моделей механизма образования событий. Разработан компонент для платформы бизнес-аналитики Loginom, реализующий предложенный новый подход прогнозирования событий с помощью восстановления механизмов их

образования. Полученные результаты открывают возможность широкого применения разработанной методологии в практике прогнозирования экономических событий (С. 159; 206; 211; 217–255).

Теоретическая и практическая значимость работы. Формирование комплексной методологии и разработка соответствующих технологий для прогнозирования экономических событий может создать критически важный научно-практический прорыв, закладывающий новое направление исследований (эвентометрика) в области анализа данных. Построение алгоритмических моделей и оценка их параметров является дальнейшим развитием научной мысли в экономических исследованиях, переводящим фокус с непрерывности изменения изучаемых закономерностей значений и показателей на дискретность явлений, возникающих в результате функционирования сложных процессов или механизмов. В дальнейшем это может привести к образованию направления обучения студентов, которые специализировались бы на построении моделей механизмов формирования экономических событий и оценке их параметров.

Полученные результаты позволяют поставить на новые научно-технологические рельсы и повысить уровень процесса анализа и прогнозирования экономических событий, в том числе редких, способствуют получению новых знаний и закономерностей в области происходящих экономических процессов, приводящих к возникновению событий.

На основе созданного фундамента можно сконструировать целое множество моделей, предсказывающих появление событий, в том числе редких, в различных прикладных областях. Разработанная методология способствует лучшему пониманию причин возникновения отдельных событий в экономике, что позволяет получать более точные и обоснованные прогнозы экономических событий. В бизнесе это позволит к ним подготовиться и извлечь определенную выгоду, например, более эффективно сформировать план производства, план закупки сырья, план пополнения запаса, оптимизировать логистические расходы, уменьшить объемы продуктовых запасов товаров, лежащих на складах, повысить уровень сервиса, добиться большего удовлетворения клиентов, применять новые маркетинговые приемы и т.д. Также это дает возможность органам государственного управления разрабатывать планы, отвечающие вызовам современного времени. Разработанные технологии могут способствовать быстрому обучению государственных служащих прогнозированию экономических событий для последующей выработки управленческих воздействий на протекающие в экономике и социуме процессы. Результаты исследования могут способствовать достижению национальных целей развития: комфортная и безопасная среда для жизни; достойный, эффективный труд и успешное предпринимательство; цифровая трансформация.

Методология и методы исследования. При адаптации существующих методов машинного обучения к прогнозированию событий рассматриваются методы опорных векторов, логистической регрессии, ближайших соседей, деревьев решений, градиентного бустинга, нейронных сетей, классические методы регрессии, теория потоков событий, методы прогнозирования прерывистого спроса.

Формирование моделей механизмов образования событий опирается на экономическую теорию, математический анализ, теорию управления запасами, системный анализ.

Восстановление параметров механизмов образования событий аналитическими методами опирается на теорию вероятностей, методы регрессии, математический анализ, сплайновую коллокацию (восстановление функции по функционалам), методы регуляризации, такие как кросс-валидация, метод L-кривой и невязки Морозова. Обеспечение положительности восстанавливаемой функции опирается на квадратичное программирование, метод множителей Лагранжа, алгоритм Лемке, метод внутренней точки.

Восстановление параметров механизмов образования событий численными методами базируется на имитационном моделировании, методах глобальной и локальной оптимизации. Автоматический подбор алгоритмических моделей опирается на комбинаторные методы и метод генетического программирования.

Экстраполяция восстановленных параметров основывается на регрессионных методах или анализе временных рядов, на алгоритме Куина-Фернандеса, который в свою очередь базируется на гармоническом анализе с помощью разложения в ряд Фурье.

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

Разработка программного инструментария базируется на языках программирования R и Python. При разработке программ на Python задействуются средства оптимизации Cython и распараллеливания вычислений.

Степень достоверности, апробация и внедрение результатов исследования. Достоверность исследования подтверждается корректным использованием математического аппарата, моделей и методов. Достоверность полученных в исследовании математических выражений демонстрируется на представленном в работе сквозном примере. Справедливость предложенных методик обосновывается компьютерным моделированием и примерами, основанными как на синтетических, так и на реальных данных.

Результаты исследования были доложены и обсуждены на различных научных мероприятиях: 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 IV Международной научно-практической конференции-биеннале «Системный анализ в экономике-2016» (Москва, Финансовый университет, 9–11 ноября 2016 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» (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-2020» (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 Научном семинаре в Департаменте анализа данных и машинного обучения на Факультете информационных технологий и анализа больших данных (Москва, Финансовый университет, 20 января 2021 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 заседании секции «Эконометрика» в Департаменте математики Финуниверситета (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.); на VII Международной научно-практической конференции-биеннале «Системный анализ в экономике-2022» (Москва, Финансовый университет, 7–8 декабря 2022 г.). на Международной научно-практической конференции «Системная экономика, социально-экономическая кибернетика и мягкие измерения в экономике» (Москва, Финансовый университет, 15–16 июня 2023); на X Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 21 июня 2023 г.); на XI Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 19 июня 2024 г.); на Ежегодной всероссийской научно-практической конференции с международным участием «Системная экономика, социально-экономическая кибернетика, мягкие измерения в экономике-2024» (Москва, Финансовый университет, 20–21 июня 2024 г.); на Круглом столе «Экономико-математическое моделирование как инструмент исследования динамических процессов» в рамках Всероссийского фестиваля науки «NAUKA 0+» (Москва, Финансовый университет, 12 октября 2024 г.); на VI Международной научной конференции «Развитие современной экономической науки: проблемы, тенденции, перспективы» (Москва, МГИМО, 27–29 ноября 2024 г.); на VIII Международной научно-практической конференции-биеннале «Системный анализ в экономик-2024е» (Москва, Финансовый университет, 12–14 декабря 2024 г.); на XI Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 11 июня 2025 г.); на Всероссийской научно-практической конференции с международным участием «Системное моделирование в экономике, финансах и управлении» (Москва, Финансовый университет, 24–25 июня 2025 г.).

Результаты исследования использованы при выполнении научно-исследовательской работы (далее – НИР) по теме проекта исследований гранта Российского фонда фундаментальных исследований «Емкостный метод анализа редких событий в экономике» в 2019–2021 гг., договор № 19-010-00154, в части подготовки всех разделов и формирования отчета о НИР.

Результаты исследования были использованы при выполнении НИР Финуниверситета по теме «Разработка методологии моделирования и прогнозирования динамики пространственного развития Российской Федерации на основе data-driven management» (Государственное задание ПИ-20, приказ от 26.12.2024 № 3268/о) в части подготовки раздела

«Система управления пространственным развитием регионов на основе имитационных моделей с использованием агентного подхода и визуализацией посредством геоинформационных систем».

Получено Свидетельство о государственной регистрации программы для ЭВМ № 2025688461 дата регистрации 21.10.2025 «Программный продукт, реализующий метод восстановления функции одновременно по значениям, значениям первой и второй производной, определенным интегралам, наблюдаемым с погрешностью, с условием обеспечения положительности или монотонности».

Материалы диссертационной работы внедрены в практическую деятельность компании ООО «АУМЕД» в целях улучшения планирования логистических процессов. Разработанные методики и программный инструментарий расширяют набор возможностей для анализа и прогнозирования таких событий, как будущие заказы отдельных корпоративных клиентов, появление ненулевого спроса или же определение величины запасов, что способствует определенности в принятии решений и в результате улучшению уровня сервиса.

Материалы диссертационной работы внедрены в практическую деятельность компании ООО «Квайссер Фарма» в целях совершенствования процессов планирования реализации фармацевтической продукции. Разработанная методология прогнозирования редких событий в экономике позволяет точнее определять будущие потребности в поставках фармацевтической продукции. Отдельные положения нашли свое применение при обосновании спада или повышения спроса на биологически активные добавки в период пандемии COVID-19.

Материалы диссертации используются Кафедрой бизнес-информатики Факультета информационных технологий и анализа больших данных Финансового университета в преподавании учебных дисциплин «Платформы бизнес-аналитики», «Технологии продвинутой аналитики» по направлению 38.03.05 «Бизнес-информатика» профиль обучения: «Технологии цифровых бизнес-моделей», «ИТ-менеджмент в бизнесе».

Апробация и внедрение результатов исследования подтверждены соответствующими документами.

Публикации. Материалы исследования опубликованы в 32 научных публикациях общим объемом 61,61 п.л. (авторский объем – 58,07 п.л.), в том числе одной авторской монографии объемом 18,5 п.л., 18 статьях общим объемом 25,09 п.л. (авторский объем – 21,55 п.л.), опубликованных в рецензируемых научных изданиях, определенных ВАК при Минобрнауки России, из которых 4 статьи общим объемом 8,25 п.л. (авторский объем – 7,65 п.л.) опубликованы в изданиях, входящих в цитатно-аналитическую базу RSCI, и 2 статьи авторским объемом 4,1 п.л. опубликованы в изданиях, индексируемых в цитатно-аналитической базе «Scopus» (Q1 и Q2), а также 4 статьи авторским объемом 7,22 п.л. опубликованы в рецензируемом научном издании «Известия Кабардино-Балкарского научного центра РАН», включенном на момент публикации статей в перечень ВАК при Минобрнауки России по научной специальности 08.00.13 – Математические и инструментальные методы экономики (физико-математические науки).

Структура и объем диссертации обусловлены целью и задачами исследования. Диссертация состоит из введения, пяти глав, заключения, списка литературы, состоящего из 219 наименований, и одного приложения. Текст диссертации изложен на 358 страницах, содержит 31 таблицу и 124 рисунка.

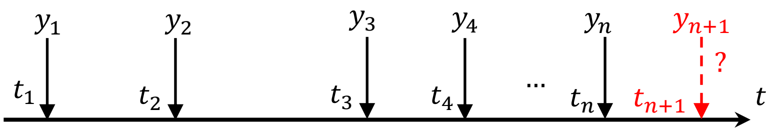
II Основное содержание работы

В **первой главе** диссертации получены следующие результаты:

1) Дано определение используемым терминам, описывается актуальность темы исследования, проведен анализ и адаптация существующих методов для прогнозирования редких событий.

Под термином «событие» понимается факт или явление, которое происходит в определенный момент времени как результат проявления экономических, политических, социальных, физических и других процессов, протекающих в той или иной сфере. Под прилагательным «экономическое» событие понимается область влияния события на мир, то есть на экономические процессы (которыми и может быть порождено), на экономических агентов, на общее экономическое состояние в целом и т.д. Термины «редкость» или «частота» показывают, насколько количество или вероятность одних событий отличается от количества или вероятности других событий. Однако понятия «редкость» или «частота» являются такими же диалектическими понятиями, как понятия «много» или «мало». Где заканчиваются частые события и начинаются редкие события, зависит от контекста. На основе проведенного исследования дано количественное определение: «редкие события» – это события, которые появляются в среднем в количестве два и менее за период требуемой гранулярности, где гранулярность данных – степень детализации данных, которая может существенно повлиять на точность, надежность и удобство использования прогнозов.

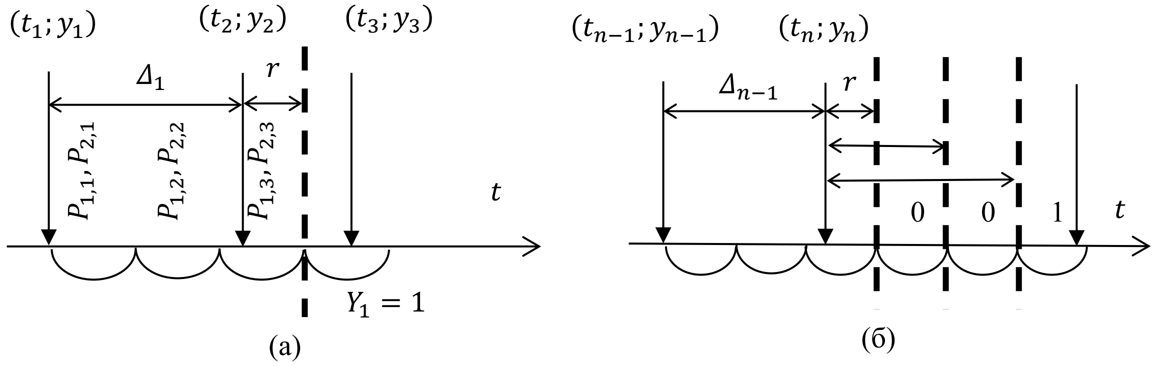
В работе делается акцент на дискретности событий, термин «редкие события» еще раз это подчеркивает. Предложенные в работе методы могут быть применены как для нередких, так и для редких событий, в то время как многие существующие методы работы с частыми событиями не подходят или малоэффективны для работы с редкими событиями. Самым ближайшим представлением событий является представление в виде потока событий из теории случайных процессов, рисунок 1, где t_i – моменты времени возникновения событий, y_i – характеристики или воздействия этих событий на мир.



Источник: составлено автором.
Рисунок 1 – Основная задача

В итоге термин «редкие события» говорит о представлении событий в виде потока событий. Но если говорить в контексте неожиданных/значимых/удивительных событий, то редкие события – это часть от событий, которые в результате работы механизма приобретут неожиданные (для стороннего наблюдателя) значения. Основная задача заключается в том, чтобы научиться предсказывать будущие события $(t_{n+1}; y_{n+1})$, рисунок 1.

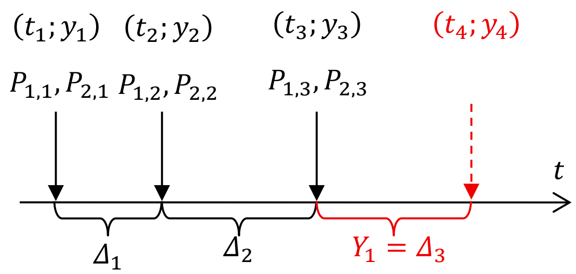
2) Произведен анализ и адаптация существующих методов для задачи прогнозирования событий. В методах бинарной классификации по векторам наблюдаемых признаков $X_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,m})$ некоторый объект относят к одному из классов $Y_i \in \{0,1\}$ (или $\{-1,1\}$). Для задачи прогнозирования событий необходимо задать интервалы времени определенной ширины (гранулярность), появление события на этих интервалах обозначать как $Y_i = 1$, а отсутствие как $Y_i = 0$. Чтобы модель могла предсказывать события в будущем, а не в настоящем, входящие в вектор X_i признаки $x_{i,j}$ должны относиться к предшествующим интервалам (быть лаговыми) по сравнению с интервалами i, на которые выпадают или не выпадают события. Поэтому первым наблюдением не может быть самое первое событие в обучающей выборке. В качестве признаков $x_{i,j}$ можно выбирать значения некоторых наблюдаемых индикаторов $P_{1,t}, P_{2,t}$, ... (если такие есть) на начало интервала, или на конец интервала, или среднее значение индикатора за интервал. Можно задействовать разницу времени между предыдущими событиями Δ_k, прошедшее время от предыдущего события до начала следующего интервала r, задействовать характеристики y_i или $\frac{y_i}{t_{i+1}-t_i}$ предшествующих событий. Например, в случае использования трех интервалов для формирования лаговых переменных, рисунок 2(а), вектор признаков может выглядеть как $X_1 = (P_{1,1}, P_{2,1}, P_{1,2}, P_{2,2}, P_{1,3}, P_{2,3}, \Delta_1, r, y_2, \frac{y_1}{t_2-t_1})$.



Источник: составлено автором.
Рисунок 2 – Пример составления наблюдений: а) первого наблюдения X_1 для обучающей выборки; б) пример трех наблюдений X_i для получения прогноза

Для получения прогноза нужно на обученную модель подать вектор признаков X_i нужной структуры, соответствующий будущим интервалам, рисунок 2(б). Для будущих интервалов часть признаков $x_{i,j}$, связанных с событиями (например, $\Delta_{n-1}, y_n, \frac{y_{n-1}}{t_n-t_{n-1}}$), меняться не будет, так как выборка событий не меняется. В качестве прогноза в зависимости от используемого метода классификации можно получить либо последовательность нулей, за которой следует последовательность единиц, например (0; 0; 0; 0; 1; 1; 1; 1; …), либо последовательность вероятностей наступления события на соответствующих интервалах, например (0; 0,05; 0,17; 0,38; 0,51; 0,65; 0,79; 0,9; 0,99). Заметим: в методе ближайших соседей 0 и 1 могут чередоваться, например (0; 0; 0; 1; 0; 0; 1; 1; 1). В качестве прогноза события следует брать самый первый интервал, на котором модель предсказывает 1 или вероятность прогноза события оказывается больше некоторого порога (например, 0,5, если обеспечивается сбалансированность выборки). Также модель может дать нулевой прогноз для всех рассматриваемых интервалов, т.е. последовательность только из нулей (или низких вероятностей). Это будет означать, что модель не прогнозирует событие в ближайшую перспективу. Зная номер интервала спрогнозированного события, можно ориентировочно определить дату события t_{n+1}. Однако предсказать характеристики событий y_{n+1} методами классификации не получится (если не рассматривать значения как метки классов).

3) В регрессионных методах строятся математические модели $Y_i = f(X_i) + u_i$, где $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,m})$ – вектор признаков, Y_i – объясняемая переменная, u_i – ошибка (как правило, с нормальным законом распределения с математическим ожиданием 0 и конечной дисперсией). Для прогнозирования событий регрессионными методами в качестве объясняемой переменной нужно выбрать время между событиями $Y_i = \Delta_i = t_{i+1} - t_i$. Наблюдения можно формировать из признаков $x_{i,j}$, которые известны в момент времени возникновения событий t_i. Это опять могут быть какие-нибудь индикаторы $P_{1,t}, P_{2,t}$, … (временные ряды, если такие есть) или сконструированные признаки, например $\Delta_{i-1} = t_i - t_{i-1} = Y_{i-1}$ (лаговые эндогенные переменные), характеристики предшествующих событий y_i или $\frac{y_i}{t_{i+1}-t_i}$. Признаки можно формировать не по одному событию, а по нескольким, например по трем событиям (в этом случае первое наблюдение уже не будет соответствовать первому событию), рисунок 3.



Источник: составлено автором.

Рисунок 3 – Пример составления первого наблюдения

Все векторы X_i группируются в одну матрицу X, а интервалы времени между событиями Y_i в столбец Y. Всего в обучающей выборке будет $n-k$ наблюдений, где n – количество событий, k – количество событий, участвующих в формировании одного наблюдения. Далее модель обучается на основе одного из регрессионных методов, например, с помощью МНК, градиентного бустинга, нейросетевых моделей или других методов машинного обучения. Для прогноза события надо сформировать еще один вектор признаков X_{n+1-k}, который соответствует последнему событию. Формируется он так же, как и для обучающей выборки. Далее с помощью обученной модели прогнозируется время до следующего события $\hat{\Delta}_n$. Наконец, получается прогноз момента появления будущего события $\hat{t}_{n+1}$. В отличие от методов классификации методами регрессии можно предсказать характеристики будущего события $\hat{y}_{n+1}$. Для этого необходимо строить отдельную модель (можно ограничиться одной моделью, если метод регрессии способен прогнозировать несколько выходных переменных, например, нейросети с несколькими выходными нейронами). Наблюдениями будут все те же векторы X_i, а объясняемой переменной значения $Y_i = y_i$. После обучения модели, подставляя наблюдение X_{n+1-k}, получаем прогноз $\hat{y}_{n+1}$.

4) При использовании метода на основе потоков событий из теории случайных процессов по выборке событий определяется закон распределения для времени между событиями T_i, которые предполагаются независимыми случайными величинами. Для получения прогноза будущего события надо откладывать моду (наиболее вероятное значение) закона распределения. Для пуассоновского потока, у которого интервалы между событиями подчиняются экспоненциальному распределению, строго говоря, надо ожидать событие через 0 единиц времени, т.е. сразу же. С теоретической точки зрения пуассоновские потоки не подходят для прогнозирования событий (из-за отсутствия последействия). Но при сильном желании можно от предыдущего события откладывать математическое ожидание $\hat{t}_{n+1} = t_n + 1/\lambda(t_n)$, где $\lambda(t_n)$ – интенсивность потока в момент возникновения последнего события. Интенсивность потока $\lambda(t)$ можно оценить с помощью сглаживания ядерными функциями по выборке событий. Если использовать поток Пальма, который всегда стационарен, с нормальным распределением, то достаточно откладывать среднее значение $\overline{\Delta}$ интервалов между событиями $\hat{t}_{n+1} = t_n + \overline{\Delta}$. Если использовать поток Пальма с распределением Эрланга, то прогнозом события будет $\hat{t}_{n+1} = t_n +$ $+\overline{\Delta}\,(k-1)/k$, где k – характеристика распределения Эрланга (количество складываемых экспоненциальных величин). Для других законов распределения надо тем или иным способом определять характеристики и откладывать моду от последнего события. Если известно количество времени, прошедшее спустя предыдущее событие, то можно откладывать от текущего времени моду распределения, сдвинутого на соответствующую величину влево, но

результат будет таким же, как если бы от последнего события откладывать моду несдвинутого распределения.

5) Рассматриваются методы Кростона и Виллемейна, которые в основном используются для определения объемов товарных запасов при прерывистом спросе. Для предсказания события с помощью метода Кростона в начале надо представить события в виде временного ряда. Для этого разбить ось времени на малые интервалы, желательно, чтобы на интервал попадало не более одного события, и подсчитать количество событий в каждом интервале. Рассчитать количество интервалов между ненулевыми интервалами τ_i (например, количество дней между событиями $\tau_i = t_{i+1} - t_i$). Сгладить полученный ряд τ_i с помощью экспоненциального сглаживания $T_i = (1-\alpha)T_{i-1} + \alpha\tau_i$. Отложить последнее значение сглаженного ряда T_i от последнего ненулевого интервала, $\hat{t}_{n+1} = t_n + T_N$, где N – индекс последнего значения сглаженного ряда. В методе Виллемейна аналитические расчеты упрощаются до того, что от последнего события необходимо откладывать $\frac{1}{P_{01}}$, где P_{01} – вероятность, что после нулевого спроса возникнет ненулевой спрос, $\hat{t}_{n+1} = t_n + \frac{1}{P_{01}} \cdot \Delta$, где Δ – ширина интервалов времени, например день.

6) Рассмотрены границы применимости и недостатки существующих методов: 1 - При конструировании искусственных признаков (частота, среднее время) теряется информация, так как теряются сами индивидуальные события. 2 - В методах классификации не учитывается, на какой участок интервала выпадает событие (когда $Y_i = 1$) или как сильно событие промахивается мимо интервала (когда $Y_i = 0$). 3 - Сложно выбирать ширину интервала для фиксации факта появления события. 4 - Очень часто методы машинного обучения при прогнозировании событий применяются неправильно. Строки обучающей матрицы являются не событиями, а записями о клиентах, относящимися к тому же интервалу времени. Тем самым модели учатся по записям об одних клиентах предсказывать события для других клиентов, причем на том же интервале времени. 5 - При валидации часто совершают ошибку, получая высокие метрики качества, когда модель проверяется на способность предсказывать события в настоящем, а не в будущем. 6 - Оцененные параметры моделей, например $a_0, a_1, a_2, a_3, \ldots$, показывают среднее влияние признаков за все время наблюдений, запечатленное в обучающей матрице. Но нет гарантии, что влияние признаков со временем не изменится. Логично, чтобы параметры были функцией времени $a_0(t), a_1(t), a_2(t), a_3(t), \ldots$ (добавление признака времени не решает проблему). 7 - Большинство моделей машинного обучения имеют характер черного ящика, и их затруднительно интерпретировать. 8 - Сами признаки могут не нести информацию о возникновении событий в будущем либо просто отсутствовать. 9 - Потоками событий нельзя объяснить причины, почему интервал между событиями отклоняется от своего среднего

значения в большую или меньшую сторону. 10 - Пуассоновскими потоками некорректно предсказывать будущие события. 11 - Потоки Пальма всегда стационарны, поэтому все будущие события будут идти с одинаковыми интервалами. 12 - Определение параметров для нестационарных потоков может быть затруднительным. 13 - В потоках событий не строятся модели, никакие признаки не используются, игнорируется какая-либо причина появления событий. 14 - Методы Кростона и Виллемейна предназначались в основном для предсказания агрегированных характеристик, таких как суммарный спрос за интересующий период времени, т.е. сразу сумма многих событий. 15 - События в них предполагаются случайными, анализа причин возникновения не происходит. 16 - Если событие не выпало в прошлый период, то вероятность выпадения в следующем периоде не изменяется. 17 - В методе Виллемейна события моделируются марковской цепью с помощью разыгрывания случайных чисел. 18 - Переходные вероятности не меняются со временем (статичные). Рассмотрены другие методы, отмечено, что они не подходят для прогнозирования событий.

Во второй главе диссертации получены следующие результаты:

1) Описывается новый подход для прогнозирования событий, который основывается на построении и восстановлении параметров алгоритмических моделей механизмов образования событий. Под механизмом понимается детерминированная последовательность операторов над внутренними переменными и параметрами. Внутренние переменные X_i изменяются операторами внутри механизма. Параметры P_i изменяются вовне механизма (от внешних условий). Параметры P_i могут быть статичные или динамичные. Часть из параметров P_i может быть известными параметрами, другая часть – неизвестными (искомыми) параметрами. Внутренние переменные X_i инициализируются в самом начале работы (моделирования) механизма. Механизм может состоять из арифметических операторов, операторов ветвления, элементарных функций и оператора формирования событий, являющегося ключевым элементом. Механизм функционирует во времени, т.е. реализован некоторый цикл, в каждой итерации которого продвигается время на один шаг вперед, например на день.

Предлагаемый метод прогнозирования событий заключается в выполнении пяти этапов:

1. подготовить данные, то есть разделить события по разным выборкам в зависимости от того, каким механизмом они образованы;
2. предположить модель механизма $Program(X, P)$, в результате которой образуются события в источнике (внутренние переменные $X = \{X_i\}$ инициализируются внутри модели);
3. по имеющейся выборке событий $(t_i; y_i)$ восстановить неизвестные статичные или динамичные параметры $P = \{P_i\}$ (*для одних моделей* $Program(X, P)$ *восстановление неизвестных параметров* $\{P_i\}$ *возможно осуществить аналитическими методами, для других моделей возможно понадобиться использовать численную оптимизацию*);

4. произвести экстраполяцию параметров механизма P;

5. получить прогноз будущих событий $(t_i; y_i)$, моделируя механизм их образования $Program(X, P)$ с установленными значениями параметров.

Первый этап зачастую выполнен автоматически, если имеющиеся события относятся только к одному источнику (например клиенту).

На втором этапе надо описать модель механизма образования событий (словами, математическими формулами или в виде алгоритма). В качестве **сквозного примера** рассматривается модель, как в системах управления запасами, которую можно описать словами: что ежедневно из запасов вычитается величина спроса, в конце дня осуществляется проверка оставшегося запаса, и если запасы опустились ниже критического уровня, то формируется событие – заказ на пополнение запасов до фиксированного максимума (который для простоты прибывает мгновенно и восполняет запасы). Запись алгоритма будет следующей:

1) Пока $(t \le t_{end})$
2) $X = X - P_1(t)$
3) Если $(X \le P_2)$, то
4) Создать событие $(t, P_3 - X)$
5) $X = P_3$
6) Продвинуть время t и обновить параметр $P_1(t)$
7) Конец цикла

где X – запасы (внутренняя переменная), $P_1(t)$ – это нестационарный и, возможно, случайный спрос (динамический параметр), P_2, P_3 – критический уровень и максимальный уровень запасов (статические параметры).

В математической форме можно сказать, что момент возникновения следующего события t_{i+1} определяется в соответствии с функцией верхнего предела (1)

$$t_{i+1} = t, \qquad t\colon \int_{t_i}^{t} P_1(\tau)d\tau = y_i, \tag{1}$$

где y_i – величина предыдущего заказа в событии $(t_i; y_i)$.

Заметим, что в этом примере не моделируется расход собственных запасов, а, наоборот, по выборке событий $(t_i; y_i)$ требуется определить динамический параметр $P_1(t)$ и два статических параметра P_2 и P_3 у неподконтрольного клиента. То есть необходимо решить обратную задачу.

Описанная модель механизма в сквозном примере представляет собой нечто больше, чем просто случай из задачи логистики. Многие механизмы, где событие образуется в результате сравнения внутренней переменной с некоторым значением, можно описать этой моделью (так как изменение внутренней переменной приближенно можно представить как накопление значений ее первой производной). В ходе работы над диссертацией были рассмотрены различные вариации этого механизма образования событий в торговле, в сфере услуг, при анализе социальных или исторических событий, в сфере финансов. Многие модели являются частным случаем или развитием этой простой модели. Такими моделями могут быть тривиальные, например, модель таймера, где запас времени убывает с постоянной скоростью, или более сложные. Например, модель с изменяющимся максимальным уровнем, модель с периодической проверкой запаса, модель с дополнительными событиями и др.

На третьем этапе необходимо по выборке событий $(t_i; y_i)$ восстановить параметры механизма образования событий (параметры $P_1(t)$, P_2, P_3 в сквозном примере). Параметры P_2 и P_3 связаны: если одновременно изменить каждый из них на одинаковую величину, то результат не поменяется. Поэтому в качестве параметра P_2 можно брать произвольное положительное значение, например $P_2 = 0$. Параметр P_3 рассчитывается по простой формуле (2)

$$P_3 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \frac{P_1(t_i)}{2}\right). \tag{2}$$

Для восстановления динамического параметра $P_1(t)$ приходим к задаче восстановления функции по последовательности интегралов y_i. Рассмотрено несколько способов. Наглядно показано, что классические регрессионные методы и современные методы машинного обучения плохо подходят для этой задачи. Для этого хорошо подходят методы сплайновой коллокации.

2) Представлен способ восстановления функции одновременно по разным функционалам как непосредственно самим значениям, так и первым производным, вторым производным и определенным интегралам. Решается следующая оптимизационная задача (3).

$$\begin{gathered} S(g) = \sum_{i=1}^{n_f} w_i^f \big(y_i - g(t_i)\big)^2 + \mu \sum_{j=1}^{n_{df}} w_j^{df} \left(y_j' - g'(t_j)\right)^2 + \nu \sum_{l=1}^{n_{d^2f}} w_l^{d^2f} \big(y_l'' - g''(t_l)\big)^2 + \\ + \psi \sum_{u=1}^{n_{int}} w_u^{int} \left(Y_u - \int_{t_u^a}^{t_u^b} g(t)dt\right)^2 + \alpha \int_{t_{start}}^{t_{end}} \big(g''(t)\big)^2 dt \to \min , \end{gathered} \tag{3}$$

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

μ – вес всей группы наблюдений первых производных;

ν – вес всей группы наблюдений вторых производных;

ψ – вес всей группы наблюдений интегралов;

последнее слагаемое $\int_{t_{start}}^{t_{end}} \big(g''(t)\big)^2 dt$ – штраф на нелинейность (шероховатость);

α – коэффициент сглаживания (регуляризации);

t_{start} и t_{end} – соответственно границы, в которых происходит восстановление функции;

y_i, y'_j, y''_l, Y_u – наблюдаемые (с погрешностью) значения, значения первой и второй производной, значения определенных интегралов;

t_i, t_j, t_l, t_u^a, t_u^b – моменты наблюдений и границы интегралов;

n_f, n_{df}, n_{d^2f}, n_{int} – количество соответствующих наблюдений;

$g(t)$ – кубический сплайн, который задается через само значение $g_k = g(s_k)$ и значение его второй производной $\gamma_k = g''(s_k)$ в узлах сплайна $s_1 < s_2 < \cdots < s_m$, где m – количество узлов (такое представление также называют как *value-second derivative representation* или как представление через наклоны и моменты).

Все функционалы выражаются через параметры сплайна g_k и γ_k.

Функционал $g(t_i)$ определяется по формуле (4)

$$g(t) = \frac{h_k^{-t}}{h_k} g_{k+1} + \frac{h_k^{+t}}{h_k} g_k - \frac{h_k^{-t} h_k^{+t} (h_k + h_k^{-t})}{6h_k} \gamma_{k+1} - \frac{h_k^{-t} h_k^{+t} (h_k + h_k^{+t})}{6h_k} \gamma_k, \qquad (4)$$
$$k{:}\, s_k \le t < s_{k+1};$$

где $h_k = s_{k+1} - s_k$ – шаг между узлами (возможно, разный);

$h_k^{-t} = t - s_k$ – расстояние от узла s_k до момента t;

$h_k^{+t} = s_{k+1} - t$ – расстояние от момента t до следующего узла s_{k+1}.

Функционал $g'(t_j)$ определяется по формуле (5)

$$g'(t_j) = \frac{g_{k+1}}{h_k} - \frac{g_k}{h_k} - \left(\frac{h_k}{6} - \frac{\left(h_k^{-t_j}\right)^2}{2h_k} \right) \gamma_{k+1} + \left(\frac{h_k}{6} - \frac{\left(h_k^{+t_j}\right)^2}{2h_k} \right) \gamma_k \,. \qquad (5)$$

Функционал $g''(t_j)$ определяется по формуле (6)

$$g''(t_l) = \frac{h_k^{-t_l}}{h_k} \gamma_{k+1} + \frac{h_k^{+t_l}}{h_k} \gamma_k \,. \qquad (6)$$

Функционал $\int_{t_u^a}^{t_u^b} g(t)dt$ рассчитывается по сложной формуле (7). В начале надо определить, на какой интервал k выпал нижний предел интегрирования t_u^a и на какой интервал $k + L$ выпал верхний предел интегрирования, где L – количество интервалов между ними ($L = 0$, если оба предела выпали на один интервал)

$$\int_{t_u^a}^{t_u^b} g(t)dt = \sum_{l=0}^{L}\left[\frac{h_{k+l}}{2}g_{k+l+1} + \frac{h_{k+l}}{2}g_{k+l} - \frac{h_{k+l}^3}{24}\gamma_{k+l+1} - \frac{h_{k+l}^3}{24}\gamma_{k+l}\right] -$$
$$-\frac{\left(h_k^{-t_u^a}\right)^2}{2h_k}g_{k+1} - \frac{h_k^2 - \left(h_k^{+t_u^a}\right)^2}{2h_k}g_k -$$
$$-\frac{\left(h_k^{-t_u^a}\right)^2\left(\left(h_k^{-t_u^a}\right)^2 - 2h_k^2\right)}{24h_k}\gamma_{k+1} + \frac{\left(h_k^{-t_u^a}\right)^2\left(h_k^{+t_u^a} + h_k\right)^2}{24h_k}\gamma_k - \tag{7}$$
$$-\frac{h_{k+L}^2 - \left(h_{k+L}^{-t_u^b}\right)^2}{2h_{k+L}}g_{k+L+1} - \frac{\left(h_{k+L}^{+t_u^b}\right)^2}{2h_{k+L}}g_{k+L} +$$
$$+\frac{\left(h_{k+L}^{+t_u^b}\right)^2\left(h_{k+L}^{-t_u^b} + h_{k+L}\right)^2}{24h_{k+L}}\gamma_{k+L+1} - \frac{\left(h_{k+L}^{+t_u^b}\right)^2\left(\left(h_{k+L}^{+t_u^b}\right)^2 - 2h_{k+L}^2\right)}{24h_{k+L}}\gamma_{k+L},$$
$$L: s_{k+L} < t_u^b \leq s_{k+L+1},$$
$$k: s_k \leq t_u^a < s_{k+1},$$

где $h_k^{-t_u^a} = t_u^a - s_k$, $h_k^{+t_u^a} = s_{k+1} - t_u^a$, $h_k = s_{k+1} - s_k$, $h_{k+L}^{-t_u^b} = t_u^b - s_{k+L}$, $h_{k+L}^{+t_u^b} = s_{k+L+1} - t_u^b$, $h_{k+L} = s_{k+L+1} - s_{k+L}$.

Выражения (4), (5), (6) и (7) линейны по отношению к неизвестным параметрам сплайна g_k и γ_k. Это позволяет выразить оптимизационную задачу (3) в матричном виде (8)

$$\begin{gathered} S(g) = \left(Y_f - V_f g + P_f\gamma\right)^T W_f\left(Y_f - V_f g + P_f\gamma\right) + \\ +\mu\left(Y_{df} - V_{df}g + P_{df}\gamma\right)^T W_{df}\left(Y_{df} - V_{df}g + P_{df}\gamma\right) + \\ +\nu\left(Y_{d^2f} - 0g + P_{d^2f}\gamma\right)^T W_{d^2f}\left(Y_{d^2f} - 0g + P_{d^2f}\gamma\right) + \\ +\psi(Y_{int} - V_{int}g + P_{int}\gamma)^T W_{int}(Y_{int} - V_{int}g + P_{int}\gamma) + \\ +\alpha g^T K g \to \min, \end{gathered} \tag{8}$$

где $Y_f, Y_{df}, Y_{d^2f}, Y_{int}$ – столбцы наблюдений;

V_f, V_{df}, V_{int} – матрицы коэффициентов при неизвестных g_k (размерностью $n_f \times m$, $n_{df} \times m$ и $n_{int} \times m$ соответственно);

P_f, P_{df}, P_{d^2f}, P_{int} – матрицы коэффициентов при неизвестных γ_k (размерностью $n_f \times (m-2)$, $n_{df} \times (m-2)$, $n_{d^2f} \times (m-2)$ и $n_{int} \times (m-2)$ соответственно);

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

$K = QR^{-1}Q^T$ симметричная матрица размерностью $m \times m$;

Q – трехдиагональная матрица размерностью $m \times (m-2)$ с элементами $Q_{j-1,j} = h_{j-1}^{-1}$, $Q_{j,j} = -h_{j-1}^{-1} - h_j^{-1}$, $Q_{j+1,j} = h_j^{-1}$ (при $j = 2, \dots, m-1$);

R – трехдиагональная матрица размерностью $(m-2) \times (m-2)$ с элементами $R_{j,j} = \frac{(h_{j-1}+h_j)}{3}$, $R_{j+1,j} = R_{j,j+1} = \frac{h_j}{6}$ (при $j = 2, \dots, m-1$).

Как заполняется каждая матрица, подробно показано в работе.

Благодаря условиям непрерывности сплайна в его узлах, которые выражаются через уравнение в матричной форме $Q^T g = R\gamma$, можно выразить одну из переменных $\gamma = R^{-1}Q^T g$. В результате в оптимизационной задаче остается только одна неизвестная g, формула (9)

$$\begin{gathered} S(g) = \left(Y_f - C_f g\right)^T W_f \left(Y_f - C_f g\right) + \\ + \mu\left(Y_{df} - C_{df} g\right)^T W_{df}\left(Y_{df} - C_{df} g\right) + \\ + \nu\left(Y_{d^2f} - C_{d^2f}\, g\right)^T W_{d^2f}\left(Y_{d^2f} - C_{d^2f} g\right) + \\ + \psi(Y_{int} - C_{int} g)^T W_{int}(Y_{int} - C_{int} g) + \\ + \alpha g^T K g \to min, \end{gathered} \tag{9}$$

где $C_f = V_f - P_f R^{-1} Q^T$, $C_{df} = V_{df} - P_{df} R^{-1} Q^T$, $C_{d^2f} = 0 - P_{d^2f} R^{-1} Q^T$, $C_{int} = V_{int} - P_{int} R^{-1} Q^T$ – матрицы размерностями $n_f \times m$, $n_{df} \times m$, $n_{d^2f} \times m$, $n_{int} \times m$ соответственно.

Решением этой задачи будет выражение (10)

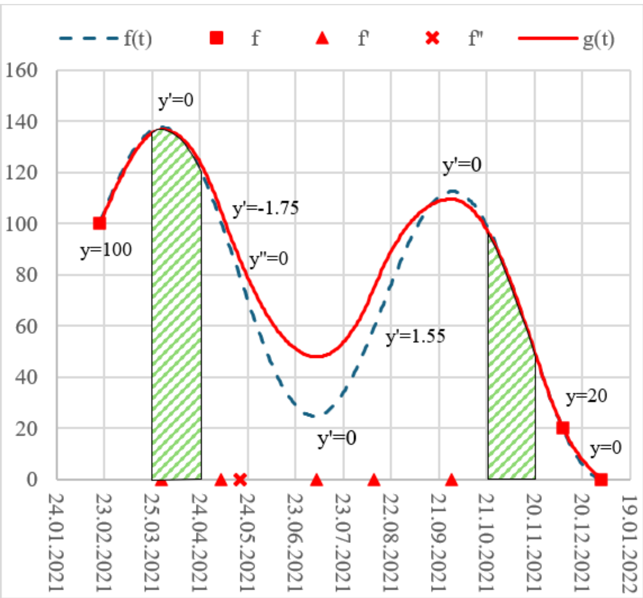
$$\begin{gathered} g = \left(C_f^T W_f C_f + \mu C_{df}^T W_{df} C_{df} + \nu C_{d^2f}^T W_{d^2f} C_{d^2f} + \psi C_{int}^T W_{int} C_{int} + \alpha K\right)^{-1} \times \\ \times \left(C_f^T W_f Y_f + \mu C_{df}^T W_{df} Y_{df} + \nu C_{d^2f}^T W_{d^2f} Y_{d^2f} + \psi C_{int}^T W_{int} Y_{int}\right). \end{gathered} \tag{10}$$

Зная g и рассчитав $\gamma = R^{-1}Q^T g$, можно определить значение сплайна $g(t)$ в любой точке t по выражению (4). Пример восстановления функции по данным, таблица 1, изображен на рисунке 4.

Таблица 1 – Данные разных функционалов

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 4 – Восстановление функции по разным функционалам

3) Проведено исследование по выбору оптимального коэффициента сглаживания α. Оказалось, что Метод L-кривой не работает для нашей задачи. Выражение для оценки значения кросс-валидации $CV(\alpha)$ при заданном коэффициенте сглаживания α будет следующим (11)

$$\begin{gathered} CV(\alpha) = n_f^{-1} \sum_{i=1}^{n_f} w_i^f \left(\frac{y_i - \sum_{k=1}^m C_{ik}^f g_k}{1 - \sum_{k=1}^m C_{ik}^f A_{ki}^f(\alpha)} \right)^2 + n_{df}^{-1} \sum_{j=1}^{n_{df}} w_j^{df} \left(\frac{y_j' - \sum_{k=1}^m C_{jk}^{df} g_k}{1 - \sum_{k=1}^m C_{jk}^{df} A_{kj}^{df}(\alpha)} \right)^2 + \\ + n_{d^2f}^{-1} \sum_{l=1}^{n_{d^2f}} w_l^{d^2f} \left(\frac{y_l'' - \sum_{k=1}^m C_{lk}^{d^2f} g_k}{1 - \sum_{k=1}^m C_{lk}^{d^2f} A_{kl}^{d^2f}(\alpha)} \right)^2 + n_{int}^{-1} \sum_{u=1}^{n_{int}} w_u^{int} \left(\frac{Y_u - \sum_{k=1}^m C_{uk}^{int} g_k}{1 - \sum_{k=1}^m C_{uk}^{int} A_{ku}^{int}(\alpha)} \right)^2, \end{gathered} \tag{11}$$

где матрицы $C^f, C^{df}, C^{d^2f}, C^{int}$ такие же, как в формуле (9). Матрицы $A^f(\alpha) = A(\alpha) C_f^T W_f$, $A^{df}(\alpha) = A(\alpha)\mu C_{df}^T W_{df}$, $A^{d^2f}(\alpha) = A(\alpha)\nu C_{d^2f}^T W_{d^2f}$, $A^{int}(\alpha) = A(\alpha)\psi C_{int}^T W_{int}$,

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

Минимизация $CV(\alpha)$ по α дает искомое значение параметра сглаживания α. Однако при восстановлении функции по интегралам в половине случаев α оказывается либо сильно завышенным, либо сильно заниженным.

Другим способом выбора α является принцип невязки Морозова (или рекомендация Реинша), в котором предлагается брать коэффициент сглаживания α на таком уровне, чтобы он сглаживал наблюдения на величину, соответствующую уровню шума σ_ε в исходных данных. В результате исследования оптимальный коэффициент сглаживания приглушал исходные наблюдения приблизительно на 65% от уровня шума σ_ε. Например, если уровень шума исходных наблюдений σ_ε ожидается как 5%, то надо выбрать такой коэффициент сглаживания α, чтобы после сглаживания наблюдения приглушились на $0{,}65 \cdot 5 = 3{,}25\%$. Под приглушением наблюдений $\sigma_y(\alpha)$ имеется в виду, насколько уменьшилось прогнозное значение наблюдения $\hat{y}_i$ (например, интегралов) относительно исходного y_i. Причем приглушение $\sigma_y(\alpha)$ лучше отсчитывать относительно прогнозных значений $\hat{y}_i$, формулы (12) и (13)

$$\hat{y}_i = [Cg]_i = \sum_{k=1}^m C_{ik} g_k, \tag{12}$$

$$\sigma_y(\alpha) = \left(n^{-1} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2 \right)^{0,5} \times 100\%, \tag{13}$$

где под матрицей C понимается одна из матриц C_f, C_{df}, C_{d^2f}, C_{int}, а под n – соответствующее количество наблюдений. Другой подход выбора оптимального гиперпараметра α основывается на точности прогноза самих событий, о чем будет рассказано ниже.

4) В прикладных задачах требуется, чтобы восстанавливаемая функция принимала лишь неотрицательные значения или была монотонно возрастающей. Под монотонностью функции можно понимать неотрицательность ее производной. В диссертационном исследовании

исследовано несколько способов обеспечения положительности функции. Первый способ – берется экспонента от сплайна $f(x) = e^{g(x)}$. Однако у такого подхода есть недостатки: тяжело аналитически выражать функционалы, оптимизационная задача становится нелинейной по параметрам, вид восстанавливаемой функции не всегда визуально удовлетворительный.

В другом подходе для обеспечения положительности/монотонности задача сводится к задаче квадратичного программирования, в которой добавляются условия неотрицательности в узлах сплайна $g(s_k) = g_k \geq 0$, формула (14)

$$\begin{gathered} S(g) \to \min, \\ g_k \geq 0, \qquad k = 1, \dots, m. \end{gathered} \tag{14}$$

В классической форме задача квадратичного программирования будет следующей (15)

$$\begin{gathered} S(g) = c^T g + \frac{1}{2} g^T D g \to \min \\ -g \leq 0 \end{gathered}, \tag{15}$$

где $c = -\left(C_f^T W_f Y_f + \mu C_{df}^T W_{df} Y_{df} + \nu C_{d^2f}^T W_{d^2f} Y_{d^2f} + \psi C_{int}^T W_{int} Y_{int}\right),$

$D = C_f^T W_f C_f + \mu C_{df}^T W_{df} C_{df} + \nu C_{d^2f}^T W_{d^2f} C_{d^2f} + \psi C_{int}^T W_{int} C_{int} + \alpha K.$

После применения метода множителей Лагранжа и теоремы Каруша-Куна-Такера получена система уравнений (16)

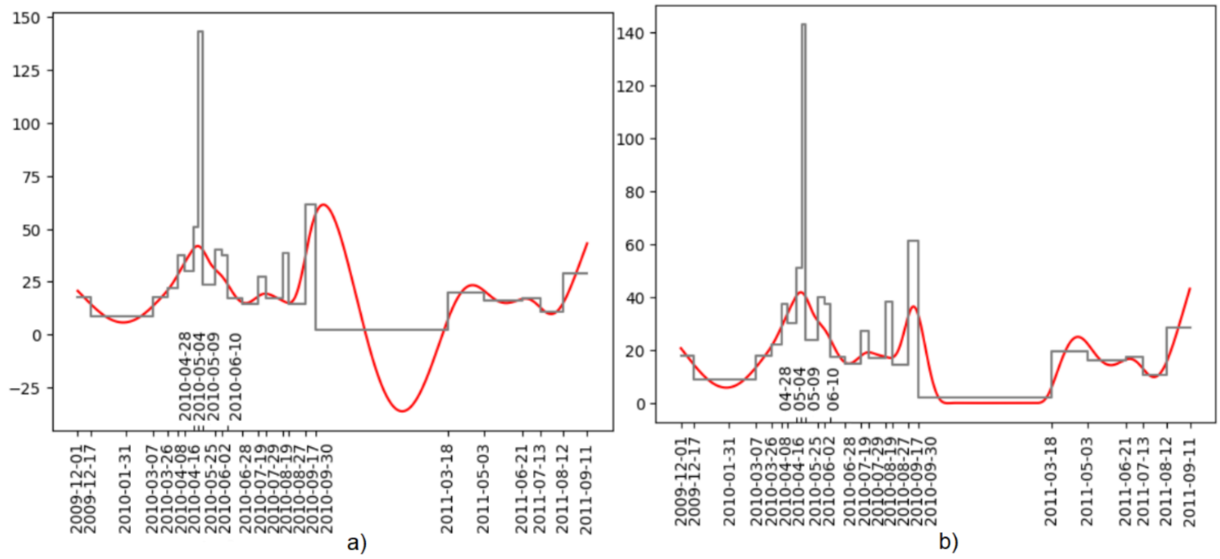
$$\begin{cases} c + Dg - \lambda = 0 \\ \lambda^T g = 0 \end{cases}. \tag{16}$$

Данная система уравнений относится к линейной задаче о дополнительности (linear complementarity problem, LCP), которая решается аналитически методом Лемке или приближенно методом внутренней точки. В результате определяются параметры g сплайна, а вслед за ними $\gamma = R^{-1} Q^T g$. Пример восстановления функции по последовательности интегралов, таблица 2, показан на рисунке 5.

Таблица 2 – Наблюдаемые с погрешностью интегралы

t_i	Y_i	t_i	Y_i	t_i	Y_i	t_i	Y_i
01.12.2009	285,06	28.04.2010	305,2	19.07.2010	274	03.05.2011	797,26
17.12.2009	391,44	04.05.2010	715,03	29.07.2010	361,16	21.06.2011	379,72
31.01.2010	304,47	09.05.2010	380,7	19.08.2010	307,46	13.07.2011	317,4
07.03.2010	337,74	25.05.2010	321,02	27.08.2010	303	12.08.2011	860,98
26.03.2010	286,58	02.06.2010	300,25	17.09.2010	797,14	11.09.2011	905,99
08.04.2010	297,94	10.06.2010	311,26	30.09.2010	354,03	-	-
16.04.2010	363,12	28.06.2010	308,1	18.03.2011	904,26	-	-

Источник: составлено автором.



Ступенчатая линия показывает среднее значение $Y_i/(t_{i+1} - t_i)$, площадь под каждой ступенькой является значением определенного интеграла Y_i.

Источник: составлено автором.

Рисунок 5 – Восстановленная по интегралам функция: a) без условий неотрицательности, b) с условиями неотрицательности.

Тем не менее между узлами сплайн по-прежнему может принимать отрицательные значения. Чтобы гарантировать неотрицательность, между узлами сплайна можно либо добавлять новые узлы с условием неотрицательности в них, либо добавлять новые условия неотрицательности без добавления новых узлов сплайна.

Между двумя узлами s_k и s_{k+1} у полинома третьей степени возможна только одна экстремальная точка минимума. Для ее нахождения решается квадратное уравнение (17)

$$g'(s_k + \Delta t) = g'_k + g''_k \Delta t + g'''_k \Delta t^2/2 = 0\,, \tag{17}$$

относительно Δt, где $g'_k = (g_{k+1} - g_k)/h_k - (\gamma_{k+1} + 2\gamma_k)h_k/6$, $g''_k = \gamma_k$, $g'''_k = (\gamma_{k+1} - \gamma_k)/h_k$. При неотрицательном дискриминанте $d = \gamma_k^2 - 2g'_k g'''_k$ возможно два действительных корня (18)

$$\Delta t = \left(-\gamma_k \pm \sqrt{d}\right)/g'''_k\,. \tag{18}$$

Если $\Delta t \in [0, h_k]$, то рассчитываем значение сплайна $g(s_k + \Delta t)$ по формуле (4), иначе переходим к следующей паре узлов s_k, s_{k+1}. Если рассчитанное значение $g(s_k + \Delta t)$ меньше нуля (или некоторого порога, например минус 0.0001), то в точке $s_k + \Delta t$ добавляем либо новый узел, либо новое условие неотрицательности без добавления нового узла. При добавлении новых узлов просто меняется последовательность $\{s_k\}$. Для добавления условий неотрицательности без добавления новых узлов записываем $g(s_k + \Delta t)$ в матричном виде (19)

$$g(s_k + \Delta t) = V_{cond}g - P_{cond}\gamma = (V_{cond} - P_{cond}R^{-1}Q^T)g\,, \tag{19}$$

где V_{cond} – строка размерностью $1 \times m$ с двумя не нулевыми элементами $V_{cond,k} = \frac{h_k - \Delta t}{h_k}$ и $V_{cond,k+1} = \frac{\Delta t}{h_k}$,

P_{cond} – строка размерностью $1 \times (m-2)$ с двумя не нулевыми элементами $P_{cond,k} = \frac{\Delta t(h_k - \Delta t)(h_k + \Delta t)}{6h_k}$ и $P_{cond,k+1} = \frac{\Delta t(h_k - \Delta t)(2h_k - \Delta t)}{6h_k}$.

Выражение $V_{cond} - P_{cond}R^{-1}Q^T$ является строкой размерностью $1 \times m$. Если добавляется несколько условий (пусть количество условий r), то из этих строк собирается матрица Z размерностью $r \times m$. Тогда оптимизационная задача запишется как (20)

$$\begin{gathered} S(g) = c^T g + \frac{1}{2} g^T D g \to \min \ , \\ Mg \leq 0 \end{gathered} \tag{20}$$

где $M = \begin{bmatrix} -I \\ -Z \end{bmatrix}$ блочная матрица размерностью $(m+r) \times m$, где I – единичная матрица размерностью $m \times m$.

Оптимизационная задача решается либо аналитически методом Лемке, либо численно методом внутренней точки.

Описанная выше процедура добавления узлов или добавления дополнительных условий неотрицательности является многошаговой, так как после одной итерации между старым узлом и новым узлом/условием неотрицательности функция по-прежнему может опуститься хоть ненамного, но ниже нуля. В работе подробно описано, как заполняются все матрицы.

Для получения интегральной функции $F(t) = \int_{-\infty}^{t} f(t)dt$, которая будет монотонно возрастающей, достаточно адаптировать (7) к произвольной точке t (полученное выражение является полиномом четвертой степени), формула (21)

$$\begin{gathered} F(t) = \int_{s_1}^{t} g(t)dt = \sum_{l=1}^{L} \left[\frac{h_l}{2} g_{l+1} + \frac{h_l}{2} g_l - \frac{h_l^3}{24} \gamma_{l+1} - \frac{h_l^3}{24} \gamma_l \right] - \\ - \frac{h_L^2 - (h_L^{-t})^2}{2h_L} g_{L+1} - \frac{(h_L^{+t})^2}{2h_L} g_L + \\ + \frac{(h_L^{+t})^2 (h_L^{-t} + h_L)^2}{24 h_L} \gamma_{L+1} - \frac{(h_L^{+t})^2 ((h_L^{+t})^2 - 2h_L^2)}{24 h_L} \gamma_L, \\ L: s_L < t \leq \ s_{L+1}, \end{gathered} \tag{21}$$

где L – номер интервала, содержащий t, $h_L = s_{L+1} - s_L$, $h_L^{-t} = t - s_L$, $h_L^{+t} = s_{L+1} - t$.

Программная реализация описанного способа построения положительного/монотонного сплайна реализована на языке Python в виде функции FunctionalSmoothingSpline и доступна на GitHub по адресу: https://github.com/YuAKorablev/FunctionalSmoothingSpline.

Вышесказанное относилось к третьему этапу предлагаемого подхода прогнозирования событий, на котором по имеющейся выборке событий $(t_i; y_i)$ восстанавливаются параметры механизма образования событий $P = \{P_i\}$. С помощью монотонной сплайновой коллокации можно восстанавливать динамические параметры, такие как $P_1(t)$ в приведенном выше примере.

5) Предложено несколько способов экстраполяции динамических параметров (этап экстраполяции является четвертым в предлагаемом подходе, в сквозном примере надо произвести экстраполяцию параметра $P_1(t)$):

- Если при восстановлении параметров использовались регрессионные методы или методы машинного обучения, то достаточно на вход обученных моделей подавать новые значения t или сформированный вектор признаков X, соответствующий будущему времени (заметим: сплайн на концах обращается в прямую линию, экстраполяция подстановкой другого значения t не годится при использовании сплайновой коллокации).

- Можно переносить тенденции, привязанные к праздничным датам или временам года, с предыдущего года на следующий год.

- Раскладывать функцию на сумму ограниченного количества гармонических функций с помощью алгоритма Куинна и Фернандеса, но при этом отбрасывать по n_1 и n_2 точек сплайна слева и справа (n_1 и n_2 будут являться очередными гиперпараметрами). Имеется программная реализация на языках R и Python.

- Использовать экспертные оценки других исследователей и специалистов.

На пятом этапе происходит моделирование самого механизма образования событий с установленными значениями параметров для получения прогноза будущего события. Предварительно надо установить значения внутренних переменных такими, какими они должны быть в момент появления последнего события обучающей выборки (для примера выше внутренняя переменная инициализируется как $X = y_n - 0.5P_1(t_n)$).

6) Для определения гиперпараметров (таких как α, n_1 и n_2) разработана специальная процедура, базирующаяся на методе глобальной оптимизации с помощью поиска на сетке и на методе локальной оптимизации с помощью алгоритма Нелдера-Мида. В качестве критерия используется ошибка прогноза $Score$ самого события $(\hat{t}_{n+1}, \hat{y}_{n+1})$, формула (22)

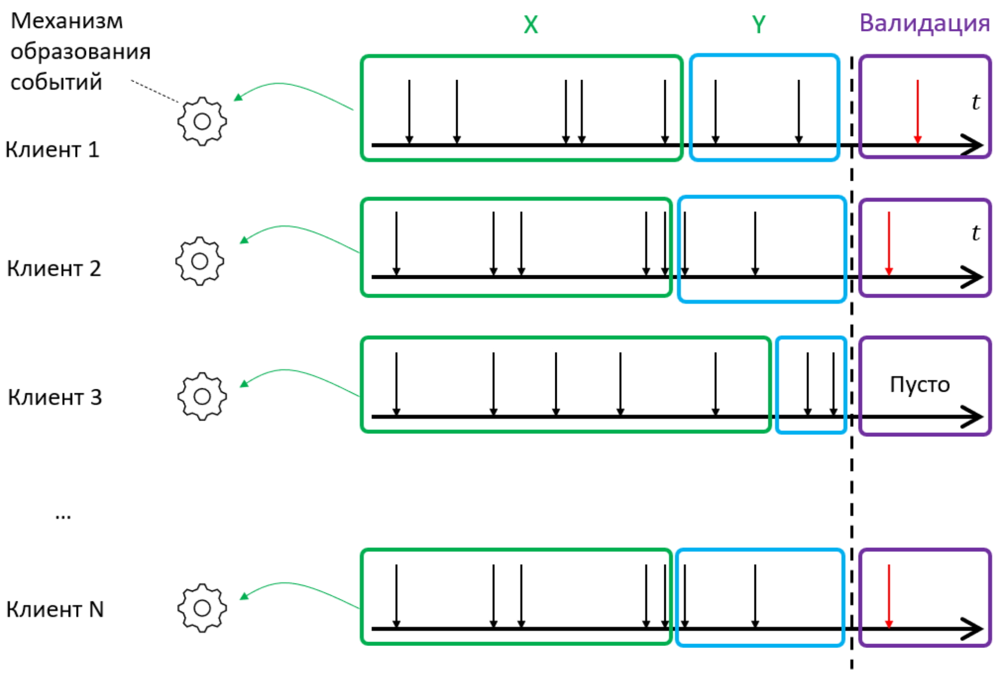
$$Score = \left|\frac{\hat{t}_{n+1} - t_{n+1}}{t_{n+1} - t_n}\right| + \mu \cdot \left|\frac{\hat{y}_{n+1} - y_{n+1}}{y_{n+1}}\right|, \tag{22}$$

где t_{n+1} и y_{n+1} – фактические значения будущего события;

$t_{n+1} - t_n$ – фактическая ширина интервала между событиями;

μ – весовой коэффициент, выбирается эмпирически, например 0.1.

Однако надо использовать ошибку прогноза не одного события, а нескольких, например $(Score1 + Score2)/2$. Чем больше событий используется для оценки ошибки, тем меньше их остается в обучающей выборке. Схема обучения в новом подходе изображена на рисунке 6. Возможны другие метрики, например, среднее квадратичное абсолютное или относительное отклонение, причем характеристики событий могут являться векторами $y_i = (y_{i,1}, y_{i,2}, \dots, y_{i,k})$, таблица 3.



Источник: составлено автором.
Рисунок 6 – Схема обучения в новом подходе прогнозирования событий

Таблица 3 – Метрики оценки качества (погрешности) прогноза N событий

Метрика	y_i – скаляр	$y_i = (y_{i,1}, y_{i,2}, \dots, y_{i,k})$ – вектор
mean absolute error	$\frac{1}{\sum w_i}\sum_{i=1}^{N} w_i(\lvert\hat{t}_i - t_i\rvert + \mu\lvert\hat{y}_i - y_i\rvert)$	$\frac{1}{\sum w_i}\sum_{i=1}^{N} w_i\left(\lvert\hat{t}_i - t_i\rvert + \mu\frac{1}{k}\sum_{j=1}^{k}\lvert\hat{y}_{i,j} - y_{i,j}\rvert\right)$
mean relative error	$\frac{1}{\sum w_i}\sum_{i=1}^{N} w_i\left(\left\lvert\frac{\hat{t}_i - t_i}{t_i - t_{i-1}}\right\rvert + \mu\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert\right)$	$\frac{1}{\sum w_i}\sum_{i=1}^{N} w_i\left(\left\lvert\frac{\hat{t}_i - t_i}{t_i - t_{i-1}}\right\rvert + \mu\frac{1}{k}\sum_{j=1}^{k}\left\lvert\frac{\hat{y}_{i,j} - y_{i,j}}{y_{i,j}}\right\rvert\right)$
root mean squared error	$\sqrt{\frac{1}{\sum w_i}\sum_{i=1}^{N} w_i(\lvert\hat{t}_i - t_i\rvert + \mu\lvert\hat{y}_i - y_i\rvert)^2}$	$\sqrt{\frac{1}{\sum w_i}\sum_{i=1}^{N} w_i\left(\lvert\hat{t}_i - t_i\rvert + \mu\sqrt{\frac{1}{k}\sum_{j=1}^{k}(\hat{y}_{i,j} - y_{i,j})^2}\right)^2}$
root mean squared relative error	$\sqrt{\frac{1}{\sum w_i}\sum_{i=1}^{N} w_i\left(\left\lvert\frac{\hat{t}_i - t_i}{t_i - t_{i-1}}\right\rvert + \mu\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert\right)^2}$	$\sqrt{\frac{1}{\sum w_i}\sum_{i=1}^{N} w_i\left(\left\lvert\frac{\hat{t}_i - t_i}{t_i - t_{i-1}}\right\rvert + \mu\sqrt{\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\hat{y}_{i,j} - y_{i,j}}{y_{i,j}}\right)^2}\right)^2}$

Источник: составлено автором.

7) Для проверки адекватности модели механизма образования событий предложены три способа построения доверительного интервала для момента времени будущего события $\hat{t}_{n+1}$.

Первый способ основан на расчете погрешности восстановления параметров сплайна во время третьего этапа. Если в сквозном примере функция восстанавливается по интегралам, то ковариационная матрица для прогнозируемых интегралов будет (23)

$$\mathrm{Cov}(\hat{y}_{int}) = C_{int}^{new} D^{-1}(C_{int}^{T} W_{int} W_{int}^{T} C_{int} \hat{\sigma}_{int}^{2}) D^{-1^{T}} C_{int}^{new^{T}} + \hat{\sigma}_{u}^{2}, \tag{23}$$

где C_{int}^{new} – матрица размерностью $n^{new} \times m$, которая заполняется, как и C_{int} в формуле (9) (где $k = m - 1$, $L = 0$, $t_u^a = s_m$, $t_u^b = \hat{t}_{n+1}$), но для новых n^{new} наблюдений, $D = C_{int}^T W_{int} C_{int} + \alpha K$, $\sigma_u^2 = \sum_{i=1}^{n} u_i^2 / (n - m - 1)$, где $u_i = y_i - \hat{y}_i$ – разница наблюдаемых и прогнозных интегралов.

Однако, так как на концах сплайн ведет себя как прямая линия, для экстраполяции применялись другие методы (раскладывали функцию на гармоники), расчет $\hat{y}_{n+1}^{new} = [C_{int}^{new} g]_{i=n+1}$ будет некорректным для точки $\hat{t}_{n+1}$. Также m превышает n (расчет σ_u^2 становится невозможен), и для оценки m параметров по n наблюдениям используется регуляризация, из-за чего оценки становятся смещенными, формулу (23) для смещенных оценок использовать некорректно. Но можно допустить, что дисперсия нового интеграла $\hat{y}_{n+1}$ будет соответствовать средней дисперсии прогнозов обучающих интегралов $\sigma_u^2 = \sum_{i=1}^{n} u_i^2 / n$. Другим подходом будет вмешательство в алгоритм Куинна и Фернандеса разложения функции на гармоники на этапе экстраполяции. После определения регрессионными методами амплитуды всех синусов и косинусов, можно аналогично формуле (23) определить дисперсию функции в новых точках, откуда посчитать дисперсию всего интеграла $\hat{y}_{n+1}$ как сумму дисперсий за интервал времени между t_n и $\hat{t}_{n+1}$. Зная дисперсию интеграла $\hat{y}_{n+1}$, можно предположить, что и момент времени возникновения события $\hat{t}_{n+1}$ будет обладать такой же относительной дисперсией. То есть, если отклонение интеграла составляет 5%, то и отклонение времени составляет 5%.

Второй способ построения доверительного интервала основан на имитационном моделировании, когда в зависимости от разного уровня шума в наблюдениях определялись погрешности $R_{sqr}(\alpha)$ восстановления функции в виде сплайна $g(t, \alpha)$, а также уровень приглушения наблюдений $\sigma_y(\alpha)$ (в процентах) при разных α, формулы (24) и (25).

$$R_{sqr}(\alpha) = \sqrt{\frac{1}{t_n - t_1} \sum_{t=t_1}^{t_n} \left(\frac{f(t) - g(t, \alpha)}{f(t)} \right)^2}, \tag{24}$$

$$\sigma_y(\alpha) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} \left(\frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2} \times 100\%. \tag{25}$$

Построена зависимость погрешности $R_{sqr}(a^*)$ восстановления функции от уровня приглушения $\sigma_y(a^*)$ при оптимальном коэффициенте сглаживания $a^* = \arg\min_{\alpha} R_{sqr}(\alpha)$, формула (26)

$$ln(\hat{R}_{sqr}) = \begin{matrix} -3{,}0175 + 0{,}49116 \cdot ln(\sigma_y) + \varepsilon \\ (0{,}02948) \quad (0{,}01414) \qquad\quad (0{,}41482) \end{matrix}. \tag{26}$$

После определенных аналитических выкладок получена дисперсия для восстановленной функции в зависимости от уровня приглушения наблюдений σ_y, формула (27)

$$D_{\hat{f}(t)} = \left(0{,}04892 \cdot \sigma_y^{0{,}49116}\right)^2 \left(1{,}17294 + 0{,}01414^2 \cdot \left[Ln\left(\sigma_y\right)\right]^2\right). \qquad (27)$$

Предположив, что между событиями восстановленная функция $\hat{f}(t)$ отклоняется от исходной $f(t)$ в одну и ту же сторону, получаем, что квадрат относительных отклонений интеграла и функции совпадает $D_{\hat{y}_{n+1}} = D_{\hat{f}(t)}$. Квадрат относительного отклонения для времени будущего события $D_{\hat{t}_{n+1}}$ соответствует квадрату относительного отклонения для интеграла $D_{\hat{t}_{n+1}} = D_{\hat{y}_{n+1}} = D_{\hat{f}(t)}$. Для построения доверительного интервала используем $\sigma_{\hat{t}_{n+1}} = \sqrt{D_{\hat{t}_{n+1}}}$. Например, если уровень приглушения наблюдений σ_y составит 5%, то $\sigma_{\hat{t}_{n+1}} = 0{,}11683$ или 11,7% (а при $\sigma_y = 10\%$ имеем $\sigma_{\hat{t}_{n+1}} = 16{,}42\%$). Доверительный интервал в 95% приблизительно будет (28)

$$t_{n+1} \in \left[\hat{t}_{n+1} - 2 \cdot \widehat{\Delta t} \cdot \sigma_{\hat{t}_{n+1}}; \hat{t}_{n+1} + 2 \cdot \widehat{\Delta t} \cdot \sigma_{\hat{t}_{n+1}}\right], \qquad (28)$$

где $\widehat{\Delta t} = \hat{t}_{n+1} - t_n$ – предполагаемый интервал между событиями.

При уровне приглушения наблюдений σ_y, равном 5%, доверительный интервал составляет $\pm 23{,}4\%$ от интервала между событиями (при $\sigma_y = 10\%$ получим $\pm 32{,}84\%$). Однако при больших σ_y нижняя граница доверительного интервала может оказаться в прошлом. Предлагается нижнюю границу считать мультипликативно, формула (29)

$$t_{n+1} \in \left[\hat{t}_{n+1} - \widehat{\Delta t} \cdot \left(\frac{\sigma_{\hat{t}_{n+1}}}{1 + 2\sigma_{\hat{t}_{n+1}}}\right); \hat{t}_{n+1} + \widehat{\Delta t} \cdot 2\sigma_{\hat{t}_{n+1}}\right]. \qquad (29)$$

Третий способ основан на простой идее: достаточно дважды получить прогноз будущего события, моделируя механизм образования событий со значениями параметров, которые отклоняются от своих средних значений на величину, соответствующую заданной доверительной вероятности (при занижении в меньшую сторону, чтобы не получить отрицательных значений, лучше занижать мультипликативно).

Также судить об адекватности модели можно по средней ошибке прогноза событий Score для нескольких последних событий (Score рассчитывается, как во время поиска гиперпараметров). Например, если среднее отклонение фактических событий от прогнозных событий меньше 30%, то считать модель адекватной (это будет уже четвертый способ, причем в нем не строятся доверительные интервалы).

В третьей главе получены следующие результаты:

1) Разработан способ восстановления параметров модели механизма образования событий с помощью численных методов. Модель задается в виде алгоритма, в которой есть оператор формирования события, а также функция «Продвинуть время t и обновить параметры».

Динамические параметры задаются кубическим сплайном $g(t)$, который определяется через изменяющиеся скачками значения третьей производной в узлах сплайна s_k и начальные условия. Для этого необходимо:

а) задать узлы сплайна $s_1 < s_2 < \cdots < s_m$;

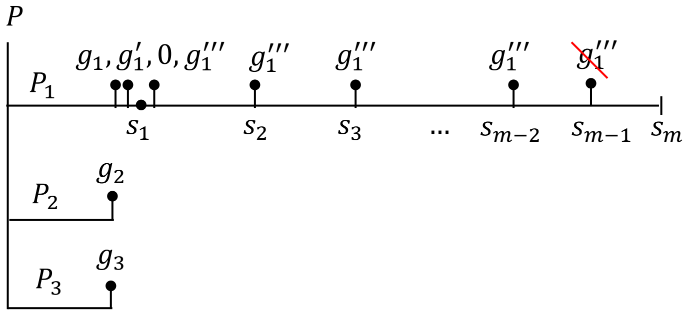
б) задать $g(s_1)$, $g'(s_1)$, $g'''(s_1)$ в самом первом узле s_1 ($g''(s_1)$ можно не задавать, так как для натурального сплайна $g''(s_1) = 0$);

в) задать $g'''(s_k)$ во всех узлах s_k, предшествующих моменту времени t, (k: $s_k < t$).

Такое представление позволяет ограничивать количество неизвестных параметров сплайна временем t, нет необходимости определять другие параметры сплайна, если время t не наступило. Значения $g'''(s_{m-1})$ в предпоследнем узле s_{m-1} не задается, так как его можно выразить через предыдущие по формуле (30)

$$g'''(s_{m-1}) = -\frac{\sum_{k=1}^{m-2} h_k g'''(s_k)}{h_{m-1}}. \tag{30}$$

Для сквозного примера схематичное изображение определяемых значений представлено на рисунке 7.



Источник: составлено автором.

Рисунок 7 – Пример схемы определяемых значений для описания параметров механизма формирования событий, P_1 – динамический параметр, P_2 и P_3 – статичные параметры

В функции «Продвинуть время t и обновить параметры» реализуется расчет нового значения для всех динамических параметров. Значения $g(t)$, $g'(t)$ и $g''(t)$ обновляются относительно значений на предыдущем шаге по формулам (31), (32), (33) (и в указанном порядке)

$$g(t) = g(t) + g'(t) + \frac{1}{2}g''(t) + \frac{1}{6}g'''(s_k), \tag{31}$$

$$g'(t) = g'(t) + g''(t) + \frac{1}{2}g'''(s_k), \tag{32}$$

$$g''(t) = g''(t) + g'''(s_k), \tag{33}$$

где k – номер узла, предшествующий моменту времени t.

Вводится функция потерь (34)

$$S(P) = \sum_{i=2}^{n} \left(\frac{t_i - t_i'}{t_i - t_{i-1}} \right)^2 + \mu \sum_{i=2}^{n} \left(\frac{y_i - y_i'}{y_i} \right)^2 + \sum_{j=1}^{N_P} \alpha_j Pen_j \underset{P}{\to} \min \ , \tag{34}$$

где t_i и t_i' – моменты появления фактических событий и событий, полученных в результате моделирования;

y_i и y_i' – характеристики фактических событий и событий, полученных в результате моделирования,

μ – весовой коэффициент, например $\mu = 0{,}1$ (означает большее внимание к первому критерию);

n – размер выборки событий;

α_j – коэффициент, с которым учитывается штраф;

N_P – количество динамических параметров;

Pen_j – штраф на гладкость для параметра j, который рассчитывается по формуле (35)

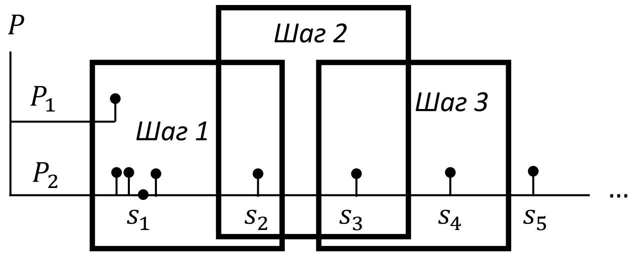
$$Pen = \int_{s_1}^{s_m} \left(g''(t) \right)^2 dt = \sum_{k=1}^{m-1} \frac{\left(g''(s_k) + g'''(s_k) h_k \right)^3 - \left(g''(s_k) \right)^3}{3 g'''(s_k)}, \tag{35}$$

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

При функционировании модели механизма образования событий может быть получена выборка событий, которая по размеру отличается от ожидаемого количества событий (или вообще может быть пустой), в этом случае функция потерь обращается в бесконечность. Моделировать механизм образования событий необходимо чуть дольше, чем время последнего события, чтобы дать шанс сформироваться нужному количеству событий.

Для определения параметров минимизируем функцию потерь $S(P) \to \min$. Однако она является недифференцируемой, и возможно множество локальных оптимумов. Используется комбинация глобальной оптимизации с помощью поиска на сетке и локальной оптимизации с помощью метода Нелдера-Мида для поиска экстремума внутри ячейки сетки. Классический алгоритм Нелдера-Мида модифицирован определенным образом, локальный поиск останавливается, если происходит выход за границы ячейки.

Если для всех неизвестных значений сразу применять оптимизацию на сетке, то получится задача огромной размерности. Для снижения размерности оптимизационная задача превращена в многошаговую. Задается скользящее окно из N событий. На первом шаге подбираются на сетке значения таким образом, чтобы первые N сформировавшихся моделью событий хорошо соответствовали первым N событиям исходной выборки (первое образованное событие будет вторым событием вторым). На следующем шаге скользящее окно как бы сдвигается на одно событие и подбираются на сетке значения, влияющие на образование событий со второго по $N+1$. На рисунке 8 изображены значения параметров, перебираемые на сетке на каждом новом шаге для случая $N=2$ (при условии, что узлы сетки s_k совпадают с наблюдениями).

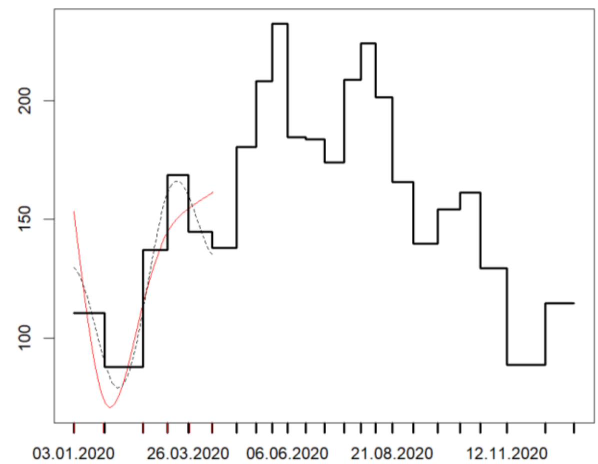


Прямоугольной рамкой изображено скользящее окно. На первом шаге на сетке перебираются комбинации пяти значений. На втором шаге – комбинации двух значений, на третьем – тоже комбинации двух значений.

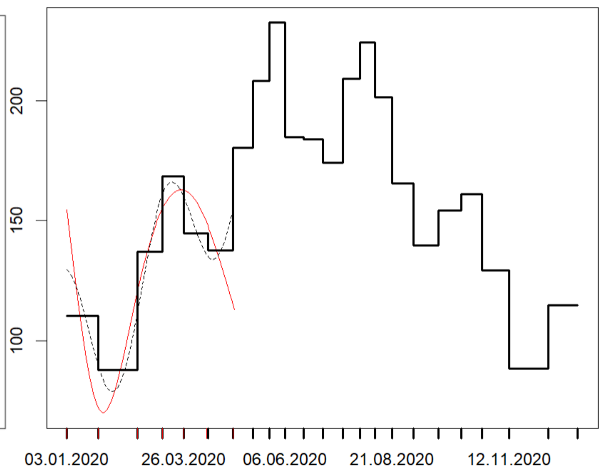
Источник: составлено автором.

Рисунок 8 – Подбор значений на $N=2$ события вперед

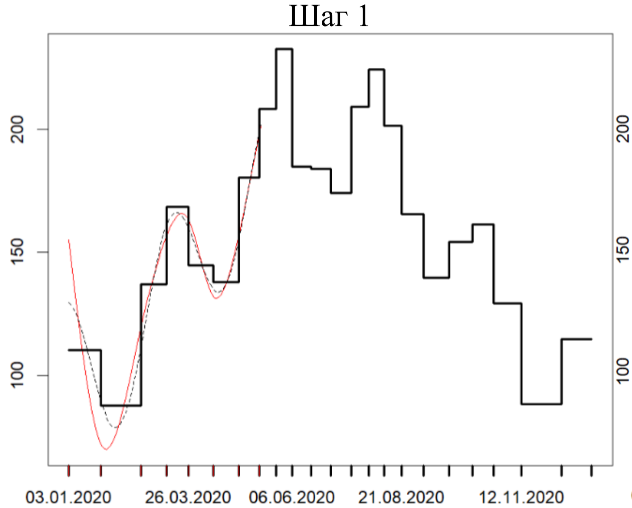
Описанный прием позволяет обойти проблему экспоненциального роста объема вычислений. Пример восстановления динамического параметра ($P_1(t)$ по интегралам y_i) описанным способом представлен на рисунке 9.



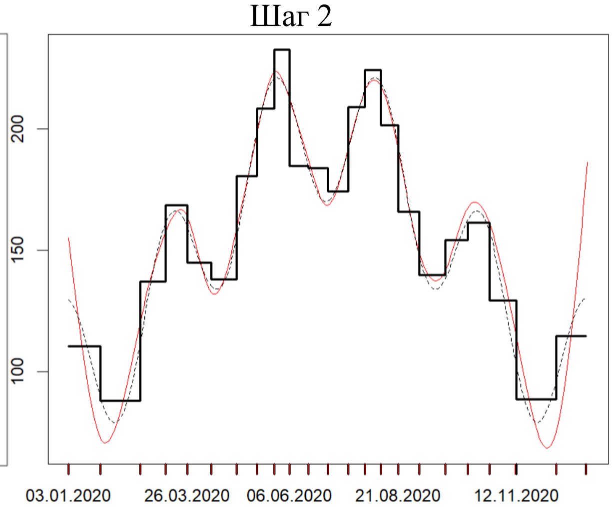
Шаг 1



Шаг 2



Шаг 3



Шаг 18

Источник: составлено автором.

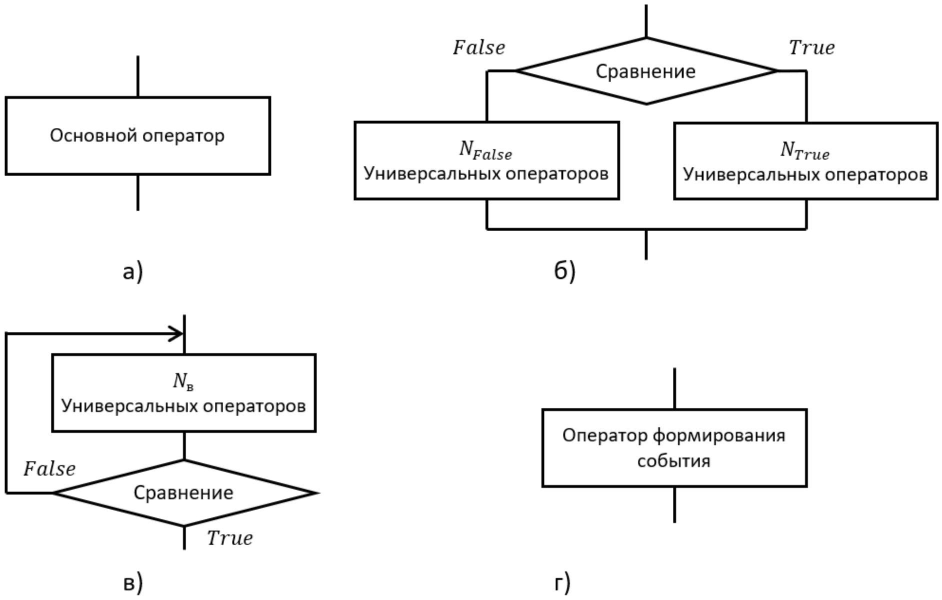
Рисунок 9 – Пример работы алгоритма

2) Предложены способы подбора структуры алгоритмических моделей, в которых кроме параметров модели P надо подобрать модель механизма образования событий $Program(X, P)$, состоящую из операторов. Имеем оптимизационную задачу (36)

$$S(Program, P) \to \min \quad . \tag{36}$$

Первый способ основывается на комбинаторном подходе. В нем предполагается перебирать все возможные конструкции из операторов для составления модели механизма образования событий $Program$ при заданном ограничении на количество операторов, количество внутренних переменных и параметров. Для каждой модели $Program$ происходит восстановление параметров P по имеющейся выборке событий, используя критерий (34).

Модель $Program$ конструируется рекурсивным образом из универсальных операторов, которые могут быть одного из четырех видов: основной оператор, оператор ветвления, оператор цикла и оператор формирования события, рисунок 10. Внутри оператора ветвления могут располагаться N_{true} и N_{false} (в операторе цикла $N_{в}$) других универсальных операторов, которые в свою очередь опять могут быть одного из четырех видов.



а) основной оператор; б) оператор ветвления; в) оператор цикла; г) оператор формирования события.

Источник: составлено автором.

Рисунок 10 – Четыре вида универсального оператора

Основной оператор выполняет одну арифметическую операцию между парой операндов с сохранением результата во внутреннюю переменную. Операндами могут быть внутренние переменные X и параметры P. В операторах ветвления и цикла обязательно осуществляется операция сравнения двух операндов. Оператор формирования события, как и основной оператор, осуществляет одну из арифметических операций над операндами, но вместо сохранения результата во внутреннюю переменную формирует новое событие (t_i', y_i'), значением y_i' которого является полученный результат.

Ограничив количество основных операторов и количество операторов сравнений числом N_{opers} и N_{if}, рекурсивным образом перебираются все возможные конструкции модели. Количество конструкций быстро растет с увеличением N_{opers} и N_{if}, таблица 4.

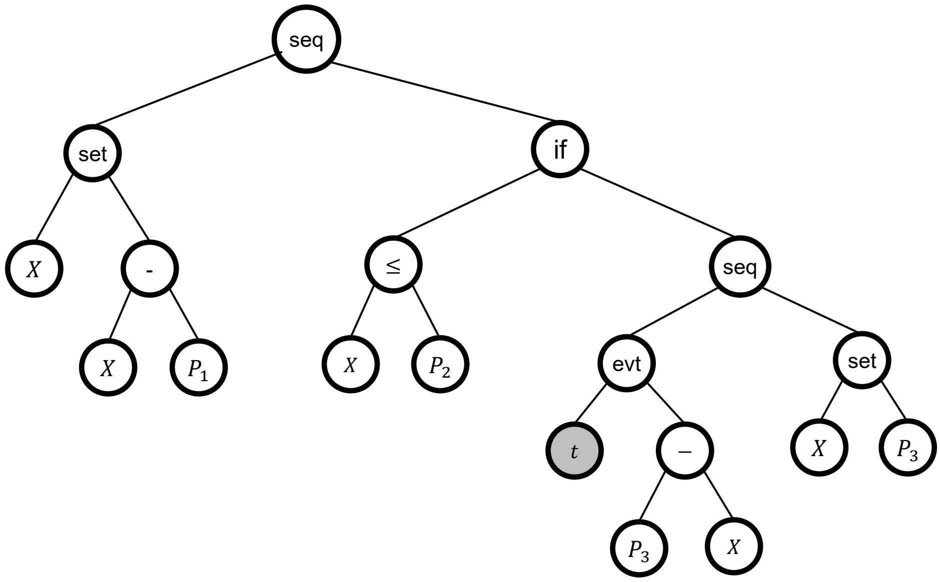
Таблица 4 – Количество конструкций моделей при ограничении на N_{opers} и N_{if}

N_{opers}	N_{if}	Количество конструкций
3	1	30
3	2	450
4	1	70
4	2	1638
5	1	140
5	2	4761

Источник: составлено автором.

Далее для каждой фиксированной схемы перебираются комбинации арифметических операций и операндов (имея N_X внутренних переменных и N_P параметров). Количество комбинаций зависит от составленной схемы и может быть огромным. Например, для схемы, составленной из $N_{opers} = 3$ операторов и $N_{if} = 1$ сравнения, если $N_X = 2$ и $N_P = 2$, получается 10 892 880 комбинаций операторов и операндов. Однако можно значительно уменьшить количество комбинаций, если убрать повторяющиеся, например, когда происходит перестановка операторов сложения местами, перестановка операндов в операциях сложения и умножения и т.д. Но это требует дополнительных исследований. Имеется программная реализация на языке R рекурсивного алгоритма автоматического перебора модели механизма образования событий $Program$ (без фильтрации повторяющихся моделей).

Второй способ основывается на генетическом программировании. В генетическом программировании в результате операций скрещивания и случайных мутаций получается лучший экземпляр компьютерной программы. Модели $Program$ представляются деревьями, где в терминальных вершинах располагаются внутренние переменные X или параметры P, а в узлах дерева располагаются разные функции, среди которых обязательно должен быть оператор формирования событий. Качество построенной программы $Program$ – оценивается во время определения параметров P (в предыдущем разделе) по формуле (34). Дерево программы для сквозного примера изображено на рисунке 11, где set – оператор присвоения, seq – оператор последовательности, evt – оператор формирования событий (t не является терминалом, событие всегда привязывается к текущему времени). Роль генетического программирования в том, чтобы в результате скрещивания (обмена поддеревьями), мутаций и отбора составилось дерево, рисунок 11.



Источник: составлено автором.
Рисунок 11 – Пример модели механизма образования событий в виде дерева

Для инициализации внутренних переменных (нулем или одним из параметров) можно обойтись без генетического программирования. Функционирование механизма во времени программируется циклом с продвижением времени и обновлением динамических параметров после каждой итерации. Поэтому для генетического программирования достаточно оставить только подбор тела цикла. Так как генетическое программирование полагается на случайные числа, есть вероятность, что определенная комбинация операторов не сложится при ограниченном числе циклов эволюции.

В результате появляется принципиальная возможность исследовать совершенно неизвестные события с помощью автоматического подбора алгоритмической модели механизма образования событий, после чего заниматься интерпретацией модели. Однако это требует дальнейших исследований.

В четвертой главе разрабатываются информационные технологии для прогнозирования событий:

1) Разработана библиотека eventometrics на языке Python для прогнозирования событий. Библиотека размещена на официальный репозитарий PyPI для свободного использования. Библиотека имеет интерфейс, похожий на библиотеку машинного обучения sklearn, с тем отличием, что надо использовать связку fit – extrapolate – predict. Основные вычисления оптимизированы с помощью Cython. Используется гибридный подход, когда все классы написаны с помощью языка программирования Python, а критические функции, требующие большого объема вычислений, на языке Cython. Проведена дополнительная оптимизация, пропускающая повторные вычисления у выполненных блоков кода при повторном вызове функций.

Установка библиотеки:

pip install eventometrics

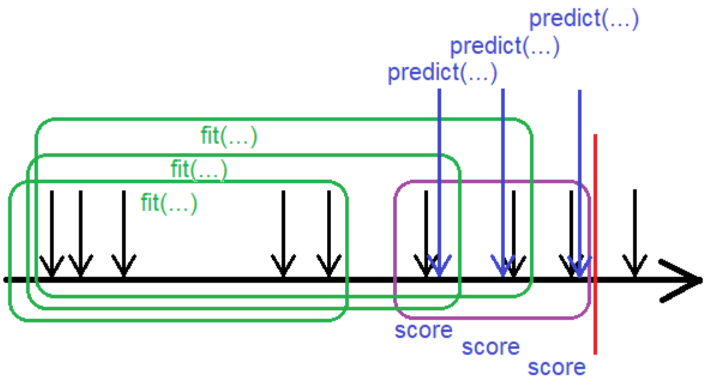
Создание модели, обучение, экстраполяция параметров и прогноз (если этап экстраполяции не вызывается явно, он запускается при вызове функции прогнозирования).

```
from eventometrics.CapacityModels import SimpleCapacity
from eventometrics.Extrapolators import QuinnFernandesExtrapolator

model = SimpleCapacity(t, Y,
                       knots_number=n*3,
                       alpha=10**5,
                       max_recalculate_positive_iterations=1,
                       extrapolator=QuinnFernandesExtrapolator(n1=0, n2=0, Nharm=7)
                       )
model.fit()
t_pred, y_pred = model.predict(n_predict=200, n_events=1)
```

Этап экстраполяции можно было выполнить явно.

```
model.fit()
model.extrapolate_parameters(n_predict=300, n1=10, n2=15, Nharm=8)
t_pred, y_pred = model.predict(n_events=3)
```

Для поиска оптимальных гиперпараметров используется комбинация глобальной оптимизации с помощью поиска на сетке и локальной оптимизации из центра каждой ячейки сетки с помощью алгоритма Нелдера-Мида. На рисунке 12 изображена схема использования данных и сравнения прогнозируемых событий с фактическими (кросс-валидации).



Источник: составлено автором.

Рисунок 12 – Логика использования набора данных при кросс-валидации в eventometrics

Пример определения оптимальных гиперпараметров.

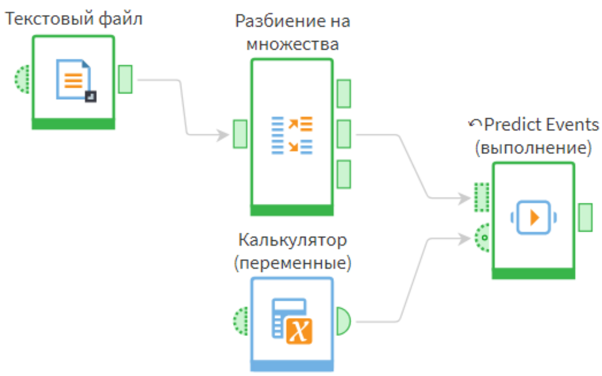
```
from eventometrics.CapacityModels import SimpleCapacity
from eventometrics.Extrapolators import QuinnFernandesExtrapolator
from eventometrics.Search import GridSearchCV
model = SimpleCapacity(t, Y,
                       knots_number=m,
                       extrapolator=QuinnFernandesExtrapolator()
)

param_grid = {"alpha": [0, 10**3, 10**4, 10**5, 10**6, 10**7, np.inf],
              "n1": [0, 10, 20, 30],
              "n2": [0, 10, 20, 30]
              }
gs = GridSearchCV(estimator= model,
                  param_grid=param_grid,
                  scorer="re",
                  extra_scorer="rms",
                  n_jobs=-2,
                  refit=True,
                  num_val_events=3,
                  n_predict=200,
                  mix_weight=0.1,
```

```
                     is_int=["n1","n2"],
                     verbose=1
                     )
gs.fit()
print("best_params = ", gs.best_params)
print("best_value = ", gs.best_value)
t_pred, y_pred = gs.best_estimator.predict(n_predict=200)
```

Разработанная библиотека eventometrics является новым фреймворком прогнозирования событий.

2) Создан компонент прогнозирования событий для популярной платформы бизнес-аналитики Loginom (аналитическая low-code платформа). Компонент называется «Predict Events» и реализует описанную методологию на примере простейшей модели из сквозного примера. Для получения прогноза с помощью созданного компонента надо составить сценарий из узлов, как показано на рисунке 13.

В узле «Калькулятор (переменные)» (или в переменных пользователя) задаются управляющие переменные n_predict и n_events, которые отвечают за количество предсказываемых событий.

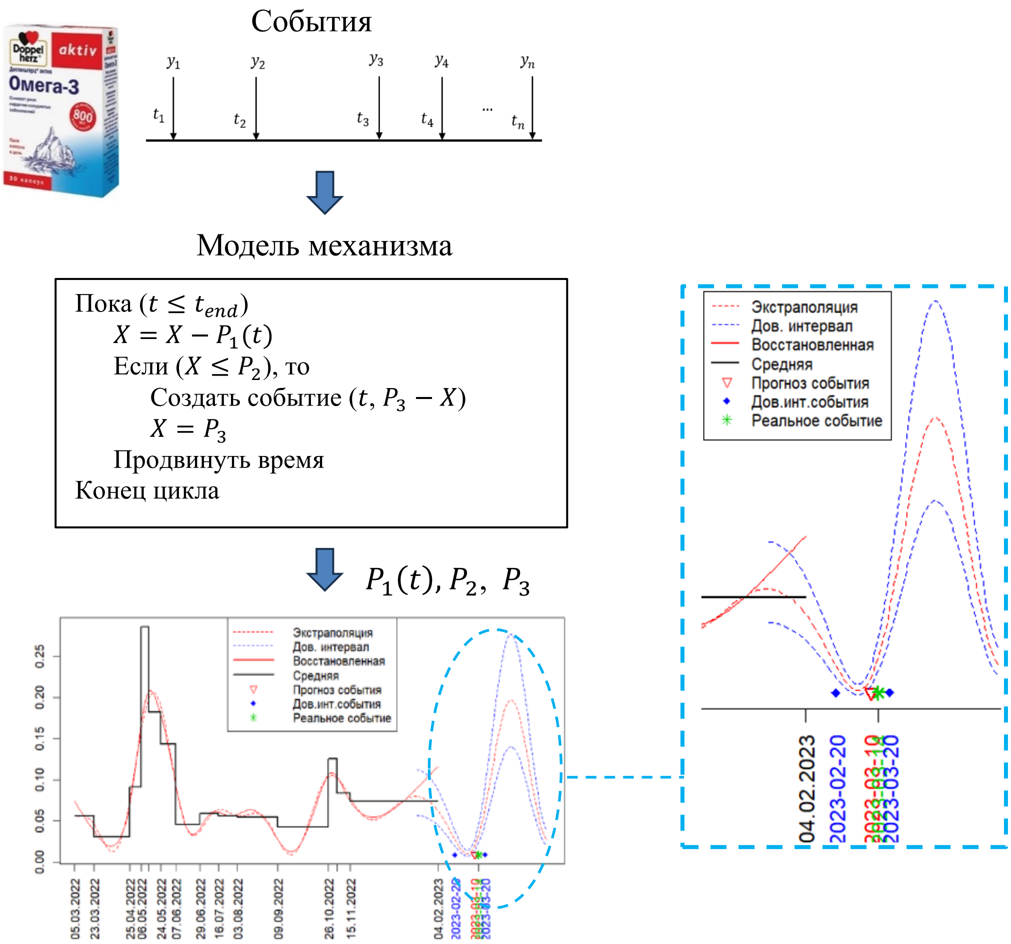


Источник: составлено автором.
Рисунок 13 – Расположение и соединение узлов

Создание компонента подробно описано в работе, благодаря компоненту появляется возможность прогнозировать будущие события сразу из системы бизнес-аналитики Loginom.

В пятой главе получены следующие результаты:

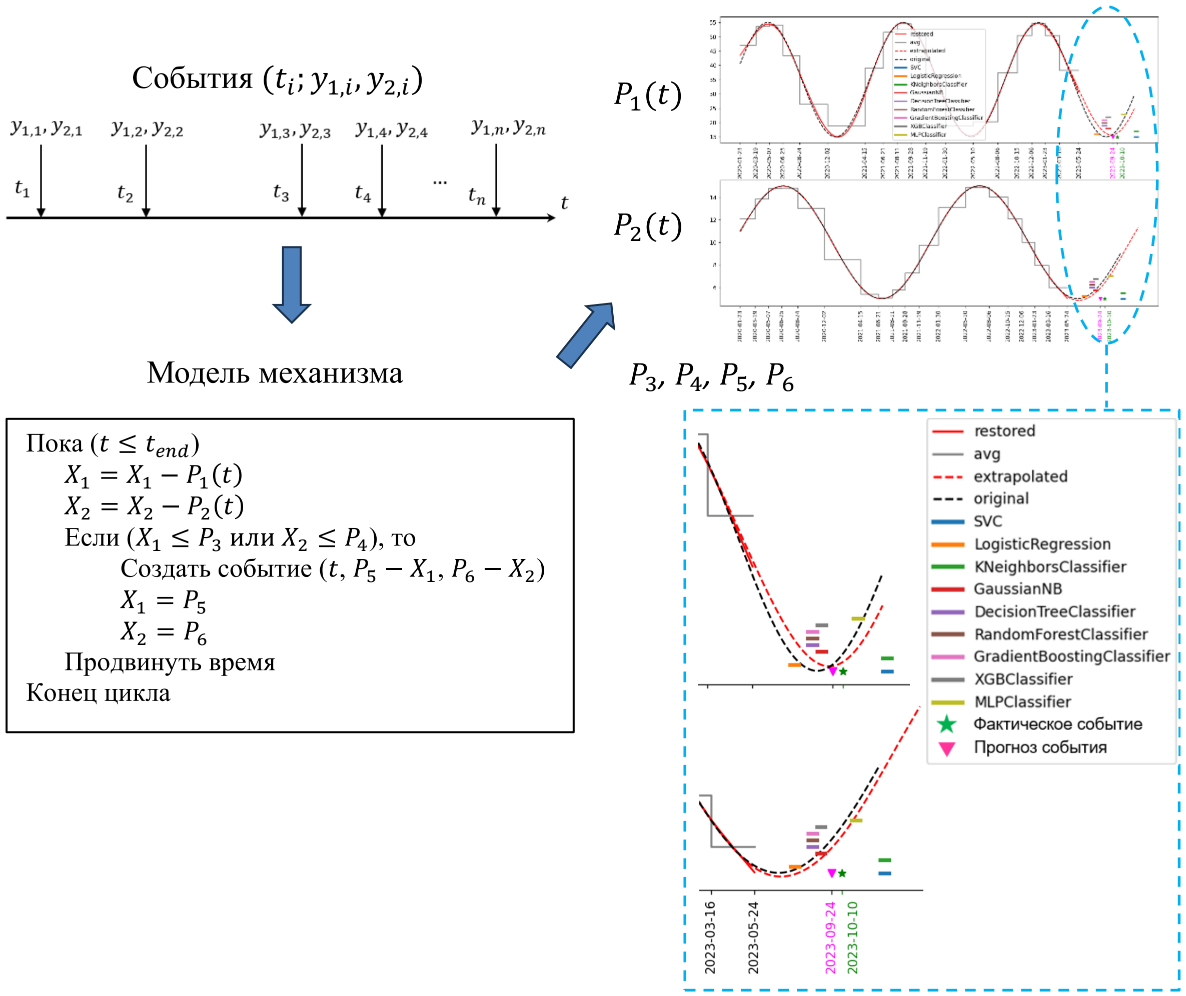
1) Представлены результаты апробации разработанной методологии на данных о поставках биологически активных добавок Доппельгерц Актив Омега-3 российского представительства компании ООО «Квайссер Фарма» в аптечные пункты регионов Российской Федерации, во всех примерах использовалась простейшая модель, рисунок 14.



Источник: составлено автором.

Рисунок 14 – Прогнозирование поставки биологически активных добавок и построение доверительного интервала

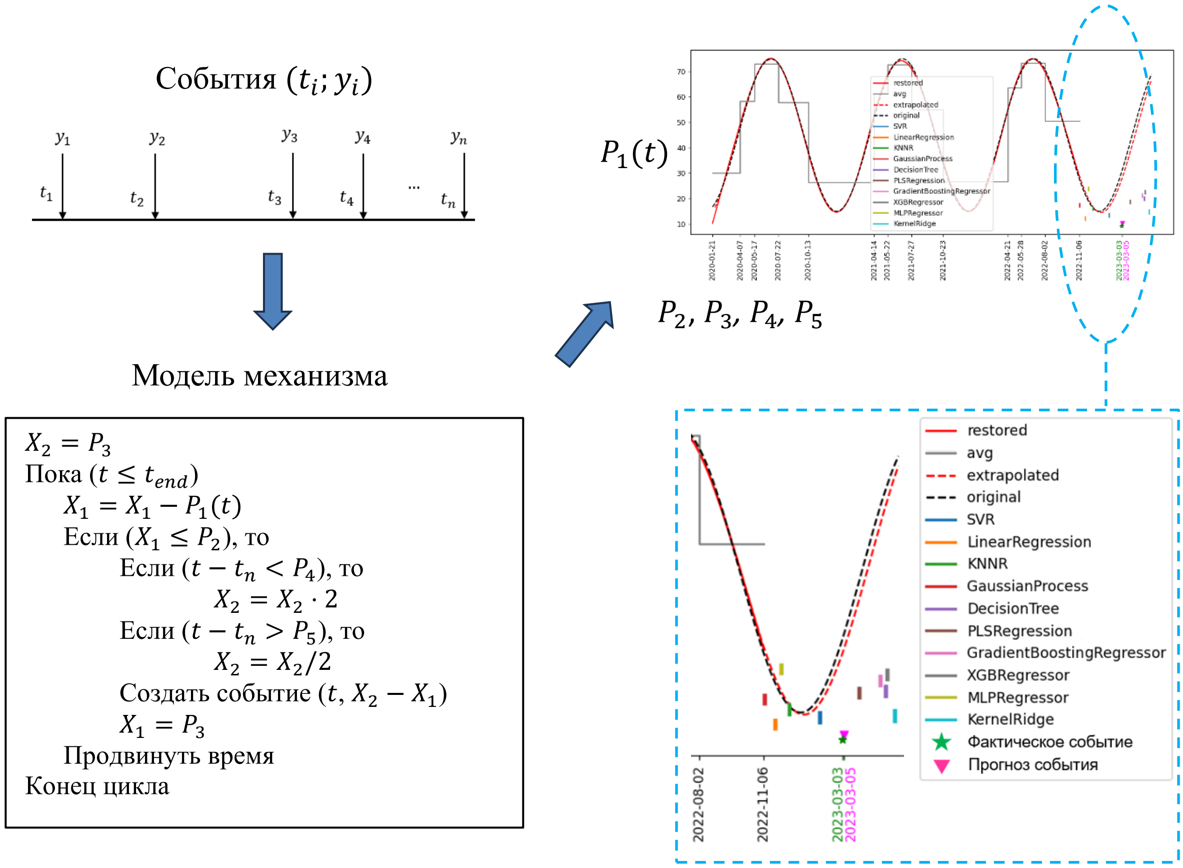
2) Проведено сравнение на различных примерах предлагаемого подхода, основанного на восстановлении механизмов образования событий, с современными методами классификации, такими как SVC, LogisticRegression, KNeighborsClassifier, GaussianNB, DecisionTreeClassifier, RandomForestClassifier, GradientBoostingClassifier, MLPClassifier, XGBClassifier, адаптированными для прогнозирования событий. На рисунке 15 изображен пример, в котором использовалась модель с двумя продуктами (событие возникает, когда закончился любой из двух продуктов, при этом пополняются запасы обоих продуктов). Горизонтальными отрезками изображены интервалы, на которые методами классификации прогнозируется событие. Фактическое событие обозначается зеленой звездочкой ★, прогноз новым подходом – пурпурным треугольником ▼. Методы классификации плохо справляются с предсказанием будущего события (причем используется 24 признака). Дополнительные примеры и подробные пояснения представлены в работе.



Источник: составлено автором.

Рисунок 15 – Прогноз события по модели с двумя продуктами и сравнение с методами классификации

3) Проведено сравнение с современными методами регрессии, такими как LinearRegression, SVR, KernelRidge, KNeighborsRegressor, GaussianProcessRegressor, PLSRegression, DecisionTreeRegressor, GradientBoostingRegressor, MLPRegressor и XGBRegressor, адаптированными для прогнозирования событий. На рисунке 16 изображен пример прогнозирования событий по модели с изменяющимся максимумом (максимальный запас изменяется в зависимости от того, как быстро он расходуется). Вертикальными черточками обозначен прогноз $\hat{t}_{n+1}$ различными регрессионными методами. Все методы регрессии дали очень плохой прогноз, так как не смогли разгадать сложную закономерность. Дополнительные примеры и подробные пояснения представлены в работе.

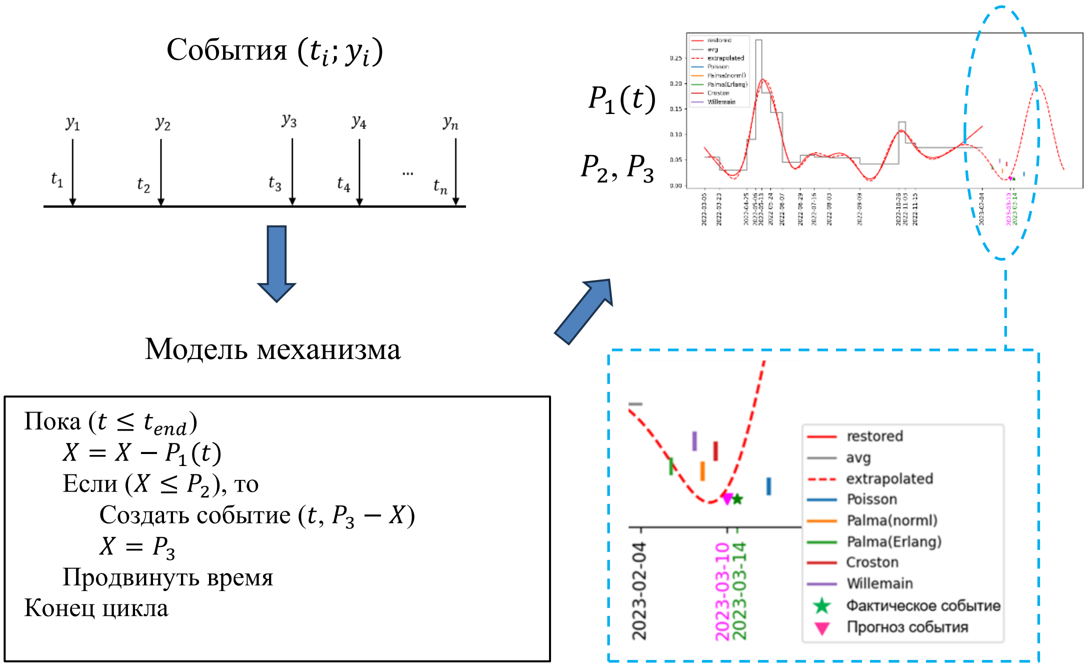


Источник: составлено автором.

Рисунок 16 – Прогноз события по модели с изменяющимся максимумом и сравнение с методами регрессии

4) Проведено сравнение с методами на основе потоков событий (поток Пуассона, поток Пальма с распределением Эрланга и поток Пальма с нормальным законом распределения интервалов между событиями), а также с адаптированными для прогноза событий методами Кростона и Виллемейна. На рисунке 17 изображен пример предсказания события поставки биологически активных добавок в аптечные пункты. Вертикальными черточками обозначен прогноз $\hat{t}_{n+1}$ указанными методами. Прогноз, полученный с помощью восстановления механизмов образования событий, оказался точнее.

В работе представлены дополнительные примеры как на синтетических, так и на реальных данных: пример модели механизма с дополнительными событиями (когда появляются события, не связанные с механизмом образования); модель поставки запчастей для авиатранспорта (когда по данным пополнения запаса запчастей получается определить количество самолетов в воздухе, интенсивность замены деталей как по налету, так и из-за случайных поломок); пример с прогнозом выпуска облигаций; пример прогноза покупок клиентов онлайн-магазина в следующем месяце. Во всех примерах по выборке дискретных событий восстанавливались параметры механизма и делался прогноз будущих событий. Новый подход во всех примерах по точности был конкурентным передовым современным методам и даже превосходил их.



Источник: составлено автором.

Рисунок 17 – Сравнение с методами регрессии, модель с изменяющимся максимумом

5) Представлен экономический эффект применения разработанной методологии прогнозирования событий. Отмечается возможность широкого применения во множестве отраслей, на множестве уровней, в различных аспектах деятельности.

Согласно расчетам, в торговле 1 день содержания запасов обходится в среднем в 0,082% от стоимости товаров, т.е. приблизительно каждые 12 дней хранения обходятся приблизительно в 1% их стоимости. В одном примере было показано, что прогноз (будущей покупки), полученный в соответствии с предложенной методологией, может быть на 62 дня точнее, чем у передового метода XGBRegressor. Такое увеличение точности прогноза эквивалентно снижению затрат в среднем на 5,1% от стоимости хранимых товаров. В другом примере точность прогноза возрастает на 16 дней, что может соответствовать сокращению затрат в среднем на 1,3% стоимости товаров. Причем планировать запасы можно не только со стороны производителя/продавца, а также со стороны поставщика сырья или деталей. Если в среднем содержание запасов составляет 4,4% мирового ВВП, то сокращение их хотя бы на процент может привести к огромной экономии.

Прогнозирование событий открывает новые возможности в маркетинге. Позволяет к дате прогнозируемого события предупредить и скоординировать действия клиентов, например, заблаговременно пополнить те же запасы, причем в нужном объеме. Можно осуществлять маркетинговые действия, привлекать новых клиентов, опираясь их потребности и на прогноз событий. На торговых площадках информация о грядущих событиях позволит игрокам использовать ситуацию в свою пользу, воздействовать на рынок или подстраиваться к будущим изменениям. На уровне государства возможно обеспечить более высокий уровень снабжения территорий за счет более точного планирования в связи с повышением точности прогнозирования событий. Позволяет снизить экономические потери, сохранить рабочие места, обеспечить стабильность критической инфраструктуры и социальную защищенность населения, обеспечить экономическую безопасность. Разработанная методология закладывает основы нового направления исследований – «эвентометрика», что дает мощный толчок в исследовании экономических процессов, позволяет получить новые качественные или количественные выводы в экономических исследованиях.

III Заключение

В диссертационной работе разработана новая методология прогнозирования экономических событий, в том числе редких. Проведен анализ и адаптация методов классификации, регрессии, методов на основе потоков событий, а также методов Кростона и Виллемейна для прогнозирования событий, рассмотрены их недостатки и границы применимости. Предложен новый подход, основанный на восстановлении алгоритмических моделей механизмов образования событий, в которых участвуют внутренние переменные, инициализирующиеся и изменяющиеся внутри механизма, и параметры, изменяющиеся вовне механизма. Параметры могут быть не только статичными, но и динамическими, т.е. функцией времени. Разработаны методы восстановления неизвестных параметров этих алгоритмических моделей. Для относительно простых моделей можно использовать аналитические методы. Для этих целей разработаны математические и инструментальные средства на основе сплайновой коллокации восстановления функции одновременно по разным функционалам, таким как значения функции, значения первой и второй производной, определенные интегралы. Обеспечивается положительность восстанавливаемой функции с помощью сведения к задаче квадратичного программирования или помещения сплайна в степень экспоненты. Для восстановления параметров сложных алгоритмических моделей надо использовать численную оптимизацию. Для этого разработан специальный метод, в котором динамические параметры определяются через третью производную в узлах сплайна и через начальные условия. Эти значения подбираются с помощью комбинации глобальной оптимизации поиска на сетке и локальной оптимизации алгоритмом Нелдера-Мида. Функцией потерь выступает метрика качества прогноза событий как смесь модулей относительных отклонений прогнозируемых событий от фактических и штраф на нелинейность. Используется специальный прием, позволяющий уменьшить размерность получающейся оптимизационной задачи. После того как параметры механизма образования событий восстановлены, необходимо их экстраполировать на будущее. Для этого предлагается несколько способов: использовать регрессионные методы; переносить тенденцию прошлого года на следующий; раскладывать функцию на сумму ограниченного количества гармоник с помощью алгоритма Куинна-Фернандеса. Последний способ используется в большинстве примеров в работе. Наконец, моделируя механизм образования событий с установленными значениями параметров, получаем прогноз будущего события. Для определения оптимальных гиперпараметров применяемых методов восстановления и экстраполяции параметров механизма образования событий используется комбинация поиска на сетке и локальной оптимизации алгоритмом Нелдера-Мида. В качестве целевой функции выступает модуль относительных отклонений нескольких последних событий от фактических. Предложены еще несколько метрик на основе абсолютных и относительных средних или среднеквадратичных отклонений. Для проверки адекватности построенной модели механизма образования событий предложены три способа построения доверительного интервала для момента времени будущего события. Предложены способы подбора самих алгоритмических моделей механизмов образования событий, основанные на комбинаторном подходе и на генетическом программировании. Проведено сравнение на разных примерах нового предлагаемого подхода прогнозирования событий с адаптированными существующими методами. Почти во всех примерах новый подход значительно обходит по точности существующие методы. Делается вывод о том, что новый метод является конкурентным для передовых методов машинного обучения. Все разработанные в диссертационной работе методы и алгоритмы имеют программную реализацию на современных языках программирования. Отдельно надо отметить создание библиотеки eventometrics на языке Python, которая является фреймворком для прогнозирования событий. Библиотека опубликована в репозитарии PyPi для свободного использования. Разработан компонент для системы бизнес-аналитики Loginom, реализующий предложенный новый подход прогнозирования событий с помощью восстановления механизмов их образования.

Поставленные в диссертационном исследовании задачи были решены полностью, а исследование получилось комплексным. В результате исследования заложены основы фундамента для нового направления исследований, которое было названо эвентометрика. Созданы соответствующие технологии для прогнозирования экономических событий. Разработанная методология может создать критически важный научно-практический прорыв в экономических исследованиях. В будущем возможно открытие соответствующего направления обучения студентов.

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. 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.

8. 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.

9. 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.

10. 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.

11. 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).

12. 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.

13. 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.

14. 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.

15. 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).

16. 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).

17. 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.

18. Кораблев Ю.А. Распространенные ошибки использования машинного обучения при прогнозировании событий и новый подход на основе моделей механизмов образования событий / Ю.А. Кораблев, В.А. Судаков // Экономика и математические методы. – 2025. – № 1. Т. 61.– С. 25–37. – ISSN 0424-7388. (RSCI)

Публикации в рецензируемых научных изданиях,
определенных ВАК при Минобрнауки России (физико-математические науки):

19. 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.

20. 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

21. 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.

22. 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.

Публикации в других научных изданиях:

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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).

31. 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).

32. Кораблев, Ю.А. Емкостный метод анализа и прогнозирования редких событий в экономике : монография / Ю.А. Кораблев. – Москва : РУСАЙНС, 2023. – 296 с. – 100 экз. – ISBN 978-5-466-04159-0.