

Правительство Российской Федерации  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»  
(НИУ ВШЭ)

УДК 33:550.311.6  
Рег. № НИОКТР 121051200272-1  
Рег. № ИКРБС

УТВЕРЖДАЮ  
Проректор НИУ ВШЭ,  
канд. экон. наук  
\_\_\_\_\_  
М.М. Юдкевич  
«\_\_\_» \_\_\_\_\_ 2021 г.

ОТЧЕТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

АНАЛИЗ ДАННЫХ И ВЫБОР РЕШЕНИЙ В ЗАДАЧАХ ИССЛЕДОВАНИЯ  
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ И ПОЛИТИЧЕСКИХ СИСТЕМ  
(заключительный)

Руководители НИР:

Директор, д-р техн. наук, проф. \_\_\_\_\_ Ф.Т. Алескеров

Гл. науч. сотр., PhD \_\_\_\_\_ Э.С. Маскин

Москва 2021

## СПИСОК ИСПОЛНИТЕЛЕЙ

Руководители НИР:

Директор, д-р техн. наук, проф.

Ф.Т. Алескеров  
(введение, заключение)

подпись, дата

Гл. науч. сотр., PhD

Э. Маскин  
(подраздел 2.1)

подпись, дата

Исполнители:

Зам. директора центра,  
д-р физ.- мат. наук, проф.

А.Е. Лепский  
(подраздел 3.4)

подпись, дата

Вед. науч. сотр.,  
д-р техн. наук, проф.

А.С. Беленький (под-  
раздел 2.1)

подпись, дата

Вед. науч. сотр.,  
д-р физ.-мат. наук, проф.

А.А. Лазарев  
(подраздел 2.1)

подпись, дата

Вед. науч. сотр., д-р техн. наук,  
проф.

Б.Г. Миркин  
(подраздел 3.1)

подпись, дата

Вед. науч. сотр., д-р техн. наук

В.В. Подиновский  
(подразделы 1.1, 1.2)

подпись, дата

Ст. науч. сотр., канд. экон. наук

Д.С. Карабекян  
(подраздел 1.5)

подпись, дата

Ст. науч. сотр., канд. экон. наук

А.В. Карпов  
(подраздел 1.3)

подпись, дата

Ст. науч. сотр., канд. наук по  
прикл. мат.

Т.А. Ломоносов  
(полраздел 6.2)

подпись, дата

Ст. науч. сотр., канд. физ.-мат.  
наук

В.С. Молостков  
(подраздел 2.2)

подпись, дата

Ст. науч. сотр., канд. техн. наук,  
доц.

А.Л. Мячин  
(подраздел 3.2)

подпись, дата

Ст. науч. сотр., канд. экон. наук,  
доц.

Г.И. Пеникас  
(подраздел 4.6)

подпись, дата

Ст. науч. сотр., канд. физ.-мат.  
наук

А.Н. Субочев  
(подраздел 1.4)

Ст. науч. сотр., канд. наук по прикл. мат.	_____	С.В. Швыдун (раздел 5)
Ст. науч. сотруд., канд. техн. наук	_____	В.И. Якуба (подраздел 1.5)
Науч. сотр., PhD	_____	Э. Доган (реферат) подпись, дата
Науч. сотр., канд. наук по прикл. мат.	_____	М.А. Марон (подраздел 2.3)
Науч. сотр.	_____	Н.Г. Мещерякова (раздел 5)
Мл. науч. сотр., канд. физ.-мат. наук	_____	Ю.А. Веселова (подраздел 1.5)
Мл. науч. сотр.	_____	С.С. Демин (подраздел 4.4)
Мл. науч. сотр., канд. физ.-мат. наук	_____	Л.Г. Егорова (подраздел 4.5)
Мл. науч. сотр.	_____	Ю.В. Зонтов (подраздел 1.5)
Мл. науч. сотр., канд. комп. наук	_____	Д.С. Фролов (подраздел 2.1)
Вед. эксперт, канд. техн. наук	_____	В.А. Саркисянц (аннотации)
Эксперт, канд. техн. наук	_____	А.А. Рубчинский (подраздел 3.3)
Стажёр-исследователь	_____	Г.Т. Броницкий (список исп. источников)
Стажёр-исследователь	_____	И.В. Дунилов (подраздел 6.1)

Стажёр-исследователь	_____	М.Д. Ермолова (подраздел 4.6)
Стажёр-исследователь	_____	М.М. Земцова (подраздел 4.3)
Стажёр-исследователь	_____	А.А. Иванов (подраздел 1.5)
Стажёр-исследователь	_____	А.И. Казачинская (раздел 5)
Стажёр-исследователь	_____	Л.Г. Никорашвили (подраздел 4.1)
Стажёр-исследователь	_____	В.М. Семчукова (подраздел 4.1)
Стажёр-исследователь	_____	Ю.Б. Скрупская (подраздел 4.2)
Стажёр-исследователь	_____	М.М. Снегирева (подраздел 6.1)
Стажёр-исследователь	_____	М.А. Сурков (подраздел 4.6)
Стажёр-исследователь	_____	Д.С. Ткачев (подраздел 1.6)
Стажёр-исследователь	_____	Д.А. Шадрин (подраздел 4.2)
Стажёр-исследователь	_____	А.В. Юдина (подраздел 1.4)
Нормоконтроль	_____	М.М. Земцова

## РЕФЕРАТ

Отчет 323 с., 1 кн., 55 рис., 41 табл., 225 источн., 1 прил.

ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ, ТЕОРИЯ КОЛЛЕКТИВНОГО ВЫБОРА,  
КЛАСТЕРНЫЙ АНАЛИЗ, СЕТЕВОЙ АНАЛИЗ, МЕТОДЫ  
ОПТИМИЗАЦИИ, ТЕОРИЯ ФУНКЦИЙ ДОВЕРИЯ

Объектами исследования были задачи оптимизационного моделирования, задачи теории коллективного выбора, оболочечный анализ данных, банковское регулирование, задачи теории предпочтений, сетевой и кластерный анализ данных, текстовые коллекции научных статей, задачи теории много-критериального выбора, задачи теории функций доверия.

Целью проекта была разработка и исследование новых методов анализа данных и принятия решений с применением разрабатываемых моделей, методов и алгоритмов к задачам анализа социально-экономических и политических систем.

Методологической базой проекта были: современные методы теории принятия решений, теории многокритериального и коллективного выбора, теория графов и сетевой анализ данных, методы интеллектуальной обработки и анализа данных, теории функций доверия, теория игр и экономических механизмов, методы оптимизации и др.

Область применения полученных результатов: управление наукой, управление образованием, антиковидное регулирование, региональные транспортные системы, банковское регулирование, теория коллективного выбора, сетевой анализ данных, многокритериальное принятие решений и др.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	12
1 Исследования по анализу и выбору решений .....	20
1.1 Исследование ПН-средних величин.....	20
1.1.1 Сведения о ПН-средних величинах.....	20
1.1.2 Аналитический метод нахождения средних по Р <sub>ЕΔ</sub> .....	23
1.1.3 Об устойчивости ПН-средних .....	25
1.1.4 Случай данных с повторениями .....	26
1.1.5 Случай порядковых данных.....	29
1.1.6 Некоторые выводы и возможные обобщения.....	31
1.2 Исследование задач регрессионного анализа методами многокритериального принятия решений .....	32
1.2.1 Постановка задачи регрессионного анализа как многокритериальной.....	32
1.2.2 Регрессия при различных видах информации Г .....	34
1.2.3 Некоторые выводы.....	38
1.3 Исследование структурированных дихотомических предпочтений ....	38
1.3.1 Введение.....	38
1.3.2 Основные результаты .....	40
1.3.3 Некоторые выводы.....	46
1.4 Исследование турнирных способов выбора оптимальных вариантов ..	47
1.4.1 Введение.....	47
1.4.2 Общие теоретические сведения.....	50
1.4.3 Свойства существенного множества и его связь с другими решениями .....	60
1.4.4 Эмпирическое сравнение методов ранжирования.....	62
1.4.5 Некоторые выводы.....	65

1.5 Исследование влияния неполноты информации и неопределенности на манипулируемость в теории коллективного выбора.....	66
1.5.1 Введение.....	67
1.5.2 Неполнота информации и одновременное манипулирование .....	67
1.5.3 Разработка алгоритмов для точной оценки сильной и слабой манипулируемости процедур агрегирования.....	77
1.5.4 Некоторые выводы.....	82
1.6 Аксиоматизация функций интенсивности связи в индексах влияния...	82
1.6.1 Введение.....	82
1.6.2 Основная часть .....	83
1.6.3 Некоторые выводы.....	88
<b>2 Исследование оптимизационных моделей .....</b>	<b>89</b>
2.1 Исследование моделей оптимального административного управления89	
2.1.1 Введение.....	89
2.1.3 Математическая модель задачи организации переработки твердых бытовых отходов в крупных городах.....	106
2.1.4 О математической модели задачи оптимизации использования альтернативных источников энергии в системах электроснабжения крупных городов .....	116
2.1.5 Краткие выводы.....	117
2.2 Исследование оптимизационной задачи минимизации гарантированного риска .....	118
2.2.1 Введение.....	118
2.2.2 Постановка проблемы.....	119
2.2.3 Принцип минимаксного сожаления Сэвиджа .....	121
2.2.4 Построение гарантированного по риску решения.....	123
2.2.5 Алгоритм построения гарантированного по риску решения .....	133
2.2.6 Некоторые выводы.....	134

2.3 Модификация метода расчёта приоритетов стратегических целей.....	135
2.3.1 Введение.....	135
2.3.3 Некоторые выводы.....	142
3 Исследования по анализу текстов и кластерному анализу.....	143
3.1 Разработка программного обеспечения для таксономического анализа текстовых коллекций .....	143
3.1.2 Введение.....	143
3.1.2 Таксономический контент-анализ.....	145
3.1.3 Оптимальный подъем нечеткого тематического кластера в таксономии: модель и метод .....	147
3.1.4 Описание и структура программного пакета GOT .....	154
3.1.5 Некоторые выводы.....	160
3.2 Диффузионно-инвариантная паттерн-кластеризация на основе сортировки данных .....	161
3.2.1 Введение.....	161
3.2.2 Анализ паттернов: общая постановка задачи .....	162
3.2.3 Возможность использования алгоритмов сортировки данных .....	165
3.2.4 Альтернативная реализация диффузионно-инвариантной паттерн-кластеризации .....	166
3.2.5 Корректировка результатов диффузионно-инвариантной паттерн-кластеризации .....	167
3.2.6 Некоторые выводы.....	169
3.3 Декомпозиционная сложность графов и её применение для анализа фондовых рынков.....	169
3.3.1 Введение.....	169
3.3.3 БАД и показатели сложности графа.....	176
3.3.5 Вычисление глубины графов для фондового рынка .....	186
3.3.6 Некоторые выводы.....	192

3.4 Разработка и исследование метода кластеризации тел свидетельств .	193
3.4.1 Введение.....	193
3.4.2 Необходимые сведения из теории функций доверия.....	194
3.3.4 Численный пример.....	202
4 Исследования по анализу данных .....	208
4.1 Анализ влияния мирового изменения климата на экономики арктических стран .....	208
4.1.1 Введение.....	208
4.1.2 Основная часть .....	209
4.1.3 Некоторые выводы.....	211
4.2 Анализ среднесрочных стратегий развития Арктического региона....	212
4.2.1 Введение.....	212
4.2.2 Основная часть .....	213
4.2.3 Некоторые выводы.....	219
4.3 Исследование причин вододефицита в странах региона Северной Африки .....	220
4.3.1 Введение.....	220
4.3.2 Оценка водообеспеченности.....	221
4.3.3 Торговля виртуальной водой .....	224
4.3.4 Модель линейной регрессии .....	225
4.3.5 Результаты регрессионного анализа .....	226
4.3.6 Использование охлажденной воды .....	228
4.3.7 Некоторые выводы.....	229
4.4 Интервальный метод оболочечного анализа данных и его применение для анализа эффективности карантинных мер.....	230
4.4.1 Введение.....	230
4.4.2 Основная часть .....	231

4.4.3 Некоторые выводы.....	235
4.5 Анализ предпочтений студентов .....	235
4.5.1 Введение.....	235
4.5.2 Основная часть .....	237
4.5.3 Краткие выводы.....	240
4.5.4 Заключение .....	242
4.6 Анализ политик регулирования банковской системы .....	242
4.6.1 Введение.....	242
4.6.2 Основная часть .....	244
4.6.3 Некоторые выводы.....	251
5 Исследования по сетевому анализу данных.....	252
5.1 Введение.....	252
5.2 Сетевой анализ сети международного рынка заимствований.....	254
5.4 Сетевой анализ двусторонней торговли в условиях асимметрии данных .....	269
5.5 Некоторые выводы.....	275
6 Исследование моделей распространения .....	276
6.1 Модели распространения COVID-19 .....	276
6.1.1 Анализ и разработка моделей для прогнозирования распространения COVID-19 по странам мира .....	276
6.1.2 Оценка эффективности карантинных мер на примере Великобритании .....	279
6.2.1 Введение.....	283
6.2.2 Модель апоптоза .....	283
6.2.3 Разностная схема для численного решения модели апоптоза.....	287
6.2.4 Реализация разностной схемы .....	288
6.2.5 Линеаризация разностных схем с использованием теоремы о гомоморфизме.....	290

6.2.6 Некоторые выводы.....	291
ЗАКЛЮЧЕНИЕ .....	292
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	296
ПРИЛОЖЕНИЕ А Таблицы к подразделу 4.5 .....	320

## **ВВЕДЕНИЕ**

В отчете приведены результаты работы в 2021 году сотрудников Международного центра анализа и выбора решений (МЦАВР) НИУ ВШЭ над проектом «Анализ данных и выбор решений в задачах исследования социально-экономических и политических систем».

Основными научными направлениями исследований были:

- модели анализа и выбора решений;
- оптимационные модели;
- анализ текстов и кластерный анализ данных;
- модели сетевого анализа данных;
- анализ данных и принятие решений в задачах социально-экономического развития регионов;
- модели распространения (в том числе COVID-19).

В настоящем проекте были продолжены исследования 2020 г. по выполнению НИР «Принятия решений и анализ данных в социально-экономических и политических системах» [1]. В частности, в 2020 г. были получены следующие результаты:

- разработаны и исследованы новые средние значения выборок в рамках теории важности критериев, основанные на идее многокритериальной оптимизации;
- в рамках исследования по оцениванию степени манипулируемости правил агрегирования предпочтений произведены расчеты, показывающие, какая процедура агрегирования будет наименее манипулируемой для различных параметров (моделей голосования, расширения предпочтений, числа альтернатив, числа участников); разработано соответствующее программное обеспечения и найдены оценки вычислительной сложности алгоритмов; найдены условия, когда манипулирование всегда безопасно и когда существует небезопасно манипулируемый профиль для двухступенчатой мажоритарной системы и для правила Коупленда;

- в рамках исследования профилей предпочтений найдены вероятности однопиковых профилей предпочтений для разных моделей генерирования таких профилей;
- найдены достаточные условия непустоты множества обобщенных турнирных решений, основанных на принципе устойчивости; показано, что эти же условия будут справедливы для некоторых турнирных решений и в случае бесконечного множества вариантов;
- применены методы анализа латентных классов и дисперсионного анализа к обработке мониторинговых данных, выделены и проанализированы различные группы студентов по их отношению к переводу занятий в онлайн формат;
- построены обобщения разработанных ранее индексов дальних взаимодействий (LRIC), учитывающие взаимозависимость между элементами сети; эти индексы применены к анализу сетей международной торговли и взаимодействия террористических групп;
- разработаны индексы центральности, учитывающие индивидуальные и групповые влияния вершин сети; эти индексы применены к анализу сетей международной миграции и межстрановой торговли;
- разработана модель по оценке схожести слоёв многослойной динамической сетевой структуры; эта модель применена к анализу сети международного экспорта/импорта продовольствия;
- рассмотрена модификация гравитационной модели, включающей в себя индексы центральности сети; эта модель применена к анализу многосторонних торговых потоков;
- разработан и исследован способ оценивания центральностей многослойной сети методами теории функций доверия; доказаны взаимосвязи полученных оценок с агрегированием классических послойных центральностей в некоторых частных случаях;
- разработан гибридный метод экономного подъема нечеткого тематического кластера в таксономии предметной области; этот метод

применен к выявлению исследовательских тенденций по анализу статей, опубликованных в журналах издательств Springer и Elsevier по тематике Наука о Данных;

- разработан и исследован метод альтернирующих дихотомий для кластеризации данных; этот метод применен для анализа фондового рынка;
- проведено исследование государственной состоятельности в странах мира в период 1996-2015гг с помощью двух различных методов (иерархической кластеризации и порядково-инвариантной паттерн-кластеризации);
- при решении задачи оценивания экономической целесообразности со-здания нового грузового транспортного узла сформулировано и доказа-но достаточное условие конкурентоспособности тарифов на услуги в та-ком узле;
- продемонстрирована возможность решения задачи об оценке финансо-вой целесообразности строительства и эксплуатации нового транспорт-ного узла, как задачи нахождения максимума максимины разности двух билинейных функций;
- предложена модификация ранее разработанной модели оптимального размещения зарядных станций автомобилей вдоль участка шоссе для решения аналогичной задачи размещения городских сетей таких стан-ций;
- проведен анализ изменения населения в Арктике из-за естественного прироста и чистой миграции населения;
- проанализированы существующие методы распределения спорных территорий и предложен ряд новых модификаций этих методов; эти ме-тоды апробированы на примере раздела спорных территорий стран Арк-тического региона;
- разработаны интервальные модификации оболочечного анализа дан-ных; эти методы апробированы на решении задачи оценки эффективно-сти превентивных противопожарных мероприятий в регионах РФ;

- проанализирована и модифицирована популярная модель образования теневой банковской системы; доказано, что политика регулятора не может основываться на поддержке банковского теневого сектора;
- разработана и исследована модификация модели специализации в многоклеточных системах в случае структурных ограничений, задаваемых неравенствами; произведено сравнение решений задачи нахождения стратегии, максимизирующей эффективность функционирования системы, для разных случаев ограничений;
- предложен матричный метод линеаризации систем уравнений в частных производных на постоянных решениях, позволяющий автоматически получить линеаризованную систему, подавая на вход лишь исходную нелинеаризованную систему.

В этом году некоторые исследования 2020 года были продолжены. В частности, это исследования новых средние величин в рамках теории важности критериев, исследования по оцениванию степени манипулируемости правил агрегирования предпочтений, исследования по паттерн-кластеризации, по интервальному оболочечному анализу данных и др.

В то же время ряд исследований этого года предполагается продолжить в следующем году.

Методологической базой исследования данного проекта были современные методы анализа данных и теории принятия решений, теории колективного выбора, методы оптимизации, теории вероятностей, теории функций доверия, методы исследования операций и др.

Ряд исследований имеет не только фундаментальный, но и прикладной характер. При решении некоторых прикладных задач было использовано прикладное программное обеспечения и разработаны комплексы программ.

Отчет состоит из 6 разделов.

В первом разделе представлены результаты исследований по анализу и выбору решений.

В подразделе 1.1 приведены результаты продолженного с 2020 года исследования ПН-средних величин в рамках теории важности критериев, основанных на идеях многокритериальной оптимизации. В частности, исследованы вопросы устойчивости таких средних, рассмотрены некоторые частные случаи и найдены простые правила вычисления средних величин.

В подразделе 1.2 приведены результаты исследования задач регрессионного анализа методами многокритериального принятия решений.

Подраздел 1.3 посвящен исследованию структурированных дихотомических предпочтений. В частности, получены некоторые характеристики доменов через запрещенные подпрофили.

В подразделе 1.4 приведены результаты продолжающегося в отчетном году исследования турнирных способов выбора оптимальных вариантов. Исследование этого года посвящено содержательному описанию и нахождению так называемого существенного множества, как обобщения двухпартийного множества.

Подраздел 1.5 отчета посвящен новым результатам, полученным в ходе продолжающегося исследования влияния неполноты информации и неопределенности на манипулируемость в теории коллективного выбора. В частности, в этом отчете приведены результаты по разработке и исследованию модели, учитывающей одновременно и неполноту информации, и неопределенность относительно действий других избирателей на манипулируемость.

В последнем подразделе 1.6 приведены результаты предварительного исследования по аксиоматизации функций интенсивности связи в индексах влияния.

Во втором разделе отчета приведены результаты исследования оптимизационных моделей.

Так, подраздел 2.1 посвящен исследованию моделей оптимального административного управления. В частности, предложена и исследована теоретико-игровая модель задачи организации переработки твердых бытовых отходов города.

В подразделе 2.2 приведены результаты исследования задачи оптимальной диверсификации мультивалютных депозитов с неопределенными будущими обменными курсами как проблема минимизации упущеной из-за неопределенности выгоды.

В подразделе 2.3 предложена модификация метода расчёта приоритетов стратегических целей по таблице парных сравнений их значимости.

В третьем разделе отчета приведены результаты исследований по анализу текстов и кластерному анализу данных.

Так в подразделе 3.1 дано описание нового программного комплекса, который был разработан участниками проекта, для таксономического контент-анализа коллекций текстовых документов, реализующий методологию обобщения в таксономии предметной области.

В подразделе 3.2 представлены результаты по разработке вычислильно экономичной модификации алгоритма диффузионно-инвариантной паттерн-кластеризации на основе сортировки данных.

В подразделе 3.3 предложена новая модификация дихотомии графов на основе разработанного ранее частотного алгоритма декомпозиции. С помощью модифицированного алгоритма проанализировано функционирования рынка S&P-500 в период 1990–2019 гг.

Наконец в последнем подразделе 3.4 третьего раздела приведены результаты по разработке и исследованию двухэтапного алгоритма кластеризации тел свидетельств в рамках теории функций доверия.

В четвертом разделе приведены результаты исследований, которые можно отнести к тематике «анализ данных». Так подраздел 4.1 посвящен анализу влияния мирового изменения климата на экономики арктических стран.

В подразделе 4.2 приведен анализ среднесрочных стратегий развития Арктического региона.

Подраздел 4.3 посвящен исследованию причин вододефицита в странах региона Северной Африки.

В подразделе 4.4 рассмотрена модификация интервального метода оболочечного анализа данных и с её помощью оценена эффективность карантинных мер в разных странах мира.

Подраздел 4.5 посвящен анализу предпочтений студентов относительно формата онлайн-обучения.

Последний подраздел 4.6 четвертого раздела посвящен исследованию задач банковского регулирования. В частности, проанализировано влияние перехода на основанный на внутренних рейтингах (ПВР) подход на котировки банковских акций. Предложена параметризация регулирования достаточности капитала исламских банков как функция от макроэкономических параметров.

Пятый раздел отчета посвящен результатам сетевого исследования данных. Раздел состоит из введения (подраздел 5.1), в котором дан общий обзор результатов, полученных в рамках работы над проектом по сетевому анализу данных, и ряда подразделов, посвященных отдельным исследованиям.

Так в подразделе 5.2 приведены результаты сетевого исследования международного рынка заимствований за 2005–2020 гг. с точки зрения анализа структуры сети, выделения наиболее важных элементов и сообществ, анализа устойчивости сети её динамического изменения.

В подразделе 5.3 приведены результаты исследования сети цитирования 466 российских экономических журналов РИНЦ за период 2016–2020 гг. с использованием индекса центральности Коупленда, индекс группового влияния BI и индекс влияния по ключевым вершинам.

В подразделе 5.4 рассматривается проблема построения сетевой структуры международной торговли продовольствием с учетом неполноты (неточности, неопределенности) и противоречивости данных.

Последний, шестой раздел отчета посвящен исследованию моделей распространения (в широком смысле). В подразделе 6.1 приведены некоторые результаты по анализу и моделированию распространения COVID-19.

В частности, 19 выполнен анализ эффективности антковидных мер в различных странах мира, выявлены группы стран со схожими показателями интенсивности введенных мер, законопослушности, уровня образования, вакцинационной политики. Проведен анализ изменения экономических показателей Великобритании с изменением интенсивности заболеваемостью COVID-19 и с принимаемыми антковидными мерами

А в подразделе 6.2 приведены результаты по программной имплементации разностной схемы высокого порядка точности для решения модели клеточного апоптоза.

Представленные в настоящем отчете результаты были апробированы на семинарах и конференциях (в этом году онлайн) различного уровня. Многие из представленных материалов уже опубликованы, направлены в печать или подготовлены к публикации.

## **1 Исследования по анализу и выбору решений**

В этом разделе отчета приведены результаты исследований по анализу и выбору решений. Работа проводилась в следующих направлениях:

- исследование ПН-средних величин;
- исследование задач регрессионного анализа методами многокритериального принятия решений;
- исследование структурированных дихотомических предпочтений;
- исследование турнирных способов выбора оптимальных вариантов;
- исследование влияния неполноты информации и неопределенности на манипулируемость в теории коллективного выбора;
- аксиоматизация функций интенсивности связи в индексах влияния.

### **1.1 Исследование ПН-средних величин**

В этом подразделе отчета представлены результаты изучения средних величин, продолживших исследования 2020 года и опубликованные в [2]. Работы выполнены в рамках многокритериального подхода с использованием теории важности критериев [3].

#### **1.1.1 Сведения о ПН-средних величинах**

Пусть имеется совокупность  $X$ , состоящая из  $n \geq 2$  действительных чисел, называемых далее данными, или точками, и являющихся результатами измерения интенсивности некоторого выделенного признака:

$$X = \{x_1, x_2, \dots, x_n\}. \quad (1.1)$$

Эти данные являются однородными в том смысле, что измерения производились по одной и той же шкале, которая не менее совершенна, чем шкала интервалов. Упорядоченные соответственно по неубыванию и невозрастанию множества

$$X_{\uparrow} = \langle x_{(1)}, x_{(2)}, \dots, x_{(n)} \rangle; X_{\downarrow} = \langle x_{[1]}, x_{[2]}, \dots, x_{[n]} \rangle,$$

где  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  и  $x_{[1]} \geq x_{[2]} \geq \dots \geq x_{[n]}$ , получаются из совокупности чисел (1.1) при помощи соответствующих перестановок.

Пусть  $x$  – произвольное фиксированное число – точка на числовой прямой  $\text{Re}$ . Удаленность ее от отдельной точки  $x_i$  из множества  $X$  можно оценить расстоянием  $y_i = |x - x_i|$ . Тогда удаленность точки  $x$  от совокупности всех точек из множества  $X$  характеризуется вектором  $y = (y_1, y_2, \dots, y_n)$ , составленным из таких расстояний. Его можно считать значением векторного критерия  $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$ , где  $f_i(x) = |x - x_i|$ . Областью значений  $Z$  этого векторного критерия является положительный ортант  $\text{Re}_+^n = [0, +\infty)^n$  – множество  $n$ -мерных векторов с неотрицательными компонентами. Значение  $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$  векторного критерия  $f$ , называемое векторной оценкой точки  $x$ , для краткости будем обозначать также  $y = (y_1, y_2, \dots, y_n)$ , где  $y_i = f_i(x)$ ,  $i \in N = \{1, 2, \dots, n\}$ .

Пусть на множестве  $Z$  задано отношение предпочтения – строгий частичный порядок  $P^\Gamma$ , где  $\Gamma$  – информация о предпочтениях ЛПР, касающаяся удаленности: если верно  $y' P^\Gamma y''$ , то точка  $y' = f(x')$  ближе ко множеству значений векторного критерия  $Y = \{y \in Z \mid y = f(x), x \in X\}$ , чем  $y'' = f(x'')$ . Отношение  $P^\Gamma$  порождает имеющий аналогичный смысл отношение  $P_\Gamma$  на числовой прямой:  $x' P_\Gamma x'' \Leftrightarrow y' P^\Gamma y''$ , где  $y' = f(x')$ ,  $y'' = f(x'')$ . На роль наиболее близких к  $X$  и представляющих все множество  $X$  могут претендовать те и только те точки, которые недоминируемые по  $P_\Gamma$ . (Точка  $x$  недоминируема по  $P_\Gamma$ , если не существует точки  $x'$  такой, что верно  $x' P_\Gamma x$ .) Если множество таких точек  $G^\Gamma(X)$  внешне устойчиво (т. е. для каждой доминируемой точки  $x$  найдется недоминируемая точка  $x'$  такая, что верно  $x' P_\Gamma x$ ), то все они именуются ПН-средними, а более конкретно (для рассматриваемой информации  $\Gamma$ ) и кратко – средними по  $P_\Gamma$ .

Естественно полагать, что предпочтения с увеличением значений критериев  $f_i$  убывают или, иными словами, что критерии желательно минимизировать. При отсутствии иной информации о предпочтениях на множестве  $Z$  предпочтения описывает отношение Парето  $P^\emptyset$ , определяемое так:  $y P^\emptyset z \Leftrightarrow (y_i \leq z_i, i = 1, 2, \dots, n$ , причем хотя бы одно из неравенств является строгим).

Отношение  $P^\emptyset$  порождает на числовой прямой  $\mathbf{Re}$  отношение Парето  $P_\emptyset$ :  $xP_\emptyset x' \Leftrightarrow yP^\emptyset y'$ . Оказывается, что средними по  $P_\emptyset$  являются все точки отрезка с концами  $x_{(1)} = \min_{i \in N} x_i$  и  $x_{(n)} = \max_{i \in N} x_i$ , т. е.  $G^\emptyset(X) = \bar{X} = [x_{(1)}, x_{(n)}]$ . Таким образом, понятие средних по  $P_\emptyset$  оказывается эквивалентным понятию средних по Коши.

Пусть все критерии имеют равную важность (информация  $E$ ) [3]. В этом случае удаленность точки  $x$  от множества  $X$  оценивается отношением  $P_E$  на числовой прямой  $\mathbf{Re}$ . Оно порождается отношением  $P^E$  на множестве  $\mathbf{Re}^n$ , которое определяется каждым из двух равносильных решающих правил:

$$\begin{aligned} yP^E z &\Leftrightarrow y_{(1)} \leq z_{(1)}, y_{(2)} \leq z_{(2)}, \dots, y_{(n)} \leq z_{(n)} \\ yP^E z &\Leftrightarrow [y_{[1]} \leq z_{[1]}, y_{[2]} \leq z_{[2]}, \dots, y_{[n]} \leq z_{[n]}], \end{aligned} \quad (1.2)$$

причем хотя бы одно из нестрогих неравенств является строгим.

ПН-средними (по  $P_E$ ), составляющими множество  $G^E(X)$ , здесь являются недоминируемые по  $P_E$  точки числовой прямой.

Пусть теперь увеличение удаленности точки  $x$  от одних точек  $x_i$  не компенсируется уменьшением ее удаленности от других. Это означает, что если в произвольной векторной оценке  $y$ , в которой  $y_i > y_j$ , заменить  $y_i$  на  $y_i - \delta$ , а  $y_j$  – на  $y_j + \delta$ , где  $\delta$  – положительное число такое, что  $y_i - \delta \geq y_j + \delta$ , то полученная таким образом векторная оценка  $z$  будет предпочтительнее, чем исходная  $y$ . Отношение строгого предпочтения  $P^{E\Delta}$ , порождаемое такой информацией на  $Z$  (которая соответствует случаю, когда шкалой критериев является шкала первой порядковой метрики), задается так (сравни с (1.2)):

$$yP^{E\Delta} z \Leftrightarrow (y_{[1]} \leq z_{[1]}, y_{[1]} + y_{[2]} \leq z_{[1]} + z_{[2]}, \dots, y_{[1]} + y_{[2]} + \dots + y_{[n]} \leq z_{[1]} + z_{[2]} + \dots + z_{[n]}),$$

причем хотя бы одно из неравенств является строгим.

Здесь ПН-средними являются точки, недоминируемые по  $P_{E\Delta}$ . Так как  $P_{E\Delta} \supset P_E \supset P_\emptyset$ , то  $G^{E\Delta}(X) \subseteq G^E(X) \subseteq G^\emptyset(X) = \bar{X}$ . Структура множества  $G^E(X)$  может быть весьма сложной – оно может состоять из нескольких числовых

промежутков и быть незамкнутым и не открытым. Например, для  $X = \{1, 2, 5, 9, 11\}$  имеем  $G^E(X) = [1,5; 7,5) \cup (8,5; 9,5)$ . А вот структура множества  $G^{E\Delta}(X)$  оказывается совсем простой – это отрезок  $[\alpha, \beta]$ , где  $x_{(1)} \leq \alpha \leq \beta \leq x_{(n)}$ . Так, если  $X = \{1, 2, 5, 9, 11\}$ , то  $G^{E\Delta}(X) = [5; 6]$ . В вырожденном случае, когда  $\alpha = \beta$ , отрезок  $[\alpha, \beta]$  стягивается в одну точку. Для расчета величин  $\alpha$  и  $\beta$  были разработаны специальные решающие правила. Существенно более простое такое правило предлагается в следующем пункте.

### 1.1.2 Аналитический метод нахождения средних по $P_{E\Delta}$

Нам далее понадобится множество  $H = \{1, 2, \dots, h\}$ , где  $h = \lfloor (n+1)/2 \rfloor$  – целая часть  $(n+1)/2$ . Следующее утверждение позволяет достаточно просто найти среднюю величину  $G^{E\Delta}(X) = [\alpha, \beta]$ .

Теорема 1.1. Справедливы формулы:

$$\alpha = \frac{1}{2} \min_{p \in H} (x_{(p)} + x_{(n+1-p)}), \beta = \frac{1}{2} \max_{p \in H} (x_{(p)} + x_{(n+1-p)}). \quad (1.3)$$

Формула (1.3) дает простой и точный аналитический метод нахождения средних по  $P_{E\Delta}$ .

Пример 1.1. При  $n = 5$  имеем  $h = \lfloor (n+1)/2 \rfloor = 3$ , так что  $H = \{1, 2, 3\}$ . Для  $X = \{1, 2, 7, 8, 11\}$ , пользуясь теоремой 1, последовательно вычисляем:

$$\begin{aligned} \alpha &= \frac{1}{2} \min \{x_{(1)} + x_{(5)}, x_{(2)} + x_{(4)}, x_{(3)} + x_{(3)}\} = \frac{1}{2} \min \{1 + 11, 2 + 8, 7 + 7\} = \\ &= \frac{1}{2} \min \{12, 10, 14\} = 5; \end{aligned}$$

$$\beta = \frac{1}{2} \max \{x_{(1)} + x_{(5)}, x_{(2)} + x_{(4)}, x_{(3)} + x_{(3)}\} = \frac{1}{2} \max \{12, 10, 14\} = 7;$$

Таким образом,  $G^{E\Delta}(X) = [\alpha, \beta] = [5, 7]$ .

Для упрощенного применения средней  $G^{E\Delta}$  отрезок  $[\alpha, \beta]$  можно представлять одной точкой – его серединой  $\gamma = \frac{1}{2}(\alpha + \beta)$ , которую можно назвать сконденсированной средней (по  $P_{E\Delta}$ ).

Пример 1.2. ВВП Европы на душу населения (Real GDP per capita) представлены в Таблице 1.1 [4].

Таблица 1.1 – ВВП Европы на душу населения

The screenshot shows a data visualization interface with a table titled "Real GDP per capita (online data code: SDG\_08\_10)". The table includes columns for "TIME" (2007 to 2020) and "GEO" (countries). The data is sorted by GDP per capita in descending order. The table shows significant differences between countries, with some like Luxembourg having very high values (e.g., over 100k in 2020) and others like Moldova having much lower values (e.g., around 1k in 2020). The interface includes various filters and settings at the top.

GEO	TIME	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
European Union - 27 countries (from 2000)		25 580	25 580	24 400	24 598	25 320	25 890	25 868	25 428	25 958	26 410	27 118	27 638	28 656	28 376
European Union - 28 countries (2013-2020)		26 220	26 240	25 630	25 510	25 980	25 770	26 178	26 798	27 140	27 880	28 288	28 698	-	-
Euro area - 19 countries (from 2015)		29 360	29 340	27 930	28 460	28 870	28 550	28 440	28 780	29 288	29 730	30 450	30 930	31 310	29 288
Belgium		33 760	33 640	32 760	33 330	33 450	33 490	33 570	34 360	34 560	34 620	35 650	35 510	35 950	33 560
Bulgaria		4 820	5 140	4 990	5 650	5 390	5 350	5 400	5 530	5 700	6 050	6 310	6 550	6 840	6 600
Czechia		15 250	15 580	14 690	15 620	15 310	15 170	15 100	15 480	16 290	16 670	17 450	17 990	18 460	17 340
Denmark		46 210	45 760	43 220	43 840	44 240	44 170	44 410	44 890	45 630	45 720	47 740	49 450	49 270	49 150
Germany (until 1990 former territory of the FRG)		31 920	32 320	30 580	31 940	33 290	33 330	33 330	33 920	34 130	34 610	35 410	35 690	34 310	34 310
Estonia		13 230	12 590	10 770	11 860	11 890	12 320	12 540	12 960	13 230	13 620	14 410	14 970	15 510	15 610
Ireland		41 240	38 550	36 240	36 700	36 940	36 770	37 650	40 610	40 620	50 660	53 930	53 100	60 130	62 980
Greece		22 590	22 370	21 350	20 150	18 130	16 940	16 680	16 820	16 870	16 850	17 100	17 400	17 750	16 360
Spain		24 380	24 200	23 100	23 040	22 770	22 880	21 840	22 210	23 080	23 700	24 430	24 880	25 280	22 350
France		31 480	31 310	30 250	30 690	31 210	31 160	31 170	31 320	31 540	31 770	32 360	32 820	33 320	30 610
Croatia		11 250	11 460	10 640	10 520	10 530	10 310	10 300	10 310	10 630	11 100	11 660	12 040	12 450	11 500
Italy		28 740	28 250	26 600	26 940	27 030	26 160	25 620	25 920	25 600	26 240	26 730	27 030	27 210	24 890
Cyprus		24 420	24 680	23 550	23 490	22 990	21 780	20 400	20 250	21 050	22 310	23 460	24 430	25 370	23 770
Latvia		10 280	10 650	8 770	8 550	8 540	9 680	9 980	10 260	10 750	11 110	11 590	12 140	12 530	12 150
Lithuania		9 770	10 130	8 720	9 650	9 620	10 330	10 810	11 290	11 620	12 070	12 760	13 400	14 050	14 030
Luxembourg		88 120	86 330	82 920	83 650	82 490	81 540	80 400	82 590	82 820	84 750	84 020	84 040	85 030	82 250
Hungary		10 410	10 530	9 850	9 980	10 200	10 120	10 330	10 800	11 220	11 500	12 030	12 690	13 270	12 680
Malta		15 470	15 960	15 660	16 440	16 450	16 970	17 650	18 610	19 920	20 210	21 810	22 350	22 720	20 380
Netherlands		39 120	39 810	38 160	38 470	38 880	38 340	38 180	38 580	39 170	39 810	40 730	41 450	41 990	40 160
Austria		35 870	36 280	34 830	35 390	36 380	36 390	36 180	36 130	36 140	36 980	37 720	38 110	35 590	-
Poland		8 550	8 910	9 670	9 400	9 650	9 980	10 100	10 440	10 890	11 240	11 790	12 420	12 790	12 790
Portugal		17 230	17 260	16 710	16 990	16 720	16 110	16 650	16 260	16 620	17 810	17 650	18 190	18 670	17 670
Romania		6 650	6 730	6 410	6 200	6 550	6 500	6 770	7 640	7 290	7 670	8 280	8 700	9 120	8 810
Slovenia		18 570	19 190	17 570	17 750	17 670	17 360	17 160	17 620	17 990	18 530	19 440	20 240	20 720	19 720
Slovakia		10 650	12 610	11 160	12 610	13 420	13 180	13 600	14 450	14 600	14 600	15 100	15 100	15 100	-
Finnland		37 210	37 530	34 150	35 980	35 140	34 610	34 540	34 440	34 530	34 380	34 740	37 200	36 070	-
Sweden		48 590	49 180	38 830	39 650	40 350	40 280	40 510	41 390	42 380	42 920	43 760	44 180	42 440	-
Iceland		34 360	34 230	33 450	32 650	33 150	33 230	34 510	34 710	35 970	37 410	39 280	39 990	39 160	34 690
Norway		69 140	68 610	66 580	66 220	66 910	66 980	66 780	67 340	67 940	68 990	69 130	69 440	69 560	68 630
Switzerland		57 720	58 420	56 750	58 170	58 730	58 820	59 290	59 290	60 220	60 780	61 170	62 590	62 880	62 880
United Kingdom		31 280	30 340	29 460	29 630	29 060	28 190	28 660	31 290	31 780	32 860	32 420	32 440	32 910	-
Montenegro		4 900	5 230	4 920	5 650	5 280	5 660	5 230	5 230	5 590	5 660	5 920	6 230	-	-
North Macedonia		3 280	3 370	3 350	3 460	3 530	3 510	3 410	3 340	3 370	3 420	3 420	3 420	3 420	-
Serbia		4 130	4 380	4 280	4 330	4 450	4 440	4 590	4 540	4 640	4 820	4 950	5 280	5 460	5 440
Turkey		8 810	7 980	7 400	8 600	8 760	9 070	9 710	10 650	10 520	10 730	11 380	11 560	11 580	11 680

Средние реального ВВП Европы на душу населения (евро) за 2022 по 2019 годы, рассчитанные на основе таблицы 1, представлены в Таблице 1.2.

Таблица 1.2 – Средние реального ВВП Европы на душу населения (евро)

Годы	2012	2013	2014	2015	2016	2017	2018	2019
$m$	25741	25825	26257	27063	27564	28274	28909	29249
$\mu$	21780	20400	20250	21020	22270	23200	24120	24570
$\alpha$	19720	19745	20250	21020	21995	22840	23245	23485
$\beta$	40840	41310	42015	42970	43850	43750	44335	44545
$\gamma$	30280	30528	31133	31995	32923	33295	33790	34015

Таблица 1.2 показывает, что в период с 2012 по 2019 годы средний душевой доход  $m$  возрастает, но медианный доход  $\mu$  вначале падает (до 2014 года), а затем растет. Поэтому однозначный вывод о росте душевого дохода в Европе за этот период сделать затруднительно. Однако средняя  $G^{\Delta}$  (её границы  $\alpha$  и  $\beta$ ) возрастает (есть лишь незначительное падение  $\beta$  в 2017 году), а сконденсированная средняя  $\gamma$  увеличивается постоянно. Это служит основанием для утверждения о росте душевого дохода в странах Европы в 2012 – 2019 годы. Результаты расчетов представлены на рисунке 1.1.

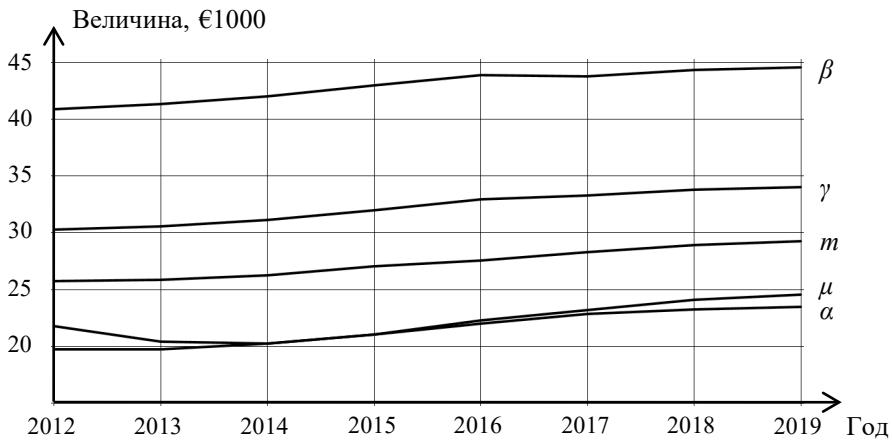


Рисунок 1.1 – Средние реального ВВП Европы на душу населения

### 1.1.3 Об устойчивости ПН-средних

Практически важным и теоретически интересным представляется вопрос о том, насколько сильно могут изменяться средние величины при небольшом изменении исходных данных (1.1).

Поскольку  $G^\emptyset(X) = \bar{X} = [x_{(1)}; x_{(n)}]$ , то малое изменение величин  $x_i$  может привести лишь к малому изменению чисел  $x_{(1)}$  и  $x_{(n)}$ , так что множество средних  $G^\emptyset(X)$  является устойчивым.

А вот средняя по  $P_E$  может быть неустойчивой в том смысле, что сколь угодно малое изменение (сдвиг) даже только одной из точек в множестве  $X$  может привести к «большому» изменению множества  $G^E(X)$ . Следующие примеры иллюстрируют такую возможность.

Пример 1.3. Для  $X = \{1, 2, 3\}$  имеем:  $G^E(X) = [1,5; 2,5]$ . Однако при сколь угодно малом положительном числе  $\varepsilon$  для  $X^\varepsilon = \{1, 2 - \varepsilon, 3\}$  оказывается, что  $G^E(X^\varepsilon) = [1,5 - 0,5\varepsilon; 2]$ . Правая граница отрезка средних по  $P_E$  изменилась сразу на 0,5.

Пример 1.4. Для  $X = \{10, 25, 40, 110\}$  имеем:  $G^E(X) = [25, 60]$ . Однако при сколь угодно малом положительном числе  $\varepsilon$  для  $X^\varepsilon = \{10, 25, 40 + \varepsilon, 110\}$  оказывается, что  $G^E(X^\varepsilon) = [17,5; 60]$ . В этом примере примечательно ещё и то,

что левая граница средних по  $P_E$  уменьшилась на 7,5 при том, что координата одной из точек множества  $X$  увеличилась на  $\varepsilon$ .

Выясним теперь, устойчива ли средняя по  $P_{E\Delta}$ .

Пример 1.5. В условиях примера 1.3 имеем:  $G^{E\Delta}(X) = \{2\}$  и  $G^{E\Delta}(X^\varepsilon) = [2 - \varepsilon, 2]$ . Здесь отклонение одной из точек множества  $X$  на  $\varepsilon$  приводит к отклонению одной из границ средних по  $P_{E\Delta}$  также на  $\varepsilon$ .

Пример 1.6. В условиях примера 1.4 имеем:  $G^{E\Delta}(X) = [32,5; 60]$  и  $G^{E\Delta}(X^\varepsilon) = [32,5 + 0,5\varepsilon; 60]$ . В этом примере отклонение одной из точек множества  $X$  на  $\varepsilon$  приводит к отклонению одной из границ средних по  $P_{E\Delta}$  на  $0,5\varepsilon$ .

Рассмотрим общий случай. Пусть множество  $X$  точек (1.1) изменилось до множества  $X^\varepsilon = \{x_1 + \varepsilon_1, x_2 + \varepsilon_2, \dots, x_n + \varepsilon_n\}$ , где  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  – произвольные вещественные числа.

*Теорема 1.2. Средняя по  $P_{E\Delta}$  является устойчивой в таком смысле: при изменении множества точек  $X$  до  $X^\varepsilon$  границы множества средних  $G^{E\Delta}(X) = [\alpha, \beta]$  изменяются не более, чем на величину*

$$\max \{|\varepsilon_1|, |\varepsilon_2|, \dots, |\varepsilon_n|\}. \quad (1.4)$$

Таким образом, средние по  $P_\emptyset$  и по  $P_{E\Delta}$  являются устойчивыми к малым изменениям данных (1.1), а средние по  $P_E$  могут быть весьма неустойчивыми.

#### 1.1.4 Случай данных с повторениями

Пусть имеется совокупность данных с повторениями: величина  $x_1$  повторяется  $\beta_1$  раз, величина  $x_2$  повторяется  $\beta_2$  раз, ..., величина  $x_n$  повторяется  $\beta_n$  раз, т. е. вместо данных (1.1) имеется Таблица 1.3.

Таблица 1.3 – Данные с повторениями

Значение $x_i$	$x_1$	$x_2$	...	$x_n$
Вес $\beta_i$	$\beta_1$	$\beta_2$	...	$\beta_n$

В статистике числа  $\beta_i$  называются весами, или (абсолютными) частотами, и они используются для расчетов взвешенных средних. Например, средняя взвешенная арифметическая рассчитывается по формуле:

$$g^1(X) = \frac{1}{\sigma} \sum_{i=1}^n \beta_i x_i,$$

где  $\sigma = \beta_1 + \beta_2 + \dots + \beta_n$ .

Все результаты, полученные ранее, включая определения ПН-средних, сразу переносятся и на этот, более общий случай, так как можно рассматривать совокупность величин  $x_1, x_1, \dots, x_1$  (всего  $\beta_1$  штук),  $x_2$  (всего  $\beta_2$  штук) и т. д., т. е. представить данные из Таблицы 1.1 в виде (1.1):

$$\underbrace{(x_1, \dots, x_1)}_{\beta_1}, \underbrace{(x_2, \dots, x_2)}_{\beta_2}, \dots, \underbrace{(x_n, \dots, x_n)}_{\beta_n}.$$

Однако применение прежних методов построения ПН-средних здесь может оказаться весьма обременительным, если размерность задачи резко возрастет. Поэтому целесообразно пользоваться решающими правилами, разработанными в теории количественной важности критериев с непрерывной (континуальной) шкалой [5], [6], рассматривая натуральные числа  $\beta_i$  как количественные величины важности критериев, а вместо обозначений  $P^E$  и  $P^{E\Delta}$  соответствующих отношений использовать обозначения  $P^\beta$  и  $P^{\beta\Delta}$ .

Для формулировки указанных правил для векторных оценок  $y$  и  $z$  введем в рассмотрение такие множество и величины:

$$W(y, z) = \{y_1\} \cup \{y_2\} \cup \dots \cup \{y_m\} \cup \{z_1\} \cup \{z_2\} \cup \dots \cup \{z_m\} = \{w_1, w_2, \dots, w_q\},$$

$$w_1 > w_2 > \dots > w_q; b_k(y) = \sum_{i:y_i \geq w_k} \beta_i, b_k(z) = \sum_{i:z_i \geq w_k} \beta_i \quad k = 1, 2, \dots, q - 1. \quad (1.5)$$

Решающее правило для отношения  $P^\beta$ :

$$y P^\beta z \Leftrightarrow b_k(y) \leq b_k(z), \quad k = 1, 2, \dots, q - 1, \quad (1.6)$$

причем хотя бы одно из неравенств является строгим.

Решающее правило для отношения  $P^{\beta\Delta}$ :

$$yP^{\beta\Delta}z \Leftrightarrow \sum_{j=1}^k b_j(y)(w_j - w_{j+1}) \leq \sum_{j=1}^k b_j(z)(w_j - w_{j+1}), k = 1, 2, \dots, q-1.$$

(1.7)

причем хотя бы одно из неравенств является строгим.

Пример 1.7. Исходные данные заданы в Таблице 1.4.

Таблица 1.4 – Данные к примеру 1.7

Значение $x_i$	1	2	4	5	7	9	11
Вес $\beta_i$	2	1	4	1	2	3	1

С помощью решающих правил (1.6) и (1.7) сравним точки 5 и 3, имеющие векторные оценки

$$y = f(5) = (4, 3, 1, 0, 2, 4, 6) \text{ и } z = f(3) = (2, 1, 1, 2, 4, 6, 8).$$

Здесь  $\sigma = 14$ ,  $W = (8, 6, 4, 3, 2, 1, 0)$ , так что  $q = 7$ .

Для сокращения записи с учетом формул (1.5) и (1.7) введем в рассмотрение векторы:

$$b(y) = (b_1(y), b_2(y), \dots, b_6(y)), b(z) = (b_1(z), b_2(z), \dots, b_6(z));$$

$$d(y) = (d_1(y), d_2(y), \dots, d_6(y)), d(z) = (d_1(z), d_2(z), \dots, d_6(z)),$$

где

$$d_k(y) = \sum_{j=1}^k b_j(y)(w_j - w_{j+1}), d_k(z) = \sum_{j=1}^k b_j(z)(w_j - w_{j+1}), k = 1, 2, \dots, 6.$$

Согласно формулам (1.5) и (1.7) имеем:

$$b(y) = (0, 1, 6, 7, 9, 13), b(z) = (1, 4, 6, 6, 9, 14);$$

$$d(y) = (0, 2, 8, 15, 24, 37), d(z) = (2, 10, 16, 22, 31, 45).$$

Поскольку  $b_1(y) = 0 < b_1(z) = 1$ , но  $b_4(y) = 7 > b_4(z) = 6$ , то, согласно правилу (1.6), неверно ни  $yP^{\beta}z$ , ни  $zP^{\beta}y$ . Однако, так как все шесть неравенств (1.7) выполнены, и среди них есть строгое, то верно  $yP^{\beta\Delta}z$ .

А формулу (1.3) удобнее использовать, предварительно представив данные с повторениями в виде (1.1).

Пример 1.8. Представим данные из примера 1.7 (из Таблицы 1.4) в виде (1.1), для удобства сведя их в Таблицу 1.5, в которой указаны и номера точек  $i$  (для удобства введено обозначение  $x_{(i)}$ ).

Таблица 1.5 – Данные к примеру 1.8

Номер $i$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Значение $x_{(i)}$	1	1	2	4	4	4	4	5	7	7	9	9	9	11

Пользуясь формулами (1.3), последовательно вычисляем:

$$\begin{aligned} \alpha &= \frac{1}{2} \min \{x_{(1)} + x_{(14)}, x_{(2)} + x_{(13)}, x_{(3)} + x_{(12)}, \\ &\quad x_{(4)} + x_{(11)}, x_{(5)} + x_{(10)}, x_{(6)} + x_{(9)}, x_{(7)} + x_{(8)}\} = \\ &= \frac{1}{2} \min \{1 + 11, 1 + 9, 2 + 9, 4 + 9, 4 + 7, 4 + 7, 4 + 5\} = \\ &= \frac{1}{2} \min \{12, 10, 11, 13, 11, 9\} = 4,5; \\ \beta &= \frac{1}{2} \max \{x_{(1)} + x_{(14)}, x_{(2)} + x_{(13)}, x_{(3)} + x_{(12)}, x_{(4)} + x_{(11)}, \\ &\quad x_{(5)} + x_{(10)}, x_{(6)} + x_{(9)}, x_{(7)} + x_{(8)}\} = \frac{1}{2} \max \{12, 10, 11, 13, 11, 11, 9\} = 6,5. \end{aligned}$$

Таким образом,  $G^{\beta\Delta}(X) = [4,5; 6,5]$ .

### 1.1.5 Случай порядковых данных

Пусть шкала, в которой измерены данные, образующие множество  $X = \{x_1, x_2, \dots, x_n\}$ , является порядковой. Удаленность («расстояние»)  $\delta(x', x'')$  точки  $x'$  от точки  $x''$ , среди которых есть точка из множества  $X$ , будем оценивать числом шагов «по кочкам» – точкам из множества  $X$ . Пусть  $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$ , где  $f_i(x) = \delta(x, x_i)$ . Здесь имеет смысл рассматривать только ПН-средние по  $P_\emptyset$  и  $P_E$ .

Теорема 1.3. Средними по  $P_\emptyset$  являются все заданные точки, и только они:  $G^\emptyset(X) = X$ .

Теорема 1.4. При нечетном  $n$  медиана  $\mu_X = x_{(\frac{n+1}{2})}$  является и средней по  $P_E$ , а при четном  $n$  средними по  $P_E$  являются точки  $x_{(\frac{n}{2})}$  и  $x_{(\frac{n}{2}+1)}$ . Если в множестве  $X$  нет совпадающих точек, то указанные одна точка (при нечетном  $n$ ) или две точки (при четном  $n$ ) исчерпывают множество средних  $G^E(X)$ .

Пример 1.9. Пусть  $n = 3$  и  $x_1 < x_2 < x_3$  (можно записать любые числа, удовлетворяющие этим неравенствам). Обозначим через  $x_0, x_{12}, x_{23}$  и  $x_4$  произвольные точки, лежащие соответственно левее  $x_1$ , между  $x_1$  и  $x_2$ , между  $x_2$  и  $x_3$  и правее  $x_3$ . Тогда

$$f(x_0) = (1, 2, 3), f(x_1) = (0, 1, 2), f(x_{12}) = (1, 1, 2), f(x_2) = (1, 0, 1), f(x_{23}) = (2, 1, 1), \\ f(x_3) = (2, 1, 0), f(x_4) = (3, 2, 1).$$

Множество точек, недоминируемых по  $P_\emptyset$ , есть  $X = \{x_1, x_2, x_3\}$ . Далее:

$$f\uparrow(x_1) = (0, 1, 2), f\uparrow(x_2) = (0, 1, 1), f\uparrow(x_3) = (0, 1, 2),$$

и поэтому единственной недоминируемой по  $P_E$  точкой является  $\mu_X = x_2$ .

Пример 1.10. Пусть  $n = 3$  и  $x_1 = x_2 < x_3$ . Аналогично примеру 1.9, можно убедиться в том, что множество точек, недоминируемых по  $P_\emptyset$ , есть  $X = \{x_1, x_2, x_3\}$  и двумя недоминируемыми по  $P_E$  точками являются  $x_1$  и  $x_2$ .

Пример 1.11. Пусть  $n = 4$  и  $x_1 < x_2 < x_3 < x_4$ . Используя обозначения, аналогичные обозначениям из примера 1.9, имеем:

$$f(x_0) = (1, 2, 3, 4), f(x_1) = (0, 1, 2, 3), f(x_{12}) = (1, 1, 2, 3), f(x_2) = (1, 0, 1, 2), \\ f(x_{23}) = (2, 1, 1, 2), f(x_3) = (2, 1, 0, 1), f(x_{34}) = (3, 2, 1, 1), f(x_4) = (3, 2, 1, 0), \\ f(x_5) = (4, 3, 2, 1).$$

Множество недоминируемых по  $P_\emptyset$  точек есть  $X = \{x_1, x_2, x_3, x_4\}$ . Далее:

$$f\uparrow(x_1) = (0, 1, 2, 3), f\uparrow(x_2) = (0, 1, 1, 2), f\uparrow(x_3) = (0, 1, 1, 2), f\uparrow(x_4) = (0, 1, 2, 3), \\ \text{и поэтому недоминируемыми по } P_E \text{ точками являются } x_2 \text{ и } x_3.$$

Пример 1.12. Пусть  $n = 5$  и  $x_1 = x_2 < x_3 < x_4 < x_5$ . Тогда

$$f(x_1) = f(x_2) = (0, 0, 1, 2, 3), f(x_3) = (1, 1, 0, 1, 2), f(x_4) = (2, 2, 1, 0, 1), \\ f(x_5) = (3, 3, 2, 1, 0).$$

$$f\uparrow(x_1) = f\uparrow(x_2) = (0, 0, 1, 2, 3), f\uparrow(x_3) = (0, 1, 1, 1, 2), f\uparrow(x_4) = (0, 1, 1, 2, 2), \\ f\uparrow(x_5) = (0, 1, 2, 3, 3).$$

Здесь недоминируемыми по  $P_E$  точками являются  $x_1, x_2$  и медиана  $x_3$ .

### 1.1.6 Некоторые выводы и возможные обобщения

Выше рассматривалась проблема определения средних для точек на прямой. Однако предложенный подход к определению ПН-средних очевидным образом распространяется и на более общие случаи.

Пусть, например,  $x_1, x_2, \dots, x_n$  – точки в  $m$ -мерном евклидовом пространстве  $E^m$ . Тогда близость точки  $x$  к совокупности указанных  $n$  точек можно характеризовать значением векторного критерия  $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$ , где  $f_i(x) = d(x, x_i)$  – расстояние между точками  $x$  и  $x_i$ . А далее можно формулировать определения ПН-средних вполне аналогично предложенным выше для одномерного случая. Однако свойства средних и методы построения множеств средних здесь оказываются более сложными. Рассмотрение этих вопросов является предметом дальнейших исследований. Здесь отметим лишь, что средние по  $P_\emptyset$  образуют выпуклую оболочку заданных  $n$  точек.

Многокритериальный подход к определению средних оказался эффективным для самых различных видов исходных данных, в том числе порядковых. Среди введенных таким образом средних наиболее перспективными для практического применения являются средние по  $P_{E\Delta}$ : множество таких средних  $G^{E\Delta}(X)$  имеет простую структуру – является отрезком вида  $[\alpha, \beta]$  – и устойчиво к малым изменениям исходных данных (при изменении множества точек  $X$  до  $X^\varepsilon$  границы множества средних  $G^{E\Delta}(X) = [\alpha, \beta]$  изменятся не более, чем на величину (1.4)). Кроме того, разработан простой и точный аналитический метод построения множества  $G^{E\Delta}(X)$ : найдены формулы для расчета границ отрезка – величин  $\alpha$  и  $\beta$ . Для упрощенного применения указанной средней отрезок  $[\alpha, \beta]$  можно представлять одной точкой – его серединой  $\gamma = \frac{1}{2}(\alpha + \beta)$ , которую можно назвать сконцентрированной, или стянутой средней (по  $P_{E\Delta}$ ).

Полученные результаты о средних опубликованы в статье [7].

## **1.2 Исследование задач регрессионного анализа методами много- критериального принятия решений**

### **1.2.1 Постановка задачи регрессионного анализа как многоокритериаль- ной**

Задача регрессионного анализа состоит в установлении формы корреляционной связи между зависимой и независимыми переменными, т. е. вида функции регрессии (линейная, квадратичная, показательная и т. д.), и оценке её параметров [8]. В случае парной регрессии исходными данными для построения функции регрессии являются пары наблюденных значений независимой переменной (фактора)  $x$  и зависимой переменной (отклика)  $y$ :

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n). \quad (1.8)$$

Уравнение линейной регрессии имеет такой вид:

$$y = ax + b, \quad (1.9)$$

где  $a$  и  $b$  – коэффициенты регрессии, которые нужно оценить с использованием исходных данных (1.8). Для такого оценивания вводятся в рассмотрение невязки, или уклонения теоретических значений независимой переменной – рассчитанных по формуле (1.9) – от наблюденных:

$$\delta_1 = |ax_1 + b - y_1|, \delta_2 = |ax_2 + b - y_2|, \dots, \delta_n = |ax_n + b - y_n|. \quad (1.10)$$

Наиболее распространенным методом оценивания является метод наименьших квадратов (МНК), согласно которому значения коэффициентов регрессии определяются из условия:

$$\sum_{i=1}^n \delta_i^2 \rightarrow \min_{a, b}. \quad (1.11)$$

Для расчета значений коэффициентов регрессии, минимизирующих (1.11), используются простые формулы [8].

Достаточно распространенным, особенно в последние годы, стал метод наименьших модулей (МНМ), согласно которому значения коэффициентов регрессии определяются из условия:

$$\sum_{i=1}^n |\delta_i| \rightarrow \min_{a, b}. \quad (1.12)$$

Разработаны специальные алгоритмы нахождение значений коэффициентов регрессии [9]. Отметим, что нелинейную оптимизационную задачу (1.12) можно свести к эквивалентной задаче линейного программирования.

По своей сути, задача оценивания параметров регрессии является многокритериальной: критериями  $f_i$ , подлежащими минимизации, служат невязки  $\delta_i$  из (1.10), составляющие векторный критерий  $f(a, b) = (f_1(a, b), f_2(a, b), \dots, f_n(a, b))$ , где  $f_i(a, b) = \delta_i$ . МНК и МНМ основаны, по существу, на свертывании векторного критерия в один обобщенный критерий. В общем постановке выбор вида обобщенного критерия – непростая проблема. В регрессионном анализе она решается обоснованно лишь для некоторых специальных частных случаев, когда распределение вероятностей ошибки измерений подчиняется закону Гаусса, или нормальному (и тогда рекомендуется МНК) или закону Лапласа (и тогда рекомендуется МНМ).

Пусть на множестве векторных оценок  $Z = \mathbb{R}_+^n = [0, +\infty)^n$  задан строгий частичный порядок  $P^\Gamma$ , где  $\Gamma$  – информация о предпочтениях ЛПР, касающаяся удаленности: если верно  $z' P^\Gamma z''$ , то точка  $z' = f(a', b')$  ближе ко множеству пар точек (двумерных векторов)  $\{x_i, y_i\}$  (см. (1.8)), чем  $z'' = f(a'', b'')$ . Отношение  $P^\Gamma$  индуцирует на плоскости  $\mathbb{R}^2$  аналогичное по смыслу отношение  $P_\Gamma$ :

$$(a, b)P_\Gamma(c, d) \Leftrightarrow f(a, b)P^\Gamma f(c, d).$$

В качестве общего подхода к оцениванию параметров регрессии можно рассматривать выделение в качестве решения множество точек (векторов)  $(a, b) \in \mathbb{R}^2$ , которые недоминируемые по отношению  $P_\Gamma$ .

При указанном многокритериальном подходе выделяется не одно оптимальное значение коэффициентов регрессии ( $a^*, b^*$ ), а множество недоминируемых значений  $U^\Gamma$ . Каждой паре  $(a, b) \in U^\Gamma$  соответствует «своя» линия регрессии – график функции (1.9), а множеству  $U^\Gamma$  – «пучок» таких линий.

Множественность выделяемых значений параметров регрессии – плата за то, что при многокритериальной постановке задачи их оценивания не делается никаких дополнительных ограничительных допущений о характере (свойствах) исходных данных (1.8) типа фиксирования вида закона распределения вероятностей ошибок измерений и т. п.

### 1.2.2 Регрессия при различных видах информации $\Gamma$

Рассмотрим вначале случай, когда никакой информации  $\Gamma$  нет:  $\Gamma = \emptyset$ . Здесь в роли частичного порядка  $P^\Gamma$  выступает отношение Парето  $P^\emptyset$ , которое определяется (с учетом желательности минимизации критериев) так:

$$z' P^\emptyset z'' \Leftrightarrow z'_i \leq z''_i, i = 1, 2, \dots, n, \text{ причем хотя бы одно из неравенств строгое.}$$

Нетрудно проверить, что значения параметров регрессии ( $a^*, b^*$ ), определяемые из условия:

$$\sum_{i=1}^n \delta_i^s \rightarrow \min_{a, b}, \quad (1.13)$$

где  $s > 0$ , входят во множество  $U^\emptyset$ , так как функция  $\sum_{i=1}^n t_i^s$  возрастает по  $P^\emptyset$  на  $\mathbb{R}_+^n$ . В частности, в  $U^\emptyset$  входят точки, выделяемые МНК и МНМ.

Пусть теперь все критерии имеют равную важность (информация  $E$ ). В этом случае имеем отношение  $P^E$ , которое определяется следующим решающим правилом:

$$z' P^E z'' \Leftrightarrow [z'_{[1]} \leq z''_{[1]}, z'_{[2]} \leq z''_{[2]}, \dots, z'_{[n]} \leq z''_{[n]}], \quad (1.14)$$

причем хотя бы одно из нестрогих неравенств является строгим.

Нетрудно проверить, что значения параметров регрессии  $(a^*, b^*)$ , определяемые из условия (1.13), входят во множество  $U^E$ , так как функция  $\sum_{i=1}^n t_i^s$ , где  $s > 0$ , возрастает по  $P^E$  на  $\text{Re}_+^n$ . В частности, в это множество входят точки, выделяемые МНК и МНМ.

Пусть, наконец, все критерии имеют равную важность, и шкала критериев является шкалой первой порядковой метрики (информация  $E\Delta$ ). В этом случае имеем частичный порядок  $P^{E\Delta}$ , задаваемый следующим решающим правилом (сравни с (1.14)):

$$z' P^{E\Delta} z'' \Leftrightarrow (z'_{[1]} \leq z''_{[1]}, z'_{[1]} + z'_{[2]} \leq z''_{[1]} + z''_{[2]}, \dots, z'_{[1]} + z'_{[2]} + \dots + z'_{[n]} \leq z''_{[1]} + z''_{[2]} + \dots + z''_{[n]}) \quad (1.15)$$

причем среди неравенств есть строгое.

Нетрудно проверить, что значения параметров регрессии  $(a^*, b^*)$ , определяемые из условия (1.13) при  $s > 1$ , входят во множество  $U^{E\Delta}$ , так как функция  $\sum_{i=1}^n t_i^s$ , где  $s > 1$ , возрастает по  $P^{E\Delta}$  на  $\text{Re}_+^n$  (см. (1.15)). В частности, в это множество входят точки, выделяемые МНК. Так как функция  $\sum_{i=1}^n t_i$  является лишь неубывающей по  $P^{E\Delta}$  на  $\text{Re}_+^n$ , то значение  $(a^*, b^*)$ , выделяемое МНМ, входит в  $U^{E\Delta}$ , если оно единственno.

Эффективные вычислительные методы построения множества  $U^\Gamma$  для рассмотренных случаев информации  $\Gamma$  пока не разработаны. Поэтому для его аппроксимации можно воспользоваться следующим «лобовым» подходом: построить «достаточно мелкую» сетку, покрывающую «подходящую» область в  $\text{Re}_+^2$  (она определяется с учетом специфики задачи, т.е. расположения точек из (1.8) на плоскости), рассматривать эту сетку вместо множества  $\text{Re}_+^2$  и выделять из её узлов те из них, которые являются недоминируемыми по  $P_{E\Delta}$ . Именно таким методом решался приведенный ниже пример с применением специально разработанной компьютерной программы.

Пример 1.13. Пусть исходные данные (1.8) заданы Таблицей 1.6.

Таблица 1.6 –Данные к примеру 1.13

$x_i$	1	2	3	4	5	7
$y_i$	3	2	5	5	4	7

Здесь уравнения регрессии, определяемые МНК и МНМ соответственно, таковы:

$$y = \frac{47}{70}x + 1\frac{61}{70} \approx 0,671x + 1,871.$$

$$y = \frac{2}{3}x + 2\frac{1}{3} \approx 0,667x + 2,333.$$

На рисунках 1.2–1.5 представлены пучки линий регрессии и множества  $U^\Gamma$  для рассмотренных выше видов информации  $\Gamma$ . При этом красным цветом выделены линия регрессии и точка  $(a^*, b^*)$ , соответствующие МНК, а желтым – МНМ.

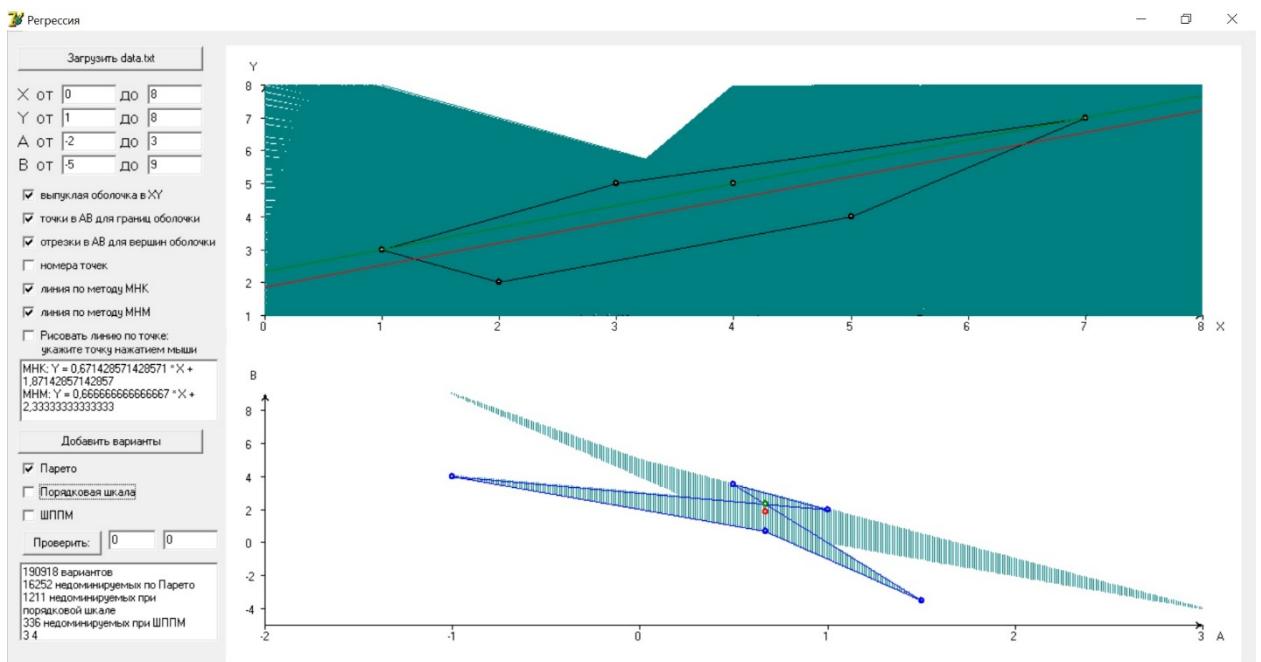


Рисунок 1.2 – Случай  $\Gamma = \emptyset$

Рисунок 1.2 иллюстрирует следующее доказанное в [10] свойство множества  $U^\emptyset$ : оно является замкнутым, связным, возможно, невыпуклым, ограниченным объединением конечного числа политопов (выпуклых оболочек конечных наборов двумерных векторов).

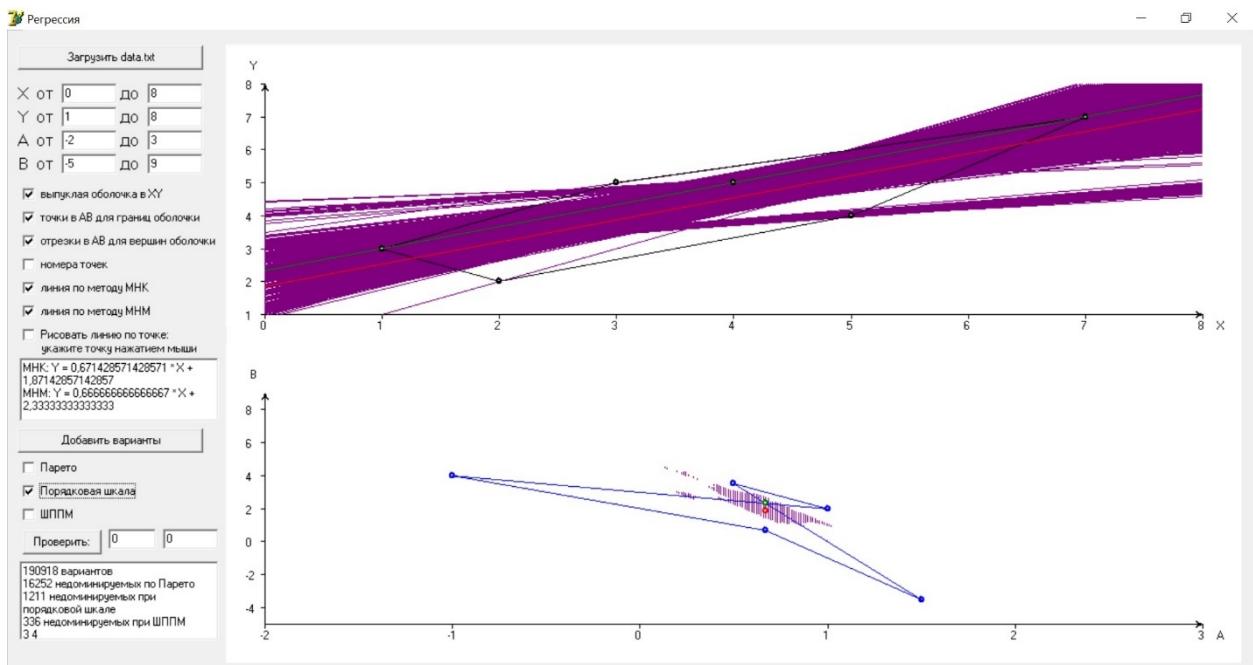


Рисунок 1.3 – Случай  $\Gamma = E$

На рисунке 1.3 видно, что множество  $U^E$  несвязно.

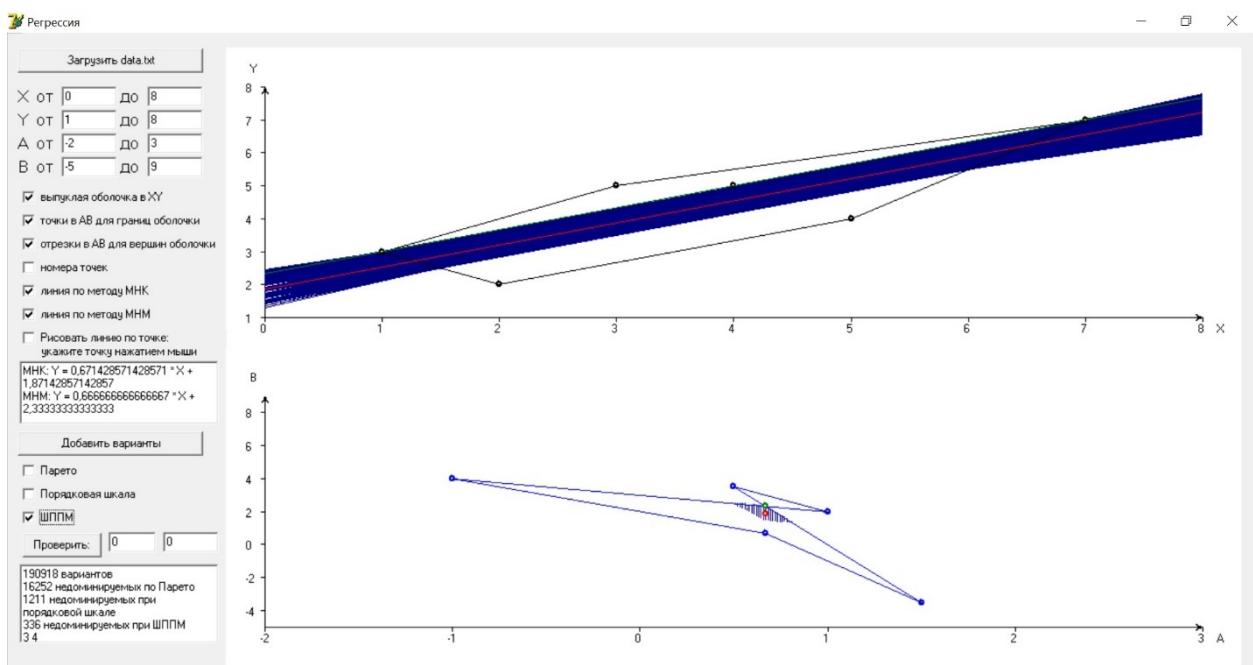


Рисунок 1.4 – Случай  $\Gamma = E\Delta$

На рисунке 1.4 видно, что множество  $U^{E\Delta}$  связно.

На рисунке 1.5 для наглядности сведены результаты расчетов для всех трех вышеуказанных случаев.

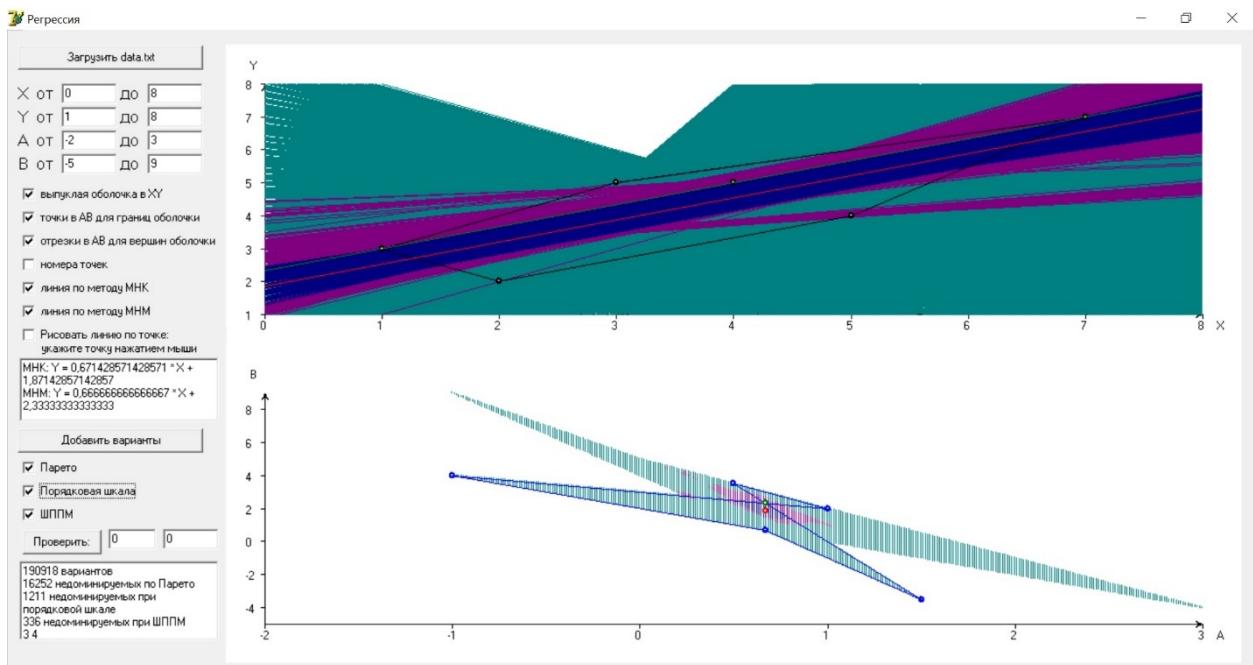


Рисунок 1.5 – Случаи  $\Gamma = \emptyset$ ,  $\Gamma = E$ ,  $\Gamma = E\Delta$

### 1.2.3 Некоторые выводы

Множественный подход к определению параметров регрессии выделяет не одну линию регрессии, а пучок таких линий. Множественность выделяемых значений параметров регрессии – плата за то, что при многокритериальной постановке задачи их оценивания не делается никаких дополнительных ограничительных допущений о характере (свойствах) исходных данных. Актуальной остается проблема разработки эффективных вычислительных методов построения множеств недоминируемых значений параметров регрессии.

## 1.3 Исследование структурированных дихотомических предпочтений

### 1.3.1 Введение

Индивидуальные предпочтения, с одной стороны, и агрегирование этих предпочтений в один коллективный выбор, с другой стороны, составляют центральные элементы исследований искусственного интеллекта [11]. С приложениями, от рекомендательных систем до электронного голосования и ав-

томатизированных персональных помощников, проблема выбора подходящих моделей предпочтений и методов агрегирования становится актуальной. Но хорошо отлаженные механизмы агрегирования не всегда легко найти – агрегирование часто является вычислительно трудной задачей (см., например, [12]). Положительные результаты обнаруживаются при условии, что предпочтения агентов соответствуют определенной структуре, также известной как ограничение домена. По указанной выше причине ограничения предпочтений избирателей привлекают все большее внимание в области вычислительной теории общественного выбора (недавний обзор см. в [13]). На концептуальном уровне ограниченные домены представляют собой структуры, которые возникают как разумные модели предпочтений во многих реальных условиях; на техническом уровне они позволяют эффективно применять несколько механизмов голосования, использование которых, как правило, является сложной вычислительной задачей.

В частности, хорошо изучены ограниченные области предпочтений, которые представляют собой линейный порядок на множестве альтернатив. Напротив, области дихотомических предпочтений, в которых избиратели имеют одобренный и не одобренный набор альтернатив – хотя и очень естественны – были разработаны только недавно. Элкинд и Лакнер [14] представили различные сети такого типа (позже обобщенные [15]); они показали, что ряд структур дихотомических предпочтений допускает полиномиальные алгоритмы в контексте двух популярных правил пропорционального представительства, для которых определение победившего комитета, как известно, является NP-сложной задачей.

Раздел содержит два основных результата, опубликованных в [16]. Во-первых, доказаны теоремы о характеризации структурированных дихотомических доменов, на которых основаны результаты Элкинда и Лакнера [14], путем выявления закономерностей, которые не позволяют профилю предпочтения проявлять определенную структуру. Во-вторых, разработаны полиномиальные алгоритмы, которые дают конструктивные ответы на вопрос о воз-

можности заполнения неполного профиля предпочтений с соблюдением определенных ограничений, приводящим к известным структурированным предпочтениям.

Можно ли заполнить неполный профиль, сохраняя нужные свойства? Если ответ положительный, можно ли эффективно обнаружить подходящее завершение.

### 1.3.2 Основные результаты

Дихотомические предпочтения задаются разбиением множества альтернатив на два подмножества: поддерживаемые альтернативы и неприемлемые альтернативы.

Дихотомические предпочтения используются в одобряющем голосовании [17], в задаче пропорционального представительства [18], в задаче нахождения паросочетаний [19], в задаче распределения ресурсов [20], [21], в гедонических играх [22] и других приложениях.

Несмотря на простую структуру многие вычислительные задачи, связанные с дихотомическими предпочтениями, оказываются NP-трудными, что привело к созданию теории доменов дихотомических предпочтений – ограничений на множество возможных предпочтений. Э. Элкинд и М. Лакнер [14] показали, что многие естественные ограничения на профиль предпочтений приводят к упрощению вычислительных задач пропорционального представительства.

Мы имеем множество альтернатив  $A = \{1, \dots, m\}$  и множество агентов  $\mathcal{N} = \{1, \dots, n\}$ . Каждый агент  $i \in \mathcal{N}$  одобряет множество  $A_i \subseteq A$ . Составим матрицу  $X$  размера  $m$  на  $n$ , в которой  $x_{ij} = 1$ , если агент  $i$  ободряет альтернативу  $j$ , и  $x_{ij} = 0$ , если агент  $j$  не ободряет альтернативу  $i$ . Матрица  $X_{m \times n}$  представляет собой профиль предпочтений. Множество всех возможных бинарных матриц размера  $m \times n$  будем обозначать как  $\mathcal{M}_{m \times n}$ . Существуют  $2^{mn}$  различных матриц и, соответственно, различных профилей предпочтений. Доменом предпочтений называется подмножество возможных столбцов в бинарной матрице.

Будем говорить, что матрицы  $X$  и  $Y$  *эквивалентны* с обозначением,  $X \equiv Y$ , если матрица  $X$  равна матрице  $Y$  после перестановки строк и столбцов:  $X \equiv Y$  тогда и только тогда, когда  $x_{ij} = y_{\sigma(i)\tau(j)}$ ,  $\sigma(i) \in S_m$ ,  $\tau(j) \in S_n$ , где  $S_k$  – множество всех перестановок на множестве  $\{1, \dots, k\}$ .

Будем говорить, что матрица  $Q_{k \times l}$  является *паттерном* в матрице  $A_{m \times n}$ , если существует подматрица  $B_{k \times l}$  матрицы  $A_{m \times n}$  такая что  $Q_{k \times l} \equiv B_{k \times l}$ . Паттерны играют роль конфигураций в ординальных предпочтениях.

Для каждого линейного порядка можно определить множество дихотомических предпочтений, которые могут быть достигнуты при некотором уровне одобрения агентом.

Для однопиковых предпочтений множества одобрения (некоторое количество наилучших альтернатив) обязательно образуют интервал на оси альтернатив. Таким образом, мы имеем домен предпочтений с последовательными единицами в столбцах (интервалы кандидатов).

**Теорема 1.5 [23].** Профиль дихотомических предпочтений удовлетворяет свойству последовательных единиц в столбцах (аналог однопиковости в случае дихотомических предпочтений) тогда и только тогда профиль не содержит паттернов 1–5, представленных на рисунках 1.6–1.10.

$$\left[ \begin{array}{cccccc|c} 1 & 0 & 0 & & & & 1 \\ 1 & 1 & 0 & \cdots & & 0 & 0 \\ 0 & 1 & 1 & & & & \\ \vdots & & \ddots & & \vdots & & \cdots \\ & & & 1 & 1 & 0 & \\ 0 & & \cdots & 0 & 1 & 1 & 0 \\ & & & 0 & 0 & 1 & 1 \end{array} \right] \in \mathcal{M}_{k+2 \times k+2}, k \geq 1.$$

Рисунок 1.6 – Паттерн 1

$$\left[ \begin{array}{cccc|c} 1 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 1 & 1 & 0 & \cdots & & 1 & 1 \\ 0 & 1 & 1 & & & & \dots \\ \vdots & & \ddots & & \vdots & & \dots \\ & & & 1 & 1 & 0 & 1 & 1 \\ 0 & \cdots & 0 & 1 & 1 & & 1 & 1 \\ 0 & 0 & 1 & & & 0 & 1 \\ \hline 0 & \cdots & & & 0 & 0 & 1 & 1 \end{array} \right] \in \mathcal{M}_{k+2 \times k+3}, k \geq 1.$$

Рисунок 1.7 – Паттерн 2

$$\left[ \begin{array}{cccc|c} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 1 & & & \dots \\ \vdots & & \ddots & & \vdots & \dots \\ & & & 1 & 1 & 0 \\ 0 & \cdots & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & & & 0 \\ \hline 0 & \cdots & & & 0 & 0 & 1 \end{array} \right] \in \mathcal{M}_{k+3 \times k+2}, k \geq 1.$$

Рисунок 1.8 – Паттерн 3

$$\left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{array} \right].$$

Рисунок 1.9 – Паттерн 4

$$\left[ \begin{array}{cccc} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right].$$

Рисунок 1.10 – Паттерн 5

Теорема 1.5 [23]. Профиль дихотомических предпочтений удовлетворяет свойству последовательных единиц в столбцах (аналог однопиковости в случае дихотомических предпочтений) тогда и только тогда профиль не содержит паттернов 1–5, представленных на рисунках 1.6–1.10.

Характеристика из теоремы 1.5 содержит бесконечное число паттернов, но так как паттерны схожи друг с другом, задача поиска паттерна минимального размера решается за полиномиальное время [24]. Существует простой алгоритм для тестирования свойства последовательных единиц в матрице, основанный на PQ-деревьях [25].

Транспонированные паттерны 1–5, представленные на рисунках 1.6–1.10, дадут характеристику домена с последовательными единицами в строках (интервалы агентов), рассмотренного в [14]. В литературе можно найти характеристизации бинарных матриц с круговым свойством последовательных единиц [26] со свойством последовательных единиц на дереве [27].

3. Терзопулу и др. [16] исследовали домены предпочтений с ограниченными интервалами агентов. Данные интервалы образуются только у краев оси (домен предпочтений с *внешними интервалами агентов, далее ВИА*) или только у одного края оси (домен предпочтений с *одним внешним интервалом агентов, далее ОВИА*). В случае голосования за кандидатов от политических партий избиратели могут быть упорядочены по степени лояльности к одной из двух партий. Наиболее лояльные избиратели будут одобрять всех кандидатов своей партии, менее лояльные избиратели могут одобрять некоторое количество кандидатов обеих партий. Примером профиля предпочтений с ВИА является

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

Каждый кандидат получает одобрение от интервала агентов, находящегося у края оси.

Теорема 1.6 [16]. Профиль дихотомических предпочтений удовлетворяет свойству ВИА тогда и только тогда профиль не содержит паттернов 6–10, представленных на рисунке 1.11-1.15.

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Рисунок 1.11 – Паттерн 6

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Рисунок 1.12 – Паттерн 7

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Рисунок 1.13 – Паттерн 8

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Рисунок 1.14 – Паттерн 9

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

Рисунок 1.15 – Паттерн 10

Теорема 1.7 [16]. Профиль дихотомических предпочтений удовлетворяет свойству ОВИА тогда и только тогда профиль не содержит паттерна 11, представленного на рисунке 1.16.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Рисунок 1.16 – Паттерн 11

Матрицы, не содержащие паттерна 11, представленного на рисунке 1.16, могут быть однозначно восстановлены по их суммам строк и столбцов [28].

Теорема 1.8 [28]. Количество профилей дихотомических предпочтений, удовлетворяющих свойству ОВИА, равно

$$\#SVEI(m, n) = \sum_{k=1}^{\min(m,n)} k!^2 S(n+1, k+1) S(m+1, k+1),$$

где  $S(a, b)$  – число Стирлинга второго рода.

В [16] рассмотрены также достаточные и необходимые условия нахождения структурированных дихотомических предпочтений в случае неполной информации, когда мы не имеем информацию о некоторых ячейках в профиле предпочтений. В [15] расширено возможное применение интервальных на оси предпочтений до интервальных на деревьях.

Другим примером структурированных дихотомических предпочтений является разбиение. Профиль дихотомических предпочтений удовлетворяет свойству *разбиения*, если множество одобрения каждого агента совпадает с одним из множеств, принадлежавших некоторому разбиению множества альтернатив.

Теорема 1.9 [16]. Профиль дихотомических предпочтений удовлетворяет свойству ОВИА тогда и только тогда профиль не содержит паттерна 12, представленного на рисунке 1.17.

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

Рисунок 1.17 – Паттерн 12

Теорема 1.10 [29]. Количество профилей дихотомических предпочтений, удовлетворяющих свойству ОВИА, равно

$$\#PART(m, n) = \sum_{k=1}^{\min(m,n)} k! S(n, k)S(m, k),$$

где  $S(a, b)$  – число Стирлинга второго рода.

Свойство разбиения является очень ограничивающим и количество профилей, удовлетворяющих этому свойству меньше, чем количество профилей, удовлетворяющих свойству ОВИА.

Неполный профиль  $P \in I_{m \times n}$  содержит “0”, “1”, и “?”, где ячейки “?” отражает отсутствие данных, эти ячейки нужно заполнить “0” или “1”.

Теорема 1.11 [16]. Существует алгоритм полиномиальной сложности, который отвечает на вопрос о возможном заполнении профиля со свойством ВИА или ОВИА и представляет пример такого заполнения, если он существует.

В работе [16] представлен также алгоритм полиномиальной сложности, который отвечает на вопрос сохранении свойств ВИА или ОВИА при всевозможных заполнениях профиля и представляет пример такого заполнения, если он существует.

### 1.3.3 Некоторые выводы

В этом подразделе отчета представлены результаты исследования структурированных дихотомических предпочтений в двух направлениях. Во-первых, в модели с полной информацией получены характеристики доменов через запрещенные подпрофили, расширяющие предыдущие работы по теме. Во-вторых, в случаях с полной информацией определено может ли неполный

профиль допускать возможное или необходимое завершение в определенных ограниченных доменах.

Тем не менее, один из вопросов остается открытым: какова сложность определения того, обязательно ли завершение неполного дихотомического профиля предпочтений удовлетворяет свойству последовательных единиц в столбцах? Хотя начало этому исследованию положено, ограниченные области дихотомических предпочтений, безусловно, заслуживают дальнейшего анализа. Для полных профилей, которые нарушают данное свойство, было бы интересно узнать минимальное количество значений, которые необходимо поменять местами, чтобы свойство сохранялось.

Такой результат будет актуален для приложений с зашумленными данными [30]. Что касается случаев неполной информации, можно также изучить вероятность соответствия профиля определенной структуре после случайного заполнения или ограничить количество ячеек с неизвестными значениями и провести параметризованный анализ сложности. Все эти проблемы, конечно же, следует исследовать с учетом различных ограничений домена.

## **1.4 Исследование турнирных способов выбора оптимальных вариантов**

### **1.4.1 Введение**

Проблема оптимального выбора заключается в выборе в некотором смысле лучших альтернатив из множества, предъявленного выбирающему. Важным частным случаем является выбор, основанный на результатах попарного сравнения вариантов. Подобную проблему можно представить как задачу выбора победителя (победителей) в спортивном турнире. Следует отметить, что, определив понятие наилучшей альтернативы, методом сортировки можно также решить и проблему ранжирования альтернатив (распределения всех участников турнира по упорядоченным от лучшего к худшему местам).

В литературе предложено несколько различных решений данной задачи (турнирных решений). Свойства этих решений, в целом, хорошо исследо-

ваны в случае, когда орграф турнира является полным, то есть тогда, когда нет равноценных или несравнимых альтернатив, и в любой паре альтернатив (в любой игре турнира) можно определить один и только один наилучший вариант (единственного победителя). В частности, особенную важность имеет такое свойство решения, как устойчивость выбора, понимаемая как независимость результатов выбора от наличия или отсутствия неоптимальных альтернатив в доступном для выбора множестве (так называемая независимость от посторонних альтернатив по Нэшу). Данное свойство обладают отнюдь не все известные турнирные решения, и это обстоятельство придаёт особую ценность тем из них, у кого оно есть. Благодаря наличию устойчивости одним из кандидатов на роль наилучшего турнирного решения является так называемое двухпартийное множество (*bipartisan set*), введённое в [31].

К сожалению, в практических задачах обычны ситуации, когда в предъявлении есть пары равноценных альтернатив (несравнимые альтернативы без потери общности можно считать равноценными). Эти ситуации соответствуют турнирам, в которых разрешены ничьи. Соответственно, возникает естественная необходимость обобщения турнирных решений, предложенных для полных орграфов, на случай выбора победителя (победителей) в турнирах, чей орграф неполон.

Данная работа посвящена рассмотрению так называемого существенного множества (*essential set*), предложенного в качестве обобщения двухпартийного множества. Также рассматривается и так называемое консервативное обобщение двухпартийного множества.

Основными задачами исследования были, во-первых, проверка сохранения существенным множеством ряда важных свойств двухпартийного множества, наличием которых оправдывается его использование в качестве инструмента оптимального выбора (сохранение свойств турнирного решения его консервативным обобщением доказано в [32]), во-вторых, установление теоретико-множественных отношений существенного множества и консервативного обобщения с обобщениями других важных турнирных решений, и в-

третьих, проверка на реальных данных качества ранжирований, полученных методом сортировки с использованием существенного множества в качестве инструмента отбора.

В результате проведённого исследования было получено достаточно полное описание свойств существенного множества, в частности доказано наличие свойства устойчивости. Были найдены теоретико-множественные отношения двух рассматриваемых обобщений с различными турнирными решениями, в частности, с версиями такого популярного турнирного решения как непокрытое множество. На основании ранее известных и полученных в ходе данного исследования теоретических результатов был усовершенствован и программно реализован алгоритм вычисления существенного множества и консервативного обобщения. Созданный в ходе исследования комплекс программ был применен, во-первых, для проверки примеров и поиска контрпримеров для доказательства теоремы, устанавливающей теоретико-множественные отношения турнирных решений. Во-вторых, с его помощью была реализована третья, эмпирическая часть исследования – проверка качества ранжирования, получаемого сортировкой с помощью существенного множества. Для этой проверки были использованы эмпирические данные и методология работы [33], в которой рассматривалась задача построения агрегированного ранжирования научных журналов по менеджменту методами теории коллективного выбора. Оценка относительного качества ранжирований, проведённая по методу, предложенному авторами статьи [33], основанному на анализе ранговых корреляций, показала что полученное в ходе данного исследования ранжирование 93 ведущих журналов по менеджменту, отобранных в [33], представляет совокупность семи агрегируемых библиометрических рейтингов лучше, чем любое из приведённых в [33] ранжирований, как агрегируемое, так и агрегированное.

Пункт 1.4.2 посвящен описанию рассматриваемой теоретической модели. В данном пункте определены основные понятия, в частности понятия

турнирного решения и его свойств. Формулировки определений, используемые в данном разделе, заимствованы в основном из [16].

Пункт 1.4.3 содержит формулировки и доказательства теорем, устанавливающих наличие или отсутствие у рассматриваемого обобщения свойств, определённых в предыдущем разделе. Также в нем рассматривается связь существенного множества и консервативного обобщения двухпартийного множества с другими турнирными решениями, в частности, с различными версиями непокрытого множества.

В пункте 1.4.4 приводится результат сравнения сортировки с помощью существенного множества с пятью другими методами ранжирования альтернатив, также связанными с турниром. Сравнение методов основано на корреляционном анализе ранжирований научных журналов по менеджменту и использует данные и методологию статьи [33].

В пункте 1.4.5 кратко резюмируются основные результаты работы и даётся оценка существенного множества как инструмента оптимального выбора.

## 1.4.2 Общие теоретические сведения

### 1.4.2.1 Выбор и предпочтения

Объекты выбора (альтернативы, варианты) в совокупности составляют *универсальное множество*  $A$  всех потенциально доступных для выбора альтернатив,  $|A| > 2$ . *Предъявлением* называется конечное непустое множество  $X$ , являющееся подмножеством  $A$ ,  $X \subseteq A$ . Предъявление  $X$  представляет собой множество альтернатив, актуально доступных для выбора, и является переменной величиной. Выбор есть разделение предъявления на два подмножества – подмножество выбранных и подмножество отвергнутых альтернатив (последнее может быть пустым). *Функция выбора* или *решение задачи выбора*  $S(X)$  – это функция, которая каждому предъявлению ставит в соответствие подмножество выбранных альтернатив  $S \subseteq X$ .

Предположим, что для любого предъявления, содержащего только две альтернативы, результат оптимального выбора всегда известен. Выбор из пары однозначно определяет на  $A$  бинарное отношение  $\gtrsim$ .

Асимметричную часть  $>$  этого отношения составляют упорядоченные пары  $(a, b)$  элементов  $A$  такие, что при выборе из  $X = \{a, b\}$  выбирается  $a$  и отвергается  $b$ , то есть вариант  $a$  строго предпочитается варианту  $b$ . Из этого определения следует, что отношение  $>$  является полным,  $\forall a, b \in A, a > b \vee b > a$ , антирефлексивным,  $\forall a \in A, \neg a \sim a$ , и асимметричным  $\forall a, b \in A, a > b \Rightarrow \neg b > a$ . Оно репрезентирует строгие предпочтения субъекта выбора.

Симметричную часть  $\sim$  отношения  $\gtrsim$  составляют упорядоченные пары  $(a, b)$  элементов  $A$  такие, что при выборе из  $X = \{a, b\}$  альтернативы  $a$  и  $b$  рассматриваются как равноценные. Без потери общности можно считать, что в этом случае выбираются обе альтернативы.

Из этих определений следует, что отношение  $\gtrsim$  является полным,  $\forall a, b \in A, a \gtrsim b \vee b \gtrsim a$ , рефлексивным,  $\forall a \in A, a \sim a$ , и антисимметричным,  $\forall a, b \in A, (a \gtrsim b \wedge b \gtrsim a) \Rightarrow a = b$ . Отношение  $\gtrsim$  представляет нестрогие предпочтения выбирающего.

Задача оптимального выбора из предъявления, в котором больше двух альтернатив, имеет естественное решение в том случае, когда у сужения отношения  $\gtrsim$  на данное предъявление есть максимальный элемент (или максимальные элементы), то есть альтернатива, которая не хуже (при парном сравнении), чем любая альтернатива, доступная для выбора. Однако хорошо известно, что в задаче коллективного выбора определение предпочтений коллективного субъекта через бинарный выбор, основанный на правиле большинства, может приводить к отношению  $\gtrsim$  без максимальных элементов, что является следствием наличия циклов в его асимметричной части  $>$ . Этот результат известен как парадокс Кондорсе. Следствием данного парадокса является необходимость определить иной, нетривиальный способ выбора наилучших альтернатив, каковым является любое турнирное решение.

#### 1.4.2.2 Турнир и турнирные решения

*Турниром*  $T$  называют пару  $(X, \succ_X)$ , где  $\succ_X$  обозначает сужение отношения  $\succ$  на предъявление  $X$ . Турнир  $T(X, \succ_X)$  можно представить полным ориентированным графом (орграфом), вершины которого представляют элементы множества  $X$ , а дуги – элементы множества  $\succ_X$ . *Слабым турниром*  $W$  называют пару  $(X, \lesssim_X)$ . Граф слабого турнира, в отличие от графа турнира, может быть неполным в случае, если  $\exists(a, b \in X \wedge a \neq b): a \sim b$ .

Функция выбора  $S$ , которая каждому турниру ставит в соответствие некоторое *множество выбора*  $S(T) \subseteq X$  (использующиеся далее обозначения  $S(T)$ ,  $S(X, \succ_X)$ ,  $S(X, \succ)$ ,  $S(\succ_X)$ ,  $S(X)$  являются эквивалентными), называется *турнирным решением*, если она удовлетворяет следующим трём свойствам:

- 1)  $\forall T, S(T) \neq \emptyset$  (*Непустота*);
- 2)  $\forall T, \text{MAX}(\lesssim_X) \subseteq S(T)$ , где  $\text{MAX}(\lesssim_X)$  – множество максимальных элементов  $\lesssim_X$  (*Согласованность по Кондорсе*);
- 3)  $S(\pi(X)) = \pi(S(X))$ , где  $\pi(X)$  – произвольная перестановка альтернатив из  $X$  (*Нейтральность*).

#### 1.4.2.3 Свойства турнирных решений

Турнирные решения оценивают и сравнивают по наличию у них важных свойств, отсутствие которых заставляет сомневаться в разумности и оптимальности такого способа выбора.

##### 1) Монотонность.

Пусть турниры  $\lesssim_X$  и  $\lesssim'_X$  совпадают для всех пар, не содержащих альтернативу  $a$ :  $\lesssim_{X \setminus \{a\}} = \lesssim'_{X \setminus \{a\}}$ . Пусть  $\forall b \in X \setminus \{a\}, (a \gtrsim b \Rightarrow a \gtrsim' b) \wedge (a \succ b \Rightarrow a \succ' b)$ . Тогда решение  $S$  является *монотонным*, если из того, что  $a \in S(\lesssim_X)$  следует, что  $a \in S(\lesssim'_X)$ .

Монотонность турнирного решения означает, что если, при прочих равных условиях, мнение выбирающего об оптимальной альтернативе улучшится, то она останется для него оптимальной. Нарушение данного естественного свойства является сильным аргументом против использования не-монотонного решения в качестве инструмента выбора.

## 2) Устойчивость.

Турнирное решение  $S$  является *устойчивым*, если для любых двух предъявлений  $Y, Z \subseteq A$  таких, что  $Y \cap Z \neq \emptyset$ , имеет место эквивалентность:

$$S(Y, \succsim_Y) = S(Z, \succsim_Z) = X \Leftrightarrow X = S(Y \cup Z, \succsim_{Y \cup Z}).$$

Это утверждение можно рассматривать как конъюнкцию прямой и обратной импликаций. Прямая импликация в работе [35] называется свойством  $\gamma$ , а обратная – свойством  $\alpha$ . Свойство  $\alpha$  также называется свойством отбрасывания [36].

## 3) Композиционная согласованность.

Орграф, репрезентирующий турнир, можно разделить на компоненты. Альтернативы  $x, y \in X$  принадлежат одной *компоненте* орграфа тогда и только тогда, когда они одинаково относятся ко всем остальным альтернативам из других компонент. Турнирное решение называется *композиционно-согласованным*, если для любых турниров  $T, T_1, \dots, T_n$ , и турнира компонент  $\tilde{T}$  таких, что  $T = \Pi(\tilde{T}, T_1, \dots, T_n)$ , верно

$$S(T) = \bigcup_{i \in S(\tilde{T})} S(T_i).$$

Другими словами, турнирное решение является композиционно-согласованным, если оно выбирает лучшие альтернативы из лучших компонент.

В слабом турнире множество  $Y \subseteq X$  можно назвать *компонентой* тогда и только тогда, когда  $|Y| = 1$  или  $\forall x \in X \setminus Y, Y > x \vee x > Y$ . Турнирное решение называется *слабо композиционно-согласованным*, если для любых двух турниров  $T = (X, \succ_X)$  и  $T' = (X, \succ'_X)$ , различающихся только отношением доминирования над альтернативами из некоторой компоненты  $Z$  турнира  $T = (X, \succ_X)$ , верно следующее утверждение:

$$(S(T) \setminus Z = S(T') \setminus Z) \wedge (S(T) \cap Z \neq \emptyset \Leftrightarrow S(T') \cap Z \neq \emptyset).$$

Очевидно, что если турнирное решение удовлетворяет свойству композиционной согласованности, то оно удовлетворяет и свойству слабой композиционной согласованности.

В литературе было введено много различных турнирных решений, но, как уже было сказано, далеко не все из них обладают перечисленными выше свойствами, важными для принятия решения о рациональности той или иной процедуры выбора.

В данной работе рассматривается двухпартийное множество и его обобщения на случай слабого турнира. Остальные решения, такие как, например, непокрытое множество и различные его версии в слабых турнирах, рассматриваются только в связи с двухпартийным множеством и его обобщениями.

#### 1.4.2.4 Непокрытое множество (*UC*)

Непокрытое множество основано на отношении покрытия, для определения которого, требуются понятия верхнего и нижнего срезов и горизонта данной альтернативы.

*Верхний срез*  $D(x)$  альтернативы  $x$  есть множество всех альтернатив, которые строго предпочтительнее  $x$ ,  $D(x) = \{y \in X | y > x\}$ . *Нижний срез*  $L(x)$  альтернативы  $x$  есть множество всех альтернатив, которым строго предпочтается  $x$ ,  $L(x) = \{y \in X | x > y\}$ . *Горизонт*  $H(x)$  альтернативы  $x$  есть множество всех альтернатив равнозначных  $x$ ,  $H(x) = \{y \in X | x \sim y\}$ .

Альтернатива  $x$  покрывает альтернативу  $y$  (обозначим это отношение как  $xCy$ ), если  $x$  не только лучше, чем  $y$ , но и лучше любой альтернативы, которая хуже, чем  $y$ ,  $(x > y) \wedge (L(y) \subset L(x))$  [37]. Альтернатива называется *непокрытой* тогда и только тогда, когда в предъявлении нет ни одной альтернативы, её покрывающей. *Непокрытое множество UC* есть множество всех непокрытых альтернатив предъявления,  $UC(X, >) = \{x \in X | \forall y \in X, \neg yCx\}$ .

Следует отметить, что существуют и другие версии определения отношения покрытия [38], но в турнирах все они эквивалентны. Различия версий

непокрытого множества проявляются при рассмотрении слабых турниров, и некоторые из них описаны в следующем разделе данной работы.

#### 1.4.2.5 Версии непокрытого множества

Существует 15 различных версий определения отношения покрытия [38], а следовательно, и непокрытого множества. Следует отметить, что первые три версии ( $Ia, Ib, Ic$ ) не принято рассматривать в рамках изучения турнирных решений, так как они могут приводить к пустому выбору, что противоречит определению турнирного решения. Все версии отношения покрытия определяются в [38], в данной работе мы ограничимся версиями  $IIa, IIb, IIIa, IIIb, IVa$  и  $Va$ .

Альтернатива  $x$  покрывает альтернативу  $y$  по Миллеру [37] (по версии  $IIa$ ), если  $x$  не только лучше, чем  $y$ , но и лучше любой альтернативы, которая хуже, чем  $y$ ,  $xC^{IIa}y \Leftrightarrow (x > y) \wedge (L(y) \subset L(x))$ . Альтернатива  $x$  покрывает альтернативу  $y$  по версии  $IIb$ , если  $x$  лучше любой альтернативы, которая хуже, чем  $y$ , но для  $y$  это неверно,  $xC^{IIb}y \Leftrightarrow L(y) \subset L(x)$  (включение строгое). Альтернатива  $x$  покрывает альтернативу  $y$  по Фишберну [39] (по версии  $IIIa$ ), если  $y$  не только хуже  $x$ , но и хуже любой альтернативы, которая лучше, чем  $x$ ,  $xC^{IIIa}y \Leftrightarrow (x > y) \wedge (D(x) \subset D(y))$ . Альтернатива  $x$  покрывает альтернативу  $y$  по версии  $IIIb$ , если  $y$  хуже любой альтернативы, которая лучше, чем  $x$ , но для  $x$  это неверно,  $xC^{IIIb}y \Leftrightarrow D(x) \subset D(y)$  (включение строгое). Альтернатива  $x$  покрывает альтернативу  $y$  по МакКельви (по версии  $IVa$ , обозначим это отношение как  $xC^{IVa}y$ ), если  $(x > y) \wedge ((L(y) \subset L(x) \wedge D(x) \subseteq D(y)) \vee (L(y) \subseteq L(x) \wedge D(x) \subset D(y)))$ . Альтернатива  $x$  покрывает альтернативу  $y$  по Дуггану (по версии  $Va$ ), если  $x$  лучше всех альтернатив, которые хуже или равноценны  $y$ ,  $xC^{Va}y \Leftrightarrow H(y) \cup L(y) \subset L(x)$ . По аналогии с определением непокрытого множества, непокрытое по версии  $Na$  множество  $UC^{Na}$  ( $N = I \div V$ ) представляет из себя объединение всех непокрытых по версии  $Na$  альтернатив предъявления,  $UC^{Na}(X, \gtrsim) =$

$$\{x \in X \mid \forall y \in X, \neg y C^{Na} x\}. \quad \text{Аналогично} \quad UC^{Nb}(X, \gtrsim) = \{x \in X \mid \forall y \in X, \neg y C^{Nb} x\}, \text{ где } N = \{II, III\}.$$

#### 1.4.2.6 Двухпартийное множество (*BP*)

Двухпартийное множество основано на теоретико-игровом понятии равновесия по Нэшу в смешанных стратегиях, поэтому для того, чтобы дать его определение, требуется привести ряд фактов из теории игр. Определения некооперативной игры  $G$  и равновесия по Нэшу  $NE$  предполагаются уже известными читателю [31].

Определим функцию сравнения  $g: X \times X \rightarrow \mathbb{R}$

$$g(x, y) = \begin{cases} 1, & x \succ y, \\ -1, & y \succ x, \\ 0, & x \sim y. \end{cases}$$

Определим игру  $G(X, \succ)$  следующим образом. Множество игроков  $N$  состоит из игроков 1 и 2 ("кандидаты от двух конкурирующих партий"),  $N = \{1, 2\}$ . Стратегиями каждого игрока являются альтернативы из  $X$  ("предвыборные обещания"), то есть множества стратегий игроков совпадают с предъявлением задачи выбора. Игроки одновременно и независимо выбирают одну из альтернатив предъявления ("дают предвыборное обещание"). При этом они могут выбрать одну и ту же альтернативу ("дать идентичные обещания"). Функция сравнения воплощает мнение большинства избирателей о сделанных обещаниях (предполагается, что у коллектива избирателей есть способ заставить избранного политика выполнить обещанное). В электо́ральном соревновании побеждает политик, чьё обещание предпочтительнее для большинства избирателей. Если предложенные альтернативы равноценны для общества, то выбор победителя определяется случайным образом. Такая ситуация хуже, чем гарантированная победа, но лучше, чем гарантированное поражение. Также предполагается, что политиков интересует только победа на выборах, и они готовы дать любое обещание, приносящее электо́ральный успех. Соответственно, в такой игре выигрыш игрока 1, если он выбирает  $x$ , а игрок 2 выбирает  $y$ , равен  $g(x, y)$ , а выигрыш игрока 2, если он выбирает  $y$ , а

игрок 1 выбирает  $x$ , равен  $-g(x, y)$ . Таким образом,  $G$  является *игрой с нулевой суммой*,  $\forall x, y \in X, g(x, y) + g(y, x) = 0$ .

Определённая выше игра  $G(X, \succ)$ , называется *турнирной*. Известно, что в любой турнирной игре есть равновесие по Нэшу  $p$  в смешанных стратегиях, также называемое *максимальной лотерей*, причём только одно [40].

*Двухпартийное множество BP* определяется как носитель максимальной лотереи  $p$  в турнирной игре  $G(X, \succ)$ ,  $BP(X, \succ) = \{x \in X | p(x) > 0 \wedge p = NE(G(X, \succ))\}$ , то есть как множество таких альтернатив, для которых вероятность быть выбранными не равна нулю.

Было доказано, что двухпартийное множество является монотонным, устойчивым и композиционно согласованным турнирным решением, что делает его привлекательным инструментом выбора в турнире [31].

#### 1.4.2.7 Существенное множество ( $ES$ )

В [41] было введено существенное множество  $ES$  – естественное обобщение двухпартийного множества на случай, когда турнир может быть слабым. В турнирной игре, соответствующей слабому турниру, равновесие по Нэшу в смешанных стратегиях существует, но может не быть единственным. Следует отметить, что если существует равновесие Нэша в смешанных стратегиях вида  $(p_1, p_2)$ , где  $p_1$  и  $p_2$  – это разные распределения вероятностей выбора альтернатив из предъявления, то равновесными также являются профили стратегий  $(p_1, p_1), (p_2, p_1), (p_2, p_2)$ . Поэтому можно рассматривать лишь симметричные равновесия вида  $(p_1, p_1), (p_2, p_2)$ . Подразумевая этот факт, далее для простоты будем говорить, что равновесными по Нэшу являются распределения  $p_1, p_2$ .

Итак, из всех возможных распределений вероятностей на  $X$  выбираются такие, которые являются равновесиями по Нэшу в смешанных стратегиях  $NE$  в турнирной игре  $G(X, \gtrless)$ . Альтернатива, выбирающаяся с положительной вероятностью хотя бы в одном равновесии Нэша, называется *существенной*. Множество  $ES$  существенных альтернатив называется *существенным множеством*,

$$ES(X, \geq) = \{x \in X \mid p(x) > 0 \wedge p = NE(G(X, \geq))\}.$$

Чтобы определять, входит ли  $x$  в  $ES$  или нет, удобно использовать следующее обобщение функции сравнения  $g$ . Пусть  $p$  – произвольное распределение вероятностей на  $X$ . Положим  $g(x, p) = \sum_{y \in X} g(x, y) \cdot p(y)$ . Для  $g(x, p)$  справедливы следующие утверждения:

- $p = NE(G(X, \geq)) \Leftrightarrow \forall x \in X, g(x, p) \leq 0$ ;
- если  $x \in ES(X, \geq)$ , то  $\forall p = NE(G(X, \geq)), g(x, p) = 0$ ;
- если  $x \notin ES(X, \geq)$ , то  $\exists p = NE(G(X, \geq)): g(x, p) < 0$ .

#### 1.4.2.8 Другие турнирные решения и их обобщения

Для сравнения рассмотрим иные турнирные решения и некоторые их обобщения на случай слабых турниров.

Оптимальный выбор можно определить как выбор альтернатив из «хорошего» подмножества предъявления. Сформулируем ряд определений «хорошего» множества. Множество  $Y$  называется *доминирующими*, если для каждой альтернативы из предъявления в  $Y$  найдется хотя бы одна строго предпочтительная ей альтернатива,  $\forall x \in X, \exists y \in Y: y > x$  [42]. Множество  $Y$  называется *самозащищающимся*, если оно удовлетворяет условию: если альтернатива  $x \in X$  строго предпочтительнее альтернативы  $y \in Y$ , то в  $Y$  найдется другая альтернатива  $z$ , являющаяся строго лучше, чем  $x$ ,  $\forall x \in X, (\exists y \in Y: y > x) \vee (\forall y \in Y, y \geq x)$  [43]. Множество  $Y$  называется *слабоустойчивым*, если оно удовлетворяет такому условию: если альтернатива  $x$ , не лежащая в  $Y$ , строго предпочтительнее альтернативы  $y$  из  $Y$ , то в  $Y$  найдется другая альтернатива  $z$ , являющаяся строго лучше, чем  $x$ ,  $\forall x \in X \setminus Y, (\exists y \in Y: y > x) \vee (\forall y \in Y, y \geq x)$  [44]. Множество  $Y$  называется *внешне устойчивым*, если любая альтернатива  $x$  не из множества  $Y$  является строго хуже, чем хотя бы одна альтернатива из множества  $Y$ ,  $\forall x \in X \setminus Y, \exists y \in Y: y > x$  [45]. Множество  $Y$  называется *слабо внешне устойчивым*, если любая альтернатива  $x$  не из множества  $Y$  является строго хуже, чем хотя бы одна альтернатива из множества  $Y$ , или равноцenna ей,  $\forall x \in X \setminus Y, \exists y \in Y: y \geq x$  [38]. Множество  $Y$  называется

*доминантным*, если любая альтернатива из этого множества является строго предпочтительнее любой альтернативы, в нем не лежащей,  $\forall x \in X \setminus Y, \forall x \in Y, x > y$  [46]. Множество  $Y$  называется *слабо доминантным*, если любая альтернатива из этого множества является слабо предпочтительнее любой альтернативы, в нем не лежащей,  $\forall x \in X \setminus Y, \forall x \in Y, x \gtrsim y$  [46]

Множество называется *минимальным* относительно некоторого свойства, если само множество имеет данное свойство, но ни одно его строгое подмножество его не имеет.

Мы можем определить ряд турнирных решений как объединение минимальных подмножеств предъявления, обладающих определенным «хорошим» свойством: *объединение минимальных доминирующих множеств D*, *объединение минимальных самозащищающихся множеств SP*, *объединение минимальных слабоустойчивых множеств WS*, *объединение минимальных внешне устойчивых множеств MES*, *объединение минимальных слабо внешне устойчивых множеств MRES*. Объединения минимальных доминантных и минимальных слабо доминантных множеств в литературе называются *слабым и сильным максимальным циклом (WTC и STC)*, соответственно. Заметим, что минимальное доминантное множество является единственным для любого (слабого) турнира. В случае, если ядро турнира непусто, объединение минимальных доминирующих множеств  $D$  по определению отождествляется с ядром (множеством максимальных элементов турнира).

Говорят, что альтернатива  $x$  *запирает* альтернативу  $y$ , если  $x$  лучше, чем  $y$ , и не существует пути из  $y$  в  $x$ . Альтернатива  $x$  *захватывает* альтернативу  $y$ , если  $x$  не только покрывает  $y$  по версии IVa, но и покрывает все альтернативы из нижнего среза  $y$  по версии IIIa отношения покрытия.

*Незапертым множеством UT* называют множество всех незапертых альтернатив [47]. *Незахваченным множеством UC<sub>p</sub>* называют множество всех незахваченных альтернатив [47].

Отметим, что если турнир является полным, то  $D$  совпадает с  $SP$ ,  $MRES$  и  $WS$  совпадают с  $MES$ ,  $STC$  и  $UT$  совпадают с  $WTC$ .

## 1.4.3 Свойства существенного множества и его связь с другими решениями

### 1.4.3.1 Свойства существенного множества

В [41] было доказано, что существенное множество является монотонным и удовлетворяет свойству  $\alpha$ . Следовательно, для проверки устойчивости процедуры достаточно проверить, удовлетворяет ли оно свойству  $\gamma$ , гарантирующему независимость выбора от добавления неоптимальных альтернатив в предъявление. Формально, наличие у функции выбора  $ES(Y, \succsim_Y)$  свойства  $\gamma$ , означает, что для любых двух предъявлений  $Y, Z \subseteq A$  таких, что  $Y \cap Z \neq \emptyset$  верно  $(ES(Y, \succsim_Y) = ES(Z, \succsim_Z) = X) \Rightarrow (X = ES(Y \cup Z, \succsim_{Y \cup Z}))$ . Результат проверки сформулирован в виде теоремы.

Теорема 1.12. Существенное множество удовлетворяет свойству  $\gamma$ .

Следствие 1.1. Из того, что существенное множество удовлетворяет как свойству  $\alpha$ , так и свойству  $\gamma$ , следует, что рассматриваемое турнирное решение является устойчивым.

Следствие 1.2. Турнирное решение  $S$  называется *независимым от выбора между неоптимальными вариантами*, если для любых двух турниров  $(X, \succsim)$  и  $(X, \succsim')$  таких, что  $\forall x \in S(X, \succsim), \forall y \in X, x \succsim y \Rightarrow x \succsim' y$ , имеет место  $S(X, \succsim) = S(X, \succsim')$ . Из доказательства Теоремы 1.12 непосредственно следует, что существенное множество независимо от выбора между неоптимальными вариантами, что согласуется с аналогичным результатом в [41].

Третьим свойством турнирных решений, указанным в пункте 1.4.2, является композиционная согласованность. Наличие данного свойства у существенного множества было доказано в [48]. Зная это, можно повысить эффективность вычисления существенное множество, которое, как было показано в [49], является полиномиальной проблемой.

Устойчивость, монотонность, композиционная согласованность и вычислительная простота рассматриваемого турнирного решения дают основание считать его разумным и удобным инструментом выбора и ранжирования на основе информации о парных сравнениях альтернатив.

#### 1.4.3.1 Консервативное обобщение

Естественным обобщением любого турнирных решения  $S$  на случай слабого турнира является так называемое консервативное обобщение, введенное в [32], и обозначающееся  $[S]$ . Рассмотрим множество ориентаций  $[W]$  слабого турнира  $W(X, \gtrsim)$ , то есть множество всех турниров  $T(X, >)$ , которые можно получить произвольным дополнением слабого турнира  $W(X, \gtrsim)$  до полного. Консервативным обобщением  $[S]$  турнирного решения  $S$  называется множество альтернатив, выбираемых этим решением из всех турниров, входящих в множество ориентаций  $[W]$  слабого турнира  $W$ ,

$$[S](W) = \bigcup_{T \in [W]} S(T).$$

В [32] показано, что существенное множество не совпадает с консервативным обобщением двухпартийного множества. В указанной статье были приведены примеры, показывающие, что  $ES \notin [BP] \wedge [BP] \notin ES$ . Гипотеза о том, что пересечение  $ES$  и  $[BP]$  всегда непусто, до сих пор не доказана.

#### 1.4.3.3 Теоретико-множественные отношения с другими турнирными решениями

Рассмотрим теоретико-множественные отношения консервативного обобщения двухпартийного множества и существенного множества с приведенными в данной работе обобщениями турнирных решений на случай слабых турниров.

Теорема 1.13.

- 1)  $ES$  всегда является подмножеством  $UC_{IVa}$ ,  $UC_D$ ,  $[D]$ ,  $[MES]$ ,  $UC_p$ ,  $WTC$ .  $ES$  не включено в  $UC_{IIb}$ ,  $UC_{IIa}$ ,  $UC_{IIIb}$ ,  $UC_{IIIa}$ ,  $D$ , причем их пересечение может быть пустым.  $ES$  не включено в  $SP$ ,  $WS$ ,  $MES$ ,  $MRES$ ,  $STC$ ,  $UT$ , но их пересечение всегда непусто.

2)  $[BP]$  всегда является подмножеством  $UC_{Va}$ ,  $[D]$ ,  $[MES]$ ,  $WTC$ .  $[BP]$  не включено в  $UC_{Ib}$ ,  $UC_{IIa}$ ,  $UC_{IIIb}$ ,  $UC_{IIIa}$ ,  $UC_{IVa}$ ,  $D$ ,  $MES$ ,  $MRES$ ,  $UC_p$ ,  $STC$ ,  $UT$ , но их пересечение всегда непусто.

Результаты теоремы приведены в Таблице 1.7.

Таблица 1.7 – Теоретико-множественные соотношения между турнирными решениями

	$ES$	$[BP]$	$UC_{Ib}$	$UC_{IIa}$	$UC_{IIIb}$	$UC_{IIIa}$	$UC_{IVa}$	$UC_{Va}$	$D$	$SP$	$[D]$	$WS$	$MES$	$MRES$	$[MES]$	$UC_p$	$STC$	$UT$	$WTC$
$ES$	=	?	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\subset$	$\subset$	$\emptyset$	$\cap$	$\subset$	$\cap$	$\cap$	$\cap$	$\subset$	$\subset$	$\cap$	$\cap$	$\subset$
$[BP]$	?	=	$\cap$	$\cap$	$\cap$	$\cap$	$\cap$	$\subset$	$\cap$	?	$\subset$	?	$\cap$	$\cap$	$\subset$	$\cap$	$\cap$	$\cap$	$\subset$

В Таблице 1.7 символ “ $\subset$ ” означает, что турнирное решение, расположенное в строке, всегда содержится в турнирном решении, расположенном в соответствующем столбце. Знаком “ $\emptyset$ ” обозначены независимые турнирные решения, то есть решения, пересечение которых может оказаться пустым, а символ “ $\cap$ ” обозначает турнирные решения, пересечение которых всегда не-пусто. Знаком вопроса “?” обозначены пары решений, для которых пока не найдено доказательства непустоты пересечения или примера, показывающего их независимость. Равенством обозначаются эквивалентные решения.

#### 1.4.4 Эмпирическое сравнение методов ранжирования

С точки зрения теории использовать существенное множество как инструмент выбора - разумно. Однако для оценки практической полезности метода следует также сравнить результаты его применения в какой-либо практической задаче с результатами, которые дают другие процедуры. Как уже было сказано выше, повторяя процедуру выбора, то есть сортируя альтернативы, можно решить и иную проблему – проблему построения полного ранжирования альтернатив. В статье [33], лежащей в основе настоящего раздела, приводится способ сравнения различных процедур ранжирования, основанный на попарном сравнении результатов корреляционного анализа получаемых ранжирований. Практической задачей, к которой применяются сравниваемые процедуры ранжирования, является построение агрегированного ранжирования научных журналов по менеджменту. Агрегируются рейтинги

93 журналов, основанные на семи наиболее популярных библиометрических индикаторах.

Для достижения поставленной цели была создана программа на языке программирования Python. В ее функционал входит расчет турнирных решений, описанных в настоящей работе, построение на их основе ранжирований, расчет коэффициента ранговой корреляции  $\tau_b$ -Кендалла, построение с его помощью матрицы строгого мажоритарного отношения, построение ранжирования с помощью второй версии правила Коупланда, визуализация, сохранение и скачивание результатов в подходящем формате. Заметим, что эта же программа применялась для построения всех приведенных в работе контрпримеров.

Подавляющее большинство турнирных решений вычисляется программой за полиномиальное время. К сожалению, не для всех решений найден полиномиальный алгоритм вычисления, так нахождение объединения минимальных самозащищающихся множеств или объединения минимальных слабоустойчивых множеств является NP-трудной проблемой.

В данной работе с помощью созданного комплекса программ методом сортировки с использованием существенного множества строится новое агрегированное ранжирование журналов по менеджменту, упомянутых ранее. Другие ранжирования, агрегируемые и агрегированные, к которым применен анализ, описанный ниже, взяты из [33].

#### 1.4.4.1 Корреляционный анализ и сравнение ранжирований

В корреляционном анализе используется коэффициент ранговой корреляции  $\tau_b$ -Кендалла.

Ранжирования можно сравнить попарно и упорядочить следующим образом. Будем считать, как предлагают авторы статьи [33], что одно ранжирование  $I$  репрезентирует множество агрегируемых рейтингов лучше ранжирования  $II$ , если ранжирование  $I$  коррелирует с большинством из агрегируемых рейтингов сильнее, чем ранжирование  $II$ . Будем считать, что ранжирование  $I$  коррелирует с ранжированием  $III$  сильнее, чем ранжирование  $II$ , если значе-

ние  $\tau_b$  для пары  $\{I, III\}$  больше значения  $\tau_b$  для пары  $\{II, III\}$ . Попарное сравнение определяет на множестве ранжирований мажоритарное отношение. Матрица полученного мажоритарного отношения приведена в Таблице 1.8.

Таблица 1.8 – Матрица мажоритарного отношения

	impact factor	5-year impact factor	immediacy index	article influence	Hirsch index	SNIP	SJR	Copeland rule (2 v.)	Copeland rule (3 v.)	sorting by UC	sorting by MES	Markovian method	sorting by ES	Copeland score
impact factor	0	0	1	0	0	0	0	0	0	0	0	0	0	1
5-year impact factor	1	0	1	1	1	1	1	0	0	0	0	0	0	6
immediacy index	0	0	0	0	0	0	0	0	0	0	0	0	0	0
article influence	1	0	1	0	0	0	1	0	0	0	0	0	0	3
Hirsch index	1	0	1	1	0	0	1	0	0	0	0	0	0	4
SNIP	1	0	1	1	1	0	1	0	0	0	0	0	0	5
SJR	1	0	1	0	0	0	0	0	0	0	0	0	0	2
Copeland rule (2 v.)	1	1	1	1	1	1	1	0	1	0	0	1	0	9
Copeland rule (3 v.)	1	1	1	1	1	1	1	0	0	0	0	1	0	8
sorting by UC	1	1	1	1	1	1	1	1	1	0	0	1	0	10
sorting by MES	1	1	1	1	1	1	1	1	1	1	0	1	0	11
Markovian method	1	1	1	1	1	1	1	0	0	0	0	0	0	7
sorting by ES	1	1	1	1	1	1	1	1	1	1	1	1	0	12

Для ранжирования ранжирований применим к полученному мажоритарному отношению вторую версию правила Коупланда. В соответствии с этим правилом альтернатива  $x$  будет находиться в ранжировании выше, чем  $y$ , если  $x$  лучше (по мажоритарному отношению) большего числа альтернатив, чем  $y$ . Более формально, правило Коупланда ранжирует альтернативы, расставляя их в порядке убывания значений оценок Коупланда  $s_2$ , где  $s_2(x) = |L(x)|$  (по второй версии). Значения оценок Коупланда приведены в последнем столбце Таблицы 1.8, а полученное на их основе ранжирование – в Таблице 1.9.

Таблица 1.9 – Ранжирование ранжирований

Ранг	Метод
1	Сортировка с помощью ES
2	Сортировка с помощью MES
3	Сортировка с помощью UC
4	Правило Коупланда (2 версия)
5	Правило Коупланда (3 версия)
6	Марковский метод
7	5-летний импакт-фактор

## Продолжение таблицы 1.9

Ранг	Метод
8	SNIP
9	Индекс Хирша
10	Индекс влияния статьи
11	SJR
12	Импакт-фактор (классический)
13	Индекс оперативности

На основании результатов сравнения можно сказать, что существенное множество представляет исходное множество агрегируемых ранжирований лучше, чем любое другое рассматриваемое ранжирование. Оно, как и непокрытое множество или  $MES$ , разбивает журналы на довольно крупные группы. Однако стоит отметить, что количество позиций, выделенных с помощью существенного множества, в обоих случаях оказалось всё же больше, чем число рангов, выделенных сортировкой непокрытым множеством и объединением минимальных внешне устойчивых множеств. Этого и следовало ожидать, исходя из теоретико-множественных отношений данных турнирных решений и существенного множества.

### 1.4.5 Некоторые выводы

Дано содержательное описание существенного множества – естественного обобщения такой важной и интересной функции выбора наилучших альтернатив в турнире, как двухпартийное множество. Описаны важные свойства турнирных решений (монотонность, устойчивость, композиционная согласованность и вычислительная простота), которые делают двухпартийное множество привлекательным инструментом выбора. Показано, что все эти свойства сохраняются у рассматриваемого обобщения, причём наличие одного из них было доказано впервые. Можно отметить, что турнирные решения, у которых есть все эти свойства, весьма немногочисленны.

Была исследована связь существенного множества и другого обобщения двухпартийного множества, называемого консервативным, с версиями непокрытого множества и иными турнирными решениями. Отмечено, что включённость существенного множества в версии  $IVa$  и  $Va$  [41] непокрытого множества дает возможность сократить размерность матрицы турнира, об-

легчая вычисление существенного множества. Это обстоятельство было использовано при написании программы, осуществляющей сортировку с помощью существенного множества.

В ходе исследования также была рассмотрена практическая задача, связанная с управлением в академической сфере – агрегирование противоречивых рейтингов научных журналов (по менеджменту), основанных на различных библиометрических показателях. Корреляционный анализ полученного агрегата показал, что ранжирование журналов, построенное с помощью сортировки, основанной на отборе лучших альтернатив посредством существенного множества, представляет исходную совокупность агрегируемых рейтингов лучше, чем каждый из них, и лучше, чем агрегированные ранжирования, построенные с помощью иных процедур.

Также в ходе написания настоящей работы была создана программа на языке программирования Python. В ней были реализованы полиномиальные алгоритмы для вычисления всех турнирных решений, рассматриваемых в работе, вычислительная простота которых доказана. В частности, полиномиальный алгоритм для вычисления существенного множества был взят из [49]. Разработанное решение позволило оптимизировать ряд рутинных исследовательских задач, таких как расчет турнирных решений, построение с их помощью ранжирований, проведение корреляционного анализа и поиск контрпримеров.

Можно утверждать, что текущими научно-практическими задачами становятся, во-первых, разработка полноценного десктопного приложения на основе имеющегося программного кода с целью его использования и продвижения в академических кругах, а во-вторых, расширение и обновление библиометрических данных о научных журналах разных категорий для формирования лучшего понимания принципов работы существенного множества.

## **1.5 Исследование влияния неполноты информации и неопределенности на манипулируемость в теории коллективного выбора**

### 1.5.1 Введение

В теории коллективного принятия решений широко известна проблема манипулирования: избиратели, действуя стратегически, могут намеренно исказить заявляемые ими предпочтения с целью изменить результат голосования в свою пользу. Существует большое количество различных процедур коллективного принятия решений, которые на основе множества индивидуальных предпочтений избирателей позволяют получить коллективное решение – выбор нескольких альтернатив или ранжирование. Однако, если предположить, что каждый избиратель знает предпочтения всех остальных, имеется хотя бы три альтернативы, и процедура не является диктаторской, то она оставляет возможность манипулирования для избирателей. Этот феномен формально описан теоремой Гиббарда — Саттертуэйта и её многочисленными расширениями.

Учитывая этот результат, мы приходим к необходимости оценить, насколько подвержены манипулированию различные правила. Широко используется в литературе вероятностный подход, при котором степень манипулируемости рассчитывается как доля профилей предпочтений, в которых возможно манипулирование. В рамках этого подхода по проекту МЦАВР в 2021 году исследование манипулируемости правил агрегирования предпочтений проводилось в двух направлениях: изучение влияния неполноты информации и неопределенности относительно действий других избирателей на манипулируемость; а также разработка алгоритмов для точной оценки сильной и слабой манипулируемости процедур агрегирования.

### 1.5.2 Неполнота информации и одновременное манипулирование

При этом в большинстве исследований предполагается наличие полной информации. Модель манипулирования при неполной информации была предложена недавно в [50], а индексы манипулируемости для случая с неполной информацией исследованы в [51].

Использование предпосылки о наличии полной информации позволяет упростить сравнительный анализ правил, но является достаточно сильным и

нереалистичным предположением. Введение предпосылки о неполноте информации помогает приблизить математическую модель манипулирования при голосовании к реальным ситуациям принятия коллективных решений.

Следующая дополнение к базовой модели заключается в том, что избиратели учитывают возможность манипулирования со стороны других участников голосования. В случае, если существует опасность получить менее предпочтительный исход при одновременном искажении предпочтений другими избирателями, стимул к манипулированию снижается. Таким образом, рассматриваемая модель учитывает как неопределенность, связанную с неполнотой информации, так и риск, возникающий ввиду возможности ответных действий других участников. Анализируется влияние этих факторов на относительную манипулируемость правил.

#### 1.5.2.1 Математическая модель

Пусть  $N = \{1, 2, \dots, n\}$  – множество избирателей, имеющих предпочтения на множестве альтернатив  $X$ ,  $|X| = m$ . Бинарное отношение  $P_i \subset A \times A$  – предпочтения избирателя  $i$ , являющиеся линейным порядком,  $L(X)$  – множество всех линейных порядков. Профиль предпочтений всех избирателей обозначается за  $\vec{P} = (P_1, \dots, P_n)$ , а профиль предпочтений всех избирателей, кроме  $i$  – за  $\vec{P}_{-i}$ . Множество всех профилей предпочтений есть  $L(X)^N$ , включающее  $(m!)^n$  элементов. Отображение  $F: L(X)^N \rightarrow 2^X \setminus \emptyset$  называется правилом коллективного выбора. Если результат правила включает в себя более одной альтернативы, то применяется алфавитное правило устранения множественности выбора,  $T: 2^X \setminus \emptyset \rightarrow X$ . Согласно ему, на множестве  $X$  предполагается заданным некоторый линейный порядок  $P_T$ , например,  $aP_T bP_T c\dots$ , и если по правилу коллективного выбора было выбрано несколько альтернатив, то из них выбирается альтернатива, недоминируемая по  $P_T$ , т.е.  $T(A) = \{a \in A \mid \forall x \in A, x \neq a \ (a, x) \in P_T\}$ . Композиция функций  $F$  и  $T$ , т.е.  $T(F(\vec{P}))$  обозначается как  $V(\vec{P})$ . Пусть  $v(a, \vec{P})$  – вектор распределения рангов

альтернативы  $a$ ,  $j$ -ый элемент которого,  $v_j(a, \vec{P})$ , равен количеству избирателей, в предпочтениях которых альтернатива  $a$  стоит на  $j$ -ом месте.

Обозначим мажоритарное отношение:  $a_k \mu a_l$ , если  $|\{i \in N : a_k P_i a_l\}| > |\{i \in N : a_l P_i a_k\}|$ , а матрицу мажоритарного отношения – за  $MG(\vec{P})$ , где

$$MG(\vec{P})_{kl} = \begin{cases} 1, & \text{если } a_k P_M a_l, \\ -1, & \text{если } a_l P_M a_k. \\ 0, & \text{иначе.} \end{cases}$$

Пусть  $r_i(a, P_i) = |\{b \in A | a P_i b\}|$  – ранг альтернативы  $a$  в предпочтениях  $P_i$ .

### 1.5.2.2 Моделирование неполной информации

Пусть на первом этапе проводится предварительный опрос всех избирателей (предполагается, что на опросе они заявляют искренние предпочтения), и его результаты в агрегированном виде затем становятся публичной информацией. В математической модели результат опроса представлен с помощью функции публичной информации (ФПИ)  $\pi$ , предложенной в [50]. На втором этапе проводится голосование, и избиратели, располагая информацией о результатах предварительного опроса, принимают решение о том, какие предпочтения сообщать процедуре выбора. Профиль искренних предпочтений неизвестен избирателям, однако для каждого из них имеется множество возможных профилей предпочтений, согласующихся с полученной избирателем информацией – информационное множество.

В представленном исследовании использовалось проводилось сравнение индексов манипулируемости для случая полной информации и ФПИ, вычисляющей победителя после устранения множественности выбора, т. е.  $V(\vec{P})$ .

Пусть  $W_i^{\pi(\vec{P})}$  – информационное множество избирателя  $i$ , множество всех профилей предпочтений, согласующихся с имеющейся у избирателя  $i$  информацией  $\pi(\vec{P})$ . Формально

$$W_i^{\pi(\vec{P})} = \{\vec{P}'_{-i} \in L(X)^{N \setminus \{i\}} : \pi(P_i, \vec{P}'_{-i}) = \pi(\vec{P})\}.$$

### 1.5.2.3 Модели манипулирования

#### 1.5.2.3.1 Модель 1

**Определение 1.1.** Избиратель  $i$  имеет стимул манипулировать в Модели 1 при ФПИ  $\pi$  в профиле  $\vec{P}$  при правиле коллективного выбора  $F$ , если существует  $\tilde{P}_i$  такое, что 1) либо  $V(\tilde{P}_i, \vec{P}'_{-i}) = V(\vec{P})$ , либо  $V(\tilde{P}_i, \vec{P}'_{-i}) P_i V(\vec{P})$  для всех  $\vec{P}'_{-i} \in W_i^{\pi(\vec{P})}$ ; 2) существует  $\vec{P}''_{-i} \in W_i^{\pi(\vec{P})}$  такой, что  $V(\tilde{P}_i, \vec{P}''_{-i}) P_i V(\vec{P})$ .

Иными словами, для того, чтобы у избирателя был стимул манипулировать, нужно, чтобы при некоторой его стратегии результат во всех профилях информационного множества не стал для него хуже, и существуют профили, где результат станет лучше.

Если в профиле  $\vec{P}$  существует хотя бы один избиратель, имеющий стимул манипулировать при информации  $\pi$  и правиле  $F$ , то профиль называется манипулируемым при ФПИ  $\pi$ . Множество всех избирателей, которые имеют стимул манипулировать в  $\vec{P}$  при  $\pi$ , обозначается за  $\Pi(\vec{P})$ .

В определении 1.1, предложенном в [50], избирателем не учитываются возможные действия остальных участников голосования. Этую модель манипулирования, Модель 1, мы назовем «наивной моделью». В противоположность ей рассмотрим Модель 2, в которой избиратель проверяет, будет ли стратегия, с которой он имеет стимул манипулировать в Модели 1, более предпочтительна, чем истинные предпочтения, если все остальные избиратели из  $\Pi(\vec{P})$  решают манипулировать.

#### 1.5.2.3.2 Модель 2

В общем случае у избирателя, имеющего стимул манипулировать, может быть несколько стратегий манипулирования. Для простоты мы фиксиру-

ем одну стратегию для каждого избирателя. В противном случае нам бы пришлось рассматривать все комбинации стратегий для манипулирующих избирателей. Это было бы не только сложным с точки зрения модели и вычислений, но и имело бы небольшой эффект, так как в случае трех альтернатив различные стратегии манипулирования чаще всего не отличаются с точки зрения результата.

Обозначим за  $\tilde{\vec{P}}$  профиль предпочтений, полученный из  $\vec{P}$ , в котором все избиратели из  $\Pi(\vec{P})$  манипулируют.

**Определение 1.2.** Избиратель  $i$  имеет стимул манипулировать в Модели 2 при ФПИ  $\pi$  в профиле  $\vec{P}$  при правиле коллективного выбора  $F$ , если существует  $\tilde{P}_i$  такое, что 1) избиратель  $i$  имеет стимул манипулировать в Модели 1 при  $\pi$  в  $\vec{P}$  и при правиле  $F$ ; 2) либо  $V(\tilde{P}_i, \tilde{\vec{P}}_{-i}) = V(P_i, \tilde{\vec{P}}_{-i})$ , либо  $V(\tilde{P}_i, \tilde{\vec{P}}_{-i}) P_i V(P_i, \tilde{\vec{P}}_{-i})$  для всех  $\tilde{P}_{-i}' \in W_i^{\pi(\vec{P})}$ .

#### 1.2.2.4 Правила коллективного выбора

1) **Правило относительного большинства.** Выбор по данному правилу определяется как множество альтернатив, являющихся наилучшими для наибольшего числа избирателей, т.е.

$$c \in F(\vec{P}) \Leftrightarrow c \in \arg \max_{a \in X} \left( \langle s_{Pl}, v(a, \vec{P}) \rangle \right),$$

где  $s_{Pl} = (1, 0, \dots, 0)$ , а  $\langle s_{Pl}, v(a, \vec{P}) \rangle$  – скалярное произведение.

2) **Правило вето (обратное правило простого большинства).** Выбирается альтернатива, занявшая наименьшее число последних мест в предпочтениях избирателей.

3) **Правило Борда.** По данному правилу для каждой альтернативы  $a$  вычисляется ранг Борда  $r(a, \vec{P})$  как сумма рангов  $r_i(a, P_i)$  по всем избирателям. Или, с помощью вектора очков  $s_B = (m-1, m-2, \dots, 1, 0)$ ,

$$c \in F(\vec{P}) \Leftrightarrow c \in \arg \max_{a \in X} \left( \langle s_B, v(a, \vec{P}) \rangle \right).$$

4) Двухступенчатая мажоритарная система. Выбирается альтернатива, являющаяся наилучшей больше, чем для половины избирателей. Если такой альтернативы нет, то проводится второй тур голосования, в котором участвуют две альтернативы, набравшие наибольшее количество голосов в первом туре.

5) Правило передачи голосов (Процедура Хара) Выбирается альтернатива, имеющая более 50% первых мест в предпочтениях избирателей. Если таковой нет, то из профиля исключается альтернатива, имеющая наименьшее число первых мест. Процедура повторяется до тех пор, пока не будет выбрана какая-либо альтернатива.

6) Правило Коупленда. Выбирается альтернатива, для которой количество побед минус количество поражений по мажоритарному отношению наибольшее, т.е.

$$CS(a, \vec{P}) = |\{b \in X \mid a \mu b\}| - |\{b \in X \mid b \mu a\}|,$$

$$c \in F(\vec{P}) \Leftrightarrow c \in \arg \max_{a \in X} CS(a, \vec{P}).$$

7) Максиминная процедура. Для каждой альтернативы  $a$  вычисляется минимальное количество избирателей, которые предпочитают эту альтернативу некоторой другой, а затем выбирается альтернатива, для которой это значение максимально

$$c \in F(\vec{P}) \Leftrightarrow c \in \arg \max_{a \in X} \min_{b \in X} |\{i \in N : a P_i b\}|.$$

8) Процедура Болдуина (обратная процедура Борда). Для каждой альтернативы подсчитывается ранг Борда  $r(a, \vec{P})$ , затем из профиля удаляется альтернатива, имеющая наименьший ранг Борда. Процедура исключения альтернатив продолжается до тех пор, пока не останутся неисключаемые альтернативы, которые и являются множеством выбора в процедуре Болдуина.

9) Правило Нансона. Рассчитываются значения ранга Борда для каждой альтернативы, затем из профиля удаляются все альтернативы, имеющие ранг Борда ниже среднего ранга,

$$\bar{r} = \frac{1}{m} \sum_{a \in X} r(a, \vec{P}).$$

10) Процедура Блэка. Выбирается победитель Кондорсе – альтернатива, недоминируемая по мажоритарному отношению, т.е.  $a \in F(\vec{P}) \Leftrightarrow \neg \exists b \in A, b \mu a$ , если он существует, в противном случае победитель определяется по правилу Борда.

11) Правило Кемени. Введем сначала функцию расстояния по Кемени между линейными порядками  $P_1$  и  $P_2$ . Пусть  $\{d_{ij}^{(k)}\}$  – матрица бинарного отношения  $P_k$ : если  $a P_k b$ , то элемент этой матрицы  $d_{ab}^{(k)} = 1$ ,  $d_{ba}^{(k)} = -1$ . Расстояние между двумя предпочтениями вычисляется как сумма модулей разности элементов матриц

$$d(P_1, P_2) = \sum_{a \in X} \sum_{b \in X} |d_{ab}^{(1)} - d_{ab}^{(2)}|.$$

Результат правила Кемени – линейный порядок, для которого сумма расстояний до предпочтений всех избирателей минимальна, т.е.

$$P_K = \arg \min_{P \in L(X)} \sum_{i \in N} d(P, P_i).$$

Победитель – кандидат, занимающий первое место в найденном упорядочении.

12) Правило порогового агрегирования. Альтернативы упорядочиваются на основании сравнения векторов распределения рангов. Строится бинарное отношение  $R$ :  $a R b$ , если  $v_m(a, \vec{P}) < v_m(b, \vec{P})$  или если существует такое  $k \leq m$ , что  $v_i(a, \vec{P}) = v_i(b, \vec{P})$ ,  $i = k-1, \dots, m$  и  $v_k(a, \vec{P}) < v_k(b, \vec{P})$ . Альтернативы, недоминируемые в  $R$ , являются победителями:  $a \in F(\vec{P}) \Leftrightarrow \neg \exists b \in A, b R a$ .

### 1.5.2.5 Результаты

Для сравнения степени манипулируемости правил в случае неполной информации существует несколько различных индексов [51]. В данной работе использовался самый простой индекс – доля профилей предпочтений, в которых хотя бы один избиратель имеет стимул манипулировать при заданной ФПИ.

Пусть  $I^M(m, n, \pi, F)$  – доля профилей предпочтений, в которых хотя бы один избиратель имеет стимул манипулировать в модели  $M$  при ФПИ  $\pi$  и правиле  $F$ .

Точные значения индексов  $I^{M^1}(m, n, \pi, F)$  и  $I^{M^2}(m, n, \pi, F)$  были вычислены в MATLAB для  $m=3$  в случае полной информации, т.е. *Profile*-ФПИ,  $\pi(\vec{P}) = \vec{P}$ , и для информации о единственном победителе, т.е. 1 Winner-ФПИ,  $\pi(\vec{P}) = V(\vec{P})$ .

Сравнивая значения  $I^{M^1}(m, n, Profile, F)$  и  $I^{M^1}(m, n, 1Winner, F)$ , мы изучаем влияние неполноты информации на манипулируемость. Сравнение  $I^{M^1}(m, n, Profile, F)$  и  $I^{M^2}(m, n, Profile, F)$ , позволяет оценить влияние неопределенности относительно действий других избирателей на вероятность манипулирования. В значениях индекса  $I^{M^2}(m, n, 1Winner, F)$  учтены оба типа неопределенности.

Рассмотрим сначала манипулируемость для Модели 1 в условиях полной информации (рисунок 1.18). Как видно из графиков, все вычисленные значения не превышают 0,4. Для большинства правил они постепенно снижаются, а среди наименее манипулируемых – правило передачи голосов, двухступенчатая мажоритарная система и правило Болдуина.

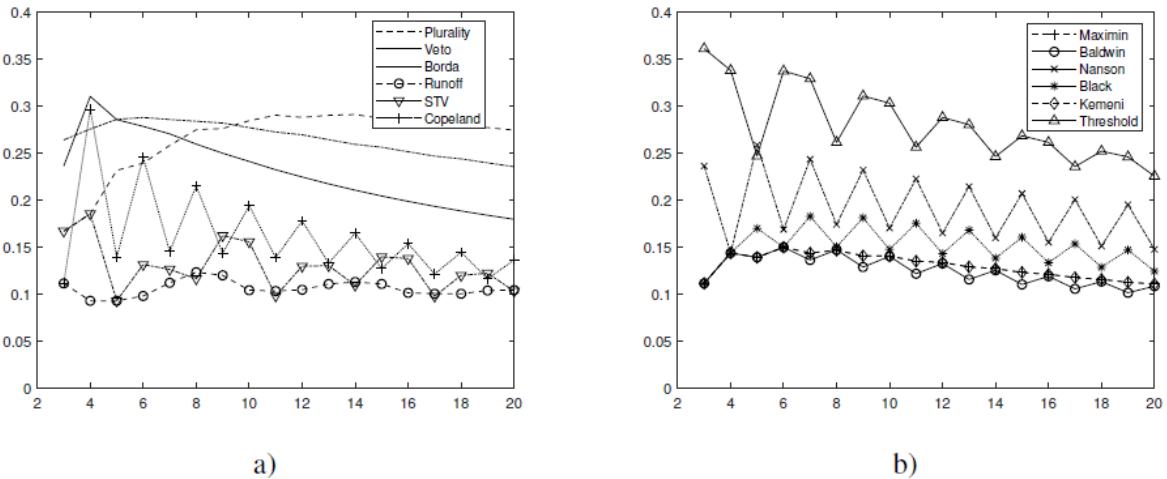


Рисунок 1.18 – Значения индекса  $I^{M^1}(3,n,Profile,F)$

Для функции информации о единственном победителе различия в индексе манипулируемости для правил практически исчезают, так как  $I^{M^1}(3,n,1Winner,F)$  для всех правил, кроме Коупленда, приближается к 1 с ростом числа избирателей (см. рисунок 1.19). Для правила Коупленда наблюдается периодичность значений, хотя пики  $I^{M^1}(3,n,1Winner,Copeland)$  приближаются к 1.

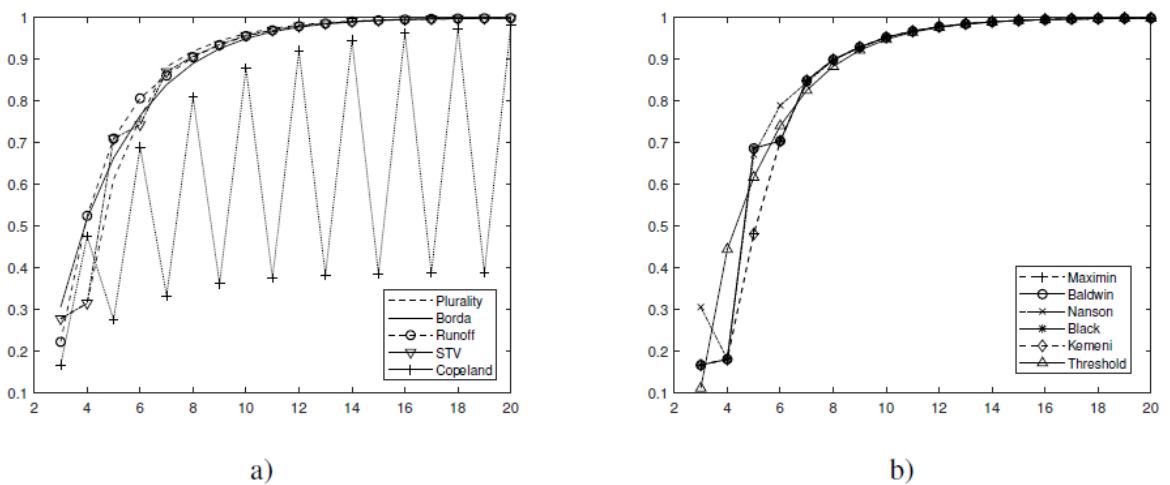


Рисунок 1.19 – Значения индекса  $I^{M^1}(3,n,1Winner,F)$

Объясним увеличение значений индекса. Согласно определению манипулирования при неполной информации, применение стратегии  $\tilde{P}_i$  избирате-

лем  $i$  в некоторых профилях информационного множества приведет к более предпочтительному результату, в остальных же результат не изменится. Так как избиратель не имеет возможности различать профили из одного информационного множества, то избиратель будет иметь стимул манипулировать также и в тех профилях, где результат фактически не изменится. Поэтому индекс манипулируемости становится выше несмотря на то, что избиратель располагает меньшей информацией. Более подробно эта проблема раскрыта в [51].

Теперь перейдем к Модели 2. Так как значения индексов  $I^{M^1}(3, n, Profile, F)$  и  $I^{M^2}(3, n, Profile, F)$  отличаются незначительно, мы приведем здесь не графики, а точные значения разности  $I^{M^1}(3, n, Profile, F) - I^{M^2}(3, n, Profile, F)$  (Таблица 1.10). Они иллюстрируют, в какой доле профилей предпочтений, манипулируемых «наивно» (в Модели 1), угроза получить менее предпочтительный результат из-за одновременного манипулирования других избирателей, приведет к исчезновению стимула манипулировать. Как видно из Таблицы 1.10, доля таких профилей невелика. Если исключить правило вето и пороговое правило, эта доля не превышает 0,07 для остальных процедур. Это наблюдение демонстрирует, что опасность проиграть из-за действий других избирателей сама по себе не может рассматриваться как серьезное препятствие к стратегическому поведению, в силу того что это достаточное редкое явление.

Таблица 1.10 – Значения  $I^{M^1}(3, n, Profile, F) - I^{M^2}(3, n, Profile, F)$

1 Winner	3	5	7	9	11	13	15	17	19	20
Plurality	0,06	0,03	0,03	0,03	0,02	0,02	0,01	0,01	0,01	0,01
Veto	0,06	0,08	0,09	0,08	0,07	0,07	0,06	0,05	0,05	0,04
Borda	0,00	0,03	0,02	0,01	0,01	0,01	0,01	0,00	0,00	0,00
Run-off	0,00	0,00	0,00	0,01	0,00	0,01	0,00	0,00	0,00	0,00
STV	0,06	0,00	0,01	0,03	0,00	0,00	0,01	0,00	0,00	0,00
Copeland	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Maximin	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Baldwin	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Nanson	0,00	0,07	0,06	0,04	0,03	0,02	0,01	0,01	0,01	0,01
Black	0,00	0,02	0,03	0,03	0,04	0,03	0,03	0,03	0,02	0,01
Kemeny	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Threshold	0,17	0,02	0,05	0,04	0,07	0,05	0,04	0,05	0,03	0,04

Рассмотрим теперь влияние неполноты информации и неопределенности относительно действий других избирателей в комбинации, т.е. индекс  $I^{M^2}(m, n, 1\text{Winner}, F)$  (Таблица 1.11). Сочетание двух типов неопределенности сильно снижает манипулируемость, так как в информационном множестве появляются ситуации, в которых возможны потери. Достаточно одного только профиля, где с учетом действий других избирателей для того, кто принимает решение о манипулировании, выгоднее не манипулировать, чтобы все профили его информационного множества стали неманипулируемыми этим избирателем. Поэтому комбинация неполноты информации и одновременности манипулирования резко снижает стимулы и делает некоторые правила вообще не манипулируемыми.

Таблица 1.11 – Значения  $I^{M^2}(m, n, 1\text{Winner}, F)$

1Winner	3	5	7	9	11	13	15	16	17	18	19	20
Plurality	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Veto	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Borda	0,31	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Run-off	0,22	0,71	0,65	0,55	0,25	0,35	0,30	0,00	0,00	0,00	0,00	0,00
STV	0,00	0,71	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Copeland	0,17	0,00	0,00	0,00	0,00	0,00	0,00	0,29	0,00	0,30	0,00	0,31
Maximin	0,17	0,21	0,26	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Baldwin	0,17	0,43	0,17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Nanson	0,31	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Black	0,17	0,43	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kemeny	0,17	0,21	0,26	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Threshold	0,11	0,62	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Для всех правил, кроме Коупленда, выявлен определенный порог в числе агентов, после которого правила становится неманипулируемым в нашей модели. Более детальное изучение этого вопроса, включая поиск неманипулируемых правил для других случаев неполной информации, а также аналитическое решение для общего случая позиционных правил, остается возможной темой будущих исследований.

### 1.5.3 Разработка алгоритмов для точной оценки сильной и слабой манипулируемости процедур агрегирования

### 1.5.3.1 Постановка задачи

В работах [52], [53] было показано, что любая недиктаторская процедура агрегирования является манипулируемой, т.е. существуют ситуации, когда участник пишет в бюллетень неискренние предпочтения, чтобы получить более хороший для себя результат процедуры. Возник вопрос: какие процедуры агрегирования являются наименее манипулируемыми? Для оценки степени манипулируемости чаще всего используется индекс Нитцана-Келли, который был предложен в работах [54], [55].

С тех пор вышел ряд исследований по оценке манипулируемости процедур агрегирования, которые можно разделить на три типа.

1) Исследования, где выводятся формулы для расчёта индексов манипулируемости. Одна из ключевых проблем в том, что процедур агрегирования известно несколько десятков, и для каждой нужно заниматься выводом формулы. Для простых процедур, например, Правила относительного большинства, вывод формулы может быть несложным. Однако есть ряд процедур, например, мажоритарные процедуры ( $k$ -устойчивые множества) и процедуры с несколькими раундами (Процедура Хара), где механизмы процедур таковы, что вывод формулы для них почти невозможен, с практической точки зрения.

2) Исследования, где вычисляются точные индексы для небольшого количества участников с помощью компьютерного моделирования. Например, в [44] исследованы 26 процедур агрегирования с точными расчетами для нескольких участников. Но возникает вопрос: как получить такие же значения для большего количества участников, например, для 50-100?

3) Исследования, где генерируется набор случайных профилей для примерной оценки манипулируемости. Например, в [56], [57] генерируется 1 миллион случайных профилей. Здесь может возникнуть вопрос: если во время примерной оценки манипулируемости возникла небольшая разница в индексах между двумя правилами, можно ли утверждать, что это следствие

меньшей манипулируемости или это может быть следствием примерной модели расчётов?

Мы решали задачу расчёта точных значений индексов манипулируемости для случая 3 альтернатив, от 3 до 100 участников. Разработанные алгоритмы позволили рассчитать не приближенные, а точные значения индекса Нитцана-Келли для 27 известных процедур агрегирования.

### 1.5.3.2 Процедуры агрегирования

Рассматривались 27 процедур агрегирования. В Таблице 1.12 приведено их асимптотическое время расчёта. Реализации и алгоритмы процедур были взяты из [58].

Таблица 1.12 – Процедуры агрегирования и асимптотическое время их расчёта

Процедура агрегирования	Асимптотическое время расчета
Правило относительного большинства	$O(n+m)$
Одобряющее голосование с $q=2$	$O(n+m)$
Правило Борда	$O(n \cdot m)$
Процедура Блэка	$O(n \cdot m^2)$
Обратное правило Борда	$O(n \cdot m^2)$
Пороговое правило	$O(m \cdot (n+m))$
Процедура Хара	$O(n \cdot m^2)$
Обратное правило относительного большинства	$O(n+m)$
Процедура Нансона	$O(n \cdot m^2)$
Процедура Кумбса	$O(n \cdot m^2)$
Минимальное доминирующее множество	$O(m^2(n+2^m))$
Минимальное недоминируемое множество	$O(m^2(n+2^m))$
Непокрытое множество I	$O(n \cdot m^2)$
Непокрытое множество II	$O(n \cdot m^2)$
Правило Ричельсона	$O(n \cdot m^2)$
Минимальное слабоустойчивое множество	$O(m^2(n+2^m))$
Правило Фишберна	$O(n \cdot m^2)$
Правило Коупланда I	$O(n \cdot m^2)$
Правило Коупланда II	$O(n \cdot m^2)$
Правило Коупланда III	$O(n \cdot m^2)$
Процедура Симпсона	$O(n \cdot m^2)$
Процедура Минмакс	$O(n \cdot m^2)$
Сильное $q$ -Паретовское правило простого большинства	$O(m(n+m))$
Сильное $q$ -Паретовское правило относительного большинства	$O(m(n+m))$
1-устойчивое множество	$O(m^2(n+m+2^m))$
2-устойчивое множество	$O(m^2(n+m+2^m))$
3-устойчивое множество	$O(m^2(n+m+2^m))$

### 1.5.3.3 Вероятностные модели профилей

Профиль – это совокупность участников и их предпочтений на множестве альтернатив. Возникает вопрос: какова вероятность каждого конкретного профиля? В литературе известны различные подходы. Самые популярные – это Impartial Culture (IC) и Impartial Anonymous Culture (IAC).

В IC предполагается, что каждый профиль равновероятен. Если у нас 3 альтернативы, то у каждого участника может быть одно из  $3! = 6$  предпочтений на множестве альтернатив. Общее количество различных профилей в IC для  $n$  участников для случая 3 альтернатив будет равно  $3!^n$ .

В IAC предполагается, что каждый профиль равновероятен с учётом анонимности. Количество различных профилей для случая IAC равно  $C_{m!+n-1}^n$ , если у нас случай  $m=3$  альтернатив, то это будет  $C_{n+5}^n$ .

Можно заметить, что общее количество профилей для случая IC слишком большое: уже для случаев более 30 участников осуществить расчёты на компьютере было бы сложно. Поэтому мы разработали алгоритм генерации всех профилей для случая IAC, а затем для каждого профиля пересчитали его вероятность в модели IC. После чего посчитали индекс Нитцана-Келли для случая IC.

### 1.5.3.4 Алгоритмы расчёта и результаты

Прямолинейные алгоритмы расчёта сталкиваются с проблемой большой вычислительной сложности. Предположим, что использовалась бы такая последовательность действий:

- 1) генерация всех профилей;
- 2) для каждого профиля считается искренний результат для каждой процедуры агрегирования;
- 3) для каждого профиля генерируются все возможные попытки манипуляции;
- 4) для каждой попытки манипуляции считается новый результат (при неискренних предпочтениях) для каждой процедуры агрегирования;

- 5) сравниваются искренний и неискренний результаты: если хотя бы одна попытка манипулирования даёт лучший результат, то профиль помечается как манипулируемый при заданной процедуре агрегирования;
- 6) рассчитываются индексы Нитцана-Келли для каждой процедуры агрегирования.

Тогда вычислительная сложность была бы равна количеству профилей, умноженному на количество попыток манипулирования, процедур агрегирования и вычислительной сложности каждой процедуры.

Мы можем уменьшить вычислительную сложность, если разобьём процесс на две стадии:

1) стадия предварительных расчётов:

- генерируем все возможные профили для IAC;
- считаем для каждого профиля результат каждой процедуры агрегирования;

2) стадия расчёта манипулируемости:

- генерируем все возможные попытки манипулирования;
- рассчитываем индекс Нитцана-Келли для IAC и IC (с учетом вероятности профилей).

Главное ускорение такого алгоритма в том, что на второй стадии, когда мы генерируем все возможные попытки манипулирования, нам не потребуется для каждого нового профиля (текущий профиль, где один из участников вписал неискренние предпочтения) пересчитывать результаты процедуры агрегирования, так как они уже были заранее посчитаны на первой стадии.

Такой алгоритм с предварительными расчётами позволяет значительно сократить вычислительную сложность. Как результат, даже случая с 100 участниками голосования был посчитан на одной ЭВМ за несколько дней при переборе всех профилей.

В результате для 27 процедур агрегирования для случаев от 3 до 100 участников были сгенерированы все возможные профили, а также рассчитаны значения индексов Нитцана-Келли для каждой процедуры агрегирования.

#### 1.5.4 Некоторые выводы

В рамках проекта МЦАВР в 2021 году в исследовании манипулируемости правил коллективного выбора были сделаны важные шаги. Впервые в модели манипулирования учитывались одновременно наличие неполной информации у избирателей и влияние на принятие решения о манипулировании избирателем наличия стимулов к манипулированию других участников голосования. В частности, было показано, что по отдельности эти факторы не могут служить серьезным препятствием к манипулированию, тогда как их сочетание резко снижает вероятность манипулирования, а в некоторых случаях приводит к неманипулируемости. Кроме того, были разработаны и реализованы эффективные алгоритмы вычисления точных значений сильной и слабой манипулируемости. По сравнению с предыдущими исследованиями, было сделано значительное сокращение времени вычислений за счет оптимизации алгоритма, благодаря чему впервые появилась возможность получения не приближенных, а точных значений вероятности манипулирования даже для большого количества избирателей.

Результаты исследований отражены в работе [59].

### 1.6 Аксиоматизация функций интенсивности связи в индексах влияния

#### 1.6.1 Введение

В настоящее время индексы влияния являются распространенным способом для изучения избирательных органов и институционального баланса власти в этих органах [60]. Например, в [61] исследуется влияние политических партий в Государственной Думе Российской Федерации за период с 2000 по 2003 год.

В большинстве исследований по индексам влияния отмечается, что основным недостатком известных индексов влияния является тот факт, что ин-

дексы не учитывают предпочтения рассматриваемых участников [60]. В статье [60] вводятся индексы влияния, учитывающие предпочтения участников по формированию коалиций. Индексы влияния определяются при помощи функций интенсивности связи, которые каждому участнику и любой коалиции, в которую входит рассматриваемый участник, ставят в соответствие число, характеризующее силу связи участника с коалицией. В [60] определяется два способа определения функции интенсивности связи – через кардинальные и ординальные индексы. Вводятся 16 функций интенсивности связи, определяемых через кардинальные индексы, поставлена проблема об аксиоматизации рассмотренных функций интенсивности связи.

Аксиоматизация является актуальной задачей на сегодняшний день, так как введение набора свойств, составляющих аксиоматику, помогает однозначно определить рассматриваемую функцию, а также показывает качественное различие функций интенсивности связи между собой.

В ходе исследования была построена аксиоматика для 4 из 16 функций интенсивности связи, определяемых через кардинальные индексы. А также подготовлена к обсуждению аксиоматика для 4 функций интенсивности, основанных на линейных функциях.

### 1.6.2 Основная часть

Множество участников обозначается через  $N := \{1, \dots, n\}, n > 1$ . Коалицией  $w$  называется любое непустое подмножество  $N$ , то есть  $w \subseteq N, w \neq \emptyset$ . Функция интенсивности  $f(i, w)$  определяется как неотрицательное действительное число, характеризующее силу связи участника  $i$  с коалицией  $w, i \in w$ . Индекс влияния  $\alpha(i)$  участника  $i$  определяется как  $\alpha(i) = \frac{\sum_w f(i, w)}{\sum_i \sum_w f(i, w)}$ .

В [60] вводится способа определения функции интенсивности  $f(i, w)$  – через ординальные индексы, кардинальные индексы. В данном исследовании рассматриваются функции интенсивности, определенные через кардинальные индексы.

Считается, что желание участника  $i$  образовать коалицию с участником  $j$  определяется через действительное число  $p_{ij}$ ,  $\sum_j p_{ij} = 1$ ,  $i, j = 1, \dots, n$  [60]. В общем случае  $p_{ij} \neq p_{ji}$ . Согласно определению значения  $p_{ij}$ , его можно рассматривать как вероятность того, что участник  $i$  образует коалицию с участником  $j$ .

В [60] с помощью кардинальных индексов  $p_{ij}$  вводятся следующие функции интенсивности связи:

- средняя интенсивность связи участника  $i$  с остальными участниками коалиции  $w$

$$f^+(i, w) = \frac{\sum_{j \in w} p_{ij}}{|w|};$$

- средняя интенсивность связи остальных участников коалиции  $w$  с участником  $i$

$$f^-(i, w) = \frac{\sum_{j \in w} p_{ji}}{|w|};$$

- средняя интенсивность для участника  $i$

$$f(i, w) = \frac{1}{2}(f^+(i, w) + f^-(i, w));$$

- средняя положительная интенсивность в коалиции  $w$

$$f^+(w) = \frac{\sum_{i \in w} f^+(i, w)}{|w|};$$

- средняя отрицательная интенсивность в коалиции  $w$

$$f^-(w) = \frac{\sum_{i \in w} f^-(i, w)}{|w|};$$

- средняя интенсивность в коалиции  $w$

$$f(w) = \frac{\sum_{i \in w} f(i, w)}{|w|};$$

- минимальная интенсивность связей участника  $i$  с остальными участниками коалиции  $w$

$$f_{min}^+(i, w) = \min_j p_{ij};$$

- максимальная интенсивность связей участника  $i$  с остальными участниками коалиции  $w$

$$f_{max}^+(i, w) = \max_j p_{ij};$$

- максимальное колебание связей участника  $i$

$$f_{mf}(i, w) = \frac{1}{2} (\min_j p_{ij} + \max_j p_{ij});$$

- минимальная интенсивность связей участников коалиции  $w$  с участником  $i$

$$f_{min}^-(i, w) = \min_j p_{ji};$$

- максимальная интенсивность связей участников коалиции  $w$  с участником  $i$

$$f_{max}^-(i, w) = \max_j p_{ji};$$

- S-средняя интенсивность связей участника  $i$  с другими участниками в  $w$

$$f_{sm}^+(i, w) = \frac{1}{|w|} \sqrt[s]{\sum_j p_{ij}^s};$$

- S-средняя интенсивность связей участников в  $w$  с участником  $i$

$$f_{sm}^-(i, w) = \frac{1}{|w|} \sqrt[s]{\sum_j p_{ji}^s};$$

- максиминная интенсивность в коалиции  $w$

$$f_{maxmin}(w) = \max_i \min_j p_{ij};$$

- минимаксная интенсивность в коалиции  $w$

$$f_{minmax}(w) = \min_i \max_j p_{ji};$$

- максимальное колебание в коалиции  $w$

$$f_{mf}(w) = \frac{1}{2}(f_{maxmin}(w) + f_{minmax}(w)).$$

Ниже представлена аксиоматика для индекса  $f^+(i, w) = \frac{\sum_{j \in w} p_{ij}}{|w|}$  [60].

- свойство 1: для любого набора значений  $(p_{i1}, \dots, p_{im})$  существует функция  $f(i, w)$  такая, что  $0 \leq f(i, w) \leq 1$ ,  $f(i, w)$  – непрерывная дифференцируемая функция по каждому аргументу;
- свойство 2: если  $p_{ij} = 0$  для любого  $j$ , то  $f(i, w) = 0$ ;
- свойство 3 (монотонность): значение  $f(i, w)$  возрастает (убывает), если возрастает (убывает) любое значение  $p_{ij}$ ; кроме того,  $\frac{\partial f_i}{\partial p_{ij}} = \mu_i$  для любого  $j$ ,  $\frac{\partial f_i}{\partial p_{lj}} = 0$  для любого  $l \neq j$ .

Теорема 1.5 [60]. Функция интенсивности связи  $f(i, w)$  удовлетворяет свойствам 1, 2, 3 тогда и только тогда, когда  $f(i, w) = \frac{\sum_{j \in w} p_{ij}}{|w|}$ .

В ходе настоящего исследования была построена аксиоматика для следующих функций интенсивности связи:  $f_{min}^+(i, w) = \min_j p_{ij}$ ,  $f_{max}^+(i, w) = \max_j p_{ij}$ ,  $f_{min}^-(i, w) = \min_j p_{ji}$ ,  $f_{max}^-(i, w) = \max_j p_{ji}$ .

Кроме того, подготовлена аксиоматика для функций интенсивностей связи:  $f^-(i, w) = \frac{\sum_{j \in w} p_{ji}}{|w|}$ ,  $f^+(w) = \frac{\sum_{i \in w} f^+(i, w)}{|w|}$ ,  $f^-(w) = \frac{\sum_{i \in w} f^-(i, w)}{|w|}$ ,  $f(w) = \frac{\sum_{i \in w} f(i, w)}{|w|}$ . Однако, полученный результат требует дальнейшего обсуждения.

Аксиоматика для функций интенсивности связи  $f_{min}^+(i, w) = \min_j p_{ij}$ ,  $f_{max}^+(i, w) = \max_j p_{ij}$ ,  $f_{min}^-(i, w) = \min_j p_{ji}$ ,  $f_{max}^-(i, w) = \max_j p_{ji}$ :

- свойство 1:  $f(i, w)$  зависит только от связей участника  $i$  с остальными участниками коалиции  $w$ , то есть только от значений  $p_{ij}, j \in w$ ;
- свойство 1̄:  $f(i, w)$  зависит только от связей остальных участников коалиции  $w$  с участником  $i$ , то есть только от значений  $p_{ji}, j \in w$ ;

- свойство 2: для каждого набора значений  $(p_{11}, \dots, p_{1m}, p_{21}, \dots, p_{2m}, \dots, p_{i1}, \dots, p_{im}, \dots, p_{mm})$  существует значение  $0 \leq f(i, w) \leq 1$ , где  $w$  – произвольная коалиция;
- свойство 3: если существует  $p_{ij} = 0, j \in w$ , то  $f(i, w) = 0$ ;
- свойство 3̄: если существует  $p_{ji} = 0, j \in w$ , то  $f(i, w) = 0$ ;
- свойство 4: если существует  $p_{ij} = 1, j \in w$ , то  $f(i, w) = 1$ ;
- свойство 4̄: если существует  $p_{ji} = 1, j \in w$ , то  $f(i, w) = 1$ ;
- свойство 5 (усиленное свойство монотонности): если для коалиции  $\hat{w}$  все значения связей участника  $s \in \hat{w}$  с остальными участниками коалиции  $\hat{w}$  больше (меньше) всех значений связей участника  $i \in w$  с остальными участниками коалиции  $w$  на  $k \in [0,1]$ , то  $f(s, \hat{w}) = f(i, w) + k$  ( $f(i, w) = f(s, \hat{w}) + k$ );
- свойство 6 (Null player): если существует  $p_{ij} = 0, j \in w$ , то  $f(i, w) = f(i, w \setminus \{j\})$ ;
- свойство 6̄ (Null player): если существует  $p_{ji} = 0, j \in w$ , то  $f(i, w) = f(i, w \setminus \{j\})$ .

Теорема 1.6.

1)  $f(i, w)$  удовлетворяет свойствам 1, 2, 3, 5 тогда и только тогда, когда

$$f(i, w) = f_{min}^+(i, w);$$

2)  $f(i, w)$  удовлетворяет свойствам 1, 2, 4, 5, 6 тогда и только тогда, когда  $f(i, w) = f_{max}^+(i, w)$ ; 3.  $f(i, w)$  удовлетворяет свойствам 1̄, 2, 3̄, 5 тогда и только тогда, когда  $f(i, w) = f_{min}^-(i, w)$ ; 4.  $f(i, w)$  удовлетворяет свойствам 1̄, 2, 4̄, 5, 6̄ тогда и только тогда, когда  $f(i, w) = f_{min}^-(i, w)$ .

Свойства 1, 1̄, 2 взяты из [60] и основаны на аксиоме независимости ПА. Идея для свойств 3, 3̄ основана на понятии nullifying agent, свойстве Nullifying Agents Get Nothing рассмотренных в [62]. Свойства 6, 6̄ основаны на понятии null agent, рассмотренном в [62]. Свойство 5 является усилением свойства монотонности аксиоматики, рассмотренной в [60]. Идея свойств 4, 4̄

основана на нормировке функций интенсивности связи, свойстве монотонности.

### 1.6.3 Некоторые выводы

В проведенном исследовании были рассмотрены индексы влияния, учитывающие предпочтения участников по созданию коалиции, введенные в [60]. В качестве цели исследования была поставлена задача аксиоматизации функций интенсивности связи. Изучен способ построения функции интенсивности связи через кардинальные индексы. В результате получена аксиоматика для 4 функций интенсивности, основанных на функциях минимума и максимума. А также рассмотрены аксиомы, основанные на других исследованиях в области кооперативных игр.

В дальнейшем планируется провести разработку аксиоматики функций интенсивности связи, основанных на линейных функциях. Кроме того, планируется провести аксиоматизацию всех 16 функций интенсивности связи.

## **2 Исследование оптимизационных моделей**

В этом разделе отчета приведены результаты исследований оптимизационных моделей в следующих направлениях:

- исследование моделей оптимального административного управления;
- исследование оптимизационной задачи минимизации гарантированного риска;
- модификация метода расчёта приоритетов стратегических целей.

### **2.1 Исследование моделей оптимального административного управления**

#### **2.1.1 Введение**

В 2021 планировалось исследование класса игр на множестве связанных стратегий игроков, в виде которых могут формулироваться некоторые задачи административного управления. Как и в предыдущих исследованиях в области игровых задач, которые возникают в практических системах планирования и управления, основное внимание предполагалось уделить тем системам административного управления, в которых для моделирующих их (разрешимых) игр можно было бы построить: а) проверяемые условия равновесия; б) методы отыскания оптимальных стратегий игроков, позволяющие использовать эти условия и эти методы в прикладных задачах, особенно в таких задачах большого размера (к которым относятся и задачи, изучаемые в рамках настоящего исследования) на основе стандартного программного обеспечения, эффективного при решении таких задач. Системам административного управления, связанным с охраной окружающей среды, планировалось уделить особое внимание, и в качестве таких систем предполагалось рассмотреть системы, регулирующие использование: а) электромобилей; б) любых альтернативных источников энергии; в) переработку твердых бытовых отходов в повседневной жизни городов. По исследованию проблем математического моделирования функционирования именно этих систем участниками проекта были получены обнадеживающие результаты в последние годы [63], [64], [65], [66], [67].

В предыдущих исследованиях, в частности, в [63], [64] было доказано, что, например, в задаче использования альтернативных источников энергии в системах электроснабжения городов, формулируемой в виде игры на выпуклом многограннике связанных стратегий конечного (более трех) числа игроков, отыскание как условий равновесия, так и оптимальных стратегий игроков, может осуществляться решением трех вспомогательных задач линейного программирования, две из которых образуют двойственную пару. Аналогичная возможность указывалась в [65] для рассмотренной там задачи организации переработки твердых бытовых отходов в городах, содержательная постановка которой приведена в [65] (хотя ни математическая модель, ни играя задача, формулируемая на ее основе, автором ранее не публиковались). Наконец, для задачи о размещении зарядных станций для электромобилей на участке автомагистрали, в [66] была предложена математическая модель, на основе которой эта задача формулируется в [66] в виде задачи отыскания максимума суммы билинейной и линейных функций векторных переменных, часть из которых принимает только целочисленные значения для всех своих координат. Было доказано [66], что отыскание оптимальных стратегий игроков в этой задаче сводится к отысканию решений в задаче квадратичного программирования, в задаче линейного программирования и в задаче смешанного математического программирования, требуя решения более сложных задач, чем только задачи линейного программирования, к которым удается свести указанные выше задачу для систем электроснабжения города и задачу, связанную с переработкой твердых бытовых отходов города.

Изучение ряда публикаций по системам административного управления, связанным с наличием электромобилей, как в личном пользовании у населения, так и в транспортных системах городов, регионов и других административных единиц (см., в частности, например, [68], [69]), показало, что задача организации страхования этих электромобилей (в каких бы административных единицах оно ни рассматривалось) требует более детального обсуждения и освещения правовых вопросов этого страхования, чем то, ко-

торое было проведено в имеющихся опубликованных материалах, и не может считаться достаточным для построения прикладных математических моделей, имеющих перспективу быть использованными на практике. Поэтому, в рамках настоящей работы, задача об оценке инвестиционной привлекательности проектов, связанных с размещением зарядных станций для электромобилей в городе с позиций потенциального инвестора, рассматривалась в качестве задачи административного управления, связанной с электромобилями, в которой изучалась сама возможность ее формулировки в виде игровой задачи на множестве связанных стратегий игроков (как одной из задач, для которых можно было бы строить системы планирования и управления, обладающие указанными выше свойствами). Целесообразность исследования принципиальной возможности такого рода формулировки была продиктована успешным моделированием задачи размещения таких станций для электромобилей на участке автомагистрали; математическая модель сама и игра, сформулированная на ее основе, были рассмотрены в [66].

В настоящей работе рассмотрена задача о размещении зарядных станций для электромобилей в городе, при решении которой была предпринята попытка использовать математическую модель и формулировку задачи, предложенные в [66] для решения задачи размещения зарядных станций для электромобилей на участке автомагистрали. Было показано, что структура математической модели, предложенной в [66], может быть использована и при построении математической модели задачи, являющейся предметом исследования в настоящей работе.

В то же время, задача, формулируемая в настоящей работе на основе математической модели, предложенной в [66], существенно отличается от задачи, рассматриваемой в [66]. В обоих случаях рассматриваемая задача математически формулируется как задача отыскания максимина функции двух групп векторных переменных. Переменные одной из этих групп управляются организацией, оценивающей экономическую привлекательность инвестирования в предлагаемые варианты размещения зарядных станций в городе для

потенциального инвестора, а другую группу этих переменных можно считать отражающей поведение рынка. Однако, все переменные из каждой группы считаются в [66] изменяющимися независимо от всех остальных переменных, в то время как в задаче, рассматриваемой в настоящем исследовании, все векторные переменные, входящие в первую группу, считаются удовлетворяющими некоторой системе линейных неравенств, т.е. являются (линейно) связанными.

В настоящей работе показано, что:

- 1) задача оценки инвестиционной привлекательности проектов, связанных с размещением зарядных станций для электромобилей в городе, для потенциального инвестора может формулироваться как указанная выше задача отыскания максимина со связанными векторными переменными (в рамках одной из двух групп этих переменных);
- 2) решение указанной в а) задачи может быть сведено к решению задачи квадратичного программирования, задачи линейного программирования и задачи математического программирования со смешанными переменными и линейной структурой системы ограничений.

Формулировку рассматриваемой задачи, предложенную автором в настоящем исследовании, следует рассматривать как первый шаг в построении игровой модели более общего варианта этой задачи, в которой связанными являются векторные переменные не только из первой группы, но и векторные переменные из разных групп и в которой, как при рассмотрении задачи, исследованной в настоящей работе, будут использованы результаты автора по исследованию игр двух лиц на множестве связанных стратегий игроков [67].

Кроме задачи размещения зарядных станций для электромобилей в городе, в настоящей работе, как и планировалось, рассматривались задача организации переработки твердых бытовых отходов в крупных городах, описание которой приведено в [65], и задача оптимизации использования всех альтернативных источников энергии в системах электроснабжения крупных го-

родов, которая (в близкой постановке) изучалась в [63], [64]. Однако, в то время как для каждой из этих двух задач в игровой постановке показана возможность сведения отыскания в них точек равновесия Нэша к решению трех вспомогательных задач линейного программирования, рассмотрение связанных стратегий игроков в задаче организации переработки твердых бытовых отходов в крупных городах приводит к более сложным играм на множествах связанных стратегий, чем игра, рассмотренная в настоящем исследовании (при формализованном описании задачи размещения зарядных станций для электромобилей в городе). Этот факт вызывает вопрос о том, что же следует считать более целесообразным с прикладной точки зрения:

- 1) поиск условий равновесия в игре, являющейся математическим описанием рассмотренной в настоящей работе и в [70] задачи размещения зарядных станций для электромобилей в городе, позволяющих сводить отыскание точек равновесия в игре к решению конечного числа задач линейного программирования;
- 2) поиск практически применимых формулировок задачи организации переработки твердых бытовых отходов в городах и задачи использования альтернативных источников энергии в системах электроснабжения городов в виде игр со смешанными стратегиями игроков, в которых проверяемые условия (хотя бы достаточные) для точек равновесия Нэша и для оптимальных стратегий игроков имели бы такую же (или, по крайней мере, в каком-либо практическом смысле близкую) структуру, что и такие условия для задачи размещения зарядных станций для электромобилей в городе.

Более подробно эта проблема, с анализом плюсов и минусов каждого из этих вариантов, будет рассмотрена в [70].

### 2.1.2 Математическая модель задачи размещения зарядных станций для электромобилей в городе

Первая группа векторных переменных в рассматриваемой ниже задаче о размещении зарядных станций для электромобилей в городе включает в себя

- ожидаемый спрос на электроэнергию (объем) на каждой из станций сети в пределах каждого из 23 часовых сегментов в пределах суток в каждом из 4 сезонов;
- ожидаемую цену, по которой электроэнергия будет покупаться для каждой из станций сети в пределах каждого из 23 часовых сегментов в пределах суток в каждом из 4 сезонов;
- ожидаемую розничную цену, по которой электричество (электроэнергия) будет продаваться потребителям (клиентам) на каждой из станций сети в пределах каждого из 23 часовых сегментов в пределах суток в каждом из 4 сезонов;
- объем электроэнергии, ожидаемой для покупки для каждой из станций сети в пределах каждого из 23 часовых сегментов в пределах суток в каждом из 4 сезонов.

Вторая группа векторных переменных включает в себя:

- ожидаемый суточный объем финансовой поддержки каждой из станций сети в каждом из 4 сезонов (со стороны государства и/или частного сектора);
- ожидаемую суточную стоимость обслуживания станции в каждом из 4 сезонов (относящуюся к обслуживанию зарядных станций для электромобилей);
- число зарядных станций каждого типа, которые планируется установить на каждой из станций сети в каждом из 4 сезонов.

Параметрами, присутствующими в математической модели наряду с векторными переменными, являются:

- число типов зарядных станций, которые могут быть закуплены, установлены на станциях сети и пущены в эксплуатацию в городе;

- дневная мощность зарядной станции каждого типа на каждой из станций сети;
- ожидаемое максимальное значение суточного объема электроэнергии, которое планируется закупить для зарядных станций, планируемых быть установленными на каждой из станций сети;
- ожидаемые затраты, связанные с приобретением и установкой зарядной станции каждого типа в городе (которые считаются одинаковыми для всех станций сети);
- ожидаемая суточная стоимость обслуживания зарядной станции каждого типа на каждой из станций сети в каждом из 4 сезонов;
- число дней в каждом из 4 сезонов;
- ожидаемый максимальный суточный объем электроэнергии, который потребуется на каждой зарядной станции в каждом из 4 сезонов;
- ожидаемый минимальный суточный объем электроэнергии, который потребуется на каждой зарядной станции в каждом из 4 сезонов;
- число лет, на которое оценивается привлекательность требуемого инвестирования для потенциального инвестора;
- размер налоговых отчислений (от дохода), устанавливаемый для сети станций города, на которых планируется установка зарядных станций;
- коэффициент, отражающий потери электроэнергии, которую ожидается купить у сети (или у сетей) электроснабжения и доставить к станциям сети, на которых планируется установка зарядных станций, включая потери от передачи электроэнергии к станциям сети.

Связанность перечисленных векторных переменных в математической модели, описывающей функционирование зарядных станций, планируемых к установке на сети пунктов города, требует некоторого обоснования.

Ясно, что приобретение электромобиля как организацией, так и частным лицом в любом городе предполагает наличие в нем мест, в которых было бы возможно “заправить” этот электромобиль электричеством. Поэтому, неслучайно, среди пунктов города, в которых обычно рассматривается воз-

можность установки зарядных станций для обслуживания электромобилей, находятся многие из действующих станций по заправке бензиновых двигателей. Прежде всего, это объясняется а) привычкой водителей заправлять свои машины бензином в определенных местах города и б) наличием гибридных автомобилей, в которых имеется как электрический, так и бензиновый двигатель. Однако, в отличие от автомобилей, работающих только на бензиновых двигателях, многие владельцы электромобилей могут организовать заправку своих машин или в домашних условиях (особенно имеющие собственные дома или арендующие такие дома в пределах городской черты или на близкой к ней территории). Даже если такая заправка (по каким-либо причинам) производится ими не всегда, сама возможность такой заправки достаточным числом владельцев электромобилей делает спрос на электроэнергию на зарядных станциях любого города нестабильным при любом варианте размещения зарядных станций в городе, превращая этот спрос в векторную переменную (поскольку для разных мест установки зарядных станций, для разных временных сегментов в пределах суток и для разных сезонов этот спрос, вообще говоря, разный).

Указанный ожидаемый нестабильный спрос на электроэнергию на зарядных станциях сети, является одним из двух факторов, определяющих экономику функционирования сети зарядных станций. Другим фактором являются цены, по которым электроэнергия может покупаться у систем электроснабжения города для каждой из станций сети в пределах каждого из 23 часовых сегментов в пределах суток в каждом из 4 сезонов (также образующие векторную переменную в рамках модели, рассматриваемой в настоящем исследовании). Оба фактора очевидно относятся к числу тех, которые не могут быть точно указаны в любом из 23 часовых сегментов в пределах суток в каждом из 4 сезонов, и их значения в указанных выше часовых промежутках удобно рассматривать как стратегию одного из игрока в игре двух лиц. В этой игре один из игроков предлагает услуги в виде возможности заправить электромобили, формирующие некоторый парк электромобилей, электро-

энергией на заправочных станциях города, а другой игрок запрашивает услуги по заправке электроэнергией электромобилей из этого парка и готов эти услуги оплачивать по приемлемым (для этого игрока) тарифам.

Ясно, что предлагающий услугу игрок может выбирать тарифы, по которым услуга будет предлагаться запрашивающему услугу игроку в разные временные сегменты в пределах каждого суток в каждом из 4 сезонов, а также выбирать объем предлагаемой услуги в каждом из указанных временных промежутков. Одной из основных проблем организации сети зарядных станций в городе является проблема отыскания указанных выше приемлемых тарифов (для запрашивающего услугу игрока), которые делали бы предоставление этой услуги прибыльным бизнесом для игрока, предлагающего услуги по заправке электромобилей.

При разумной организации размещения сети зарядных станций в городе, объем предлагаемой услуги по заправке электромобилей должен быть не меньше, чем спрос на нее при таких тарифах на заправку, которые позволяют игроку, предоставляющему эту услугу, не испытывать никаких финансовых потерь при самом неблагоприятном (для этого игрока) стечении обстоятельств. Для того, чтобы определить какие стечения обстоятельств в городе с точки зрения предоставления конкретной транспортной услуги его населению возможны для этого игрока и какое из них (или какие из них) являются для этого игрока самым неблагоприятным (самыми неблагоприятными) необходимо рассмотреть условия жизни города, в которых такое определение производится.

Прежде всего, к таким условиям следует отнести: а) наличие общественного транспорта в городе (метро, автобусов и микроавтобусов, в том числе троллейбусов и микроэлектробусов); б) наличие конкурента (личного транспорта с бензиновыми двигателями); в) социальный состав владельцев личных электромобилей, использующих свои электромобили в городе (многие из этих владельцев имеют собственные дома и гаражи, в которых они могут осуществлять заправку и/или дозаправку электромобилей как в дневное

время, так и в ночное время); г) ожидаемую почасовую динамику изменения спроса на услуги по заправке электромобилей в течение суток и посutoчную динамику этого изменения в каждом сезоне со стороны как общественного так и личного электротранспорта; д) ожидаемые тарифы на услуги по заправке электромобилей электроэнергией в городе, политику местных властей по развитию наземного электрического транспорта в городе, ожидаемые тарифы на электроэнергию в городе со стороны систем электроснабжения города.

Даже когда перечисленные (и другие похожие) условия удается описать математически, нельзя говорить о множествах допустимых значений векторов:

- ожидаемого спроса на электроэнергию (объем) на каждой из станций сети в пределах каждого из 23 часовых сегментов в пределах суток в каждом из 4 сезонов;
- ожидаемую цену, по которой электроэнергия будет покупаться для каждой из станций сети в пределах каждого из 23 часовых сегментов в пределах суток в каждом из 4 сезонов, без учета их связи с векторами как ожидаемого объема закупок электроэнергии для сети станций города, на которых предполагается установка станций по зарядке электромобилей, так и тарифов, по которым электроэнергия для заправки электромобилей будет отпускаться на этих станциях в розничной торговле.

Хотя такое математическое описание является достаточно трудной задачей, требующей более детального изучения, чем анализ набора сведений, имеющихся на сегодняшний день, связанность перечисленных 4 групп векторов в рамках любого формализованного описания функционирования сети зарядных станций города должна быть отражена.

Множество векторов ожидаемых объемов закупок электроэнергии для сети станций города, на которых предполагается установка зарядных станций для электромобилей, в значительной мере определяется типом электрооборудования, устанавливаемого на этих зарядных станциях, числом зарядных станций каждого типа, устанавливаемых на каждой из станций сети, а также

наличием систем хранения электроэнергии, закупаемой для зарядных станций сети.

Множество векторов ожидаемых розничных цен, по которым электроэнергия будет продаваться потребителям (клиентам) на каждой из зарядных станций сети в значительной мере определяется приемлемостью этих цен для потребителя. С одной стороны, такие приемлемые цены не должны превышать некоторых значений, а, с другой стороны, эти приемлемые цены должны обеспечивать предприятию (предприятиям), предоставляющему (предоставляющим) электромобилям в городе возможность заправки электроэнергией на зарядных станциях города, требуемый уровень дохода (и прибыли) от этой деятельности.

Поскольку значения векторов в каждой из этих перечисленных выше 4 групп векторов не могут быть описаны точно, в наиболее общем случае, каждый вектор из каждой группы можно считать принадлежащим некоторому параллелепипеду с известными границами изменения каждого из векторов, либо некоторому выпуклому многограннику, являющемуся подмножеством этого параллелепипеда. Для каждого из 4 указанных выше векторов такой выпуклый многогранник является множеством допустимых решений некоторой совместной системы линейных неравенств балансового типа.

Ясно, что значения границ изменения а) каждого из векторов ожидаемых розничных цен, по которым электроэнергия будет продаваться потребителям (клиентам) на каждой из зарядных станций сети, б) ожидаемых объемов закупок электроэнергии для станций города, на которых предполагается установка зарядных станций для электромобилей, определяются (задаются) предприятием (или предприятиями), предоставляющему (предоставляющим) электромобилям возможность заправки электроэнергией на зарядных станциях города. То же справедливо и для соотношений, определяющих указанные выпуклые многогранники, являющиеся подмножествами указанных выше параллелепипедов, и описываемых указанными выше совместными системами линейных неравенств балансового типа.

Совершенно аналогичные утверждения можно высказать по поводу значений границ изменения в) ожидаемого спроса на электроэнергию (объем) на тех станциях сети, на которых предполагается установка зарядных станций для электромобилей, г) ожидаемой цены, по которой электроэнергия будет покупаться для каждой из станций сети, указанных в в), с той разницей, что эти границы определяются рынком, на котором водители электромобилей стараются наиболее выгодно для себя заправить эти электромобили электроэнергией. Очевидно, что соотношения, определяющие указанные выше выпуклые многогранники, являющиеся подмножествами соответствующих параллелепипедов (о которых шла речь применительно к указанным векторным переменным в) и г) выше), также описываются совместными системами линейных неравенств балансового типа.

Центральными соотношениями, предлагаемыми в рамках математической модели для задачи размещения зарядных станций для электромобилей в городе, являются соотношения, связывающие переменные, упомянутые выше в а) и б) (с одной стороны) и переменные, упомянутые выше в в) и г) (с другой стороны). Природа этих связей ясна на примере соотношений, где рассматриваются векторные переменные из всех упомянутых выше 4 групп этих векторных переменных.

Именно, пусть:

$u$  – вектор ожидаемого спроса на электроэнергию (объем электроэнергии) на тех станциях сети, на которых предполагается установка зарядных станций для электромобилей в городе;

$v$  – вектор ожидаемых цен на электроэнергию, по которым электроэнергия будет покупаться у систем электроснабжения города для последующей продажи на зарядных станциях сети;

$u$  – вектор ожидаемых розничных цен, по которой электроэнергия будет продаваться клиентам зарядных станций сети, и которые окажутся приемлемыми для клиентов этих зарядных станций;

$h$  – вектор объемов электроэнергии, которую предполагается купить у систем электроснабжения города для каждой из зарядных станций сети.

Тогда  $\langle y, u \rangle$  – скалярное произведение векторов  $y$  и  $u$  определяет доход предприятия (возможно предприятий), обеспечивающего (обеспечивающих) город электроэнергией для зарядных станций сети для заправки на них электромобилей, в то время как  $\langle v, h \rangle$  – скалярное произведение векторов  $v$  и  $h$  определяет затраты этого предприятия (или этих предприятий) на приобретение у систем электроснабжения города электроэнергии, необходимой для функционирования системы зарядных станций города. Деятельность, связанная с созданием и поддержанием работоспособности сети зарядных станций города, может привлекать потенциального инвестора если разность этих двух скалярных произведений, т.е.,  $\langle y, u \rangle - \langle v, h \rangle$ , выражающая прибыль предприятия (а часть этой прибыли является прибылью потенциального инвестора) за какой-либо конкретный промежуток времени (например, за сутки, за неделю, за месяц, и т.д.), будет не ниже некоторого (приемлемого для инвестора) конкретного значения.

Нетрудно заметить, что векторы  $y$  и  $h$  не могут считаться меняющимися независимо, поскольку векторное неравенство  $h > y$  должно выполняться. Аналогично, векторы  $u$  и  $v$  не могут считаться меняющимися независимо, поскольку векторное неравенство  $u > v$  также должно выполняться (для неубыточного бизнеса). В то же время наличие таких неравенств в системе ограничений на все 4 группы переменных  $y$ ,  $v$ ,  $u$ ,  $h$  делает эти переменные (линейно) связанными, причем эти переменные могут удовлетворять и более сложным системам линейных ограничений.

Для оценки инвестиционной привлекательности для потенциального инвестора какого-либо проекта, связанного с созданием сети зарядных станций, необходимо определить значение этой прибыли при наиболее неблагоприятном стечении для проекта обстоятельств. Такое, стечение обстоятельств, в частности, будет иметь место, если оба вектора  $y$  и  $v$  будут принимать значения из некоторого множества допустимых значений, структура которо-

го учитывает связанность этих переменных с переменными  $h$ ,  $u$ , соответственно, не позволяя значению скалярного произведения векторов  $y$  и  $u$  (доход предприятия) “сильно отклоняться” от значения скалярного произведения векторов  $h$  и  $v$  (расход предприятия на закупку электроэнергии у систем электроснабжения города и обеспечение функционирования сети зарядных станций города). Математически, это требует рассмотрения: а) функции  $f(y, v, u, h)$ , являющейся указанной выше разностью скалярных произведений на множестве  $W$  допустимых значений векторных переменных  $y$ ,  $v$ ,  $u$ ,  $h$ ; б) функции минимума этой функции на всех допустимых значениях пар векторов  $(y, v)$ ; в) максимизации указанной в б) функции минимума на множестве всех допустимых пар векторов  $(u, h)$ , т.е., требует решения задачи отыскания максимина функции  $f(y, v, u, h)$

$$\begin{aligned} \min_{(y,v)} f(y, v, u, h) &\rightarrow \max_{(u,h)} \\ T_1(y, v) + T_2(u, h) &\geq \varepsilon \\ y \in Y, v \in V, u \in U, h \in H \end{aligned} \tag{2.1}$$

где  $Y, V, U, H$  – выпуклые многогранники в соответствующих векторных пространствах, которым принадлежат векторные переменные  $y$ ,  $v$ ,  $u$ ,  $h$  соответственно,  $T_1$ ,  $T_2$  – матрицы и  $\varepsilon$  – вектор с действительными элементами соответствующих размеров, формализующие систему линейных ограничений, которым удовлетворяют указанные выше связи между парами векторных переменных  $(y, v)$  и  $(u, h)$ .

Здесь  $T_1(y, v)$  представляет собой вектор, являющийся произведением матрицы  $T_1$  на вектор  $(y, v)$  справа, а  $T_2(u, h)$  – вектор, являющийся произведением матрицы  $T_2$  на вектор  $(u, h)$  справа, а функция  $f(y, v, u, h)$  является билинейной функцией пар векторов  $(y, v)$  и  $(u, h)$ , т.е.,  $f(y, v, u, h) = \langle (y, v), D(u, h) \rangle$ , где  $D$  – матрица соответствующей структуры и размеров с действительными элементами.

Задача (2.1) является частным случаем задачи

$$(x^{***}, y^{***}) \in \arg_{(x,y) \in S} \max_y \min_x,$$

$$S = \{(x, y) \geq 0, Ax + By \geq d\},$$

$$\phi(x, y) = \langle p, x \rangle + \langle x, Dy \rangle + \langle q, y \rangle$$
(2.2)

где  $A$  и  $B$  – некоторые матрицы, а  $p$  и  $q$  – некоторые векторы с действительными элементами,  $S$  – выпуклый многогранник в конечномерном пространстве, рассмотренной в [67], где была показана возможность сведения отыскания решения этой задачи к отысканию решения в некоторой задаче квадратичного программирования.

Именно, в [67] приведена формулировка следующего утверждения

$$(x^{***}, y^{***}) \in \arg_{(x,y) \in S} \max_y \min_x \phi(x, y)$$
(2.3)

тогда и только тогда, когда существуют векторы

$$w_1^{***} \geq 0, w_2^{***} \geq 0, w_3^{***} \geq 0, w_4^{***} \geq 0, z_1^{***} \geq 0, z_2^{***} \geq 0,$$
(2.4)

которые вместе с векторами  $x^{***}, y^{***}$  образуют решение задачи

$$\langle h_2, H_2 h_2 \rangle + \langle p_2, h_2 \rangle \rightarrow \max_{h_2 \geq 0: T_2^{***} h_2 \leq \delta_2},$$
(2.5)

где  $H_2$  и  $T_2^{***}$  – матрицы и  $p_2$  – вектор с действительными элементами и соответствующих размеров,  $h_2 = (w_1^{***}, w_2^{***}, w_3^{***}, w_4^{***}, x^{***}, y^{***}, z_1^{***}, z_2^{***}) \geq 0$  и выполняется равенство

$$\max_{h_2 \geq 0: T_2^{***} h_2 \leq \delta_2} \langle h_2, H_2 h_2 \rangle + \langle p_2, h_2 \rangle = \phi(x^{***}, y^{***}).$$
(2.6)

В [67] отмечено, что в предположении о выполнении естественного условия, в соответствии с которым в задаче (2.3) для каждого  $y$ , для которо-

го  $(x, y) \in S$ , существует  $\min_{x:(x,y) \in S} \phi(x, y)$  это утверждение может быть доказано по той же схеме, по которой в [67] аналогичное утверждение (при аналогичном предположении) доказано для задачи

$$\begin{aligned} (x^{**}, y^{**}) &\in \operatorname{Arg}_{(x,y) \in S} \min_x \max_y \phi(x, y), \\ S &= \{(x, y) \geq 0, Ax + By \geq d\}, \\ \phi(x, y) &= \langle p, x \rangle + \langle x, Dy \rangle + \langle q, y \rangle \end{aligned} \quad (2.7)$$

(Здесь, соотношения (2.3) – (2.7) сформулированы в обозначениях, использованных в [67]).

Пусть:

$\pi$  – вектор ожидаемых объемов финансовой поддержки создания и функционирования сети зарядных станций города;

$d$  – вектор ожидаемой суточной стоимости обслуживания зарядных станций города;

$\omega$  – вектор, компонентами которого являются числа зарядных станций каждого типа, планируемых быть установленными в местах установки зарядных станций города.

Тогда, как показано в [66], для функции вида

$$f(u, h, \omega, y, v, \pi, d) = \langle y, u \rangle - \langle h, v \rangle + \langle \hat{\rho}, \pi \rangle - \langle \hat{\varepsilon}, d \rangle - \langle \hat{c}, \hat{\omega} \rangle - \langle g, \tilde{\omega} \rangle, \quad (2.8)$$

где  $\hat{\omega}$  и  $\tilde{\omega}$  – образы вектора  $\omega$  при некоторых линейных отображениях, справедливо равенство

$$\begin{aligned} \max_{u, h, w} \min_{y, v, \pi, d} f(u, h, \omega, y, v, \pi, d) &= \\ \max_{u, h, z_1, z_2, z_3, z_4, \omega} (\langle \tilde{\theta}, z_1 \rangle + \langle \tilde{p}, z_2 \rangle + \langle \tilde{\delta}, z_3 \rangle + \langle \tilde{\tau}, z_4 \rangle - \langle \beta, \omega \rangle) & \end{aligned} \quad (2.9)$$

где  $z_1 \geq 0, z_2 \geq 0, z_3 \geq 0, z_4 \geq 0$ , – неотрицательные векторы в конечно-мерных пространствах соответствующего размера удовлетворяющие системе линейных неравенств

$$(z_1, z_2, z_3, z_4) \begin{pmatrix} \tilde{D} & 0 & 0 & 0 \\ 0 & \tilde{P} & 0 & 0 \\ 0 & 0 & \tilde{\Psi} & 0 \\ 0 & 0 & 0 & \tilde{V} \end{pmatrix} \leq \begin{pmatrix} u \\ -h \\ \tilde{\rho} \\ \tilde{\varepsilon} \end{pmatrix}, \quad (2.10)$$

где  $\tilde{D}, \tilde{P}, \tilde{\Psi}, \tilde{V}$  – матрицы с действительными элементами, 0-матрицы соответствующих размеров, все элементы которых нулевые,  $\tilde{p}, \tilde{\delta}, \tilde{\tau}, \beta$  – векторы в конечно-мерных пространствах соответствующих размеров с действительными координатами, и  $\omega$  – вектор в конечно-мерном пространстве соответствующего размера с полностью целочисленными координатами.

В задаче, рассматриваемой в настоящей работе, векторы  $(y, v)$  и  $(u, h)$  связаны линейными неравенствами, присутствующими в (2.1), однако, как и в [66], эти векторы, образующие первую группу векторных переменных, рассмотренных выше, меняются независимо от векторных переменных, образующих вторую группу векторных переменных. Таким образом, отыскание значений векторных переменных в задаче (2.8) – (2.10), как и в [66], может быть осуществлено в результате решения некоторой задачи математического программирования со смешанными переменными. Возможность отыскания значений векторных переменных в результате преобразования задачи вида (2.2) к задаче квадратичного программирования вида (2.6) позволяет использовать стандартное программное обеспечение типа CPLEX и MILP для решения практических задач большого размера более эффективно, чем при решении задачи, рассмотренной в [66], где задача математического программирования со смешанными переменными, к которой сводится рассматриваемая в [66] задача, имеет более общую структуру. Отметим также, что проверка некоторых предположений, делающих возможными сведение задачи, рассмат-

риваемой в настоящей работе, к задаче (2.8) – (2.10) требует решения некоторых вспомогательных задач линейного программирования, что также может осуществляться с использованием стандартного программного обеспечения, в частности, CPLEX и MILP.

### 2.1.3 Математическая модель задачи организации переработки твердых бытовых отходов в крупных городах

Ниже рассматривается задача об отыскании резервных цен для двух тендеров, связанных с организацией мест захоронения твердых бытовых отходов города, проводимых администрацией города (содержательная постановка которой приведена, в частности, в [65]).

Первый тендер объявляется на проведение строительных работ, связанных с возведением новых и/или перепрофилированием существующих зданий как для перерабатывающего завода, предполагаемого быть построенным на каком-либо участке для производства какого-либо полезного продукта, так и для сортировочного комплекса, предполагаемого быть построенным на этом участке.

Второй тендер объявляется на поставку технологии переработки ТБО города в какой-либо полезный продукт, и оборудования, реализующего эту технологию, на перерабатывающий завод, который планируется построить на каком-либо участке, включая управление этим заводом и продажу этого продукта победителем этого тендера, и оборудования для сортировочного комплекса, который планируется построить на этом участке.

Эта задача формулируется в виде некооперативной игры трех лиц на выпуклых многогранниках допустимых стратегий игроков.

Далее в этом разделе, АЛС – администрация крупного города, СА – концессионное соглашение [71].

Пусть:

$n$  – число участков, на которых ТБО города в настоящее время выгружается и на которые будет выгружаться в ближайшие годы;

$c_i$  – объем ТБО города (в тоннах), который в настоящее время находится на участке с номером  $i \in \overline{1, n}$ ;

$D$  – число полезных продуктов, которые могут быть извлечены из городских ТБО, находящихся на всех  $n$  участках, в результате их переработки, включая компоненты для цементных блоков, получаемые из пепла;

$y_i^d$  – годовая мощность завода, который предполагается построить на участке  $i \in \overline{1, n}$ , для переработки части городских ТБО в полезный продукт  $d \in \overline{1, D}$ , т.е., объем (в тоннах) части городских ТБО, которые могут быть переработаны в полезный продукт  $d$  из ТБО, которое ожидается быть выгруженным на участке  $i \in \overline{1, n}$ ;

$\underline{\mu}_i^d$  – наименьшая нижняя граница возможных значений  $y_i^d$  (в тоннах),  $i \in \overline{1, n}, d \in \overline{1, D}$ ;

$\overline{\mu}_i^d$  – наибольшая верхняя граница возможных значений  $y_i^d$  (в тоннах),  $i \in \overline{1, n}, d \in \overline{1, D}$ ;

$y_i^o$  – годовая мощность сортировочного комплекса, который предполагается построить на участке с номером  $i \in \overline{1, n}$  для сортировки ТБО города в  $D$  порций, подлежащих дальнейшей переработке для извлечения из них полезных продуктов (в тоннах);

$r > 0$  – бюджет, который инвестор может зарезервировать для оплаты всех строительных работ, связанных с возведением новых и/или пере-профилированием имеющихся зданий для всех перерабатывающих заводов, которые предполагается построить на всех ( $n$ ) участках;

$\pi > 0$  – бюджет, который инвестор может зарезервировать для приобретения технологий для переработки ТБО города в полезные продукты и оборудования, реализующего эти технологии для всех перерабатывающих заводов, которые предполагается построить на всех ( $n$ ) участках;

$\beta_i^d > 0$  – стоимость работ, связанных с построением новых фундаментов для зданий и/или перепрофилированием фундаментов существующих зданий и с обеспечением всех коммунальных услуг для перерабатывающего завода, который предполагается построить на участке  $i \in \overline{1, n}$  для производства полезного продукта  $d \in \overline{1, D}$ ;

$\beta_i^0 > 0$  – стоимость работ, связанных с построением новых фундаментов для зданий и/или перепрофилированием фундаментов существующих зданий и с обеспечением всех коммунальных услуг для сортировочного комплекса, который предполагается построить на участке  $i \in \overline{1, n}$ ;

$x_i^d$  – рыночная стоимость строительных работ, связанных как с возведением новых, так и с перепрофилированием имеющихся зданий для перерабатывающего завода, который планируется построить на участке  $i \in \overline{1, n}$  (на уже построенных новых и/или на перепрофилированных фундаментах существующих зданий) для производства полезного продукта  $d \in \overline{1, D}$ , на единицу годовой мощности этого завода;

$x_i^0$  – рыночная стоимость строительных работ, связанных как с возведением новых, так и с перепрофилированием имеющихся зданий для сортировочного комплекса, который планируется построить на участке  $i \in \overline{1, n}$  (на уже построенных новых и/или на перепрофилированных фундаментах существующих зданий) для производства полезного продукта  $d \in \overline{1, D}$ , на единицу годовой мощности этого сортировочного комплекса;

$\underline{\rho}_i^d$  – наименьшая нижняя граница возможных значений  $x_i^d$ ,  $i \in \overline{1, n}$ ,  $d \in \overline{0, D}$ ;

$\overline{\rho}_i^d$  – наибольшая верхняя граница возможных значений  $x_i^d$ ,  $i \in \overline{1, n}$ ,  $d \in \overline{0, D}$ ;

$\alpha^d > 0$  – коэффициент, отражающий соотношение между объемом строительных работ, связанных с возведением новых и/или перепрофилированием существующих зданий как для перерабатывающего завода, предполагаемого быть построенным на каком-либо участке для производства полезного продукта  $d \in \overline{1, D}$ , так и для сортировочного комплекса, предполагаемого быть построенным на этом участке, и годовой производительностью (мощностью) этого завода и годовой производительностью (мощностью) этого комплекса, соответственно;

$\lambda > 0$  – доля годового дохода от продажи какого-либо полезного продукта  $d \in \overline{1, D}$ , производимого каждым перерабатывающим заводом, который предполагается построить, идущая совместному предприятию из полного годового дохода, который предполагается получаться победителем тендера на а) обеспечение перерабатывающей технологией этот завод и оборудованием, реализующим эту технологию, б) управление функционированием этого завода и продажей этого полезного продукта;

$u_i^d$  – рыночная цена, которую планируется платить за а) технологию, которая будет приобретаться или для переработки фракции ТБО города в полезный продукт  $d \in \overline{1, D}$  на перерабатывающем заводе, который предполагается построить на участке  $i \in \overline{1, n}$ , или для сортировки ТБО города (при  $d=0$ ), и б) оборудование, реализующее эту технологию, на единицу (годовой) мощности перерабатывающего завода или на единицу (годовой) мощности сортировочного комплекса, соответственно;

$\underline{\delta}_i^d$  – наименьшая нижняя граница возможных значений  $u_i^d$ ,  $i \in \overline{1, n}$ ,  $d \in \overline{0, D}$ ;

$\overline{\delta}_i^d$  – наибольшая верхняя граница возможных значений  $u_i^d$ ,  $i \in \overline{1, n}$ ,  $d \in \overline{0, D}$ ;

$w_i^d$  – годовой объем фракции ТБО города, который ожидается быть выгруженным на участок  $i \in \overline{1, n}$  и который может быть переработан в полезный продукт  $d$ ,  $d \in \overline{1, D}$  на этом участке в течение первых трех лет (в тоннах);

$1 > \tau_i^d > 0$  – доля тонны ТБО города, которая может быть переработана в полезный продукт  $d$  на перерабатывающем заводе для переработки фракции ТБО города в полезный продукт  $d$  на участке  $i$ ,  $d \in \overline{1, D}$ ,  $i \in \overline{1, n}$ ;

$\underline{\sigma}_i^d$  – наименьшая нижняя граница возможных значений  $w_i^d$  (в тоннах),  $i \in \overline{1, n}$ ,  $d \in \overline{0, D}$ ;

$\overline{\sigma}_i^d$  – наибольшая верхняя граница возможных значений  $w_i^d$  (в тоннах),  $i \in \overline{1, n}$ ,  $d \in \overline{0, D}$ ;

$\underline{\psi} > 0$  – наибольшая нижняя граница ожидаемого суммарного ежегодного спроса на размещение ТБО города на всех участках в течение периода, охватываемого СА (в тоннах);

$\overline{\psi} > 0$  – наименьшая верхняя граница ожидаемого суммарного ежегодного спроса на размещение ТБО города на всех участках в течение периода, охватываемого СА (в тоннах);

$k > 0$  – доля финансовых средств, которые должны быть выплачены инвестором а) победителю тендера на поставку технологии переработки ТБО и оборудования, реализующего эту технологию, управление (планируемыми к постройке) заводами, перерабатывающими фракцию ТБО города в какой-либо полезный продукт  $d \in \overline{1, D}$ , и продажу продукта  $d$  (в рамках второго тендера), б) победителю тендера на поставку технологии сортировки ТБО города и оборудования, реализующего эту технологию, которую инвестор предлагает победителю тендера на первый тендер (на возведение новых и/или перепрофилирование существующих зданий для перерабатывающих заводов, которые планируется построить на всех

( $n$ ) участках) в качестве компенсации за дополнительные услуги по обеспечению поддержания работоспособности всех перерабатывающих заводов и всех сортировочных комплексов, которые планируется построить на каждом из ( $n$ ) участков;

$\gamma_i^d$  – штраф за “хранение” (т.е., за не переработку) единицы объема (одной тонны) ТБО города, выгруженных на участок  $i \in \overline{1, n}$ , связанное с загрязнением грунтовых вод;

$\gamma_i^0$  – стоимость вышеуказанных дополнительных услуг, требуемых для сортировочных комплексов, которые планируется построить на участке  $i, i \in \overline{1, n}$ , на единицу мощности сортировочного комплекса;

$\tilde{\theta}_i^d$  – доход на единицу объема (одну тонну), который планируется быть полученным победителем тендера на поставку технологии переработки ТБО города в полезный продукт  $d \in \overline{1, D}$  и оборудования, реализующего эту технологию, на перерабатывающий завод, который планируется построить на участке  $i \in \overline{1, n}$ , включая управление этим заводом и продажу продукта  $d$  победителем этого тендера;

$\Delta_i$  – ожидаемый штраф, который совместное предприятие заплатит вышестоящим властям (региональным и/или федеральным, непосредственно или через АЛС в рамках СА) за выброс в атмосферу одного кубического метра газа, который будет выделяться в результате “хранения” одной тонны ТБО города, находящегося на участке  $i \in \overline{1, n}$ ;

$\zeta$  – число кубических метров газа, который ожидается выделиться в результате “хранения” одной тонны ТБО города, находящегося на каждом участке;

$\nu_i^d$  – штраф, который совместное предприятие должно заплатить более высоким (региональным и/или федеральным властям непосредственно или через АЛС в рамках СА) за выброс в атмосферу одного кубического метра газа, который ожидается быть выброшенным в атмосферу в ре-

зультате переработки фракции ТБО города в полезный продукт  $d \in \overline{1, D}$  на перерабатывающем заводе, который планируется построить на участке  $i \in \overline{1, n}$ ;

$\zeta^{d^*}$  – число кубических метров газа, который ожидается выделиться в результате переработки одной тонны фракции ТБО города в полезный продукт  $d \in \overline{1, D}$ ;

$\underline{\eta}^d$  – наименьшая нижняя граница возможных значений минимума годовой мощности всех заводов, перерабатывающих фракцию ТБО города в полезный продукт  $d \in \overline{1, D}$ , которые планируется построить на всех ( $n$ ) участках (в тоннах);

$\bar{\eta}^d$  – наибольшая верхняя граница возможных значений максимума годовой мощности всех заводов, перерабатывающих фракцию ТБО города в полезный продукт  $d \in \overline{1, D}$ , которые планируется построить на всех ( $n$ ) участках (в тоннах);

$\Theta_i^d$  – ожидаемые ежегодные выплаты, связанные со стандартным обслуживанием зданий, в которых или перерабатывающий завод, производящий полезный продукт  $d \in \overline{1, D}$ , или сортировочный комплекс (при  $d=0$ ) предположительно будет размещаться на участке  $i \in \overline{1, n}$ , которые совместное предприятие будет делать победителю тендера на первый тендер, с момента начала работы этого завода (или перерабатывающего комплекса) на единицу мощности перерабатывающего завода (или сортировочного комплекса);

$\Psi > 3$  – продолжительность функционирования всех перерабатывающих заводов и сортировочных комплексов, оговоренная в СА.

Вводя векторные переменные [71]

$$y_i = (y_i^0, \dots, y_i^D), \quad x_i = (x_i^0, \dots, x_i^D), \quad u_i = (u_i^0, \dots, u_i^D),$$

$$w_i = (0, w_i^1, \dots, w_i^D), \quad \gamma_i = (\gamma_i^0, \dots, \gamma_i^D), \quad \Theta_i = (\Theta_i^0, \dots, \Theta_i^D),$$

$$c_i = (0, c_i^1, \dots, c_i^D), \quad \tilde{\omega}_i = (0, \tilde{\omega}_i^1, \dots, \tilde{\omega}_i^D), \quad \tilde{\theta}_i = (0, \tilde{\theta}_i^1 \tau_i^1, \dots, \tilde{\theta}_i^D \tau_i^D),$$

$$v_i = (0, v_i^1, \dots, v_i^D), \quad \Delta_i = (0, \Delta_i^1, \dots, \Delta_i^D), \quad \zeta^* v_i = (0, \zeta^{1*} v_i^1, \dots, \zeta^{D*} v_i^D)$$

в пространстве  $R_+^{D+1}, i \in \overline{1, n}$ , и рассматривая некоторые выпуклые многогранники  $M(x_1, \dots, x_n), H(u_1, \dots, u_n), \Omega(y_1, \dots, y_n), W(w_1, \dots, w_n)$  в пространстве  $R_+^{n(D+1)}$ , можно показать, что рассматриваемая задача определения структуры контрактов для каждого из  $n$  мест и резервных цен для тендеров на эти контракты может быть сформулирована в виде игры трех лиц

### Игра 1

$$\begin{aligned} f_1(x_1, \dots, x_n; u_1, \dots, u_n; y_1, \dots, y_n) &= \\ &= \max_{(w_1, \dots, w_n) \in W} \sum_{i=1}^n \left( \sum_{d=0}^D \beta_i^d + \sum_{d=0}^D \alpha_i^d y_i^d x_i^d + \langle y_i, u_i \rangle + \right. \\ &\quad \left. \kappa \langle \tilde{\omega}_i + \zeta \Delta_i, w_i - y_i \rangle + k \langle \zeta^* v_i + \Theta_i - \lambda \tilde{\theta}_i, y_i \rangle \right) \\ &\rightarrow \min_{(y_1, \dots, y_n) \in \Omega} \end{aligned}$$

$$\begin{aligned} f_2(x_1, \dots, x_n; u_1, \dots, u_n; y_1, \dots, y_n) &= \\ &= \sum_{i=1}^n \left( \sum_{d=0}^D \beta_i^d + \sum_{d=0}^D \alpha_i^d y_i^d x_i^d + k \langle y_i, u_i \rangle - \kappa \langle \Theta_i, y_i \rangle - \kappa \langle \gamma_i, y_i \rangle \right) \rightarrow \\ &\rightarrow \max_{(x_1, \dots, x_n) \in M} \end{aligned}$$

$$\begin{aligned} f_3(x_1, \dots, x_n; u_1, \dots, u_n; y_1, \dots, y_n) &= (1-k) \sum_{i=1}^n \langle y_i, u_i \rangle + \\ &\quad \kappa (1-\lambda) \sum_{i=1}^n \langle \tilde{\theta}_i, y_i \rangle \rightarrow \max_{(u_1, \dots, u_n) \in H} \end{aligned}$$

где

$$\begin{aligned} f_1(x_1, \dots, x_n; u_1, \dots, u_n; y_1, \dots, y_n), f_2(x_1, \dots, x_n; u_1, \dots, u_n; y_1, \dots, y_n), \\ f_3(x_1, \dots, x_n; u_1, \dots, u_n; y_1, \dots, y_n) \end{aligned}$$

являются платежными функциями игроков. В этой игре первым игроком является инвестор, вторым игроком является потенциальная строительная компания, а третьим игроком является потенциальный оператор, который: а) обеспечит поставку технологий сортировки и переработки ТБО города и оборудование, реализующее эти технологии, как на перерабатывающих заводах, так и на сортировочных комплексах; б) будет управлять всеми этими

предприятиями и продавать все полезные продукты, которые будут получаться в результате переработки ТБО города.

Здесь, все параметры в математических описаниях целевых функций игроков определяются с учетом веденных обозначений и в предположении о выполнении ряда (естественных) соотношений между этими параметрами и переменными, используемыми в указанных математических описаниях, которые приведены в [71].

Пусть  $\sum_{d=0}^D \beta_i^d = \beta_i, i \in \overline{1, n}$ , а

$$x = (\alpha^0 x_1^d, \dots, \alpha^D x_1^d; \dots; \alpha^0 x_n^d, \dots, \alpha^D x_n^d), \quad y = (y_1, \dots, y_n),$$

$$u = (u_1, \dots, u_n), \quad w = (w_1, \dots, w_n), \quad \Delta = (\Delta_1, \dots, \Delta_n),$$

$$\tilde{\omega} = (\tilde{\omega}_1, \dots, \tilde{\omega}_n), \quad \zeta^* v = (\zeta^* v_1, \dots, \zeta^* v_n), \quad \tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_n), \\ \Theta = (\Theta_1, \dots, \Theta_n), \quad \gamma = (\gamma_1, \dots, \gamma_n)$$

векторы из  $R^{n(D+1)}$ , и пусть  $x \in \tilde{M}$ , где  $\tilde{M}$  – выпуклый многогранник, являющийся образом выпуклого многогранника  $M$  при линейном отображении, определяемом  $n(D+1) \times n(D+1)$  диагональной матрицей, в которой главная диагональ образована вектором  $(\alpha_1^0, \dots, \alpha_1^D; \dots; \alpha_n^0, \dots, \alpha_n^D) \in R^{n(D+1)}$ .

Тогда при некоторых дополнительных предположениях, приведенных в [71], полагая  $\tilde{\omega} + \zeta \Delta + \lambda \tilde{\theta} - \zeta^* v = \omega, \tilde{\gamma} = \gamma + \Theta, \Lambda = \sum_{i=1}^n \beta_i$  и  $\Omega(\omega) = \{y \in \Omega : y \leq \omega\}$ , игру (2.1) можно записать в виде

## Игра 2

$$f_1(x, y, u) = \Lambda + \langle y, x \rangle + \langle y, u \rangle - \kappa \langle \omega, y \rangle + \max_{w \in W} \kappa \langle \tilde{\omega} + \zeta \Delta, w \rangle \rightarrow \min_{y \in \Omega(\omega)},$$

$$f_2(x, y, u) = \Lambda + \langle y, x \rangle + k \langle y, u \rangle - \kappa \langle \tilde{\gamma}, y \rangle \rightarrow \max_{x \in M},$$

$$f_3(x, y, u) = (1 - k) \langle y, u \rangle + \kappa(1 - \lambda) \langle \tilde{\theta}, y \rangle \rightarrow \max_{u \in H}.$$

Проверка совместности систем ограничений, описывающих выпуклые многогранники  $\tilde{M}, H, \Omega$  и  $W$  при конкретных значениях всех коэффициентов, присутствующих в этих ограничениях, может осуществляться известными методами, в частности, методом описанным в [72]. Следует также заме-

тить, что выражение  $\kappa \langle \tilde{\omega} + \zeta \Delta, w \rangle$  является неотрицательным числом, которое не зависит от значения вектора  $y$ .

В [73] доказаны

**Теорема 2.1.** Тройка векторов  $(y^*, x^*, u^*)$  является точкой равновесия Нэша в игре 2 тогда и только тогда, когда

- 1) имеет место включение  $(y^*, x^*, u^*), w^* \in \text{Arg max}_{w \in W} \kappa \langle \tilde{\omega} + \zeta \Delta, w \rangle$ ;
- 2) пара  $(y^*, (x^*, u^*))$  является седловой точкой во вспомогательной антагонистической игре игроков А и В с платежной функцией

Игра 3

$$\langle y, x \rangle + \langle y, u \rangle - \kappa \langle \omega, y \rangle$$

на множестве допустимых стратегий игроков  $\Omega(w^*)$  (для игрока А) и на множестве допустимых стратегий игроков  $\tilde{M}xH$  (для игрока В). В этой антагонистической игре платежная функция максимизируется по паре векторов  $(x, u)$  и минимизируется по вектору  $y$ .

**Теорема 2.2.** Пусть

$$\tilde{M} = \{x \geq 0 : Gx \geq g\}, \Omega(w^*) = \{y \geq 0 : Sy \geq s\}, H = \{u \geq 0 : Lu \geq l\},$$

где  $G, S, L$  – матрицы и  $g, s, l$  – векторы соответствующих размеров. Тогда векторы  $y^*, x^*, u^*$ , образующие седловую точку в игре 3, могут быть найдены из решения задач линейного программирования, образующих двойственную пару

$$\begin{aligned} & \langle s, z \rangle \rightarrow \max_{(z, (u, x)) \in Q}, \\ & Q = \{(z(u, x)) \geq 0 : zS \leq -\kappa w + x + u, Gx \geq g, Lu \geq l\} \end{aligned}$$

и

$$\begin{aligned} & -\kappa \langle w, y \rangle - \langle g, t^1 \rangle - \langle l, t^2 \rangle \rightarrow \min_{(y, (t^1, t^2)) \in P}, \\ & P = \{(y, (t^1, t^2)) \geq 0 : t^1 G \leq -y, t^2 L \leq -y, Sy \geq s\}. \end{aligned}$$

Пусть, наконец,  $W = \{w \geq 0 : Kw \geq e\}$ . Отыскание вектора, являющегося компонентой точки равновесия Нэша в игре 2, может осуществляться решением задачи линейного программирования

$$\langle \kappa(\tilde{\omega} + \zeta\Delta), w \rangle \rightarrow \max_{w \geq 0: Kw \geq e}.$$

Таким образом, отыскание резервных цен в двух рассматриваемых тендерах и значений других параметров задачи управления, рассматриваемой в настоящей работе, которые являются компонентами точки равновесия Нэша в игре 2, сводится к решению трех задач линейного программирования, которое может осуществляться с использованием стандартных пакетов прикладных программ CPLEX и MILP.

#### 2.1.4 О математической модели задачи оптимизации использования альтернативных источников энергии в системах электроснабжения крупных городов

Взаимодействие между а) электростанцией (генератором электроэнергии), б) всеми крупными региональными потребителями электроэнергии из электросети, получающими электроэнергию от как от генератора, так и из имеющихся в регионе возобновляемых источников энергии (как напрямую, так и через системы хранения электроэнергии), и (с) компанией, осуществляющей доставку электроэнергии к потребителям по электрической сети, моделируется в [63] в виде игры с конечным (более трех) числом игроков на выпуклом многограннике связанных стратегий игроков.

Рассматривается вспомогательная игра трех лиц на выпуклых многогранниках несвязанных стратегий игроков, связанная с исходной игрой и доказывается, что точка равновесия в этой вспомогательной игре является точкой равновесия в исходной игре с конечным числом игроков (и связанными стратегиями). В [63] предложены проверяемые необходимые и достаточные условия равновесия во вспомогательной игре трех лиц, и эти условия позволяют находить точки равновесия в этой (вспомогательной) разрешимой игре

трех лиц, из решения трех задач линейного программирования, две из которых образуют двойственную пару.

Подробное описание содержательной постановки задачи и математической модели, с помощью которой эта задача формализуется в виде игры нескольких лиц на множестве связанных стратегий игроков, содержится в [63], [64].

### 2.1.5 Краткие выводы

В настоящей работе получены следующие научные результаты.

1) Предложена математическая модель для задачи организации переработки твердых бытовых отходов (ТБО) города, на основе которой эта задача сформулирована в виде игры трех лиц на выпуклых многогранниках допустимых (несвязанных) стратегий игроков в конечномерных пространствах. Для этой игры установлены проверяемые необходимые и достаточные условия точек равновесия Нэша, позволяющие отыскивать эти точки из решения трех вспомогательных задач линейного программирования, две из которых образуют двойственную пару.

2) возможность отыскания точек равновесия Нэша на основе использования только линейного программирования установлена для более широкого класса игровых задач административного управления, в частности, для игры трех лиц на (общего вида) выпуклых многогранниках (несвязанных) допустимых стратегий игроков (ранее, такая возможность была установлена в [63], [64] для задачи об использовании альтернативных источников энергии в городских системах электроснабжения, которая формулировалась в игровой постановке как задача на выпуклом многограннике смешанных стратегий игроков, имеющем специальную структуру).

3) показана возможность использования структуры математической модели, предложенной в [66] для формулировки задачи размещения зарядных станций для электромобилей на участке автомагистрали, при математической формулировке задачи оценки целесообразности размещения зарядных станций для электромобилей в крупном городе с позиций потенциального

инвестора. С использованием математической модели указанной структуры задача размещения зарядных станций для электромобилей в крупном городе формулируется в виде игры двух лиц на выпуклом многограннике связанных стратегий игроков. Для этой игры доказана возможность отыскания точек равновесия Нэша из решения задачи квадратичного программирования, задачи линейного программирования и задачи смешанного математического программирования с линейной структурой системы ограничений.

## **2.2 Исследование оптимационной задачи минимизации гарантированного риска**

### **2.2.1 Введение**

Одной из основных причин риска является неполнота информации при принятии решений. Если известны стохастические характеристики неопределенных параметров, задача оптимизации обычно сводится к детерминированной со средними значениями оптимизируемых показателей. Более сложным случаем неполноты информации является неопределенность, когда нет ни точных значений, ни каких-либо стохастических характеристик некоторых параметров, и они известны только с точностью до некоторого множества. Например, Лицу, Принимающему Решения (ЛПР), известны только границы возможных значений неопределенных параметров. Именно этот случай рассмотрен ниже в задаче оптимальной диверсификации депозита по трем видам валют.

В данном исследовании риск понимается в соответствии с [74] (риск по Сэвиджу или сожаление), что может быть истолковано как потеря возможного результата (сожаление) из-за незнания неопределенных значений.

Мы исходим из точки зрения ЛПР, допускающего риск и стремящегося минимизировать его. При этом используем подход наилучшего гарантированного результата, чтобы получить явную форму оптимального решения с минимальным гарантированным риском (решение с минимаксным сожалением по Сэвиджу) в задаче трехвалютной диверсификации депозита.

Другие точки зрения (ЛПР – не допускающие риска и нейтральные) изучались ранее в [75]. В частности, в явном виде было получено решение для наилучшего гарантированного дохода в рассматриваемой задаче с пятью типами валют. Нейтральное ЛПР, которое принимает во внимание как доход, так и риски, стремясь и повысить доход, и уменьшить риск, должно в то же время учитывать, что могут реализоваться любые возможные значения неопределенных параметров. Соответствующая бикритериальная задача в условиях неопределенности изучалась в [75], где был получен явный вид оптимального решения в задаче диверсификации с двумя валютами. Другие многокритериальные задачи в условиях неопределенности подробно изучались в [76], [77], [78]. Различные аспекты многокритериальной оптимизации в условиях неопределенности были исследованы как для статических, так и для динамических случаев в публикациях [79], [80], [81], [82], [83], [84], [85]. В работе [75] исследовалось решение минимаксного сожаления Сэвиджа в задаче диверсификации депозита по двум видам валют.

Данная работа посвящена поиску решения с оптимальным гарантированным риском (по Сэвиджу) в аналогичной задаче с тремя типами валют. Получен явный вид оптимального решения, построен конструктивный алгоритм его вычисления, приведены рекомендации для ЛПР.

### 2.2.2 Постановка проблемы

Очевидно, что оптимальная структура диверсификации депозита на некоторую сумму денег (в рублях) полностью определяется оптимальным размещением одного рубля. В начале некоторого временного интервала (например, года) ЛПР распределяет один рубль между тремя видами депозита, для определенности, в рублях, долларах и евро, стремясь получить максимально возможную ценность результата (в рублях) по окончании срока депозита.

Итак, пусть  $K_d$  и  $K_e$  – курсы доллара и евро по отношению к рублю на начало года,  $(1 - x_d - x_e, x_d, x_e)$  – размеры вкладов в рублях, долларах и евро соответственно (в рублевом выражении). Предполагается, что известны процентные ставки по всем видам вкладов  $r, d_d, d_e$ . Однако ЛПР не знает

точных значений обменных курсов долларов  $y_d$  и euro  $y_e$  в конце периода депозита, и нет доступных статистических характеристик, касающихся их возможных значений. Для этих неопределенных параметров известны только пределы их возможных значений:

$$y_d \in [a_d, b_d], y_e \in [a_e, b_e].$$

Консолидированный результат (доход) в конце года после перевода в рубли зависит как от плана диверсификации  $x = (1 - x_d - x_e, x_d, x_e)$ , так и от обменных курсов на конец периода – неопределенностей  $y = (y_d, y_e) \in Y = [a_d, b_d] \times [a_e, b_e]$ .

Этот результат представляет собой сумму будущих значений различных компонентов депозита после обратного пересчета в рубли. Поэтому его можно представить в следующем виде:

$$f(x, y) = (1 + r)(1 - x_d - x_e) + x_d \frac{1+d_d}{K_d} y_d + x_e \frac{1+d_e}{K_e} y_e. \quad (2.11)$$

Задача ЛПР состоит в нахождении оптимальной стратегии  $x = (x_d, x_e) \in X = \{x_d + x_e \leq 1, x_i \geq 0 (i = d, e)\}$  с целью получения наибольшего результата  $f(x, y)$ .

Однако ЛПР должен учитывать возможность реализации любых значений неопределенности  $y \in Y$ .

Таким образом, математическая модель задачи диверсификации представлена упорядоченной тройкой  $\Gamma = < X, Y, f(x, y) >$ , где  $f(x, y)$  – функция полезности (2.11) депозитора (ЛПР), а множество  $X$  стратегий  $x$  ЛПР и множество  $Y$  неопределенностей  $y$  имеют следующий вид:

$$X = \{x = (x_d, x_e) | x_d + x_e \leq 1, x_i \geq 0 (i = d, e)\},$$

$$Y = \{y = (y_d, y_e) | y_i \in [a_i, b_i] (i = d, e)\}. \quad (2.12)$$

Задача  $\Gamma$  – это однокритериальная задача принятия решений в условиях неопределенности: максимизировать линейную функцию от  $x$  с неопреде-

ленными коэффициентами на многограннике  $X$ , учитывая неопределенные факторы. Наличие неопределенности приводит к концепции риска как возможности отклонения некоторых результатов от их желаемых или ожидаемых значений.

### 2.2.3 Принцип минимаксного сожаления Сэвиджа

Пусть  $f(x, y)$  – целевая функция (доход),  $X$  – множество стратегий ЛПР,  $Y$  – множество неопределенностей. Тогда  $\max_{z \in X} f(z, y)$  будет наилучшим возможным результатом, если будет иметь место неопределенность  $y$ . Однако ЛПР не знает заранее, каким будет значение неопределенного параметра  $y$ . Разница

$$\Phi(x, y) = \max_{z \in X} f(z, y) - f(x, y) \quad (2.13)$$

называется функцией риска (сожаления) по Сэвиджу. Она представляет собой потерю из-за незнания – разницу между наилучшим результатом, который мог бы быть получен при известной неопределенности  $y$ , и реальным результатом с некоторой стратегией  $x$ . Этот риск зависит как от стратегии  $x$ , так и от неопределенности  $y$ .

Пытаясь минимизировать риск, ЛПР может использовать концепцию наилучшего гарантированного результата (принцип Вальда) [86].

Эти соображения приводят к следующему определению (принцип минимаксного сожаления Сэвиджа [74]).

*Определение 2.1. Назовем решение  $x^r$  гарантированным по риску решением (ГРР) Задачи  $\Gamma$ , если*

$$\Phi^r = \max_{y \in Y} \Phi(x^r, y) = \min_{x \in X} \max_{y \in Y} \Phi(x, y), \quad (2.14)$$

где функция риска (потерь)  $\Phi(x, y)$  определена в (2.13).

Отметим некоторые известные свойства:

- в силу определения (2.13) функция риска неотрицательна:

$$\Phi(x, y) \geq 0 \quad \forall x \in X, \forall y \in Y;$$

– если функция  $f(x, y)$  непрерывна, а множества  $X$  и  $Y$  компактны, то ГРР существует.

Операция внутреннего максимума в (2.14) определяет для каждой стратегии  $x \in X$  гарантированный риск  $\Phi[x] = \max_{y \in Y} \Phi(x, y)$ ,  $\Phi[x] \geq \Phi(x, y) \quad \forall y \in Y$ . Операция внешнего минимума в (2.14)  $\min_{x \in X} \Phi[x] = \Phi[x^r] = \Phi^r$  выделяет наилучшую(наименьшую) гарантию, поскольку  $\Phi^r \leq \Phi[x] \quad \forall x \in X$  и  $\Phi^r = \Phi[x^r] \geq \Phi(x, y) \quad \forall y \in Y$ .

Следовательно, ЛПР стремится уменьшить свой риск, выбирая стратегию  $x \in X$  и предполагая при этом возможность реализации любой неопределенности  $y \in Y$ .

Построение ГРР состоит из четырех шагов.

Шаг 1. Вычислить  $f[y] = \max_{z \in X} f(z, y)$  для каждой неопределенности  $y \in Y$ .

Шаг 2. Вычислить функцию риска  $\Phi(x, y) = f[y] - f(x, y)$ .

Шаг 3. Вычислить внутренний максимум в (2.14) – гарантированный риск  $\max_{y \in Y} \Phi(x, y) = \max_{y \in Y} (f[y] - f(x, y)) = \Phi[x] \geq \Phi(x, y)$  для каждой стратегии  $x \in X$ .

Шаг 4. Вычислить внешний минимум в (2.14) – наилучший гарантированный риск

$$\Phi^r = \min_{x \in X} \Phi[x] = \Phi[x^r].$$

В итоге  $x^r$  и будет гарантированным по риску решением Задачи Г.

Рассмотрения этого раздела справедливы для произвольных  $f, X$  и  $Y$ . Для конкретной задачи (2.11) – (2.12) план диверсификации  $(1 - x_d^r - x_e^r, x_d^r, x_e^r)$  будет оптимальным – гарантированным по риску решением, соответствующее  $\Phi^r$  – минимальным гарантированным риском.

## 2.2.4 Построение гарантированного по риску решения

Пусть множества стратегий  $X$  и неопределенностей  $Y$  определены в (2.12), а целевая функция  $f(x, y)$  задана в (2.11). Выполним шаги 1—4 построения ГРР.

Шаг 1. Функция  $f(x, y)$  линейна по переменной  $x = (x_d, x_e)$ . Поэтому при фиксированном  $y \in Y$  она достигает максимального значения на многоугольнике  $X$  и в одной из его вершин  $(0,0)$ ,  $(1,0)$  или  $(0,1)$ . С учетом (2.11) имеем:

$$f[y] = \max_{z \in X} f(z, y) = \max\{(1+r), \frac{1+d_d}{K_d}y_d, \frac{1+d_e}{K_e}y_e\}. \quad (2.15)$$

Шаг 2. Функция риска принимает вид:

$$\left\{ \begin{array}{l} \Phi_1(x, y) = (1+r) - f(x, y) = \\ = \left[1 + r - \frac{1+d_d}{K_d}y_d\right]x_d + \left[1 + r - \frac{1+d_e}{K_e}y_e\right]x_e, \quad y \in Y_1, \\ \Phi_2(x, y) = \frac{1+d_d}{K_d}y_d - f(x, y) = \\ = (1-x_d)\left[\frac{1+d_d}{K_d}y_d - (1+r)\right] + \left[1 + r - \frac{1+d_e}{K_e}y_e\right]x_e, \quad y \in Y_2, \\ \Phi_3(x, y) = \frac{1+d_e}{K_e}y_e - f(x, y) = \\ = \left[1 + r - \frac{1+d_d}{K_d}y_d\right]x_d + \left[\frac{1+d_e}{K_e}y_e - (1+r)\right](1-x_e), \quad y \in Y_3, \end{array} \right. \quad (2.16)$$

где  $Y_1, Y_2, Y_3$  суть подмножества множества  $Y$ , определяемые линейными неравенствами:

$$Y_1 = \left\{ y \in Y, \quad 1+r \geq \frac{1+d_d}{K_d}y_d, \quad 1+r \geq \frac{1+d_e}{K_e}y_e \right\},$$

$$Y_2 = \left\{ y \in Y, \quad \frac{1+d_d}{K_d}y_d \geq 1+r, \quad \frac{1+d_d}{K_d}y_d \geq \frac{1+d_e}{K_e}y_e \right\},$$

$$Y_3 = \left\{ y \in Y, \quad \frac{1+d_e}{K_e}y_e \geq 1+r, \quad \frac{1+d_e}{K_e}y_e \geq \frac{1+d_d}{K_d}y_d \right\}.$$

Очевидно, что  $Y_1 \cup Y_2 \cup Y_3 = Y$ , некоторые из множеств  $Y_1, Y_2, Y_3$  могут быть пустыми. Функции  $\Phi_i(x, y)$  ( $i = 1, 2, 3$ ) являются билинейными функциями переменных  $x$  и  $y$ . Поэтому функция риска  $\Phi(x, y)$  при любой фиксированной стратегии  $x$  является кусочно-линейной функцией неопределенности  $y$ . Множества  $Y_1, Y_2, Y_3$  являются областями (многоугольниками) линейности по  $y$  функции риска. Иными словами, функция риска  $\Phi(x, y)$  совпадает с линейной (по  $x$ ) функцией  $\Phi_i(x, y)$  при любом фиксированном  $y \in Y_i$ . В этом случае говорят, что функция риска  $\Phi(x, y)$  задана в точке  $(x, y)$  функцией  $\Phi_i(x, y)$ .

Шаг 3. Для каждой стратегии  $x \in X$  вычисляется гарантированный риск  $\Phi[x]$ .

В силу естественных предположений об исходных параметрах рассматриваемой задачи  $r > 0, d_i > 0, b_i > a_i > 0, K_i > 0$  ( $i = d, e$ ) получаем:

$$\Phi[x] = \max_{y \in Y} \Phi(x, y) =$$

$$= \max_{y \in Y} \begin{cases} \left[1 + r - \frac{1 + d_d}{K_d} y_d\right] x_d + \left[1 + r - \frac{1 + d_e}{K_e} y_e\right] x_e, & y \in Y_1, \\ \left[\frac{1 + d_d}{K_d} y_d - (1 + r)\right] (1 - x_d) + \left[1 + r - \frac{1 + d_e}{K_e} y_e\right] x_e, & y \in Y_2, \\ \left[1 + r - \frac{1 + d_d}{K_d} y_d\right] x_d + \left[\frac{1 + d_e}{K_e} y_e - (1 + r)\right] (1 - x_e), & y \in Y_3, \end{cases}$$

или

$$\Phi[x] = \max_{y \in Y} \Phi(x, y) = \max_{i=1,2,3} \Phi_i[x], \quad (2.17)$$

где

$$\Phi_1[x] = \left[1 + r - \frac{1 + d_d}{K_d} a_d\right] x_d + \left[1 + r - \frac{1 + d_e}{K_e} a_e\right] x_e = \Phi_1(x, a_d, a_e),$$

$$\begin{aligned}\Phi_2[x] &= \left[ \frac{1+d_d}{K_d} b_d - (1+r) \right] (1-x_d) + \left[ 1+r - \frac{1+d_e}{K_e} a_e \right] x_e \\ &= \Phi_2(x, b_d, a_e),\end{aligned}$$

$$\begin{aligned}\Phi_3[x] &= \left[ 1+r - \frac{1+d_d}{K_d} a_d \right] x_d + \left[ \frac{1+d_e}{K_e} b_e - (1+r) \right] (1-x_e) \\ &= \Phi_3(x, a_d, b_e).\end{aligned}$$

Введем для краткости записи следующие обозначения:

$$\begin{aligned}\alpha_d &= \left[ 1+r - \frac{1+d_d}{K_d} a_d \right], \quad \alpha_e = \left[ 1+r - \frac{1+d_e}{K_e} a_e \right], \\ \beta_d &= \left[ \frac{1+d_d}{K_d} b_d - (1+r) \right], \quad = \left[ \frac{1+d_e}{K_e} b_e - (1+r) \right].\end{aligned}\tag{2.18}$$

С учетом этих обозначений формула (2.17) принимает вид:

$$\begin{aligned}\Phi[x] &= \max\{\Phi_1[x], \Phi_2[x], \Phi_3[x]\} = \\ &= \max\{\alpha_d x_d + \alpha_e x_e, \beta_d (1-x_d) + \alpha_e x_e, \alpha_d x_d + \beta_e (1-x_e)\}, x \in X.\end{aligned}\tag{2.19}$$

Шаг 4. Вычисляется наилучший гарантированный риск  $\Phi^r = \min_{x \in X} \Phi[x] = \Phi[x^r]$  путем рассмотрения задачи о  $\min_{x \in X} \Phi[x]$  отдельно на границе и на внутренности множества (треугольника)  $X$  с последующим выбором итогового решения. Формально:

$$\min_{x \in X} \Phi[x] = \min\{\min_{x \in [0,A]} \Phi[x], \min_{x \in [0,B]} \Phi[x], \min_{x \in [A,B]} \Phi[x], \min_{x \in \text{int}X} \Phi[x]\}.\tag{2.20}$$

Случай 1 (катет ОА треугольника АОВ):  $[O,A] = \{x_d \in [0,1], x_e = 0\}$ , стратегия  $x = (x_d, 0)$ , и  $\varphi(x_d) = \Phi[x_d, 0] = \max\{\alpha_d x_d, \beta_d (1-x_d), \alpha_d x_d + \beta_e\} = \max\{\varphi_1(x_d), \varphi_2(x_d), \varphi_3(x_d)\}$ .

Заметим, что  $\varphi_3(x_d) - \varphi_1(x_d) = \beta_e = \text{const}$ .

Подслучай 1а. Пусть  $\beta_e \geq 0$ , тогда  $\varphi_3 \geq \varphi_1$ , и функция  $\varphi_1$  может быть удалена из определения функции  $\varphi$  как несущественная. График функции  $\varphi$

на отрезке  $[0,1]$  состоит из одного или двух отрезков. Минимум  $\varphi$  должен достигаться в точке  $A=(1,0)$ , или  $O=(0,0)$ , или в точке  $x_d^{OA}$  пересечения графиков функций  $\varphi_2$  и  $\varphi_3$  (при условии, что  $x_d^{OA} \in [0,1]$ ). В последнем случае из уравнения  $\beta_d(1 - x_d) = \alpha_d x_d + \beta_e$  получаем выражение для третьей возможной точки частного минимума на гипотенузе  $[O,A]$ :

$$x_d^{OA} = \frac{\beta_d - \beta_e}{\alpha_d + \beta_d}.$$

Необходимое условие  $0 \leq x_d^{OA} \leq 1$  эквивалентно следующему соотношению между параметрами задачи:  $a_d \leq \frac{(1+d_e)}{(1+d_d)} \frac{K_d}{K_e} \leq b_d$ .

Подслучай 1b. Если  $\beta_e \leq 0$ , то  $\varphi_3 \leq \varphi_1$ , и функция  $\varphi_3$  может быть исключена из рассмотрения. График  $\varphi$  на  $[0, 1]$  состоит из одного или двух сегментов. Минимум функции  $\varphi$  на отрезке  $[O,A]$  должен достигаться в точке  $A=(1,0)$ , или  $O=(0,0)$ , или в точке  $x_d^{OA}$  пересечения графиков  $\varphi_1$  и  $\varphi_2$  (при условии, что  $x_d^{OA} \in [0,1]$ ). Соответствующее уравнение  $\alpha_d x_d = -\beta_d x_d + \beta_d$  дает

$$x_d^{OA} = \frac{\beta_d}{\alpha_d + \beta_d}.$$

Необходимое условие  $0 \leq x_d^{OA} \leq 1$  эквивалентно следующему соотношению между параметрами задачи:  $a_d \leq \frac{1+r}{1+d_d} K_d \leq b_d$ .

Обе формулы можно объединить в одно выражение:

$$x_d^{OA} = \frac{\beta_d - \beta_e (\text{sign} \beta_e + 1)/2}{\alpha_d + \beta_d}.$$

Дальнейшее исследование того, какой из трех «кандидатов» действительно является точкой максимума  $\Phi[x]$  на отрезке  $[OA]$ , может быть выполнено путем изучения увеличения и уменьшения функций  $\varphi_1, \varphi_2, \varphi_3$ . Другой возможный подход – вычислить и сравнить значения функции  $\Phi[x]$ . Это сравнение может быть произведено после подсчета всех «кандидатов» в ГРР.

Утверждение 2.1. Функция гарантированного риска  $\Phi[x]$  достигает минимального на отрезке  $[OA]$  значения в некоторой точке из множества

$$[0,1] \cap \{O = (0,0), A = (1,0), x^{OA} = (x_d^{OA}, 0)\},$$

$$\text{где } x_d^{OA} = \begin{cases} \frac{\beta_d - \beta_e}{\alpha_d + \beta_d}, & \text{если } \beta_e \geq 0, \\ \frac{\beta_d}{\alpha_d + \beta_d}, & \text{если } \beta_e \leq 0. \end{cases}$$

Иначе говоря, если существует ГРР вида  $x^r = (x_d, x_e = 0)$ , то множество точек  $[0,1] \cap \{O = (0,0), A = (1,0), x^{OA} = (x_d^{OA}, 0)\}$  содержит ГРР.

Случай 2 (катет ОВ:  $[OB] = \{x_d = 0, x_e \in [0,1]\}$ ), стратегия  $x = (0, x_e)$ ,

$$\psi(x_e) = \Phi[0, x_e] = \max\{\alpha_e x_e, \beta_e(1 - x_e), \alpha_e x_e + \beta_d\} \quad (2.21)$$

В этом случае рассуждения аналогичны предыдущему. Приведем соответствующие результаты.

Подслучай 2a. Если  $\beta_d \geq 0$ , то минимум функции  $\psi(x_e)$  на  $[OB]$  будет в точке  $B=(0,1)$ , или  $O=(0,0)$ , или  $x_e^{OB} = \frac{\beta_e - \beta_d}{\alpha_e + \beta_e}$ , при условии, что  $x_e^{OB} \in [0,1]$ .

Условие  $0 \leq x_e^{OB} \leq 1$  эквивалентно неравенствам:  $a_e \leq \frac{(1+d_d)}{(1+d_e)} K_e \leq b_e$ .

Подслучай 2b. Если  $\beta_d \leq 0$ , то минимум функции  $\psi(x_e)$  на  $[OB]$  будет в точке  $B=(0,1)$ , или  $O=(0,0)$ , или  $x_e^{OB} = \frac{\beta_e}{\alpha_e + \beta_e}$ , при условии, что  $x_e^{OB} \in [0,1]$ .

Условие  $0 \leq x_e^{OB} \leq 1$  эквивалентно неравенствам:  $a_e \leq \frac{1+r}{1+d_e} K_e \leq b_e$ .

Здесь также обе формулы могут быть объединены одним выражением:

$$x_e^{OB} = \frac{\beta_e - \beta_d(\text{sign}(\beta_d) + 1)/2}{\alpha_e + \beta_e}.$$

Дальнейший выбор будет сделан после подсчета всех «кандидатов» в ГРР.

Утверждение 2.2. Функция гарантированного риска  $\Phi[x]$  достигает минимального на отрезке  $[OB]$  значения в некоторой точке из множества

$$[0,1] \cap \{O = (0,0), B = (0,1), x^{OB} = (0, x_e^{OB})\},$$

$$\text{где } x_e^{OB} = \begin{cases} \frac{\beta_e - \beta_d}{\alpha_e + \beta_e}, & \text{if } \beta_d \geq 0, \\ \frac{\beta_e}{\alpha_e + \beta_e}, & \text{if } \beta_d \leq 0. \end{cases}$$

Иначе говоря, если существует ГРР вида  $x^r = (x_d = 0, x_e)$ , то множество точек  $[0,1] \cap \{O = (0,0), B = (0,1), x^{OB} = (0, x_e^{OB})\}$  содержит ГРР.

Случай 3 (гипотенуза треугольника АОВ):  $[A,B] = \{x_d + x_e = 1, x_d \geq 0, x_e \geq 0\}$ , стратегия  $x = (1 - x_d, x_e)$ , функция гарантированного риска (2.20) имеет вид:

$$\chi(x_e) = \Phi[1 - x_e, x_e] = \max\{\chi_1(x_e), \chi_2(x_e), \chi_3(x_e)\}, \quad (2.22)$$

где  $x_e \in [0,1]$  и

$$\chi_1(x_e) = \Phi_1[1 - x_e, x_e] = (\alpha_e - \alpha_d)x_e + \alpha_d,$$

$$\chi_2(x_e) = \Phi_2[1 - x_e, x_e] = (\beta_d + \alpha_e)x_e,$$

$$\chi_3(x_e) = \Phi_3[1 - x_e, x_e] = (\alpha_d + \beta_e)(1 - x_e) = -(\alpha_d + \beta_e)x_e + (\alpha_d + \beta_e).$$

Пусть  $l_i$  – график линейной функции  $\chi_i$  ( $i = 1, 2, 3$ ). Функция  $\chi(x_e)$  является выпуклой (вниз) кусочно-линейной функцией одной переменной  $x_e \in [0,1]$ . Её график является нижней огибающей семейства функций  $\{l_1, l_2, l_3\}$  и содержит один, два или три отрезка.

Такая функция достигает минимума на  $[0,1]$  в конце отрезка или в точке пересечения некоторых пар прямых  $l_1, l_2, l_3$ . Следовательно, «кандидатов» в возможные точки минимума функции  $\chi(x_e)$  можно найти из следующих уравнений:

$$\chi_1(x_e) = \chi_2(x_e), \quad \chi_1(x_e) = \chi_3(x_e), \quad \chi_2(x_e) = \chi_3(x_e).$$

Решения этих уравнений соответственно:

$$x_e^{12} = \frac{\alpha_d}{\alpha_d + \beta_d}, \quad x_e^{13} = \frac{\beta_e}{\alpha_e + \beta_e}, \quad x_e^{23} = \frac{\alpha_d + \beta_e}{\alpha_e + \alpha_d + \beta_e + \beta_d}. \quad (2.23)$$

Решения вне отрезка  $[0, 1]$  должны быть исключены. Оставшиеся точки из множества (2.23) вместе с двумя внутренними точками — «кандидатами» из случаев 1 и 2 и тремя вершинами треугольника  $X$  составляют «достаточное» множество возможных точек минимума функции  $\Phi[x]$  на границе множества  $X$ .

Утверждение 2.3. Функция гарантированного риска  $\Phi[x]$  достигает минимального на отрезке  $[AB]$  значения в некоторой точке из множества  $\{(0,1), (1,0), x^{12} = (x_e^{12}, 1 - x_e^{12}), x^{13} = (x_e^{13}, 1 - x_e^{13}), x^{23} = (x_e^{23}, 1 - x_e^{23})\} \cap [0,1]$ .

Случай 4 (внутренность треугольника АOB):  $intX = \{x_d + x_e < 1, x_d, x_e > 0\}$ , стратегия  $x = (x_d, x_e)$ , функция гарантированного риска (2.20) имеет вид:

$$\begin{aligned}\Phi[x] &= \max\{\Phi_1[x], \Phi_2[x], \Phi_3[x]\} = \\ &= \max\{\alpha_d x_d + \alpha_e x_e, \beta_d(1 - x_d) + \alpha_e x_e, \alpha_d x_d + \beta_e(1 - x_e)\}, x \in intX.\end{aligned}\quad (2.24)$$

Будем говорить о полной диверсификации, если функция гарантированного риска  $\Phi[x]$  достигает минимального значения  $\Phi^*$  на множестве  $X$  в некоторой внутренней точке  $X$ , и для всех точек границы  $frX$   $\Phi[x] > \Phi^*$ .

Это условие означает, что оптимальный план диверсификации обязательно использует все три валюты. Далее в этом пункте ограничимся этим условием, поскольку другие возможные ситуации охватываются утверждениями 2.1–2.3.

Гарантированный риск  $\Phi[x]$  является выпуклой непрерывной функцией и может достигать минимума в некоторой внутренней точке выпуклого множества  $X$ . Этот минимум также будет глобальным минимумом для  $X$ . Функцию  $F[x] = \Phi[x]$ ,  $x \in intX$ , определенную предыдущими формулами для  $\Phi[x]$ , можно рассматривать на всем пространстве  $\mathbf{R}^2$ . Любая точка  $x^\wedge$  (глобального) минимума  $F[x]$  on  $\mathbf{R}^2$  (если она существует) является точкой минимума на  $intX$ , если  $x^\wedge \in intX$ .

Пусть  $P_1$ ,  $P_2$  и  $P_3$  – двумерные плоскости, определяемые уравнениями  $z = \Phi_i(x)$ ,  $x \in \mathbb{R}^2$  ( $i = 1, 2, 3$ ). График функции риска представляет собой нижнюю огибающую семейства плоскостей  $\{P_1, P_2, P_3\}$ , точнее, ее часть, расположенную над треугольником  $X$ . Надграфик функции  $\Phi[x]$  при  $x \in X$  есть перевернутое обелискообразное трехмерное тело. Вычисление минимума функции  $\Phi[x]$  на  $X$  равносильно нахождению наименшей точки этого тела. Эта геометрическая интерпретация позволяет упростить нашу задачу о минимизации негладкой функции  $\Phi[x]$  при ограничениях типа равенств и неравенств. Именно, она подсказывает, что строгий минимум функции  $\Phi[x]$  на  $X$  может быть в точке  $x^*$  пересечения всех трех плоскостей  $P_1, P_2, P_3$ .

Так как граница множества  $X$  исследована выше, займемся внутренностью  $X$ .

Предположим, что  $\Phi[x] > \min_{x \in X} \Phi[x] = \Phi^*$  для всех точек границы  $frX$ .

В этом случае непрерывная выпуклая неотрицательная функция  $\Phi[x]$  имеет строгий глобальный минимум в некоторой внутренней точке множества  $X$ . Нам нужно указать всех возможных «кандидатов» для этого решения.

**Лемма 2.1.** Пусть для некоторого  $i \in \{1, 2, 3\}$   $\Phi_i[x] \neq const$  и в некоторой внутренней точке  $x^0 \in intX$   $\Phi_i(x^0) > \Phi_j(x^0) \wedge \Phi_i(x^0) > \Phi_k(x^0)$  ( $j \neq i, k \neq i$ ). Иначе говоря, функция  $\Phi[x]$  определяется в точке  $x^0$  только функцией  $\Phi_i[x]$ . Тогда  $x^0$  не является точкой минимума  $\Phi[x]$  на  $X$ .

Действительно, в силу непрерывности функций  $\Phi_i[x]$  данные неравенства будут выполняться и в некоторой окрестности точки  $x^0$ . Поэтому в этой окрестности функция риска совпадает с линейной функцией ( $\Phi[x] = \Phi_i(x) \neq const$ ) и не может иметь экстремум (даже локальный) во внутренней точке  $x^0$ .

Далее, если какая-либо из функций  $\Phi_i(x) = const = A$  и в некоторой внутренней точке  $x^0$  только она определяет функцию  $\Phi[x]$ , то очевидно, что минимум функции риска на  $X$  равен  $A$  и он достигается и на границе. Исключаем этот случай из дальнейшего рассмотрения. Тогда, в соответствии с

Леммой, минимум функции  $\Phi[x]$  на множестве  $int X$  может быть только на пересечении некоторой пары плоскостей  $P_i (i = 1,2,3)$ . Поэтому далее рассматриваем только те внутренние точки множества стратегий  $X$ , в которых функция риска определяется более чем одной функцией  $\Phi_i$ :  $\Phi_1[x] = \Phi_2[x] \vee \Phi_1[x] = \Phi_3[x] \vee \Phi_2[x] = \Phi_3[x]$ . Обозначим соответствующие множества через  $L_1, L_2, L_3$ . Это в силу линейности функций  $\Phi_i$  открытые интервалы (возможно, пустые).

Рассмотрим функцию  $\Phi[x]$  на интервале  $L_1$ , где  $\Phi_1[x] = \Phi_2[x]$ . Тогда возможны две ситуации.

1)  $\Phi_1[x] = \Phi_2[x] > \Phi_3[x]$  для всех  $x \in L_1$  (функция  $\Phi_3$  несущественна). Если на открытом интервале  $L_1$  линейная функция  $\Phi[x] = \Phi_1[x] = const$ , то она достигает минимума и на границе  $int X$ . Эта точка минимума будет обнаружена при исследовании границы множества  $X$ . Если же на открытом интервале  $L_1$  линейная функция  $\Phi[x] = \Phi_1[x] \neq const$ , то она не имеет точек минимума на  $L_1$ . Следовательно, внутренних точек минимума в этом случае нет.

2)  $\Phi_1[x^*] = \Phi_2[x^*] = \Phi_3[x^*]$  для некоторой точки  $x^* \in L_1$ . Эта точка соответствует пересечению плоскостей  $P_1, P_2, P_3$  – графиков функций  $z = \Phi_i(x) (i = 1,2,3)$ . Единственность такой точки при естественных предположениях о параметрах будет установлена ниже.

Итак, внутренняя точка минимума функции гарантированного риска должна удовлетворять следующей системе линейных равенств и неравенств:

$$\Phi_1[x] = \Phi_2[x], \Phi_3[x] = \Phi_2[x], x_d + x_e < 1, x_d > 0, x_e > 0. \quad (2.25)$$

С учетом выражений для функций  $\Phi_i(x) (i = 1,2,3)$  подсистема линейных равенств в (2.25) принимает вид:

$$\alpha_d x_d + \alpha_e x_e = \beta_d (1 - x_d) + \alpha_e x_e ,$$

$$\alpha_d x_d + \beta_e (1 - x_e) = \beta_d (1 - x_d) + \alpha_e x_e .$$

Ее решение

$$x^* = (x_d^*, x_e^*) = \left( \frac{\beta_d}{\alpha_d + \beta_d}, \frac{\beta_e}{\alpha_e + \beta_e} \right) \quad (2.26)$$

при выполнении неравенств в (2.25) является точкой – «кандидатом» в ГРР. Подчеркнем:  $x^*$  может, но не обязана быть точкой минимума на  $\text{int}X$ .

В силу обозначений (2.8)  $\alpha_d + \beta_d > 0, \alpha_e + \beta_e > 0$  при естественных предположениях  $b_d > a_d, b_e > a_e$ . Поэтому решение системы линейных уравнений в (2.15) существует, единственно и определяется формулой (2.26).

**Утверждение 2.4.** Если функция  $\Phi[x]$  достигает минимума на открытом множестве  $\text{int}X$ , то точка  $x^* = (x_d^*, x_e^*) = \left( \frac{\beta_d}{\alpha_d + \beta_d}, \frac{\beta_e}{\alpha_e + \beta_e} \right)$  есть точка минимума функции  $\Phi[x]$  на множестве  $\text{int}X$ , при условии, что  $0 < x_d, x_e < 1$  и  $x_d + x_e < 1$ .

Напомним, что любая точка локального минимума выпуклой функции на выпуклом множестве одновременно является точкой глобального (на этом множестве) минимума. Следовательно, если такая точка внутреннего минимума (2.25) существует и  $x^* = (x_d^*, x_e^*) \in \text{int}X$ , это дает ГРР. Существует она или нет, зависит от взаимосвязи параметров рассматриваемой задачи. Очень часто ГРР оказывается на границе множества планов диверсификации. Это означает, что оптимальный план диверсификации содержит только две (или даже одну) из трех валют.

Неравенства для  $(x_d^*, x_e^*)$  в (2.25) в терминах исходных параметров принимают вид:

$$\begin{aligned} \frac{1+d_i}{K_i} a_i &< (1+r) < \frac{1+d_i}{K_i} b_i \quad (i = d, e), \\ \left[ \frac{1+d_d}{K_d} b_d - (1+r) \right] \left[ \frac{1+d_e}{K_e} b_e - (1+r) \right] &< \\ &< \left[ \frac{1+d_d}{K_d} a_d - (1+r) \right] \left[ \frac{1+d_e}{K_e} a_e - (1+r) \right]. \end{aligned}$$

Подчеркнем еще раз, что  $x^*$  только “кандидат”, но не обязательно точка минимума функции гарантированного риска.

### 2.2.5 Алгоритм построения гарантированного по риску решения

Объединение результатов утверждений 2.1–2.4 приводит к следующей схеме вычисления гарантированной по риску стратегии диверсификации и наименьшего гарантированного риска.

Шаг 1. Записать числовые значения процентных ставок  $r, d_d, d_e$  и текущих обменных курсов  $K_d, K_e$ . Установить численные значения границ возможных изменений (через год) курсов доллара  $[a_d, b_d]$  и евро  $[a_e, b_e]$ . При этом может использоваться любая доступная информация – оценки и прогнозы экспертов, собственные предположения ЛПР и т.д.

Шаг 2. Вычислить вторичные параметры  $\alpha_d, \alpha_e, \beta_d, \beta_e$ , определенные в (2.18).

Шаг 3. Вычислить девять точек-кандидатов на оптимальное решение. Соответствующие формулы для этих точек представлены в приведенной ниже Таблице 2.1.

Шаг 4. Удалить точки, не принадлежащие множеству допустимых стратегий  $X$ .

Шаг 5. Вычислить значения гарантированного риска в точках, оставшихся после предыдущего шага, выбрать решение с наилучшим (минимальным) гарантированным риском.

Предложенный алгоритм реализован с использованием Excel.

Таблица 2.1 – Кандидаты в решения и проверка условий их допустимости

	Кандидаты в решения	Формулы	Проверка условий
11	Вершина О=(0,0) (только рубли)	(0,0)	Не требуется
22	Вершина А=(1,0) (только доллары)	(1,0)	Не требуется
33	Вершина В=(0,1) (только евро)	(0,1)	Не требуется
44	Точка $(x_d^{OA}, 0)$ катета (ОА) (рубли и доллары)	$x_d^{OA} = \frac{\beta_d - \beta_e(\text{sign}\beta_e + 1)/2}{\alpha_d + \beta_d}$	$0 < x_d^{OA} < 1$
55	Точка $(0, x_e^{OB})$ катета(ОВ) (рубли и евро)	$x_e^{OB} = \frac{\beta_e - \beta_d(\text{sign}\beta_d + 1)/2}{\alpha_d + \beta_d}$	$0 < x_e^{OB} < 1$
66	Точка $(1 - x_e^{12}, x_e^{12})$ гипотенузы (АВ) (доллары и евро)	$x_e^{12} = \frac{\alpha_d}{\alpha_d + \beta_d}$	$0 < x_e^{12} < 1$

## Продолжение таблицы 2.1

	Кандидаты в решения	Формулы	Проверка условий
77	Точка $(1 - x_e^{13}, x_e^{13})$ гипотенузы (AB) (доллары и евро)	$x_e^{13} = \frac{\beta_e}{\alpha_e + \beta_e}$	$0 < x_e^{13} < 1$
88	Точка $(1 - x_e^{23}, x_e^{23})$ гипотенузы (AB) (доллары и евро)	$x_e^{23} = \frac{\alpha_d + \beta_e}{\alpha_e + \alpha_d + \beta_e + \beta_d}$	$0 < x_e^{23} < 1$
99	Внутренняя точка в X (рубли, доллары и евро)	$(x_d^*, x_e^*) = \\ = \left( \frac{\beta_d}{\alpha_d + \beta_d}, \frac{\beta_e}{\alpha_e + \beta_e} \right)$	$0 < x_d, x_e < 1$ $x_d + x_e < 1$

### 2.2.6 Некоторые выводы

Рассмотрена задача оптимальной диверсификации мультивалютных депозитов с неопределенными будущими обменными курсами как проблема минимизации упущеной из-за неопределенности выгоды. Предполагалось, что известны только пределы возможных изменений этих неопределенных параметров, а какие-либо статистические характеристики недоступны. Использовалась концепция минимаксного сожаления Сэвиджа. В соответствии с ней исходная проблема была сведена к минимаксной задаче с упущеной выгодой в качестве целевой функции (функции риска). Следует отметить, что критерий Сэвиджа сложнее, чем более часто используемый максиминный критерий Вальда. Соответственно усложняются математические конструкции. Тем не менее в рассматриваемой задаче удалось найти функции риска и гарантированного риска в явном виде. После этого задача свелась к нахождению точки минимума кусочно-линейной функции при линейных ограничениях. Для ее решения была применена следующая методика.

Нахождение минимума на множестве всех допустимых планов диверсификации выполнялось в два этапа: на границе и внутри этого множества. Найдены явные формулы для девяти «репрезентативных» точек – кандидатов в оптимальное решение. Окончательный выбор делается прямым сравнением значений критерия в этих точках. Предлагаемый алгоритм реализован в виде расчетной таблицы с использованием Excel.

При определенных соотношениях параметров оптимальное решение не единственно. Не останавливаясь более подробно на этом вопросе, отметим

только, что в случае неединственности можно рассматривать дополнительный критерий (например, гарантированный доход). Это приводит к соответствующим задачам многокритериальной оптимизации в условиях неопределенности. Некоторые из них (для случая двух валют) рассматривались в [75].

Получен полный набор аналитических формул для вычисления оптимального решения. Эти результаты могут быть использованы в дальнейшем при анализе проблем управления финансами в условиях неполной информации.

## **2.3 Модификация метода расчёта приоритетов стратегических целей**

### **2.3.1 Введение**

В современном мире наличие стратегических целей является обязательным условием даже не развития, а выживания компании. Если цели не определены, то нельзя объективно оценить эффективность работы. Также становится непонятным какие проекты следует выполнять. Все признанные международные стандарты управления проектами устанавливают правило, что реализуемые проекты должны соответствовать целям [87], [88], [89]. Стратегические цели не равнозначны. Для правильного распределения ресурсов их надо ранжировать, или, по-другому, необходимо правильно установить приоритеты. Это сложная задача, которая не может быть решена простыми методами. На основе опыта выполненных консалтинговых проектов, наиболее эффективным при приоритизации целей является подход Т. Саати [90]. Однако, применение этого подхода будет оправдано только в том случае, если консультант имеет в своём распоряжении удобный программный инструмент, осуществляющий высокоточный расчёт. Причём метод расчёта не должен быть скрытым в «чёрном ящике» программы. Такой метод найден для расчёта приоритета целей по таблице парных сравнений их значимости. Разработана программа, в которой этот метод реализован. Программа компактная и имеет дружелюбный интерфейс.

### **2.3.2 Постановка проблемы**

Непосредственная оценка экспертами является исторически первым и наиболее применяемым методом ранжирования целей. Эксперту предлагается оценить степень важности (значимости) каждой цели некоторым количеством баллов. По сумме баллов осуществляется упорядочивания целей по степени важности. Процедура упорядочивания может быть усложнена приятием различных весов оценкам различных экспертов (ранжирование экспертов) и определением степени согласованности их оценок. Во всех случаях подход непосредственного ранжирования имеет ряд недостатков. Остановимся на тех, которые представляются основными. Цели формулируются на качественном лингвистическом уровне. Человеку предлагается однозначно оценивать данные качественные характеристики числами. Это не соответствует нашему восприятию достаточно сложных явлений. Проблема усугубляется тем, что значность шкалы определяет не эксперт, а лицо, заказавшее экспертизу. Если шкала имеет малую значность, то много целей получат одинаковую оценку. Соответственно, количественный подход даёт ничтожный результат. Лицу, который должен принять решение, придётся обратиться к другим, зачастую исключительно субъективным, критериям оценки. Если же значность шкалы высока, то эксперту трудно адекватно отразить различие своих оценок. Вспоминаются предложения оценивать знания школьников, причём начиная с младших классов, по 100 бальной шкале. Как может учитель определить, кто из детей нарисовал на 80, а кто на 85 баллов?

Эти проблемы вызваны тем, что традиционный подход непосредственного ранжирования не соответствует реальной сложности задачи приоритизации целей. Приоритизация – это выбор из  $n!$  вариантов, где  $n$  – количество целей. Она не может быть решена «арифметическими методами». Из работ Н. Винера известно, что простых методов решения сложных задач не существует, или, как ещё говорят специалисты: «У каждой сложной задачи есть простой неверный способ решения». Другим подходом является методология, предложенная недавно ушедшим от нас Томасом Саати – великим математиком, работавшим и над задачами менеджмента. Эта методология вклю-

чает в себя метод анализа иерархий (AHP) и метод анализа сетей (ANP). Они разработаны для приоритизации различных возможных решений по достижению поставленных целей. При этом ANP является развитием AHP, позволяющим учитывать зависимости между целями. Имеется пакет SuperDecision, реализующий AHP и ANP [91]. Метод AHP реализован и в современных версиях MS – Project Server [92]. Пакет SuperDecision не слишком удобен в использовании из-за недружелюбного интерфейса, а MS – Project Server большая программа, использовать которую при стратегическом консалтинге затруднительно.

### Модифицированный метод приоритизации проектов на основе АНР

Основная идея подходов Саати состоит в том, что эксперт должен не пытаться упорядочить весь список целей, а осуществлять попарное сравнение целей. После чего упорядочение может быть осуществлено с помощью релевантной математической процедуры. При этом, желательно дать возможность эксперту выражать результаты сравнения не с помощью чисел, а на речевом – лингвистическом уровне. Исследования психологов показали, что точность экспертных оценок, выраженных на лингвистическом уровне, намного выше, чем при непосредственном использовании экспертами числовых баллов.

Возникают следующие вопросы:

- 1) сколько степеней превосходства использовать или, что тоже, какой должна быть значность шкалы сравнения при применении лингвистических переменных?
- 2) как перейти от парных сравнений к упорядочению всего списка?

В качестве ответа на первый из них, Т. Саати предложил использовать девятизначную шкалу, имеющую значения симметрично противоположные относительно значения «Эквивалентны». Он назвал её фундаментальной шкалой, желая подчеркнуть универсальность. Как это часто бывает в случаях, когда мы имеем дело с языками, выяснилось, что для русскоязычных экспертов, и не только, более удобной является семизначная шкала, в которой

нет промежуточного значения между значениями «Намного более важно» и «Несравненно более важно» и, соответственно, нет промежуточного значения между противоположными им значениями «Намного менее важно» и «Несравненно менее важно». Ответом на второй вопрос стало доказательство того, что приоритеты целей можно рассчитать, как собственный вектор, соответствующий главному собственному значению матрицы парных сравнений. Естественно, перед расчётом лингвистические переменные должны быть заменены на соответствующие числовые значения. В работах Т. Саати предложен простой метод приближённого расчёта приоритетов целей по матрице парных сравнений, который даёт очень значительную погрешность, во многих случаях влияющую на результат упорядочивания целей [93]. Следующие ему разработчики SuperDecision, и разработчики MS – Project Server не раскрывают того, как они реализовали численный расчёт приоритетов целей. В этих условиях, было принято решение реализовать процедуру максимально точного расчёта приоритетов целей, по таблице сравнения с семизначной шкалой. Основой этой процедуры является ускоренный метод расчёта наибольшего собственного значения матрицы и соответствующего собственного вектора [94].

Пронумеруем цели компании числами от 1 до  $n$ . Запишем цели в строки и столбцы квадратной таблицы размерности  $n \times n$ . На пересечении строки  $i$  и столбца  $j$  ( $j > i$ ) запишем результат важности цели  $i$  по сравнению с целью  $j$ . При этом будем использовать следующие значения:

- 1) несравненно более важно;
- 2) намного более важно;
- 3) более важно;
- 4) эквивалентны;
- 5) менее важно;
- 6) намного менее важно;
- 7) несравненно менее важно.

Преобразуем таблицу в числовую форму. Для этого заменим в заполненных клетках указанные значения на числа: 7, 5, 3, 1, 1/3, 1/5, 1/7; соответственно. Заполним главную диагональ единицами. Заполним пустые клетки так, чтобы получить обратно симметричную матрицу, то есть, если  $a_{ij} = r$ , то  $a_{ji} = \frac{1}{r}$ . Полученную матрицу обозначим через  $A$ .

Для нахождения приоритетов целей применим следующую процедуру.

- 1) Положим  $k = 1$ .
- 2) Возведём матрицу  $A$  в степень  $m = 2k$ .
- 3) Вычислим

$$\lambda_k = \sqrt[m]{SpA^m}, \quad (2.27)$$

где  $SpA^m$  – след матрицы  $A^m$ , то есть сумма элементов её главной диагонали.

- 4) Увеличим  $k$  на 1, то есть  $k := k+1$ .
- 5) Возведём матрицу  $A$  в степень  $2^{(k+1)}$  или, что тоже, возведём в квадрат матрицу, полученную на предыдущем шаге.
- 6) Произведём расчёт  $\lambda_{k+1}$  по формуле (2.27) при  $m = 2^{(k+1)}$ .
- 7) Если

$$|\lambda_{k+1} - \lambda_k| \leq \varepsilon, \quad (2.28)$$

где  $\varepsilon$  – значение разности собственных значений матрицы  $A$ , при которой процесс считается установившимся, то прекратим итерации и примем значение  $\lambda_{k+1}$  в качестве главного (максимального) собственного значения  $\lambda_{max}$ . В противном случае продолжим итерации, увеличив  $k$  на 1.

- 8) Вычислим индекс согласованности оценок в матрице  $A$  по формуле

$$I = \frac{\lambda_{max}-n}{n-1}. \quad (2.29)$$

9) Найдём  $\vec{Y}$  – собственный вектор, соответствующий  $\lambda_{max}$ . Для этого воспользуемся формулой

$$\vec{Y} = A^m e_n, \quad (2.30)$$

где  $e_n$  – вектор-столбец из  $n$  единиц.

Из формулы (2.30) следует, что элемент  $y_i$  ( $i = 1, 2, \dots, n$ ) вектора  $\vec{Y}$  равен сумме элементов строки  $i$  матрицы  $A^m$ .

10) Проведём нормирование, разделив каждый элемент  $y_i$  на сумму всех элементов вектора  $\vec{Y}$ , то есть вычислим для  $i$  от 1 до  $n$  значения

$$y_i^* = \frac{y_i}{\sum_{j=1}^n y_j}. \quad (2.31)$$

11) Вычислим выраженный в процентах приоритет  $g_i$  цели  $G_i$  ( $i = 1, 2, \dots, n$ ) по формуле

$$g_i = y_i^* 100\%. \quad (2.32)$$

Разработана программа S-Target, в которой реализован этот метод расчёта приоритетов целей. Входной является заполненная по заданному шаблону в Excel таблица парных сравнений целей.

Приведём пример. Рассмотрим условную, но имеющую реальный прототип, компанию «Добрая пицца». Производственная деятельность компании – изготовление и доставка пиццы, заказанной по Интернету. Благотворительная деятельность компании – бесплатная доставка пиццы детям из многодетных семей района. Особенностью является то, что ингредиенты определяет заказчик, из имеющихся и сочетаемых продуктов.

Результаты сравнения целей компании записаны в Таблицу 2.2.

Таблица 2.2 – Результаты сравнения целей компании

	1	2	3	4	5	6
Цели	Увеличить	Повысить	Увеличить кол-во	Снижать издержки,	Достичь высокого	Повышать квалифика

		доход	рентабельность	постоянных клиентов	не влияющие на кач-во	кач-ва продукции	цию сотрудников
1	Увеличить доход		Эквивалентны	Эквивалентны	Менее важно	Эквивалентны	Эквивалентны
2	Повысить рентабельность			Более важно	Эквивалентны	Более важно	Более важно
3	Увеличить кол-во постоянных клиентов				Менее важно	Эквивалентны	Эквивалентны
4	Снижать издержки, не влияющие на кач-во					Менее важно	Более важно
5	Достичь высокого кач-ва продукции						Эквивалентны
6	Повышать квалификацию сотрудников						

Заметим, что количество целей уменьшено, относительно реальный проектов. В реальной компании, являющейся прототипом, сформулировано 16 стратегических целей. Они разнесены по четырём стратегическим перспективам: «Финансы», «Клиенты», «Внутренние процессы», «Обучение и рост сотрудников» в соответствии с методологией Нортон и Каплана [95]. Переядя от лингвистическим к цифровым значениям, получим Таблицу 2.3.

Таблица 2.3 – Результаты сравнения целей компании в численном виде

	Цели	1	2	3	4	5	6
	Цели	Увеличить доход	Повысить рентабельность	Увеличить кол-во постоянных клиентов	Снижать издержки, не влияющие на кач-во	Достичь высокого кач-ва продукции	Повышать квалификацию сотрудников
1	Увеличить доход	1	1	1	1/3	1	1
2	Повысить рентабельность	1	1	3	1	3	3
3	Увеличить кол-во постоянных клиентов	1	1/3	1	1/3	1	1
4	Снижать издержки, не влияющие на кач-во	3	1	3	1	1/3	3
5	Достичь высокого кач-ва продукции	1	1/3	1	3	1	1
6	Повышать квалификацию сотрудников	1	1/3	1	1/3	1	1

Программа S-Target реализует такой переход и принимает полученную таблицу в качестве таблицы А, над которой проводятся операции, указанные выше. Программа S-Target обеспечивает большую точность расчётов по сравнению с аналогами. Она компактна и имеет простой дружелюбный интерфейс. Вместе с тем, это ещё не промышленный продукт, а удобный инструмент консультанта. Результаты расчёта приоритетов приведены на рисунке 2.1.

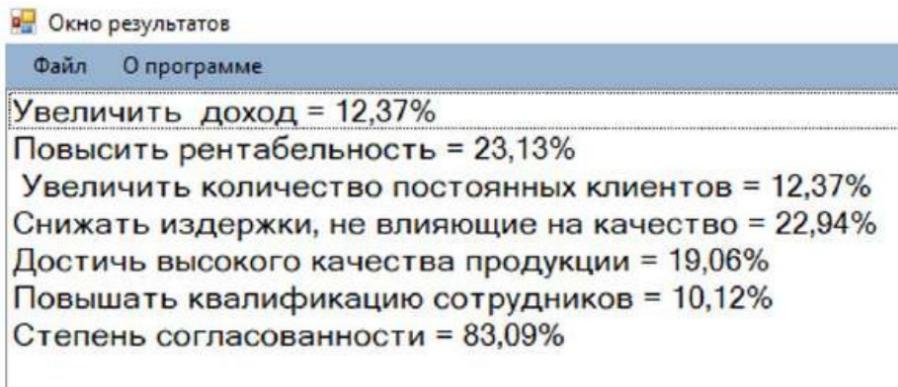


Рисунок 2.1 – Результаты расчёта приоритетов

В расчётах значение  $\varepsilon$  принято равным 0,005. Как видим, индекс согласованности выше 80%. Это позволяет считать, что проведенное руководством компании сравнение целей, является достаточно непротиворечивым.

### 2.3.3 Некоторые выводы

- 1) Предложена модификация метода расчёта приоритетов стратегических целей по таблице парных сравнений их значимости.
- 2) Разработана программа, в которой этот метод реализован.
- 3) Полученные результаты расширяют возможности применения метода анализа иерархий при проведении стратегического консалтинга.
- 4) Результаты опубликованы в статье [96].

### **3 Исследования по анализу текстов и кластерному анализу**

В этом разделе отчета приведены результаты исследований по анализу текстов и кластерному анализу, которые осуществлялись в следующих направлениях:

- разработка программного обеспечения для таксономического анализа текстовых коллекций;
- диффузионно-инвариантная паттерн-кластеризация на основе сортировки данных;
- декомпозиционная сложность графов и её применение для анализа фондовых рынков;
- разработка и исследование метода кластеризации тел свидетельств.

#### **3.1 Разработка программного обеспечения для таксономического анализа текстовых коллекций**

##### **3.1.2 Введение**

Автоматизация анализа текста – одна из основных задач искусственного интеллекта. Несмотря на серьезные успехи, связанные с решением отдельных задач анализа текстов, таких как категоризация, аннотирование, классификация и др., анализ содержания текстовых коллекций в целом остается серьезной проблемой для исследователей.

Одним из популярных направлений исследований является так называемый «контент-анализ» [97]. Это общий метод выявления закономерностей в текстах, основанный на выборе определенных категорий и анализе их появления. Этот подход был разработан в доцифровую эпоху, прежде всего как метод социологического исследования. Однако в настоящее время он используется для анализа коллекций текстов в любой области. В настоящее время этот метод значительно обогатился использованием методов многомерного анализа данных, включая так называемое тематическое моделирование [98]. Модель тематического моделирования применяется к текстовым документам, представленным в виде «мешков слов» (bag of words), и основана на извлечении так называемых «скрытых»,

или «латентных» тем, которые вероятностным образом связаны с наблюдаемыми частотами слов. В результате применения метода получаются списки слов, соответствующих отдельным темам, значение которых восстанавливается вручную из анализа общего значения слов, входящих в них.

Типичное исследование по тематическому моделированию описано в \cite [99]. Для научного направления «экологическая социология» авторы сформировали набор из 815 научных публикаций и проанализировали содержание аннотаций к этим публикациям методом тематического моделирования. Они определили, используя так называемый индекс перплексии, что наилучшее количество скрытых тем для данной выборки будет от 25 до 31, из которых они выбрали 25, потому что полученные темы лучше всего интерпретируются. Для наглядности приведем примеры тем и их наиболее вероятных слов, полученных в [99]: «Экологическая заботливость» (окружающая среда, заботливость, отношение, результат); «Ценности» (социальные ценности, исследования, концепции, отношения), «Политика и управление» (общество, менеджмент, группа, решение, план). Еще раз акцентируем внимание, что названия этих тем были предложены авторами работы.

Темы, формируемые тематическим моделированием, в отличие от классического контент-анализа, определяются эмпирически, а не теоретически, что создает определенные проблемы на пути теоретического осмыслиения результатов. Подход таксономического контент-анализа, предложенный нашей командой в [100], [101], [102], включает теоретический компонент явно. Этот подход предполагает заранее разработанную таксономию предметной области. Таксономия – это древовидная иерархия основных понятий. Тот факт, что  $A$  является «родителем»  $B$  в таксономии, означает отношение « $B$  является частью или конкретным случаем  $A$ ».

Отметим также два факта, важные для нашей методологии. Во-первых, таксономия предоставляет список основных категорий, то есть концепций

самого нижнего уровня иерархии, структура распределения которых по отношению к рассматриваемому собранию текстов является предметом анализа. Во-вторых, таксономия раскрывает свои концепции верхнего уровня как общие темы, присущие определенной группе (кластеру) простых категорий.

### 3.1.2 Таксономический контент-анализ

Метод таксономического контент-анализа, описанный в [100], устроен следующим образом. В первую очередь (как и всегда в контент-анализе) формируется набор текстовых документов, представляющих определенную область знаний или деятельности. Затем (или независимо) формируется таксономия предметной области («домена»), листовые темы которой представляют категории для последующего анализа. После этого последовательно выполняются следующие три этапа:

- расчет матрицы релевантности листовых тем и текстов коллекции;
- получение нечетких кластеров листовых тем по структуре матриц релевантности и со-релевантности;
- оптимальный подъем каждого из полученных кластеров.

На первом этапе вычисляется степень представления листовых тем таксономии в текстах коллекции. Предлагаемая нами мера релевантности между текстом и ключевой фразой категории основана использовании механизма так называемых аннотированных суффиксных деревьев (АСД). Формально говоря, АСД вычисляет среднее значение условной вероятности правильного предсказания букв ключевой фразы на основе знания фрагментов текста, с которым происходит сопоставление. Эта характеристика средней условной вероятности используется в качестве меры релевантности между ключевой фразой и текстом. Данный метод, представленный в [103], показал свою эффективность в ряде экспериментов. Этот метод особенно хорошо работает в тех довольно распространенных ситуациях, когда не все темы системы понятий таксономии присутствуют в рассматриваемых текстах. Формирование матрицы релевантности между

текстами с одной стороны, и темами с другой, завершает первый этап нашего метода.

Второй этап – поиск нечетких кластеров листовых тем. Для этого матрица релевантности между текстами и листовыми темами преобразуется в матрицу значений ко-релевантности между листовыми темами. После этого применяется аддитивно-спектральный метод FADDIS [104]. Чтобы сделать кластерную структуру матрицы более «резкой», каждый текст взвешивается обратным значением количества соответствующих ему тем. Кроме того, применяется псевдо-лапласова нормализация из спектрального кластерного анализа. Метод FADDIS извлекает кластеры последовательно, один за другим, так что всегда можно остановить процесс получения кластеров, и таким образом определить само количество кластеров. Этому способствуют характеристики вклада кластеров в разброс данных.

Третий этап – поиск обобщений полученных кластеров. В отличие от других подходов, где определение обобщения кластера является предметом воображения и изобретательности авторов, метод, предложенный в [100], делает это автоматически с использованием специального алгоритма, называемого ПарГеНМ. ПарГеНМ находит оптимальный набор «головных тем» для данного нечеткого кластера листовых тем в таксономии. Набор «головных тем» (одной или нескольких) содержит концепты таксономии высокого ранга, которые включают (то есть «покрывают») темы из данного кластера, возможно, с ошибками двух типов: небольшим количеством «пробелов» и «выбросов» (подробнее см. [105]). «Пробел» – это тема дерева таксономии, которая покрывается головной темой, но не принадлежит рассматриваемому исходному кластеру, а «выброс» – это тема из кластера, которая не охвачена головной. Во многих случаях такой оптимальный подъем переводит рассматриваемую структуру, кластер листовых тем, на некоторый значительно более «грубый» уровень грануляции, что позволяет увидеть такие идеи и связи, которые не очевидны при более интуитивных подходах.

Эффективность этой гибридной методологии была продемонстрирована в нескольких публикациях. В [100] и [101] методология была применена для извлечения исследовательских тенденций из двух коллекций исследовательских работ: (а) около 18000 научных статей, опубликованных в журналах Springer по Науке о Данных за 20 лет, и (б) около 27000 научных статей, полученных из журналов Springer и Elsevier в ответ на запросы, связанные с Наукой Данных. Авторы использовали таксономию науки о данных (TDS), основанную на классификации вычислительных систем Ассоциации вычислительной техники (ACM-CCS 2012) [106]. В [105] алгоритм обобщения был применен для расширения рекламной аудитории в Интернете.

### 3.1.3 Оптимальный подъем нечеткого тематического кластера в таксономии: модель и метод

Остановимся на описании метода оптимального обобщения нечетких тематических кластеров в таксономии несколько подробнее.

Говоря математическим языком, таксономия – это корневое дерево, чьи узлы аннотированы темами таксономии. Мы будем рассматривать следующую задачу: дано нечеткое множество  $S$  из листьев таксономии, необходимо найти узел  $t(S)$  более высокого ранга в таксономии, который покрывает множество  $S$  наиболее “тесным” образом. Такая задача “подъема” – математическая постановка человеческой способности делать обобщение, то есть, процесса формирования “концептуальной” формы представленного явления, в данном случае, нечеткого множества листьев дерева.

Задача не так проста, как могло бы показаться. Для краткости, рассмотрим четкое множество  $S$ , представленное пятью черными прямоугольниками на фрагменте дерева на рисунке 3.1. Рисунок 3.2 иллюстрирует ситуацию, при которой множество поднято в корень, что показано закрашиванием соответствующего ему прямоугольника, как и его потомков.

Если мы допустим, что множество  $S$  может быть обобщено корнем, это приведет к тому, что корнем будет покрыто и четыре белых прямоугольника, и, таким образом, попадет в то же обобщение, что и  $S$ , при том, что они не принадлежат множеству  $S$ . Такую ситуацию будем далее называть появлением пробелов. Очевидно, что вариант подъема, при котором возникают пробелы, должен штрафоваться моделью подъема. В целом, число концептуальных элементов, введенных для обобщения  $S$ , здесь равно 1 головной вершине (то есть корню), который мы теперь присвоили переменной  $S$ , и 4 пробела, образовавшиеся из-за топологии дерева. При этом за подобный подъем, как уже было сказано, должно быть назначено определенное значение штрафа. Другой возможный вариант подъема показан на рисунке 3.3: здесь заданное множество поднимается только до корня левой ветви дерева. Мы можем видеть, что число пробелов резко сократилось – до 1. Однако возник еще один тип неточности подъема: черный прямоугольник справа, принадлежащий  $S$ , не покрыт корнем левой ветви, в которую отображается множество  $S$ . Этот тип ошибки будем называть выбросом. При этом варианте подъема появляются три новых объекта: одна головная вершина, один выброс и один пробел. Это меньше, чем количество объектов, появившихся при поднятии множества к корню (одна головная вершина и четыре пробела, то есть совокупно – пять), что делает его более предпочтительным. Конечно, этот вывод верен только в том случае, если относительный штраф за один выброс меньше суммарного штрафа за три пробела.

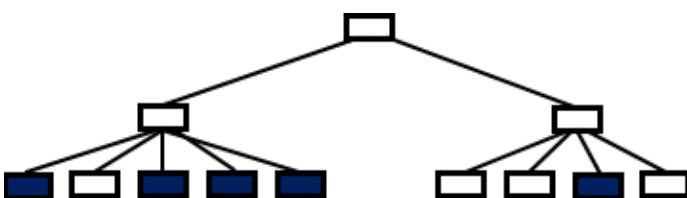


Рисунок 3.1 – Фрагмент таксономии, черные прямоугольники соответствуют элементам, принадлежащим множеству

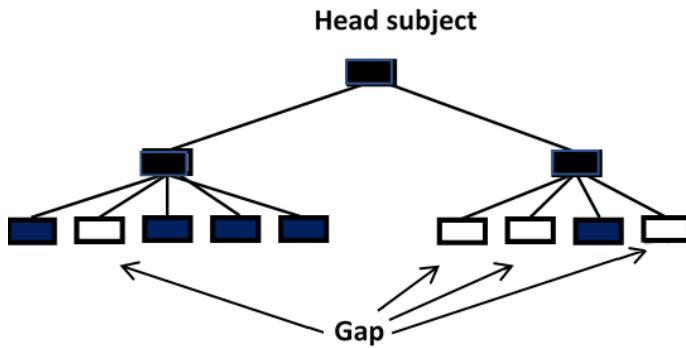


Рисунок 3.2 – Обобщение нечеткого множества запроса, показанного на рисунке 3.1 с помощью его отображения в корень дерева. Цена этого обобщения – четыре пробела (gap)

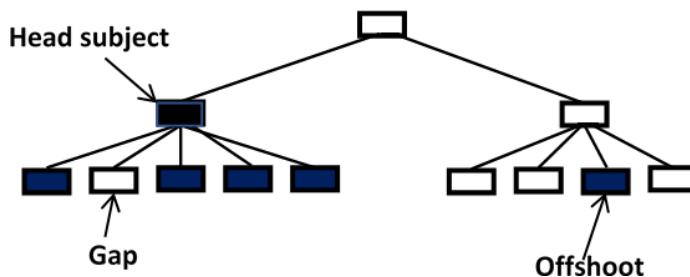


Рисунок 3.3 – Обобщение нечеткого множества запроса, показанного на рисунке 3.1 с помощью его отображения в корень левого поддерева. Цена такого обобщения - один пробел (gap) и один выброс (offshoot)

Нас интересует, можно ли обобщить нечеткое множество  $S$  в узел  $t$  принадлежащий более высокому рангу таксономии, так, чтобы можно было считать, что  $S$  попадает в покрытие, охватываемое узлом  $t$ . Если это будет сделано напрямую, например, путем выбора последнего общего предка для узлов в  $S$  получение желаемого охвата будет маловероятным. Чтобы сделать процесс обобщения более адекватным, следует допустить наличие в итоговом обобщении не только пробелов, но и выбросов, являющихся листьями, принадлежащими обобщению, но не исходному множеству.

Цель нахождения интерпретируемого обобщения для  $S$  в рамках таксономии может быть формализована как нахождение одного или нескольких “головных тем” для покрытия  $S$  с минимальным количеством новых элементов всех типов, получившихся в процессе обобщения: главных

вершин, пробелов и выбросов. Такая постановка цели реализует принцип максимальной экономии (Maximum Parsimony, MP) при описании данного явления, популярный в некоторых областях, таких как биоинформатика.

Рассмотрим корневое дерево  $T$ , представляющее иерархическую таксономию, такую, что его узлы снабжены ключевыми фразами, обозначающими различные понятия. Обозначим множество его листьев через  $I$ . Соотношение между узлами в иерархии обычно выражается с использованием генеалогических терминов: каждый узел  $t \in T$  называется родителем узлов, непосредственно ответвляющихся от  $t$ , его потомков. Будем использовать  $\chi(t)$  для обозначения множества потомков этого узла. Предполагается, что каждый внутренний узел  $t \in T - I$ , то есть узел, не являющийся листом таксономии, соответствует понятию, обобщающему темы, соответствующие листьям  $I(t)$ , выходящим из  $t$ , то есть, листья поддерева  $T(t)$  имеют общий корень в узле  $t$ , называющийся кластером листьев.

Нечеткое множество на  $I$  – это отображение  $u$  из множества  $I$  в множество неотрицательных действительных чисел, которое присваивает значение функции принадлежности (или поддержки)  $u(i) \geq 0$  каждому узлу  $i \in I$ . Мы называем множество, где  $S_u \subset I$ , где  $S_u = \{i \in I : u(i) > 0\}$ , поддержкой  $u$ . Никаких других предположений о виде функции  $u$  не делается. Отметим только то, что для удобства целесообразно положить такое ограничение, чтобы ее значение не превышало единицу. Классические, или четкие множества соответствуют бинарным функциям принадлежности – таким, что  $u(i)=1$ , если  $i \in S_u$  и  $u(i)=0$  иначе.

Рассматривая нечеткое множество запроса  $u$ , определенное на листьях  $I$  дерева  $T$ , можно считать  $u$  (возможно, сделанной с погрешностями) проекцией понятия более высокого ранга, “головной темы”, на соответствующий кластер листьев. В этом предположении должен существовать головной узел  $h$  среди внутренних узлов дерева  $T$  – такой, что его кластер листьев  $I(h)$  в определенной степени совпадает (с небольшими

ошибками) с  $S_u$ . Этот головной узел – обобщение  $u$ , которое и необходимо найти.

Возможны два типа ошибок, связанных с поиском головного узла, если он точно не покрывает поддержку, которые представляют собой ложные срабатывания и ложные отрицания, уже упомянутые ранее как пробелы (gaps) и выбросы (offshoots) – соответственно, иллюстрируются на рисунках 3.2 и 3.3. Общее количество головных узлов, пробелов и выбросов должно быть как можно меньше. Для этого мы вводим штраф за каждый из этих элементов. Предполагая, для простоты, что все черные прямоугольники на рисунке 3.1 имеют значения функции принадлежности, равные единице, можно легко увидеть, например, что общее значение штрафа, например, для подъема на рисунке 3.3 равно  $1 + \gamma + \lambda$ , где  $\gamma, \lambda$  и  $1$  – штрафы за выброс, пробел и головную тему соответственно.

Рассмотрим узел-кандидат  $h$  в  $T$  и его значение принадлежности в нечетком множестве  $u$ .  $h$ -пробел – это узел  $g$  из  $T(h)$ , отличный от  $h$ , в котором произошла потеря значения, то есть  $g$  является максимальным  $u$ -иррелевантным узлом в том смысле, что его родитель не является  $u$ -иррелевантным. И наоборот, установление узла  $h$  в качестве головного можно рассматривать как приобретение значения  $u$  в узле. Множество всех  $h$ -выбросов обозначим как  $G(h)$ .

Чтобы правильно определить идею использования пробелов, введем понятие  $u$ -иррелевантных узлов. Узел  $t$  называется  $u$ -иррелевантным, если его кластер листьев  $I(t)$  не пересекается с поддержкой  $u$ . Очевидно, что если узел принадлежит  $u$ , все его потомки также принадлежат  $u$ .

$h$ -выброс – это лист  $i \in S_u$ , который не покрывается  $h$ , то есть  $i \notin I(h)$ . Множество всех  $h$ -выбросов – это  $S_u - I(h)$ .

Если задано нечеткое тематическое множество  $u$  над  $I$ , множество узлов  $H$  будет называться  $u$ -покрытием, если: (а)  $H$  покрывает  $S_u$ , то есть,  $S_u \subseteq \bigcup_{h \in H} I(h)$ , и (б) узлы в  $H$  не связаны, то есть  $I(h) \cap I(h') = \emptyset$  для любых  $h, h' \in H$ , таких, что  $h \neq h'$ . Внутренние узлы  $H$  будут называться

головными темами, а листовые узлы – выбросами, поэтому множество выбросов в  $H$  равно  $H \cap I$ . Множество пробелов в  $H$  является объединением  $G(h)$  по всем головным темам  $h \in H - I$ .

Определим штрафную функцию  $p(H)$  для  $u$ -покрытия  $H$  следующим образом:

$$p(H) = \sum_{h \in H - I} u(h) + \sum_{h \in H - I} \sum_{g \in G(h)} \lambda v(g) + \sum_{h \in H \cap I} \gamma u(h). \quad (3.1)$$

Задача, которую мы решаем, состоит в том, чтобы найти такое  $u$ -покрытие  $H$ , которое минимизирует штрафную функцию  $p(H)$ . Такое  $u$ -покрытие и будет наиболее оптимальным обобщением нечеткого множества  $u$ .

Перед применением алгоритма для минимизации суммарного штрафа необходимо выполнить предварительное преобразование дерева. Для этого необходимо сделать следующие шаги. Из дерева нужно удалить все немаксимальные  $u$ -иррелевантные узлы, то есть потомки пробелов, а также найти множества  $G(t)$  и значения суммарной важности пробелов  $V(t) = \sum_{g \in G(t)} v(g)$  для всех внутренних узлов  $t$ . Отметим, что все элементы  $S_u$  находятся в множестве листьев обрезанного дерева, а остальные листья обрезанного дерева – это именно пробелы.

Теперь мы можем применить алгоритм подъема ПарГеНМ. Для каждого узла  $t$  алгоритм ПарГеНМ вычисляет два набора –  $H(t)$  и  $L(t)$ , содержащие такие узлы в  $T(t)$ , в которых, соответственно, происходят приобретения и потери головных узлов (включая выбросы). Соответствующий штраф обозначается  $p(t)$  и вычисляется так, как описано ниже.

Предположение алгоритма состоит в том, что после потери узла не может быть последующего ее приобретения. Следовательно,  $H(t)$  и  $L(t)$  вычисляются при условии, что головной узел не был ни приобретен (ни потерян) ни у одного из предков  $t$ . Алгоритм ПарГеНМ рекурсивно

вычисляет  $H(t)$ ,  $L(t)$  и  $p(t)$  из соответствующих значений для дочерних узлов в  $\chi(t)$ . Для каждого листового узла, который не принадлежит  $S_u$ , мы определяем как  $L(\cdot)$  так и  $H(\cdot)$  пустыми, а штраф равным нулю. Для каждого листового узла, который находится в  $S_u$ ,  $L(\cdot)$  устанавливается пустым, тогда как  $H(\cdot)$ , содержит только листовой узел, и штраф определяется как произведение зафиксированного параметра  $\gamma$  на значение его принадлежности. Чтобы вычислить  $L(t)$  и  $H(t)$  для любого внутреннего узла  $t$  возможны два случая:

- (a) головной узел был приобретен в  $t$ ;
- (b) головной узел не был приобретен в  $t$ .

В случае (a) множества  $H(\cdot)$  и  $L(\cdot)$  у его потомков далее оказываются ненужными. В этом случае,  $H(t)$ ,  $L(t)$  и  $p(t)$  определяются как:

$$\begin{cases} H(t) = \{t\}, \\ L(t) = G(t), \\ p(t) = u(t) + \lambda V(t). \end{cases} \quad (3.2)$$

В случае (b) множества  $H(t)$  и  $L(t)$  являются просто объединениями соответствующих множеств из их потомков, а  $p(t)$  – сумма всех штрафов из потомков:

$$\begin{cases} H(t) = \bigcup_{w \in \chi(t)} H(w), \\ L(t) = \bigcup_{w \in \chi(t)} L(w), \\ p(t) = \sum_{w \in \chi(t)} p(w). \end{cases} \quad (3.3)$$

Чтобы получить оптимальный подъем, из этих случаев выбирается тот, где достигается меньшее значение  $p(t)$ . В ситуации, когда оба случая дают одинаковые значения  $p(t)$ , мы можем выбрать любой из них произвольно – в формулировке алгоритма ниже, мы выбрали (a). Выходные данные алгоритма включают значения рассчитываемых множеств в корне, а именно:  $H$  – множество головных узлов и выбросов,  $L$  – множество пробелов и  $p$  –

назначенное значение штрафа. Мы доказали, что алгоритм ПарГеНМ действительно приводит к глобально оптимальному обобщению [107].

### 3.1.4 Описание и структура программного пакета GOT

Программный пакет, реализующий методологию таксономического анализа контента, описанную выше, GOT (Generalization Over Taxonomies) можно загрузить из репозитория GitHub (<https://github.com/dmitsf/GOT/>). Пакет предусматривает реализацию всех этапов методологии и позволяет пользователю:

- вычислить значения релевантности текстового документа, составить матрицы релевантности и совпадения;
- выполнить нечеткую кластеризацию, используя реализацию алгоритма FADDIS;
- сделать оптимальное обобщение нечетких кластеров в таксономии, используя реализацию метода ParGenFS;
- визуализировать таксономии и результаты обобщения нечетких кластеров.

В состав пакета входят три подраздела-модуля, содержащие программные файлы:

- asts;
- relevance\_analysis;
- taxonomies.

В следующих подразделах мы опишем структуру инструментария GOT.

#### asts

Этот модуль включает программную реализацию аннотированного суффиксного дерева (AST), которая позволяет вычислять релевантность текстового документа. Файл ast.py содержит реализацию АСД. Чтобы вычислить релевантность между заданной строкой и текстовым документом, необходимо:

- создать АСД для текста с помощью конструктора класса AST();

- вызвать метод *score* с аргументом – строкой, для которой необходимо хотите вычислить релевантность.

#### *relevance\_analysis*

Этот модуль содержит реализации программ для получения матриц релевантности и ко-релевантности, преобразований LAPIN и нечеткой кластеризации FADDIS.

Файл *relevance.py* позволяет получить матрицу релевантности заданного набора строк заданной коллекции текстов на основе метода аннотированного суффиксного дерева (АСД). Этот файл содержит функции для вычисления матрицы релевантности: функция *get\_relevance\_matrix* позволяет получить матрицу релевантности между списком текстовых документов и списком строк.

Файл *corelevance.py* включает функции для вычисления матрицы релевантности между списком текстовых документов и списком строк. Функция *get\_corelevance\_matrix* позволяет получить матрицу ко-релевантности. Файл *lapin.py* содержит функцию, которая выполняет псевдообратное преобразование Лапласа (LAPIN), а программа *faddis.py* содержит функцию поиска нечетких кластеров с помощью метода FADDIS.

#### *taxonomies*

В этот модуль входят программы для работы с таксономиями, обобщения нечетких тематических кластеров с использованием алгоритма ПарГеНМ и визуализации результатов. Файл *taxonomy.py* объединяет функции для работы с таксономией. В частности, он сочетает в себе извлечение таксономии из файла *.fvtr* и подготовку базовых внутренних структур данных для работы с ней. *fvtr* – (flat-view taxonomy representation, представление таксономии в плоском виде) – это csv-подобный формат для представления таксономий. Формат основан на представлении, используемом в таксономии ACM CCS с официального сайта ACM (<https://www.acm.org/publications/class-2012>).

Рассмотрим в качестве примера фрагмент таксономии контента Бюро интернет-рекламы (IAB) <https://www.iab.com/guidelines/content-taxonomy/> в формате *.fvtr*. Эта таксономия представляет собой иерархию тем контента в Интернете.

579.,Men's Fashion,

579.580.,,Men's Accessories,

579.580.581.,,,Men's Jewelry and Watches,

579.582.,,Men's Clothing Style,

579.582.583.,,,Men's Business Wear,

579.582.584.,,,Men's Casual Wear,

579.582.585.,,,Men's Formal Wear,

579.582.586.,,,Men's Outerwear Style,

579.582.587.,,,Men's Sportswear,

579.582.588.,,,Men's Underwear and Sleepwear,

579.589.,,Men's Shoes and Footwear,

Этот фрагмент представляет дерево таксономии, показанное на рисунке 3.4. В каталоге *test\_files/* репозитория GOT можно найти больше примеров файлов *.fvtr*.

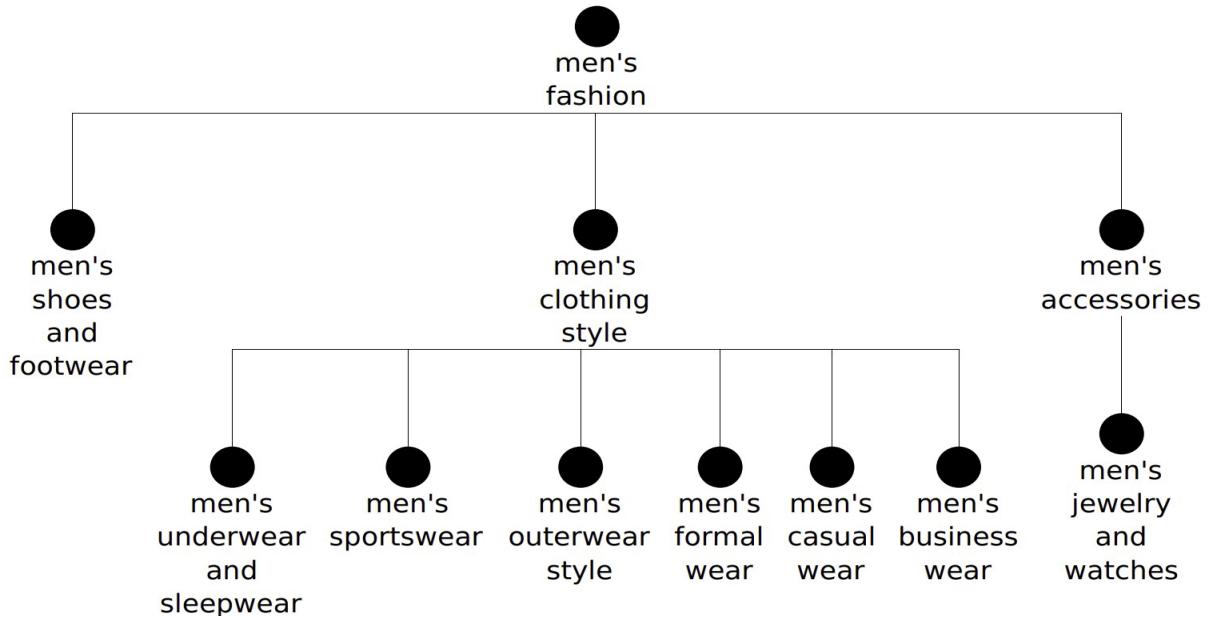


Рисунок 3.4 – Фрагмент Таксономии IAB

Для работы с таксономиями реализованы два класса: *Tree* представляет дерево таксономии, *Node* представляет узел дерева таксономии (см. техническую документацию [https://got-](https://got-docs.readthedocs.io/en/latest/got.taxonomies.html)  
[docs.readthedocs.io/en/latest/got.taxonomies.html](https://got-docs.readthedocs.io/en/latest/got.taxonomies.html)).

Файл `pargenfs.py` содержит функции, реализующие алгоритм оптимального обобщения нечеткого кластера в таксономии (ПарГенМ). После запуска он создает два файла, таблицу со значениями всех переменных на всех этапах алгоритма и файл с деревом подъема в формате `.ete` (будет описан ниже). Некоторые примеры обобщений кластеров по таксономии IAB представлены в Таблице 3.1. Эти кластеры содержат интересы пользователей в Интернете, поэтому мы можем получить представление об интересах пользователей в терминах категорий высокого уровня. Это свойство было использовано в специальном алгоритме расширения пользовательской аудитории, см. статьи Фролова, Таран и Миркина [105], [108], как упоминалось выше.

Таблица 3.1 – Примеры обобщений нечетких кластеров по таксономии IAB

Нечеткий кластер со значениями принадлежности	Обобщение после применения ParGenFS
{Cloud Computing (0.596), Web Development (0.481), Internet for Beginners (0.432), IT and Internet Support (0.356), Social Networking (0.312)}	{Internet (1.0)}
{Men's Jewelry and Watches (0.662), Men's Business Wear (0.514), Men's Casual Wear (0.443), Men's Outerwear (0.320)}	{Men's Fashion (1.0)}
{3-D Graphics (0.678), Video Software (0.571), Graphics Software (0.570), Operating Systems (0.308), Business Accounting and Finance (0.351)}	{Computer Software and Applications (0.902), Business Accounting and Finance (0.351)}

.ete – это формат представления дерева, используемый в программном пакете ETE (<http://etetoolkit.org/>) и других подобных пакетах. Формат позволяет определять топологию дерева вместе с атрибутами узлов дерева. Этот формат предполагает скобочные обозначения, что затрудняет чтение людьми, но, с другой стороны, он чрезвычайно компактен и легко анализируется программами. Пример представления дерева в .ete из документации ETE toolkit приведен ниже.

```
"(A,B,(C,D));"
#
#   /-A
#
#   |
#---|--B
#
#   |
#   |   /-C
#   |   \
#   |   \---|
#       \--D
```

Файл visualize.py содержит функции для визуализации подъема кластеров в таксономиях и самих деревьев в формате .ete. Модуль может создавать изображения в формате .pdf. Пример визуализации подъема

клUSTERA в упомянутом выше фрагменте таксономии IAB показан на рисунке 3.5.

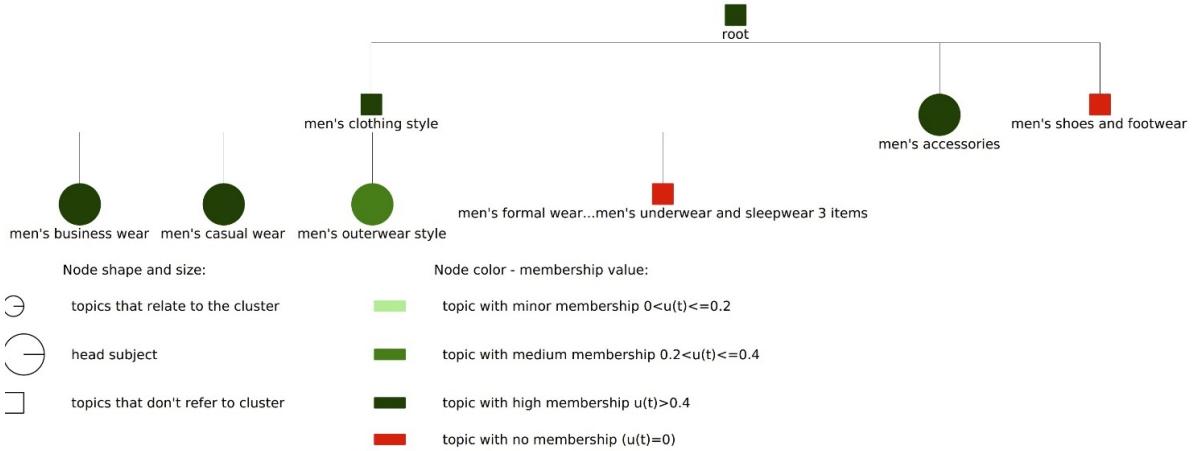


Рисунок 3.5 – Пример визуализации подъема кластера над фрагментом таксономии IAB

Модуль имеет широкий набор настроек. Он также позволяет визуализировать произвольные деревья в формате .ete (см. рисунок 3.4, полученный с помощью данного модуля), а также получить расширенную визуализацию (см. рисунок 3.6).

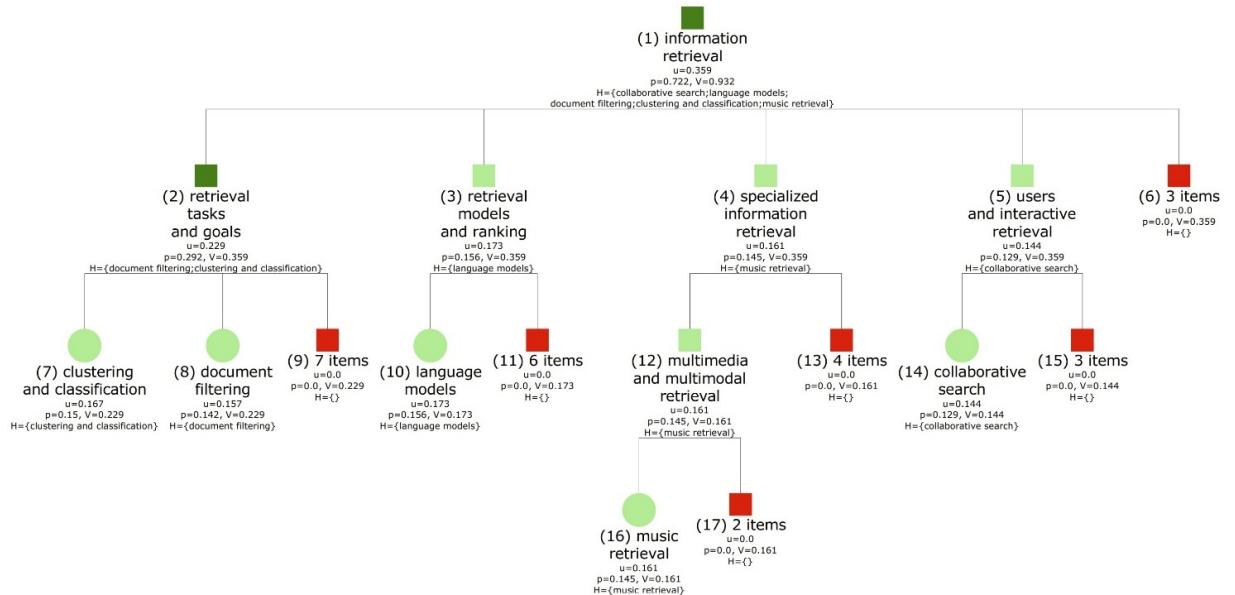


Рисунок 3.6 – Расширенная визуализация подъема кластера во фрагменте таксономии Науки Данных. Легенда обрезана

Подробное руководство и туториал находится в репозитории GOT. Туториал включает ряд примеров для каждого этапа гибридной методологии, а также файлы для экспериментов. Техническая документация для разработчиков доступна по адресу в ReadTheDocs: <https://got-docs.readthedocs.io>.

Рекомендуемый способ использования GOT – работать с ним непосредственно из программного кода. Однако функционал модуля taxonому можно использовать как из кода Python3, так и из командной строки. Эта функция была реализована для удобства на основе практического опыта работы с текстовыми коллекциями и таксономиями доменов. Для работы с инструментарием GOT необходим Python3 версии 3.5 или новее.

Для базового использования можно пропустить установку инструментария: достаточно будет клонировать репозиторий локально и запускать программы из папки репозитория. Если же необходимо импортировать модули GOT в свой код Python, можно установить пакет через любой менеджер пакетов Python3. Пакет не имеет никаких специальных требований к программному обеспечению окружения, кроме пакета *ete* и подмодулей *Qt* для визуализации. Подробные инструкции по установке находятся в репозитории GitHub.

### 3.1.5 Некоторые выводы

Мы представили новый программный пакет Python3 для таксономического контент-анализа коллекций текстовых документов. Программный пакет реализует методологию, которая предусматривает обобщение, то есть отображение значений на более абстрактных уровнях. Эта особенность уникальна и не может быть смоделирована с помощью традиционных подходов. Поскольку инструментарий GOT предусматривает независимое использование его отдельных частей, пользователь получает гибкие возможности, которые можно использовать в других смежных областях.

В наши планы на будущее входит расширение инструментария пакета. В частности, мы собираемся добавить реализации других методов кластеризации. Также мы будем работать над внедрением альтернативных подходов к обобщению в таксономиях. Конечно, мы собираемся расширить использование инструментария в различных прикладных областях.

Результаты работы были опубликованы в 2021 году в [108], [109].

## **3.2 Диффузионно-инвариантная паттерн-кластеризация на основе сортировки данных**

### **3.2.1 Введение**

Анализ паттернов является современным методом анализа данных, весьма востребованным на сегодняшний день. К примеру, запрос в ресурсе Google Academy “pattern analysis” дает 5 390 000 результатов, 136 000 из которых с 2019 года (результаты представлены на 24/06/2020). Для сравнения, запрос “machine learning” предоставит 4 600 000 (89 900 с 2019 года), а запрос “cluster analysis” – 4 240 000 (51 100 с 2019 года). Методы анализа паттернов применяются при решении множества разнородных задач (к примеру, при анализе государственной состоятельности [110], банковского сектора [111], исследование мнения экономического сообщества [112]). В анализ паттернов входит, в зависимости от постановки исходной задачи: поиск групп объектов, имеющих определенную «структурную схожесть»; поиск типичного объекта определенного класса; нахождение динамических групп объектов. При этом, важным является само определение термина «паттерн», которое может отличаться в зависимости от конкретной решаемой задачи. В настоящей работе используется определение из [113]: «под паттерном будем понимать комбинацию определенных качественно похожих признаков». Важным здесь является именно «качественная», а не «количественная» схожесть исследуемых объектов. К примеру, при исследовании объектов  $x_1, x_2, x_3$  и  $x_4$  со значениями показателей (60, 90, 60), (80, 50, 80), (6, 9, 6) и (8, 5, 8) соответственно, анализ паттернов, в терминологии, рассматриваемой в настоящей работе, предложит следующее разбиение:  $\{x_1, x_3\}$  и  $\{x_2, x_4\}$ . Объяснение здесь простое: при рас-

смотрении абсолютных значений данных показателей, объекты  $x_1$  и  $x_2$  являются «более близкими» (если рассматривать в качестве меры близости, к примеру, расстояние Евклида). Однако, несложно заметить, что структуры показателей объектов  $x_1$  и  $x_3$  весьма схожи: при умножении значений показателей  $x_3$  на 10 получим значения, аналогичные объекту  $x_1$ . Подобная методология разбиения частично объясняет различие между классическими методами кластерного анализа и анализа паттернов.

В этой части отчета рассматриваются два метода анализа паттернов: порядково-инвариантная [113] и диффузионно-инвариантная [114] паттерн-кластеризации. Рассмотрены альтернативные алгоритмические решения диффузионно-инвариантной паттерн-кластеризации, а также исследованы различия существующих подходов в ее реализации.

### 3.2.2 Анализ паттернов: общая постановка задачи

Исходная постановка задачи анализа паттернов весьма похожа на задачу кластерного анализа: имеется некоторая выборка объектов и несколько параметров. Требуется найти схожие по некоторой метрики объекты и сгруппировать их. Как правило, при использовании методов анализа паттернов не известно ни исходное количество групп, ни типичный представитель каждой группы. Формально, исследуется множество объектов X (наиболее часто используются тип “объект\_время”), каждый из которых описывается набором данных. Для определенности объекты представим в векторном виде  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ . Исходная последовательность параметров заранее не задается (считается произвольной). Необходимо выявить группы структурно близких объектов. «Структурная близость» определяется выбранной мерой схожести. Подробнее, о методах анализа паттернов [115], [116].

Основные этапы анализа паттернов:

- постановка задачи: определение корректности использования предложенных в работе показателей и целесообразности сбора новых данных;
- предварительная обработка данных;
- исключение выбросов в данных;

- работа с пропусками (выбор методологии их заполнения или исключение объектов, содержащих пропуски в данных);
- корреляционных анализ используемых параметров;
- исключение (при необходимости) высоко коррелированных параметров;
- нормировка показателей (к примеру, с использованием «линейного масштабирования»);
- построение кусочно-линейных функций (в системе параллельных координат) для визуализации исследуемых объектов;
- выявление закономерностей в данных на основе выбранных для работы алгоритмов.

В настоящей работе речь пойдет о методах анализа паттернов, в основе которых лежит парное сравнение показателей: порядково-инвариантная и диффузионно-инвариантная паттерн-кластеризация. У данных методов есть ещё одна важная особенность: конечные результаты их разбиения не зависят от входной последовательности данных. Заметим, что аналогичное верно и для методов кластерного анализа, и подобная зависимость является спецификой методов анализа паттернов, использующих системы параллельных координат [117].

Кратко, опишем базовые реализующие алгоритмы порядково-инвариантной и диффузионно-инвариантной паттерн-кластеризаций. Для порядково-инвариантной паттерн-кластеризации каждому строятся полные взвешенные ориентированные графы, вершины которых соответствуют базовым показателям, а ребра – парным сравнениям значений данных показателей. Целесообразным является парное сравнение типа «больше», «равно» и «меньше». Используя данные сравнения, формируются кодировки каждого объекта (обозначим как  $q_i$ ) и сравниваются данные кодировки. В качестве меры близости выступает расстояние Хемминга. Принципом объединения объектов в группы является нулевое расстояние Хемминга между кодировками объектов. Группы объектов, полученных на основе порядково-

инвариантной паттерн-кластеризации или реализующих ее алгоритмов, будем называть порядково-инвариантными паттерн-кластерами. Подробнее, порядково-инвариантная паттерн-кластеризация, описана в [113], а практическое применение на конкретных задачах – в [110].

Диффузионно-инвариантная паттерн-кластеризация основана на порядково-инвариантной паттерн-кластеризации, однако используется при необходимости объединения объектов, не только схожих по структуре, но и по самим значениям показателей. В основе данного метода лежит следующая идея: при близости структур объектов взаимный обмен одним из параметров не должен существенным образом сказываться на их схожести. Для поиска подобных групп объектов, для начала используется порядково-инвариантная паттерн-кластеризация. Далее, рассматривается каждый полученный порядково-инвариантный паттерн-кластер по отдельности. К примеру, в первый порядково-инвариантный паттерн-кластер попали несколько объектов, в т.ч.  $x_k = (x_{k1}, x_{k2}, \dots, x_{kn})$  и  $x_m = (x_{m1}, x_{m2}, \dots, x_{mn})$ . Попарно меняются значения показателей (для определенности, сначала первого). Таким образом, получаем объекты  $x_k^* = (x_{m1}, x_{k2}, \dots, x_{kn})$  и  $x_m^* = (x_{k1}, x_{m2}, \dots, x_{mn})$ . После парного обмена показателей повторно формируются кодировки объектов и сравнивается расстояние Хемминга. Подобная процедура повторяется до тех пор, пока не произойдет парный обмен всех показателей между объектами. При нулевом расстоянии Хемминга на каждом этапе объекты объединяются в единую группу.

Проиллюстрируем данный метод на гипотетическом примере. Допустим, исследуются 6 объектов по 4 параметрам, представленным ниже в системе параллельных координат.

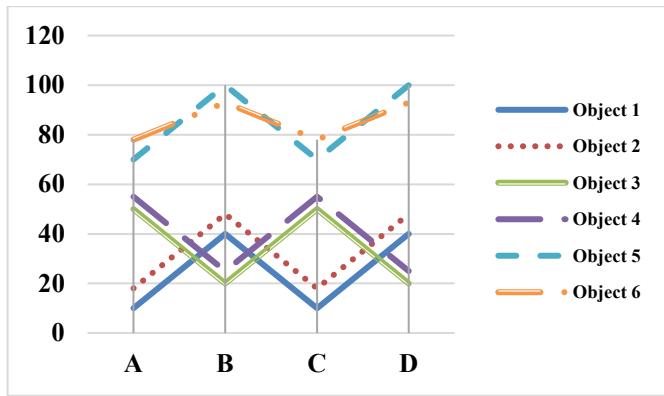


Рисунок 3.7 – Пример использования параллельных координат

Из рисунка 3.7 видно, что объекты 1, 2, 5 и 6 схожи по структуре, однако существенным образом отличаются по абсолютным значениям показателей. Диффузионно-инвариантная паттерн-кластеризация предполагает разбиение для объектов 1, 2, 5 и 6 на 2 группы: {1, 2} и {5, 6}. Далее, подобные группы будем называть диффузионно-инвариантными паттерн-кластерами.

Подобнее, диффузионно-инвариантная паттерн-кластеризация описана в [114].

### 3.2.3 Возможность использования алгоритмов сортировки данных

В условиях больших объемов данных весьма важным условием возможности применения алгоритмов анализа данных является скорость их обработки. Несмотря на относительно невысокую вычислительную сложность порядково-инвариантной и диффузионно-инвариантной паттерн-кластеризаций имеется возможность получения аналогичного разбиения с использованием алгоритм сортировки данных. Для этого, используем утверждение, доказанное в [113]: “Для объектов, принадлежащих одному порядково-инвариантному паттерн-кластеру, существует порядок расположения показателей, образующих монотонно возрастающую/убывающую последовательность, для каждого объекта кластера”. Данное утверждение предполагает возможность альтернативного использование порядково-инвариантной (а значит и диффузионно-инвариантной) паттерн-кластеризации.

Замечание 3.1. Для линейного упорядочивания справедливы следующие условия: а) закон трихотомии – выполняется только одно из следующих

условий –  $a > b$ ,  $a = b$  или  $a < b$ ; б) закон транзитивности – для всех  $a$ ,  $b$ ,  $c$ :  $a > b$  и  $b > c \Rightarrow a > c$ .

Таким образом, при возможности поиска объектов, для которых единая последовательность показателей образует монотонно возрастающую/убывающую последовательность, подобные объекты возможно объединить в единый порядково-инвариантный паттерн-кластер.

Однако, вопрос возникает о возможности (и целесообразности) использования подобных алгоритмов для поиска диффузионно-инвариантных паттерн-кластеров. Данный вопрос усложняется и необходимостью обмена значениями показателей для реализации диффузионно-инвариантной паттерн-кластеризации – а значит и необходимостью повторения процедуры поиска монотонно возрастающих/убывающих последовательностей.

### 3.2.4 Альтернативная реализация диффузионно-инвариантной паттерн-кластеризации

Для использования диффузионно-инвариантной паттерн-кластеризации совместно с алгоритмами сортировки данных, следует определить реализующий алгоритм. В [113] доказано, что разбиение, полученное на основе порядково-инвариантной паттерн-кластеризации – единственное (поскольку парные сравнения определяются однозначно). Использование алгоритмов сортировки для данного метода подразумевает нахождение одного объекта, расположение его показателей в виде монотонно возрастающей/убывающей последовательности, и поиск других объектов с аналогичной последовательностью показателей.

Исходя из этого, определим альтернативный алгоритм для диффузионно-инвариантной паттерн-кластеризации. Для практической реализации также рассматривается один порядково-инвариантный паттерн-кластер, в который попали несколько объектов (при наличии единого объекта подобный паттерн будет автоматически являться диффузионно-инвариантным паттерн-кластером, состоящий из единого объекта). Для определенности возьмем те же объекты  $x_k = (x_{k1}, x_{k2}, \dots, x_{kn})$  и  $x_m = (x_{m1}, x_{m2}, \dots, x_{mn})$ . Поскольку данные

объекты уже объединены на основе порядково-инвариантной паттерн-кластеризации, можно утверждать, что для них существует единая монотонно неубывающая/невозрастающая последовательность расположения показателей. Далее, обменяем первый показатель между объектами:  $x_k^* = (x_{m1}, x_{k2}, \dots, x_{kn})$  и  $x_m^* = (x_{kl}, x_{m2}, \dots, x_{mn})$ . Произвольно, выбираем любой из них (для определенности, первый). Для данного объекта производится парное сравнение показателей и определяется возрастающая/убывающая последовательность. Данная последовательностью сравнивается с расположением показателей объекта  $x_m^*$ . При их несовпадении объекты разделяем в разные группы. Если совпадают – производим взаимный обмен второго показателя между объектами. В таком случае, получаем объекты  $x_k^* = (x_{kl}, x_{m2}, \dots, x_{kn})$  и  $x_m^* = (x_{m1}, x_{k2}, \dots, x_{mn})$ , а на j-ом шаге  $x_k^* = (x_{kl}, x_{k2}, \dots, x_{mj}, \dots, x_{kn})$  и  $x_m^* = (x_{m1}, x_{m2}, \dots, x_{kj}, \dots, x_{mn})$ . Далее, фиксируем один из них и снова образуем монотонно возрастающую/убывающую последовательность. При совпадении расположения показателей данный алгоритм повторяется, пока не произошел обмен  $n$  показателей между объектами.

### 3.2.5 Корректировка результатов диффузионно-инвариантной паттерн-кластеризации

Накопление большого рода различной информации в настоящее время позволяет извлекать и практически применять разработанные алгоритмы в самых различных областях: повышение уровня продаж, выявление целевых групп и т.д. Однако, весьма важным является вопрос о точности предоставленных для обработки данных информации. При наличии погрешностей, возникающих на этапе сбора и обработки информации возможно накопление ошибки на различных этапах анализа данных, что в свою очередь может привести к ошибочным результатам. Данное утверждение особенно справедливо для диффузионно-инвариантной паттерн-кластеризации, поскольку данный метод базируется на парном сравнении показателей. Очевидным решением кажется округление данных до определенного значения. Однако, недостатком данного метода является снижение точности, что может быть критичным

для определенных задач. Разумеется, здесь следует исходить из конкретной постановки исходной задачи и предполагаемых результатов. Альтернативным подходом может являться корректировка полученных результатов с использованием центроидов диффузионно-инвариантных паттерн-кластеров. В [113] доказано утверждение, позволяющее использовать центроиды для данной задачи: “Если объект  $x_l = (x_{l1}, \dots, x_{lj}, \dots, x_{ln})$  принадлежит порядково-инвариантному паттерн-кластеру  $v_{inv}$ , тогда для любого значения  $\alpha (\alpha > 0)$  объект  $x_\alpha = \alpha x_l$  также принадлежит данному кластеру”. Таким образом, некоторый «средний объект», который образуется как среднеарифметическое значение всех показателей, вошедших в определенный порядково-инвариантный паттерн-кластер, будет принадлежать данному кластеру (при использовании аналогичной процедуры разбиения исследуемых объектов).

Приведем алгоритмическую реализацию корректировки возможной корректировки (в зависимости от поставленной задачи) результатов диффузионно-инвариантной паттерн-кластеризации. Предположим, при использовании данного метода были сформированы  $z$  диффузионно-инвариантных паттерн-кластеров  $v_{diff}$ . Для каждого подобного кластера возможен расчет центроида (или «среднего объекта») согласно формуле:

$$x_{centroid}^{v_l} = \frac{\sum_{i=1}^{|v_l|} x_i}{|v_l|},$$

где  $l$  – номер диффузионно-инвариантного паттерн-кластера.

Далее, берется необходимая в конкретной задачи мера близости (приведем пример с использованием евклидовой метрики) и рассчитывается расстояние от полученных центроидов до всех исследуемых объектов  $d_E^w(x_{centroid}^{v_l}, x_i)$ . В таком случае результаты разбиения корректируются согласно минимальному расстоянию до центроидов диффузионно-инвариантных паттерн-кластеров. Формально, рассчитывается дополнительный параметр согласно формуле:

$$g_i = \min(d_E^w(x_{centroid}^{v_1}, x_i), \dots, d_E^w(x_{centroid}^{v_z}, x_i)).$$

При этом, объект соотносим к кластеру  $v_l$ , если  $g_i = d_E^w(x^{vl}_{centroid}, x_i)$ .

Следует отметить, что в ряде случаев, применение данного метода может не изменять существенным образом результаты. Это может происходить по причине четкого выделения определенных кластеров в конкретной рассматриваемой задаче, относительно небольшом внутри кластерном расстоянии и большом меж кластерным.

### 3.2.6 Некоторые выводы

В этой части отчета продемонстрирована возможность использования различных алгоритмов сортировки данных для уменьшения временных затрат на реализацию диффузионно-инвариантной паттерн-кластеризации. Данная реализация позволяет получать относительно высокую точность разбиения при наличии выборки объектов больших размеров. Описанный алгоритм корректировки результатов позволяет работать с данными, имеющими определенную погрешность, при этом несущественно теряя точность конечного разбиения. Отметим, однако, что подобное алгоритмическое решение увеличивает вычислительную сложность рассматриваемого метода анализа паттернов.

## 3.3 Декомпозиционная сложность графов и её применение для анализа фондовых рынков

### 3.3.1 Введение

Большинство публикаций, посвящённых фондовым рынкам, в той или иной степени связаны с попытками предсказания поведения этих рынков, в том числе предсказаний больших кризисов. Анализ этих предсказаний требует отдельного обзора и не является целью данной работы. Более того, настоящая работа вообще не посвящена каким бы то ни было предсказаниям поведения фондовых рынков. Представляется, что, прежде чем приступить к задачам предсказания, оптимизации, управления для сколько-нибудь сложных систем, желательно постараться понять – хотя бы частично – как устроены эти системы. Подход к сложной системе, как к чёрному ящику, при всей своей привлекательности, вряд ли является эффективным при решении упо-

мянущих и многих других задач. В то же время разумные (пусть и неполные) представления о конкретной системе позволяют сформулировать обоснованные гипотезы о её поведении, выбрать несколько важных параметров и т. д., короче говоря, позволяют понять, «как устроен мир в этом месте» (по выражению выдающегося математика И.И. Пятецкого-Шапиро).

Перейдём от этих общих и поэтому весьма расплывчатых рассуждений к конкретной системе – фондовому рынку. В настоящей работе вводится в рассмотрение новый численный показатель, характеризующий ежедневное состояние фондового рынка. Оказалось, что этот показатель разделяет все рабочие дни на две очень сильно отличающиеся группы, которые условно можно назвать стабильной и переходной. Указанное разделение установлено для достаточно длительного периода 01.01.1990 – 30.09.2019. Это разделение не соответствуют кризисам на рынке и даже какому-то заметному росту или падению цен, но представляется, что в дни, входящие в переходную группу, происходит заметное изменение кластерной структуры множества акций.

Не менее важным является распределение этих групп во времени. С начальной точки наблюдений и до 1 июля 2014 года число стабильных дней в десятки раз превышало число переходных. А после 1 июля 2014 года, наоборот – число переходных дней в десятки раз превышает число стабильных. При этом сколько-нибудь заметных кризисов, да и вообще резкого изменения цен акций летом 2014 года не наблюдалось.

Следует отметить, что рассматриваемый показатель строится по хорошо известному рыночному графу, описывающему поведение фондового рынка. Однако все понятия и алгоритмы могут быть применены к произвольным неориентированным графикам и, тем самым, к анализу многочисленных систем, моделируемых такими графиками. Сделанные же выводы относятся именно к фондовому рынку S&P-500. Нет никаких оснований автоматически переносить эти экспериментальные результаты на другие системы, моделируемые графиками.

### 3.3.2 Базовый алгоритм построения семейства дихотомий

Центральное понятие работы – понятие глубины графа – тесно связано с проблемой декомпозиции графов. Эта проблема является одной из очень важных и всесторонне изученных проблем в теории графов, информатике, интеллектуальном анализе данных и других областях прикладной дискретной математики. Декомпозиция графа на небольшое число слабо связанных подграфов может привести к важным выводам, касающимся исходной системы, моделируемой графиком.

Однако в некоторых случаях такой единственной декомпозиции не существует совсем или существует несколько различных декомпозиций такого типа. Эти случаи не являются досадными ошибками того или иного метода. Более того, можно утверждать, что разнообразные декомпозиции одного и того же графа естественным образом возникают при изучении многих сложных систем, в том числе систем, функционирование которых определяется деятельностью человека. Поэтому представляется целесообразным рассмотреть – в качестве важного обобщения традиционной задачи декомпозиции – построение семейства декомпозиций вместо одной. Как и в традиционной постановке, построенное семейство характеризует исходную систему. Более того, в таких ситуациях сами декомпозиции, образующие вышеупомянутое семейство, не представляют заметного интереса. Оказалось, что целесообразнее сосредоточить внимание на определении и использовании специальных числовых показателей, характеризующих построенные семейства декомпозиций «в целом». Предлагаемые показатели описывают такие свойства исходной системы, которые не могут быть выявлены с помощью традиционного подхода к построению единственной «правильной» декомпозиции. Они имеют различные интерпретации в разных ситуациях, но в целом они описывают сложность, запутанность, неясность и другие подобные, с трудом определяемые формально, хотя и важные, свойства различных моделируемых реальных систем. Эти показатели можно назвать показателями декомпозиционной сложности рассматриваемых графов и моделируемых ими систем.

Первый показатель такого рода, названный сцепленностью графа (graph entanglement), был введён в препринте [118]. Там же был приведён пример использования этого понятия для анализа голосований в Думе РФ в 1996–2007 гг. Применение этого же показателя для анализа фондового рынка было продемонстрировано в публикациях [119], [120].

В настоящей статье впервые предлагается новый показатель сложности графа, названный здесь глубиной графа (graph depth). Он позволяет по-новому взглянуть на ежедневные данные по фондовому рынку и сдвинуть внимание с больших кризисов на более локальные и значительно более частые явления, которые также могут представлять интерес для финансовых аналитиков.

Ниже кратко излагается алгоритм построения семейства дихотомий произвольного неориентированного простого связного графа, опубликованный и проанализированный в публикациях [118], [119], [120]. Поскольку графы других типов в работе не рассматриваются, будем далее пользоваться термином «граф» без дополнительных уточнений. Выходом данного алгоритма является семейство разбиений множества вершин исходного графа на два подмножества (такие разбиения принято называть дихотомиями). В следующих разделах демонстрируется, как к данному алгоритму «подключаются» другие операции. В результате работы таких модификаций определяются некоторые формальные объекты и характеристики, которые могут использоваться для описания исходного графа (и, следовательно, моделируемой им системы) с различных позиций. В силу такой структуры изложения данный алгоритм назван базовым. Для краткости он обозначен через БАД (базовый алгоритм построения семейства дихотомий).

Прежде всего введём основные обозначения и определим основные параметры БАД. Через  $G(V, E)$  обозначается исходный граф с множеством вершин  $V$  и множеством рёбер  $E$ . Без ограничения общности можно считать, что  $V = \{0, 1, \dots, n-1\}$ , где  $n$  – число вершин данного графа;  $E = \{e_0, e_1, \dots, e_{m-1}\}$ , где  $m$  – число его рёбер.

БАД состоит из предварительного этапа и повторяющегося затем  $T$  раз основного этапа. Число повторений  $T$  является одним из параметров БАД. Остальные необходимые параметры и обозначения будут вводиться по ходу изложения.

Предварительный этап. На этом этапе каждому ребру исходного графа сопоставляется случайное целое число от 1 до  $F_0$  включительно ( $F_0$  является параметром БАД). Максимальное из этих случайных значений находится на данном этапе; оно обозначается через  $F_{\max}$ . Заметим, что  $F_{\max}$  может быть меньше, чем  $F_0$ , хотя обычно эти числа совпадают. Далее в процессе выполнения БАД натуральные числа, сопоставляемые рёбрам  $e_j$  графа, будут обозначаться через  $f_j$  и называться частотой в ребре  $e_j$ . Сам исходный граф  $G(V, E)$  является входом БАД.

Основной этап. Входом основного этапа являются вышеупомянутые текущие значения частот во всех рёбрах графа и текущая величина  $F_{\max}$ .

Выход основного этапа будет описан ниже (после описания выполняемых на нём операций). Блок-схема основного этапа представлена на рисунке 3.8. Ниже даётся подробное описание выполняемых шагов.

Шаг 1. С помощью стандартного датчика равномерно распределённых случайных чисел выбираются две различные вершины графа.

Шаг 2. Для двух выбранных вершин алгоритмом Дейкстры находится соединяющий их кратчайший путь. Длиной ребра является его текущая частота. Длина пути равна длине его самого длинного ребра, а не сумме длин всех его рёбер. Хорошо известно, что алгоритм Дейкстры применим и в подобных случаях, с единственным изменением: при определении продолженного пути вместо суммы длин начального участка и добавляемого ребра записывается максимальное из этих же двух чисел.

Шаг 3. Определяется максимальная частота рёбер  $F_p$  в найденном на шаге 2 пути.

Шаг 4. Если  $F_p < F_{\max}$ , то переходим к шагу 5. Иначе переходим к шагу 6.

Шаг 5. Осуществляется модификация частот: число 1 прибавляется к частотам всех рёбер пути, найденного на шаге 2. Переходим к шагу 1.

Шаг 6. Находим компоненты связности исходного графа, которые получились бы при удалении всех рёбер с максимальной частотой. Заметим, что реально никакие рёбра из графа не удаляются и исходный граф остаётся тем же самым.

Шаг 7. Компоненту с максимальным числом вершин объявляем 1-й частью очередной дихотомии графа, а другую компоненту (если она одна) или объединение всех остальных компонент (если их больше одной) объявляем 2-й частью очередной дихотомии. Запоминаем её.

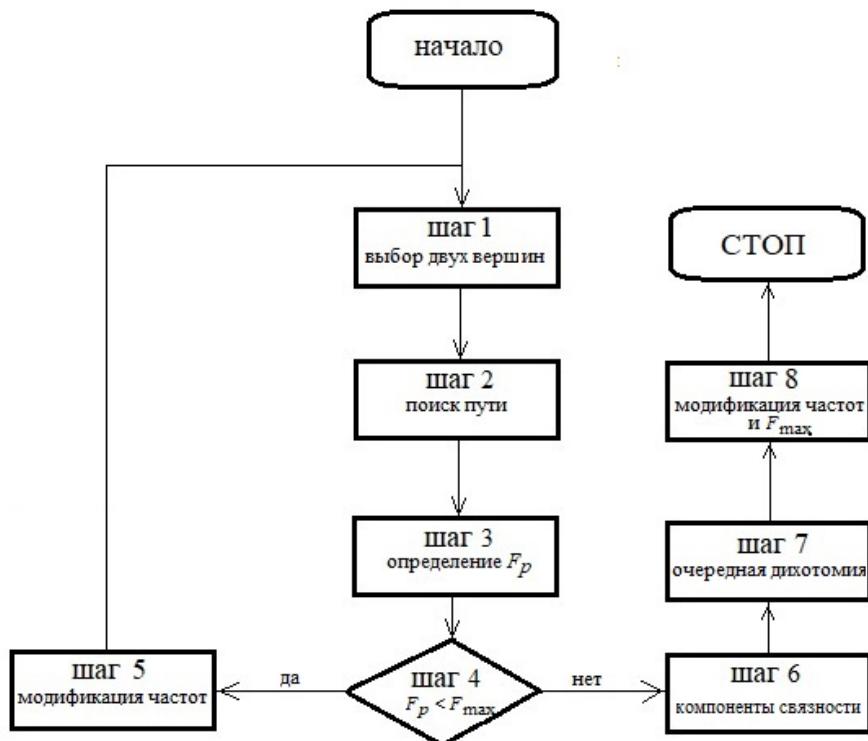


Рисунок 3.8 – Блок-схема БАД

Шаг 8. Как и на шаге 5, осуществляется модификация частот: число 1 прибавляется к частотам всех рёбер последнего пути, найденного на шаге 2. Кроме этого, увеличивается и максимальная частота в рёбрах:  $F_{\max} = F_{\max} + 1$ .

Текущее выполнение основного этапа завершено. Его выходом являются:

- модифицированные текущие значения частот во всех рёбрах графа;
- модифицированное значение текущей максимальной частоты  $F_{\max}$ ;
- очередная дихотомия исходного графа.

Если число выполнений основного этапа меньше заданного параметра  $T$ , он выполняется снова (с новым входом, определяемым выходом предыдущего выполнения основного этапа или выходом предварительного этапа при первом выполнении). В противном случае БАД останавливается. Его выходом является семейство  $T$  дихотомий множества вершин  $V$  исходного графа  $G(V, E)$ , каждая из которых была построена на очередном выполнении основного этапа.

Дадим некоторые пояснения и комментарии к основному этапу БАД. Имеется три разных случая непосредственно перед выполнением сравнения частот на шаге 4, обозначенные как случаи А, В, С. На рисунке 3.9 жирные отрезки представляют собой рёбра с максимальной частотой, в то время как тонкие линии представляют собой пути, соединяющие пару вершин  $a$  и  $b$ .

В случае А (рисунок 3.9) множество всех рёбер с максимальной частотой не содержит никакого разреза исходного графа. Поэтому найденный на шаге 2 кратчайший путь не содержит рёбер с максимальной частотой в силу минимаксного определения длины пути. Следовательно, найденная на шаге 3 максимальная частота  $F_p$  меньше, чем  $F_{\max}$ , и мы переходим на шаг 5, на котором частоты во всех рёбрах найденного пути увеличиваются на 1, после чего возвращаемся на шаг 1 основного этапа. Это соображение является центральным в данном алгоритме. Ведь если максимальная частота  $F_p$  в рёбрах на найденном пути равна  $F_{\max}$ , то это значит, что множество рёбер, частота в которых равна  $F_{\max}$ , содержит разрез графа, так что построенный путь пересекает этот разрез. Если бы эти рёбра не содержали разрез, то алгоритм Дейкстры на минимакс нашёл бы путь, во всех рёбрах которого частота была бы меньше, чем  $F_{\max}$ .

В случае В (рисунок 3.9) множество всех рёбер с максимальной частотой всё же содержит разрез исходного графа, но найденный путь не содержит

рёбер с максимальной частотой, так как оба его конца расположены с одной стороны разреза. Процесс продолжается, как в случае А.

В случае С (рисунок 3.9) множество всех рёбер с максимальной частотой содержит разрез исходного графа, и концы найденного пути расположены по разные стороны разреза. Поэтому на данном пути есть по крайней мере одно ребро, которое входит в указанный разрез. И частота в этом ребре максимальна, т. е. она совпадает с  $F_{\max}$ . Следовательно, после сравнения на шаге 4 процесс пойдёт другим образом (по шагам 6–8), так что на шаге 7 будет построена очередная дихотомия исходного графа.

Заметим, что все рассуждения в данном разделе относятся к произвольным графикам, а не только к графикам, моделирующим системы какого-либо специального вида.

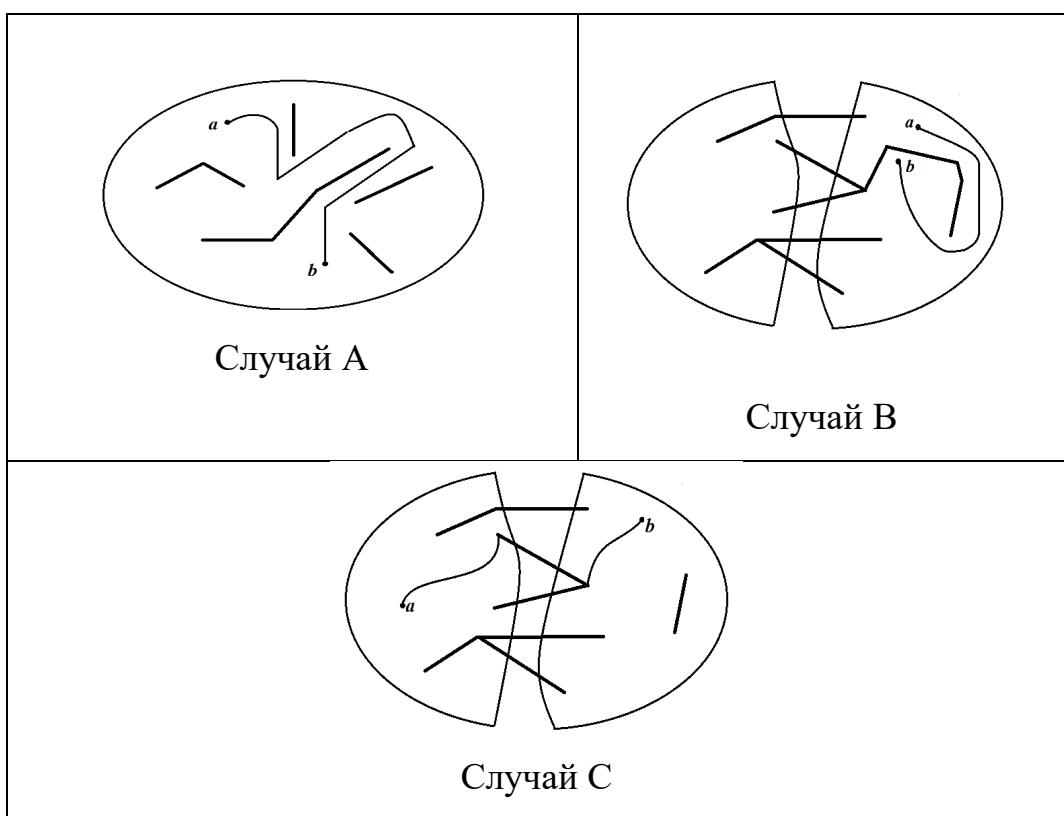


Рисунок 3.9 – Разрезы и пути на графике

### 3.3.3 БАД и показатели сложности графа

В данном разделе вводятся в рассмотрение числовые показатели построенных БАД семейств дихотомий. Во многих случаях, особенно при ана-

лизе систем, изменяющихся во времени, такие показатели характеризуют эти системы значительно полнее, чем отдельные дихотомии. Рассматриваемые ниже показатели описывают – в том или ином смысле – сложность моделирующих реальные системы графов. Поэтому их можно интерпретировать, как показатели сложности исходных моделируемых систем. Конкретный пример приводится ниже.

Определение рассматриваемых индексов сложности связано с работой вышеописанного БАД. Они существенно отличаются от многих известных показателей сложности графов, сколько-нибудь подробный обзор которых выходит за рамки данного отчёта. Ниже в этом пункте будут описаны два различных индекса сложности графа – их сцепленность и глубина.

**Сцепленность графа.** Это понятие в рамках данной тематики было введено в работе [118]. Сам по себе термин «сцепленность графа» (graph entanglement) используется достаточно давно, но он описывает совершенно другие (и по-другому определяемые) свойства графа.

Введём необходимые понятия и определения, позволяющие описать интересующие нас свойства построенного семейства дихотомий. Семейство всех дихотомий, найденных БАД, состоит из  $T$  элементов. Некоторые из них могут совпадать, а некоторые – быть различными. Пусть среди  $T$  дихотомий дихотомия  $d_p$  встречается  $m_p$  раз ( $p = 1, \dots, t$ ), где

$$\sum_{p=1}^t m_p = T. \quad (3.4)$$

Числа  $m_p$  ( $p = 1, \dots, t$ ), входящие в (3.4) подсчитываются непосредственно по семейству  $T$  дихотомий, найденных БАД. Для этого достаточно воспользоваться очевидным алгоритмом сравнения двух дихотомий. Расположим числа  $m_p$  ( $p = 1, \dots, t$ ) в невозрастающем порядке. Этот упорядоченный набор представляет собой распределение дихотомий для данного графа  $G$  при заданном параметре  $T$ .

Интуитивно понятно, что свойства этих распределений отражают свойства исходных графов и, следовательно, свойства рассматриваемых систем, представленных этими графами. Они имеют различные содержательные интерпретации, но в общем описывают сложность, запутанность, неясность, неопределенность и другие трудно формализуемые, хотя и весьма важные свойства разнообразных систем. Примеры минимального ( $t = 1$ ) и максимального ( $t = T$ ) распределений приводятся в публикациях [118], [119], [120].

Введём формальный показатель, который количественно выражает свойства этих распределений в произвольном графе. Положим

$$E = -\sum_{p=1}^t \mu_p \ln(\mu_p), \quad (3.5)$$

где  $\mu_p = m_p / T$ . Число  $E$  представляет собой обычную энтропию разбиения  $T$  элементов на  $t$  частей. Очевидно, что минимально возможное значение  $E$  равно 0, в то время как максимально возможное значение равно  $\ln(T)$ . В первом случае имеется только одна группа совпадающих элементов, во втором случае имеется  $T$  групп, состоящих из одного элемента каждого. Положим

$$I = E / \ln(T), \quad (3.6)$$

где  $E$  определено формулой (3.5).  $I$  является случайной величиной, поскольку случайный датчик многократно использовался как на предварительном, так и на каждом из  $T$  основных этапов БАД. По построению,  $0 \leq I \leq 1$ . Чтобы подчеркнуть зависимость  $I$  от числа  $T$ , будет использовать обозначение  $I(T)$ . Математическое ожидание  $\bar{I}(T)$  этой случайной величины можно назвать сцепленностью рассматриваемого графа.

Предполагается, что введённое понятие сцепленности отражает важные свойства многих систем, упомянутые ранее в этом разделе. Но само это значение априорно неизвестно и предложенный БАД даёт только его аппроксимацию  $I(T)$ . Экспериментальная оценка дисперсии этой аппроксимации была рассмотрена в статье [120]. Сравнительно низкие значения дисперсии обес-

печивают воспроизводимость и надёжность экспериментальных результатов, описывающих поведение фондовых рынков и голосующих органов. Подробнее в данном отчёте эти опубликованные материалы не рассматриваются. Обратим ещё раз внимание на то, что сцепленность графов определяется только после выполнения БАД, а оценки дисперсии – только *после* некоторого числа повторений этого алгоритма.

Глубина графа. Переходим к изложению центрального нового понятия данной работы – глубины графа. Содержательные обоснования его введения примерно те же, что и выше – попытки придать точный смысл содержательным понятиям сложности, запутанности, неясности графов. Однако формальное определение и свойства этого понятия являются другими, а возможности его использования, в частности, при анализе фондовых рынков, оказываются большими, что будет продемонстрировано ниже.

Переходим к формальным конструкциям. Для этого вернёмся к БАД и его блок-схеме на рисунок 3.7. Рассмотрим подробнее внутренний цикл  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1$ . Этот цикл многократно повторяется, так что при всех его проходах, кроме последнего, имеют место случаи А или В и выполняется проверяемое на шаге 4 условие  $F_p < F_{\max}$ . А на последнем проходе имеет место случай С и равенство  $F_p = F_{\max}$ . Заметим также, что на каждом проходе цикла  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1$  при выполнении шага 2 находится (алгоритмом Дейкстры) кратчайший путь между случайно выбранными вершинами графа, т. е. на этот шаг приходится подавляющее большинство вычислительных операций данного цикла.

Проведём содержательные рассуждения. Предположим, что структура графа достаточна ясна, т. е. имеется два соизмеримых по числу элементов множества вершин, соединённых сравнительно небольшим числом рёбер. При начале каждого выполнения основного этапа – начиная со второго раза – частоты во всех этих рёбрах совпадают, кроме частоты в одном ребре, через которое проходит путь, найденный последним на предыдущем основном этапе. В этом ребре частота на единицу больше, чем в остальных соединяющих

рёбрах (см. шаг 8 БАД). Поскольку ситуации, когда одна из случайно выби-раемых на шаге 1 вершин лежит в одной части, а другая вершина – в другой части, будут достаточно частыми (в силу соизмеримости числа вершин в каждой из частей), то столь же частой будет и необходимость проведения соответствующего пути по одному из соединяющих рёбер. И в этом ребре частота станет равна максимальной, откуда следует, что при следующем таком пути частота станет максимальной в другом соединяющем ребре и т.д. Но ведь мы предположили, что число соединяющих рёбер невелико. Поэтому и число прохождений цикла  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1$  также сравнительно невелико.

Однако такая благостная картина имеет место далеко не всегда. Например, если в одной части лежит 495 вершин, а в другой части – 5, то в значительном большинстве случаев на шаге 1 будут выбираться вершины из большей части, так что до завершения внутреннего цикла понадобится много шагов. Примерно то же самое будет и в другом случае, когда велико число соединяющих части рёбер, и пр. Самое главное состоит в том, что большие значения глубины графа указывают на отсутствие чёткой структуры кластеров. Заметим, что речь идёт не о фиксированном числе кластеров. Понятно, что наличие нескольких чётких кластеров определяет и наличие двух чётких кластеров (путём их частичного объединения). Поэтому отсутствие двух чётких кластеров говорит и об отсутствии чёткой кластерной структуры вообще.

Таким образом, качество дихотомии в значительной степени определяется числом прохождений упомянутого цикла  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1$ . А это число легко определяется одновременно с выполнением БАД. Достаточно ввести ещё одну переменную  $S$ , присвоить ей значение 0 на предварительном этапе и вставить шаг  $S = S + 1$  между шагами 5 и 1 во внутреннем цикле блок-схемы БАД. Будем считать величину  $S$  ещё одним элементом входа на основной этап, а модифицированную на основном этапе величину  $S$  – ещё одним элементом его выхода. Наконец, после последнего выполнения основного этапа

надо положить  $S = S / T$ . Чтобы указать на зависимость  $S$  от  $T$ , будем пользоваться обозначением  $S(T)$ .

Как и нормализованная энтропия  $I(T)$ , величина  $S(T)$  является случайной. Глубиной графа назовём её математическое ожидание  $\bar{S}(T)$ . Конечно, вычисляемые по указанной модификации БАД значения  $S(T)$  являются оценками глубины  $\bar{S}(T)$ . Остановимся подробнее на стабильности этой величины.

Три вида стабильности глубины графа. Как уже упоминалось, понятие глубины определяется для произвольных графов. На самом нижнем уровне можно говорить о стабильности экспериментально определяемой величины  $S(T)$  при фиксированном  $T$ . Речь идёт о близости чисел  $S(T)$  при многократном повторении БАД. На следующем уровне можно задуматься о стабильности вычисления  $S(T)$  при изменении  $T$ .

В обоих случаях речь идёт об ограниченной случайной величине, и принципиальная сходимость следует из центральной предельной теоремы. «Техническая» сходимость с точностью 2–3% при разумных значениях параметров БАД имеет место для рассматриваемых конкретных графов, порождённых фондовым рынком S&T-500.

Наконец, имеет смысл посмотреть более широкую картину – что произойдёт со значениями  $S(T)$  при небольших изменениях исходного графа. В данном случае под изменением графа можно понимать удаление сравнительно небольшого числа вершин и инцидентных им рёбер.

Нетрудно понять, что для произвольных графов в последнем случае стабильность не гарантируется, что может сделать использование указанной характеристики малоэффективной. Общее теоретическое исследование этой проблематики в настоящей работе не осуществляется. Однако приводимые ниже экспериментальные результаты для рыночных графов дают увереные положительные ответы о данном виде стабильности, что обосновывает возможность использования глубины графов при анализе фондовых рынков.

### 3.3.4 Матрица расстояний и построение отдельных дихотомий

Матрица расстояний. В двух предыдущих разделах предполагалось, что исходная информация о рассматриваемой системе представлена в виде графа. И в приведённом БАД, и при определении обоих индексов сложности никакая другая информация не использовалось. Однако во многих случаях имеется значительно более детальная информация об исходной системе. Именно, в системе выделяются отдельные объекты, для которых определены расстояния между ними. Такими объектами в голосующих органах являются депутаты; для фондовых рынков такими объектами являются акции отдельных компаний и т. д. Упомянутые расстояния соответствуют непохожести между объектами, которая определяется по-разному для различных систем. В данном пункте предполагается, что непохожесть между объектами задаётся матрицей попарных расстояний

$$D = (d_{ij}) \quad (i, j = 1, \dots, n), \quad (3.7)$$

где через  $d_{ij}$  обозначено расстояние между объектами  $i$  и  $j$  (здесь и далее считается, что номера объектов в общем списке и являются их именами). Предполагается, что все числа в матрице расстояний неотрицательны;  $d_{ij} = 0$  означает минимально возможное расстояние между объектами  $i$  и  $j$ . Чем больше число  $d_{ij}$ , тем меньше похожи друг на друга объекты  $i$  и  $j$ . Заметим сразу, что  $d_{ij} = 0$  не означает, что объект  $i$  полностью совпадает с объектом  $j$ . Например, для двух депутатов  $i$  и  $j$  равенство  $d_{ij} = 0$  означает только то, что они голосовали по всем рассматриваемым в течение некоторого периода времени вопросам одинаково. Поводы к именно такому голосованию у них могли быть заметно различными. Для удобства положим  $d_{ii} = \infty$  ( $i = 1, \dots, n$ ). Это позволит избежать появления петель при вершинах во всех дальнейших алгоритмах. Наконец, предполагается, что  $d_{ij} = d_{ji}$  (симметричность). Заметим, что построение матрицы расстояний, как и многих других моделей реальных систем, является зачастую весьма сложным и творческим делом. Однако для упомянутых выше систем – голосующих органов и фондового рынка – по-

строение матриц расстояний подробно описано в литературе. Эти известные конструкции описаны ниже.

**Граф соседства.** Из подробного описания исходной системы матрицей расстояний нетрудно получить её более сжатое (но во многих случаях не менее полезное) описание графом. Приведём известный алгоритм построения этого графа, обычно называемого графом соседства.

#### Алгоритм построения графа соседства

Входом данного алгоритма является матрица расстояний  $D$  размера  $n \times n$ . Напомним, что через  $n$  обозначено число объектов в рассматриваемой системе.

Шаг 1. Определим целочисленную матрицу  $A$  размера  $n \times n$  и положим в ней все элементы  $a_{ij}$  равными 0.

Шаг 2. Для каждого  $i = 1, \dots, n$  выполним следующие операции.

2.1. Определим (по матрице расстояний  $D$ ) индексы  $i_1, i_2, i_3, i_4, k_1, \dots, k_s$ , такие, что

$$d_{ii_1} \leq d_{ii_2} \leq d_{ii_3} \leq d_{ii_4} = d_{ik_1} = \dots = d_{ik_s}, \quad (3.8)$$

и все остальные элементы  $i$ -ой строки матрицы  $D$  строго больше, чем выписанные в цепочке (3.8). Таким образом, объекты  $i_1, i_2, i_3, i_4, k_1, \dots, k_s$  являются ближайшими к объекту  $i$ . При этом объекты с номерами  $k_1, \dots, k_s$  находятся от объекта  $i$  на одном и том же расстоянии  $d_{ii_4}$ . Заметим, что само наличие объектов  $k_1, \dots, k_s$ , занимающих места после четвёртого в цепочке (3.8) не гарантируется: объектов, удовлетворяющих требуемым неравенствам, может просто не быть. В то же время существование объектов  $i_1, i_2, i_3, i_4$ , занимающих первые четыре места в цепочке 5, гарантируется (если  $n > 4$ ).

2.2. Положим  $a_{ij} = 1$  для всех  $j = i_1, i_2, i_3, i_4, k_1, \dots, k_s$ .

Шаг 3. Для всех  $i, j = 1, \dots, n$  выполним следующие действия

Если  $a_{ij} = 1$ , положим  $a_{ji} = 1$ . В противном случае переходим к следующей паре объектов  $i', j'$ .

Выходом рассмотренного алгоритма является построение матрицы смежности графа соседства. Легко видеть, что построенный граф не зависит от конкретных нумераций, удовлетворяющих вышеуказанным условиям.

Построение отдельных дихотомий. Во многих случаях требуется не только описать некоторые общие свойства исходной системы, моделируемой графом, но и найти дихотомии этого графа, в том или ином смысле оптимальные. Оказывается, что наличие одновременно двух способов описания исходной системы – с помощью матрицы расстояний и с помощью построенного по ней алгоритмом из предыдущего подраздела 4.2 графа соседства – позволяет найти разумные дихотомии. Соответствующий алгоритм будет описан ниже в этом разделе.

Введём необходимые формальные понятия. Пусть некоторая система описана матрицей расстояний  $D$ ,  $X$  – произвольное множество объектов системы. Поскольку рассматриваемая система описана одновременно и графом, множество вершин которого совпадает с множеством объектов системы, будет удобно считать, что  $X$  – множество вершин графа соседства, построенно-го описанным выше алгоритмом.

Обозначим через  $S(X)$  среднее значение расстояния между вершинами в множестве  $X$ . Формально

$$S(X) = \frac{1}{g} \sum d_{ij}, \quad (3.9)$$

где суммирование идёт по всем неупорядоченным парам различных вершин из множества  $X$ ,  $g$  – общее число таких пар. Очевидно, что  $g = 0.5p(p+1)$ , где  $p = |X|$ . Положим также  $S(X) = 0$  для одноэлементных множеств  $X$ .

Содержательно понятно, что  $S(X)$  характеризует плотность объектов из множества  $X$ , их близость друг к другу и т. д. Теперь можно сформулировать критерий для выбора одной дихотомии из заданного множества дихотомий. Каждая дихотомия состоит из двух подмножеств  $(X_1, X_2)$ . Рассмотрим плот-

ности  $S(X_1)$  и  $S(X_2)$ . За критерий качества дихотомии  $(X_1, X_2)$  примем наибольшее значение из двух чисел –  $S(X_1)$  и  $S(X_2)$ . Соответственно, в качестве оптимальной из всех имеющихся дихотомий выберем ту, которая минимизирует данный критерий.

Дадим общее описание алгоритма. Он состоит из двух последовательных этапов:

- 1) построение семейства дихотомий описанным в разделе 2 БАД;
- 2) нахождение среди найденного семейства оптимальной дихотомии.

Поскольку построение семейства дихотомий (обозначим его через  $Z$ ) подробно описано выше, то перейдём сразу к этапу В.

На этапе В последовательно просматриваются все дихотомии из построенного на этапе А множества дихотомий. Напомним, что их общее число равно  $T$  – одному из параметров БАД. Для каждой дихотомии двукратным (для каждой части по отдельности) применением формулы (3.9) находится значение указанного критерия. Запоминаются текущее минимальное значение критерия и соответствующая дихотомия вплоть до просмотра всех  $T$  дихотомий. Обратим внимание на то, что на этапе А дихотомии строятся разделением графа на две части на основе построения разреза между ними без использования метрической информации, (если не учитывать, что сами графы построены на основе этой информации). Но на этапе В свойства разрезов уже не учитываются, зато учитывается плотность частей, которая в явном виде никак не используется на этапе А. Представляется, что именно использование разнородной информации приводит к достаточно разумным результатам.

Построение общих декомпозиций. До сих пор рассматривались только дихотомии, т. е. разделение исходных графов или графов соседства на две части. Как и ранее в этом разделе, будем предполагать, что исходная система описана матрицей расстояний и построенным по ней графом соседства.

Рассматриваемая в этом подразделе задача состоит в разбиении исходной системы на заданное число частей. В рассматриваемых здесь постанов-

ках, как и выше, речь идёт о разбиениях графа соседства на  $K$  частей с использованием данных из матрицы расстояний.

Собственно говоря, решение состоит в многократном применении алгоритма построения одной дихотомии из предыдущего подраздела. Суть дела такова. Делим граф указанным алгоритмом на две части. После этого выбираем одну из частей и делим её на две части. Выбираем – теперь уже одну из трёх частей, делим её и т. д., вплоть до получения заданного числа  $K$  частей. Ясно, что для этого требуется  $K - 1$  дихотомий.

Единственный оставшийся вопрос состоит в том, какую именно из имеющихся частей выбирать для очередного деления. Здесь в качестве ответа можно выбирать одно из нескольких простых правил. Например, выбирать часть с наибольшим числом вершин. Или выбирать часть с наибольшим средним расстоянием  $S(X)$ . Или чередовать эти правила. Как показывают эксперименты, результаты получаются достаточно похожими. Естественно также вместо фиксированного числа  $K$  частей осуществлять дихотомии вплоть до достижения некоторого заранее определённого уровня – по максимальному числу вершин в одной части, или по максимально допустимому среднему расстоянию, или по обеим границам сразу.

Ещё раз обратим внимание на то, что все простые алгоритмы данного пункта включают в себя – в качестве центральной части – базовый алгоритм дихотомий (БАД).

### 3.3.5 Вычисление глубины графов для фондового рынка

Хотя все введённые выше понятия носят общий характер и относятся к произвольным неориентированным графикам, специальные – так называемые рыночные графы, порождённые фондовыми рынками, в этом разделе рассмотрены подробно.

Рассмотрен рынок S&P-500 (500 крупнейших компаний в США). Прежде всего дадим описание хорошо известной графовой модели произвольного фондового рынка. Объекты соответствуют рассматриваемым (в те-

чение некоторого периода) акциям. Расстояние между двумя акциями определяется следующим образом.

1) Определим базисный минимальный период, состоящий из  $l$  последовательных рабочих дней. Все данные, найденные для периода  $x, x-1, \dots, x-l+1$ , сопоставляются дню  $x$ . Положим длину  $l$  рассматриваемого периода равной 15. Такой выбор связан с содержательными соображениями: для короткого периода данные слишком вариативны, для длинного периода – слишком сглажены.

2) Цены всех акций на момент закрытия биржи рассмотрены для дней  $x, x-1, \dots, x-l+1$ . Матрица  $R$  попарных коэффициентов корреляции считается, исходя из этих цен.

3) Расстояния  $d_{ij}$  между двумя акциями (например,  $i$  и  $j$ ) определяется формулой  $d_{ij} = 1 - r_{ij}$ , где  $r_{ij}$  – соответствующий элемент матрицы  $R$ . Указанное расстояние  $d$  близко к 0 для «очень похожих» акций и близко к 2 для «очень непохожих» акций. Поэтому матрица  $D = (d_{ij})$  рассматривается как матрица расстояний.

4) По матрице расстояний алгоритмом из предыдущего пункта определяется граф соседства. Этот граф также сопоставляется последнему дню  $x$  из рассматриваемого периода.

На первый взгляд, описанная хорошо известная конструкция не вызывает никаких вопросов. Предположим теперь, что начальный день  $x$  – это сегодняшнее число. Это означает, что ничего о поведении биржи в последующем неизвестно. Заметим, что это относится не только к курсам акций, но и к самому составу S&P-500. Зато о прошлом, естественно, известно всё.

Поэтому для построения рыночного графа для текущего (сегодняшнего) дня  $x$  нужно взять все акции из сегодняшнего списка S&P-500, которые существуют на этом рынке не менее 15 последних дней. Поскольку изменение списка S&P-500 случается достаточно редко (5–15 раз в год), то таких акций обычно бывает столько же, сколько в день  $x$ . В редких случаях их бывает на одну или две меньше.

Описанный выбор акций хорошо подходит для постоянного наблюдения за курсами. Однако для рассматриваемого в статье большого периода почти в 30 лет такой анализ прошлого потребует обработки большого объёма данных, поскольку каждый день войдёт в 15 списков (напомним, что длина базисного периода  $l = 15$ ) и соответствующего большого объёма скачивания данных из финансовой базы данных Блумберга. Суть дела в том, что для скачивания данных надо указать последний день периода и его первый день. При этом будут представлены данные (цены на момент закрытия) по всем акциям, которые были на рынке в последний день периода и во все предшествующие дни указанного в запросе периода. Естественно, что при изменении последнего дня весь список незначительно, но всё же может измениться.

Поэтому для анализа данных за длительный период предлагается некоторое упрощение. В качестве конечных дней были рассмотрены последние рабочие дни в каждом квартале, т. е. непосредственно перед 31 марта, 30 июня, 30 сентября и 31 декабря. Скачивались данные о ежедневных курсах акций, начиная с одного из этих дней, на квартал + ещё 14 дней назад. То есть данные о курсах в рабочие дни, начиная с 31 марта по 1 января + ещё 14 последних рабочих дней в предшествующем декабре; с 30 июня по 1 апреля + ещё 14 последних рабочих дней в предшествующем марте; с 30 сентября по 1 июля + ещё 14 последних рабочих дней в предшествующем июне; с 31 декабря по 1 октября + ещё 14 последних рабочих дней в предшествующем сентябре, и т. д.

Рассматривались все акции, торговавшиеся в последний рабочий день квартала и присутствовавшие ранее на рынке непрерывно в течение всего квартала + 14 последних рабочих дней в предшествующем месяце. Число таких акций во многих случаях меньше, чем число акций, торговавшихся в последний день периода + предшествующие 14 дней. Дело в том, что некоторые акции, торговавшиеся в последние 15 дней периода, поступили на рынок позднее, чем три с лишним месяца назад. И хотя их число обычно очень невелико (2–5), было бы желательно убедиться, что исключение этих акций ма-

ло влияет на глубину обоих графов, построенных по данным за разное число дней и содержащих разное число вершин.

Конкретнее, рассмотрим для примера 3-й квартал 1993 года. Одна акция (с номером 131) поступила на рынок 14 июля, а вторая (с номером 193) – 27 июля. Для одновременного анализа всего квартала обе акции были удалены, и на всём квартале рассматривались остальные 498 акций, которые присутствовали на рынке всё время – с 14-го рабочего дня до начала квартала (11 июня) вплоть до его последнего рабочего дня (30 сентября). Таким образом, глубина графа была посчитана для всех 64-х рабочих дней с 1 июля до 30 сентября (напоминаем, что курсы акций для дней до начала квартала использованы только для вычисления коэффициентов корреляции, отнесённых к дням из собственно квартала).

Для сравнения рассматривался период с 27 июля до конца квартала (30 сентября). На нём присутствовали эти же 498 акций + 2 упомянутые акции, поступившие позднее начала квартала, но не позднее 27 июля. Указанный сокращённый период состоит из 47-ми рабочих дней, из которых первые 14 используются для вычисления коэффициентов корреляции, а глубина графов подсчитывается для последних 33-х рабочих дней. Именно эти 33 числа (см. нижнюю часть Таблицы 3.2) сравниваются с уже подсчитанными последними 33 из 64-х чисел, подсчитанных для всего квартала, но без двух позднее поступивших акций (см. верхнюю часть Таблицы 3.2).

Таблица 3.2 – Глубина графов для сокращённого и полного списка акций

84.07	92.48	103.60	103.69	89.72	63.49	102.45	104.66
107.54	95.96	70.69	96.55	88.77	103.39	123.76	87.99
79.99	91.85	87.33	91.23	83.39	81.68	94.83	83.96
103.57	109.19	110.13	104.47	100.44	78.30	51.06	80.98
91.81							
82.07	92.30	103.85	102.15	87.32	65.17	102.07	101.41
108.63	99.76	75.12	99.46	93.33	102.84	124.38	89.70
81.92	93.78	87.25	90.75	83.44	81.37	94.89	87.76
103.69	108.89	113.16	102.33	98.40	78.48	49.44	81.57
91.59							

В других случаях наблюдается примерно такая же техническая близость, вполне оправдывающая многократную экономию времени для проведения вычислений.

Вернёмся к рассматриваемому длительному периоду 01.01.1990 – 30.09.2019. Глубина графов подсчитана для всех 7492-х рабочих дней этого периода. В Таблице 3.3 приведены значения глубины графа соседства для первых кварталов 1990 года и 2019 года.

Обращает на себя внимание, что до 1 июля 2014 года подавляющее большинство значений глубины не превосходили 210, т. е. были малыми. Большие и средние значения почти не встречались. А после 1 июля 2014 года ситуация резко «перевернулась». В этот период, наоборот, почти все значения стали большими (превосходят 270), а малые и средние значения встречаются редко. Ниже (в Таблице 3.4) будут приведены средние значения для указанных двух периодов.

Таблица 3.3 – Глубина графов соседства для каждого рабочего дня в 1-х кварталах 1990 и 2019 годов

1990, 1 квартал																		
98.89	122.30	100.88	111.69	83.07	88.46	79.83	59.15	80.55	98.26	91.68	83.32	104.81	104.23					
76.13	47.49	90.78	99.49	73.42	138.42	140.78	136.03	100.25	97.39	102.70	98.95	105.45	103.90					
97.71	99.33	93.42	85.41	78.87	97.64	93.34	90.22	80.10	92.12	100.51	84.74	89.30	110.38	115.49				
100.09	63.69	87.60	96.53	105.43	114.27	106.14	122.31	105.46	114.11	116.97	112.40	108.48	92.56					
83.41	103.65	103.06	75.86	97.54	116.74													
2019, 1 квартал																		
416.95	398.71	397.98	398.76	394.90	391.60	396.56	398.29	399.97	400.67	395.19	395.48	398.08						
400.15	417.05	413.39	416.16	413.23	376.20	372.96	370.49	376.53	377.75	397.78	397.82	394.61						
391.87	400.37	401.60	58.76	373.97	376.07	375.87	91.18	379.96	375.99	378.53	135.33	134.62						
386.54	379.59	387.11	309.49	382.94	381.98	373.55	376.58	374.41	396.91	398.16	398.36	400.16						
400.35	399.59	401.92	395.56	393.42	392.01	368.59	377.55	197.13										

Начнём с формальных величин, непосредственно вычисляемых при построении Таблицы 3.3. Прежде всего, посмотрим на стабильность подсчитанных значений. Таковы результаты проведённых по 5 раз расчётов для двух разных дней:

Таблица 3.4 – Стабильность глубины рыночных графов

13.05.1997	391.34	391.28	391.23	391.34	391.44
15.05.1997	157.17	156.63	156.24	156.66	156.71

Примерно таковы же результаты для других дней. Можно сделать вывод, что случайная величина  $S(T)$  обладает высокой стабильностью и, следовательно, глубина графа является его достаточно устойчивой характеристикой.

В следующей Таблице 3.5 приведены средние значения глубины для указанных под-периодов, а также встречаемость разных значений.

Таблица 3.5 – Показатели для периодов

Периоды	Числа и значения	Малые значения	Промежуточные значения	Большие значения
1-ый период	Число случаев	6019	3	148
	Средние значения	93.85	244.16	384.85
2-ой период	Число случаев	143	44	1135
	Средние значения	94.76	229.32	388.17

Перейдём к анализу чисел из Таблицы 3.3, которые являются устойчивыми значениями глубины графов. Уже при поверхностном взгляде на записанные в ней значения видно, что в ней есть числа порядка 300–400 и числа порядка 50–150. Определим условный «промежуточный» уровень между 210 и 270 и разделим числа на три группы – те, которые больше 270, те, которые между 210 и 270, и те, которые меньше 210. Выбор этих границ является неформальным, как и выбор числа  $T$  повторений основного этапа в БАД равным 2000, выбор 15-дневного периода для построения матрицы корреляций, и выбор некоторых других параметров рассматриваемых алгоритмов. Как и при построениях формальных моделей многих сложных систем, для нас важны только выводы, которые можно сделать о системе на основании исследования предложенных моделей. Заметим, что в отличие от многих других работ, посвящённых фондовым рынкам, никаких предположений об исходных данных (в данном случае – курсах акций) здесь не делается.

Причины установленного резкого деления фондового рынка на две части пока непонятны (доля промежуточного уровня совсем мала). Ясно только, что это – не случайное явление. Может возникнуть иллюзия, что это связано с увеличением числа акций в S&P-500 с 500 до 505, которое происходило в течение 2013–2016 годов, как это демонстрируется в Таблице 3.6. Но это не так. Уже Таблица 3.2 показывает, что небольшое изменение числа акций, т. е. вершин в соответствующих графах, мало изменяет их глубину. Аналогичные эксперименты были проведены и в других кварталах, например, в квартале 1803, в котором было удалено 5 вершин, но характер данных не изменился.

Таблица 3.6 – Число акций в S&P-500 по кварталам

Квартал	1304	1401	1402	1403	1404	1501	1502	1503	1504	1601	1602	1603
Число акций	500	500	501	502	502	502	502	505	504	504	505	505

Возможно придумать искусственные примеры графов, в которых удаление небольшого числа вершин может сильно повлиять на их глубину, но такие графы окажутся мало похожими на реальные рыночные графы. Конечно, желательно проанализировать более широкие наборы данных. Пока можно сказать, что данное свойство рыночных графов экспериментально выглядит правдоподобным.

### 3.3.6 Некоторые выводы

Многие случаи большой глубины графов соседства с их «размазанной» кластерной структурой не связаны явно с какими бы то ни было важными событиями на фондовом рынке. Этим они и интересны. Понятно, что традиционные методы анализа поведения фондовых рынков ориентированы на большие кризисы. Однако большие кризисы случаются достаточно редко, в то время как фондовый рынок работает почти ежедневно. И именно поэтому представляется важным понять, что происходит на этом рынке в моменты,

когда никакого кризиса нет, но в определённом (и формально описанном) смысле ситуация существенно изменяется.

Желательно проанализировать работу известных программ для работы на фондовом рынке в определяемые предложенным методом даты. Конечно, для этого потребуется серьёзное сотрудничество со специалистами по фондовому рынку. Конечно же, надо провести аналогичный анализ для других фондовых рынков – в первую очередь *NASDAQ* и *Dow Jones*.

### **3.4 Разработка и исследование метода кластеризации тел свидетельств**

#### **3.4.1 Введение**

В прикладных задачах теории функций доверия тело свидетельств может иметь довольно сложную структуру: состоять из большого числа фокальных элементов со сложной структурой пересечения. Такие свидетельства трудно интерпретировать. Кроме того, поскольку многие операции теории функций доверия (например, правила комбинирования свидетельств), являются вычислительно трудными, то обработка сложных по структуре тел свидетельств с большим числом фокальных элементов становится алгоритмически невозможной.

Поэтому возникают задачи нахождения структуры сложного тела свидетельств (или соответствующей функции доверия  $Bel$ ), получение укрупненной (упрощенной) структуры тела свидетельств, нахождения (приближенной) функции доверия  $\tilde{Bel}$  более простой по структуре, но близкой в некотором смысле к функции доверия  $Bel$  и т.д.

В этом подразделе отчета приведены результаты решения этих задач на основе кластеризации множества фокальных элементов. При этом мы предполагаем, что сложная структура тела свидетельств часто является результатом агрегирования разнородной информации. Эта информация, получена из разных источников, может быть противоречивой (конфликтной). Поэтому общим подходом к кластеризации тела свидетельств может быть следующий. В полученном разбиении на кластеры исходного множества фокальных эле-

ментов противоречивость внутри кластеров должна быть минимальной, а между кластерами – максимальной. В терминах функций доверия этот принцип можно сформулировать так: в полученном разложении исходной функции доверия  $Bel$  на небольшой число функций доверия  $Bel_i$  внутренний конфликт функций  $Bel_i$  должна быть минимальным, а внешний конфликт между функциями доверия  $Bel_i$  должна быть максимальным.

В этой части отчета приведены результаты по разработке и исследованию алгоритма кластеризации тела свидетельств на основе выбора в исходном теле свидетельств так называемых «центров конфликта», т.е. таких фокальных элементов, которые не пересекаются с большим числом других фокальных элементов с большими массами. Основываясь на этих выбранных «центрах конфликта» с помощью решения оптимизационной задачи строится функция доверия  $\bar{Bel}$  близкая к исходной функции доверия  $Bel$ .

### 3.4.2 Необходимые сведения из теории функций доверия

Приведем необходимые сведения из теории функций доверия [121], [122].

Пусть  $X = \{x_1, \dots, x_n\}$  — конечное множество и  $2^X$  — множество всех подмножеств из  $X$ . Базовым доверительным назначением или функцией масс называют функцию множеств  $m: 2^X \rightarrow [0,1]$ , которая удовлетворяет условию  $\sum_{A \in 2^X} m(A) = 1$ .

Подмножество  $A \subseteq X$  называют фокальным элементом, если  $m(A) > 0$ . Пару  $F = (A, m)$  из множества всех фокальных элементов  $A \equiv \{A\}$  и соответствующей функции масс  $m(A)$ ,  $A \in A$  называют телом свидетельства. Пусть  $F(X)$  – множество всех тел свидетельств на  $X$ , а  $P(X)$  – множество всех вероятностных мер на  $X$ .

Телу свидетельств  $F = (A, m)$  можно поставить во взаимно однозначное соответствие функцию доверия (belief function)  $Bel(A) = \sum_{B \subseteq A} m(B)$  и

функцию	правдоподобия	(plausibility	function)
---------	---------------	---------------	-----------

$$Pl(A) = 1 - Bel(A^c) = \sum_{A:B \cap A \neq \emptyset} m(B), \text{ где } A^c - \text{дополнение множества } A.$$

Функция доверия (и тело свидетельств) называется:

- категоричной (categorical), если она имеет только один фокальный элемент  $A \in 2^X$ ; пусть  $\eta_{\langle A \rangle}(B) = \begin{cases} 1, & A \subseteq B, \\ 0, & A \not\subseteq B. \end{cases}$
- бессодержательной (vacuous), если единственным фокальным элементом этой функции является всё множество  $X$ ,  $F_X = (X, 1)$ ;
- консонантной (consonant), если её фокальные элементы являются вложеными, т.е.  $\forall A, B \in A : A \subseteq B$  или  $B \subseteq A$ ;
- простой, если БДН имеет не более двух фокальных элементов и, если их два, то  $X$  – один из них.

Любую функцию доверия  $Bel(A) = \sum_{B \subseteq A} m(B)$  можно записать в виде

$$Bel = \sum_B m(B) \eta_{\langle B \rangle}.$$

В теории функций доверия развит инструментарий комбинирования тел свидетельств. Под правилом комбинирования понимают некоторую операцию  $\otimes : F(X) \times F(X) \rightarrow F(X)$ . Наиболее популярным является (ненормализованное) правило Демпстера [121]  $\otimes_{ND} : m_{ND}(A) = \sum_{B \cap C = A} m_1(B)m_2(C)$ ,  $\forall A \in 2^X$ .

Величина  $K = K(F_1, F_2) = m_{ND}(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B)m_2(C) \in [0, 1]$  называется канонической мерой (внешнего) конфликта и характеризует степень конфликтности источников информации, описываемых телами свидетельств  $F_1$  и  $F_2$ : чем больше значение этого параметра, тем более противоречивую информацию предоставляют источники.

### 3.4.3 Общая схема построения упрощенной функции доверия путем кластеризации тела свидетельств

В основе предлагаемого алгоритма лежит идея, что структуры исходного тела свидетельств и упрощенного (со значительно меньшим числом фо-

кальных элементов) должны быть близкими. В частности, они должны иметь близкие «центры конфликта», т.е. таких фокальных элементов, которые не пересекаются с большим числом других фокальных элементов с большими массами.

Поэтому предлагается разложить (декомпозировать) исходное свидетельство на небольшое число простых по структуре тел свидетельств, состоящих из достаточно далеко расположенных относительно друг друга «центров конфликта».

Рассмотренный в этом разделе алгоритм кластерной декомпозиции свидетельств является в некотором смысле аналогом популярного в анализе данных алгоритма Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [123].

Основная идея алгоритма состоит в выборе тех фокальных элементов, в которых функция плотности распределения конфликта имеет наибольшее значение и соответствующие им функции доверия достаточно далеко расположены друг от друга.

Информацию по другим подходам к аппроксимации функций доверия можно найти в [124], [125].

Рассмотрим понятия плотности распределения конфликта, построение функции доверия по выбранным фокальным элементам и вычисление расстояния между функциями доверия

#### 3.4.3.1 Функция плотности распределения конфликта

Пусть  $\mathcal{A}$  – тело свидетельств функции доверии  $Bel$ . Введем в рассмотрение функцию плотности распределения конфликта «внутри» тела свидетельств  $\psi_{Bel} : 2^X \rightarrow [0,1]$ . Значение  $\psi_{Bel}(A)$ ,  $A \in \mathcal{A}$  должна быть тем большим, чем больше фокальных элементов  $B \in \mathcal{A}$  с большими массами не пересекаются с  $A$ . Исходя из этого интуитивного предположения рассмотрим функции  $\psi_{Bel}$ , удовлетворяющую следующим условиям:

- 1)  $\psi_{Bel}(A) = 0$ , если  $\forall B \in \mathcal{A}, B \cap A \neq \emptyset$ ;

- 2)  $\psi_{Bel}(A)=1$ , если  $A \notin \mathcal{A}$ ,  $A \neq \emptyset$  и  $\forall B \in \mathcal{A} / B \cap A = \emptyset$ ;
- 3)  $\psi_{Bel}(A)=\varphi(m(A))$ , если  $A \in \mathcal{A}$  и  $\forall B \in \mathcal{A} \setminus \{A\}$ ,  $B \cap A = \emptyset$ , где функция  $\varphi : [0,1] \rightarrow [0,1]$  удовлетворяет условиям: а)  $\varphi(0)=\varphi(1)=0$ ; б)  $\varphi$  – строго вогнута на  $[0,1]$ .

Будем также считать, что  $\psi_{Bel}(\emptyset)=1$ .

Примерами функций  $\varphi$  являются:  $\varphi(t)=1-|2t-1|$ ,  $\varphi(t)=t(1-t)$ ,  $\varphi(t)=-Ct \log_a t$ ,  $a > 1$ ,  $C > 0$  (нормирующий множитель).

Из условий 1) и 2) следует, что  $\psi_{\eta_{\langle B \rangle}} \equiv 1 - \eta_{\langle B \rangle}^d$  для  $B \neq \emptyset$ , т.е. для любого  $A \in 2^{\mathcal{X}}$  верно:  $\psi_{\eta_{\langle B \rangle}}(A)=1-\eta_{\langle B \rangle}^d(A)=\eta_{\langle B \rangle}(\bar{A})=\begin{cases} 1, & \text{если } A \cap B = \emptyset, \\ 0, & \text{если } A \cap B \neq \emptyset. \end{cases}$

Примерами функции плотности распределения конфликта, удовлетвроящих условиям 1)-3), являются:

$$\text{а) } \psi_{Bel}(A)=\begin{cases} 1, & \text{if } A \notin \mathcal{A} \& B \cap A = \emptyset \ \forall B \in \mathcal{A} / \\ & m(A) \sum_{B \in \mathcal{A} \setminus \{A\}} m(B) \text{ otherwise,} \end{cases} \quad A \neq \emptyset \ (\text{здесь } \varphi(t)=t(1-t));$$

$$\text{б) } \psi_{Bel}(A)=\begin{cases} 1, & \text{if } A \notin \mathcal{A} \& B \cap A = \emptyset \ \forall B \in \mathcal{A} \\ -Cm(A)\log_2\left(1-\sum_{B \in \mathcal{A} \setminus \{A\}} m(B)\right) & \text{otherwise,} \end{cases}, \quad A \neq \emptyset,$$

где  $C \leq 2^{\frac{1}{\ln 2}} \ln 2 \approx 1.884$  (здесь  $\varphi(t)=-Ct \log_2 t$ ).

Заметим, что в случае функции плотности распределения конфликта а) имеем:  $\sum_{A \in \mathcal{A}} \psi_{Bel}(A) = \sum_{A \cap B = \emptyset} m(A)m(B) = U_{aut}(Bel)$ . Величина  $U_{aut}(Bel)$  – это так называемая мера автоконфликта [126] (величина внешнего конфликта функции доверия  $Bel$  с самой собой относительно меры конфликта, генерируемой правилом Демпстера).

Функция плотности распределения конфликта б) связана с одним из так называемых энтропийных функционалов [127]:  $\frac{1}{C} \sum_{A \in \mathcal{A}} \psi_{Bel}(A) = E(Bel)$  – мера диссонанса (measure of dissonance) [128].

Назовем функцию  $\psi_{Bel}$  квазилинейной, если функция

$$\tilde{\psi}_{Bel}(A) = \begin{cases} 1 - \frac{1}{m(A)} \psi_{Bel}(A), & A \in \mathcal{A} / \\ 0, & A \notin \mathcal{A} \end{cases}$$

является линейной, т.е.  $\tilde{\psi}_{\alpha Bel_1 + \beta Bel_2} = \alpha \tilde{\psi}_{Bel_1} + \beta \tilde{\psi}_{Bel_2} \quad \forall \alpha, \beta \in [0,1], \alpha + \beta = 1.$

**Теорема 3.1.** Функция  $\psi_{Bel}$ , удовлетворяющая условиям 1) и 2), будет квазилинейной тогда и только тогда, когда  $\tilde{\psi}_{Bel}(A) = Pl(A)$  для всех  $A \in \mathcal{A}$ .

**Следствия 3.1.** Если  $\psi_{Bel}$  квазилинейная функция распределения конфликта, удовлетворяющая условиям 1) и 2), то:

- a)  $\psi_{Bel}(A) = (1 - Pl(A))m(A)$  для всех  $A \in \mathcal{A}$ ;
- б) для  $\psi_{Bel}$  выполняется условие 3), где  $\varphi(t) = t(1-t)$ ;
- в)  $\psi_{Bel}(A) \leq m(A)$  для всех  $A \in \mathcal{A}$ .

Если  $G: [0,1]^k \rightarrow [0,1]$  – агрегирующая функция (т.е.  $G$  не убывает по каждому аргументу и  $G(\mathbf{0}) = 0$ ,  $G(\mathbf{1}) = 1$ ), а функции  $\psi_{Bel}^{(i)}$ ,  $i = 1, \dots, k$  удовлетворяют условиям 1), 2), то  $\psi_{Bel} = G(\psi_{Bel}^{(1)}, \dots, \psi_{Bel}^{(k)})$  также удовлетворяет условиям 1), 2). Кроме того, если для функций  $\varphi^{(i)}$ , соответствующих функциям  $\psi_{Bel}^{(i)}$  в условии 3),  $i = 1, \dots, k$ , существует такое  $t_0 \in (0,1)$ , что все  $\varphi^{(i)}$  не убывают на  $[0, t_0)$  и не возрастают на  $(t_0, 1]$ , то  $\psi_{Bel} = G(\psi_{Bel}^{(1)}, \dots, \psi_{Bel}^{(k)})$  удовлетворяет и условию 3) с  $\varphi = G(\varphi^{(1)}, \dots, \varphi^{(k)})$ .

В частности, функциями плотности распределения конфликта будут:

$$\psi_{Bel} = \sum_{i=1}^k \alpha_i \psi_{Bel}^{(i)}, \text{ если } \sum_{i=1}^k \alpha_i = 1, \quad \alpha_i \in [0,1] \quad \forall i = 1, \dots, k; \quad \psi_{Bel} = \min_i \psi_{Bel}^{(i)};$$

$$\psi_{Bel} = \max_i \psi_{Bel}^{(i)}; \quad \psi_{Bel} = \prod_{i=1}^k \psi_{Bel}^{(i)}.$$

Для алгоритма кластерной декомпозиции нам понадобится также процедура сужения и продолжения функции масс.

### 3.4.3.2 Сужение и продолжение функции масс

Пусть  $\mathcal{A}$  – тело свидетельств функции доверия  $Bel = \sum_B m(B)\eta_{\langle B \rangle}$ . Рассмотрим некоторое подмножество  $\mathcal{A}' \subseteq \mathcal{A}$ . Сужением функции масс  $m$  на

$A' \subseteq A$  назовем функцию множеств  $m': 2^X \rightarrow [0,1]$ ,  $m'(A) = m(A) \quad \forall A \in A'$  и  $m'(A) = 0 \quad \forall A \notin A'$ . В общем случае функция множеств  $m'$  не удовлетворяет условию нормировки  $\sum_A m'(A) = 1$ . Поэтому пара  $(A', m')$  не определяет некоторое свидетельство (и соответствующую функцию доверия). Чтобы построить функцию доверия  $Bel'$ , отражающее распределение  $m'$ , необходимо продолжить функцию множеств  $m'$  до функции масс  $\tilde{m}'$ . Это можно сделать разными способами.

Примеры продолжений:

- 1) пропорциональное продолжение:  $\tilde{m}'(A) = m'(A) / \sum_{B \in A'} m'(B) \quad \forall A \in A'$ ;
- 2) несодержательное продолжение:  $\tilde{m}'(A) = m'(A) \quad \forall A \in A \setminus X$ ,  
 $\tilde{m}'(X) = m'(X) + 1 - \sum_{B \in A'} m'(B)$ ;
- 3) оптимизационное продолжение. В этом случае значения функции масс  $\tilde{m}'(A)$ ,  $\forall A \in A'$  находятся как решение оптимизационной задачи, минимизирующей расхождение между новой функцией доверия  $Bel'$  и исходной функцией доверия  $Bel$ .

Заметим, что различные продолжения функции множеств до функции масс некоторого свидетельства используются в правилах комбинирования свидетельств. Например, в конъюнктивных правилах пропорциональное продолжение применяется в правиле Демпстера [121], а несодержательное – в правиле Ягера [129].

Если некоторое правило продолжения функции масс фиксировано, то для подмножества  $A' \subseteq A$  новое свидетельство с функцией доверия  $Bel'$  будет однозначно определяться по исходной функции доверия  $Bel$  и множеству  $A' \subseteq A$ . Поэтому такую функцию доверия будем обозначать как  $Bel_{A'}$ .

В частности, если используется несодержательное продолжение, то для любого множества  $A \in A$  функцию доверия  $Bel_A = m(A)\eta_{\langle A \rangle} + (1 - m(A))\eta_{\langle X \rangle}$  будет простой функцией доверия.

### 3.4.3.3 Расстояние между функциями доверия

Существуют различные способы оценивания расстояний между функциями доверия  $Bel_i = \sum_B m_i(B) \eta_{\langle B \rangle}$ ,  $i = 1, 2$ . Например, это можно сделать с помощью следующей популярной метрики [130]:

$$d_J(Bel_1, Bel_2) = \frac{1}{\sqrt{2}} \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)^T S (\mathbf{m}_1 - \mathbf{m}_2)}, \quad (3.10)$$

где  $\mathbf{m}_i = (m_i(A))_A$ ,  $i = 1, 2$ ,  $S = (S_{A,B})$ ,  $S_{A,B} = \frac{|A \cap B|}{|A \cup B|}$  – индекс Жаккара. В

частности, для простых функций доверия, которые в качестве продолжения мы будем рассматривать ниже, это расстояние имеет следующий вид.

Лемма 3.1. Если

$$Bel_1 = m_1(A) \eta_{\langle A \rangle} + (1 - m_1(A)) \eta_{\langle X \rangle}, \quad Bel_2 = m_2(B) \eta_{\langle B \rangle} + (1 - m_2(B)) \eta_{\langle X \rangle},$$

то

$$\begin{aligned} d_J^2(Bel_1, Bel_2) &= m_1^2(A) + m_2^2(B) - m_1(A)m_2(B) \left(1 + \frac{|A \cap B|}{|A \cup B|}\right) - \\ &\quad -(m_2(B) - m_1(A)) \frac{|B|m_2(B) - |A|m_1(A)}{|X|}. \end{aligned}$$

Следствие 3.2. Если  $A = \bar{B}$ , то  $d_J^2(Bel_1, Bel_2) = \frac{|B|m_1^2(A) + |A|m_2^2(B)}{|X|}$

В частности, если, кроме того,  $m_1(A) = m_2(B) = m$ , то  $d_J(Bel_1, Bel_2) = m$ .

Следствие 3.3. Если  $A = B$ , то  $d_J(Bel_1, Bel_2) = |m_2(B) - m_1(A)| \sqrt{\frac{|X \setminus A|}{|X|}}$ .

Следствие 3.4. Для категориальных функций доверия (т. е.  $m_1(A) = m_2(B) = 1$ )  $d_J(Bel_1, Bel_2) = \sqrt{\frac{|A \Delta B|}{|A \cup B|}}$ , где  $A \Delta B = (A \cup B) \setminus (A \cap B)$  – симметрическая разность множеств.

Другой, в определенном смысле оптимальной метрикой, является метрика Вассерштейна (Wasserstein metric), введенная в [131]. Эта метрика определяется как

$$d_w(Bel_1, Bel_2) = \min \sum_{A \in 2^X} \sum_{B \in 2^X} m(A, B) d(A, B), \quad (3.11)$$

где  $d(A, B) = |A \Delta B|$ ,  $A \Delta B = (A \cup B) \setminus (A \cap B)$  – симметрическая разность множеств. Минимум берется по всем числам  $m(A, B) \geq 0 \quad \forall A, B \in 2^X$ ,  $\sum_{A, B \in 2^X} m(A, B) = 1$ , удовлетворяющих условиям согласования

$$\begin{cases} \sum_{B \in 2^X} m(A, B) = m_1(A), A \in 2^X, \\ \sum_{A \in 2^X} m(A, B) = m_2(B), B \in 2^X. \end{cases}$$

Для простых функций доверия эта метрика будет иметь следующий вид.

Лемма 3.2. Если

$$Bel_1 = m_1(A)\eta_{\langle A \rangle} + (1 - m_1(A))\eta_{\langle X \rangle}, \quad Bel_2 = m_2(B)\eta_{\langle B \rangle} + (1 - m_2(B))\eta_{\langle X \rangle},$$

то

$$\begin{aligned} d_w(Bel_1, Bel_2) &= (m_2(B) - \min\{m_1(A), m_2(B)\})|X \setminus B| + \\ &+ (m_1(A) - \min\{m_1(A), m_2(B)\})|X \setminus A| + \min\{m_1(A), m_2(B)\}|A \Delta B|. \end{aligned} \quad (3.12)$$

Замечание 3.2. Чтобы метрика не зависела от мощности базового множества  $X$ , нормируем ее:  $d_{w,0} = \frac{1}{|X|} d_w$ . В этом случае  $d_{w,0} \in [0, 1]$ .

Следствие 3.5. Если  $A = \bar{B}$ , то  $d_{w,0}(Bel_1, Bel_2) = \frac{|A|m_1(A) + |B|m_2(B)}{|X|}$ . В частности, если, кроме того,  $m_1(A) = m_2(B) = m$ , то  $d_w(Bel_1, Bel_2) = m$ .

Следствие 3.6. Если  $A = B$ , то  $d_{w,0}(Bel_1, Bel_2) = \frac{|X \setminus A|}{|X|} |m_2(B) - m_1(A)|$ .

Следствие 3.7. Для категориальных функций доверия (т.е.

$$m_1(A) = m_2(B) = 1 \quad d_{W,0}(Bel_1, Bel_2) = \frac{|A \Delta B|}{|X|}.$$

### 3.3.4 Численный пример

3.4.4.1 Алгоритм кластерной декомпозиции свидетельств на основе функции плотности конфликта

Алгоритм кластерной декомпозиции свидетельств функции доверия  $Bel$ , основанной на использовании функции плотности распределения конфликта  $\psi_{Bel}$ , будет состоять из следующих шагов.

Алгоритм.

1) Вычисляем значения функции  $\psi_{Bel}(A)$  для всех  $A \in \mathcal{A}$ . Рассматриваем только те фокальные элементы, для которых  $\psi_{Bel}(A) \geq h_1 > 0$ , где  $h_1$  – некоторое пороговое значение. Если  $\psi_{Bel}(A) < h_1$  для всех  $A \in \mathcal{A}$ , то алгоритм останавливаем и считаем, что внутренний конфликт  $U_C(Bel) = 0$ . В противном случае упорядочиваем фокальные элементы по убыванию значений функции  $\psi_{Bel}$ :  $\psi_{Bel}(A_1) \geq \psi_{Bel}(A_2) \geq \dots$ . Получим последовательность  $\{A_1, A_2, \dots\}$ .

2) Просматривая последовательность  $\{A_1, A_2, \dots\}$  в порядке возрастания индекса, находим фокальные элементы  $\{A_{i_1}, \dots, A_{i_p}\}$ , где  $A_{i_1} = A_1$ , а каждый следующий фокальный элемент – это такой первый из нерассмотренных фокальных элементов, для которого расстояние от соответствующей функции доверия до функции доверия предыдущего элемента будет не меньше, чем  $h_2 > 0$ . Другими словами, если найдены  $A_{i_1}, \dots, A_{i_{s-1}}$ ,

$$\text{то } A_{i_s} = \arg \min \left\{ k \geq i_{s-1} \mid d\left(Bel_{\{A_k\}}, Bel_{\{A_{i_{s-1}}\}}\right) \geq h_2 \right\}.$$

Найденное множество фокальных элементов  $\boxed{\mathcal{A}} = \{A_{i_1}, \dots, A_{i_p}\}$  можно интерпретировать как «центры конфликта». В алгоритме фиксирован способ

построения функции доверия по выделенному фокальному элементу и способ измерения расстояния между функциями доверия.

Вычислительная сложность алгоритма равна  $O(|\mathcal{A}|^2)$ .

Заметим, что если в исходном теле свидетельств  $A \cap B = \emptyset$  для всех  $A, B \in \mathcal{A}$ ,  $A \neq B$  (абсолютный конфликт), то это же свойство сохраняется и для выделенных после кластеризации свидетельств:  $A_{i_k} \cap A_{i_l} = \emptyset$  для всех  $k, l \in \{1, \dots, p\}$ ,  $k \neq l$ . И наоборот, если  $\bigcap_{A \in \mathcal{A}} A \neq \emptyset$  (отсутствие конфликта), то, согласно 1)  $\psi_{Bel}(A) = 0$  для всех  $A \in \mathcal{A}$  и алгоритм остановит свою работу на первом шаге, не выделив ни одного фокального элемента.

#### 3.4.4.2 Построение функции доверия по выделенному множеству фокальных элементов и оценивание внутреннего конфликта

Чтобы построить функцию доверия по множеству выделенных фокальных элементов  $\mathbb{A} = \{A_{i_1}, \dots, A_{i_p}\}$ , необходимо определить значения функции масс  $a_{i_k} = m(A_{i_k})$ ,  $k = 1, \dots, p$ . Это можно сделать с помощью минимизации функционала расхождения между исходной функцией доверия  $Bel = \sum_{A \in \mathcal{A}} m(A) \eta_{\langle A \rangle}$  и функцией доверия  $\bar{Bel} = \sum_{k=1}^p a_{i_k} \eta_{\langle A_{i_k} \rangle}$  с неизвестными коэффициентами  $a_{i_k}$ ,  $k = 1, \dots, p$ . Функционал минимизации должен удовлетворять условиям низкой вычислительной сложности, независимости от мощности базового множества и количества выделенных фокальных элементов. Например, функционалы, основанные на метриках (3.10) или (3.11), являются вычислительно сложными.

Будем рассматривать следующие функционалы среднеквадратичной невязки:

$$D_{BetP}(a_{i_1}, \dots, a_{i_p}) = \sqrt{\frac{1}{2} \sum_{x \in X} (BetP_{Bel}(x) - BetP_{\bar{Bel}}(x))^2}$$

– невязка пигнистических вероятностей  $BetP_{Bel}(x) = \sum_{A:x \in A} \frac{m(A)}{|A|}$ , совпадающая со значением Шепли для функции множеств  $Bel$ ;

$$D_{Bel}(a_{i_1}, \dots, a_{i_p}) = \sqrt{\frac{1}{2^{|X|}-1} \sum_{A \subseteq X} (Bel(A) - \bar{Bel}(A))^2}$$

– невязка функций доверия;

$$D_{U_C}(a_{i_1}, \dots, a_{i_p}) = k(Bel, \bar{Bel}) = \sum_{k=1}^p a_{i_k} \sum_{B: B \cap A_{i_k} = \emptyset} m(B)$$

– мера конфликта правила Демпстера между функциями доверия  $Bel$  и  $\bar{Bel}$ ;

$$D_{U_C,2}(a_{i_1}, \dots, a_{i_p}) = \sqrt{\sum_{k=1}^p a_{i_k}^2 \sum_{B: B \cap A_{i_k} = \emptyset} m(B)}$$

– мера квадратичного конфликта между функциями доверия  $Bel$  и  $\bar{Bel}$ .

Все невязки нормированы так, что соответствующие функции принимают значения из промежутка  $[0,1]$ . Заметим, что функция  $D_{U_C}$  является линейной. Поэтому ее минимум будет достигаться в вершине симплекса  $S_p = \{(a_{i_1}, \dots, a_{i_p}) : a_{i_k} \geq 0 \forall k, \sum_{k=1}^p a_{i_k} = 1\}$ . И при использовании такого минимизирующего функционала  $Bel$  будет иметь только один фокальный элемент. Поэтому вместо функции  $D_{U_C}$  будем использовать функцию  $D_{U_C,2}$ , близкую к  $D_{U_C}$  по смыслу, но имеющую «среднеквадратичный» вид.

Заметим, что функция  $D_{BetP}$  имеет самую низкую вычислительную сложность  $O(|A|)$ . В то время как функции  $D_{Bel}$ ,  $D_{U_C}$  и  $D_{U_C,2}$  –  $O(|A| |A|)$ .

Решая задачу минимизации одного из указанных функционалов в симплексе  $S_p$ , найдем значения масс  $a_{i_k} = m(A_{i_k})$ ,  $k = 1, \dots, p$  функции доверия  $Bel$ . При этом число фокальных элементов в  $Bel$  может уменьшиться.

Замечание 3.3. В силу условия 1)  $\psi_{Bel}(X)=0$ . Поэтому если  $h_1 > 0$ , то множество  $X$  не попадет в число выделенных алгоритмом фокальных элементов (даже если  $X \in A$ ). Тогда влияние не попавших в число выделенных фокальных элементов можно учесть, если искать функцию доверия  $Bel$  в виде  $Bel = \sum_{k=1}^p a_{i_k} \eta_{\langle A_{i_k} \rangle} + a_0 \eta_{\langle X \rangle}$ .

### 3.3.4.3 Кластеризация тела свидетельств ответов партий

Рассмотрим функцию доверия  $Bel$  свидетельства ответов партий Германии перед выборами в Бундестаг в 2013, построенного на основе минимизации расхождения между агрегированными ответами партий и результатами голосования. Подробное описание методики получения этого свидетельства можно найти в [132]. Рассматривается множество из 8-ми вопросов,  $|X|=8$ . Функция масс  $m(A)$  характеризует значимость подмножества вопросов  $A$  для голосования. Тело свидетельств состоит из 255 фокальных элементов,  $|A|=255$ . Функции масс принимают значения в промежутке  $[0, 0.0072]$ .

В кластерной декомпозиции будем использовать несодержательное продолжение для построения функции доверия по выделенному множеству фокальных элементов и метрику  $d_{w,0}$  для вычисления расстояния между функциями доверия.

Параметры алгоритма кластеризации  $h_1$  и  $h_2$  (значение  $h_2$  – доля от максимального значения функции плотности) подбирались так, чтобы число выделенных фокальных элементов было не меньше 10.

В Таблице 3.7 приведены результаты кластерной декомпозиции и вычисления приближенной функции доверия  $Bel$  для различных значений параметров:  $p$  – число выделенных фокальных элементов после алгоритма кластеризации,  $q$  – число фокальных элементов, оставшихся после минимизации соответствующего функционала и нахождения функции  $Bel$ . Все результаты получены для порога  $h_1 = 0.003$ . Столбцы таблицы соответствуют тому функционалу, относительно которого вычислялась функция доверия

$\Box_{Bel}$ . В ячейках таблицы приведены количества оставшихся фокальных элементов  $q$  и значения невязок между вычисленной функцией доверия  $\Box_{Bel}$  и исходной функции доверия  $Bel$  относительной трех разных функционалов.

Таблица 3.7 – Результаты кластерной декомпозиции и вычисления приближенной функции доверия  $\Box_{Bel}$

	$D_{BetP}$	$D_{Bel}$	$D_{U_{C,2}}$
$h_2 = 0.05$ $p = 84$	$q = 25$	$q = 36$	$q = 80$
	$D_{BetP} = 0.0005$	$D_{BetP} = 0.0072$	$D_{BetP} = 0.019$
	$D_{Bel} = 0.097$	$D_{Bel} = 0.025$	$D_{Bel} = 0.041$
	$D_{U_{C,2}} = 0.102$	$D_{U_{C,2}} = 0.033$	$D_{U_{C,2}} = 0.024$
$h_2 = 0.1$ $p = 36$	$q = 15$	$q = 20$	$q = 36$
	$D_{BetP} = 0.022$	$D_{BetP} = 0.042$	$D_{BetP} = 0.092$
	$D_{Bel} = 0.185$	$D_{Bel} = 0.1$	$D_{Bel} = 0.133$
	$D_{U_{C,2}} = 0.155$	$D_{U_{C,2}} = 0.065$	$D_{U_{C,2}} = 0.051$
$h_2 = 0.15$ $p = 19$	$q = 9$	$q = 10$	$q = 19$
	$D_{BetP} = 0.065$	$D_{BetP} = 0.094$	$D_{BetP} = 0.134$
	$D_{Bel} = 0.221$	$D_{Bel} = 0.158$	$D_{Bel} = 0.197$
	$D_{U_{C,2}} = 0.146$	$D_{U_{C,2}} = 0.099$	$D_{U_{C,2}} = 0.08$

Из анализа Таблицы 3.7 можно сделать следующие выводы:

- 1) при вычислении функции доверия  $\Box_{Bel}$  с помощью минимизации функционала  $D_{U_{C,2}}$  практически не меняется число фокальных элементов, полученное после кластеризации; при использовании других функционалов число фокальных элементов уменьшается в 1.5÷3 раза;
- 2) при сравнении результатов, полученных при одних и тех же параметрах кластеризации, видно, что более точные оценки  $\Box_{Bel}$  получаются, когда используется функционал  $D_{Bel}$ , наименее точные – при использовании функционала  $D_{BetP}$ ; в тоже время, вычислительная сложность минимизации функционала  $D_{Bel}$  является наибольшей, а функционала  $D_{BetP}$  – наименьшей;

3) при сравнении результатов, в которых получается примерно одинаковое число фокальных элементов функции  $\bar{Bel}$ , более точной оказывается оценка, полученная при минимизации функционала  $D_{Bel}$ .

### 3.4.5 Некоторые выводы

Разработан и исследован двухэтапный алгоритм кластеризации тел свидетельств. На первом этапе выделяется небольшое множество фокальных элементов на основе вычисления функции плотности конфликта. Этот этап аналогичен известному кластерному анализу алгоритму DBSCAN. На втором этапе решается задача оптимального относительно некоторого функционала невязки определения масс фокальных элементов, выделенных на первом этапе. В результате реализации алгоритма получим новое тело свидетельств (и функцию доверия), которые можно рассматривать как аппроксимацию исходного тела свидетельств (и функции доверия).

Результаты этого исследования отражены в [133].

## **4 Исследования по анализу данных**

В этом разделе отчета приведены результаты исследований в рамках проекта в области анализа данных. Эти исследования, в частности, осуществлялись в следующих направлениях:

- анализ влияния мирового изменения климата на экономики арктических стран;
- анализ среднесрочных стратегий развития Арктического региона;
- исследование причин вододефицита в странах региона Северной Африки;
- интервальный метод оболочечного анализа данных и его применение для анализа эффективности карантинных мер;
- анализ предпочтений студентов;
- анализ политик регулирования банковской системы.

### **4.1 Анализ влияния мирового изменения климата на экономики арктических стран**

#### **4.1.1 Введение**

Поскольку изменение климата в Арктике происходит быстрее, чем в среднем по планете, и эти изменения выражены уже сейчас, Арктическому региону посвящено множество исследований. Однако на данный момент не представлено комплексных исследований экономики, отражающих полную картину мирового Арктического региона. Задачей работы является сбор региональных показателей либо показателей по стране (в случае Исландии) для анализа экономики арктических территорий и дальнейшего определения отраслевой структуры экономики и занятости, а также перспективных проектов в Арктике. Мы уже наблюдаем процесс изменения климата, который оказывает влияние на экономику стран, в том числе арктических, в плане экономики, основанной на природных ресурсах. Примером может послужить рыбная промышленность, где вследствие изменения климата меняются объем и структура вылова по странам. Однако новизна темы заключается в том, что в текущей ситуации к анализу добавляется климатическая политика. Дело в

том, что все больше стран объявляют намерения о переходе на нулевые нетто-выбросы. Например, посредством Европейского зеленого курса, амбициозный проекта Европейского союза, в рамках которого Европа станет первым климатически нейтральным континентом. Безусловно, это будет способствовать вводу ограничений в экономике, посредством которых и предстоит сократить выбросы [134]. Таким образом, к климатическим изменениям, которые и так имеют влияние на экономику, в скором времени добавятся ограничительные меры, которые также будут влиять на экономику Арктики.

#### 4.1.2 Основная часть

В ходе работы были обработаны данные из государственных статистических баз данных, были рассмотрены показатели за 2000–2019 годы. Основными показателями являются: состав валового регионального продукта, структура занятости по отраслям, структура экспорта, а также специфические показатели по отраслям, наиболее развитым на арктической территории в каждой северной стране, например, объем добычи полезных ископаемых, объем и структура вылова рыбы. Также рассматриваются крупные инвестиционные проекты, показывающие ограничения и возможности развития регионов. Общей чертой для экономики всех арктических территорий является присутствие в той или иной степени горнодобывающей промышленности. Горнодобывающая промышленность может составлять основу экономики, как это происходит в Канаде, РФ и Швеции, либо же может находиться на начальных этапах развития, как в Финляндии или Норвегии. Другой отраслью, особенно развитой в Норвегии, Гренландии, Аляске, Исландии и РФ является рыбная промышленность. Например, в Гренландии морепродукты составляют 90% всего экспорта [135]. Другим направлением, активно развивающимся в Гренландии, Аляске, Исландии, Финляндии является туризм. Что касается Арктической зоны Российской Федерации, то регион крайне важен в общероссийской структуре экономики, а также в национальной безопасности и укреплении международных связей. В настоящее время доля дохода общероссийской экономики от Арктики составляет 11%, а также на долю региона

приходится до 22% общероссийского экспорта [136]. Основой арктической экономики РФ являются нефтегазодобывающая и горнодобывающая промышленность, а также энергетика (атомная энергетика, гидроэнергетика) и транспорт. В российской Арктике особенно важна минерально-сырьевая база, поскольку на территории сосредоточено большое количество углеводородов и многочисленных металлов.

В 2021 году Межправительственная группа экспертов по изменению климата опубликовала шестой оценочный доклад о состоянии климата, в который также входят различные сценарии изменения климата до 2100 года [137]. На данный момент целью является приложение странами усилий для удержания глобального потепления ниже 2°C, при этом точкой невозврата является порог 1,5°C [138]. В последнем докладе МГЭИК представлены доказательства того, что именно антропогенное влияние является главным драйвером изменения климата. В зависимости от уровня антропогенного воздействия, ученые МГЭИК выделяют 5 климатических сценариев. Лишь два из пяти сценариев предполагают выполнение обязательств по удержанию потепления на уровне до 2°C, однако оба сценария подразумевают, что начать трансформацию экономики для выхода на нулевой уровень нетто-выбросов к середине 21 века необходимо уже сейчас. Таким образом, на сегодняшний день декарбонизация экономики является одним из важнейших вызовов. Трансформация экономики приведет к тому, что некоторые сектора экономики могут исчезнуть, в то время как другие будут процветать. В работе приведен анализ на примере нефтегазодобывающей отрасли и рыбной промышленности, которые составляют основу экономического развития в некоторых Арктических регионах.

Наконец, в этом году отдельно была проанализирована арктическая экономика в 2020–2021 годы для выявления последствий экономических эффектов пандемии COVID-19 в регионе. Можно выделить высокую зависимость Арктики от мировой торговли, в частности из-за экспорта нефти, газа, морепродуктов, а также от импорта жизненно необходимых товаров, напри-

мер продуктов питания и оборудования [139]. В работе показаны ограничения экономического развития Арктики на фоне кризиса. Снижение экономической активности в Арктике из-за пандемии COVID-19 зависело непосредственно от ограничений, которые вводила каждая страна. Так, в некоторых странах, таких как Швеция, горнодобывающая промышленность замедлилась, а в России деятельность не останавливалась на протяжении года. Арктические страны, в особенности Дания (Гренландия) и Исландия, использовали закрытие границ как эффективную меру в борьбе с COVID-19 в силу своей изолированности, несмотря на то что, это привело к коллапсу международного туризма, значимого в регионах. С другой стороны, в Арктике особенно важно было не допустить распространение вируса, поскольку в отдаленных районах нельзя было рассчитывать на локальную систему здравоохранения в силу лимитированной вместительности госпиталей.

#### 4.1.3 Некоторые выводы

В работе была рассмотрена экономика арктических стран для выявления перспектив и ограничений, связанных с мировым изменением климата. Более того, внимание уделено тому, как экономика стран может поменяться в связи с ограничениями и намерением стран перейти на нулевой уровень нетто-выбросов. Работа предоставляет комплексное исследование того, как экономика арктических территорий изменилась под воздействием пандемии COVID-19. Несмотря на то, что экономические ограничения отличались в зависимости от страны, весна 2020 года характеризовалась общим снижением экономической активности в регионе. Теперь перед правительствами стоит следующим вопрос: арктические страны могут либо использовать замедление темпов роста как первый шаг к декарбонизации экономики, либо отдать приоритет восстановлению экономики любой ценой, не учитывая долгосрочное устойчивое развитие. Исследования по данной теме будут продолжены. На текущий момент результаты работы нашли отражение в курсовой работе «Уязвимость социально-экономического развития Арктики в условиях изменения климата».

## **4.2 Анализ среднесрочных стратегий развития Арктического региона**

### **4.2.1 Введение**

Арктика обладает значительным потенциалом в области добычи углеводородных ресурсов и может стать основным местом в мире для удовлетворения глобального спроса на энергию в будущем. Для транспортировки нефти, природного газа и других грузов может широко использоваться Северный Морской Путь. Кроме того, Арктика – это уникальная экосистема с широким биоразнообразием, и на ее территории региона проживает более 5 миллионов человек.

Однако сейчас арктические города и добывающие компании сталкиваются с различными неопределенностями и рисками, связанными с экологическими, экономическими и социальными факторами, такими как изменение климата, таяние вечной мерзлоты, миграция, меняющаяся институциональная среда, спрос на арктические ресурсы и другие. Особую роль в изменении климата Арктики играет фактор таяния вечной мерзлоты, который может разрушить инфраструктуру и трубопроводы. Данный фактор был изучен в рамках участия в программе «Arctic Opportunity Explorers» совместно со студентами из США и Великобритании и в результате был создан масштабируемый проект ArcticRewid, позволяющий остановить эффект таяния вечной мерзлоты.

Устойчивое развитие арктического региона возможно только при должном учете всех рисков и неопределенностей. «Фундаментом» данного исследования и послужил анализ перспектив развития Арктики и связанных с этими рисками. Для составления среднесрочного сценария и стратегий развития Арктического региона на 15 лет использовался метод сценарного планирования.

Цель исследования: создание практических рекомендаций по устойчивому развитию Арктики.

Задачи исследования:

- 1) анализ литературы о существующих перспективах и рисках, связанных с развитием Арктики;
- 2) предложение методологического подхода к оценке будущего Арктического региона;
- 3) предоставление 4 сценариев, сделанных методом сценарного планирования;
- 4) формирование практических среднесрочных стратегий с рекомендациями по развитию Арктического региона на основе методов тройного критерия.

Профессиональное значение и актуальность: многие недавние исследования ставят похожие вопросы, однако вопрос развития Арктики в ближайшие 15 лет нельзя считать широко изученным. Пандемия коронавируса показала, что нас ждет неопределенное будущее. Основные факторы, касающиеся Арктики, также неопределены: спрос и цена на ресурсы Арктики, влияние изменения климата и таяния вечной мерзлоты на регион, инновации и создание новых технологий обществом, качество институциональной среды и другие. Понимание потенциальных рисков и сценариев развития Арктики, поможет инвесторам, компаниям и организациям в Арктике принимать устойчивые и перспективные решения.

В исследовании рассмотрены основные факты об Арктике и перспективы развития региона, разобраны риски, связанные с развитием Арктики в ближайшие 15 лет, проект ArcticRewild, представляющий один из вариантов замедления таяния вечной мерзлоты. По итогу построены 4 сценария развития будущего Арктического региона, а также составлены стратегии методом сценарного планирования.

#### 4.2.2 Основная часть

Арктика – это уникальная местность с впечатляющим, но очень хрупким биоразнообразием. Существуют разные подходы к определению арктического региона в зависимости от рамок и сферы исследований [140].

В данном исследовании границы Арктики определены по следующему принципу: объединение определения границ по полярному кругу и по температурному режиму так, что площадь региона будет наибольшей.

Огромный потенциал Арктики заключается в том, что там находится много природных ресурсов от природного газа и нефти до меди, железа и золота [141]. Сообщества коренных народов Арктики сталкиваются с огромными миграционными изменениями, которые происходят во всем мире. Сами в Скандинавии, а также коренные народы из России и Аляски (США) покидают свою естественную среду обитания и теряют свою историческую идентичность. Многие российские арктические города становятся заброшенными, и там остаются только пожилые люди и вахтовики.

Несмотря на свое удаленное местоположение, Арктика глубоко интегрирована в мировую экономику. Место и роль экономики Арктики в глобальной экономике, и в экономике России на данный момент в первую очередь заключается в обеспечении энерго и биоресурсами, и сырьем для промышленности, поэтому Арктика в современной экономике играет роль «территории будущего» углеводородов [142].

Несмотря на все перспективы развития Арктического региона, существует огромное количество различных рисков. Наиболее важные экологические и социо-экономические риски с потенциальными вариантами их уменьшения и предотвращения были проанализированы: демографические факторы, изменение климата, таяние вечной мерзлоты, разрушение инфраструктуры, финансовое стимулирование и инвестиции. Эти риски и факторы использовались для дальнейшего создания сценариев развития Арктического региона.

Одним из важнейшим эффектов изменения климата в Арктике стало таяние вечной мерзлоты [143]. Данный экологический риск развития Арктики был широко изучен во время участия в программе Arctic Opportunity Explorers. Программа была успешно закончена, и наша международная команда заняла 1 место по итогам финальной защиты проектов. Мы представили ини-

циативу Rewilding на Аляске, как естественного решения проблемы таяния вечной мерзлоты и создали “Arctic ReWild” как природоохранное решение проблемы таяния вечной мерзлоты.

Далее были построены четыре возможных сценария развития российской Арктики к 2036 годам (Таблица 4.1). Сценарии основаны на двух критических неопределенностях: качество институциональной среды и спрос на ресурсы Арктики. Для прогнозирования использовался метод сценарного прогнозирования, так как он является наиболее эффективным в средней и долгосрочной перспективе и при средней или высокой степени неопределенности. Также выбор этого метода связан с тем, что невозможно в полной мере провести прогноз количественными методами прогнозирования (статистическими и математическими моделями) из-за недоступности полного объема данных для анализа [144].

**Таблица 4.1 – Этапы сценарного планирования**

Номер этапа	1	2	3	4	5	6
Этап	Начало	Подготовка сценария	Неопределенности	Сценарии	Стратегии	Контролировать и применять
Описан ие	Определен ие рамок	Анализ ключевых факторов	Выявление двух факторов наибольшей неопределенности	Построение сценариев	Создание стратегий	Оценить стратегические последствия каждого сценария
Действие	Определить основную проблему	Изучить ключевые движущие силы и риски, влияющие на сценарий	Обсудить и оценить тенденции и неопределенности	Построить 4 сценария и детализировать их	Разработать стратегии и планы действий для сценариев	Постоянно сравнивать сценарии с реальным миром. Реализовывать и совершенствовать стратегии

**Этапы сценарного планирования для нашей задачи:**

- этап 1: рамки исследования определены во вступлении к работе; рассматривается Арктический регион и его развитие;
- этап 2: анализ ключевых перспектив и рисков для развития Арктики проведен;
- этап 3: выявление основных неопределенностей.

В контексте развития Арктики существует огромное количество многофакторных неопределенностей (Таблица 4.2). Сгруппируем все неопреде-

ленности Арктики на несколько категорий и подкатегорий, чтобы далее выбрать 2 основные из них.

Таблица 4.2 – Категории неопределенностей, факторы, влияющие на развитие Арктики, и оценка неопределенности факторов и уровня влияния

Неопределенность	Фактор, зависящий от неопределенности	(n, m) – уровень неопределенности от 1 до 4; уровень влияния на развитие Арктики от 1 до 4)
Скорость изменения климата	Физическая доступность ресурсов и доступность использования СМП Разрушение инфраструктуры Таяние вечной мерзлоты	(4,4) (2,4) (2,4)
Качество институциональной среды	Экологическое регулирование Финансовое стимулирование и инвестиции Демографическая и социальная политика Эффективность реагирования на катастрофы (экологические/пандемии и т.п.) Международное сотрудничество Возможности для развития бизнеса Динамика населения	(3,4) (3,4) (3,3) (4,4) (3,4) (4,3) (1,2)
Динамика geopolитики	Скорость глобализации Международные соглашения по Арктике	(1,1) (2,2)
Спрос на ресурсы Арктики	Цены на нефть Иновации в добывающих технологиях Развитие альтернативной энергетики Стоимость добычи ресурсов Арктики Экологические стандарты добычи нефти и газа	(4,4) (3,4) (2,3) (4,4) (2,3)

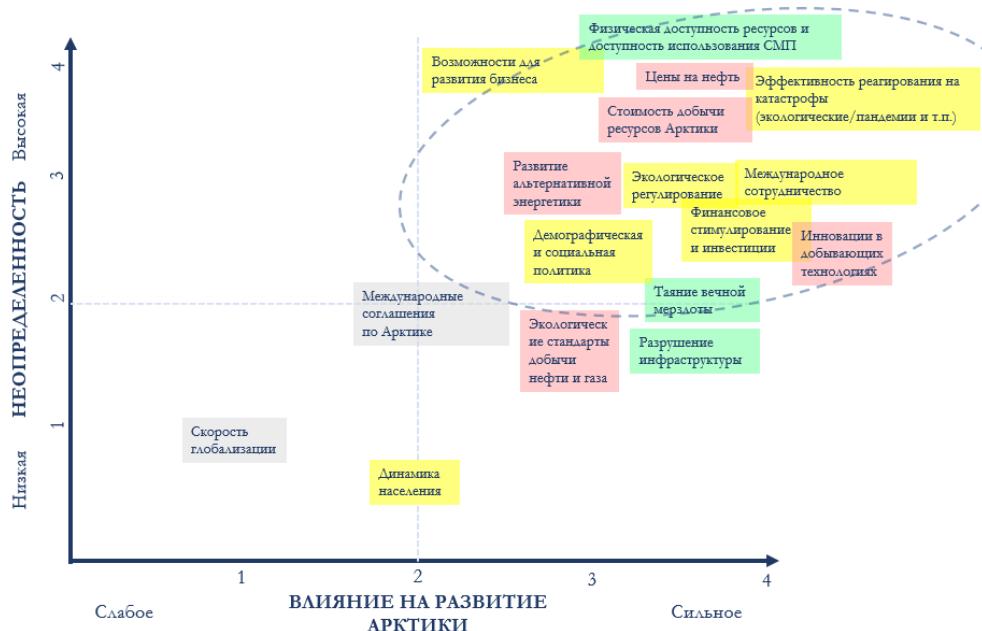


Рисунок 4.1 – Выявление ключевых неопределенностей. Построено самостоятельно на основе [144]. Желтый – качество институциональной среды, красный – спрос на ресурсы Арктики, зеленый – скорость изменения климата, серый – динамика geopolитики.

Таким образом, изобразив все факторы на графике (рисунок 4.1) с учетом их уровня неопределенности и уровня влияния на развитие Арктики, видим, что ключевые неопределенности – это качество институциональной среды и спрос на ресурсы Арктики. Основные неопределенности выявлены.

#### Этап 4.

Построим сценарии, принимая за основу две выявленные ключевые неопределенности.

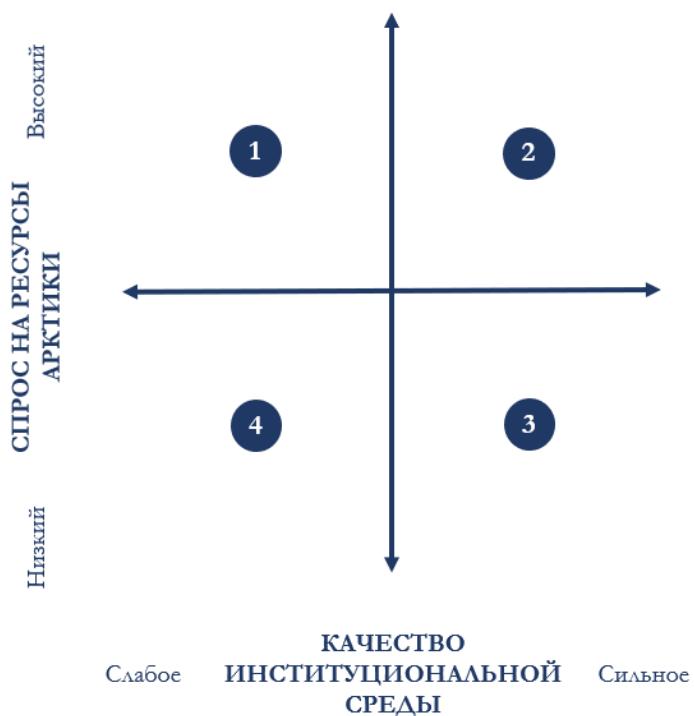


Рисунок 4.2 – Матрица сценариев в виде SWOT диаграммы. Построен на основе [144].

Далее были детализированы 4 сценария развития Арктики к 2036:

- высокий спрос на ресурсы Арктики и слабое качество институциональной среды;
- высокий спрос на ресурсы Арктики и сильное качество институциональной среды;
- низкий спрос на ресурсы Арктики и сильное качество институциональной среды;

– низкий спрос на ресурсы Арктики и слабое качество институциональной среды.

#### Этап 5.

На данном этапе были созданы стратегии и планы действий для наиболее желаемых сценариев развития Арктики к 2036 году. Сценарии построили на основе концепции тройного критерия (Triple bottom line) и тактике 3Р (People, Profit, Planet. Люди, прибыль, планета), которая использовалась при создании проекта ArcticRewild. Каждый из сценариев как раз был построен на трех составляющих: экономика и geopolитика (прибыль), экосистема Арктики (планета) и общество (люди). Представим сценарии на графике 3Р (рисунок 4.3) для наглядности.

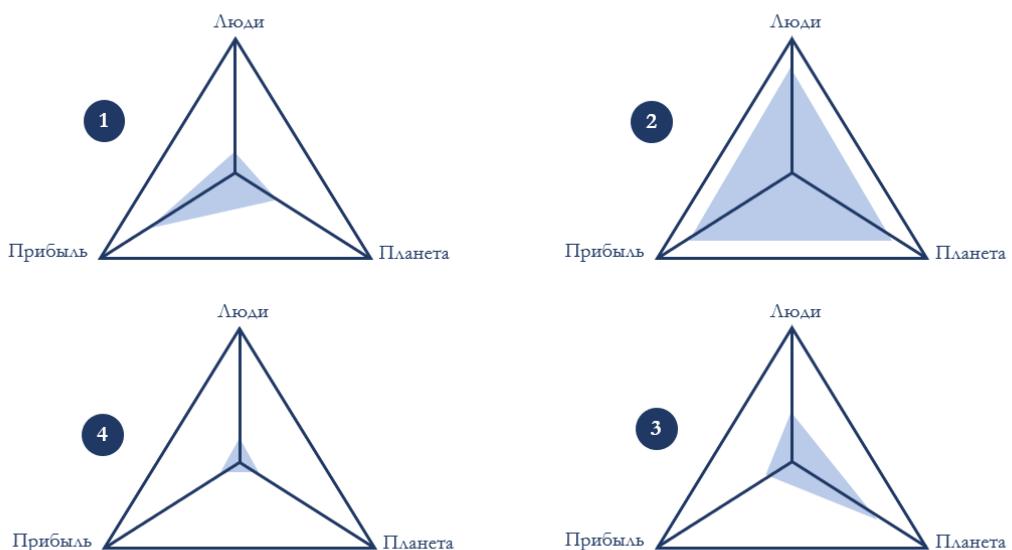


Рисунок 4.3 – Четыре сценария развития Арктики на графике 3Р (прибыль, люди, планета)

Наиболее предпочтительным и благоприятным сценарием развития Арктики является сценарий номер 2, когда существует высокий спрос на ресурсы Арктики и сильное качество институциональной среды. В этом случае происходит равномерное развитие во всех трех измерениях, что и является устойчивым развитием.

Способность мыслить в долгосрочной перспективе и поддерживать баланс всех трех измерений – это устойчивое мышление, которое необходимо

для лидеров и инвесторов Арктики. При балансе данных трех составляющих в будущем существуют большие шансы того, что сегодняшняя Арктика приблизится к 2036 году к ситуации, похожей на второй сценарий, где соблюдаются баланс между добывающей арктической деятельностью с использованием новейших технологий, социальной политикой и защитой окружающей среды. Это позволит сделать арктический бизнес устойчивым, выйти на новый уровень производительности, создать благоприятные условия для жителей севера, развить международное сотрудничество и сохранить уникальную среду Арктики.

#### 4.2.3 Некоторые выводы

В ходе данной работы был проведен анализ перспектив развития Арктики. Проведен анализ рисков, влияющих на данный процесс, показан пример противодействия одному из таких рисков через устойчивый проект ArcticRewild, который позволит уменьшить скорость таяния вечной мерзлоты. Для этого проекту также была построена финансовая модель, показывающая финансовую осуществимость и масштабируемость идеи.

В рамках следующей практической части работы были смоделированы 4 сценария развития Арктики и с помощью применения концепции тройного критерия сформирована оптимальная стратегия для развития Арктики на ближайшие 15 лет.

Ключевые рекомендации по развитию Арктики:

- 1) привлечение государственных и частных инвестиций в развитие инфраструктуры Арктики;
- 2) государственная политика, направленная на поддержку коренных народов;
- 3) развитие инноваций и применение экологических методов добычи и переработки полезных ископаемых;
- 4) строительство долгосрочных социально-экономических планов по развитию региона на основе мониторинга экологических и демографических тенденций;

5) развитие безуглеродных источников дохода в Арктике, включая экотуризм.

### **4.3 Исследование причин вододефицита в странах региона Северной Африки**

#### **4.3.1 Введение**

Проблема водного кризиса признана одной из серьезнейших угроз современного мира согласно докладу о глобальных рисках 2020 года, в котором проблема водного [145]. К сожалению, ситуация с водными ресурсами в мире только продолжит ухудшаться. По оценкам уже к концу 2040 года разрыв между водопотреблением и водообеспеченностью составит 40% [146]. Кроме того, около 700 миллионов человек будут вынуждены покинуть свои страны к 2030 году ввиду обострения проблемы вододефицита [147].

Рост населения и глобальное потепление только усугубляют проблему доступа людей к водным ресурсам. Довольно очевидно, что общий объем имеющихся водных запасов определяет устойчивое политическое и социально-экономическое развитие страны. Тем не менее, ситуация с водными ресурсами обостряется ввиду того, что само распределение воды в мире достаточно неравномерно. Наиболее критическую картину можно наблюдать в Северной Африке, которая является некой глобальной горячей точкой [148] в контексте её уязвимости и подверженности к изменению климата. Водообеспеченность же на душу населения упала примерно на 30 процентов за последние 20 лет [149], едва достигая уровня в 1 000 кубических метров на душу населения.

Совместный контроль над водными ресурсами создаёт условия сильной зависимости государств нижнего течения от любого решения, принимаемого странами верховья. Отсутствие рационального управления водными ресурсами может привести к усилению водного кризиса и эскалации конфликта между странами. Таким образом, доступ к водным ресурсам превратился в основной инструмент политических спекуляций и причину разногласий между государствами.

Целью данной работы является оценка уровня обеспеченности водой в странах региона Северной Африки. Кроме того, исследуются основные причины дефицита воды и предпринимается попытка предложить подходы, которые бы устранили негативные последствия нехватки водных ресурсов.

Гипотезой данной работы является допущение о том, что уровень водообеспеченности является основным фактором, определяющим торговую ориентацию и направленность стран. Иными словами, чем большим объемом водных ресурсов страна располагает, тем ниже уровень импорта зерновых культур ввиду того, что страна обладает возможностями самостоятельно производить сельскохозяйственную продукцию внутри собственных локальных рынков. Для того, чтобы проверить данную гипотезу, были проведены регрессионный и корреляционный анализ.

Данная работа основана на анализе и обобщении существующей литературы и официальных данных о гидрографических условиях стран Северной Африки с целью анализа водообеспеченности и причин нехватки водных ресурсов в этом регионе. Для того, чтобы проверить выдвинутую гипотезу были проведены регрессионный и корреляционный анализ. Необходимо отметить, что разработка линейной регрессии довольно ограничена из-за отсутствия актуальных данных о водных ресурсах. Наблюдение за динамикой изменения уровня водообеспеченности также не представляется возможным ввиду непредставления необходимой информации официальными организациями, вовлеченными в процесс мониторинга наличия водных ресурсов в странах Северной Африки. Следовательно, результаты проведенного анализа могут быть нерепрезентативными из-за небольшого размера выборки.

#### 4.3.2 Оценка водообеспеченности

Для определения водообеспеченности стран обычно используется индикатор водного стресса, который был разработан Малин Фалькенмарк. Показатель основывается на объеме возобновляемых водных ресурсов доступного для населения в конкретном регионе. Согласно данному индикатору, страны могут быть поделены на 4 категории [150]:

- страны, не испытывающие водный дефицит (когда объем водообеспеченности превышает 1700 кубических метров на душу населения);
- страны, подверженные водному стрессу, количество водных ресурсов находится между 1 000 и 1 700 кубических метров;
- страны, испытывающие дефицит водных ресурсов, когда объем водных ресурсов находится между 500 и 1 000 кубических метров;
- страны, страдающие от хронического недостатка водных ресурсов, когда уровень водообеспеченности принимает значения ниже 500 кубических метров.

Таблица 4.3 – Классификация стран в соответствии с индикатором водного стресса

Water scarcity (500-1000 m <sup>3</sup> /year/capita)	
Country	Water endowment per capita
Egypt	589.4
Morocco	811.4
Sudan	932.6
Absolute water scarcity (<500 m <sup>3</sup> /year/capita)	
Country	Water endowment per capita
Algeria	282.4
Libya	109.8
Tunisia	400.2

Как видно из Таблицы 4.3 Алжир, Тунис и Ливия испытывают хронический дефицит водных ресурсов в то время, как в остальных странах региона ситуация с водными ресурсами немного лучше.

Кроме того, мы решили ввести собственный индекс водообеспеченности (WEI) для того, чтобы оценить уровень реального водоснабжения. Данный индекс показывает разницу между установленным уровнем водообеспеченности и фактическим количеством водных ресурсов, доступных для потребления.

$$WEI = \frac{S - P \cdot R}{P \cdot R} = \frac{S}{P \cdot R} - 1,$$

где S – фактический объем доступных водных ресурсов на душу населения (АКВАСТАТ); Р – общая численность населения; R – минимальная рекомендованная норма водопотребления (должна составлять 50 литров в день

на душу населения для того, чтобы удовлетворить простые базовые нужды людей).

Распределение значений:

- если значение WEI находится в интервале  $[-1;0)$ , то страна испытывает недостаток водных ресурсов;
- если значение WEI находится в интервале  $(0;+\infty)$ , население страны хорошо обеспечено водными ресурсами;
- если значение WEI близко к нулю, то можно предположить, что водо-снабжение страны находится в пределах жизнеспособности человека.

Как видно из Таблицы 4.4 уровень водоснабжения чрезвычайно низкий, доступный объем водных ресурсов не обеспечивает все население выделенных стран минимальным количеством воды, которое могло бы удовлетворить базовые потребности человека.

Таблица 4.4 – Индекс вообеспеченности

Country	The deviation of water supply
Algeria	-0.9999996
Egypt	-0.9999997
Libya	-0.9999991
Morocco	-0.9999999
Sudan	-0.9999997
Tunisia	-0.9999981

Основным ограничением этого показателя является отсутствие доступной и актуальной информации о фактическом количестве водных ресурсов. АКВАСТАТ позволяет получать данные только до 2017 года.

Теоретическая часть работы посвящена факторам, провоцирующим нехватку воды. Среди наиболее распространённых причин водных дефицитов климатические условия, устаревшая водная инфраструктура и быстрый процесс урбанизации. Данные причины играют основополагающую роль в замедлении процесса расширения доступа населения к воде. Кроме того, данные факторы могут ухудшать социально-экономическую и политическую стабильность стран, создавая возможность для возникновения конфликтов как внутри одной страны, так и между несколькими государствами.

Для того, чтобы предотвратить развитие разногласий, возникающих на почве нехватки водных ресурсов, необходимо разрабатывать политику, которая была бы направлена на сокращение или устранение разрыва между спросом на водные ресурсы и их предложением. В данной работе представлены два подхода, которые могли бы стать довольно эффективными инструментами по решению водных кризисов – концепт торговли виртуальной водой и использование морской воды.

#### 4.3.3 Торговля виртуальной водой

Применение концепта торговли виртуальной водой, как способа снижения уровня водного дефицита, в научной литературе остается довольно спорным подходом. Тем не менее, мы полагаем, что имплементация данного подхода может стать полезным инструментом не только по обеспечению более устойчивого объема водных ресурсов, но и по отслеживанию зависимости между доступными водными ресурсами и торговой ориентацией. Под термином виртуальная вода понимается вода, которая используется в процессе производства импортируемой продукции [151].

Данный подход достаточно полезный инструмент по снижению уровня водного стресса. Внедрение данного концепта позволит странам Северной Африки, который страдают от дефицита водных ресурсов, перейти от производства сельскохозяйственной продукции, которая требует довольно устойчивого объема водных ресурсов, внутри собственных рынков к импортированию таких водоемких товаров из регионов, обеспеченных водой. Импортеры виртуальной воды могут выиграть от подобной торговой специализации благодаря возможности перенаправить ресурсы, полученные от экономии воды, на другие нужды [152]. Кроме того, виртуальная торговля водой выгодна с точки зрения ее экономической целесообразности [153] в условиях ограниченных бюджетов североафриканских стран.

Виртуальная торговля водой определенно эффективный механизм по минимизации вероятности возникновения конфронтаций между государствами, поскольку приводит к некоему балансу неравномерно распределён-

ные водные ресурсы, тем самым снижая конкуренцию за данные ресурсы. Тем не менее, эта политическая невидимость, о которой говорил Джон Аллан [154], может привести к исключению проблемы водных дефицитов из политической повестки и к замедлению водных реформ ввиду снижения значимости водного баланса на локальном уровне. Лица, отвечающие за принятие решений и разработку политики, уверенные в том, что есть некая подушка безопасности, которая образуется благодаря сбережениям от виртуальной торговли воды, могут испытывать соблазн не разрабатывать и реализовать долгосрочную политику в области водных ресурсов, поскольку виртуальная вода уже удовлетворяет краткосрочные потребности [155]. И хотя концепт виртуальной воды снижает зависимость от условий, которые возникают благодаря ограниченным водным ресурсам, данный подход провоцирует дополнительную зависимость от колебаний глобальной экономической среды за счет интеграции локальных рынков в глобальные торговые сети [155].

На вектор сельскохозяйственной торговли влияет не только уровень водообеспеченности, но и степень обеспеченности земельными, трудовыми ресурсами и капиталом, наличие или отсутствие различных субсидий и налогов. Таким образом, потоки международной торговли нельзя полностью соотнести с уровнем обеспеченности водными ресурсами [156]. Тем не менее, концепция виртуальной торговли водой может быть полезной с точки зрения ее использования как способа анализа, позволяющего отследить общие тенденции и разработать политику для предотвращения эскалации водных конфликтов.

#### 4.3.4 Модель линейной регрессии

Данная работа сконцентрирована на выявлении зависимости между торговлей виртуальной водой и дефицитом водных ресурсов. Для того чтобы объяснить, почему объем зернового импорта на душу населения используется в качестве индикатора торговли виртуальной водой, стоит отметить, что страны Северной Африки являются одними из ключевых импортеров вирту-

альной воды, в особенности в области сельского хозяйства, находясь в сильной зависимости от зернового импорта [157].

Предлагаемая модель рассматривает объем импорта зерновых на душу населения как зависимую переменную, ВВП на душу населения и водоснабжение на душу населения относятся к независимым переменным. Кроме того, мы ввели в качестве независимой переменной долю сельского населения, которая будет отражать предложение трудовых ресурсов.

$$Y_c = \alpha + \beta_1 g_c + \beta_2 e_c + \beta_3 r_c,$$

где  $Y_c$  – объем импорта зерновых на душу населения,  $g_c$  – ВВП на душу населения,  $e_c$  – общий объем возобновляемых водных ресурсов на душу населения,  $r_c$  – процент сельского населения.

Данные из Таблицы 4.5 были использованы для проведения регрессионного анализа.

Таблица 4.5 – Статистические данные

Country	Population in 2017	Cereal import per capita in 2017, kg/inhab	Rural population in 2017	Percentage of rural population, %	GDP in 2017, per capita, US\$/inhab
Algeria	41 389 000	311.705	11 548 000	28	4 051
Egypt	96 443 000	192.158	55 893 000	58	2 000
Libya	6 581 000	423.105	1 287 000	20	3 941
Morocco	35 581 000	181.108	13 614 000	38	3 070
Sudan	40 813 000	55.716	26 602 000	65	3 060
Tunisia	11 433 000	323.291	3 616 000	32	3 452

#### 4.3.5 Результаты регрессионного анализа

Из Таблицы 4.7 мы получаем следующее уравнение линейной регрессии:

$$Y_c = 589.785 - 0.016 g_c - 0,28 e_c - 345,78 r_c.$$

Р квадрат примерно равен 0.99 (Таблица 4.6). Это означает, что условная дисперсия модели достаточно мала и весьма вероятно, что модель неплохо описывает данные. Тем не менее, мы получили высокие р-значения по удельному весу сельского населения и ВВП, что что может быть объяснено тем фактом, что наши данные неубедительны, чтобы предположить, что эф-

фект существует. Однако, можно заметить, что с увеличением на одну единицу общего объема водообеспеченности, уровень зернового импорта снижается на 0.28 единиц.

Таблица 4.6 – Статистики регрессии

Multiple R	R <sup>2</sup>	Adjusted R <sup>2</sup>	Standard Error	Observations
0,999157	0.998	0.993	11.611	5

Таблица 4.7 – Результаты регрессионного анализа: коэффициента

	Coefficients	Standard Error	t Stat	p-value	Lower 95%	Upper 95%
Intercept	589.785	57.794	10.205	0.062	-144.555	1324.12
Water supply, per capita	-0.28	0.03	-9.274	0.068	-0.663	0.104
Percentage of rural population	-345.78	69.949	-4.943	0.127	-1234.57	543.011
GDP, per capita	-0.016	0.013	-1.212	0.439	-0.179	0.147

Таким образом, ни процент сельского населения, ни уровень ВВП не влияют на объем зернового импорта. Обратная зависимость между водообеспеченностью и вектором торговли подтверждает наше предположение о наличии зависимости между водным дефицитом и увеличением объемов торговли виртуальной воды. Северная Африка, наряду с Ближним Востоком, остаются единственными регионами, где наличие водных ресурсов оказывает довольно значительное влияние на торговую специализацию.

Таблица 4.8 – Матрица корреляции

	Cereal import, per capita	Water supply, per capita	Percentage of rural population	GDP, per capita
Cereal import, per capita	1			
Water supply, per capita	-0.965	1		
Percentage of rural population	-0.916	0.824	1	
GDP, per capita	0.628	-0.607	-0.787	1

Кроме того, следует пояснить взаимосвязь между объемом импорта зерновых и соотношением ВВП. Результаты корреляционного анализа (Таблица 4.8) показывают, что спрос на импорт зерновых культур соответствует общему уровню экономической активности в стране.

#### 4.3.6 Использование охлажденной воды

Другим инструментом по минимизации негативных последствий водного дефицита, является использование альтернативных источников воды, в частности использование соленой воды. Несмотря на то, что водозабор из этих ресурсов оценивается менее чем в 5 процентов в Египте, Марокко и Тунисе, существует значительный потенциал для использования нетрадиционных ресурсов [158]. Основным препятствием на пути использования альтернативных водных ресурсов является относительно высокая стоимость их очистки. Затраты на охлаждение покрываются лишь в прибрежной зоне курортов Красного моря, где прибыль от использования данного ресурса превышает затраты [159].

Стоит учесть опыт одного из пионеров в области внедрения технологий охлаждения. Мы рассмотрели пример Ливии, которая запустила несколько охлаждающих заводов с 1960-х годов [160]. Этот вид нетрадиционных водных ресурсов используется в основном для промышленных и муниципальных нужд. Водозабор составляет 58 и 29 процентов от общей установленной мощности, соответственно. Широкое использование охлажденных водных ресурсов можно объяснить тем фактом, что Ливия страдает от остройшего дефицита воды. Более того, традиционные водные ресурсы ограничены из-за загрязнения и ухудшения состояния водотоков [161].

Внедрение процесса охлаждения в Северной Африке кажется разумным решением по следующим причинам:

- подавляющее большинство стран являются прибрежными государствами, имеющими непосредственно прямой доступ к возобновляемым водным ресурсам;
- в регион Северной Африки входят страны, которые имеют достаточно большой опыт в области нетрадиционного водопользования.

Охлаждение воды не является ключевым решением водного кризиса, тем не менее, это полезный инструмент, позволяющий снизить уровень вод-

ного стресса за счет предоставления дополнительных альтернативных водных ресурсов.

#### 4.3.7 Некоторые выводы

Проблема водных ресурсов настолько же остра, насколько и сложна. В соответствии с целью данной работы, были проанализированы страны Северной Африки, что помогло понять истинные причины нехватки воды. Основными причинами нехватки воды являются глобальное потепление или изменение климата, рост населения и как следствие увеличение его давления на водные ресурсы, что приводит к чрезмерной эксплуатации водоносных горизонтов и к их загрязнению. Кроме того, мы ввели индекс водообеспеченности, который показал значительное отклонение между фактическим объемом водоснабжения и установленной нормой обеспеченности водными ресурсами.

Мы рассмотрели несколько подходов, которые могли бы смягчить последствия потенциальных конфликтов, связанных с нехваткой воды. Эти концепции были выбраны в соответствии со спецификой региона. Однако реализации только этих подходов было бы недостаточно. Концепция опреснения воды и виртуальной воды должна быть реализованы и внедрены параллельно вместе с политикой государственной поддержки водных проектов.

Более того, данная работа продолжила рассмотрение довольно спорной гипотезы о зависимости торговой специализации и степени обеспеченности водой. Разработанная модель линейной регрессии также включает две дополнительные независимые переменные – ВВП на душу населения и долю сельского населения. Тем не менее, большие р-значения, связанные с этими переменными, отражают незначительность их влияния на объем импорта. Высокие р-значения могут свидетельствовать о довольно небольшой выборке, что является результатом отсутствия данных.

Наше предположение о взаимосвязи между нехваткой воды и специфической торговли подтвердилось: если страна будет испытывать дефицит воды, уровень импорта зерновых превысит объем экспортirуемой сельскохозяй-

ственной продукции. Однако применение подхода виртуальной торговли водой в качестве решения водного кризиса довольно нецелесообразно. Эта концепция может быть полезным инструментом, позволяющим отслеживать изменения в водообеспеченности в странах.

#### **4.4 Интервальный метод оболочечного анализа данных и его применение для анализа эффективности карантинных мер**

##### **4.4.1 Введение**

COVID-19 – глобальная пандемия коронавирусной инфекции COVID-19, вызванной коронавирусом SARS-CoV-2. Для сдерживания распространения неизученного вириуса были приняты беспрецедентные меры: страны закрыли границы, прекратили воздушное и морское пассажирское сообщение и были вынуждены ввести локдаун – закрыть магазины, бары, рестораны, музеи, театры, места отдыха, остановить общественный транспорт, перевести сотрудников на удаленную работу и значительно ограничить передвижение в городах.

Однако разные стратегии введения ограничений приводят к разным последствиям. Кроме того, некоторые местные факторы могут влиять на уровень инфицирования коронавирусом.

Например, хорошо известно, что карантин в Швеции был гораздо более мягким, чем в других странах. У этого подхода есть свои преимущества и недостатки. Экономические потери, вызванные закрытием бизнеса в стране, были очень низкими. Но, с другой стороны, в течение длительного периода 2020 года в Швеции был самый высокий уровень смертности в Европе.

В свою очередь, Китай активно отреагировал на угрозу и оперативно ввел очень строгие карантинные меры. Суровое наказание, вплоть до смертной казни, грозило тем, кто был осужден за изготовление и распространение поддельных лекарств, злонамеренное заражение коронавирусом других людей и причинение серьезного вреда врачам во время вспышки коронавируса. Это дало возможность остановить дальнейшее распространение коронавируса.

са. Однако это привело к серьезным экономическим потерям – ВВП Китая сократился на 6,8 процента в первом квартале 2020 года [162].

Именно из-за этой разницы в подходах важно уметь сравнивать и оценивать эффективность тех или иных карантинных мер. Более того, поскольку пандемия до сих пор продолжается, методология сравнения позволит найти оптимальные стратегии поведения и дать рекомендации для стран с наихудшей обстановкой.

#### 4.4.2 Основная часть

Классический вариант оболочечного анализа данных заключается в решении многоцелевой оптимизационной задачи (минимизацию затрат и максимизацию достигнутых результатов), предложенной Чарнсом, Купером и Роудсом в 1978 году [163]. Для этого для каждого объекта в выборке вычисляется отношение взвешенных сумм достигнутых результатов и затраченных ресурсов. А также для интерпретируемости результатов добавляется ограничение, гарантирующее, что нахождение полученных оценок эффективности между 0 и 1 всех сравниваемых объектов.

В результате получается следующая оптимизационная задача:

$$\max_{u_i, v_j} \frac{\sum_{i=1}^M u_i y_{ik}}{\sum_{j=1}^N v_j x_{jk}}.$$

С учётом ограничений  $\forall i, j, k$

$$\begin{cases} \frac{\sum_{i=1}^M u_i y_{ik}}{\sum_{j=1}^N v_j x_{jk}} \leq 1, \\ u_i \geq 0, \\ v_j \geq 0. \end{cases}$$

где,  $y_{ik}$  и  $x_{jk}$  – это результаты и ресурсы  $k$ -го объекта, в то время как  $u_i$  и  $v_j$  – весовые коэффициенты, отражающие относительную важность критериев.

В рамках исследования было предложено модифицировать интервальный метод оболочечного анализа данных с оптимальной трубкой [164]. Изначально в методе максимальную оценку эффективности в 100% получали

только объекты, находящиеся на границы оптимальной эффективности и нesравнимые с ними.

Однако, часто бывают ситуации, когда объект находится чуть ниже границы абсолютной эффективности, но далеко от объектов, получивших 100% оценку эффективности согласно классическому оболочечному анализу (объект F на рисунке 4.4). И тот факт, что в подобной ситуации объект получит оценку эффективности ниже 100%, не отображает всю полноту картины. Поэтому был предложен новый алгоритм определения объектов, которые получат наивысший балл. Для этого после построения границы оптимальной эффективности рассматривается нахождение оптимальной версии каждого объекта с учётом интервальных значений критериев (минимальные значения в интервалах ресурсов и максимальные – в интервалах достигнутых результатов). Таким образом, объекты A, A' и F получат 100% эффективность. Остальные же объекты оцениваются с помощью классического оболочечного анализа данных.

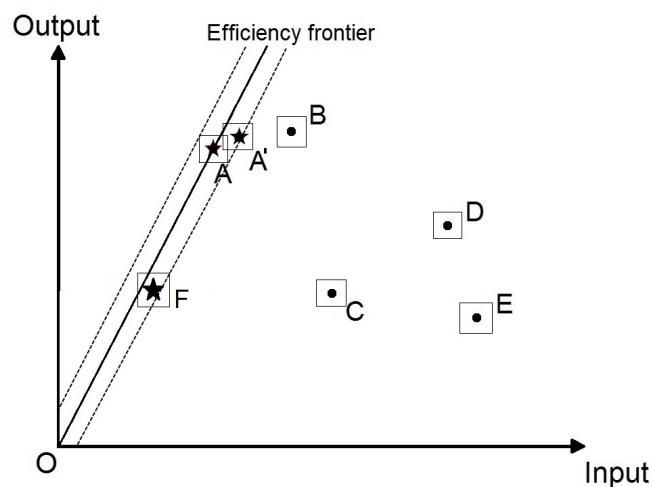


Рисунок 4.4 – Пример применения полной версии интервального оболочечного анализа с оптимальной трубкой

Для проверки данного интервального метода оболочечного анализа данных мы оценили эффективность противокоронавирусных карантинных мер в различных странах мира.

Мы взяли данные по строгости введения ограничительных мер по нескольким основным направлениям: закрытие предприятий, закрытие учебных заведений, отмена публичных мероприятий, ограничения на международные поездки, ограничения на внутренние поездки, ограничения на выход из дома, остановка общественного транспорта, ограничения на многочисленные собрания граждан [165].

Кроме того, помимо самих ограничительных мер не менее важно то, насколько строго эти указания выполняются. И для этого мы воспользовались мнением экспертов, которые оценили законопослушность крупнейших стран мира, которые будут исследованы.

В качестве выходных параметров было решено использовать количество новых инфицированных пациентов. Однако было бы неверно сравнивать 100 новых инфицированных пациентов для стран с населением 3 миллиона и 300 миллионов человек. Поэтому мы используем долю новых случаев заболевания по отношению к населению страны, а не просто количество новых случаев. Также, увеличение числа инфицированных со 100 до 200 и с 10 100 до 10 200 не равнозначны с точки зрения эффективности карантинных мер. Следовательно, мы включаем соотношение новых случаев заболевания к общему числу инфицированных людей в качестве еще одного выходного параметра.

Кроме того, важно подчеркнуть, что влияние карантинных мер на число новых случаев заболевания не является немедленным. Поэтому для оценки эффективности мы использовали карантинные мероприятия на определенную дату и количество новых случаев заболевания спустя две недели.

В результате оценки эффективности в течение всего периода пандемии каждые две недели мы получили временные ряды для всех исследуемых стран, что позволяет нам анализировать основные тенденции в терминах эффективности карантинных мер. Используя эти данные, мы смогли разделить все страны на несколько групп.

Первая группа состоит из стран, которые были далеки от основных центров распространения COVID-19: Ангола, Непал, Сирия. Первые инфицированные в этих странах появились только весной 2020-го. Однако после этого карантинные мероприятия не были хорошо организованы. В результате эти страны имели высокую эффективность в соответствии в начале весны (март, апрель) и низкую эффективность в остальное время.

Вторая группа, в свою очередь, состоит всего из двух стран – Вьетнама и Китая. Китай стал отправной точкой для COVID-19 [166]. В стране были серьёзные проблемы с количеством инфицированных, однако правительство приняло строгие карантинные меры. В результате распространение коронавируса было достаточно быстро остановлено. И до сих пор эффективность ограничительных мероприятий остается высокой (рисунок 4.5). Таким образом, именно Китай и Вьетнам могут быть использованы в качестве ориентира для всех многих стран.

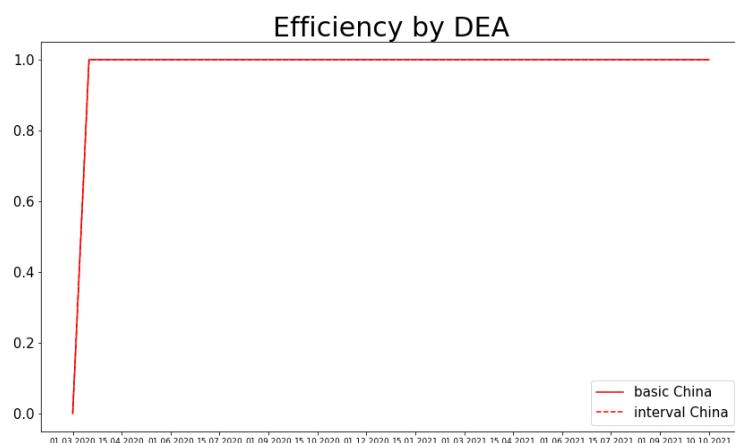


Рисунок 4.5 – Эффективность карантинных мер в Китае

У всех остальных стран ситуация с эффективность карантинных мер схожа – у них возникало несколько волн пандемии, и в каждую из них эффективность мер падает, возрастая по окончании очередной волны пандемии (рисунок 4.6).

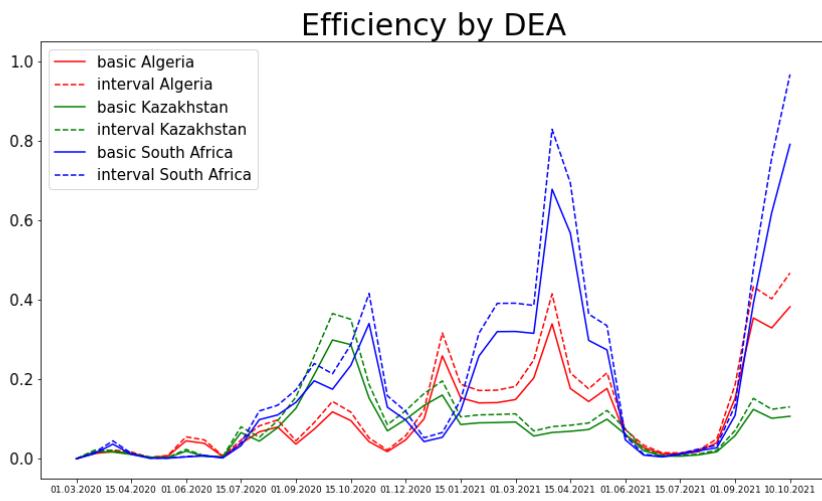


Рисунок 4.6 – Эффективность карантинных мер в странах с несколькими волнами пандемии

#### 4.4.3 Некоторые выводы

По итогам работы можно сказать, что поставленные задачи выполнены в полном объёме. Во-первых, были модифицированы существующие интервальные методы оболочечного анализа данных.

Во-вторых, с помощью предложенной улучшенной версии метода была оценена эффективность карантинных мер в разных странах мира. В свою очередь, это позволило выделить группы стран, со схожими стратегиями внедрения ограничений. Результаты данного исследования были представлены на нескольких международных конференциях и научных семинарах.

### 4.5 Анализ предпочтений студентов

#### 4.5.1 Введение

Современные технологии все больше меняют нашу жизнь, в том числе и сферу образования. Школы и университеты не только внедряют современные технологические решения в поддержку классического стиля обучения и преподавания, но и трансформируют саму модель обучения, переходя от многовековой традиционной модели живого общения «учитель-ученик» к более обезличенной модели онлайн-обучения. Существуют как гибридные форматы, когда лишь часть лекций и занятий по дисциплине заменяется ра-

ботой в дистанционном режиме, так и полноформатные онлайн-курсы на специализированных платформах, наиболее известными из которых являются Coursera, edX, Национальная платформа открытого образования и др. Массовый открытый онлайн-курс (МООК) стандартного формата включает в себя определенный набор видеолекций, в которых объясняется весь материал, и задания для проверки качества усвоения прослушанной темы (обычно в тестовом формате). Кроме того, есть форумы, где можно задать вопрос преподавателю для лучшего понимания темы или обсудить что-то с другими слушателями.

Очевидны преимущества МООК: 1) отсутствие привязки ко времени и месту проведения занятий, что позволяет студентам совмещать их с постоянной работой; 2) возможность прослушать курсы ведущих специалистов из различных университетов (в том числе зарубежных), что позволяет им учиться, не затрачивая средства на переезд для обучения; 3) онлайн-курсы не предполагают ограничений по количеству мест, что открывает дополнительные возможности для студентов в изучении любого предмета. Особенно актуально стало использование дистанционных форм обучения с пандемией COVID-19, когда по всему миру в целях предотвращения распространения заболевания школьники и студенты находились и обучались дома. С каждым годом растет число МООК и все больше элементы онлайн-обучения внедряются в школьное и университетское образование.

При этом далеко не всем обучающимся нравится подобная замена формата обучения ввиду серьезных минусов онлайн-образования. Одним из главных негативных факторов исследователи называют отсутствие прямого контакта и общения не только в рамках взаимодействия ученика с преподавателем, но и общения внутри группы. Онлайн-студентам сложнее задавать по ходу лекции интересующие их вопросы, уточнять информацию, организовывать коллективное обсуждение – хотя эти формы активности пытаются компенсировать чатами и форумами, студенты отмечают недостаточность таких форматов [167]. Кроме того, преподавателям крайне сложно полно-

стью контролировать самостоятельность выполнения оценочных заданий слушателем, как следствие полученные сертификаты об окончании курсов не будут достоверно отражать знания студента, получившего его.

Параллельно внедрению элементов онлайн-обучения в традиционную модель обучения в университетах, когда онлайн-курсы целиком или частично встраиваются в учебные планы студентов, также развивается направление так называемых *Micromaster's degree*, которые представляют собой цикл/серию курсов по одной или смежным тематикам, результаты которого эквивалентны 25-50% кредитам в учебном плане и часто являются частью образовательной программы университетов. И, конечно, онлайн-курсы выбираются слушателями для повышения своей квалификации, изучения нового материала и просто расширения кругозора. Поэтому была поставлена задача всестороннего изучения предпочтений разных групп слушателей онлайн-курсов.

#### 4.5.2 Основная часть

Для проведения исследования были взяты данные двух опросов за 2018 год, проводимых НИУ ВШЭ: внутреннего опроса студентов (МСЖ, Мониторинг студенческой жизни) в НИУ ВШЭ, а также опроса слушателей курсов, предлагаемых НИУ ВШЭ на платформе Coursera не только студентам, но и всем желающим. Второй опрос является пре-опросом, то есть он проводился до начала курса; пост-опросы не проводились из-за малого отклика респондентов. В рамках МСЖ было опрошено 6631 студента НИУ ВШЭ, в рамках пре-опроса на Coursera – 6456 слушателей, однако очень многие респонденты не ответили на все вопросы анкеты и для анализа были оставлены только полные анкеты 4338 респондентов.

Поскольку опросы проводились на разных категориях респондентов и с разными целями, то вопросы анкет во многом не совпадают и поэтому для построения регрессий будут на разных базах данных будут задействованы разные регрессоры. Анкета опроса МСЖ является избыточной для целей нашего исследования, поэтому были отобраны часть вопросов, непосред-

ственno относящихся к онлайн-курсам и предпочтениям относительно формата обучения, а также социально-демографические характеристики и некоторые другие данные: о курсе обучения, наличии работы, материальном положении и т.п. В опросах слушателей онлайн-курсов НИУ ВШЭ содержится меньше вопросов, большая часть из которых была использована в анализе; удалены лишь технические вопросы (например, аксесс-код респондентов, время прохождения опроса и т. п.) и специфические вопросы, на которые было мало ответов (например, вопросы на английском языке к англоязычным онлайн-курсам).

Прежде всего, обратимся к анкете МСЖ, которую заполняли студенты НИУ ВШЭ. В 2018 году анкета содержала большой блок вопросов, связанных с внедрением элементов онлайн-обучения в учебные планы студентов и отдельный вопрос о поддержке идеи перевода всех лекций в онлайн формат, которую до этого ранее публично озвучил ректор НИУ ВШЭ Я. И. Кузьминов в одном из интервью. Вопрос задавался следующим образом: «Как Вы оцениваете идею перевести все лекционные занятия в Вышке в онлайн формат?» Вариантами ответа были: 1) полностью поддерживаю, 2) скорее поддерживаю, 3) скорее не поддерживаю, 4) точно не поддерживаю, 5) затрудняюсь ответить. Вариант 1 выбрали 1151 студентов (17,4%), вариант 2 – 1579 студента (23,8%), вариант 3 – 1483 (22,4%), вариант 4 - 2101 (31,7%), затруднились ответить 317 студентов (4,8%). Объединим варианты ответов 1 и 2, а также 3 и 4, и отнесем неопределившихся студентов к не поддерживающим переход на онлайн-формат, таким образом, 41,2% студентов предложили онлайн-формат обучения, а 58,8% студентов не высказали поддержку онлайн-формату обучения.

В анкете слушателей онлайн-курсов НИУ ВШЭ на платформе Coursera не было аналогичного вопроса, поэтому выберем вопрос, наиболее подходящий для анализа предпочтений респондентов, а именно, вопрос «Оцените, пожалуйста, Ваш опыт обучения на массовых открытых онлайн-курсах по шкале от 1 до 5, где ‘1’ – опыт обучения был крайне неудачным, ‘5’ – опыт

обучения был очень удачным». Аналогичным образом, объединим варианты ответов, считая ответы 4 и 5 позитивным отношением к формату онлайн-курсов, а остальные ответы – негативным. Если отсутствие ответа на этот вопрос также будем трактовать как негативное отношение, то из 4338 респондентов положительно (то есть выбрав варианты 4 и 5) на этот вопрос ответили 1074 респондента, что составляет всего 24,75%. Стоит отметить при этом, что всего 1922 респондент ответили на этот вопрос, то есть доля имеющих положительный опыт онлайн-обучения есть у 55,87% обучающихся, что более соответствует процентному соотношению в опросе студентов НИУ ВШЭ. В [168] при исследовании отношения к онлайн-образованию в двух группах студентов, у одной из которых уже был опыт такого обучения, а у другой – нет, уже отмечали, что студенты, не имеющие подобного опыта, менее уверены в своих силах и способности преуспеть в онлайн-обучении по сравнению с первой группой.

Основная задача анализа предпочтений относительно формата обучения заключается в построении регрессионной модели, позволяющей выделить социально-экономические, демографические и прочие характеристики респондентов, лежащих в основе выбора ответа относительно предпочтений по формату обучения. Зависимой переменной в регрессионных уравнениях по первой и второй базам ответов респондентов выступает бинарная переменная поддержки идеи перевода лекций в онлайн (для респондентов МСЖ) и наличия положительного опыта онлайн-обучения (для респондентов опроса на Coursera).

Перечень независимых переменных для оптимальной спецификации модели логистической регрессии на данных МСЖ, а также коэффициенты этой регрессии, робастные стандартные ошибки коэффициента регрессии, а также значимость на 10- и 5-процентном уровне (\* и \*\*, соответственно) указаны в Таблице А.1 Приложения А.

На данных опроса слушателей курсов НИУ ВШЭ на платформе Coursera были аналогичным образом построены различные логистические регрессии для

объяснения выбора респондентами положительного или отрицательного ответа об опыте прохождения онлайн-курсов, результаты оптимальной спецификации представлены в Таблице А.2 Приложения А.

#### 4.5.3 Краткие выводы

Построенные модели позволяют выявить значимые характеристики респондентов, определяющие их предпочтения по отношению к онлайн-обучению. В выборке студентов НИУ ВШЭ таковыми являются:

- 1) Пол – мужчины оказались более лояльно настроенными по отношению к переводу всех лекций в онлайн-формат (также разное отношение мужчин и женщин к онлайн-курсам отмечали в [169]);
- 2) Средняя оценка – студенты с более низкими оценками более склонны поддержать переход на онлайн-обучение;
- 3) Оценка студентом его образовательной программы – чем выше студент ценит полезность получаемого образования для будущей работы / профессиональной деятельности и условия аудиторного обучения, тем менее он склонен поддерживать замену лекций на онлайн-формат; противоположная ситуация с оценкой уровня преподавания английского языка в Вышке (вероятно, студенты с хорошим знанием языка больше ценят возможность слушать курсы профессоров со всего света, которые в традиционной аудиторной форме обучения для них менее доступны);
- 4) Альтернативное решение о поступлении в магистратуру – чем более выражено было желание магистра поступать в ВШЭ на эту же или другую специализацию, тем ниже вероятность положительного отношения к переходу на онлайн (что объясняется скорее всего тем, что такие студенты очень ценят возможность обучения в этом вузе, бренд вуза и его возможности);
- 5) Опыт обучения на онлайн-курсах также положительно влияет на поддержку такого решения, также как и выраженные жалобы на слишком узкий список онлайн-дисциплин;

- 6) Материальное положение семьи также, что довольно неожиданно, влияет на поддержку такого перехода – студенты, высоко оценившие обеспеченность своего домохозяйства, более склонны к положительному отношению к онлайн-обучению;
- 7) Наличие оплачиваемой работы на полный день вне ВШЭ и неполный день в ВШЭ также положительно влияют на поддержку перехода на онлайн.

Среди респондентов пре-опросов на Coursera можно выделить такие значимые характеристики:

- 1) Разработчики онлайн курсов/платформ более негативно оценивают свой опыт онлайн-курсов;
- 2) Желающие прослушать курс и не покупать сертификат более лояльны к онлайн-образованию;
- 3) Также положительное влияние оказывает серьезное отношение к курсу – чем больше желание просмотреть все видео, тем выше положительная оценка опыта онлайн-обучения (что соответствует результатам [170]);
- 4) Предшествующий опыт онлайн-обучения является значимым, но зависит от платформы – те, кто имел уже опыт онлайн-обучения на Coursera, EdX и Udacity настроены негативно, а слушатели курсов на ИНТУИТ, Универсариум, Национальной платформе открытого образования (openedu.ru) и других платформах скорее положительно оценивают свой опыт;
- 5) Возраст слушателя отрицательно, а наивысший уровень полученного образования положительно влияют на предпочтения по онлайн-курсам;
- 6) В части занятости также можно отметить благосклонное отношение к онлайн-образованию со стороны работающих по найму полный рабочий день, предпринимателей, фрилансеров, находящихся в отпуске работников.

#### **4.5.4 Заключение**

Был проведен регрессионный анализ результатов опросов студентов НИУ ВШЭ и слушателей курсов, предлагаемых НИУ ВШЭ на платформе Coursera. Выявлены основные факторы, определяющие предпочтения респондентов относительно формата онлайн-обучения, в разных группах слушателей онлайн-курсов.

### **4.6 Анализ политик регулирования банковской системы**

#### **4.6.1 Введение**

Ряд академических работ использует индикаторы принятия банками кредитного риска. Это могут быть кредитные рейтинги, оценки вероятности дефолта, значения нормативов достаточности капитала. Часто изменение этих показателей интерпретируют как изменение объемов принятых кредитных рисков. В числе прочего такое изменение ставят в соответствие с мерами монетарной и макропруденциальной политики. По итогам делают выводы о том, что банки склонны принимать больше кредитных рисков при смягчении политик.

Тем не менее подобного рода исследования имеют два значимых ограничения. Во-первых, в них смотрят сразу на все банки. При этом часть банков использует оценки риска, заданные регулятором, а часть – определяет их самостоятельно, с помощью ПВР. Поэтому при исследовании индикаторов кредитного риска необходимо разделять эффекты для ПВР- и не-ПВР-банков. Во-вторых, указанные исследования трактуют снижение вероятности дефолта и увеличение норматива достаточности капитала как уменьшение относительного объема кредитного риска банка. Будет показано, что такого рода выводы могут оказаться несостоятельными, в том числе из-за наличия у банка возможностей по реализации «внутреннего регуляторного арбитража».

Соглашение Базель III и стандарт МСФО 9 являются одними из самых масштабных и недавних новаций в ПБРН. Данные новации схожи в подходах к оценке кредитного риска. Соглашение Базель III сохранило требования к оценке кредитного риска практически без изменений относительно их версии

в соглашении Базель II. Таким образом, ряд уже известных недостатков Базель II и модели Васичека не был учтен. Как будет показано далее, существуют ранее невыявленные эффекты, которые существенно более значительно искажают оценку кредитных рисков.

В 2021 г. Банк России обсуждает возможность перехода всех российских СЗКО на ПВР в обязательном порядке [171]. Выводы исследования особенно важно учесть при таком переходе.

Распространение коронавируса создало существенные риски финансовой нестабильности. Всемирный Банк в июне 2020 г. прогнозировал изменение мирового ВВП на -5,2% за 2020 г. [172]. Введенные ограничения на перемещения людей привели к значимым сокращениям выручки предприятий отдельных отраслей. Для экономической поддержки правительства многих стран ввели антикризисные меры.

Для поддержки экономик в период пандемии 2020 г. Центральные Банки отдельных стран, включая Россию, перешли к мягкой денежно-кредитной политике. В части стран – например, в США [173] и Марокко [174] – впервые в истории установили норматив обязательных резервов на уровне нуля.

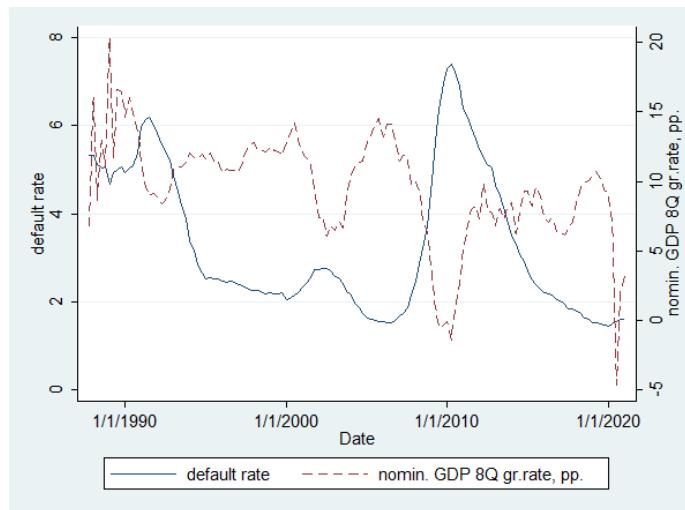


Рисунок 4.7 – Доля дефолтов в США не выросла при спаде ВВП в 2020 г. (Источник: ФРС США, квартальные данные с 1987 по 2021 гг.)

В числе прочих в Европе предположили не использовать критерии не-склонности к платежу (UTP). Это позволит банкам позднее признавать дефолты (неплатежи по кредитам). В частности, беспрецедентный спад в экономике США не нашел отражения в росте доли дефолтов по состоянию на первый квартал 2021 г., см. рисунок 4.7.

Тем не менее, есть прогнозы того, что после окончания действия указанных мер многие компании все равно не смогут вернуть кредиты [175]. Председатель Базельского комитета и Председатель Банка Испании отметил в апреле 2021 г., что еще рано утверждать, что банки прошли пиковую нагрузку по реализации кредитных рисков [176]. По итогам 2020 г. уже отмечают рост банкротств как известных крупных компаний, например, Мурманское морское пароходство [177]; Hertz, L'Occitane [178], так и двукратный рост объемов невыплаченных долгов в сравнении с 2016 г. [179]. Все это указывает на значимо возросшие кредитные риски и материальные последствия их негативных реализаций.

Тогда возникают следующие закономерные вопросы. Насколько банки – и особенно ПВР-банки – корректно учитывают, настолько возросшие кредитные риски в своих моделях? В случаях, когда в моделях существуют материальные ранее невыявленные искажения, каковы их масштабы? Какими должны быть поправки к нормативам (макронадбавки) для их учета? Какова роль банковского регулирования достаточности капитала в создании кредитных денег? Текущее исследование отвечает на данные вопросы.

#### 4.6.2 Основная часть

##### 4.6.2.1 Оценки целесообразности сохранения теневого банковского сектора при наличии регулирования достаточности капитала

В работе [180] обоснована ограниченность существующей теоретической модели [181] участия банков в теневом секторе. В ней есть нереалистичная предпосылка о проводимой регулятором политике по снижению уровня кредитного риска в системе. Обоснована модификация модели. Продемонстрировано достижение оптимального состояния рынка как следствие

такого учета. Обоснованы пороги CAR, когда у банков не возникает стимулов переходить в теневой банковский сектор. Рассчитан эффект требований к капиталу на размер теневого банковского сектора на основе модифицированной теоретической модели.

#### 4.6.2.2 Оценки степени влияния перехода на математические модели Базель II на стоимость греческих банков

В [182] был модифицирован метод разность разностей для оценки эффекта на стоимость банка от перехода на ПВР (модификация № 1). Выделяем подмассивы, где есть пары контрольных и пилотных наблюдений.

Таблица 4.9 – Данные о переходе банков Греции на ПВР и выделение трех подмассивов для сравнения

	d_irb	1	1	1	1	1
	ticker	ETE	EUROB	ATT	TPEIR	ALPHA
d_t	2001	0	0	0	0	0
	2002	0	0	0	0	0
	2003	0	0	0	0	0
	2004	0	0	0	0	0
	2005	0	0	0	0	0
	2006	0	0	0	0	0
	2007	0	0	0	0	0
	2008	1	1	0	0	0
	2009	1	1	0	0	0
	2010	1	1	0	0	0
	2011	1	1	0	0	0
	2012	1	1	0	0	0
	2013	1	1	0	0	0
	2014	1	1	1	1	0
	2015	1	1	1	1	0
	2016	1	1	1	1	0
	2017	1	1	1	1	1

Получаем три массива (см. Таблицу 4.9) – переходим от 5 банков к 11:

- 1) ETE, EUROB – ATT, TPEIP, ALPHA в 2001–2013 (красная рамка – 5 банков);
- 2) ETE, EUROB – ALPHA в 2001–2016 (жирный, курсив, подчеркнутый шрифт – 3 банка);
- 3) ATT, TPEIP – ALPHA в 2001–2016 (ячейки с заливкой – 3 банка).

Таблица 4.10 – Переход от исходных к растиражированным по объектам данным (модификация № 1)

В ячейках число наблюдений		ДО модификации		ПОСЛЕ модификации		
		d_treat (d_irb)		d_treat (d_irb)		ИТОГО
d_time	0	0	56	56	41	54
	1	0	24	24	30	36
	ВСЕГО	0	80	80	71	90
						161

В [182] показали, что переход на ПВР отрицательно сказывается на приросте стоимости банка при прочих равных.

Разработанный метод оценки эффекта для сокращающейся контрольной выборки был развит и применен ко всем американским и европейским банкам, по которым доступна информация о котировках их акций. В итоге, в работе [183] показано, что переход на внутренние модели имел негативный эффект не только в Греции, но и в целом в Европе, чего не наблюдалось в США. Причем это не просто улучшение показателя риск-доходность акций этих банков, это рост средних котировок и снижение волатильности. Заметим, что в Европе переход носил добровольный характер, а в США принудительный. Однако, относить эффект на характер перехода некорректно, поскольку вероятнее инвесторы в США более позитивно оценили тот факт, что иные банки не получили возможности осуществить такой переход, тогда как в Европе переход доступен всем банкам.

#### 4.6.2.3 Оптимальные значения параметров в модели регулирования достаточности капитала исламских банков

В [184] была обоснована модель для определения параметра  $\alpha$  в нормативе CAR для исламских банков с приложением для России.

##### Методология и данные

Зависимой переменной регрессионного анализа является параметр  $\alpha$ , значения которого берутся из работы [185]. В качестве регрессора выступает показатель доли кредитов нефинансовому сектору к ВВП за 2007 г. Данные получены из базы данных “World Development Indicators” Всемирного банка.

Независимой переменной брался также показатель «ВВП на душу населения».

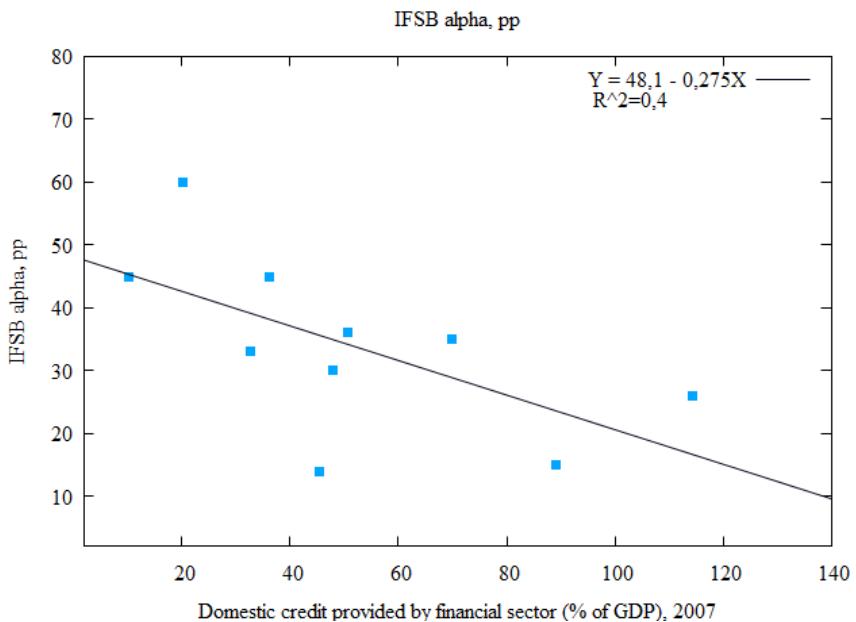


Рисунок 4.8 – Графический анализ статистической связи зависимой и независимой переменных

Рассматривались значения на иные даты от 2007 до 2019 гг. Визуально данные, по которым удалось выявить значимую взаимосвязь, представлены на рисунке 4.8.

### Результаты

Результаты парной регрессии параметра  $\alpha$  и уровня кредитов к ВВП страны демонстрируют наличие отрицательной корреляции на 10%-процентном уровне значимости (см. Таблицу 4.11). Увеличение среднего доли кредитов к ВВП на 1 пп. ведет к уменьшению параметра  $\alpha$  в среднем на 0.3 пп.

Таблица 4.11 – Параметры регрессии для определения параметра  $\alpha$  в нормативе CAR исламских банков

переменная	оценка коэф.	станд. ошибка	t-стат.	p-value	95%-дов.инт, левая граница	95%-дов.инт, правая граница
constant	48,13	7,31064	6,584	0,0002***	31,2717	64,9885
Domestic Credit to GDP	-0,275	0,122316	-2,252	0,0544*	-0,557469	0,00665248
R-квадрат	0,387901;	скор. R-квадрат		0,311388		

Такая связь параметра  $\alpha$  с показателем 10-летней давности может быть оправдана. Регуляторы принимают решения по историческим данным и, ориентируясь на докризисные показатели 2007 г., могли фиксировать параметр  $\alpha$  на уровне, достаточном для компенсации потерь в случае возникновения новых "пузырей" на финансовых рынках.

### Применение для России

В России предпринимаются активные действия по развитию исламского банкинга в Республике Татарстан, а Сбербанк объявил о своих намерениях развивать исламские финансы в Москве, Санкт-Петербурге, Поволжье и на Кавказе. Используя полученную формулу в эконометрической модели из Таблицы 4.11, параметр  $\alpha$  для России в случае введения регулирования исламских банков должен быть 33–42% (Таблица 4.12).

Таблица 4.12 – Рассчитанный для России показатель  $\alpha$

год	Domestic Credit to GDP	$\alpha$
2007	23,92	42%
2018	53,96	33%

### Эффект от регулирования капитала исламских банков

В работе [183] исследован эффект введения регулирования достаточности капитала исламских банков. Как показывает Таблица 4.13, коэффициент при переменных произведения метода разность разностей указывает, что невозможно статистически значимо утверждать, что такое регулирование имело точный количественный эффект на восприятие исламских банков инвесторами из-за незначимости коэффициента. Однако, его положительный знак и значение t-статистики более единицы позволяет утверждать, что устойчивым является положительный знак такого эффекта. Это означает, что при прочих равных котировки исламских банков стали выше, чем у традиционных банков после введения регулирования в 2013 г. и после контроля на прочие равные факторы (на типичные показатели стоимости банков).

Таблица 4.13 – Оценка регрессий

	(1)	(2)	(3)	(4)
VARIABLES	All countries	All countries	Countries with the known alpha	Countries with the known alpha
CAR	-0.2333		-0.3084	

### Продолжение таблицы 4.13

	(0.405)		(0.427)	
ROE	<b>0.5677*</b>	<b>0.5157***</b>	<b>0.4297</b>	<b>0.4447***</b>
	(0.334)	(0.099)	(0.274)	(0.110)
NPL	<b>-0.2440</b>		<b>-0.3714</b>	
	(0.474)		(0.502)	
$d_{t_i}$	<b>-0.0252</b>		<b>-0.0152</b>	
	(0.029)		(0.037)	
$d_{IB_i}$	<b>0.0126</b>		<b>0.0060</b>	
	(0.096)		(0.098)	
$d_{t_i} * d_{IB_i}$	<b>0.0699</b>		<b>0.0770</b>	
	(0.047)		(0.056)	
Bank_FE	yes	yes	yes	yes
Observations	879	879	592	592
R-squared	0.120	0.135	0.105	0.091

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

#### 4.6.2.4 Оценки устойчивости банковской системы при наличии регулирования в виде достаточности капитала

В [186] расширили АОМ банковской системы и продемонстрировали, что банковская система со 100%-м резервированием вкладов до востребования (НОР 100%) и без норматива достаточности капитала является более стабильной, чем современная с частичным резервированием всех вкладов и наличием такого норматива (рисунок 4.9). Независимо от природы кризиса (кредитного и/или ликвидности) норматив CAR является избыточным при НОР 100%, см. рисунок 4.10.

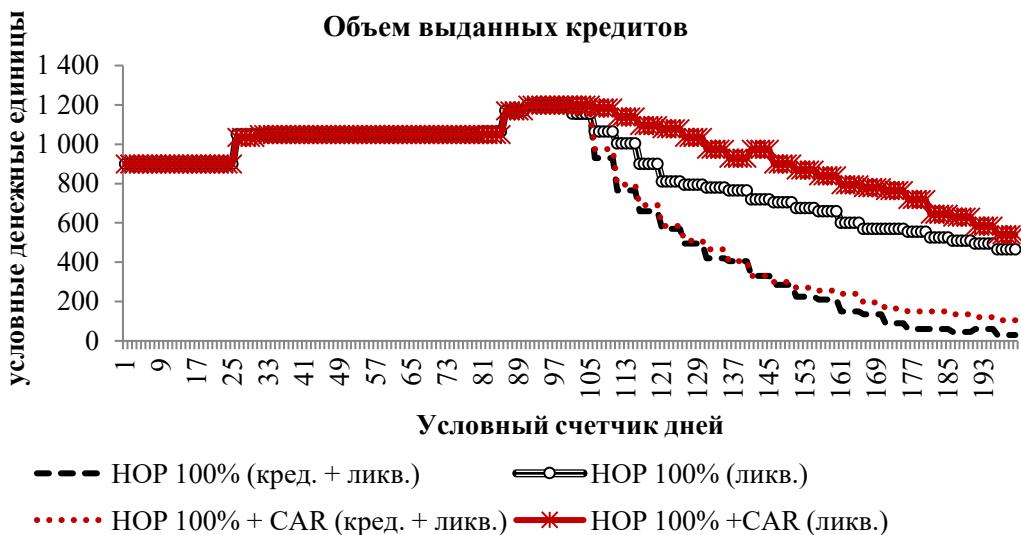


Рисунок 4.9 – Норматив CAR является избыточным и не ограничивает кредитование при наличии 100%-го резервирования депозитов до востребования.

В [187] впервые обратил внимание на то, что норматив обязательных резервов не относится к капиталу банка. Поэтому он смог показать, что норматив достаточности капитала не позволяет создать бесконечный объем кредитных денег при нулевой ставке обязательных резервов.

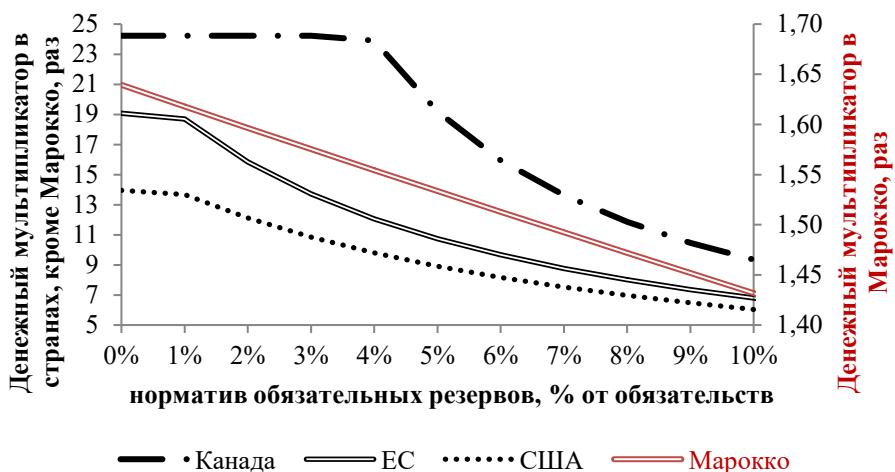


Рисунок 4.10 – Норматив CAR ограничивает создание кредитных денег только при высокой доле безналичных платежей.

При этом норматив CAR не ограничивает создание кредитных денег, когда в экономике низка доля безналичных платежей (например, в Марокко – ниже 30%), но не в США, ЕС, Канаде (где он превышает 90%), см. рисунок 4.10.

#### 4.6.3 Некоторые выводы

Показано, что корректировка порогов норматива достаточности капитала может стимулировать банки не вовлекаться в «теневой банковский бизнес» [180].

Предложена параметризация регулирования достаточности капитала для исламских банков как функция от макроэкономических параметров [184]. При этом выявлено, что введение такого регулирования в целом положительно может сказываться на стоимости исламских банков в сравнении с традиционными [183].

Выявлено, что переход на подход внутренних рейтингов (ПВР) Базель II отрицательно связан с темпами прироста стоимости акций банка [182]. Это характерно как для отдельных стран (например, для Греции), так и для целых регионов (ЕС), в которых переход осуществлялся в добровольном порядке [183].

При необходимости создания дополнительных кредитных денег в экономике с низким уровнем безналичных платежей целесообразно снизить норму обязательных резервов [183]. Это может дать стимулы к краткосрочному экономическому росту. Тем не менее, в более отдаленной перспективе это увеличит угрозу финансового кризиса [186].

## **5 Исследования по сетевому анализу данных**

### **5.1 Введение**

В настоящее время большое внимание уделяется изучению сетевых структур, описывающих множество социальных, экономических, технологических, биологических и других систем. Поскольку большинство существующих сетей неоднородны и имеют сложную структуру, а элементы сети могут воздействовать друг на друга, одной из наиболее важных задач в области сетевого анализа является выявление наиболее влиятельных участников сетевых структур. В литературе известен ряд моделей, которые основаны на количестве взаимосвязей и их интенсивности у каждого элемента, путях между различными элементами и т. д. К сожалению, большинство из них не позволяют учитывать важные характеристики сети: наличие индивидуальных атрибутов у вершин, а также возможность их непрямого и группового влияния друг на друга. Таким образом, при анализе реальных сетевых структур возникает необходимость в использовании моделей центральности, учитывающие данные особенности.

Данное направление посвящено исследованию влияния в реальных сетевых структурах на примере сети международного рынка заимствований, а также сети цитирования экономических журналов.

Международные финансовые отношения играют важную роль во взаимосвязи всех видов экономической деятельности, развитии мировой экономики, а также экономики отдельных стран. Потоки банковского межстранового заимствования всегда, особенно для развитых стран, выступали одним из главных каналов формирования глобальных изменений в ответ на различные меры монетарной политики. Тем не менее, потенциальный крах отдельных финансовых институтов может быстро нарушить работу всей системы и привести к экономическому кризису. В связи с этим возникает необходимость в оценке системной значимости отдельных элементов финансовых систем, что позволит проследить, как реализация системного риска в банковской сфере может привести к возникновению множества негативных послед-

ствий для реального сектора и, тем самым, стать причиной более глобального кризиса. Кроме того, поскольку финансовые отношения изменяются во времени, возникает необходимость в оценке устойчивости сети международного рынка заимствований и выявлении периодов, в рамках которых система находится в одинаковом состоянии.

Что касается сети цитирования экономических журналов, то отметим, что развитие баз цитирования и использование научометрических показателей в управлении наукой привлекло интерес исследователей к изучению и построению различных рейтингов журналов, анализу цитируемости и структуры различных научных областей [188], [189]. Рост роли метрик приводит к манипулированию со стороны участников процесса, в том числе появлению хищнических журналов и, как следствие, к расхождениям между экспертными и научометрическими рейтингами журналов [190]. Одним из перспективных способов исследования является сетевой анализ. В данной части отчета используются методы, которые были впервые применены к анализу научометрических данных. Чаще всего проводится анализ баз Scopus и WoS. Работ, основанных на базе РИНЦ, достаточно мало [191]. В настоящем исследовании использованы данные РИНЦ, которые взяты за основу показателей для сетевого анализа.

Наконец, в рамках сетевого анализа данных рассматривается проблема двусторонней статистики международной торговли продовольствием для построения сетевой структуры. Точность и полнота данных являются одним из главных факторов качества исследования. Тем не менее, проблемы с данными (пропуски, несоответствие различных источников и т. д.) встречаются довольно часто. Поэтому одной из главных задач этого подраздела является разработка подхода к определению числа потока между странами для построения соответствующей сетевой структуры. Предложенный подход основан на официальных открытых источниках ООН, а также на сопоставлении предоставленной информации по торговле между различными странами.

## **5.2 Сетевой анализ сети международного рынка заимствований**

В этом подразделе приведены результаты изучению международного рынка заимствований и влияния стран с точки зрения их роли на данном рынке. Для рынка данного типа можно утверждать, что заимствования определяют влияния стран с двух сторон: на уровне стран-заемщиков, которые чем больше заняли денег, тем более значимы для своих кредиторов; на уровне кредиторов, которые могут рассматриваться как носители системного кредитного риска. Оценка влияния в этом контексте будет эквивалентна определению того, кто является ключевым игроком на рынке заимствований, выстраиванию ранжирования участников таких взаимодействий. В этом случае уровень заимствования стран измеряет риск возникновения финансовых затруднений в той или иной экономике, и возможность его распространения на мировую финансовую систему.

Одним из наиболее удобных инструментов для анализа потоков межстрановых заимствований выступает статистика Банка международных расчетов (Bank for International Settlements, BIS). Статистические данные BIS содержат информацию об иностранных требованиях банков, которая включает в себя данные по кредитам, депозитам, вложениям в долговые ценные бумаги и, кроме того, данные по вкладам в не включаемых в консолидированную отчетность банках или небанковских дочерних организациях и прочим балансовым операциям [192]. Статистические данные охватывают требования отделений банка, в том числе и их зарубежных филиалов, при этом производится взаимопогашение встречных требований.

За последние годы статистика BIS постоянно совершенствуется: растет охват стран, детализация информации, вводятся новые полезные классификации для анализа. С 2005 года BIS также начал предоставлять информацию по иностранным требованиям на основе риска конечного заемщика (*ultimate risk basis*), которая учитывает страновую принадлежность головного офиса финансовых институтов, а также включает в себя часть забалансовых операций (деривативы, гарантии, условные обязательства и т.д.). Исходные данные

показывают объем иностранных требований страны  $i$  по отношению к заемщикам в стране  $j$ , которые включают в себя консолидированные прямые иностранные требования, а также обязательства дочерних компаний и филиалов в стране  $i$  по отношению к стране  $j$ . Другими словами, все требования относятся к стране, в которой находится носитель конечного риска. Таким образом, BIS постепенно стремится к наиболее полному и точному учету возникновения рисков, связанных с предоставлением финансовых кредитных ресурсов.

Статистика Банка международных расчетов предоставляет информацию о межстрановых заимствованиях за каждый прошедший квартал с 2005 по 2020 год, а все страны делятся на две группы: предоставляющие отчетность и не предоставляющие отчетность. Отметим, что если на 1-й квартал 2005 года всего 16 стран предоставляли отчетность по заимствованиям, то к 2020 году общее число стран увеличилось до 23. На данный момент в список стран, представляющих данные на регулярной основе, входят страны Большой десятки (Бельгия, Канада, Франция, Германия, Италия, Япония, Нидерланды, Швеция, Швейцария, Великобритания и США), а также Австралия, Австрия, Чили, Финляндия, Греция, Индия, Ирландия, Португалия, Республика Корея, Испания и Турция. Стоит отметить, что в данный список не входит Китай, а для остальных более чем 200 стран BIS ограничивается информацией о совокупных операциях и ролью на финансовом рынке. Тем не менее, на настоящий момент BIS отмечает, что их статистика охватывает более 90% всех международных банковских требований.

Согласно статистике BIS за 2005–2020 гг., общий объем межстрановых заимствований достиг своего максимума в 2008 году на отметке 28 триллионов долларов, после чего немного снизился и сейчас колеблется в районе 25 триллионов. На рисунке 5.1 приведена информация о структуре основных стран-кредиторов. Стоит отметить, что около 74% от общего объема займов приходится на банковские системы Франции, Германии, Японии, Великобритании и США. Кроме того, было выявлено, что Япония увеличила свое при-

существие на международном рынке заимствований с 2005 года и теперь оно оценивается в 17.5%. Наконец, заметим, что объем заимствований из Германии снизился с 24.6% в 2005 году до 7% в 2020 году.

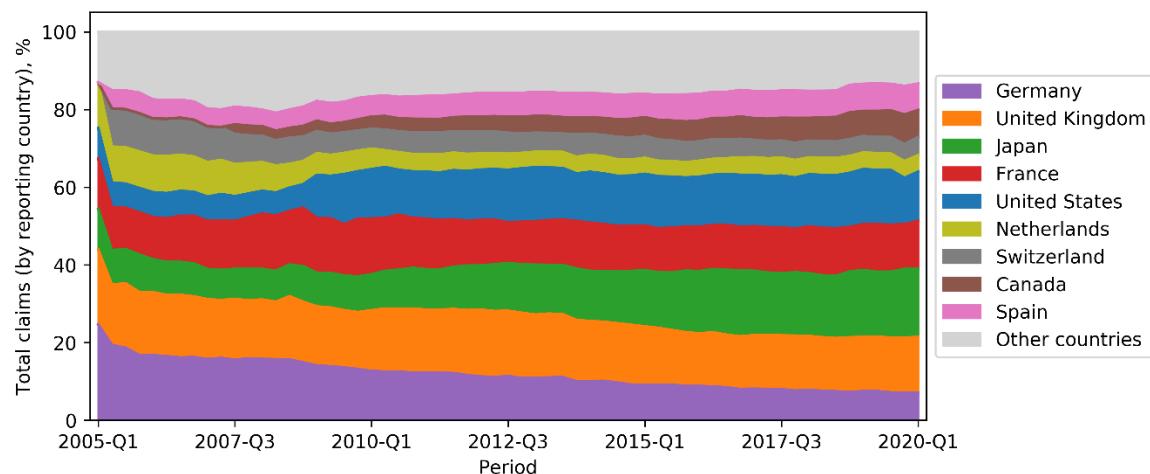


Рисунок 5.1 – Структура стран-кредиторов за 2005–2020 гг.

Структура основных стран-заемщиков приведена на рисунке 5.2. Отметим, что наибольшую долю занимает США – около 23.7% от общего объема заимствований ( $\approx 5.6$  триллиона долларов). Вторым крупнейшим заемщиком является Великобритания, объем заимствований которой снизился с 3.7 триллионов долларов (около 13%) в 2008 году до 2.2 триллионов (около 8.5%). Заметим также, что объем заимствования Германии, Каймановых островов, Франции и Японии оценивается в 4–5% от общего объема заимствований. Наконец, можно сделать вывод, что порядка 61% межстранных заимствований приходится на 9 стран.

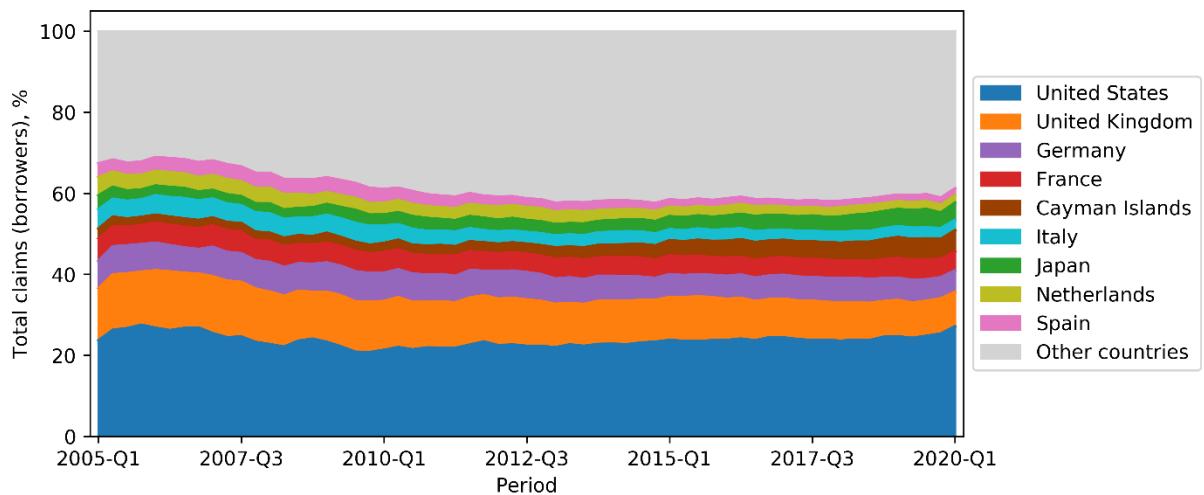


Рисунок 5.2 – Структура стран-заемщиков за 2005–2020 гг.

Проанализировав основных стран-кредиторов и стран-заемщиков, перейдем к анализу сети международного рынка заимствований. Поскольку статистике BIS предоставляет информацию по каждому прошедшему кварталу с 2005 по 2020 год, в рамках данного анализа было построено 61 сеть, каждая из которых характеризует информацию о межстрановых заимствованиях за каждый отдельный период. На рисунке 5.3 приведена иллюстрация данной сети на 1-й квартал 2020 года. Отметим, что для наглядности страны с объемом заимствований менее 200 миллиардов долларов были исключены из рисунка.

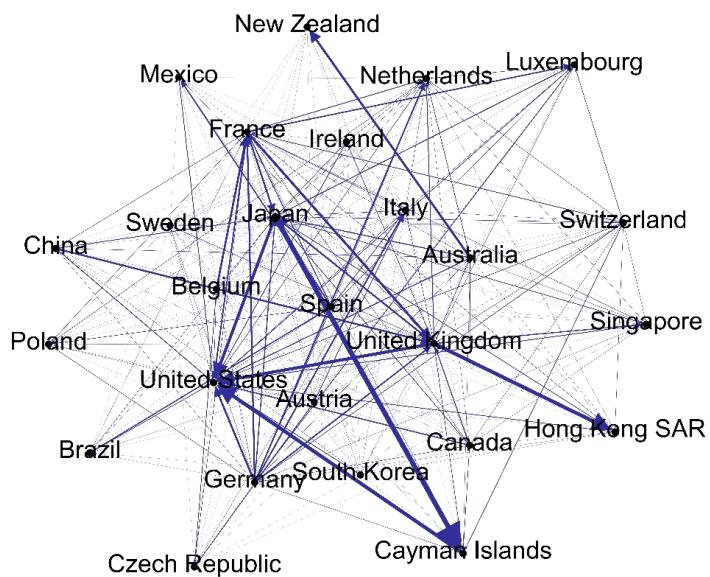


Рисунок 5.3 – Сеть международных заимствований за 1-й квартал 2020 года

Основной целью данного исследования является выявление ключевых заемщиков на финансовых рынках, которые могут повлиять на стабильность всей финансовой системы. Определение наиболее важных элементов сети может быть проведено с использованием классических индексов центральности. Например, ключевые элементы сети могут быть определены с помощью алгоритма PageRank [193]. Данная модель определяет вероятность посещение той или иной вершины в сети, что позволит выделить банковские системы, наиболее задействованные в межстранных взаимоотношениях. Информация о странах, вошедших в ТОП-5 за конкретный период по данному индексу, приведена на рисунке 5.4.

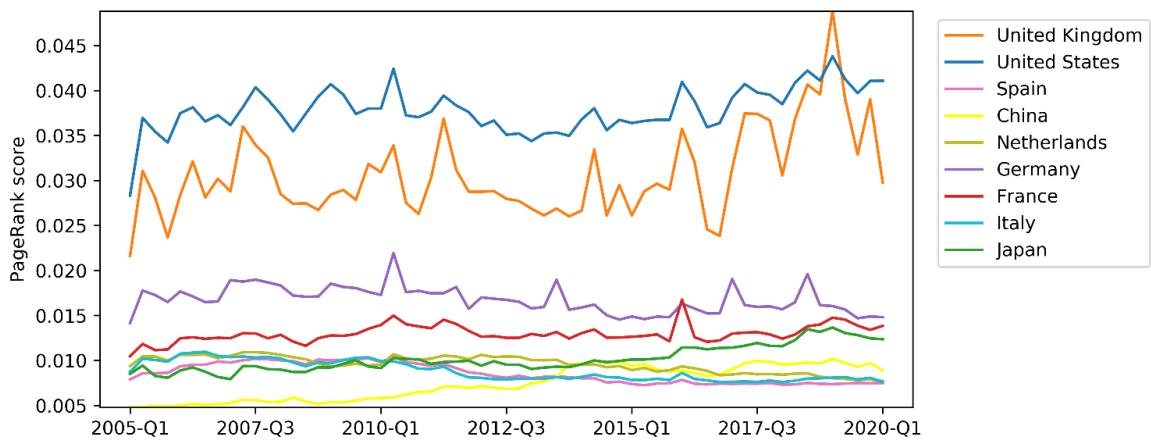


Рисунок 5.4 – ТОП-5 стран по индексу PageRank за 2005–2020 гг.

Согласно рисунку 5.4, США, Великобритания, Германия и Франция входят в список ключевых стран по алгоритму PageRank. Поскольку на данные страны приходится порядка 45–50% всех исходящих потоков и около 40–50% входящих потоков (см. рисунки 5.1–5.2), можно сделать вывод, что данные страны играют большую роль на рынке международных заимствований. Что касается остальных стран, можно заметить увеличение влияния Китая и Японии с 2013 года. Кроме того, в список ключевых стран по данному индексу вошли Италия, Испания и Нидерланды. Наконец, можно отметить, что список и ранжирование ключевых стран практически никак не меняется в период с 2005 по 2020 годы.

Стоит отметить, что классические индексы центральности не учитывают важные характеристики сети: наличие индивидуальных атрибутов у вершин, а также возможность их непрямого и группового влияния друг на друга. В связи с тем, что элементы сети международных заимствований являются неоднородными, в рамках данного исследования был применен индекс дальних взаимодействий (LRIC), который учитывает перечисленные выше особенности сети [194], [195].

Важным аспектом анализа сети международных заимствований с помощью индекса дальних взаимодействий является выбор значения квоты  $q$ , которая показывает предельно допустимую сумму займа, потеря которого будет являться чувствительной для конечного кредитора. Один из возможных способов – установить ее в соответствии с рекомендациями Базельского комитета [196] относительно кредитного риска крупных финансовых организаций (25% от капитала 1 уровня). При этом, когда мы имеем дело с межстрановыми заимствованиями, выбор соответствующего порогового уровня  $q$  (критической суммы кредита) не является столь очевидным. В итоговой модели мы используем на ребрах не просто суммы займов, а отношение суммы займов к валовому внутреннему продукту (ВВП) для того, чтобы учесть относительный размер заемщиков. Другими словами, пороговое значение для каждой страны определяется как 10% от ее ВВП.

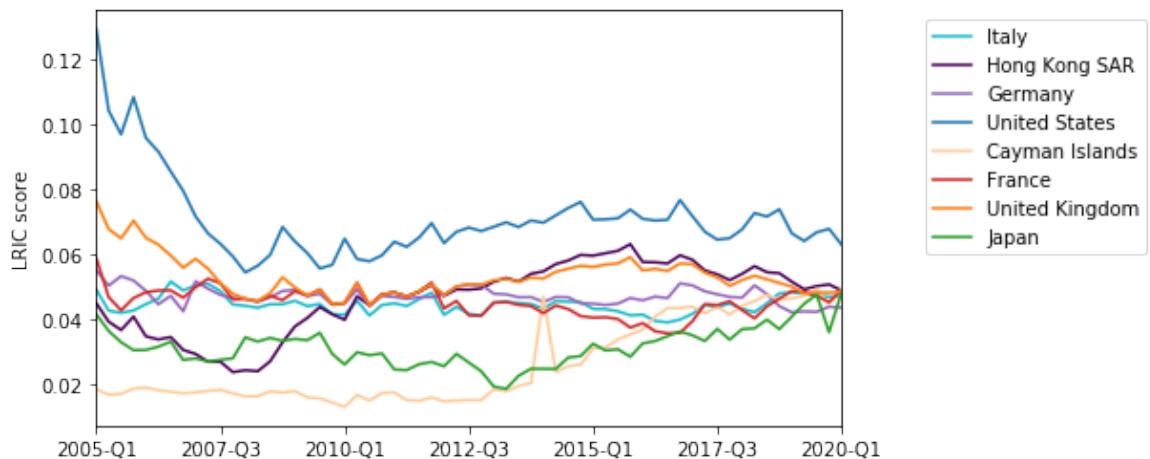


Рисунок 5.5 – ТОП-5 стран по индексу LRIC за 2005–2020 гг.

Информация о наиболее влиятельных странах по индексу дальних взаимодействий приведена на рисунке 5.5. Отметим, что в список трех наиболее влиятельных стран входят США, Великобритания и Гонконг. Несмотря на то, что Гонконг список наиболее влиятельных стран по алгоритму PageRank, его высокую позицию можно объяснить связью с Великобританией. Согласно статистике BIS, консолидированные иностранные претензии Великобритании к Гонконгу составили 435.5 миллиардов долларов в 4-м квартале 2018 года, что составляет около 15% от ВВП Великобритании. Поскольку в рамках данной модели пороговое значение определялось как 10% от общего ВВП страны, можно сделать вывод, что возможный кризис банковской системы Гонконга может привести к цепным реакциям в других странах. Отметим также присутствие в списке Каймановых островов. Благодаря своей привлекательной бизнес-среде, хорошо развитой инфраструктуре, человеческому капиталу и положительной репутации эти страны стимулируют инвесторов предоставлять им свои активы, что делает эти страны крупными системно значимыми заемщиками с высоким уровнем влияния. Наконец, можно сделать вывод, что индекс LRIC позволяет отслеживать не только страны с крупными финансовыми потоками, но и страны, для которых значим общий объем и количество операций (своего рода транзитные экономики). Исключение таких стран из мировой сети заимствований вероятнее всего напрямую не приведет к цепочке банкротств, но может стать причиной перераспределения финансовых потоков между другими участниками, что окажет влияние на стабильность всей финансовой системы.

Наконец, сеть международных заимствований была проанализирована с точки зрения изменения сообществ в данной сети. Основной целью данной работы является выявление группы вершин, которые в некотором смысле более похожи друг на друга, чем на другие вершины сети. Для выявления сообществ в сети был использован алгоритм Infomap, основанный на случайных блужданиях в графе [197]. Определение сообществ было осуществлено за каждый квартал с 2005 по 2020 годы, после чего было проведено их срав-

нение за различные периоды. Информация о распределении числа сообществ за различные периоды приведено в Таблице 5.1. Как показано ниже, большинство рассматриваемых сетей состоят из 1 и 2 сообществ.

Таблица 5.1 – Распределение числа сообществ в сети международных заимствований за 2005–2020 гг.

Число сообществ	Число сетей	Периоды
1	15	2005-Q1, 2006-Q1/2007-Q4, 2018-Q4/2020-Q1
2	30	2005-Q2/2005-Q4, 2008-Q1/2014-Q2, 2018-Q3
3	9	2014-Q3, 2015-Q3, 2016-Q4/ 2018-Q2
4	7	2014-Q4/2015-Q2, 2015-Q4/2016-Q3

Наконец, поскольку каждое сообщество соответствует определенному периоду, проведем анализ того, как сообщества меняются со временем. Результаты данного анализа приведены ниже.

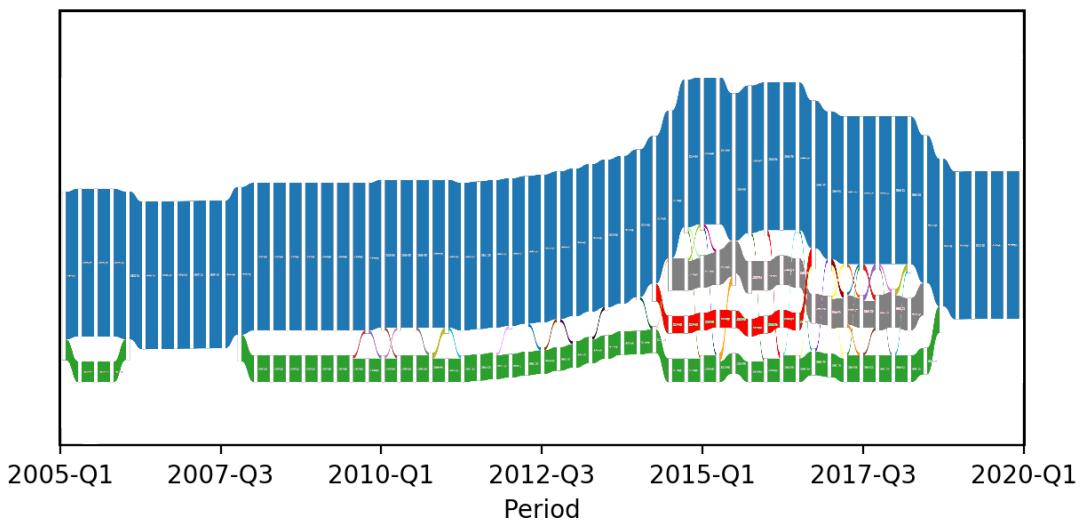


Рисунок 5.6 – Динамика сообществ за 2005–2020 гг.

Как показано на рисунке 5.6, сеть международных заимствований на протяжении почти всего периода состоит из одного большого сообщества (на рисунке отмечено синим цветом), включающее всех крупнейших игроков финансовых взаимоотношений (США, Японию, Великобританию, Германию, Францию и т. д.). Тем не менее, в период с 2005 по 2018 годы были выявлены некоторые небольшие сообщества. Первое из них (отмечено зеленым цветом) включает страны Прибалтики (Эстония, Литва, Латвия) и Скандинавии (Дания, Финляндия, Норвегия, Швеция). Присутствие данных стран может объ-

ясняться тесными связями между данными странами. Второе сообщество (отмечено серым цветом) просуществовало с 3-го квартала 2014 года по 2-й квартал 2018 года и включает преимущественно страны Океании: Новая Зеландия, Фиджи, Микронезия, Кирибати, Науру, Папуа Новая Гвинея, Соломоновы острова, Тонга, Вануату. Наконец, третье сообщество (отмечено красным цветом) включает такие страны, как Португалия, Мозамбик, Кабо-Верде, Сан-Томе и Принсипи (все – португалоговорящие страны), а также Макао, имеющую большой финансовый поток с Португалией. Наконец, можно сделать вывод, что сообщества в сети международных заимствований практически никак не меняется, что говорит об устойчивости самой сети.

### **5.3 Сетевой анализ сети цитирования российских экономических журналов**

В этой части отчета приведены результаты сетевого исследования 466 российских экономических журналов, которые цитируются в РИНЦ. Период анализа равен 5 годам – с 2016 года по 2020 год. Они были проанализированы по таким критериям, как количество цитирований, индекс Хирша, импакт-фактор и Science Index. Были выделены ТОП-5 по каждому из критерий. Часть журналов оказалась в топе в каждом из показателей, другие появлялись только в 1–2 библиометрических показателях в верхних позициях. Стоит выделить журнал «Вопросы Экономики», который входит в топ рейтинга по всем критериям. На рисунках 5.7–5.13 представлены гистограммы распределений.

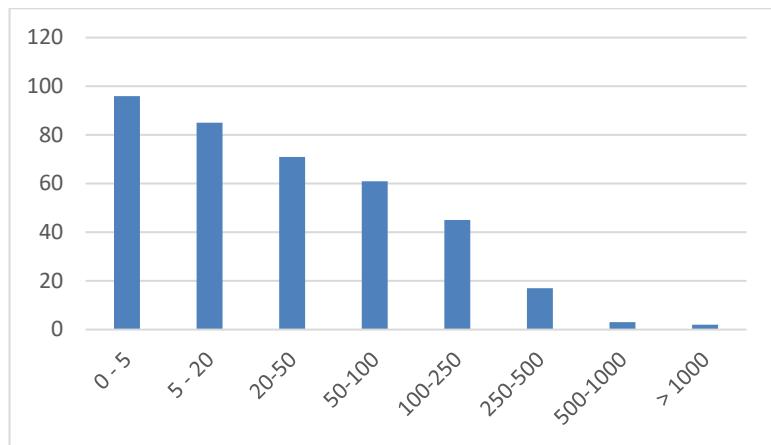


Рисунок 5.7 – Гистограмма распределений цитирований за 4 года по 2016 году

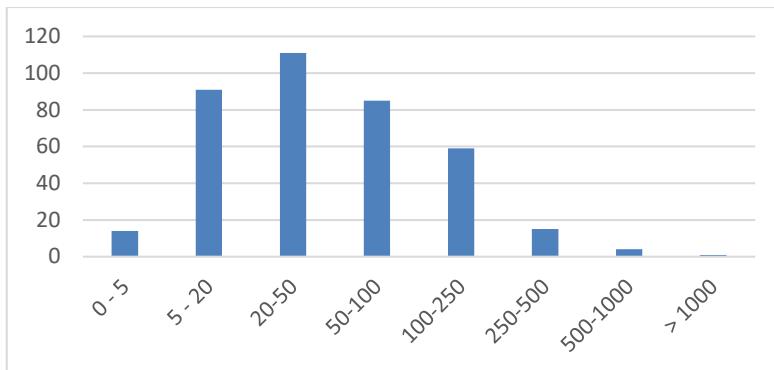


Рисунок 5.8 – Гистограмма распределений цитирований за 4 года по 2020 году

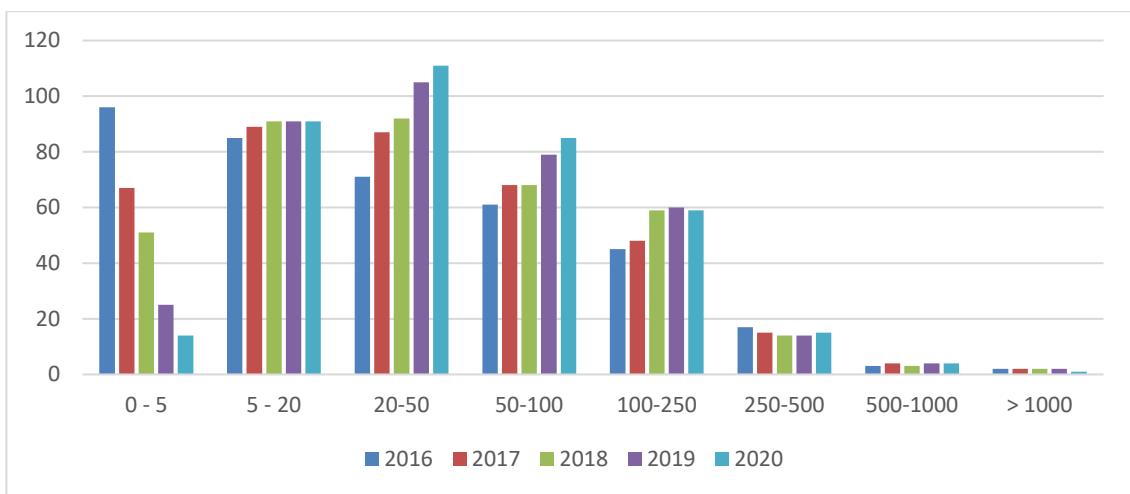


Рисунок 5.9 – Динамика распределений цитирований за 4 года по 2016–2020 годам

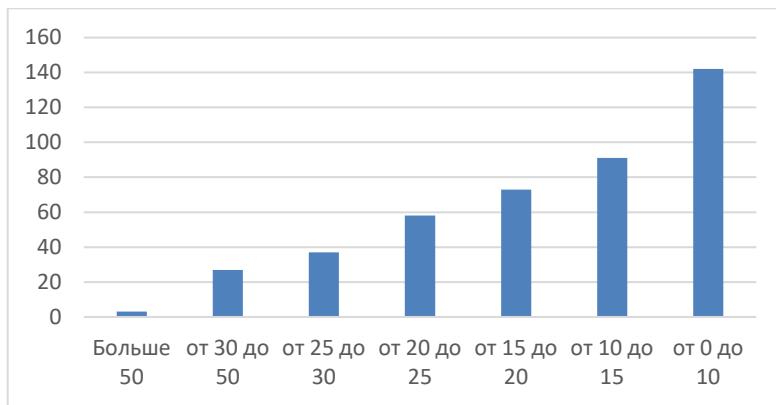


Рисунок 5.10 – Индекс Хирша на 2019 году

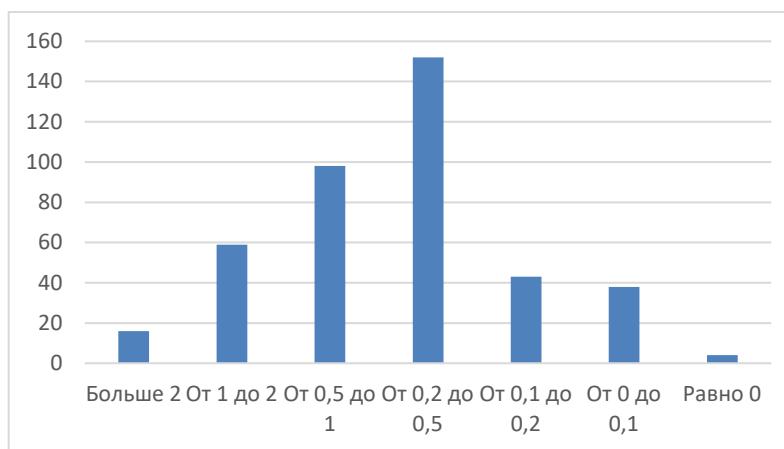


Рисунок 5.11 – Гистограмма распределения двухлетнего импакт-фактора с самоцитированиями на 2019 год

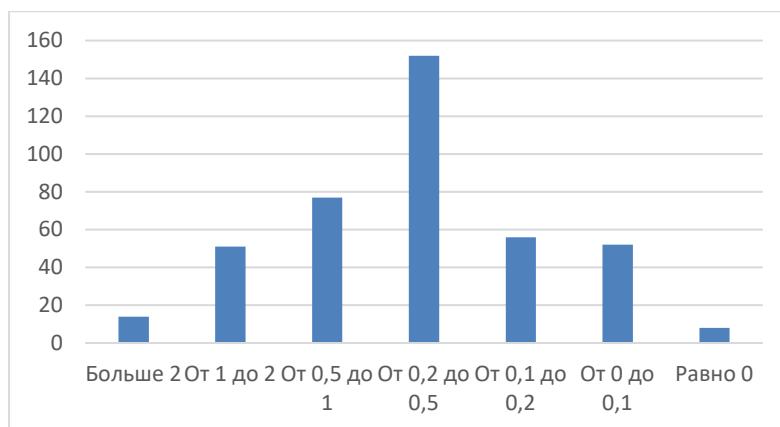


Рисунок 5.12 – Гистограмма распределения двухлетнего импакт-фактора без самоцитирований на 2019 год

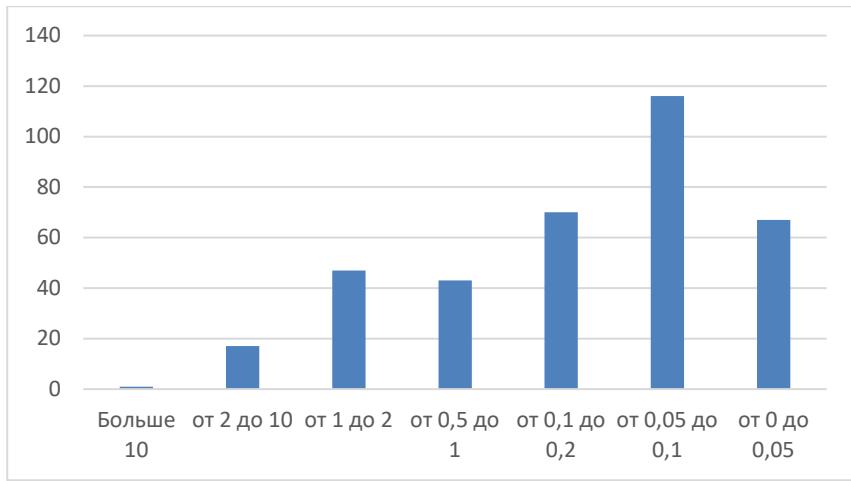


Рисунок 5.13 – Гистограмма распределения Science Index на 2019 год

В связи с неоднозначностью приведенных выше библиометрических показателей, а также появления в списках хороших показателей журналов с плохой репутаций, был предложен метод сетевого анализа [198]. Рассмотрим сеть, в которой вершинами (узлами) являются журналы, а дугами — число ссылок из одного журнала на другой за определенный период времени, получим ориентированный взвешенный граф.

Более подробно остановимся на индексах, которые были использованы для анализа цитирований — индексе Коупленда, Bundle индексе и Pivotal индексе.

Индекс Коупленда  $CI^0 = (CI_i^0)$  внутренней степени на исходном графе  $G^0$ . для каждой вершины определяется как сумма весов  $w_{ji}^0$  входящих ребер из связных вершин  $j$ , т. е.

$$CI^0(i) = \sum_j w_{ji}^0.$$

Рассмотрим пример ниже — сеть, показанную на рисунке 5.14.

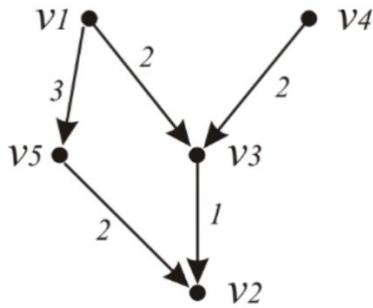


Рисунок 5.14 – Граф

Матрица смежности приведена в Таблице 5.2.

Таблица 5.2 – Матрица смежности

	v1	v2	v3	v4	v5
v1	0	0	2	0	3
v2	0	0	0	0	0
v3	0	1	0	0	0
v4	0	0	2	0	0
v5	0	2	0	0	0

Тогда мы получим следующие значения для  $CI_i^0$ , показанные в Таблице 5.3.

Таблица 5.3 – Значения  $CI_i^0$

	v1	v2	v3	v4	v5
$CI_i^0$	0	3	4	0	3

Чтобы учесть параметры вершин, сначала для каждой вершины  $i \in V$  определим квоту  $q_i$ . Также мы определяем понятие группового влияния вершин на одну вершину. Для этого мы вводим максимальное число  $k$  вершин, которые могут одновременно влиять на узел, т. е. мощность множества, влияющего на узел, не должна превышать  $k$ . Мы называем эти множества критическими группами или критическими наборами. Тогда Bundle индекс  $BI^0 = (BI_i^0)$  индекс строится следующим образом. Для каждого набора вершин,

$$S \subseteq V \setminus \{i\}, |S| \leq k, \forall j \in S, w_{ji}^0 \neq 0,$$

сумма весов  $\sum_{j \in S} w_{ji}^0$  входящих ребер из узла  $j \in S$  в узел  $i$  сравнивается с квотой  $q_i$ , а значение  $BI_i^0(S)$  рассчитывается как

$$BI_i^0(S) = \begin{cases} 1, & \text{если } \sum_{j \in S} w_{ji}^0 \geq q_i, \\ 0, & \text{в ином случае.} \end{cases}$$

Итак,  $BI_i^0(S)$  равно 1, если сумма не меньше квоты, иначе она равна 0.  $BI^0(i)$  индекс для вершины  $i$  определяется как сумма  $BI_i^0(S)$  по всем рассмотренным подмножествам  $S$ , т. е.

$$BI^0(i) = \sum_S BI_i^0(S).$$

Таким образом, для каждого узла  $i \in V$  значение индекса  $BI^0(i)$  равно количеству подмножеств входящих ребер не более, чем из  $k$  вершин, с суммой весов не менее квоты  $q_i$ .

В нашем примере рассмотрим все множества, влияющие на вершины в сети. Если  $q = 1$ , то для вершины  $v_2$  следующие критические группы влияют на эту вершину

- $\{v_3\}$  с суммой весов, равной 1,
- $\{v_5\}$  с суммой весов 2,
- $\{v_3, v_5\}$  с суммой весов, равной 3.

Тогда  $BI^0(v_2) = 3$ . Аналогично  $BI^0(v_3) = 3$  и  $BI^0(v_5) = 1$ . Поскольку нет критической группы, контролирующей  $v_1$  и  $v_4$ , то  $BI^0(v_1) = BI^0(v_4) = 0$ . Значения  $BI^0(i)$  приведены в Таблице 5.4.

Таблица 5.4 – Значения  $BI^0(i)$

	v1	v2	v3	v4	v5
$BI^0(i)$	0	3	3	0	1

Pivotal индекс  $PI^0 = (PI_i^0)$  отличается от индекса  $BI^0$ . Вместо количества наборов ребер вычисляется количество стержневых узлов. Узел  $j_p \in S$  называется стержневым для узла  $i \in V$  в множестве  $S \subseteq V \setminus \{i\}$ , с квотой  $q_i$ , если

$$\sum_{j \in S} w_{ji}^0 \geq q_i, \text{ но } \sum_{j \in S \setminus \{j_p\}} w_{ji}^0 < q_i,$$

т.е. сумма весов ребер, входящих в узел  $i \in V$  из узлов множества  $S \subseteq V \setminus \{i\}$ , больше или равна квоте  $q_i$ , но при исключении узла  $j_p$  из множества  $S$  сумма весов становится меньше квоты.

Индекс  $PI^0$  строится следующим образом. Для каждого из подмножеств узлов  $S \subseteq V \setminus \{i\}$ ,  $|S| \leq k$ ,  $\forall j \in S, w_{ji}^0 \neq 0$  рассчитывается количество стержневых узлов  $PI^0(S)$ .  $PI^0(i)$  индекс рассчитывается как сумма  $PI^0(S)$  на всех рассматриваемых подмножествах  $S$  мощности не более  $k$  вершин входящих ребер в узел  $i$ , т. е.

$$PI^0(i) = \sum_S PI_i^0(S).$$

Рассмотрим наш пример. Для узла  $v_1$  критические множества, влияющие на узел – это  $\{v3\}$ ,  $\{v5\}$  и  $\{v3, v5\}$ . В обеих группах,  $\{v3\}$  и  $\{v5\}$ , оба узла являются стержневыми. Но это не так в группе  $\{v3, v5\}$ . Следовательно,  $PI^0(v_2) = 2$ . Полученные значения  $PI^0$  приведены в Таблице 5.5.

Таблица 5.5 – Значения  $PI^0(i)$

	v1	v2	v3	v4	v5
$PI^0(i)$	0	2	2	0	1

Для анализа использовались сетевые индексы центральности: стандартный индекс центральности Коупленда, индекс группового влияния BI и индекс влияния по ключевым вершинам  $PI^0$ . Для индексов  $BI^0$  и  $PI^0$  максимальный размер критического множества взят равным  $k = 5$ .

Таким образом, была проанализирована сеть цитирования российских экономических журналов. Рассмотрены ранжирования журналов в соответствии со следующими показателями: общее число цитирований за пять лет, индекс Хирша, Science index, импакт-фактор РИНЦ (2-летний), а также — с рядом таких сетевых индексов, как индекс Коупленда и новыми индексами: Bundle и Pivotal для квот в 0,5 и 10%. Результаты исследования показали, что журнал «Вопросы экономики» входит в топ-5 журналов по всем рассмотренным показателям.

Кроме журнала «Вопросы экономики», в соответствии с сетевыми индексами, высокие значения имеют такие журналы, как: «Экономика и предпринимательство», «Экономика региона», «Фундаментальные исследования», «Экономика и управление: проблемы, решения», «Экономика и бизнес: теория и практика». Журнал «Экономика региона» — также среди топ-5 по индексу Хирша и Science index.

#### **5.4 Сетевой анализ двусторонней торговли в условиях асимметрии данных**

Для построения графа какой-либо структуры или процесса в качестве данных необходимо иметь набор вершин и набор рёбер между ними (направленные/ненаправленные, взвешенные/невзвешенные). В данном подразделе мы рассматриваем построение сети международной торговли продовольствием и проблему с двусторонней статистикой в данных. В качестве источника данных мы рассматриваем базу World Integrated Trade Solution (WITS) [199], которая является одной из самых главных баз данных по торговле продовольствием.

Существуют и другие базы данных по торговле продовольствием (WTO [200], OECD [201], BACI [202]). Но, как правило, либо эти базы неполные (содержат информацию за лимитированный период времени, по конкретным странам и продуктам), либо также основаны на данных WITS.

Торговля между странами подразумевает, что существует некоторый товар или услуга, которую одна страна продаёт другой стране. Иными словами, можно утверждать, что существует некоторый денежный поток между продавцом и покупателем, который одинаков для обоих партнёров. Но в большинстве случаев экспортёр и импортёр товара заявляют неравные суммы денежного потока между ними. С одной стороны, обе заявленных статистики могут быть настоящими заплаченными и полученными суммами с учётом косвенных факторов. С другой стороны, для построения сети торговли необходимо выбрать только одно число потока между двумя странами.

Существует несколько причин для таких несоответствий в данных [203]. Например, страны могут подавать статистику в разных системах учёта статистики, с задержкой во времени, с учётом страховых издержек для импортера (обычно страховые издержки могут давать разницу в зеркальной статистике в 10–20%) и т.д. Также в базе WITS имеется большой массив данных, когда только одна страна даёт статистику по потоку, в то время как у её партнёра стоит пропуск.

В данном случае нужно максимально точно и аккуратно выбрать необходимое число, которое будет выступать в качестве веса на соответствующем ребре. Существует множество разных подходов к решению данной проблемы. Например, можно выбрать статистику только одного из партнёров (например, всегда выбирать только статистику экспортёра или импортёра). Несмотря на то, что при таком подходе не смешиваются статистики со страховыми и без страховых издержек, данный метод может привести большому числу пропусков в данных. Более того, сеть, построенная только на статистике экспортёров, сильно отличается по ключевым показателям от сети, построенной только на статистике импортёров. Другие подходы к работе с асимметричными данными рассмотрены в [204].

Таким образом, основной целью этого исследования является разработка нового подхода к определению числа потока между странами.

В данной части мы будем рассматривать общую торговую статистику базы данных WITS согласно номенклатуре HS1996 и сфокусируемся на данных 2018 года (т. к. это одни из последних наиболее полных данных). Из рассмотрения мы убрали всех нестранных партнёров (бункеры, свободные зоны и проч.) в связи с чем у нас осталось 235 территорий и 30875 двусторонних потоков.

На первом этапе мы очистим наши данные от выбросов. Для этого мы рассмотрим среднее значение усечённой выборки данных (отдельно по экспорту и отдельно по импорту) и в качестве порога возьмём некоторый малый процент от этого среднего значения. Всё, что ниже полученного числа будем

считать аномально маленьким значением и удалим из рассмотрения. Так, при усечённой выборке между 5 и 95 перцентилем и при значении в 1% от среднего данной усечённой выборки мы удалим из рассмотрения 14212 потока, что составляет всего 0.008% от общей суммы всех весов на рёбрах.

После очистки данных от выбросов все потоки между странами находятся в одном из двух состояний: либо только один партнёр предложил свою версию статистики (и в таком случае мы будем использовать её как достоверную), либо оба партнёра заявили свою версию торговой статистики и нам необходимо выбрать одно значение для ребра.

Однаковые значения для зеркальной статистики встречаются редко. К сожалению, не существует официальной рекомендации, какому из двух заявленных чисел необходимо доверять, выбор остаётся за исследователем [203]. Но поскольку такой выбор будет субъективен, нам необходим максимально объективный подход к выбору данного числа.

Здесь мы обратимся к официальному опросу ООН касательно практики заполнения торговой статистики, который организация с некоторой регулярностью проводит среди различных стран [205]. Первый опрос был проведён в 1996 году, а второй (и на данный момент последний) был проведён в 2006 году (опубликован в 2008). В опросе приняли участие 132 страны, которым предлагалось ответить на ряд вопросов. Интересно, что по большинству вопросов ООН даёт рекомендаций, как страна должна была бы ответить, если бы она подавала торговую статистику корректно.

По результатам данного опроса мы собрали рейтинг из 40 стран, которые наиболее полно следуют рекомендациям ООН по заполнению торговой статистики. Далее будем использовать этот рейтинг для выбора соответствующего единого потока между странами. Обозначим данную группу стран как *SG* (sustainable group).

Для тех стран, которые не попали в данный рейтинг необходимо сравнить, насколько согласовано они подают свои статистики по сравнению с

группой стран из построенного рейтинга. Далее каждая страна будет проанализирована как экспортёр и как импортёр отдельно.

Рассмотрим две торгующие страны  $A$  и  $B$ . Обозначим за  $A \rightarrow_A B$  поток из страны  $A$  в страну  $B$  по версии статистики страны  $A$  (значение экспорта) и за  $A \rightarrow_B B$  поток из страны  $A$  в страну  $B$  по версии статистики страны  $B$  (значение импорта). Предположим, что обе страны согласны, что между ними существовал поток, равный  $\min(A \rightarrow_A B, A \rightarrow_B B)$ , т. е. данное значение было заявлено обеими странами, а также в дополнение одна из стран заявила что-то сверх данного числа. Полученное значение будет отражать согласованность потока из страны  $A$  в страну  $B$ .

В результате, для каждого экспортёра  $A \notin S$ , и всех его торговых партнёров-импортёров, которые в этом рейтинге присутствуют, мы можем посчитать общую долю согласованных потоков. Тогда метрика согласованности для экспортёра  $A$  будет равна следующему отношению:

$$C_{exp}(A) = \frac{2 \times \sum_{B \in SG} \min(A \rightarrow_A B, A \rightarrow_B B)}{\sum_{B \in SG} A \rightarrow_A B + \sum_{B \in SG} A \rightarrow_B B}. \quad (5.1)$$

Знаменатель в формуле (5.1) равен сумме общего экспорта из страны  $A$  в страны  $B$  по версии страны  $A$  и общего экспорта из страны  $A$  в страны  $B$  по версии страны  $B$ . Множитель 2 в числителе необходим для нормировки данной метрики в интервале  $(0; 1]$ . Когда  $C_{exp}(A) = 1$ , то статистика страны  $A$  полностью совпадает со статистиками её партнёров из  $SG$ . Аналогичную метрику можно построить для стран-импортёров  $C_{imp}(A)$ .

Отметим, что большинство значений данной метрики для экспортёров принадлежит интервалу  $[0.5; 1]$  в то время как метрика для импортёров в основном лежит в интервале  $(0; 0.5]$ . Даный факт говорит о том, что статистика экспортёров более точная, чем статистика импортёров.

В результате для определения единого числа для потока будем использовать два фактора: наличие одного из партнёров в рейтинге  $SG$  и метрику согласованности торговли со странами из  $SG$ .

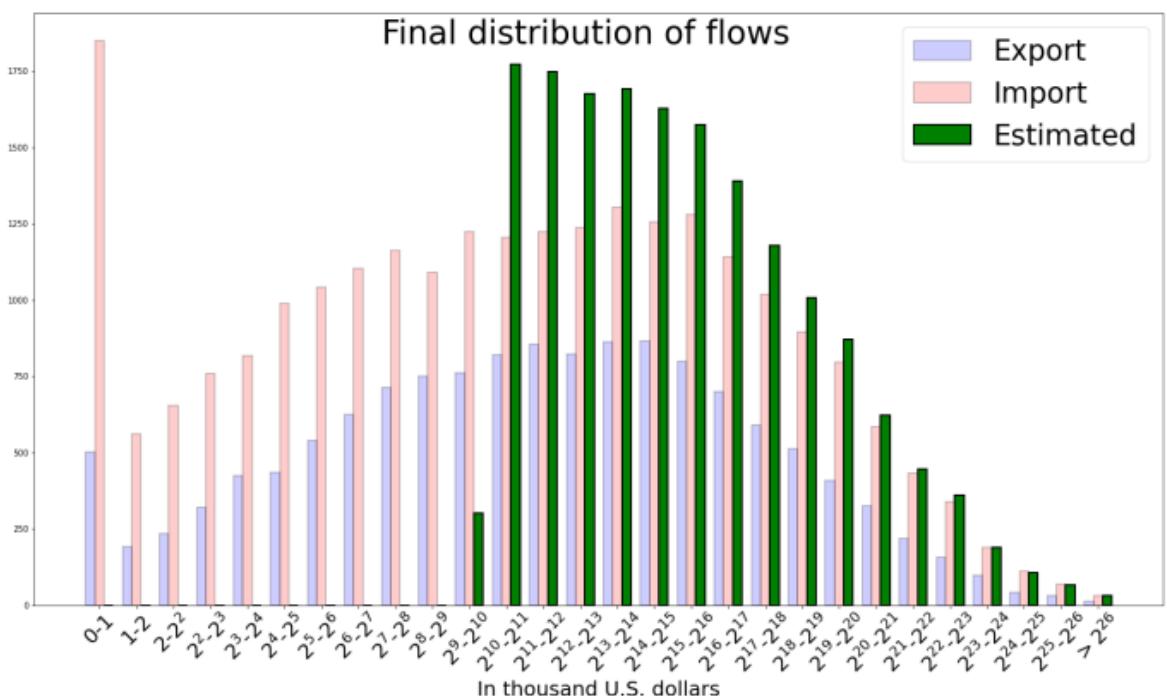
В случае, если оба партнёра принадлежат  $SG$ , то выбираем в качестве единого числа статистику той страны, которая находится выше в данном рейтинге.

Если один из партнёров принадлежит  $SG$ , а другой не принадлежит, то выбираем статистику страны из  $SG$ .

Если оба партнёра не принадлежат  $SG$ , то сравниваем их метрики и выбираем статистику той страны, чья согласованность выше.

В теории возможен случай, когда страны не принадлежат  $SG$ , а также и не торгуют со странами из  $SG$ . Тогда выбор статистики остаётся за исследователем.

На рисунке 5.15 показаны исходные распределения потоков по экспорту и импорту, а также финальное распределение, полученное в рамках построенной модели. Как видно из рисунка, в построенном распределении отсутствует левая часть (хвост с маленькими значениями), в остальном форма распределения сохранилась.



В результате мы построили сеть общей торговли продовольствием. В Таблице 5.6 приведены ключевые сетевые показатели для построенной сети, а также для сетей, построенных только на экспортных и только на импортных данных.

Таблица 5.6 – Ключевые показатели для сети торговли продовольствием, 2018

Мера	Экспорт	Импорт	Построенная
# рёбер	13629	24356	16662
Плотность	0,25	0,44	0,30
# SCC	96	95	9
# вершин в наиб. SCC	140	141	227
# WCC	3	1	4
# вершин в наиб. WCC	233	235	232
Макс. вес	\$ 563 bln.	\$ 314 bln.	\$ 563 bln.
Сумм. вес	\$ 8,4 tln.	\$ 18,3 tln.	\$ 18,7 tln.

Как видно из Таблицы 5.6 в построенной сети больше связей, чем в сети по экспорту, и чуть меньше связей, чем в сети по импорту. Суммарный вес больше всего в построенной сети. Ещё одной важной особенностью является число сильно связных компонент (SCC), которых меньше всего в построенной сети, что говорит нам о том, что новая сеть более связана.

Мы также провели статистический анализ данных сетей и сравнили рёберную корреляцию данных графов. Так, коэффициент корреляции между сетью экспорта и построенной сетью равен 0.48, а коэффициент корреляции между сетью импорта и построенной сетью равен 0.997. Данный факт подтверждает, что наша сеть больше похожа на сеть импорта. Для сетей разных годов (с 1996 по 2018) результаты схожи. Данный факт подтверждает утверждение, что импортные значения передаются более точно и аккуратно [203] и наш выбор данных является обоснованным. Тем не менее, несмотря на высокую корреляцию между построенной сетью и сетью импорта имеются сильные различия в поконтинентальном анализе. Так, сеть импорта с построенной сетью для Африканского континента совпадает на 0.49 (согласно коэффициенту корреляции Пирсона), в то время как аналогичная метрика для экспортной сети равна 0.91.

## **5.5 Некоторые выводы**

В рамках данного направления было проведено исследованию влияния в реальных сетевых структурах на примере сети международного рынка заимствований, а также сети цитирования экономических журналов. В рамках анализа сети международных займов за 2005–2020 гг. был проведен подробный анализ ее структуры, наиболее важных элементов и сообществ. Помимо прочего, была проанализирована устойчивость сети и показано, как сеть международных займов изменяется во времени.

Выявлены наиболее важные российские экономические журналы. При этом были использованы индексы, учитывающие индивидуальные атрибуты вершин, а также возможность их группового и непрямого взаимодействия.

В работе освещена проблема двусторонней статистики международной торговли продовольствием для построения сетевой структуры. Поскольку в большом количестве данных оба торговых партнёра заявляют свою версию денежного потока, существующего между ними, стоит задача выбора одного значения для построения соответствующего ребра в сети. В работе предложен подход по выбору данного значения, основанный на анализе открытых источников ООН (таких как, опрос по заполнению торговой статистики между странами), а также на сравнении согласованности подаваемой статистики участниками торгового процесса. В результате была построена новая сеть торговли продовольствием, а также был проведён сравнительный анализ ключевых сетевых метрик для новой сети и сетей, основанных на экспорте и импорте.

## **6 Исследование моделей распространения**

Исследования моделей распространения (в широком смысле) осуществлялось в этом году в двух направлениях:

- модели распространения COVID-19;
- имплементация разностной схемы высокого порядка точности для решения одной модели распространения.

### **6.1 Модели распространения COVID-19**

6.1.1 Анализ и разработка моделей для прогнозирования распространения COVID-19 по странам мира

COVID-19 начал распространяться по России и по миру в начале 2020 года, сейчас им все еще заражаются сотни тысяч людей в день во множестве разных стран, а конкретно в России более 35 тысяч человек в день – показатели заболеваемости этим вирусом в нашей стране на данный момент достигают исторических максимумов [206]. В такой нестандартной ситуации у структур государственного управления нет четкого понимания, как именно решить проблему, так как введение различных противовирусных мер влечет за собой экономические потери. Например, во втором квартале 2020 года, когда страны начали вводить противовирусные меры, ВВП Евросоюза упал на 11.3% (по сравнению с первым кварталом); сильно пострадала выручка в сферах туризма, авиаперелетов и ресторанно-гостиничных услуг [207].

Анализ собственного и иностранного опыта в борьбе с распространением COVID-19 необходим для оценки эффективности разных противовирусных мер, а мониторинг ситуации и прогнозирование нужны для своевременного введения соответствующих мер. Таким образом, в данной части исследования основными задачами являются: сбор и структуризация данных о заболеваемости и введенных мерах; анализ эффективности мер; выявление групп (кластеров) стран со схожими показателями интенсивности введенных мер, законопослушности, уровня образования, вакцинационной политики; регрессионный анализ прироста заболеваемости для каждого кластера; периодическое обновление моделей на основе новых данных.

Для исследования используются данные о количестве заболевших коронавирусом, об интенсивности введенных мер и о вакцинационной политике из источника «Oxford Covid-19 Government Response Tracker» [208]; данные о числе вакцинированных и об уровне доступа к вакцинации из источника «Our World In Data».

В исходных данных противовирусные меры разбиты на 8 видов: закрытие школ, закрытие предприятий, отмена публичных мероприятий, ограничение собраний, отмена общественного транспорта, требование не покидать жилище, ограничения на поездки внутри страны, ограничения на поездки за рубеж. Каждый вид мер принимает значение от 0 до 3, 4 или 5 в зависимости от интенсивности конкретной меры в конкретный день. Первым делом используется модель порогового агрегирования для объединения 8 представленных видов мер в 2 группы: ограничения собраний (отмена публичных мероприятий, закрытие школ, ограничение собраний, требование не покидать жилище) и государственные запреты (закрытие школ, отмена общественного транспорта, ограничения на поездки внутри страны, ограничения на поездки за рубеж); а также отдельно в одну группу агрегированных мер, в которую входят все 8 видов. Агрегирование необходимо для упрощения последующих расчетов; для построения графиков, проведения графического анализа и подготовки презентационных материалов- для этих же целей еще используется модель нестационарных временных рядов для случаев заражения людей за неделю.

Далее, с использованием метода k-means, страны делятся на несколько кластеров на основе агрегированных данных об интенсивности мер, законопослушности, уровне образования, вакцинационной политике. В таблице еженедельного прироста заболеваемости по странам проводится регрессионный анализ для каждого из полученных кластеров. На основе данных, полученных в результате регрессионного анализа, можно составить уравнение для прогнозирования прироста заболеваемости для каждого кластера стран. Ко-

эффициент детерминации ( $R^2$ -квадрат и скорректированный  $R^2$ -квадрат) используется для оценки точности модели.

Если для регрессионного анализа в качестве зависимой переменной берется неделя  $t$  (в таблице прироста заболеваемости), то данные об интенсивности введенных мер для кластеризации стран берутся за прошлую неделю  $t-1$  (данные о законопослушности и уровне образования со временем не меняются). Частью работы был ответ на вопрос о том за какую неделю лучше всего брать данные о вакцинационной политике. Вакцинационная политика – это показатель доступности вакцинации для населения из источника OxCGRt, принимающий значения от 0 до 5, где 0 обозначает недоступность для всех, а 5 – доступность для всего населения (1–4 обозначают доступность только для определенных групп, например, медработники, пожилые; чем больше групп могут вакцинироваться- тем выше показатель). В ходе работы периодически сравнивались результаты построения моделей с использованием данных о вакцинационной политике за разные недели. В большинстве случаев модели, использовавшие данные за неделю  $t-5$ , имели наивысшие показатели  $R^2$ -квадрат и скорректированный  $R^2$ -квадрат.

Для определения эффективности введенных мер используется модель оболочечного анализа данных (DEA). Модель оценивает входные данные (интенсивность введенных мер и законопослушность) и результат (прирост количества заболевших), на основе чего дает оценку эффективности мер для каждой страны и оценку весов конкретных мер.

Использование описанных выше моделей дает нам комплексную оценку ситуации: возможность мониторинга и графического анализа заболеваемости и введенных мер; определение групп стран со схожими показателями; регрессионный анализ и возможность прогнозирования заболеваемости для каждого кластера; анализ эффективности введенных мер. Одним из преимуществ моделей является их гибкость в быстро меняющейся ситуации. Например, изначально при кластеризации для последующего регрессионного анализа вакцинация никак не учитывалась, так как вакцины от COVID-19 то-

гда еще не выпускали, однако данные о вакцинации удалось внедрить, улучшив показатели R-квадрат. Сейчас в модели используются данные о вакцинационной политике за неделю  $t-5$ , так как они показали наилучшие результаты по сравнению с другими опциями.

### 6.1.2 Оценка эффективности карантинных мер на примере Великобритании

Одним из важнейших событий 2020 года стала пандемия коронавируса, распространившаяся по всему миру, которая стала причиной как кризиса в сфере здравоохранения, так и экономический кризис. Даже самые мощные экономики сталкиваются с тяжелой борьбой, не в силах противостоять давлению пандемии. Исследования по оценке экономической сферы во время пандемии выявили такие же серьезные проблемы, как и проблемы, связанные с пандемией испанского гриппа и Великой депрессией вместе взятых. Задачи, стоящие перед правительствами многих стран, отличаются своей спецификой. В короткие сроки требуется много новых видов вмешательств. Следовательно, COVID-19 поставил огромное количество отраслей во многих сферах по всему миру на грань выживания и в то же время увеличил уровень безработицы в глобальном масштабе.

Анализ ключевых показателей экономики, а также сфер экономической деятельности населения является ключевым в определении эффективности мер, принимаемых государством против распространения вируса. Таким образом, в данном исследовании на примере Великобритании сопоставлены изменения экономических показателей с заболеваемостью, мерами, ограничивающими передвижение населения, а также со строгостью государственного вмешательства.

Для исследования были использованы изменения основных экономических показателей. Они сравнивались с количеством новых случаев заболевания коронавирусом, государственными запретами и мерами, ограничивающими собрания. Для этого более подробно рассмотрены такие показатели, как ВВП, расходы домохозяйств, уровень безработицы, экспорт и импорт то-

варов. Данные о государственных запретах и мерах, ограничивающих передвижение населения брались из The Oxford COVID-19 Government Response Tracker [208].

Великобритания – десятая по величине экономика в мире. Движущей силой экономики является большой сектор услуг. В этом секторе занято более 81% рабочей силы Великобритании, что приводит к тому, что он формирует более 71,3% ВВП [209]. Но во время пандемии данный сектор оказался самым уязвимым, что и сказывалось на показателях ВВП из квартала в квартал. В 2020 году ВВП Великобритании продемонстрировал самое сильное падение с 2008 года [210]. Во втором квартале 2020 года также было зафиксировано самое резкое падение ВВП по сравнению с аналогичным периодом прошлого года – 21,5%. Он также снизился на 19,8% по сравнению с первым кварталом. Также этот период отличился тем, что отрицательный рост в этом квартале показали все 9 регионов. Наибольшее падение среди всех регионов наблюдалось в Уэст-Мидлендсе (на 21,0%), на востоке Англии (на 20,9%), на юго-востоке (на 20,6%) и на северо-западе (на 20,3%). При этом наименьшее снижение произошло в Лондоне и на Северо-Востоке – на 16,6% и 17,5% соответственно [211]. При этом правительственные ограничения оставались на прежнем довольно низком уровне.

Самые низкие расходы домохозяйств были зафиксированы во втором квартале 2019 года, что коррелирует с первой изоляцией, когда население было вынуждено оставаться дома. Соответственно снизились затраты [212].

В то же время в 2020 году наблюдался сильный рост безработицы. Пик роста пришелся на конец года. В четвертом квартале уровень безработицы увеличился до 5,1%, что составляет около 1,7 млн человек. При этом резких скачков не наблюдалось, рост происходил постепенно. В первом квартале – 4%, во втором квартале – 4,1% и в третьем квартале – 4,8%. Такие изменения могут указывать на то, что влияние платежей, производимых государством для поддержания рабочих мест, начинает ослабевать к концу года [213].

Наибольшее сокращение экспорта и импорта происходит также во втором квартале 2020. Изменения в экспорте и импорте происходят практически параллельно. В то же время экспорт достигает своей самой низкой отметки в апреле – 20,3 миллиарда фунтов, что на 7,7 миллиарда фунтов меньше, чем за тот же период 2019 года. Импорт достигнет самой низкой точки в следующем месяце, в мае. Стоимость импорта составила 27,4 миллиарда фунтов, что на 13,8 миллиарда фунтов меньше, чем за тот же период 2019 года. С июня экспорт и импорт начали восстанавливаться и уже продолжали расти до конца 2020 года, несмотря на второй локдаун, введенный в ноябре [214].

Помимо экономических показателей, важно обращать внимание и на изменения в конкретных сферах экономической жизни населения. В данной работе были рассмотрены следующие области: путешествия, высшее образование, сферы гостеприимства и искусства.

В июле 2020 года аэропорты Великобритании потеряли не менее 4 миллиардов фунтов стерлингов из-за пандемии коронавируса. Аэропорты теряли более 150 миллионов фунтов стерлингов ежедневно в течение первых четырех месяцев пандемии [215]. Выручка снизилась чуть менее чем на 2 миллиарда фунтов стерлингов по сравнению с предыдущим годом, так как количество пассажиров упало на 99% [216].

В сентябре возникла новая проблема, с которой пришлось столкнуться британским авиакомпаниям. Директор аэропорта Хитроу Джон Холланд-Кей выступил с заявлением, в котором заявил, что 14-дневный карантин, введенный правительством Великобритании для пассажиров, прибывающих из-за границы, может в дальнейшем привести к банкротству британских авиакомпаний и аэропортов [217].

Важной статьей дохода в сфере услуг Великобритании является экспорт высшего образования. Если оценивать финансовое влияние пандемии коронавируса на высшее образование, то из-за потери первокурсников, снижения платы за обучение убыток всего сектора оценивается в 2,472 миллиарда фунтов стерлингов [218].

Самым трудным временем для сферы гостеприимства и туризма оказался также второй квартал 2020 года. Иностранные резиденты в тот период совершили на 96% меньше поездок в Великобританию и потратили на 97% меньше денег, чем за тот же период прошлого года [219].

Сфера искусства оказалась наиболее пострадавшей от пандемии и мер, принимаемых государством. В период до пандемии культурный сектор Великобритании рос в пять раз быстрее, чем экономика в целом. В нем было занято более 2 миллионов человек, и его вклад в экономику составил 111,7 миллиарда фунтов – больше, чем в автомобильной, аэрокосмической, медико-биологической и нефтегазовой отраслях вместе взятых. Тем не менее музеи и галереи потеряли около 743 миллиона фунтов дохода [220].

Карантинные меры, введенные правительством Великобритании, не всегда применялись вовремя. Правительство реагировало на новые волны роста заболеваемости в отложенном режиме. Несмотря на ранние предпосылки ухудшения ситуации с коронавирусом, кардинальные меры применяются тогда, когда уровень заболеваемости достигает нового максимума. При этом второй локдаун, который старались избежать, оказался достаточно мягким для экономики, во время него даже наблюдалась положительная динамика в некоторых экономических сферах. Именно первый локдаун оказался самым коварным для экономики и различных сфер. В то же время правительственные меры, по данным Oxford's Government Response Tracker, во время второго локдауна были на самом низком уровне за весь период пандемии в Великобритании. Возможно, это связано с тем, что во втором локдауне государственные меры были уже более организованными.

Также следует учесть, что влияние на экономические показатели Великобритании оказывает Brexit. Поэтому данное исследование требует более глубокого изучения проблемы и построения расчетов.

## **6.2 Имплементация разностной схемы высокого порядка точности для решения одной модели распространения**

### 6.2.1 Введение

Программируемая клеточная гибель (programmed cell death — PCD) — процесс, при котором клетка разрушает сама себя посредством некоторых внутренних контролируемых процессов [221]. Наличие PCD в многоклеточных организмах вполне понятно — повреждённые клетки будут вредить всему организму, поэтому они уничтожают себя ради блага всего организма. Однако, такое поведение наблюдается и у одноклеточных организмов — это поведение кажется странным, поскольку смерть клетки в этом случае означает смерть всего организма.

PCD возникает в колониях одноклеточных организмов из-за родственного отбора. Организмы в этих колониях живут так, чтобы приносить максимальную пользу не себе, а всей колонии.

### 6.2.2 Модель апоптоза

#### Построение модели

Гибель клеток проявляется макроскопическими морфологическими признаками. Выделяют три основных типа: апоптоз, аутофагия и некроз [222].

При апоптозе (от др.-греч. ἀπόπτωσις «падение», от ἀπό- и πτῶσις «падение») цитоплазма клеток сокращается и образуются более мелкие везикулы — апоптические тельца, которые поглощаются соседними клетками. Поглощение апоптических телец происходит без утечки цитоплазмы, что значит, что апоптоз происходит без воспалительной реакции. Кроме того, апоптоз является хорошо регулируемым процессом и для него нужно достаточно количество молекул АТФ, которые являются универсальными источниками энергии для всех биохимических процессов, протекающих в клетках.

Аутофагия (от др.-греч. αὐτός — «сам» и φαγεῖν — «есть») характеризуется тем, что дисфункциональные органеллы и внутриклеточные компоненты доставляются в лизосомы для разложения и повторного использования.

Некрозом (от др.-греч. νέκρωσις — «смерть») называется смерть клеток, не проявляющая отличительных черт апоптоза и аутофагии. Удаление мертвых клеток происходит без поглощения соседними клетками и без участия лизосомы. При некрозе происходит утечка цитоплазмы, что вызывает воспалительные реакции. Некроз является энергетически независимой клеточной гибелью.

Рассмотрим апоптоз и его взаимодействие с ростом клеток у млекопитающих. Апоптоз — это строго контролируемое сложное явление, которое может быть вызвано двумя основными путями: внешним и внутренним. В этом процессе участвуют ферменты специального типа — каспазы, которые инициируют гибель клеток посредством расщепления пептидной связи между аминокислотами в молекулах белка.

Для построения модели используется предположение о наличии различных клеточных субпопуляций на основе активности различных каспаз:

- $x_1$  — количество здоровых клеток;
- $x_2$  — количество апоптических клеток с активной каспазой-8 (внешний путь) и неактивной каспазой-3;
- $x_3$  — количество апоптических клеток с активной каспазой-9 (внутренний путь) и неактивной каспазой-3;
- $x_4$  — количество апоптических клеток с активной каспазой-3.

Предполагается, что нормальные здоровые клетки подвергаются апоптозу посредством любого из двух путей:

- внешнего, который формирует клетки с активной каспазой-8;
- внутреннего, который формирует клетки с активной каспазой-8.

Затем эти две субпопуляции трансформируются в клетки с активной каспазой-3. После этого клетки умирают.

Жизнеспособная популяция состоит из делящихся и растущих клеток, а также клеток, находящихся в состоянии покоя, которые потребляют энергетический ресурс только для поддержания жизни, а не для роста. Доля расту-

щих и делящихся клеток описывается моделью роста, предложенной Тессье [223] следующим образом:

$$f_{gr} = 1 - e^{-\frac{S}{K_s}},$$

где  $S$  – питательное вещество и  $K_s$  – константа Тессье.

Скорость роста здоровых клеток определяется формулой

$$\mu_g f_{gr} x_1,$$

где  $f_{gr} x_1$  – количество здоровых делящихся/растущих клеток, а  $\mu_g$  – удельная скорость роста клеток, которая описывается формулой

$$\mu_g = \mu_L \left( 1 - \frac{x_1 + x_2 + x_3 + x_4}{X_{\max}} \right).$$

Здесь  $X_{\max}$  – максимально возможное количество клеток, которые могут находиться в данном объёме, и  $\mu_L$  – некоторый постоянный коэффициент.

Внешний путь апоптоза происходит из-за межклеточного взаимодействия. Клетки, находящиеся на стадии активной каспазы-3, могут передавать сигналы здоровым клеткам. Скорость начала внешнего пути апоптоза (скорость образования клеток с активной каспазой-8) зависит от количества клеток с активной каспазой-3 и задаётся формулой

$$\mu_e x_4^\alpha x_1,$$

где  $\mu_e$  и  $\alpha$  – постоянные коэффициенты, а  $\mu_e x_4^\alpha$  – удельная скорость образования клеток с активной каспазой-8.

Скорость начала внутреннего пути зависит от количества питательных веществ и определяется формулой

$$\mu_i \frac{1}{1+k_{x_3}S} x_1,$$

где  $\mu_i$  и  $k_{x_3}$  – постоянные коэффициенты.

Скорость перехода к стадии с активной каспазой-3 постоянна для обоих путей апоптоза. Тогда скорость перехода от стадии с активной каспазой-8 к стадии с активной каспазой-3 равна  $k_{x_4}x_2$ , а скорость перехода от стадии с активной каспазой-9 к стадии с активной каспазой-3 –  $k_{x_5}x_3$ , где  $k_{x_4}$  и  $k_{x_5}$  – некоторые постоянные.

Скорость смерти клеток от апоптоза также постоянна и зависит лишь от количества клеток, находящихся в стадии активной каспазы-3, т.е. равна  $\mu_d x_4$ , где  $\mu_d$  – постоянный коэффициент.

Скорость потребления питательного вещества всей популяцией равна  $-k_{\text{sup}}(x_1 + x_2 + x_3)$ , где  $k_{\text{sup}}$  — некоторая постоянная.

На основе данной модели строится динамическая модель, основанная на системе обыкновенных дифференциальных уравнений:

$$\left\{ \begin{array}{l} x_1' = \mu_g f_{gr} x_1 - \mu_e x_4^\alpha x_1 - \mu_i \frac{1}{1+k_{x_3}S} x_1 \\ x_2' = \mu_e x_4^\alpha x_1 - k_{x_4} x_2 \\ x_3' = \mu_i \frac{1}{1+k_{x_3}S} x_1 - k_{x_5} x_3 \\ x_4' = k_{x_4} x_2 + k_{x_5} x_3 - \mu_d x_4 \\ S' = -k_{\text{sup}}(x_1 + x_2 + x_3) + \gamma \mu_d x_4 \\ \mu_g = \mu_L \left( 1 - \frac{x_1 + x_2 + x_3 + x_4}{X_{\max}} \right) \\ f_{gr} = 1 - e^{-\frac{S}{K_S}} \end{array} \right.,$$

где  $\gamma$  – часть умерших клеток, которые поглощаются другими клетками и становятся дополнительным ресурсом.

### 6.2.3 Разностная схема для численного решения модели апоптоза

Пусть  $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ ,  $\mathbf{y} = \begin{pmatrix} \mathbf{x} \\ S \end{pmatrix}$  и  $\mathbf{f} : \mathbf{R}^5 \rightarrow \mathbf{R}^5$  — отображение такое, что

$$\mathbf{f}(\mathbf{y}) = \begin{pmatrix} \mu_g f_{gr} x_1 - \mu_e x_4^\alpha x_1 - \mu_i \frac{1}{1+k_{x_3} S} x_1 \\ \mu_e x_4^\alpha x_1 - k_{x_4} x_2 \\ \mu_i \frac{1}{1+k_{x_3} S} x_1 - k_{x_5} x_3 \\ k_{x_4} x_2 + k_{x_5} x_3 - \mu_d x_4 \\ -k_{\text{sup}} (x_1 + x_2 + x_3) + \gamma \mu_d x_4 \end{pmatrix},$$

$$\text{где } \mu_g = \mu_L \left( 1 - \frac{x_1 + x_2 + x_3 + x_4}{X_{\max}} \right), f_{gr} = 1 - e^{-\frac{S}{K_S}}.$$

Рассматриваемая система ОДУ тогда принимает вид  $\mathbf{y}' = \mathbf{f}(\mathbf{y})$ . Определим равномерную сетку  $\omega_{\Delta t}$  на  $\mathbf{R}$  с узлами  $t_k = k\Delta t, k \in \mathbf{Z}$  и шагом  $\Delta t = \frac{T_{\max}}{N}$ , где  $T_{\max} > 0$ . Положим  $\mathbf{y}_k = \mathbf{y}(t_k)$ .

Воспользуемся линейным многошаговым методом для решения системы. Сперва используем явный четырёхстадийный метод Рунге-Кутты четвёртого порядка точности (RK4) для нахождения первых четырёх значений решения. Метод относится к семейству хорошо известных численных методов и является одним из наиболее широко изученных методов Рунге-Кутты [224]. Метод RK4 — метод более высокой точности, который обеспечивает хорошую аппроксимацию решения. Его расчётные формулы приведены ниже.

$$\begin{aligned}
y_{n+1} &= y_n + \frac{1}{6} \Delta t \left( k_n^{(1)} + 2k_n^{(2)} + 2k_n^{(3)} + k_n^{(4)} \right), \\
k_n^{(1)} &= f(y_n), \\
k_n^{(2)} &= f \left( y_n + \frac{\Delta t}{2} k_n^{(1)} \right), \\
k_n^{(3)} &= f \left( y_n + \frac{\Delta t}{2} k_n^{(2)} \right), \\
k_n^{(4)} &= f \left( y_n + \Delta t k_n^{(3)} \right).
\end{aligned}$$

Затем, воспользуемся четырёхшаговым методом Адамса предикции-коррекции четвёртого порядка точности [224]. Это также один из наиболее широко применяемых на практике численных методов для решения систем обыкновенных дифференциальных уравнений. Расчётные формулы метода Адамса:

$$\begin{aligned}
\text{Прогноз: } y_{n+1}^{(0)} &= y_n + \frac{\Delta t}{24} (55f(y_n) - 59f(y_{n-1}) + 37f(y_{n-2}) - 9f(y_{n-3})), \\
\text{Коррекция: } y_{n+1} &= y_n + \frac{\Delta t}{24} \left( 9f(y_{n+1}^{(0)}) + 19f(y_n) - 5f(y_{n-1}) + f(y_{n-2}) \right).
\end{aligned}$$

#### 6.2.4 Реализация разностной схемы

Разностная схема была реализована на языке программирования C# в среде Visual Studio 2019 с использованием основных принципов объектно-ориентированного программирования (ООП). Для описания всех глобальных параметров (напомним, что язык C# не позволяет определять глобальные переменные – все переменные должны быть определены внутри класса) был создан статический класс GlobalParameters, в который вошли все известные постоянные коэффициенты, а также правая граница отрезка вычислений, максимальное количество узлов и шаг сетки по времени. Использование отдельного класса для глобальных параметров является достаточно распространённой практикой при использовании полностью объектно-ориентированных языков и позволяет в дальнейшем легче отлаживать исходный код программы.

Основой программы является класс Vector для операций векторной алгебры, в котором определены следующие методы:

- public Vector(int a) – конструктор нулевого вектора заданной длины;
- public Vector(double[] array) – конструктор вектора из заданного массива данных;
- public static Vector operator +(Vector c1, Vector c2) – операция сложения векторов;
- public static Vector operator -(Vector c1, Vector c2) – операция вычитания одного вектора из другого;
- public static Vector operator \*(double a, Vector c) – операция умножения вектора на скаляр;
- public double this[int i] – операция обращения доступа к вектору по индексу;
- public static Vector f(Vector y) – операция вычисления правой части системы дифференциальных уравнений.

Благодаря использованию класса, реализацию разностной схемы удаётся записать в компактном виде, максимально приближенном к её математической записи:

```
var k1 = Vector.f(y);
var k2 = Vector.f(y + GlobalParameters.deltat / 2.0 * k1);
var k3 = Vector.f(y + GlobalParameters.deltat / 2.0 * k2);
var k4 = Vector.f(y + GlobalParameters.deltat * k3);
var k = 1.0 / 6.0 * (k1 + 2 * k2 + 2 * k3 + k4);
y = y + GlobalParameters.deltat * k;
```

для метода Рунге-Кутты

и

```
var ty = y + GlobalParameters.deltat / 24.0 * (55 * Vector.f(y) - 59 * Vector.f(y3) +
37 * Vector.f(y2) - 9 * Vector.f(y1));
var ny = y + GlobalParameters.deltat / 24.0 * (9 * Vector.f(ty) + 19 * Vector.f(y) -
5 * Vector.f(y2) + Vector.f(y1));
```

для метода Адамса предикции-коррекции.

На рисунке 6.1 приведён интерфейс окна программы.

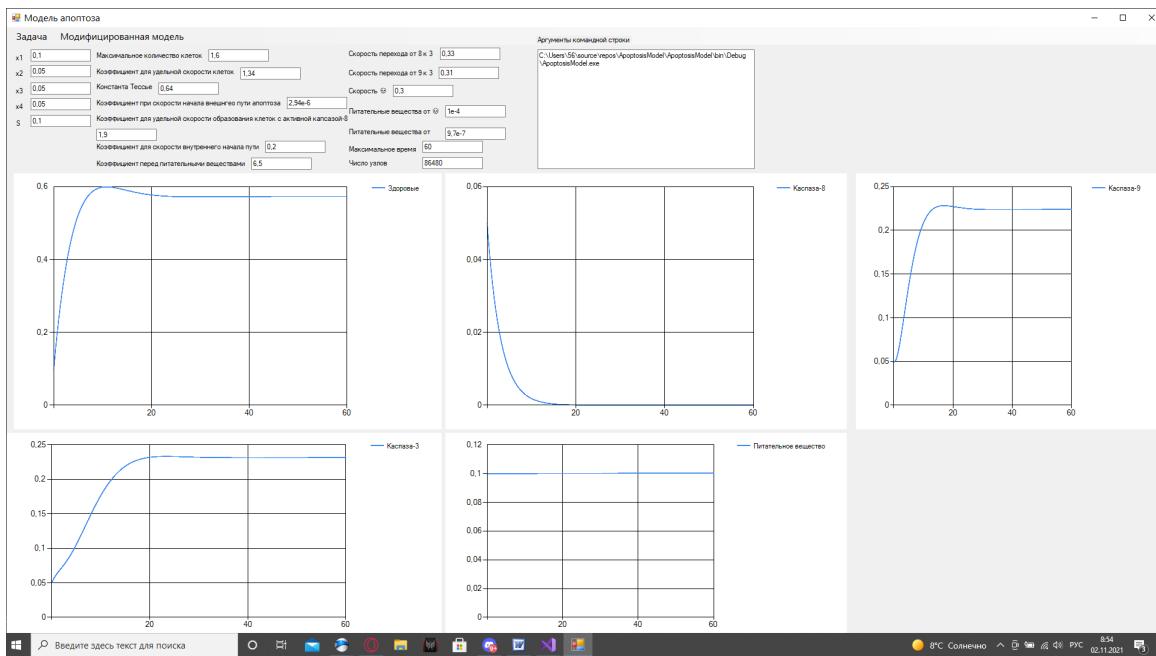


Рисунок 6.1 – Интерфейс окна программы для исследования модели апоптоза

Большую часть окна занимают элементы управления для отображения графиков пяти искомых переменных. В левом верхнем углу задаются начальные данные, наверху в середине – глобальные параметры модели. Отдельно отметим, что в текстовом поле наверху справа отображаются аргументы командной строки. В отличии от большинства оконных приложений, при программировании данного приложения была предусмотрена передача аргументов командной строки, что добавляет большей гибкости в настройки приложения.

### 6.2.5 Линеаризация разностных схем с использованием теоремы о гомоморфизме

В [225] был предоставлен алгоритм линеаризации систем уравнений в частных производных с использованием теоремы о гомоморфизме колец. Идея алгоритма состоит в том, что кольцо линеаризованных функций изоморфно кольцу функциональных матриц специального вида. Однако, при изучении разностных схем вместе с проф. А.А. Злотником авторами было обнаружено, что, вообще говоря, линеаризованная система и линеаризованная разностная схема могут не совпадать. Такое происходит, например, в

случае разнесённых сеток. Таким образом, далеко не всегда мы можем путём простой дискретизации перейти от линеаризованной системы к линеаризованной схеме.

Естественным образом возникает вопрос о расширении возможностей применения теоремы о гомоморфизме уже применительно к разностным схемам. Выясняется, что и в этом случае мы можем переходить от кольца линеаризованных сеточных функций к кольцу изоморфных им сеточно-функциональных матриц специального вида и использовать сеточные операторы сдвига, усреднения и перехода со слоя на слой в последнем.

#### 6.2.6 Некоторые выводы

На языке C# с использованием ООП программно реализована разностная схема для исследования модели апоптоза, которая задаётся системой обыкновенных дифференциальных уравнений. На основе графиков решения системы изучается поведение клеток в различных условиях.

Проведено исследование о возможности применения теоремы о гомоморфизме колец в задаче о линеаризации разностных схем.

## **ЗАКЛЮЧЕНИЕ**

При выполнении проекта получены следующие основные результаты, могут быть сделаны следующие выводы и рекомендации:

- исследованы некоторые классы ПН-средних величин, основанные на применении методов многокритериального принятия решений; исследована устойчивость таких средних к малым изменениям исходных данных;
- в рамках исследования структурированных дихотомических предпочтений в модели с полной информацией получены характеристики доменов через запрещенные подпрофили; определено может ли неполный профиль допускать возможное или необходимое завершение в определенных ограниченных доменах;
- в рамках исследования турнирных способов выбора оптимальных вариантов дано содержательное описание существенного множества, как обобщения двухпартийного множества; описаны важные свойства турнирных решений на основе такого множества; рассмотрена задача агрегирование противоречивых рейтингов научных журналов по менеджменту;
- в рамках исследования манипулируемости в теории коллективного выбора разработана и проанализирована модель, учитывающая одновременно и неполноту информации, и неопределенность относительно действий других избирателей на манипулируемость; разработаны и реализованы эффективные алгоритмы вычисления точных значений сильной и слабой манипулируемости;
- в рамках исследования индексов влияния, учитывающие предпочтения участников по созданию коалиций, найдена аксиоматика 4-х функций интенсивности, основанных на функциях минимума и максимума;
- предложена и исследована теоретико-игровая модель задачи организации переработки твердых бытовых отходов города; установлены прове-

ряемые необходимые и достаточные условия точек равновесия Нэша в такой игре;

- исследована задача оптимальной диверсификации мультивалютных депозитов с неопределенными будущими обменными курсами как проблема минимизации упущенной из-за неопределенности выгоды;
- предложена модификация метода расчёта приоритетов стратегических целей по таблице парных сравнений их значимости; разработано соответствующее программное обеспечение;
- разработан новый программный пакет для таксономического контент-анализа коллекций текстовых документов, реализующий методологию обобщения в таксономии предметной области;
- разработана вычислительно экономичная модификация алгоритма диффузионно-инвариантной паттерн-кластеризации на основе сортировки данных;
- предложена новая модификация дихотомии графов на основе разработанного ранее частотного алгоритма декомпозиции; с помощью модифицированного алгоритма проанализировано функционирования рынка S&P-500 в период 1990–2019 гг.;
- разработан и исследован двухэтапный алгоритм кластеризации тел свидетельств в рамках теории функций доверия; функцию доверия, соответствующую кластеризованному телу свидетельств, можно рассматривать как аппроксимацию исходной функции доверия;
- проанализированы механизмы изменения климата и его текущего воздействия на социально-экономическое развитие Арктики;
- проанализированы среднесрочных стратегий развития Арктического региона в условиях риска;
- проанализированы основные причины нехватки водных ресурсов в странах Северной Африки и механизмы, которые позволили бы смягчить негативные последствия нехватки воды;

- рассмотрена модификация интервального метода оболочечного анализа данных; с помощью этой модификации оценена эффективность карантинных мер в разных странах мира;
- методами регрессионного анализа выявлены основные факторы, определяющие предпочтения респондентов относительно формата онлайн-обучения;
- проанализировано влияние перехода на основанный на внутренних рейтингах (ПВР) подход на котировки банковских акций в рамках исследования банковского регулирования;
- предложена параметризация регулирования достаточности капитала исламских банков как функция от макроэкономических параметров;
- проведено сетевое исследование международного рынка заимствований за 2005–2020 гг. с точки зрения анализа структуры сети, выделения наиболее важных элементов и сообществ, анализа устойчивости сети и её динамического изменения;
- проведено исследование сети цитирования 466 российских экономических журналов РИНЦ за период 2016–2020 гг. с использованием индекса центральности Коупленда, индекс группового влияния ВІ и индекс влияния по ключевым вершинам;
- построена сетевая структура международной торговли продовольствием с учетом неполноты (неточности, неопределенности) и противоречивости данных;
- в рамках исследования моделей распространения COVID-19 выполнен анализ эффективности антковидных мер в странах мира, выявлены группы стран со схожими показателями интенсивности введенных мер, законопослушности, уровня образования, вакцинационной политики;
- проведен анализ изменения экономических показателей Великобритании с изменением интенсивности заболеваемостью COVID-19 и с принимаемыми антковидными мерами;

– разработан программный комплекс моделирования клеточного апоптоза для изучения поведения клеток в различных условиях.

## **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Отчет МНУЛ анализа и выбора решений, Т3-40 за 2020г.
2. Подиновский В.В., Нелюбин А.П. Средние величины: многокритериальный подход // Проблемы управления. – 2020. – N 5. – C. 3 – 16.
3. Подиновский В.В. Идеи и методы теории важности критериев в многокритериальных задачах принятия решений. – М.: Наука, 2019. – 103 с.
4. Eurostat (2020) Real GDP per capita.  
[https://ec.europa.eu/eurostat/databrowser/view/sdg\\_08\\_10/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/sdg_08_10/default/table?lang=en)
5. Подиновский В.В. Количественная важность критериев с непрерывной шкалой первой порядковой метрики // Автоматика и телемеханика. – 2005. – N 9. – C. 129–137.
6. Podinovski V.V. On the use of importance information in MCDA problems with criteria measured on the first ordered metric scale // Journal of multi-criteria decision analysis. – 2009. – N 15. – P. 163 – 174.
7. Подиновский В.В., Нелюбин А.П. Средние величины: многокритериальный подход. II // Проблемы управления. – 2021. – N 2. – C. 33 – 41.
8. Малов С.В. Регрессионный анализ. Теоретические основы и практические рекомендации. – СПб.: Издательство СПбГУ, 2013.
9. Мудров В.И., Кушко В.Л. Метод наименьших модулей. – М.: Либроком, 2013.
10. Зоркальцев В.И. Октаэдрические и евклидовы проекции точки на линейное многообразие // Труды института математики и механики УрО РАН. – 2012. – N 3. – C. 106 – 118.
11. Domshlak C., Hullermeier E., Kaci S., Prade H., 2011. Preferences in AI: An overview// J. Artif. Intell. Res. – 2011. – N 19(4). – P. 1037–1052.
12. Procaccia A.D., Rosenschein J.S., Zohar A. Multi-winner elections: Complexity of manipulation, control and winner-determination// In Proceedings of

the 20th International Joint Conference on Artificial Intelligence, 2007. – P. 1476–1481.

13. Elkind E., Lackner M., Peters D. Structured preferences// In Endriss, U., ed., Trends in Computational Social Choice, 2017. – P. 187–207.

14. Elkind E., Lackner M. Structure in dichotomous preferences// In Proceedings of the 24th International Joint Conference on Artificial Intelligence, 2015. – P. 2019–2025.

15. Yang Y. On the tree representations of dichotomous preferences// In Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019. – P. 644–650.

16. Terzopoulou Z., Karpov A., Obraztsova S. Restricted Domains of Dichotomous Preferences with Possibly Incomplete Information// Proceedings of the 35th AAAI Conference on Artificial Intelligence, 2021. – P. 5726–5733.

17. Brams S.J., Fishburn P.C. Approval Voting// Am. Polit. Sci. Rev. – 1978. – N 72. – P. 831–847.

18. Aziz H., Brill M., Conitzer V., Elkind E., Freeman R., Walsh T. Justified representation in approval-based committee voting// Soc. Choice Welf. – 2017. – N 48(2). – P. 461–485.

19. Bogomolnaia A., Moulin H. Random matching under dichotomous preferences// Econometrica. – 2004. – N 72. – P. 257–279.

20. Bogomolnaia, A., Moulin, H., Stong, R. Collective choice under dichotomous preferences// J. Econ. Theory. – 2005. – N 122(2). – P. 165–184.

21. Duddy C. Fair sharing under dichotomous preferences// Math. Soc. Sci. – 2015. – N 73. – P. 1–5.

22. Aziz H., Harrenstein P., Lang J., Wooldridge M. Boolean hedonic games// Proceedings of the 15th International Conference on the Principles of Knowledge Representation and Reasoning, 2016. – P. 166–175.

23. Tucker A. A structure theorem for the consecutive 1's property// J. Combin. Theory, Series B. – 1972. - N 12(2). – P. 153–162.

24. Blin G., Rizzi R., Vialette S. A faster algorithm for finding minimum Tucker submatrices// Proceedings of the 6th Conference on Computability in Europe, 2010. – P. 69–77.
25. Booth K.S., Lueker G.S. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms // J. Comput. Syst. Sci. – 1976. – N 13. – P. 335–379.
26. Safe M.D. Circularly compatible ones, D-circularity, and proper circular-arc bigraphs// SIAM J. Discrete Math. – 2021. – N 35(2). – P. 707–751.
27. Pe'er I., Pupko T., Shamir, R., Sharan R. Incomplete directed perfect phylogeny// SIAM J. Comput. – 2004. – N 33(3). – P. 590–607.
28. Brewbaker C. A combinatorial interpretation of the poly-Bernoulli numbers and two Fermat analogues// Integers. – 2008. – N 8(1). – A2.
29. Ju H.-K., S. Seo. Enumeration of (0, 1)-matrices avoiding some  $2 \times 2$  matrices// Discrete Math. – 2012. – N 312(16). – P. 2473–2481.
30. Rani M.R., Subashini R., Jagalmohanan M. 2019. Simultaneous consecutive ones submatrix and editing problems: Classical complexity and fixed-parameter tractable results// Theor. Computer Science. – 2020. – N 812. – P. 13–38.
31. Laffond G., Laslier J.-F., Le Breton M. The bipartisan set of a tournament game // Games Economic Behavior. – 1993 – N 5 – P. 182–210.
32. Brandt F., Brill M., Harrenstein P. Extending tournament solutions // Social Choice and Welfare. – 2018 – N 51(2) – P.193–222.
33. Subochev A., Aleskerov F., Pislyakov V. Ranking journals using social choice theory methods: A novel approach in bibliometrics // Journal of Informetrics. – 2018. – N 12(2). – P. 416–429.
34. Brandt F., Brill M., Harrenstein P. Tournament solutions // Handbook of Computational Social Choice / Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.D. (eds). – Cambridge: Cambridge University Press, 2016. – P. 57–84.

35. Brandt F., Harrenstein P. Set-rationalizable choice and self-stability // Journal of Economic Theory. – 2011. – N 146(4). – P. 1721–1731.
36. Aizerman M., Aleskerov F. Theory of Choice. – Amsterdam: North-Holland / Elsevier, 1995 – 324 c.
37. Miller N. A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting // American Journal of Political Science. – 1980 – N 24 – P. 68–96.
38. Aleskerov F., Subochev A. Modeling optimal social choice: Matrix-vector representation of various solution concepts based on majority rule // Journal of Global Optimization. – 2013. – N 56(2). – P. 737–756.
39. Fishburn P.C. Condorcet social choice functions // SIAM Journal on Applied Mathematics. – 1977. – N 33(3) – P. 469–489.
40. Kreweras G. Aggregation of preference orderings // Proceedings of the seminars of Menthon-Saint-Bernard, France (1–27 July 1960) and of Gösing, Austria (3–27 July 1962). – Mathematics and social sciences 1. – 1965. – P. 73–79.
41. Dutta B., Laslier J.-F. Comparison functions and choice correspondences // Social Choice and Welfare. – 1999. – N 16(4). – P. 513–532.
42. Duggan J. General conditions for the existence of maximal elements via the uncovered set // Journal of Mathematical Economics. – 2011. – N 47(6). – P. 755–759.
43. Roth A. Subsolutions and the supercore of cooperative games // Mathematics of Operations Research. – 1976. – N 1(1). – P. 43–49.
44. Aleskerov F., Kurbanov E. Degree of manipulability of social choice procedures // Current Trends in Economics: Theory and Applications / Alkan, A., Aliprantis, Ch. D., Yannelis N.C. (eds). – N.Y.: Springer-Verlag, 1999. – P. 13–27.
45. von Neumann J., Morgenstern O. Theory of Games and Economic Behavior – Princeton: Princeton University Press, 1944. – 674 c.
46. Ward B. Majority rule and allocation // Journal of Conflict Resolution. – 1961. – N 5 – P. 379–389.

47. Duggan J. A systematic approach to the construction of nonempty choice sets // Social Choice and Welfare. – 2007. – N 28. – P. 491–506.
48. Laslier J.-F. Aggregation of preferences with a variable set of alternatives // Social Choice and Welfare. – 2000. – N 17(2). – P. 269–282.
49. Brandt F. Computing the minimal covering set // Mathematical Social Sciences. – 2008. – N 56(2). – P. 254–268.
50. Reijngoud A., Endriss U. Voter Response to Iterated Poll Information // In Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems AAMAS-2012. – 2012.
51. Veselova Y. A. Does Incomplete Information Reduce Manipulability?// Group Decision and Negotiation. – 2020. – N 29(3). – P. 523–548.
52. Gibbard A. Manipulation of voting schemes// Econometrica. – 1973. – N 41. – P. 587–601.
53. Satterthwaite M. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions// Journal of Economic Theory. – 1975. – N 10. – P. 187–217.
54. Nitzan S. The vulnerability of point-voting schemes to preference variation and strategic manipulation// Public Choice. – 1985. – N 47. – P. 349 – 370.
55. Kelly J. Almost all social choice rules are highly manipulable, but few aren't// Social Choice and Welfare. – 1993. – N 10. – P. 161–175.
56. Aleskerov F., Karabekyan D., Sanver R., Yakuba V. On manipulability of positional voting rules // SERIES: Journal of the Spanish Economic Association. – 2011. – N (4). – P. 431 – 446.
57. Aleskerov F., Ivanov A., Karabekyan D., Yakuba V. Manipulability of Aggregation Procedures in Impartial Anonymous Culture// Procedia Computer Science. – 2015. – N 55, – P. 1250 – 1257.

58. Иванов А.А. Эффективные вычислительные схемы расчета манипулируемости процедур агрегирования// Информационные технологии и вычислительные системы. – 2020. – N 2. – C. 38-50.
59. Aleskerov F.T., Karabekyan D., Ivanov A., Yakuba V.I. Further Results on the Manipulability of Social Choice Rules—A Comparison of Standard and Favardin–Lepelley Types of Individual Manipulation// In: Evaluating Voting Systems with Probability Models. Cham: – Springer, 2021. – P. 231–249.
60. Aleskerov F. Power Indices Taking into Account Agent's Preferences // Mathematics and Democracy. Springer. – 2006. – 18c.
61. Алескеров Ф.Т., Очур О.А. Обобщенные индексы влияния Шепли — Оуэна и распределение влияния в Государственной думе III созыва// WP7/2007/03. М.: ГУ ВШЭ. – 2007. – 32с.
62. José M. Alonso-Mejide, Julián Costa, Ignacio García-Jurado Null, Nullifying, and Necessary Agents: Parallel Characterizations of the Banzaf and Shapley Values// Journal of Optimization Theory and Applications. Springer. – 2018. – N 180. – P. 1027–1035.
63. Belenky A.S. Finding an optimal strategy of incorporating renewable sources of energy and electricity storing systems in a regional electrical grid// Energy Systems. – 2015. – N 6(2). – P.291–308.
64. Belenky A.S. A game-theoretic approach to optimizing the scale of incorporating renewable sources of energy and electricity storing systems in a regional electrical grid// Energy Systems. – 2015. – N 6(3). – P.389–415.
65. Беленький А.С. Задача оптимизации переработки твердых бытовых отходов (ТБО) крупных городов// Раздел 1.2. отчета о научно-исследовательской работе “Анализ, выбор, и принятие решений в социально-экономических, политической и финансовой сферах: новые модели, методы и алгоритмы” (заключительный), Шифр ТЗ-72. – 2017. – С. 34–43.
66. Belenky A.S., Kornhauser A.L. Estimating the needed volume of investment in developing a chain of charging stations for electric vehicles along a

highway// International Journal of Public Administration. – 2019. – N 42(15-16). – P. 1256–1274.

67. Belenky A.S. A 2-person game on a polyhedral set of connected strategies//Computers and Mathematics with Applications. – 1997. – N 33(6). – P. 99–125.

68. Eun Kyung Kim, Legal issues of autonomous vehicle insurance// Commercial Law Review. – 2020. – N 38(4). – P. 111–150

69. Маметов Р. Электромобиль по праву// Энергосовет, 2012, N 3(14),

70. Belenky A.S., Kornhauser A.L. Optimizing the allocation of a chain of charging stations for electric vehicles in a city// To be submitted to International Journal of Public Administration in January 2022.

71. Belenky A.S. A game-theoretic approach to modeling an environmental/energy problem of utilizing city's municipal solid waste// To be submitted to Journal of Public Administration Research and Theory in December 2021-January 2022.

72. Belenky A.S. Analyzing the potential of a firm: an operations research approach// Mathematical and Computer Modelling. – 2002. – N 35. – P. 1405–1424

73. Belenky A.S. On finding Nash equilibrium points in a three-person non-cooperative game, associated with utilizing city's municipal solid waste// To be submitted to Journal of Public Administration Research and Theory in December 2021-January 2022.

74. Savage L.Y. The Foundation of Statistics. – New York: Willey, 1954.

75. Zhukovskiy V.I., Molostvov V.S., Topchishvili A.L. Multicurrency Deposit Diversification – Three Possible Approaches to Risk Accounting// International Journal of Operations and Quantitative Management. –2014. –V.20.– P.1–14.

76. Жуковский В.И., Молоствов В.С. Об оптимальности по Парето в кооперативных дифференциальных играх с многозначными целевыми

функционалами// Математические методы в теории систем. –1980. –С. 96–108.

77. Жуковский В.И., Молостков В.С. Многокритериальное принятие решений в условиях неопределенности. –М: МНИИПУ, 1988.

78. Жуковский В.И., Молостков В.С. Многокритериальная оптимизация систем в условиях неполной информации. –М., МНИИПУ, 1990.

79. Zhukovskiy V.I., Salukvadze M.E. The Vector-valued Maximin. – New York: Academic Press, 1994.

80. Zhukovskiy V.I., Salukvadze M.E. Sufficient conditions in vector-valued maximin problems// Optimization Theory and Applications. – 1996.–N3.–P.523–534.

81. Salukvadze M.E., Zhukovskiy V.I. Sufficient Conditions in the Vector-Valued Maximin Problems// Multiple Criteria Decision Making. Lecture Notes in Economics and Mathematical Systems. – 1997. – V448. – P. 74-82.

82. Molostvov V.S. Multiple-criteria optimization under uncertainty: concepts of optimality and sufficient conditions// Theory and Practice of Multiple Criteria Decision Making. North-Holland. – 1983. – P. 91–105.

83. Molostvov V.S. Savage’s principle for non-cooperative games under uncertainty// Proceedings of the international scientific conference “Problems of control and power engineering”. – Tbilisi. –2004.–P.38 – 39.

84. Molostvov V.S. Multiple criteria optimization for stochastic systems with uncertain parameters// Model Assisted Statistics and Applications. – 2011. – V3. – P. 231-237.

85. Salukvadze M., Jibladze N., Maisuradze V., Topchishvili A. Methods and algorithms for solving multicriteria problems – classification approach// Model Assisted Statistics and Applications. – 2007.–V.2.–P. 239–251.

86. Wald A. Contribution to the theory of statistical estimation and testing hypothesis// Annals Math. Statist. –1939. – N10.–P. 299-326.

87. A Guide to the Project Management Body of Knowledge// PMBOK Guide 5th edition. – 2013. – Pennsylvania, Project Management Institute. – 589p.
88. ГОСТ Р 54870-2011. Проектный менеджмент. Требования к управлению портфелем проектов. – М.: Стандарт-инфо. – 13с.
89. Сертификация специалистов по управлению проектами по модели IPMA-СОВНЕТ// [Электронный ресурс] – 2020. – URL: <http://www.sovnet.ru/specialists/> (дата обращения 8.12.2020).
90. Саати Т.Л. Принятие решений при зависимостях и обратных связях. Аналитические сети. – М.: Изд. Стереотип. – 2018. – 360с.
91. SuperDecisions software// 2020. – [Электронный ресурс] URL: <http://www.superdecisions.com/>. (дата обращения 8.12.2020).
92. Кузнецова Е.В. Управление портфелем проектов как инструмент реализации корпоративной стратегии. – М.: Юрайт. – 2016. – С. 188–201.
93. Whitaker R. Criticisms of the Analytic Hierarchy Process: Why they often make no sense// Mathematical and Computer Modelling. – 2007. – N 46. – P. 948–961.
94. Демидович Б.П., Марон И.А. Основы вычислительной математики. 5-е изд. – СПб.: Издательство «Лань». – 2006.
95. Нортон Д.П., Каплан Р. Сбалансированная система показателей. От стратегии к действию. – М.: Олимп-Бизнес. – 2017. – 320с.
96. Марон А.И., Марон М.А. Высокоточный метод расчёта приоритетов стратегических целей// В кн.: Анализ, моделирование, управление, развитие социально-экономических систем: сборник научных трудов XIV Всеросс. с межд. участием школы-симпозиума АМУР-2020, Симферополь-Судак, 14-27 сентября 2020. – ИП Корниенко А.А. – 2020. – С. 225 – 228.
97. Krippendorff K. Content analysis// E. Barnouw, G. Gerbner, W. Schramm, T. L. Worth, L. Gross (Eds.), International encyclopedia of communication. – Vol. 1. – New York, NY. – Oxford University Press. – 1989. – P. 403–407.

98. Blei D. Probabilistic topic models // Communications of the ACM. – N 55(4). – 2012. – P. 77–84.
99. Bohr J., Dunlap R.E. Key Topics in environmental sociology, 1990–2014: results from a computational text analysis // Environmental Sociology. – 2018. – N 3-4(2). – P. 181–95.
100. Frolov D., Nascimento S., Fenner T., Mirkin B. Parsimonious Generalization of Fuzzy Thematic Sets in Taxonomies Applied to the Analysis of Tendencies of Research in Data Science // Information Sciences. 2020. – N 512. – P. 595–615.
101. Mirkin B., Frolov D., Vlasov A., Nascimento S., Fenner T. A Hybrid Approach to Interpretable Analysis of Research Paper Collections // Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics 2020 Jun 30. – ACM. – 2020. – P. 184–189.
102. Mirkin B., Frolov D., Vlasov A., Nascimento S., Fenner T. A Hybrid Approach to the Analysis of a Collection of Research Papers// InInternational Conference on Intelligent Data Engineering and Automated Learning 2020 Nov 4, Springer, LNCS. – 2020. – P. 423–433.
103. Chernyak E. An approach to the problem of annotation of research publications // Proceedings of the eighth ACM international conference on web search and data mining, ACM. – 2015. – P. 429–434.
104. Mirkin B., Nascimento S. Additive spectral method for fuzzy cluster analysis of similarity data including community structure and affinity matrices // Information Sciences. – 2012. – N 183(1). – P. 16–34.
105. Frolov D., Nascimento S., Fenner T., Taran Z., Mirkin B. Computational Generalization in Taxonomies Applied to: (1) Analyze Tendencies of Research and (2) Extend User Audiences // IDEAL 2019 Proceedings. Springer. – 2020.
106. The 2012 ACM Computing Classification System// [Online]. Available: <http://www.acm.org/about/class/2012> (Accessed 2018, 30 April).

107. Frolov D., Mirkin B., Nascimento S., Fenner T. Finding an appropriate generalization for a fuzzy thematic set in taxonomy // Working paper WP7/2019/01, Moscow, Higher School of Economics Publ. House. – 2018. – 60 P.
108. Frolov D, Taran Z, Mirkin B. Using the IAB Contents Taxonomy and Optimal Lifting for Efficient Audience Extension // International Conference on Intelligent and Fuzzy Systems 2021. – Springer, Cham. – P. 596–603.
109. Frolov D, Mirkin B. GOT: Generalization over Taxonomies, a Software Toolkit for Content Analysis with Taxonomies // World Conference on Information Systems and Technologies 2021. – Springer, Cham. – P. 513–521.
110. Akhremenko A., Myachin A. The Study of Trajectories of the Development of State Capacity Using Ordinal-Invariant Pattern Clustering and Hierarchical Cluster Analysis// in: 8th International Conference on Computers Communications and Control (ICCCC) 2016. – Oradea: Agora University. – 2020.
111. Aleskerov F., Ersel H., Yolalan R. Multicriterial Ranking Approach for Evaluating Bank Branch Performance// International Journal of Information Technology and Decision Making. – 2004. – N 3(2). – P. 321–335.
112. Aleskerov F., Nurmi H. A Method for Finding Patterns of Party Support and Electoral Change: An Analysis of British Genera 1 and Finnish Municipal Elections// Mathematical and Computer Modelling. – 2008. – P. 1225–1253.
113. Myachin A. Pattern Analysis in Parallel Coordinates Based on Pairwise Comparison of Parameters // Automation and Remote Control. – 2019. – N 80(1). – P. 112–123.
114. Мячин А.Л. Анализ паттернов: диффузионно-инвариантная паттерн-кластеризация // Проблемы управления. – 2016. – N 4. – C. 2–9.
115. Shawe-Taylor J., Cristianini N. Kernel methods for pattern analysis. – Cambridge university press. – 2004.
116. Niemann H. Pattern analysis and understanding// Springer Science & Business Media. – 2013. – 4.

117. Inselberg A. The plane with parallel coordinates// The Visual Computer. – 1985. – N 1-2. – P. 69–91.
118. Rubchinsky A. Family of Graph Decompositions and Its Applications to Data Analysis// Working paper WP7/2016/09 – Moscow: Higher School of Economics Publ. House, 2016. – (Series WP7 “Mathematical methods for decision making in economics, business and politics”). – 60 p.
119. Rubchinsky A. A New Approach to Network Decomposition Problems// In: Kalyagin V., Nikolaev A., Pardalos P., Prokopyev O. (eds) Models, Algorithms, and Technologies for Network Analysis. NET 2016. Springer Proceedings in Mathematics & Statistics. –2017. – N 197. – Springer, Cham. – P. 127–152.
120. Rubchinsky A. Graph Dichotomy Algorithm and Its Applications to Analysis of Stocks Market// In: Kalyagin V., Pardalos P., Prokopyev O., Utkina I. (eds) Computational Aspects and Applications in Large-Scale Networks. NET 2017. Springer Proceedings in Mathematics & Statistics. – 2018. – N 247. – Springer, Cham. – P. 75–111.
121. Dempster A.P. Upper and lower probabilities induced by a multivalued mapping // Annals of Mathematical Statistics. – 1967. – 38. – P. 325–339.
122. Shafer G. A Mathematical theory of evidence. – Princeton, N.J.: Princeton University Press. – 1976.
123. Ester M., Kriegel H-P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise// Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). – 1996. – AAAI Press. – P.226–231.
124. Denœux T. Inner and outer approximation of belief structures using a hierarchical clustering approach// Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems. – 2001. – N 9(4). – P.437–460.
125. Tedjini T., Afifi S., Pichon F., Lefèvre E. An Extension of Specificity-Based Approximations to Other Belief Function Relations// In: Denœux T.,

Lefèvre E., Liu Z., Pichon F. (eds) Belief Functions: Theory and Applications. BELIEF 2021. Lecture Notes in Computer Science. – 2021. – N 12915. – Springer, Cham. – P. 224–233.

126. Osswald C., Martin A. Understanding the large family of Dempster-Shafer theory's fusion operators – a decision-based measure// Int. Conf. on Information Fusion. – Florence, Italy. – 2006.

127. Klir G.J. Uncertainty and Information: Foundations of Generalized Information Theory. – John Wiley, Hoboken, NJ. – 2005. – 499 pp.

128. Yager R.R. Entropy and Specificity in a Mathematical Theory of Evidence// International Journal of General Systems. N 1983. – N 9(4). – P.249–260.

129. Yager R.R. On the Dempster-Shafer Framework and New Combination Rules // Information Sciences. – 1987. – N 41. – P. 93–138

130. Jousselme A.-L., Grenier D., Bossé É. A new distance between two bodies of evidence// Information Fusion. – 2001. – N 2. – P.91–101.

131. Bronevich A., Rozenberg I.N. The Kantorovich Problem and Wasserstein Metric in the Theory of Belief Functions// in: Belief Functions: Theory and Applications 5th International Conference, BELIEF 2018, Compiègne, France, September 17-21, 2018. – N 11069. – Springer. – 2018. – P. 31–38.

132. Lepskiy A., Smolev V. Application of non-additive measures and integrals for analysis of the importance of party positions for voting// in: H. Bustince, G. Chen, R. Mesiar, G. Pasi (Eds.), Atlantis Studies in Uncertainty Modelling. – 2019. – P. 321–327.

133. Bronevich A., Lepskiy, A. Measures of conflict, basic axioms and their application to the clusterization of a body of evidence // Fuzzy Sets and Systems. – 2021. (online: <https://doi.org/10.1016/j.fss.2021.04.016>)

134. Harvey, F., Rankin, J. What is the European Green Deal and will it really cost €1tn?// The Guardian. [Электронный ресурс]. – 2020. URL: <https://www.theguardian.com/world/2020/mar/09/what-is-the-european-green-deal-and-will-it-really-cost-1tn> (Дата обращения: 04.11.2021)

135. Statistics Greenland// URL: <https://stat.gl/default.asp?lang=da> (Дата обращения: 04.11.2021)

136. Комитет Санкт-Петербурга по делам Арктики. Комплексное исследование регионов Арктической зоны Российской Федерации// Основные направления сотрудничества Санкт-Петербурга в Арктике. – 2020.

137. IPCC, 2021: Climate Change 2021: The Physical Science Basis// Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press.

138. Гладильщикова А. А., Семенов С. М. Межправительственная группа экспертов по изменению климата (МГЭИК): цикл шестого оценочного доклада // Фундаментальная и прикладная климатология. – 2017. – N 2. – C. 13–25.

139. Council A. COVID-19 in the Arctic: briefing document for senior arctic officials. – 2020.

140. Arctic Centre. University of Lapland. Basic information about the Arctic // The website of Arctic Centre [Электронный ресурс URL: <https://www.arcticcentre.org/EN/arcticregion> (дата обращения 28.04.2021)]

141. AHDR II. Arctic Human Development Report// 2015 [Электронный ресурс]. URL: <https://oaarchive.arctic-council.org/handle/11374/51> / (дата обращения 28.04.2021)

142. Wilson. Opportunities and challenges for arctic oil and gas development // Eurasia GroupReport for Wilson Center. – 2016. [Электронный ресурс]. URL: [https://www.wilsoncenter.org/sites/default/files/Artic%20Report\\_F2.pdf](https://www.wilsoncenter.org/sites/default/files/Artic%20Report_F2.pdf) (дата обращения 28.04.2021)

143. Helloscience. Arctic Opportunity Explorers// 2021 [Электронный ресурс]. URL: <https://helloscience.io/live-lab/arctic-opportunity-explorers/> (дата обращения 28.04.2021)
144. Burkhard S., Torsten Wulf. Scenario-based Strategic Planning: Developing Strategies in an Uncertain World // Springer Science & Business Media. – 2013.
145. The Global Risk Report 2020. [Online]. – 2020. – Available: [https://www3.weforum.org/docs/WEF\\_Global\\_Risk\\_Report\\_2020.pdf](https://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf) (Data of access 05.12.2020)
146. Bell M. If you want to make progress on all the major global challenges, start with water // World Economic Forum – 2021.
147. The Sustainable Development Goals Report 2019. [Online]. – 2019. – Available: <https://unstats.un.org/sdgs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf#page=36> (Data of access 17.12.2020)
148. Schilling J, Hertig E, Tramblay Y, & Scheffran, J. Climate change vulnerability, water resources and social implications in North Africa // Regional Environmental Change. – 2020. – N 20(15). – 12 P.
149. Ngounou B. North Africa: Freshwater availability has fallen by 30% according to FAO // Afrik 21. [Online]. – 2020. – Available: <https://www.afrik21.africa/en/north-africa-freshwater-availability-has-fallen-by-30-according-to-fao/> (Data of access 07.01.2021)
150. Brown A, Matlock, M D. A Review of Water Scarcity Indices and Methodologies // The Sustainability Consortium. – 2011. – 21 P.
151. Campos B C, Ren Y, & Loy J P. (2020). Scarce Water Resources and Cereal Import Dependency: The Role of Integrated Water Resources Management // Water. – 2020. – N 12(6). – 14 P.
152. Breisinger C. et al. Food Security and Economic Development in the Middle East and North Africa // IFPRI Discussion Paper 00985, International Food Policy Research Institute (IFPRI). – 2010.

153. Roth D, Warner J. Virtual water: Virtuous impact? The unsteady state of virtual water // Agriculture and Human Values. – 2007. – N 25(2). – P. 257–270.

154. Allan J. Virtual Water - the Water, Food, and Trade Nexus. Useful Concept or Misleading Metaphor? // Water International. – 2003. – N 28(1). – P. 106–113.

155. Warner J F. Virtual water - virtual benefits? Scarcity, distribution, security and conflict reconsidered // In A. Y. Hoekstra (Ed.), Virtual Water Trade. Proceedings of the International Expert Meeting on Virtual Water Trade. – 2003. – P. 125–135.

156. Hoekstra A. Y. The relation between international trade and freshwater scarcity // World Trade Organization, Economic Research and Statistics Division. – 2010. – 25 P.

157. INRAE Institutionnel. Addressing agricultural import dependence in the Middle East – North Africa region through the year 2050 // Short summary of the study. [Online]. – 2015. – Available: <https://www.inrae.fr/sites/default/files/pdf/addressing-agricultural-import-dependence-in-the-middle-east-north-africa-region-through-to-the-year-2050-doc.pdf> (Data of access 18.04.2021)

158. Suárez-Varela M. Review of the use of economic instruments in water management in Egypt, Morocco and Tunisia. [Online]. – 2018. – Available: [http://amsacta.unibo.it/6312/1/D5.1%20-%20UPM\\_Nov2019.pdf](http://amsacta.unibo.it/6312/1/D5.1%20-%20UPM_Nov2019.pdf) (Data of access 01.05.2021)

159. Ministry of Water Resources and Irrigation. Water Scarcity in Egypt: The Urgent Need for Regional Cooperation among the Nile Basin Countries // MWRI. [Online]. – 2014. – Available: [https://globalhealthnewswire.com/wp-content/uploads/2016/09/Egypt%20Water%20Resources%20Paper\\_2014.pdf](https://globalhealthnewswire.com/wp-content/uploads/2016/09/Egypt%20Water%20Resources%20Paper_2014.pdf) (Data of access 01.05.2021)

160. Brika B. Water Resources and Desalination in Libya: A Review. // Proceedings. – 2018. – N 2(11). – 586 P.

161. Abufayed A, El-Ghuel M. Desalination process applications in Libya // Desalination. – 2001. – N 138(1–3). – P. 47–53.
162. Bermingham F., Wang O. China's Economy Shrunk For The First Time Since 1976 In First Quarter // South China Morning Post. – 2020.
163. Charnes A., Cooper W.W., Rhodes E. Measuring the efficiency of decisionmaking units // Eur. J. Operat. Res. – 1978. – N 2(6). – P. 429–444.
164. Aleskerov F. T., Demin S. DEA for the Assessment of Regions' Ability to Cope with Disasters // Dynamics of Disasters. Impact, Risk, Resilience, and Solutions. – 2021. – N 2. – P. 31–37.
165. Hale T., Angrist N., Cameron-Blake E., Hallas L., Kira B., Majumdar S., Petherick A., Phillips T., Tatlow H., Webster S. Oxford COVID-19 Government Response Tracker, Blavatnik School of Government. – 2020.
166. Allam Z. The First 50 days of COVID-19: A Detailed Chronological Timeline and Extensive Review of Literature Documenting the Pandemic// Surveying the Covid-19 Pandemic and its Implications. – 2020 – P. 1–7.
167. Horspool A., Lange C. Applying the scholarship of teaching and learning: student perceptions, behaviours and success online and face-to-face // Assessment & Evaluation in Higher Education. – 2012. – N 37(1). – P. 73–88.
168. Daniels C., Feather S.R. Student Perceptions of Online Learning: A Comparison of Two Different Populations// URL: <http://proc.isecon.org/2002/223d/ISECON.2002.Daniels.pdf>.
169. Sheppard P., Smith R. What students want: using a choice modelling approach to estimate student demand // Journal of Higher Education Policy and Management. – 2016. – N 38(2). – P. 140–149.
170. Semenova T. Participation in Massive Open Online Courses: The Effect of Learner Motivation and Engagement on Achievement// Препринт WP BRP 37/EDU/2016, Высшая школа экономики. – 2016. – N 37.
171. Банк России. Переход системно значимых банков на внутренние рейтинги: консультативный доклад Банка России// 2021. [В Интернете]

Available at:

[http://cbr.ru/Content/Document/File/123098/Consultation\\_Paper\\_02062021.pdf](http://cbr.ru/Content/Document/File/123098/Consultation_Paper_02062021.pdf)

172. World Bank. Pandemic, Recession: The Global Economy in Crisis// – 2020 [Online] Available at: <https://www.worldbank.org/en/publication/global-economic-prospects>

173. FRS, 2020. Federal Reserve Actions to Support the Flow of Credit to Households and Businesses// 2020. [Online] Available at: <https://www.federalreserve.gov/newsevents/pressreleases/monetary20200315b.htm> [Accessed 10 February 2021].

174. BKAM, 2020. Monetary Policy Report No. 55. [Online] Available at: <http://www.bkam.ma/en/content/download/707707/8204070/RPM-Q2-2020-ENG.pdf>

175. Eckert F., Mikosch H., Stotz M. The corona crisis and corporate bankruptcies: Evidence from Switzerland// 2020. [Online] Available at: <https://voxeu.org/print/66172>

176. Hernández de Cos P., Evaluating the effectiveness of Basel III during Covid-19 and beyond// 2021. [Online] Available at: <https://www.bis.org/speeches/sp210420.pdf>

177. РГ, 2020. Мурманское морское пароходство признано банкротом// 2020 [В Интернете] Available at: <https://rg.ru/2020/10/08/reg-szfo/murmanskoe-morskoe-parohodstvo-priznano-bankrotom.html>

178. Goldfine J. Here are all the companies that went bankrupt during the COVID pandemic. More than ten major companies have already declared bankruptcy in 2021// 2021. [Online] Available at: <https://www.businessofbusiness.com/articles/companies-bankrupt-covid-pandemic-2021/>

179. Rystad Energy. After the costliest Chapter 11 year in history, 2021 bankruptcy numbers will normalize if WTI keeps up// 2021 [Online] Available at: <https://www.hellenicshippingnews.com/after-the-costliest-chapter-11-year-in-history-2021-bankruptcy-numbers-will-normalize-if-wti-keeps-up/>

180. Ermolova M. Shadow Banking Under Capital Regulation and The Internal Ratings-Based Approach// Chengdu, China, ITQM 2020&2021 Conference. – 2021. (in print).
181. Ordoñez G. Sustainable Shadow Banking. American Economic Journal// Macroeconomics. – 2018. – N 10(1). – P. 33–56.
182. Merika A., Merikas A., Penikas H., Surkov M. The Basel II Internal Ratings Based (IRB) Model and the Transition Impact on the Listed Greek Banks// Journal of Economic Asymmetries. – 2020. (в печати).
183. Penikas H., Savenko D., Stefanenko V. Evaluating the 2013 Islamic Banking Regulation Capital Reform Implication for the Valuation of the Islamic Banks// Conference proceedings “Rethinking Cities. Innovations and Resilience Post COVID-19. International Multidisciplinary Conference Under the Patronage of His Excellency Professor Riyad Hamzah President of University of Bahrain”, 2021. Bahrain.
184. Пеникас Г.И., Стефаненко В.Ю. Долговая нагрузка страны как определяющий фактор в регулировании достаточности капитала исламских банков// Новая экономическая ассоциация. – 2020.
185. Baldwin K., Alhalboni M., Helmi M.H. A structural model of "alpha" for the capital adequacy ratios of Islamic banks// Journal of International Financial Markets, Institutions and Money. –2019. –N 60. – P. 267–283.
186. Nechitaylo V., Penikas H. Benchmarking Banking Regulation Regimes: Application for the CBDC// Model Assisted Statistics and Applications. – 2021.
187. Penikas H. Money Multiplier Under Basel Capital Ratio Regulation: Implications for Counter-COVID-19 Stimulus// Journal of Sustainable Finance & Investment. – 2021. doi 10.1080/20430795.2021.1945348.
188. Алескеров Ф.Т., Бадгаева Д.Н., Писляков В.В., Стерлигов И.А., Швыдун С.В. Значимость основных российских и международных экономических журналов: сетевой анализ // Журнал Новой экономической ассоциации. – 2016. – N 2(30). – C. 193–205.

189. Балацкий Е.В., Екимова Н.А. Консенсусный рейтинг российских экономических журналов: идеология и опыт составления // Журнал институциональных исследований (Journal of Institutional Studies). – 2018. – N 10(1). – С. 93–106.
190. Рубинштейн А.Я. Ранжирование российских экономических журналов: научный метод или «игра в цифирь» // Журнал Новой экономической ассоциации. – 2016. – N 2(30). – С. 162–175.
191. Marina T., Sterligov I. Prevalence of potentially predatory publishing in Scopus on the country level // Scientometrics. – 2021. – N 126(6). – P. 5019–5077.
192. Bank for International Settlements. Consolidated Banking Statistics. [Электронный ресурс]. – 2020. – URL: <https://stats.bis.org/> (дата обращения 01.11.2021).
193. Brin S., Page L. The anatomy of a largescale hypertextual Web search engine// Comput. Netw. – 1998. – N 30. – P. 107–117.
194. Aleskerov F., Meshcheryakova N, Shvydun S. Centrality measures in networks based on nodes attributes, long-range interactions and group influence//Cornell Univercity. – 2016. – P.1–44.
195. Aleskerov F.T., Meshcheryakova N.G., Shvydun S.V. Power in Network Structures// Models, Algorithms, and Technologies for Network Analysis. Springer Proceedings in Mathematics & Statistics. – 2017. – N 197. – P. 79–85.
196. Bank of International Settlements. The treatment of large exposures in the Basel capital standards - Executive Summary. [Электронный ресурс]. – 2013. – URL: <https://www.bis.org/fsi/fsisummaries/largeexpos.pdf> (дата обращения 01.11.2021).
197. Rosvall M., Bergstrom, C.T. Maps of random walks on complex networks reveal community structure // Proc. Nat. Acad. Sci. USA. – 2008. – N 105. – P. 1118–1123.

198. Aleskerov F., Yakuba V. Matrix-vector approach to construct generalized centrality indices in networks. [Электронный ресурс]. – 2020. – URL: <https://ssrn.com/abstract=3597948>.
199. About WITS. [Электронный ресурс]. – 2021. – URL: [https://wits.worldbank.org/about\\_wits.html](https://wits.worldbank.org/about_wits.html) (дата обращения 05.11.2021).
200. WTO Data. [Электронный ресурс]. – 2021. – URL: <https://data.wto.org/> (дата обращения 05.11.2021).
201. OECD Statistics. [Электронный ресурс]. – 2021. – URL: <https://stats.oecd.org/> (дата обращения 05.11.2021).
202. CEPII – BACI – Presentation. [Электронный ресурс]. – 2021. – URL: [http://www.cepii.fr/CEPII/en/bdd\\_modele/presentation.asp?id=37](http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=37) (дата обращения 05.11.2021).
203. Wits User Manual [Электронный ресурс]. – 2011. – URL: [http://wits.worldbank.org/data/public/WITS\\_User\\_Manual.pdf](http://wits.worldbank.org/data/public/WITS_User_Manual.pdf) (дата обращения 05.11.2021).
204. Bilateral Asymmetries (Asymmetry). [Электронный ресурс]. – 2019. – URL: <https://unstats.un.org/unsd/tradekb/Knowledgebase/50657/> Bilateral-asymmetries (дата обращения 05.11.2021).
205. United Nations Statistics Division – International Merchandise Trade Statistics. [Электронный ресурс]. – 2008. – URL: [https://unstats.un.org/unsd/tradereport/introduction\\_MM.asp](https://unstats.un.org/unsd/tradereport/introduction_MM.asp) (дата обращения 05.11.2021).
206. Coronavirus (COVID-19) Cases [Электронный ресурс]. – 2021. – URL: <https://ourworldindata.org/covid-cases> (дата обращения 31.10.2021).
207. European statistical recovery dashboard [Электронный ресурс]. – 2021. – URL: <https://ec.europa.eu/eurostat/cache/recovery-dashboard/> (дата обращения 31.10.2021).
208. The Oxford COVID-19 Government Response Tracker [Электронный ресурс]. – 2021. – URL:

<https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker> (дата обращения 31.10.2021).

209. The economic context of the United Kingdom [Электронный ресурс]. – 2021. – URL: <https://www.nordeatrade.com/en/explore-new-market/united-kingdom/economical-context> (дата обращения: 14.03.2021).

210. GDP first quarterly estimate, UK: January to March 2020 [Электронный ресурс]. – 2020. – URL: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/bulletins/gdpfirstquarterlyestimateuk/januarytomarch2020> (дата обращения: 11.11.2020).

211. GDP, UK regions and countries: April to June 2020 [Электронный ресурс]. – 2020. – URL: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/bulletins/gdpuikregionsandcountries/apriltojune2020> (дата обращения: 01.05.2021).

212. GDP first quarterly estimate, UK: October to December 2020 [Электронный ресурс]. – 2020. – URL: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/bulletins/gdpfirstquarterlyestimateuk/octobertodecember2020> (дата обращения: 24.02.2021).

213. UK economy poised to recover after Covid-19 second wave [Электронный ресурс]. – 2021. – URL: <https://www.theguardian.com/business/2021/mar/31/uk-economy-poised-to-recover-after-covid-19-second-wave> (дата обращения: 10.04.2021).

214. UK trade: December 2020 [Электронный ресурс]. – 2020. – URL: <https://www.ons.gov.uk/economy/nationalaccounts/balanceofpayments/bulletins/uktrade/decelber2020#:~:text=The%20total%20trade%20deficit%20for,increased%20by%20%C2%A30.4%20billion>. (дата обращения: 17.04.2021).

215. Coronavirus: UK airports predict £4bn in lost revenue this year [Электронный ресурс]. – 2020. – URL: <https://yaho.it/3bwPRX4> (дата обращения: 16.12.2020).

216. ‘Worst months in the history of commercial aviation’: UK airports set to lose over \$5 BILLION in revenue this year [Электронный ресурс]. – 2020. –

URL: <https://www.rt.com/business/495502-uk-airports-losses-coronavirus/> (дата обращения: 16.12.2020).

217. Глава Хитроу предупредил о возможном банкротстве аэропортов и авиакомпаний [Электронный ресурс]. – 2020. – URL: [https://aif.ru/money/company/direktor hitrou zayavil o skorom bankrotstve aeroportov i aviakompaniy](https://aif.ru/money/company/direktor_hitrou_zayavil_o_skorom_bankrotstve_aeroportov_i_aviakompaniy) (дата обращения: 17.12.2020).

218. U.K. Universities Face Financial Loss As Brexit Hits E.U. Student Numbers [Электронный ресурс]. – 2021. – URL: <https://www.forbes.com/sites/nickmorrison/2021/02/09/uk-universities-face-financial-loss-as-brexit-hits-eu-student-numbers/?sh=21c66687e2a8> (дата обращения: 09.03.2021).

219. Coronavirus and the impact on the UK travel and tourism industry [Электронный ресурс]. – 2021. – URL: <https://www.ons.gov.uk/businessindustryandtrade/tourismindustry/articles/coronavirusandtheimpactontheuktravelandtourismindustry/2021-02-15> (дата обращения: 15.03.2021).

220. £1.57 billion investment to protect Britain's world-class cultural, arts and heritage institutions [Электронный ресурс]. – 2020. – URL: <https://www.gov.uk/government/news/157-billion-investment-to-protect-britains-world-class-cultural-arts-and-heritage-institutions> (дата обращения: 01.03.2021).

221. Nedelcu A.M., Driscoll W.W., Durand P.M., Herron M.D., Rashidi A. On the paradigm of altruistic suicide in the unicellular world// Evolution. – 2011. – N 65(1). – P.3–20.

222. Durand P.M., Ramsey G. The Nature of Programmed Cell Death// Biol Theory. – 2019. – N 14. – P.30–41.

223. Tessier G. Les lois quantitatives de la croissance// Annales de Physiologie et de Physiochimie Biologiquem. – 1936. – N 12. – P.527–573.

224. Iserles A. A First Course in the Numerical Analysis of Differential Equations. — Cambridge: Cambridge University Press. – 1996.

225. Lomonosov T. Application of the Homomorphism Theorem for Rings to Linearization of Systems of Partial Differential Equations// Journal of Mathematical Sciences. – 2021. – N 255(4). – P.459–466.

## ПРИЛОЖЕНИЕ А ТАБЛИЦЫ К ПОДРАЗДЕЛУ 4.5

Таблица А.1 – Перечень независимых переменных, коэффициенты регрессии

Обозначен ие	Определение	Среднее	Станд.откл онение	Коэф-т	Станд. откло нение
Константа				3.01**	0.27
sex	Пол: 0 – женский, 1 -мужской	0.31	0.46	0.18**	0.05
Average_mark	Средняя оценка за все время обучения	7.73	1.26	-0.10**	0.02
USL1y201 8s1_r2	Оценка обр. программы - Качество научно-образовательных мероприятий вне обязательной программы (по шкале от 1 до 5)	3.25	1.94	-0.01	0.01
USL1y201 8s1_r3	Оценка обр. программы - Полезность получаемого образования для будущей работы / профессиональной деятельности (по шкале от 1 до 5)	3.72	1.35	-0.07**	0.02
USL1y201 8s1_r4	Оценка обр. программы - Уровень преподавания английского языка в Вышке (по шкале от 1 до 5)	2.24	1.77	0.05**	0.02
USL1y201 8s2_r1	Оценка сопровождения обр. процесса - Условия для аудиторного обучения (по шкале от 1 до 5)	3.88	1.03	-0.06**	0.02
USL1y201 8s2_r3	Оценка сопровождения обр. процесса - Информационное сопровождение обучения (по шкале от 1 до 5)	4.02	1.09	-0.01	0.02
USL2By20 18	Альтернативное решение о поступлении в магистратуру	0.33	0.76	-0.13**	0.04
ONLINE1 y2018_1	Включение онлайн-дисциплин в свой план - онлайн-дисциплины: 0 – нет, 1 – да	0.34	0.47	0.07	0.08
ONLINE1 y2018_3	Включение онлайн-дисциплин в свой план - смешанное обучение (онлайн-лекции БЕЗ оффлайн-семинаров, очно только экзамен): 0 – нет, 1 – да	0.09	0.28	-0.05	0.10
ONLINE1 y2018_4	Включение онлайн-дисциплин в свой план - не включал(а) такие дисциплины в свой план: 0 – нет, 1 – да	0.46	0.49	2.20**	0.15
ONLINE0 y2018	Предпочтительный формат курсов (по шкале от 1 до 5, где 1 – определенно предпочитаю онлайн-курсы, 5 – определенно предпочитаю аудиторные занятия)	2.22	2.36	-0.65**	0.02
ONLINE2 y2018_r1	Оценка организации онлайн-курсов - Список онлайн-дисциплин для студентов довольно узкий (по шкале от 1 до 5)	1.46	1.89	0.09**	0.02

Продолжение Таблицы А.1

ONLINE2_y2018_r2	Оценка организации онлайн-курсов - Проблемы с переводом оценок в 10-балльную систему НИУ ВШЭ и/или начислением кредитов (по шкале от 1 до 5)	0.82	1.35	-0.01	0.02
ONLINE2_y2018_r3	Оценка организации онлайн-курсов - Пришлось самостоятельно оплатить (по шкале от 1 до 5)	0.60	1.03	-0.01	0.03
ONLINE2_y2018_r4	Оценка организации онлайн-курсов - При прохождении экзамена в онлайн-формате студенту легко списать (по шкале от 1 до 5)	1.28	1.78	-0.01	0.02
ONLINE2_y2018_r5	Оценка организации онлайн-курсов - Учебный офис допускает ошибки в работе с онлайн-курсами (по шкале от 1 до 5)	0.95	1.54	0.03	0.02
MS2018	Оценка материального положения семьи (по шкале от 1 до 4, где 1 – Живем крайне экономно, на ежедневные расходы хватает, а покупка одежды уже представляет трудность, 4 - Хорошо обеспечены, можем достаточно легко позволить себе покупку автомобиля или дорогостоящий отдых)	2.54	1.10	0.06**	0.02
TRUD1y2_018_2	Наличие оплачиваемой работы - работа на полный день в Вышке: 0 – нет, 1 – да	0.01	0.08	0.04	0.34
TRUD1y2_018_3	Наличие оплачиваемой работы - работа на полный день вне Вышки : 0 – нет, 1 – да	0.11	0.31	0.24**	0.09
TRUD1y2_018_4	Наличие оплачиваемой работы - работа на неполный день в Вышке: 0 – нет, 1 – да	0.03	0.18	-0.02	0.15
TRUD1y2_018_5	Наличие оплачиваемой работы - работа на неполный день в Вышке: 0 – нет, 1 – да	0.13	0.34	0.27**	0.07
TRUD1y2_018_6	Наличие оплачиваемой работы - нерегулярные приработки в Вышке: 0 – нет, 1 – да	0.04	0.19	-0.01	0.14

Таблица А.2 – Результаты оптимальной спецификации

Обозначен ие	Определение	Среднее	Станд. отклонен ие	Коэф-т	Станд. откло нение
Константа				0.002	0.31
Относитесь ли Вы к какой-либо из нижеследующих групп?					
QF1_1	Преподаватель: 0 – нет, 1 – да	0.08	0.28	0.14	0.14
QF1_2	Разработчик онлайн курсов/платформ: 0 – нет, 1 – да	0.03	0.17	-0.62**	0.24
Q2	Планируете ли Вы приобретать подтвержденный сертификат об окончании курса? 1 – уже приобрел, 2 – планирую оплатить, 3 – хочу, но не уверен, 4 – не буду покупать	3.20	0.84	0.12**	0.05
Q5	Каков уровень Вашей подготовки в предметной области курса? Оцените его по шкале от 1 до 5, где – ‘1’ – нулевой уровень подготовки, а ‘5’ – продвинутый уровень подготовки.	2.11	1.14	0.004**	0.002
Продолжите фразу: “Я участвую в курсе... - потому что ... (по шкале от 1 – совершенно не согласен, 4 – полностью согласен)					
Q1_r4	мне интересно узнать больше о преподавании данного предмета	2.26	1.26	0.002	0.002
Q1_r5	хочу доказать себе, что могу освоить сложный курс	2.42	1.22	-0.001	0.002
Q1_r6	хочу расширить свой кругозор	3.46	0.96	0.003	0.003
Q1_r7	хочу пообщаться с другими студентами, заинтересованными в данной теме	2.05	1.10	-0.001	0.002
Q1_r9	он полезен для моего обучения в вузе/подготовки к обучению в вузе	2.28	1.35	0.003	0.002
Q1_r13	он может помочь мне сменить сферу деятельности или найти новую работу	2.40	1.32	0.001	0.002
Q1_r15	курс ведет именно этот преподаватель	1.64	1.21	-0.004**	0.00
Q3	Выберите, пожалуйста, что лучше всего описывает Ваши планы относительно просмотра видео-лекций в рамках прохождения курса: от 1 – просмотреть все видеолекции до 5 – не планирую смотреть	1.38	0.92	-0.209**	0.05

Продолжение Таблицы А.2

Q4	Выберите, пожалуйста, что лучше всего описывает Ваши планы относительно выполнения заданий в рамках прохождения курса: от 1 – выполнить все задания до 5 – не планирую выполнять задания	1.62	0.96	-0.07	0.04
Обучались ли Вы ранее на платформах массовых открытых онлайн-курсов (MOOCs)? 0 – нет, 1 – да					
Q6_1	Не участвовал(а) ранее в курсах на платформах MOOCs	0.39	0.48	-3.18**	0.18
Q6_2	Coursera	0.50	0.50	-0.50**	0.12
Q6_3	EdX	0.14	0.35	-0.35**	0.11
Q6_4	Udacity	0.08	0.28	-0.67**	0.14
Q6_7	ИНТУИТ	0.06	0.24	0.76**	0.13
Q6_8	Универсариум	0.04	0.21	0.53**	0.15
Q6_9	Национальная платформа открытого образования (openedu.ru)	0.08	0.28	0.77**	0.12
Q6_10	Другое	0.08	0.27	0.63**	0.12
Q8	Пол: 1 – м, 2 – ж	1.47	0.49	-0.05	0.08
Q9	Возраст (число полных лет)	29.65	11.45	-0.02**	0.005
Q10	Укажите наивысший уровень полученного Вами образования: от 1 – еще не закончил 9 классов до 7 - высшее профессиональное и ученая степень	4.72	1.35	0.07*	0.07
Укажите характер Вашей занятости (0 – нет, 1 – да)					
Q11_1	Учусь в школе/гимназии/лицее	0.04	0.20	-0.30	0.25
Q11_4	Работаю по найму полный рабочий день	0.45	0.49	0.21**	0.10
Q11_5	Работаю по найму неполный рабочий день	0.08	0.28	0.20	0.14
Q11_6	Являюсь предпринимателем	0.06	0.25	0.31*	0.16
Q11_7	Являюсь фрилансером	0.09	0.29	0.27**	0.13
Q11_9	Нахожусь на пенсии	0.02	0.14	0.21	0.35
Q11_10	Временно нахожусь в отпуске (в т.ч. декретном)	0.03	0.17	0.90**	0.24
Q11_11	Другое	0.02	0.14	0.64**	0.29