

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

УДК: 004.8, 004.9, 519.8, 330.4  
№ госрегистрации: 115061070059  
Инв.№:

УТВЕРЖДАЮ  
Проректор НИУ ВШЭ,  
к.э.н.

М.М. Юдкевич

ОТЧЕТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ  
АНАЛИЗ ДАННЫХ И ПРИНЯТИЕ РЕШЕНИЙ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ И  
ПОЛИТИЧЕСКИХ СИСТЕМАХ

(заключительный)

Шифр: Т3-81

Руководители темы:

зав. МНУЛ анализа и выбора решений,  
д.техн.н.

Ф.Т. Алескеров

гл.н.с. МНУЛ анализа и выбора решений, д.н., PhD

Э. Маскин

Москва 2015

## СПИСОК ИСПОЛНИТЕЛЕЙ

Соруководители темы:

Зав. лабораторией, д.т.н., рук. деп., проф.	_____	Ф.Т. Алескеров (введение, заключение)
Гл. научн. сотр., PhD	_____	Э. Маскин (пункт 1.1.2)

Исполнители темы:

Вед. науч. сотр., д.ф.-м.н., проф.	_____	М.С. Агранович (пункт 2.1.4)
Вед. науч. сотр., д.т.н., проф.	_____	А.С. Беленький (пункт 1.1.1)
Ст.-исследователь, студент	_____	Д.С. Болкунов (пункт 1.1.3)
Ст.-исследователь, студент	_____	А.М. Борискова (библиография)
Ст.-исследователь, препода- вателъ, аспирант	_____	Ю.А. Веселова (пункт 2.4.3)
Ст.-исследователь, студент	_____	Е.Р. Викторова (пункты 1.2.1, 1.2.2)
Лаборант, студент	_____	С.С. Демин (пункт 1.2.8, 1.2.9)
Ст.-исследователь, студент	_____	А.А. Денисенко (подпункт 3.2.3.1)
Науч. сотр., PhD	_____	Э. Доган (пункты 2.1.6, 2.1.7)
Мл. науч. сотр., преподава- тель, аспирант	_____	Л.Г. Егорова (пункт 1.4.2)
Лаборант, студент	_____	А.Ю. Емзешева (Приложение А)
Ст.-исследователь, студент	_____	Д.Б. Ефимов (подпункт 1.3.1.1)
Ст.-исследователь, студент	_____	А.А. Иванов (пункт 2.4.1)

Ст.-исследователь, преподаватель	_____	Р.У. Камалова	(пункты 3.1.2, 3.1.3)
Ст. науч. сотр., к.э.н., доцент	_____	Д.С. Карабекян	(пункт 2.4.2)
Ст. науч. сотр., к.э.н., доцент	_____	А.В. Карпов	(пункт 3.2.2)
Мл. науч. сотр., преподаватель	_____	С.Г. Кисельгоф	(пункты 1.4.1, 1.4.3)
Ст.-исследователь, аспирант	_____	А.В. Костров	(подпункт 3.2.3.2)
Ст.-исследователь, студент	_____	М.С. Курапова	(пункт 1.2.3)
Вед. науч. сотр., д.ф.-м.н., проф.	_____	А.А. Лазарев	(пункт 2.1.5)
Ведущий эксперт, к.т.н.	_____	А.С. Лебедев	(пункт 2.2.2)
Зам. зав. лаб., д.ф.-м.н., проф.	_____	А.Е. Лепский	(пункты 3.4.3 – 3.4.5)
Ст.-исследователь, студент	_____	А.Е. Липачева	(пункты 3.1.1)
Лаборант, студент	_____	Н.Ю. Лысенко	(Приложение Б)
Ст.-исследователь, студент	_____	Н.Г. Мещерякова	(подпункты 1.3.1.2, 1.3.1.3)
Ст.-исследователь, преподаватель, аспирант	_____	А.Л. Мячин	(пункты 1.2.4, 1.2.5)
Ст.-исследователь, студент	_____	В.В. Олейник	(подраздел 3.3)
Ст. науч. сотр., к.э.н., доцент	_____	Г.И. Пеникас	(пункты 3.4.1, 3.4.2)
Ст.-исследователь, студент	_____	А.С. Петров	(Приложение В)
Ст.-исследователь, преподаватель	_____	В.В. Петрущенко	(пункты 1.2.6, 1.2.7)

ватель, студент

Вед. науч. сотр., д.т.н., проф. \_\_\_\_\_ Б.В. Подиновский (пункты 2.1.1 – 2.1.3)

Эксперт, к.т.н., доцент \_\_\_\_\_ А.А. Рубчинский (подраздел 4.3)

Ст. науч. сотр., к.ф.-м.н.,  
доц. \_\_\_\_\_ К.С. Сорокин (пункты 1.3.2, 1.3.3)

Ст. науч. сотр., к.ф.-м.н.,  
доц. \_\_\_\_\_ А.Н. Субочев (пункт 2.2.1)

Ст.-исследователь, препода-  
вател, аспирант \_\_\_\_\_ Д.Н. Тверской (аннотации)

Ст.-исследователь, препода-  
вател, аспирант \_\_\_\_\_ Г.Г. Федин (подпункт 3.2.3.3,  
пункт 3.2.4)

Лаборант, студент \_\_\_\_\_ В.А. Федотова (реферат)

Ст.-исследователь, студент \_\_\_\_\_ В.С. Цыганкова (Приложение Г)

Мл. науч. сотр., ст. препода-  
вател \_\_\_\_\_ Е.Л. Черняк (подразделы 4.1, 4.2)

Науч. сотр., к.ф.-м.н., препо-  
даватель \_\_\_\_\_ Д.А. Шварц (пункт 3.2.1)

Ст.-исследователь \_\_\_\_\_ С.В. Швыдун (пункт 2.3.1)

Ст. науч. сотр., к.т.н., \_\_\_\_\_ В.И. Якуба (пункты 2.3.2 – 2.3.4)

## РЕФЕРАТ

Отчет 321 с., 1 ч., 90 рис., 53 табл., 258 источников, 4 прил.

МНОГОКРИТЕРИАЛЬНОЕ ПРИНЯТИЕ РЕШЕНИЙ, ДИЗАЙН  
МЕХАНИЗМОВ, ТЕОРИЯ КОЛЛЕКТИВНОГО ВЫБОРА, ТЕОРИЯ ИГР,  
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, СЕТЕВОЙ АНАЛИЗ  
ДАННЫХ.

Объектом исследования были модели и методы принятия решений, теории игр, коллективного выбора и анализа данных в социально-экономических, финансовых и политических системах.

Цель работы – разработка и исследование новых моделей анализа данных и принятия решений, а также применение разработанных методов к решению ряда прикладных задач в социально-экономической, финансовой и политической сферах.

В частности, целями исследований были:

- исследования организационных моделей и механизмов;
- исследования пространственных моделей взаимодействия в экономической и политической сферах;
- исследование сетевых моделей взаимодействия;
- прогнозирования структуры приема в вузы с учетом тенденций спроса на высшее образование;
- исследование методов многокритериального принятия решений;
- сравнительное исследование и аксиоматическое описание концепций решений в задаче оптимального коллективного выбора;
- исследование нормативных свойств двухступенчатых процедур выбора;
- исследование степени манипулируемости правил агрегирования предпочтений в модели Impartial Anonymous Culture;
- анализ согласованности позиций групп в Европейском парламенте в 1979–2014 гг.;

- оценка распределения влияния в Европейском парламенте в 1979–2014 гг.;
- анализ структуры Верховной Рады Украины в 2007–2012 гг.;
- исследование конфликтности и точности рекомендаций российских финансовых аналитиков;
- исследования моделей и методов извлечения знаний из текстов.

Методологической основой исследований были теория многокритериального принятия решений, теория индивидуального и коллективного выбора, теория игр, теория функций доверия. Основными методами исследований были методы оптимизации, численные и комбинаторные методы, методы теории вероятностей и математической статистики, методы кластерного анализа данных, методы сетевого анализа данных, методы неточных вероятностей, методы анализа текстов и др.

Техническим средством исследования были информационно-вычислительные системы. Для прикладных исследований использовались интернет-данные и базы данных из открытых источников.

В результате выполнения проекта получены следующие результаты:

- предложены две математические модели для выбора стратегий инвестирования мелкими и средними участниками биржевых торгов, в которых задачи поиска оптимальных стратегий сформулированы в виде задач линейного программирования, смешанного математического программирования и задачи отыскания седловых точек в антагонистической игре на выпуклых многогранниках; доказана возможность отыскания верхней оценки значения указанной игры на выпуклых многогранниках из решения задач линейного программирования, образующих двойственную пару;
- построена математическая модель для поиска оптимального плана модернизации грузовой транспортной системы посредством добавления в неё новых грузовых транспортных узлов;
- предложена новая математическая модель для отыскания взаимоприемлемых условий взаимодействия государства и частного сектора экономики по

- выполнению крупномасштабных проектов, которые государство не может финансировать в полном объеме;
- разработана теоретико-игровая математическая модель для анализа экономической эффективности прямых контрактов между производителями и поставщиками электроэнергии на рынке мощности электроэнергии;
  - в рамках исследования моделей пространственного взаимодействия проведен анализ интенсивности потенциальных конфликтов интересов стран в Арктике с использованием правила порогового агрегирования; исследованы различные сценарии развития событий в Арктическом регионе;
  - в рамках исследования моделей сетевого взаимодействия предложены новые методы определения степени важности узлов в сложных сетях, основанные на использовании правил коллективного выбора;
  - разработана модель многоступенчатого выбора вуза абитуриентом с учетом популярности и престижности профессий, самооценки абитуриентами своих способностей и возможностей поступить в определенные вузы в соответствии с полученными баллами ЕГЭ и конкурсом в вузах;
  - проведена серия вариативных расчетов по разработанным моделям на основе данных о баллах ЕГЭ выпускников 2013 года и индексов популярности профессий, рассчитанных на основе семантического анализа текстовых данных сети Интернет и СМИ;
  - разработан метод проверки альтернатив на потенциальную недоминируемость в дискретных многокритериальных задачах принятия решений, который позволяет построить множество потенциально недоминируемых альтернатив;
  - в рамках сравнительного исследования и аксиоматического описания концепций решений в задаче оптимального коллективного выбора получены условия принадлежности альтернативы одному из минимальных а) Р-доминирующих множеств, б) Р-внешнеустойчивых множеств, в) Р-внешнеустойчивых множеств;

- установлен ряд свойств для объединения минимальных Р-внешнеустойчивых множеств, объединения минимальных R-внешнеустойчивых множеств и объединения минимальных Р-доминирующих множеств;
- проведен сравнительный анализ аксиоматических свойств процедур агрегирования, использованных для решения задачи ранжирования научных журналов на основании значений нескольких библиометрических показателей; показано, что наилучшим по релевантным свойствам способом агрегирования базовых рейтингов является сортировка с помощью объединения минимальных Р-внешнеустойчивых множеств;
- исследованы новые двухступенчатые процедуры выбора на предмет удовлетворения существующим нормативным условиям (наследования, согласия, отбрасывания, аксиоме выбора Эрроу, монотонности, пороговой некомпенсируемости); исследована вычислительная сложность вновь рассмотренных двухступенчатых процедур выбора;
- исследована коалиционная манипулируемость в модели Impartial Anonymous Culture для трех альтернатив и семи различных правил коллективного выбора; выявлены наименее манипулируемая процедура (процедуры Нансона) и наиболее манипулируемая (процедура одобряющего голосование); проведено сравнение данных результатов с результатами предыдущих исследований (коалиционного манипулирования в Impartial Culture);
- проведен анализ межгрупповой и внутригрупповой согласованности Европарламента в 1979-2014гг. на данных о поименных голосованиях по семи созывам с помощью индексов согласованности Райса и Сатарова, мер фрагментированности и поляризованности общества;
- проведен анализ распределения влияния в семи созывах Европарламента на данных о поименных голосованиях с использованием индексов влияния с учетом предпочтений по созданию коалиций; получены оценки влияния на

- уровне групп (европейских политических партий и стран), а также на уровне отдельных европарламентариев;
- исследована структура Верховной Рады Украины (2007–2012) с помощью анализа более тысяч поименных голосований депутатов; проведена оценка согласованности позиций фракций с помощью различных индексов;
  - методами теории функций доверия проанализирована согласованность (конфликтность) и точность прогнозов инвестбанков о стоимости акций российских компаний;
  - в рамках исследования моделей и методов извлечения знаний из текстов проанализированы различные модели представления текста и способы оценивания релевантности строк тексту в различных моделях; проанализирован метод оценивания релевантности строк тексту, основанный на применении нормированного аннотированного суффиксного дерева.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	13
ОСНОВНАЯ ЧАСТЬ.....	20
1 Исследования организационных, пространственных и сетевых моделей в социально-экономической, финансовой и политической сферах .....	20
1.1 Исследования организационных моделей и механизмов .....	21
1.1.1 Разработка математических моделей и методов оптимизации управления портфелем финансовых инструментов для мелких и средних участников биржевых торгов.....	21
1.1.2 Некоторые другие исследования организационных моделей и механизмов .....	37
1.1.3 Некоторые выводы.....	40
1.2 Исследования пространственных моделей взаимодействия в экономической и политической сферах.....	41
1.2.1 Географическая информация о странах .....	42
1.2.2 Взаимные интересы стран в регионе.....	44
1.2.3 Используемая модель.....	49
1.2.4 Функции полезности стран .....	50
1.2.5 Агрегирование функций полезности.....	52
1.2.6 Исследование зон конфликтов, сценарии .....	54
1.2.7 Сценарий сильной напряженности.....	58
1.2.8 Сценарий слабой напряженности .....	59
1.2.9 Некоторые выводы.....	59
1.3 Исследование сетевых моделей взаимодействия .....	60
1.3.1 Исследование взаимосвязи между правилами Копланда и различными индексами центральности .....	61
1.3.2 Расчет индексов на сети денежных переводов.....	67
1.3.3 Некоторые выводы.....	72
1.4 Прогнозирования структуры приема в вузы с учетом тенденций спроса на высшее образование .....	73
1.4.1 Анализ существующих подходов к прогнозированию контрольных цифр приема.....	73
1.4.2 Моделирование поведения семей и абитуриентов, предъявляющих спрос на высшее образование .....	74
1.4.3 Некоторые выводы.....	84
2 Исследование методов многокритериального и коллективного принятия решений .....	86
2.1 Исследование методов многокритериального принятия решений .....	86
2.1.1 Потенциально недоминируемые альтернативы и их свойства.....	87
2.1.2 Проверка потенциальной недоминируемости альтернатив .....	89
2.1.3 Полезность и стохастическое доминирование .....	93
2.1.4 Меры риска как критерии выбора .....	96
2.1.5 Модели выбора с мерами риска.....	103
2.1.6 Среднее полуотклонение и его применение как критерия выбора .....	110

2.1.7 Некоторые выводы.....	114
2.2 Сравнительное исследование и аксиоматическое описание концепций решений в задаче оптимального коллективного выбора .....	115
2.2.1 Сравнительное исследование и аксиоматическое описание некоторых турнирных решений.....	115
2.2.2 Сравнительный анализ аксиоматических свойств процедур построения агрегированных ранжирований научных журналов по менеджменту, политологии и экономике.....	121
2.3 Исследование нормативных свойств двухступенчатых процедур выбора .....	122
2.3.1 Двухступенчатые процедуры выбора .....	124
2.3.2 Исследование свойств двухступенчатых процедур выбора .....	128
2.3.3 Вычислительная сложность процедур выбора.....	128
2.3.4 Некоторые выводы.....	129
2.4 Исследование степени манипулируемости правил агрегирования предпочтений в модели Impartial Anonymous Culture .....	131
2.4.1 Описание исследуемых процедур и вычислительного эксперимента .....	132
2.4.2 Результаты компьютерного моделирования .....	134
2.4.3 Некоторые выводы.....	138
3 Исследования распределения влияния и согласованности решений в политических и финансовых системах.....	139
3.1 Анализ согласованности позиций групп в Европейском парламенте в 1979–2014 гг. .....	139
3.1.1 Описание данных .....	140
3.1.2 Анализ согласованности в Европарламенте.....	142
3.1.3 Некоторые выводы и заключение.....	158
3.2 Оценка распределения влияния в Европейском парламенте в 1979–2014 гг. ....	158
3.2.1 Европейский парламент: история института и практика работы.....	160
3.2.2 Методика оценки распределения влияния и моделирования коалиционных предпочтений.....	169
3.2.3 Описание результатов оценки распределения влияния депутатов Европарламента, политических групп и стран-членов ЕС .....	175
3.2.4 Некоторые выводы.....	195
3.3 Анализ структуры Верховной Рады Украины в 2007–2012 гг. ....	197
3.4 Исследование конфликтности и точности рекомендаций российских финансовых аналитиков.....	205
3.4.1 Обзор литературы .....	205
3.4.2 Обзор данных.....	211
3.4.3 Методология исследований .....	213
3.4.4 Результаты исследований .....	225
3.4.5 Некоторые выводы.....	231
4 Исследования моделей и методов извлечения знаний из текстов .....	233
4.1 Способы формального представления текстов и их приложения.....	233
4.1.1 Векторная модель представления текста .....	234
4.1.2 Языковая модель представления текста .....	239
4.1.3 Представление текста на основе моделей скрытых тем.....	240

4.1.4 Представление текста на основе модели суффиксных деревьев.....	246
4.2 Оценивание релевантности строки тексту с использованием метода аннотированного суффиксного дерева (АСД) .....	255
4.2.1 Проблема оценивания релевантности строки тексту и основные подходы к ее решению.....	255
4.2.2 Релевантность в векторной модели .....	256
4.2.3 Релевантность в бинарной модели независимости .....	257
4.2.4 Релевантность в вероятностной модели .....	257
4.2.5 Релевантность в тематических моделях.....	258
4.2.6 Релевантность в модели аннотированного суффиксного дерева .....	259
4.3 Метод nAST-k оценивания релевантности строки тексту с использованием нормированного АСД.....	261
4.3.1 Структура метода .....	261
4.3.2 Построение таблицы релевантности Стока-Текст (pCT) .....	268
4.4 Некоторые выводы.....	270
<b>ЗАКЛЮЧЕНИЕ.....</b>	<b>272</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....</b>	<b>276</b>
<b>ПРИЛОЖЕНИЕ А Таблицы к подразделу 1.3 .....</b>	<b>300</b>
<b>ПРИЛОЖЕНИЕ Б Таблицы к подразделу 2.3.....</b>	<b>302</b>
<b>ПРИЛОЖЕНИЕ В Рисунки к подразделу 3.1 .....</b>	<b>305</b>
<b>ПРИЛОЖЕНИЕ Г Таблицы и рисунки к подразделу 3.2 .....</b>	<b>307</b>

## ВВЕДЕНИЕ

В данном отчете представлены результаты работы в 2015 году коллектива сотрудников Международной научно-учебной лаборатории анализа и выбора решений НИУ ВШЭ над проектом «Анализ данных и принятие решений в социально-экономических и политических системах».

В соответствии с техническим заданием были проведены исследования в следующих научных направлениях:

- исследования организационных, пространственных и сетевых моделей в социально-экономической, финансовой и политической сферах;
- исследование методов многокритериального и коллективного принятия решений;
- исследования распределения влияния и согласованности решений в политических и финансовых системах;
- исследования моделей и методов извлечения знаний из текстов.

Работа над проектом этого года стала продолжением исследований, проводимых в Международной научно-учебной лаборатории анализа и выбора решений в 2014 г. в ходе работы над проектом «Теоретическое и численное исследование современных математических моделей в социально-экономической, политической и финансовой сферах» [1]. В частности, в 2014 г. были получены следующие новые теоретические результаты:

- в рамках разработки организационных моделей и механизмов построены новые математические модели для оптимизации: а) расписания движения поездов; б) стратегии перемещения порожних вагонов; в) тарифов на перевозки грузов железнодорожным транспортом; г) работы железнодорожной сортировочной станции; д) объемов выработки электроэнергии базовыми генераторами региональных сетей электроснабжения;
- в рамках теоретико-игрового исследования по библиометрии показана эквивалентность вектора Шепли и метода равного разделения долей соавторов при оценивании вклада авторов в совместные публикации;

- построены эффективные алгоритмы для расчета индексов влияния при формировании коалиций с учетом предпочтений агентов в случае, когда все игроки имеют равное количество голосов;
- в рамках теоретико-экспериментальных исследований по автоматической интерпретации текстовых документов: а) проведены экспериментальные исследования методов оценивания релевантности строк и текстов для решения задач автоматизации аннотирования статей; б) разработана и исследована мера близости между двумя предложениями на основе аннотированных суффиксных деревьев; в) разработан метод лифтинга в таксономиях на основе принципа максимального правдоподобия;
- исследованы альтернативные процедуры агрегирования исходных переменных индекса конкурентоспособности промышленного производства, разработанного экспертами ЮНИДО; вместо кардинальной агрегирующей функции применены ординальные методы ранжирования, заимствованные из теории коллективного выбора, основанные на правиле большинства; показано, что некоторые из новых версий глобального рейтинга представляют данный набор критериев лучше, чем исходная версия;
- проанализированы нечеткостные методы сравнения гистограмм применительно к возможности их использования для ранжирования вузов по гистограммам ЕГЭ поступивших абитуриентов; предложена модель «идеально-го» зачисления абитуриентов и показано, что распределение зачисленных абитуриентов в вуз в такой модели будет унимодальным при достаточно общих предположениях о функции предпочтения; поставлена и решена задача оптимального преобразования гистограмм, приводящей ее к унимодальному виду;
- найдены необходимые и достаточные условия на уровень искажений гистограмм, при выполнении которых результат сравнения гистограмм некоторыми наиболее популярными вероятностными методами не изменится;

- проведена серия экспериментов имитационного моделирования поведения игроков на бирже, в которой агенты (трейдеры) являются мелкими и/или средними игроками; найдены условия на параметры модели (опыт/удачливость трейдера, используемые инструменты торговли и др.), при которых вероятность получения прибыли трейдером достаточно высока/низка;
- в рамках развития теории важности критериев определено и исследовано понятие превосходства в важности одной группы критериев над другой; представлены решающие правила для количественной и специальных видов качественной информации о группах критериев; предложено новое аналитическое правило для многокритериальных задач, в которых критерии с порядковой шкалой нестрого упорядочены по важности;
- в рамках исследования манипулирования правил агрегирования предпочтений были найдены индексы манипулируемости при бесконечно большом числе избирателей для популярных позиционных правил коллективного выбора; была исследована вычислительная сложность различных задач и алгоритмов манипулирования правил агрегирования предпочтений;
- в рамках исследования индексов поляризованности проведено сравнение индексов поляризованности Александрова-Голубенко и Эстебана-Рэя в случаях равного распределения четного и нечетного числа групп; кроме того, проведено сравнение указанных индексов со среднеквадратичным отклонением положений групп от среднего значения; исследована устойчивость этих индексов к малым возмущениям;
- в рамках исследования свойств процедур выбора, основанных на суперпозиции, проанализировано 28 известных и 190 двухступенчатых процедур многокритериального выбора на предмет удовлетворения существующим нормативным условиям (рациональности, монотонности, пороговой некомпенсируемости), которые характеризуют качественные оценки процедур выбора; кроме того, проанализирована вычислительная сложность рассмотренных процедур выбора.

Исследования 2015 года, выполненные в ходе реализации проекта «Анализ данных и принятие решений в социально-экономических и политических системах» продолжают некоторые исследования прошлого года (в частности, исследования организационных моделей и механизмов, исследования по автоматическому анализу текстов, развитие теории многокритериального принятия решений и т.д.). В тоже время, некоторые направления исследований этого года (например, исследования пространственных и сетевых моделей взаимодействия в социально-экономической и политической сферах, исследования согласованности решений в рамках теории функций доверия) являются пионерскими. Эти исследования имеют хороший потенциал для дальнейшего развития.

Методологической основой исследований служат теория многокритериального принятия решений, теория игр (в частности, конструирование экономических механизмов), теория коллективного и индивидуального выбора, интеллектуальный анализ данных. Основными средствами и методами исследований являются комбинаторные методы, процедуры оптимизации, методы статистического и кластерного анализа данных, методы анализа текстов, методы теории вероятностей, методы сетевого анализа данных, методы теории функций доверия и др.

Техническим средством исследования служат электронные вычисильные машины. Для прикладных исследований использовались интернет-данные и открытые базы данных по различным тематикам.

Данный отчет состоит из 4 разделов, соответствующих основным направлениям исследований, проводимых в ходе выполнения данного проекта и 4-х приложений.

В первом разделе приведены результаты исследования организационных, пространственных и сетевых моделей в социально-экономической, финансовой и политической сферах. Эти исследования осуществлялись по нескольким направлениям.

Так в подразделе 1.1 рассмотрены новые организационные модели и механизмы. В частности, в этом подразделе разработаны математические модели и методы оптимизации управления портфелем финансовых инструментов для мелких и средних участников биржевых торгов, развит теоретико-игровой подход к анализу рынка мощности и рынка электроэнергии в регионе, построены математические модели для анализа и выбора схемы развития крупномасштабных транспортных систем, развит теоретико-игровой подход к анализу проблем государственно-частного партнерства.

В подразделе 1.2 приведены результаты исследований пространственных моделей взаимодействия в экономической и политической сферах. В частности, в этом подразделе проанализированы интенсивности потенциальных конфликтов интересов стран в Арктике с использованием правила порогового агрегирования. Кроме того, здесь же исследованы различные сценарии развития событий в Арктическом регионе.

В подразделе 1.3 приведены результаты исследований сетевых моделей взаимодействия в экономической и политической сферах. Основное внимание при этом было уделено разработке новых методов определения степени важности узлов в сложных сетях, основанных на использовании правил коллективного выбора. Разработанный подход был применен к анализу эмпирических данных о денежных переводах среди 214 стран мира в 2014 году.

В подразделе 1.4 разработана модель многоступенчатого выбора вуза абитуриентом с учетом популярности и престижности профессий, самооценки абитуриентами своих способностей и возможностей поступить в определенные вузы в соответствии с полученными баллами ЕГЭ и конкурсом в вузах.

Второй раздел отчета посвящен исследованиям методов многокритериального и коллективного принятия решений. Так, в подразделе 2.1 приведены новые результаты в рамках проводимого в последние годы исследования методов многокритериального принятия решений. В частности, в отчете описаны

ны разработанные методы проверки альтернатив на потенциальную недоминируемость в дискретных многокритериальных задачах принятия решений.

В подразделе 2.2 приведены новые результаты в рамках проводимого в последнее время сравнительного исследования и аксиоматического описания концепций решений в задаче оптимального коллективного выбора. В частности, с этой точки зрения были исследованы три родственных турнирных решения: объединение минимальных Р-доминирующих множеств, объединение минимальных Р-внешнеустойчивых множеств и объединение минимальных Р-внешнеустойчивых множеств.

Подраздел 2.3 посвящен исследованию нормативных свойств двухступенчатых процедур выбора и оценке вычислительной сложности таких процедур.

В подразделе 2.4 приведены результаты исследования степени манипулируемости правил агрегирования предпочтений в модели Impartial Anonymous Culture.

Третий раздел отчета посвящен исследованиям распределения влияния и согласованности решений в политических и финансовых системах. Так, в подразделе 3.1 представлены результаты анализа межгрупповой и внутригрупповой согласованности Европарламента в 1979–2014гг.

В подразделе 3.2 проанализировано распределение влияния Европарламенте на данных о поименных голосованиях в семи созывах с использованием индексов влияния с учетом предпочтений по созданию коалиций.

В подразделе 3.3 исследована структура Верховной Рады Украины (2007–2012) с помощью анализа более тысяч поименных голосований депутатов. В последнем четвертом подразделе третьего раздела проанализирована согласованность (конфликтность) и точность прогнозов инвестбанков о стоимости акций российских компаний методами теории функций доверия.

В четвертом, заключительном разделе отчета приведены результаты исследований моделей и методов извлечения знаний из текстов.

В Заключении приведены краткие выводы по результатам выполнения НИР и даны рекомендации по использованию этих результатов.

В Приложениях А,Б,В и Г приведены некоторые дополнительные таблицы и рисунки к отдельным подразделам отчета.

Все полученные результаты являются новыми. Их значимость подтверждена многочисленными аprobациями на международных научных форумах (конференциях, семинарах и пр.). Кроме того, практически все представленные результаты опубликованы, в том числе в реферируемых научных изданиях.

## ОСНОВНАЯ ЧАСТЬ

В ходе выполнения проекта «Анализ данных и принятие решений в социально-экономических и политических системах» были проведены исследования в следующих научных направлениях:

- исследования организационных, пространственных и сетевых моделей в социально-экономической, финансовой и политической сферах;
- исследование методов многокритериального и коллективного принятия решений;
- исследования распределения влияния и согласованности решений в политических и финансовых системах;
- исследования моделей и методов извлечения знаний из текстов.

### 1 Исследования организационных, пространственных и сетевых моделей в социально-экономической, финансовой и политической сферах

В этом разделе отчета приведены результаты исследований организационных, пространственных и сетевых моделей в социально-экономической и политической сферах. Эти результаты получены в следующих теоретических и прикладных направлениях:

- исследования организационных моделей и механизмов;
- исследования пространственных моделей взаимодействия в экономической и политической сферах;
- исследования сетевых моделей взаимодействия в экономической и политической сферах;
- прогнозирования структуры приема в вузы с учетом тенденций спроса на высшее образование.

## 1.1 Исследования организационных моделей и механизмов

В этом подразделе приведены результаты исследований некоторых организационных моделей и механизмов, которые осуществлялись в следующих направлениях:

- разработка математических моделей и методов оптимизации управления портфелем финансовых инструментов для мелких и средних участников биржевых торгов;
- некоторые другие исследования организационных моделей и механизмов.

### 1.1.1 Разработка математических моделей и методов оптимизации управления портфелем финансовых инструментов для мелких и средних участников биржевых торгов

Изучение законов функционирования и закономерностей поведения фондовой биржи представляет объективные трудности, поскольку фондовая биржа – это сложная система, в рамках которой взаимодействует большое число элементов с неявными взаимосвязями между ними. Более того, фондовая биржа функционирует в условиях неопределенности множества факторов, и к числу важнейших из них следует отнести неопределенность поведения самих участников биржевых торгов. Поэтому исследование поведения трейдеров при принятия ими решений о формировании и изменении инвестиционных портфелей и построение адекватных моделей их взаимодействия с биржей представляются целесообразными, прежде всего с точки зрения прояснения природы биржевых закономерностей и феноменов.

Практически во всех моделях поиска стратегий оптимального инвестирования на бирже, предлагаемых в рамках этой теории, начиная с работы Г. Марковица [2], предполагается, что трейдеру известен закон распределения будущей цены интересующих его финансовых инструментов. Однако обе указанные предпосылки, лежащие в основе этих моделей – рациональность трейдера и знание им закона распределения будущей цены актива – в реаль-

ной жизни выполняются далеко не всегда, а исследования по анализу способностей трейдеров действовать рационально вообще [3 – 6] свидетельствуют о неспособности не только значительного числа трейдеров, но даже и финансовых аналитиков принимать рациональные инвестиционные решения и предсказывать направление изменений стоимостей интересующие их финансовых инструментов на основе имеющихся у них предположений о законах распределения будущих цен на эти финансовые инструменты.

Возникает вопрос: можно ли построить модели, позволяющие трейдеру оценивать ожидаемый финансовый результат при отсутствии у него информации о каких-либо специальных законах распределения значений интересующих его финансовых инструментов, обращающихся на бирже? В частности, можно ли предложить модели, позволяющие трейдеру с подтвержденной способностью правильно оценивать в каждый момент времени направления изменения значений стоимостей интересующих его финансовых инструментов в случаях, когда: а) для каждого из указанных финансовых инструментов трейдер может указать некоторый диапазон, в котором изменяются значения этого финансового инструмента и которые трейдер считает равновозможными, и б) трейдер может указать только лишь предполагаемую область изменения значений всей совокупности интересующих его финансовых инструментов с помощью систем линейных неравенств балансового типа?

Способность трейдера предсказывать будущее поведение рынка или конкретных финансовых инструментов является ключевой характеристикой трейдера как участника биржевых торгов. Для выявления у трейдера такой способности можно воспользоваться, например, схемой Бернулли, т.е. провести серию испытаний, в каждом из которых определенное событие (угадывание направления движения цены конкретного финансового инструмента в следующий момент времени) происходит с некоторой (одной и той же) вероятностью  $p$ .

Модель 1: значения стоимостей финансовых инструментов являются случайными величинами, распределенными по закону равномерной плотности

Для удобства изложения каждому финансовому инструменту, доступному для торгов на фондовой бирже, присвоим номер из множества  $N = \{1, 2, \dots, n\}$ , где  $i \in N$  обозначает наименование финансового инструмента, находящегося во множестве  $N$  под номером  $i$ . Пусть далее:

- числа  $t_0 < \dots < t < t + 1 < t + 2 < \dots$  соответствуют моментам времени, в которые трейдер принимает решения об изменении структуры своего инвестиционного портфеля;
- $m_t$  (от слова money) – некоторая сумма наличных средств (часть капитала), оставшаяся после формирования инвестиционного портфеля к моменту времени  $t$ ;
- $W_t$  (от слова wealth) – объем капитала в виде наличных средств  $m_t$  и имеющихся в портфеле финансовых инструментов трейдера в момент времени  $t$ ;
- $s_{i,t}$  (от слова spot) – спот-цена или стоимость финансового инструмента  $i$  в момент времени  $t$ , т.е. цена, по которой продается финансовый инструмент в момент времени  $t$  в конкретном месте на условиях немедленного совершения сделки);
- $v_{i,t}$  (от слова volume) – число единиц финансового инструмента  $i$ , приобретенных трейдером в момент времени по цене  $s_{i,t}$ .

Основные предположения о поведении трейдера в момент времени  $t$ :

- 1) трейдер обладает подтвержденными (в результате испытаний по схеме Бернулли) способностями по оценке биржевой ситуации с финансовыми инструментами, формирующими его портфель, выраженными в виде известных ему вероятностей  $p_i$ , верно определять направление изменения будущей стоимости финансового инструмента  $i$  в момент времени  $t + 1$ , т.е. предсказывать будет ли эта стоимость расти или снижаться в следующий момент  $t + 1$  времени;

- 2) трейдер в момент времени  $t$  может разделить все множество финансовых инструментов  $N$  на непересекающиеся подмножества  $I_t^+, I_t^-, I_t^0$ , где:
- $I_t^+$  – множество финансовых инструментов, относительно которых трейдер уверен, что они вырастут в цене в момент времени  $t + 1$  и которые он собирается покупать в момент времени  $t$ ;
  - $I_t^-$  – множество финансовых инструментов, относительно которых трейдер уверен, что они снизятся в цене в момент времени  $t + 1$  и которые он собирается продавать в момент времени  $t$ ;
  - $I_t^0$  – множество финансовых инструментов, относительно которых трейдер считает, что они не изменятся в цене в момент времени  $t + 1$  (либо считает будущие их изменения слишком незначительными) и которые он не собирается ни продавать, ни покупать в момент времени  $t$ ;
- 3) для покупки финансовых инструментов из множества  $I_t^+$  в момент времени  $t$  трейдер может тратить наличные средства  $m_t$  и средства, вырученные от продажи в момент времени  $t$  финансовых инструментов из  $I_t^-$  (как из собственного портфеля, так и взятых взаймы, если такой кредит доступен трейдеру); аналогично, для продажи финансовых инструментов  $i \in I_t^-$  он может использовать собственные запасы этих финансовых инструментов в размере  $v_{i,t}$ , а также занимать финансовые инструменты у брокера, если это ему доступно;
- 4) трейдер не производит никаких действий с финансовыми инструментами из множества  $I_t^0$ .

Для упрощения записи математических моделей, описывающих действия трейдера, будем считать, что трейдер работает только с конкретными финансовыми инструментами — цennymi бумагами (акциями и облигациями) и не работает с производными цennymi бумагами (опционами, фьючерсами и проч.), а также что он выставляет только рыночные заявки, т.е. заявки, которые будут исполнены мгновенно по текущей рыночной цене.

При отыскании оптимальной стратегии инвестирования в финансовые инструменты трейдеру, работающему на фондовой бирже, приходится рассматривать две ситуации: а) формирования нового портфеля ценных бумаг, и б) изменения состава имеющегося у него портфеля ценных бумаг. Ниже рассматривается вторая из этих двух ситуаций (как более общая).

В момент времени  $t$  трейдер обладает портфелем ценных бумаг в размере  $v_{i,t}, i = \overline{1, n}$  и некоторым объемом наличных денег  $m_t$ , следовательно, его благосостояние на этот момент равно  $W_t = \sum_{i=1}^n v_{i,t} s_{i,t} + m_t$ . Задача трейдера состоит в выборе объемов покупки ценных бумаг  $x_{i,t}^+$  (целых чисел) из множества  $I_t^+$ , относительно которых трейдер ожидает увеличения их стоимости в момент  $t + 1$ , объемов продажи  $x_{i,t}^-$  (целых чисел) ценных бумаг из текущего портфеля ценных бумаг из множества  $I_t^-$  и объемов продажи  $z_{i,t}^-$  (целых чисел) ценных бумаг из множества  $I_t^-$ , взятых взаймы у брокера для открытия короткой позиции в момент  $t$  с возвратом ему этих ценных бумаг в момент  $t + 1$  по цене  $s_{i,t+1}$ .

Ясно, что благосостояние трейдера, ожидаемое им в момент  $t + 1$  составляет

$$W_{t+1} = \sum_{i \in I_t^0} v_{i,t} s_{i,t+1} + \sum_{i \in I_t^+} [v_{i,t} + x_{i,t}^+] s_{i,t+1} + \sum_{i \in I_t^-} [v_{i,t} - x_{i,t}^-] s_{i,t+1} + \\ + \left( m_t - \sum_{i \in I_t^+} x_{i,t}^+ s_{i,t} + \sum_{i \in I_t^-} x_{i,t}^- s_{i,t} + \sum_{i \in I_t^-} z_{i,t}^- [s_{i,t} - s_{i,t+1}] \right),$$

где первые три слагаемых определяют часть благосостояния трейдера, размещенного в принадлежащих ему ценных бумагах, а последнее слагаемое определяет сумму средств, оставшихся после совершения всех сделок по купле/продаже ценных бумаг к моменту времени  $t + 1$ , включая возврат заемных ценных бумаг.

Прирост стоимости портфеля после совершения всех сделок составит

$$\begin{aligned}\Delta W_{t+1} = & \sum_{i \in I_t^0} v_{i,t} (s_{i,t+1} - s_{i,t}) + \sum_{i \in I_t^+} (v_{i,t} + x_{i,t}^+) (s_{i,t+1} - s_{i,t}) + \\ & + \sum_{i \in I_t^-} (v_{i,t} - x_{i,t}^-) (s_{i,t+1} - s_{i,t}) + \sum_{i \in I_t^-} z_{i,t}^- (s_{i,t+1} - s_{i,t}).\end{aligned}$$

Здесь  $v_{i,t}$ ,  $s_{i,t}$ ,  $m_t$ ,  $i \in I_t^+, I_t^-$ , – известные действительные числа (числа  $v_{i,t}$  целые), а  $s_{i,t+1}, i \in I_t^+, I_t^-$  являются случайными величинами. Далее предполагается, что значения стоимостей ценных бумаг  $i, j \in N$  в момент времени  $t + 1$  являются независимыми случайными величинами.

Торговые операции трейдера с цennыми бумагами должны проводиться с учетом следующих ограничений:

- 1) условия целочисленности объемов покупаемых, продаваемых и берущихся взаймы ценных бумаг,
- 2) объем продажи ценной бумаги в количестве  $x_{i,t}^-$  из собственного портфеля ценных бумаг не может превышать имеющийся у него объем  $v_{i,t}$ ,

$$x_{i,t}^- \leq v_{i,t}, i \in I_t^-;$$

(Заметим, что если трейдер планирует продать ценную бумагу  $i$  в количестве, превышающем имеющийся у него объем  $v_{i,t}$  этой ценной бумаги, то он совершает дополнительно к продаже этого объема  $v_{i,t}$  заем ценных бумаг в количестве  $z_{i,t}$  для короткой позиции);

- 3) ограничение по доступности заемного капитала при использовании маржинальных кредитов с кредитным плечом  $1:k_t$  при условии нехватки собственного капитала (для покупки ценных бумаг или при продаже через открытие короткой позиции, т.е. продаже заемных ценных бумаг):

$$\sum_{i \in I_t^+} x_{i,t}^+ s_{i,t} + \sum_{i \in I_t^-} z_{i,t}^- s_{i,t} - \left( m_t + \sum_{i \in I_t^-} x_{i,t}^- s_{i,t} \right) \leq k_t \left( m_t + \sum_{i=1}^n v_{i,t} s_{i,t} \right),$$

где первые два слагаемых в левой части означают затраты на осуществление сделок по купле/продаже ценных бумаг, третье слагаемое – имеющийся в наличии у трейдера в момент  $t$  капитал в виде собственных денежных средств

$m_t$  и средств, полученных от продажи собственных ценных бумаг из инвестиционного портфеля, а правая часть представляет собой максимально доступный капитал с учетом заемных средств при доступном в момент  $t$  кредитном плече  $k_t$  ( $k_t = 1$  означает отсутствие маржинального кредита и использование собственного капитала при осуществлении сделок).

Также предполагается, что при принятии инвестиционных решений, трейдер исходит существования некоторого порога  $\alpha$ , относительно которого трейдер определяет критический момент остановки своей торговли на бирже  $W_{i,t+1} \geq \alpha W_{i,t}$ , т.е. трейдеру необходимо удерживать собственное благосостояние не ниже какого-то конкретного уровня  $\alpha \in [0,1]$ , в том числе при  $\alpha = 1$ :  $W_{i,t+1} \geq W_{i,t}$ .

Предположим далее, что трейдер, играющий на фондовой бирже, в момент времени  $t$  может оценить границы (пороги) изменений  $s_{i,t+1}^{\max}$  и  $s_{i,t+1}^{\min}$  будущей стоимости ценных бумаг  $i \in I_t^+, I_t^-$  в момент времени  $t + 1$ , основываясь на прошлых наблюдениях или оцененных им фундаментальных значениях стоимости каждой из ценных бумаг.

Если трейдер может оценить границы изменений, но не может сделать предположений о конкретном законе распределения будущей стоимости ценной бумаги в указанных им границах, то для оценки значений  $s_{i,t}$  естественно считать, что изменение стоимости ценной бумаги  $i$ , как в сторону увеличения, так и изменение стоимости ценной бумаги  $i$  в сторону уменьшения по отношению к текущему значению  $s_{i,t}$  являются непрерывными случайными величинами  $u$  и  $v$ , равномерно распределенными на промежутках  $[s_{i,t}, s_{i,t+1}^{\max}]$  и  $[s_{i,t+1}^{\min}, s_{i,t}]$ , соответственно, с плотностями распределения вероятностей

$$f_1(u) = \begin{cases} \frac{1}{s_{i,t+1}^{\max} - s_{i,t}}, & \text{при } u \in [s_{i,t}, s_{i,t+1}^{\max}], \\ 0, & \text{при } u \notin [s_{i,t}, s_{i,t+1}^{\max}], \end{cases}$$

$$f_1(v) = \begin{cases} \frac{1}{s_{i,t} - s_{i,t+1}^{\min}}, & \text{при } v \in [s_{i,t+1}^{\min}, s_{i,t}], \\ 0, & \text{при } v \notin [s_{i,t+1}^{\min}, s_{i,t}]. \end{cases}$$

Если трейдер прогнозирует, что стоимость ценной бумаги  $i$  вырастет по сравнению с текущей стоимостью  $s_{i,t+1} > s_{i,t}$ , то математическое ожидание значения стоимости ценной бумаги вида  $i$  составит  $Ms_{i,t+1} = \frac{s_{i,t} + s_{i,t+1}^{\max}}{2}$ , а если, напротив, он прогнозирует снижение ее стоимости  $s_{i,t+1} < s_{i,t}$ , то математическое ожидание стоимости ценной бумаги  $i$  составит  $Ms_{i,t+1} = \frac{s_{i,t+1}^{\min} + s_{i,t}}{2}$ .

Возможно также, что по ценной бумаге  $i$  трейдер не может высказать ни одного из указанных выше предположений, тогда естественно предполагать, что стоимость ценной бумаги  $i$  остается без изменения, т. е.  $s_{i,t+1} = s_{i,t}$ .

Если в момент времени  $t$  трейдер предполагает (с вероятностью  $p_i$ ), что в момент  $t + 1$  стоимость ценной бумаги  $i$  возрастет, т.е.  $i \in I_t^+$ , то среднее значение стоимости этой ценной бумаги составит  $\frac{s_{i,t} + s_{i,t+1}^{\max}}{2}$ . Если же в момент  $t$  решение трейдера относительно этой ценной бумаги ошибочно, то естественно предположить, что возможны два события: а) стоимость ценной бумаги  $i$  в момент времени  $t + 1$  уменьшится, б) стоимость ценной бумаги  $i$  в момент времени  $t + 1$  не изменится по сравнению с ее значением в момент времени  $t$ . Ясно, что эти два события несовместны, и естественно предположить, что эти события равновозможны, т.е., что каждое из них происходит с вероятностью  $\frac{1-p_i}{2}$ .

Таким образом, среднее значение стоимости ценной бумаги  $i \in I_t^+$  в результате предположений трейдера представляет собой дискретную случайную величину, принимающую три значения, ряд распределения которой задается следующей таблицей.

Таблица 1.1 – Ряд распределения случайной величины  $Ms_{i,t+1}, i \in I_t^+$

$Ms_{i,t+1}, i \in I_t^+$	$\frac{s_{i,t} + s_{i,t+1}^{\max}}{2}$	$\frac{s_{i,t+1}^{\min} + s_{i,t}}{2}$	$s_{i,t}$
---------------------------	--	--	-----------

$P$	$p_i$	$\frac{1-p_i}{2}$	$\frac{1-p_i}{2}$
-----	-------	-------------------	-------------------

Если же трейдер в момент времени  $t$  предполагает с вероятностью  $p_i$ , что в момент  $t+1$  стоимость этой ценной бумаги уменьшится, т.е.  $i \in I_t^-$ , то рассуждения, совершенно аналогичные предыдущим, позволяют заключить, что среднее значение стоимости этой ценной бумаги трейдер может рассматривать как дискретную случайную величину, принимающую три значения, ряд распределения которой задается следующей таблицей.

Таблица 1.2 – Ряд распределения случайной величины  $Ms_{i,t+1}, i \in I_t^-$

$Ms_{i,t+1}, i \in I_t^-$	$\frac{s_{i,t+1}^{min} + s_{i,t}}{2}$	$\frac{s_{i,t} + s_{i,t+1}^{max}}{2}$	$s_{i,t}$
$P$	$p_i$	$\frac{1-p_i}{2}$	$\frac{1-p_i}{2}$

Наконец, если трейдер считает, что ценная бумага  $i$  должна рассматриваться как ценная бумага из множества  $i \in I_t^0$ , аналогичные рассуждения позволяют рассматривать среднее значение стоимости этой ценной бумаги как дискретную случайную величину, принимающую три значения, ряд распределения которой задается следующей таблицей:

Таблица 1.3 – Ряд распределения случайной величины  $Ms_{i,t+1}, i \in I_t^0$

$Ms_{i,t+1}, i \in I_t^0$	$s_{i,t}$	$\frac{s_{i,t+1}^{min} + s_{i,t}}{2}$	$\frac{s_{i,t} + s_{i,t+1}^{max}}{2}$
$P$	$p_i$	$\frac{1-p_i}{2}$	$\frac{1-p_i}{2}$

Таким образом, в рассмотренных трех случаях, математические ожидания  $Ms_{i,t+1}$  вычисляются по следующим формулам

$$Ms_{i,t+1} = p_i \frac{s_{i,t} + s_{i,t+1}^{max}}{2} + \frac{1-p_i}{2} \frac{s_{i,t+1}^{min} + s_{i,t}}{2} + \frac{1-p_i}{2} s_{i,t}, i \in I_t^+,$$

$$Ms_{i,t+1} = p_i \frac{s_{i,t+1}^{min} + s_{i,t}}{2} + \frac{1-p_i}{2} \frac{s_{i,t} + s_{i,t+1}^{max}}{2} + \frac{1-p_i}{2} s_{i,t}, i \in I_t^-,$$

$$Ms_{i,t+1} = p_i s_{i,t} + \frac{1-p_i}{2} \frac{s_{i,t+1}^{min} + s_{i,t}}{2} + \frac{1-p_i}{2} \frac{s_{i,t} + s_{i,t+1}^{max}}{2}, i \in I_t^0.$$

Если же трейдер может выдвинуть какие-либо конкретные предположения о законе распределения значений будущих стоимостей ценных бумаг из множеств  $I_t^+, I_t^-$ , то эти предположения могут позволить вычислить математические ожидания этих значений, аналогично тому, как это было показано для закона распределения с равномерной плотностью.

Оптимальная стратегия изменения трейдером своего портфеля инвестиций в момент  $t$ , может быть найдена из решения задачи максимизации математического ожидания приращения стоимости портфеля при выполнении всех перечисленных выше ограничений, например, при выборе значения порога  $\alpha = 1/2$ , эта задача формулируется в виде

$$\begin{aligned}
M[\Delta W_{t+1}] &= \sum_{i \in I_t^0} v_{i,t} (Ms_{i,t+1} - s_{i,t}) + \sum_{i \in I_t^+} (v_{i,t} + x_{i,t}^+) (Ms_{i,t+1} - s_{i,t}) + \\
&+ \sum_{i \in I_t^-} (v_{i,t} - x_{i,t}^-) (Ms_{i,t+1} - s_{i,t}) + \sum_{i \in I_t^-} z_{i,t}^- (Ms_{i,t+1} - s_{i,t}) \rightarrow \max, \\
\sum_{i \in I_t^0} v_{i,t} Ms_{i,t+1} &+ \sum_{i \in I_t^+} [v_{i,t} + x_{i,t}^+] Ms_{i,t+1} + \sum_{i \in I_t^-} [v_{i,t} - x_{i,t}^-] Ms_{i,t+1} + \\
&+ \left( m_t - \sum_{i \in I_t^+} x_{i,t}^+ Ms_{i,t} + \sum_{i \in I_t^-} x_{i,t}^- s_{i,t} + \sum_{i \in I_t^-} z_{i,t}^- [s_{i,t} - Ms_{i,t+1}] \right) \\
&\geq \alpha \left[ \sum_{i=1}^n v_{i,t} s_{i,t} + m_t \right], \\
\sum_{i \in I_t^+} x_{i,t}^+ s_{i,t} &+ \sum_{i \in I_t^-} z_{i,t}^- s_{i,t} - \left( m_t + \sum_{i \in I_t^-} x_{i,t}^- s_{i,t} \right) \leq k_t \left( m_t + \sum_{i=1}^n v_{i,t} s_{i,t} \right), \\
x_{i,t}^- &\leq v_{i,t}, i \in I_t^-,
\end{aligned}$$

и  $x_{i,t}^+, i \in I_t^+$ ,  $x_{i,t}^-, i \in I_t^-$ ,  $z_{i,t}^-, i \in I_t^-$  – целые числа, т.е., является задачей целочисленного линейного программирования, в которой переменными являются  $x_{i,t}^+, i \in I_t^+$ ,  $x_{i,t}^-, i \in I_t^-$  и  $z_{i,t}^-, i \in I_t^-$ .

Как известно, при достаточно больших значениях целых переменных в задаче целочисленного линейного программирования, при решении прикладных задач, эти переменные рассматриваются как непрерывные, т.е., вместо задачи целочисленного линейного программирования решается задача линейного программирования, являющаяся ее релаксацией, и нецелочисленные компоненты решения затем округляются до целых значений [7] по какой-либо методике. Такое преобразование используется чаще всего тогда, когда ограничения задачи целочисленного линейного программирования нестрогие, т.е. являются ограничениями-неравенствами (что и имеет место в рассматриваемой задаче). Отметим, что проблема округления нецелых решений в релаксированной задаче линейного программирования и подход к анализу точности этого округления обсуждаются, в частности, в [8].

Таким образом, условия целочисленности переменных в приведенной выше задаче заменяются на условия неотрицательности объемов покупок/продаж ценных бумаг

$$x_{i,t}^+ \geq 0, i \in I_t^+,$$

$$x_{i,t}^- \geq 0, i \in I_t^-,$$

$$z_{i,t} \geq 0, i \in I_t^-,$$

что преобразует сформулированную задачу целочисленного линейного программирования в задачу линейного программирования.

Модель 2: трейдер может численно оценивать лишь область изменения значений для всей совокупности (портфеля) финансовых инструментов

Предположим теперь, что в момент времени  $t$  трейдер может лишь: 1) выбрать множество  $I_t^+$  ценных бумаг, стоимости которых, по его мнению, в момент времени  $t + 1$  вырастут по сравнению с их стоимостью в момент времени  $t$ , и 2) множество  $I_t^-$  ценных бумаг, стоимости которых, по его мнению, уменьшатся по сравнению с их стоимостью в момент времени  $t$ . Пред-

положим далее, что с вероятностью  $p^+ > 0.5$  трейдер верно определит, что значения стоимости ценных бумаг из множества  $I_t^+$  увеличатся, так что с вероятностью  $1 - p^+$  эти значения не увеличатся. Аналогично, предположим, что с вероятностью  $p^- > 0.5$  трейдер верно определит, что значения стоимостей ценных бумаг из множества  $I_t^-$  уменьшатся, так что с вероятностью  $1 - p^-$  эти значения не уменьшатся.

Если в портфеле трейдера есть ценные бумаги из множества  $I_t^0 = N \setminus (I_t^+ \cup I_t^-)$ , то их стоимость также может измениться к моменту времени  $t + 1$ , и, поскольку у трейдера относительно этих ценных бумаг нет никаких предположений о значениях их стоимостей в момент времени  $t + 1$ , естественно предположить, что с точки зрения трейдера увеличение и уменьшение значений стоимостей этих ценных бумаг равновероятны.

Трейдер может рассматривать выбор оптимальной стратегии инвестирования в двух ситуациях: а) формирования нового портфеля, и б) изменения состава имеющегося у него портфеля. Хотя ниже рассматривается лишь ситуация а), предлагаемая модель дает полное представление об идеях подхода к выбору оптимальной стратегии инвестирования в случае, когда трейдер может численно оценивать лишь область изменения значений для всей совокупности (портфеля) финансовых инструментов.

Пусть

$$-x_t = (x_t^+, x_t^-, x_t^0) \in X_t^+ \times X_t^- \times X_t^0 \in R_+^{|I_t^+| + |I_t^-| + |I_t^0|} - \text{вектор объемов поку-}$$

пок/продаж ценных бумаг из множества  $N$  в момент времени  $t$ ,

где  $x_t^+ \in R_+^{|I_t^+|}$  – вектор объемов покупок ценных бумаг из множества  $I_t^+$ ,

$x_t^- \in R_+^{|I_t^-|}$  – вектор объемов продаж ценных бумаг из множества  $I_t^-$ ,  $x_t^0 \in R_+^{|I_t^0|}$

– вектор объемов возможных покупок/продаж ценных бумаг из множества  $I_t^0$ ;

$-y_{t+1} = (y_{t+1}^+, y_{t+1}^-, y_{t+1}^0) \in Y_t^+ \times Y_t^- \times Y_t^0 \in R_+^{|I_t^+| + |I_t^-| + |I_t^0|}$  – вектор стоимо-

стей единиц продаваемых и покупаемых ценных бумаг из множества  $N$  в

момент времени  $t + 1$ , если трейдер верно определил направление изменения значений стоимостей этих ценных бумаг,

где  $y_{t+1}^+ \in R_+^{|I_t^+|}$  – вектор стоимостей ценных бумаг из множества  $I_t^+$ , по которой можно будет купить единицу ценной бумаги  $i \in I_t^+$ , если трейдер верно определил направления изменения значений этих стоимостей (с вероятностью  $p^+$ ),  $y_{t+1}^- \in R_+^{|I_t^-|}$  – вектор стоимостей ценных бумаг из множества  $I_t^-$ , по которой можно будет продать единицу ценной бумаги  $i \in I_t^-$ , если трейдер верно определил направления изменения значений этих стоимостей (с вероятностью  $p^-$ ),  $y_{t+1}^0 \in R_+^{|I_t^0|}$  – вектор стоимостей ценных бумаг из множества  $I_t^0$ , по которым (с вероятностью 1/2) они окажутся доступными в момент  $t + 1$ , если трейдер верно определил направления изменения значений этих стоимостей;

–  $z_{t+1} = (z_{t+1}^+, z_{t+1}^-, z_{t+1}^0) \in Z_t^+ \times Z_t^- \times Z_t^0 \in R_+^{|I_t^+| + |I_t^-| + |I_t^0|}$  – вектор цен покупки/продаж ценных бумаг из множества  $N$  в момент времени  $t + 1$ , если трейдер ошибся в своих прогнозах,

где  $z_{t+1}^+ \in R_+^{|I_t^+|}$  – вектор стоимостей ценных бумаг из множества  $I_t^+$ , по которой можно будет купить единицу ценной бумаги  $i \in I_t^+$ , если трейдер неверно определил направления изменения значений этих стоимостей (с вероятностью  $1 - p^+$ ),  $z_{t+1}^- \in R_+^{|I_t^-|}$  – вектор стоимости ценных бумаг из множества  $I_t^-$ , по которой можно будет продать единицу ценной бумаги  $i \in I_t^-$ , если трейдер неверно определил направления изменения значений этих стоимостей (с вероятностью  $1 - p^-$ ),  $z_{t+1}^0 \in R_+^{|I_t^0|}$  – вектор стоимостей ценных бумаг из множества  $I_t^0$ , по которым (с вероятностью 1/2) они окажутся доступными в момент  $t + 1$ , если трейдер неверно определил направления изменения значений этих стоимостей.

Пусть далее, в каждый момент времени  $t$  трейдер исходит из существования линейных ограничений балансового типа на компоненты вектора  $x_t$

(аналогичных ограничениям, описанным ранее), включая двусторонние ограничения – неравенства на каждую компоненту каждого из трех векторов, формирующих вектор  $x_t$ . Наличие этих ограничений позволяет заключить, что  $X_t^+$  (допустимое множество для  $x_t^+$ ),  $X_t^-$  (допустимое множество для  $x_t^-$ ) и  $X_t^0$  (допустимое множество для  $x_t^0$ ) являются выпуклыми многогранниками в пространствах  $R^{|I_t^+|}$ ,  $R^{|I_t^-|}$ ,  $R^{|I_t^0|}$ , соответственно. Аналогично,  $Y_{t+1}^+$ ,  $Y_{t+1}^-$ ,  $Y_{t+1}^0$  и  $Z_{t+1}^+$ ,  $Z_{t+1}^-$ ,  $Z_{t+1}^0$  (допустимые множества для  $y_{t+1}^+$ ,  $y_{t+1}^-$ ,  $y_{t+1}^0$  и  $z_{t+1}^+$ ,  $z_{t+1}^-$ ,  $z_{t+1}^0$ , соответственно) являются выпуклыми многогранниками в силу неотрицательности и конечности значений стоимостей ценных бумаг из множества  $N$ .

Пусть, наконец, трейдер уверен в том, что в каждый момент времени  $t$  направления изменения стоимостей ценных бумаг связаны внутри множеств  $I_t^+$  и  $I_t^-$  в том смысле, что значения стоимостей всех ценных бумаг внутри каждого из этих множеств меняются в одном направлении, т.е. либо все возрастают, либо все убывают, в то время как значения стоимостей ценных бумаг из группы  $I_t^0$  могут меняться разнонаправленно.

Взаимодействие трейдера с биржей в части разбиения им множества  $N$  на подмножества  $I_t^+$ ,  $I_t^-$  и  $I_t^0$  можно рассматривать как игру двух лиц – трейдера и биржи, в которой стратегией трейдера является выбор объемов продаваемых и покупаемых им ценных бумаг из множеств  $I_t^+$ ,  $I_t^-$  в момент  $t$  с намерением увеличить свой капитал в момент  $t + 1$ , в то время как «стратегией» биржи является «выбор» значений стоимостей единиц ценных бумаг из множества  $N$ . В силу сложности и непредсказуемости формирования значений стоимостей ценных бумаг на бирже эту игру можно рассматривать как игру трейдера с природой, в которой природа (фондовая биржа) как игрок предлагает трейдеру наиболее неблагоприятные (для выбора трейдером множеств  $I_t^+$ ,  $I_t^-$  и  $I_t^0$ ) сочетания значений стоимостей ценных бумаг из инвестиционного портфеля трейдера в момент  $t + 1$  в виде векторов из выпуклых многогранников  $Y_{t+1}^+$ ,  $Y_{t+1}^-$ ,  $Y_{t+1}^0$  и  $Z_{t+1}^+$ ,  $Z_{t+1}^-$ ,  $Z_{t+1}^0$ . Такая игра имеет определ-

ленную структуру, позволяющую отыскивать оптимальные стратегии игроков из решения задачи линейного программирования [9].

Теорема 1.1. В каждый момент времени  $t$  взаимодействие между трейдером и биржей может быть описано в форме антагонистической игры на выпуклых многогранниках несвязанных стратегий игроков  $M_t = \{x_t \in R_+^n : B_t x_t \geq d_t\}$  и  $\theta_{t+1} = \{w_{t+1} \in R_+^{2n} : A_t w_{t+1} \geq b_t\}$  с платежной функцией  $\langle x_t, D_t w_{t+1} \rangle$ ,

где

$D_t$

$$= \begin{pmatrix} D^{|I_t^+|}(p^+) & D^{|I_t^+|}(1-p^+) & 0 & 0 & 0 & 0 \\ 0 & 0 & D^{|I_t^-|}(p^-) & D^{|I_t^-|}(1-p^-) & 0 & 0 \\ 0 & 0 & 0 & 0 & D^{|I_t^0|}\left(\frac{1}{2}\right) & D^{|I_t^0|}\left(\frac{1}{2}\right) \end{pmatrix},$$

$x_t = (x_t^+, x_t^-, x_t^0) \in M_t = X_t^+ \times X_t^- \times X_t^0$ ,

$w_{t+1} = (w_{t+1}^+, w_{t+1}^-, w_{t+1}^0) \in \theta_{t+1} = \theta_{t+1}^+ \times \theta_{t+1}^- \times \theta_{t+1}^0$ ,

$D_t$  – матрица размера  $(|I_t^+| + |I_t^-| + |I_t^0|) \times 2(|I_t^+| + |I_t^-| + |I_t^0|)$ ,

$D^{|I|}(x)$  – диагональная матрица размера  $|I|$ , у которой все элементы главной диагонали равны  $x$ ,

$M_t$  – множество стратегий трейдера,

$\theta_{t+1}$  – множество стратегий фондовой биржи,

$\theta_{t+1}^+ = Y_{t+1}^+ \times Z_{t+1}^+$ ,  $\theta_{t+1}^- = Y_{t+1}^- \times Z_{t+1}^-$ ,  $\theta_{t+1}^0 = Y_{t+1}^0 \times Z_{t+1}^0$  – выпуклые многогранники и  $w_{t+1}^+ = (y_{t+1}^+, z_{t+1}^+) \in \theta_{t+1}^+$ ,  $w_{t+1}^- = (y_{t+1}^-, z_{t+1}^-) \in \theta_{t+1}^-$ ,  $w_{t+1}^0 = (y_{t+1}^0, z_{t+1}^0) \in \theta_{t+1}^0$  – векторы, и седловая точка этой игры может быть найдена из решения задач линейного программирования, образующих двойственную пару.

Замечание 1.1. Нетрудно показать, что в ситуации б) когда трейдер отыскивает оптимальные стратегии инвестирования в ценные бумаги, исходя из имеющегося у него в момент времени  $t$  портфеля ценных бумаг, рассмотренная игра формулируется как игра на выпуклых многогранниках  $\tilde{M}_t = \{x_t \in R_+^n : \tilde{B}_t x_t \geq \tilde{d}_t\}$  и  $\tilde{\theta}_{t+1} = \{w_{t+1} \in R_+^{2n} : \tilde{A}_t w_{t+1} \geq \tilde{b}_t\}$  с платежной функцией

цией  $\langle x_t, D_t w_{t+1} \rangle + \langle q, w_{t+1} \rangle$ , где  $q \in R_+^{2n}$  – некоторый фиксированный вектор. Как показано в [9], седловая точка в такой игре находится из решения задач линейного программирования

$$\begin{aligned}\langle \tilde{b}_t, h_t \rangle &\rightarrow \max_{(h_t, x_t) \in \tilde{Q}_t}, \\ \langle -\tilde{d}_t, s_{t+1} \rangle + \langle q, w_{t+1} \rangle &\rightarrow \min_{(s_{t+1}, w_{t+1}) \in \tilde{P}_{t,t+1}},\end{aligned}$$

где  $\tilde{Q}_t = \{(h_t, x_t) \geq 0 : h_t \tilde{A}_t \leq q + x_t D_t, \tilde{B}_t x_t \geq \tilde{d}_t\}$ ,  $\tilde{P}_{t,t+1} = \{(s_{t+1}, w_{t+1}) \geq 0 : s_{t+1} \tilde{B}_t \leq -D_t w_{t+1}, \tilde{A}_t w_{t+1} \geq \tilde{b}_t\}$ , образующих двойственную пару.

Замечание 1.2. Хотя в обеих моделях рассматривалась ситуация, когда формирование и изменение портфеля трейдера осуществлялось только из числа ценных бумаг и не включало производные финансовые инструменты (опционы и фьючерсы), можно показать, что обе предложенные модели могут быть использованы для формирования и решения задачи отыскания оптимального портфеля с учетом интересующих трейдера производных финансовых инструментов ( опционов и фьючерсов).

Замечание 1.3. Следует заметить, что так как по смыслу задачи все компоненты вектора  $x_t$  принимают целые значения, значение максимина  $\max_{x_t \in X_t} [\min_{w_{t+1} \in \theta_{t+1}} \langle x_t, D_t w_{t+1} \rangle]$ , которое достигается в седловой точке платежной функции из решения игры является, вообще говоря, лишь верхней границей значения указанного максимина. Однако, используя примененную при доказательстве Теоремы 1.1 схему рассуждения, а также тот факт, что множество точек выпуклого многогранника  $M_t \times \theta_{t+1}$  с целыми координатами является конечным, нетрудно убедиться в том, что отыскание точного значения минимакса сводится к максимизации линейной функции в задаче смешанного математического программирования с линейными ограничениями [10, 11]. С другой стороны, ясно, что решение такой задачи смешанного программирования с интересным для практики числом целочисленных переменных за приемлемое время может представить существенные трудности, в то время как решение задач линейного программирования даже большого размера при отыскании седловой точки в рассматриваемой игре не представ-

ляет никаких вычислительных трудностей. Кроме того, в теоретических исследованиях проблем взаимодействия трейдеров с биржей (к которым относится и настоящая работа), традиционно (см., например, упомянутую работу Марковица [2]), объемы приобретаемых и продаваемых акций считаются непрерывными переменными, принимающими действительные, неотрицательные значения. Наконец, поскольку коэффициенты системы линейных уравнений и неравенств, описывающих выпуклые многогранники, как правило, известны лишь приближенно, замена задачи отыскания точного значения этого максимума задачей отыскания его верхней границы представляется оправданной.

### 1.1.2 Некоторые другие исследования организационных моделей и механизмов

А. Развитие теоретико-игрового подхода к анализу рынка мощности и рынка электроэнергии в регионе

В [12, 13] рассмотрена проблема построения элементов инструментария для количественного анализа перспектив развития системы электроснабжения региона, исходя из оценки администрацией региона существующего уровня обеспеченности электроэнергией населения и предприятий региона и тенденций развития этого региона. Предложена математическая модель, с использованием которой перспективы инвестирования в строительство новых электростанций и хранилищ электроэнергии в регионе оцениваются в рамках бескоалиционной игры двух лиц на выпуклых многогранниках с платежной функцией одного из игроков в виде суммы билинейной функции и линейной функции векторных аргументов и с платежной функцией другого игрока в виде билинейной функции тех же векторных аргументов. В этой игре одним из игроков являются все производители электроэнергии в регионе, которые рассматривают целесообразность инвестирования в строительство новых электростанций и хранилищ электроэнергии и/или новых блоков дей-

ствующих электростанций и хранилищ электроэнергии в течение некоторого периода времени (более одного года). Другим игроком являются все крупные потребители электроэнергии в регионе, которые взаимодействуют с производителями в рамках двухсторонних долгосрочных контрактов на производство и поставку электроэнергии. Установлены проверяемые необходимые и достаточные условия равновесия в рассматриваемой игре, позволяющие отыскивать приемлемые для обеих сторон объемы инвестирования в систему электроснабжения региона и договорные цены на электроэнергию для потребителей из решения задач линейного программирования. Эта возможность позволяет использовать предложенную модель для соответствующих количественных расчетов по анализу перспектив развития системы электроснабжения региона с любой интересной для администрации региона степенью детализации взаимодействия региональных производителей и потребителей электроэнергии.

В [14, 15] взаимодействие всех генераторов электроэнергии в регионе (в том числе, с системами, генерирующими электроэнергию из энергии солнца и ветра) с ее потребителями в регионе (непосредственно и через системы хранения электроэнергии) и с компанией, обеспечивающей передачу электроэнергии к потребителям, в виде некооперативной игры конечного числа игроков на выпуклых многогранниках связанных стратегий этих игроков. Установлены проверяемые необходимые и достаточные условия равновесия во вспомогательной игре трех лиц на выпуклых многогранниках несвязанных стратегий игроков и показано, что точки равновесия в этой вспомогательной игре являются точками равновесия в исходной игре многих лиц. Доказано, что точки равновесия во вспомогательной игре могут быть найдены из решения трех задач линейного программирования, две из которых образуют двойственную пару.

## Б. Построение математических моделей для анализа и выбора схемы развития крупномасштабных транспортных систем

В [16, 17] рассмотрена проблема размещения и оптимизации использования новых и существующих грузовых транспортных узлов в регионе и подъездных путей к ним. Эта задача математически сформулирована как задача отыскания минимума функции векторных аргументов на выпуклом многограннике в конечномерном пространстве. В этой задаче минимизируемая функция является функцией максимума двух векторных аргументов, каждый из которых принадлежит подмножеству некоторого многогранника являющегося выпуклой оболочкой векторов, у которых все компоненты целочисленные. Доказана возможность отыскания указанного минимума этой функции (т.е. минимакса) из решения вспомогательной задачи смешанного математического программирования, у которой все целочисленные переменные булевы.

## В. Развитие теоретико-игрового подхода к анализу проблем государственно-частного партнерства

В [18] рассмотрены две проблемы государственно-частного партнерства, в которых государство, заинтересованное в выполнении конечного набора крупномасштабных проектов, пытается привлечь к их выполнению частный капитал. Первая проблема возникает в случае, когда государство не имеет достаточных средств для выполнения всех проектов и ищет выгодные для себя и привлекательные условия для инвестирования частного капитала в выполнение этих проектов. Вторая проблема связана с отысканием резервных цен на тендеры по тем проектам, которые государство может полностью финансировать. Доказано, что при некоторых естественных предположениях обе проблемы математически формализуются в виде игр на множествах допустимых стратегий игроков, являющихся выпуклыми многогранниками,

причем эти стратегии оказываются связанными в первой из проблем и несвязанными во второй проблеме.

В [19] рассмотрена проблема переработки твердых мусорных отходов в биомассу в масштабах большого города, где городская администрация намеревается построить заводы по переработке мусора в местах действующих городских мусорных свалок, в которых в настоящее время мусор сжигается. Для объявления тендеров а) на постройку зданий для указанных заводов и б) на поставку технологии и оборудования для этих заводов, городской администрации необходимо определить начальные (резервные) цены на выполнение обоих видов работ. Задача отыскания таких цен сформулирована в виде игры трех лиц на выпуклых многогранниках несвязанных стратегий игроков. Доказано, что точки равновесия в этой игре, определяющие указанные резервные цены, могут быть найдены из решения вспомогательных задач линейного программирования, образующих двойственную пару.

### 1.1.3 Некоторые выводы

Основным результатом проведенных исследований следует, по-видимому, считать продемонстрированную возможность моделирования и анализа функционирования широкого спектра крупномасштабных прикладных систем в условиях неопределенности в рамках класса игр на выпуклых многогранниках несвязанных стратегий игроков и на их подмножествах, образованных векторами, все координаты которых целочисленные. Целесообразность и перспективность такого подхода определяются возможностью отыскания равновесия в играх рассматриваемого класса из решения конечного числа задач линейного программирования или смешанного математического программирования, решения которых обладают огромными вычислительными возможностями.

В то же время анализ систем, модели которых рассмотрены в настоящем разделе отчета, показывает, что необходимо попытаться расширить класс игровых задач, в которых равновесия могут быть найдены конечными методами.

ми, обладающими значительными вычислительными возможностями, в первую очередь, методами линейного и смешанного программирования, а также методами квадратичного программирования (для специальных классов задач). Именно в этом направлении предполагается продолжить основные исследования в 2016 году. В частности, предполагается исследовать новые классы игр, в которых стратегии игроков оказываются связанными.

Результаты исследования в рамках настоящего подраздела отчета отражены в работах [10, 11, 20, 21].

## 1.2 Исследования пространственных моделей взаимодействия в экономической и политической сферах

В данном подразделе отчета приведены результаты исследований пространственных моделей взаимодействия между различными участниками в экономической и политической сферах.

Исследование пространственных моделей было осуществлено на примере арктического региона. Интерес к данной территории объясняется тем, что, несмотря на то, что она составляет всего 6% поверхности Земли, на ее территории по различным оценкам находится примерно 20% всех неразведанных запасов нефти и газа. Благодаря таянию льдов Арктика становится более доступной для добычи этих ресурсов. Более того, мировая потребность в нефти и газе возрастает, что подталкивает компании к использованию арктических гидрокарбоновых запасов. К тому же, развитие технологий, строения кораблей, оборудования для бурения, а также логистики уменьшают проблемы, связанные с доступностью шельфовых месторождений [22]. Кроме того, уменьшение общей площади поверхности льда увеличивает рыболовный сезон. Это также предоставляет новые возможности для использования арктических морских путей [23].

Точные границы стран в Арктике не были определены. В 1982 г. была принята конвенция ООН по морскому праву (UNCLOS), согласно которой к территории стран, имеющими выход к Арктическому океану, относится 12-

милльная зона (около 22 км), а исключительная экономическая зона (ИЭЗ) находится в пределах 200 морских миль от берега (около 370 км). В ИЭЗ государство имеет право на добычу и управление природными ресурсами. Оно также контролирует создание искусственных островов и других сооружений, ведение научных исследований и защиту экологической среды. Воды к северу от ИЭЗ являются международными, а их ресурсы являются общим наследием человечества. Тем не менее, в связи с выходом континентального шельфа за пределы ИЭЗ каждое государство может подать заявку на расширение данной зоны. Дания, Канада, Норвегия, Россия и США уже подали ряд заявок. Решение комиссии ООН по границам континентального шельфа может очень сильно повлиять на экономики стран. Если территория признается частью континентального шельфа и, как следствие, частью ИЭЗ страны, то ни одна другая страна не будет иметь права на минеральные и рыбные ресурсы на данной территории. Более того, отнесение ранее международных вод, через которые проходят арктические морские пути, к какой-то стране может увеличить стоимость использования этих путей другими государствами, а также международными компаниями.

Все эти факторы обуславливают важность вопроса об определении территорий, на которых возможен конфликт интересов стран, что стало первым направлением данного исследования.

Основной целью данного направления исследований был анализ интенсивности потенциальных конфликтов интересов стран в арктическом регионе, а также исследование различных сценариев развития событий на данной территории. Для начала, приведем краткую географическую информацию о рассматриваемых странах.

### 1.2.1 Географическая информация о странах

С точки зрения хозяйственно-экономических и военно-политических целей интерес к арктическому региону проявляют ряд стран. Прежде всего, это восемь так называемых арктических стран: Дания, Исландия, Канада,

Норвегия, Россия, США, Финляндия и Швеция. Пять из этих стран имеют выход к Арктическому океану.

По инициативе Финляндии в 1996 г. был создан международный форум – Арктический Совет. Он состоит из восьми арктических стран: Дании (включая Гренландию и Фарерские острова), Исландии, Канады, Норвегии, Финляндии, России, США, Швеции и Финляндии. Кроме того, двенадцать неарктических стран имеют статус наблюдателей (Великобритания, Германия, Индия, Испания, Италия, КНР, Республика Корея, Нидерланды, Польша, Сингапур, Франция, Япония).

Российское арктическое побережье является наибольшим и составляет 22 600 км (примерно 60% всей береговой линии Арктики). Россия омывается Баренцевым, Белым, Карским, Восточно-Сибирским и Баренцевым морями, а также морем Лаптевых. Эти моря не только богаты нефтяными, газовыми и другими минеральными ресурсами, через них также проходит Северный морской путь (СМП), являющийся кратчайшим между Азией и Европой. Берингов пролив разделяет Чукотку (Россия) и Аляску (США). Хотя американское побережье не является протяженным, к ИЭЗ США относятся наибольшие запасы нефти в регионе. Через Чукотское море и море Бофорта проходит часть Северо-Западного прохода, соединяющего Азию с Северной Америкой. Вторая его часть пролегает через канадские воды морей Баффина и Лабрадор. Дания представлена в Арктике островом Гренландия, который омывается этими двумя морями на западе и Гренландским – на востоке. Наконец, Исландия расположена южнее Полярного круга, а Швеция и Финляндия отделены от Арктического океана норвежскими территориями.

В исследовании также рассматривались три неарктические страны: КНР, Республика Корея и Япония, которые проявляют всё увеличивающийся интерес к региону. Он касается не только ресурсов, но и научных исследований. Так, Япония имеет пятидесятилетнюю историю полярных исследований, Япония и Китай основали институты, специализирующиеся на таких исследованиях, в 1990х и 2000х годах, соответственно. Республика Корея прини-

мает не менее активное участие, ее научная исследовательская станция «Дасан» работает на архипелаге Шпицберген с 2002 года. Важно заметить, что все три страны основывают свое участие на UNCLOS [24, 25].

Список рассматриваемых стран приведен в Таблице 1.4.

Таблица 1.4 – Базовая информация о странах

Страна	Регион	Площадь, тыс. км <sup>2</sup>	Население, млн. чел.	ВВП, \$ млрд.	ВВП на душу населения, \$ тыс.
Канада	Арктика	9 900	36	1 518	43
Дания	Арктика	2 210	5.7	347	44
Финляндия	Арктика	338	5.4	276	50
Исландия	Арктика	103	0.3	14	44
Норвегия	Арктика	385	5.1	353	68
Россия	Арктика	17 000	144	3 559	24
Швеция	Арктика	450	9.7	573	58
США	Арктика	9 800	320	16 720	52
КНР	Не Арктика	9 600	1 350	16 150	12
Япония	Не Арктика	377	126	4 800	38
Р. Корея	Не Арктика	100	51	1 790	35

### 1.2.2 Взаимные интересы стран в регионе

Увеличивающийся интерес к арктическому региону вызван объемами природных ресурсов, сконцентрированных в регионе, а также новыми технологиями, позволяющими их добывать. Хотя интересы могут быть противоположными (например, страна может стремиться увеличить/уменьшить объемы добычи ресурса на определенной территории по различным причинам), наиболее ценными ресурсами остаются нефтяные, газовые, рыбные и судоходные [26, 27].

#### Нефть и газ

Залежи нефти были обнаружены в канадской Арктике еще в XVIII веке. Ранние разведывательные бурения относятся к 1920м годам. Разведка возросла в 1940-х и 1950-х годах и переместилась на шельф в 1972 году. В 1991 году началась добыча нефти на одном из самых больших по запасам шельфовом участке около берегов острова Ньюфаундленд. В 1960-х существенные открытия были сделаны в Ямalo-Ненецком автономном округе (Россия). В

1977 г. началась добыча на месторождениях Аляски (США). В 1980-х были обнаружены залежи ископаемых в море Бофорта, однако добыча так и не началась в связи с ее огромной стоимостью. В 1980-х были пробурены геологоразведочные скважины на Фарерских островах в Дании, нефть была обнаружена в Баренцевом море недалеко от Норвегии. В то же время исследовались и запасы газа. Так, значительные залежи газа были обнаружены в Баренцевом море (Россия) в 1988 году [28, 29].

В 2008 году геологическая служба США (USGS) измерила вероятность обнаружить запасы нефти и газа в будущем. Результаты этой работы по оцениванию перспектив нефтегазоносности Арктического региона (33 региона к северу от  $66^{\circ}$  с.ш.) показаны на рисунках 1.1 и 1.2.

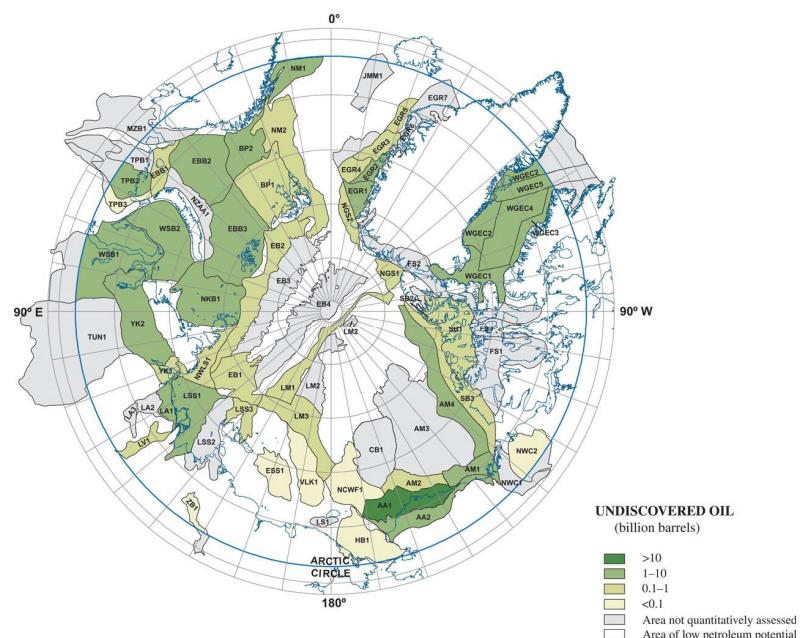
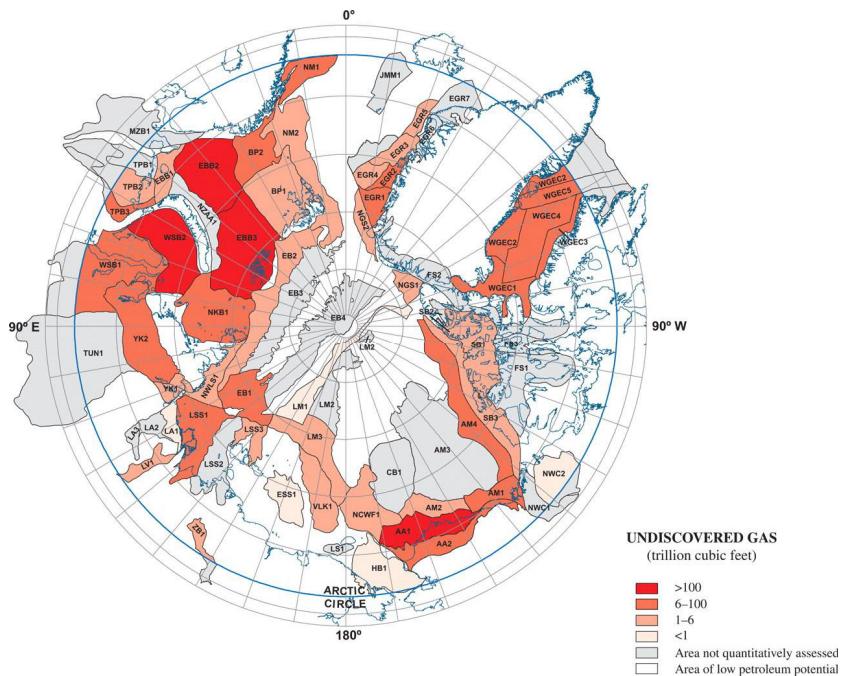


Рисунок 1.1 – Неразведанная нефть

Результаты показали, что в 25 регионах с более чем 10%-ой вероятностью ожидается наличие существенных неразведанных залежей. Существенными называют залежи объемом как минимум в 50 миллионов баррелей условного топлива. Объемы неразведанных месторождений нефти и газа оценивались только для этих 25 регионов.



## Рисунок 1.2 – Неразведанный газ

Основными игроками в шельфовой добыче нефти и газа являются пять арктических стран, поскольку большинство регионов лежат в их ИЭЗ. Согласно [22] распределение ресурсов по странам далеко неравномерное. Российская ИЭЗ охватывает 52% всех ресурсов, расположенных не в международных водах, американская – 20%, норвежская – 12%, датская – 11% и канадская – 5%. Неарктические страны (КНР, Республика Корея и Япония) соблюдают UNCLOS и ИЭЗ, однако в их интересах, чтобы международные воды оставались международными [25].

## Рыбные запасы

Экономики некоторых стран Арктики сильно зависят от экспорта рыбы и морепродуктов. Именно поэтому необходимо защищать природную среду обитания и контролировать объемы вылова рыбы, поскольку слишком большой вылов может привести к исчезновению вида. Более того, это может оказать серьезное влияние и на виды, не подлежащие вылову, в силу взаимодействия хищника и жертвы и цепей питания. Эти эффекты уже присутствуют, хотя и не в серьезных масштабах [30].

Климатические изменения в Арктике имеют ряд последствий. С увеличением температуры воды и таянием ледников, соленость воды уменьшается. Из-за уменьшения площади поверхности льдов появляются ветры и штормы, а также другие океанографические и метеорологические изменения. Все они, несомненно, влияют на экосистемы, однако это влияние очень трудно предсказать заранее. Естественная среда обитания для одних типов рыбы увеличится, в то время как для других – уменьшится. Однако невозможно предсказать, какие именно территории выиграют или проигрывают от качественных и количественных изменений в составе арктических морских экосистем [31].

Таким образом, возникает возможность конфликта интересов между странами, связанного со смещением сред обитания и риском потери способа заработка. На рисунке 1.3 показаны области обитания и вылова наиболее популярных видов рыб.

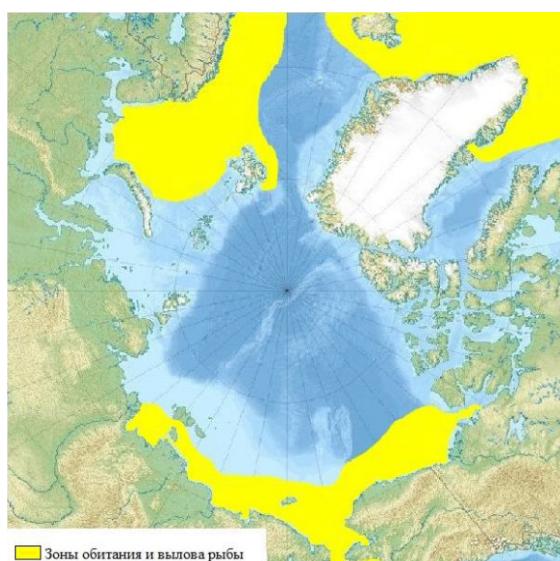


Рисунок 1.3 – Области обитания и вылова наиболее популярных видов рыб

## Мореходные ресурсы

Климатические изменения делают территории, ранее покрытые льдом, свободными от него. Поэтому все большее количество территорий становится доступным для международного судоходства. Наибольшее увеличение в судоходстве, как ожидается, будет связано с разработкой шельфовых место-

рождений, перевозки оборудования к месту добычи и ресурсов обратно на берег, а также с увеличением трансконтинентального транзитного плавания. Основные арктические пути показаны на рисунке 1.4.

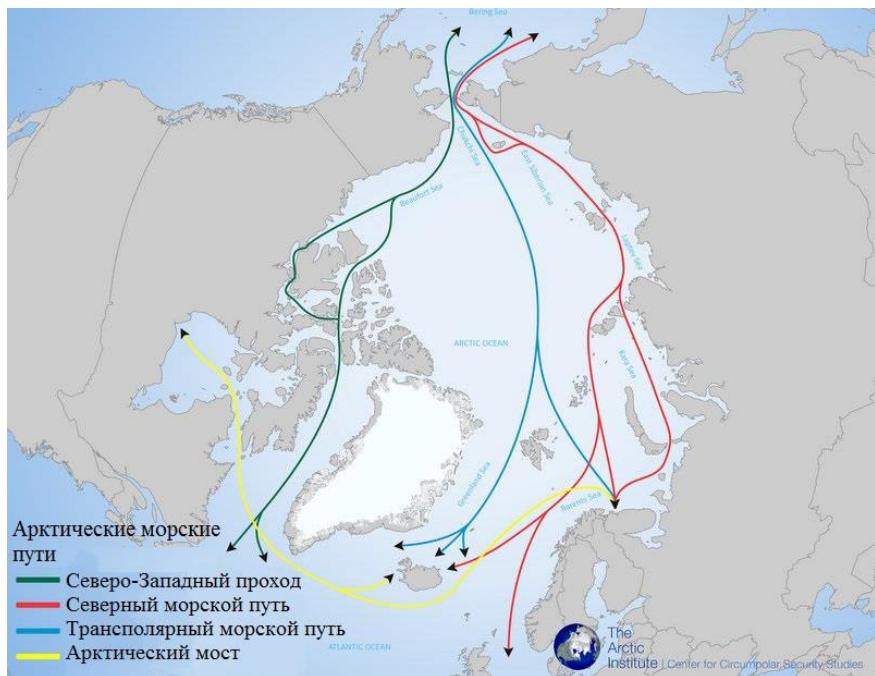


Рисунок 1.4 – Арктические морские пути

Существует четыре арктических морских пути (АМП): Северный морской путь (СМП), Северо-Западный проход (СЗП), Трансполярный морской путь (ТМП) и Арктический мост (АМ). Летом 2011 года через СМП было перевезено 850 000 тонн грузов. СМП является наиболее доступным на данный момент путем через Арктику. Для сравнения через Суэцкий канал в том же году прошло 690 миллионов тонн грузов. Общий объем грузов, перевезенных с использованием арктических морских путей, составил в 2011 году 3.1 миллион тонн, этот объем, согласно оценкам, увеличится до 50 миллионов тонн к 2020 году [32, 33].

СМП и СЗП представляют интерес, поскольку соединяют Атлантический и Тихий океаны. Ожидается, что снижение количества льда в Арктическом океане будет неравномерным. Так, оно займет больше всего времени на территории Канадского Арктического архипелага и возле Гренландии. Более того, только 2 из 6 основных путей СЗП подходят для кораблей неограниченного морского плавания, поскольку у остальных четырех максимальная глубина

бина составляет не более 10 метров, и они слишком узкие для таких судов. Наибольшая доля снижения площади ледяного покрова затронет центральную и восточную части Арктического океана, увеличивая период навигации по СМП.

Заметим, что АМП являются более короткими по сравнению с путями через Суэцкий и Панамский каналы. Например, расстояние между Санкт-Петербургом и Владивостоком сокращается на 40% с более чем 23 000 км до примерно 14 000, между Мурманском и Японией на 11 000 км (65%), Шанхаем и Гамбургом – на 5 200 км. Более того, в связи с пиратством стоимость страховок сильно возросла, некоторые компании даже приняли решение о том, чтобы оплывать Африку с юга вместо того, чтобы направляться в сторону Суэцкого канала [34].

Все эти причины делают АМП экономически привлекательными как для арктических, так и для неарктических стран, поскольку первым они приносят доход (в связи со стоимостью транзита), а вторым предоставляют возможность снизить расходы.

### 1.2.3 Используемая модель

Для решения задачи использовалась модель, состоящая из трех этапов:

- построение функций полезности;
- поиск зон конфликтов с их помощью;
- исследование зон конфликтов с учетом различных сценариев.

Источником информации для данного исследования были данные о четырех типах ресурсов, находящиеся в общем доступе. Заинтересованность стран в нефти и газе моделируется на основании отчета USGS. Все регионы делятся на пять категорий:

- 1) крайне высокого интереса с более чем 100 трлн. кубических футов неразведанного газа и более чем 10 млрд. баррелей неразведенной нефти;

- 2) высокого интереса с 6-100 трлн. кубических футов неразведанного газа и 1-10 млрд. баррелей неразведанной нефти;
- 3) среднего интереса с 1-6 трлн. кубических футов неразведанного газа и 0.1-1 млрд. баррелей неразведанной нефти;
- 4) низкого интереса с менее чем 1 трлн. кубических футов неразведанного газа и менее чем 0.1 млрд. баррелей неразведанной нефти;
- 5) безынтересные территории (запасы очень малы или не были оценены).

Заинтересованность в рыбе вычисляется исходя из данных об их областях обитания и вылова (см. рисунок 1.3). Для оценки важности морских путей используются результаты моделирования наиболее оптимальных маршрутов на данный момент и к середине XXI века [35]. На рисунке 1.5 изображены оптимальные трансатлантические пути к 2040-2059 годам.

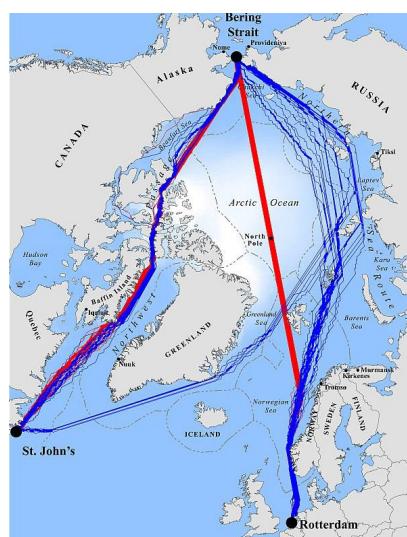


Рисунок 1.5 – Оптимальные навигационные пути

#### 1.2.4 Функции полезности стран

На основании входных данных строятся функции полезности стран. На рисунке 1.6 показано представление функции полезности. Это матрица, в которой каждый элемент (каждая ячейка) соответствует конкретной территории. Значение элемента показывает важность данной территории для страны.

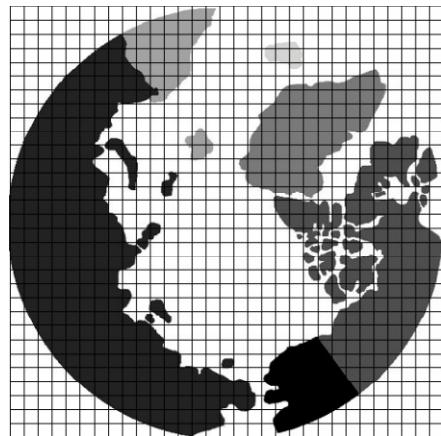


Рисунок 1.6 – Представление функции полезности

Для каждого из ресурсов строится отдельная функция полезности. Например, функция  $f_{\text{Россия,газ}}$ (территория) показывает, насколько ценной является территория для России с точки зрения запасов газа. Соответственно, для каждой страны строится по 4 матрицы в соответствии с рассматриваемыми ресурсами. Функции полезности строились согласно следующим соображениям.

1. Полезность газа, нефти и рыбных ресурсов зависит от запасов этих ресурсов на рассматриваемой территории. В то же время, она убывает с увеличением расстояния между территорией и страной.

$$f_{\text{страна,ресурс}}(t) = \frac{\text{ценность ресурса}(t)}{h(\text{расстояние}(t))},$$

где ценность ресурса определяется источниками информации,  $t$  – территория, расстояние ( $t$ ) – удаленность данной территории от страны,  $h(\cdot)$  - некоторая монотонная функция.

2. Полезность морских путей для Арктической страны высока, если путь проходит через ее исключительную экономическую зону. Пусть АС – арктическая страна, нАС – неарктическая страна, м – мореходный ресурс. Тогда:

$$f_{AC,m}(t) = \begin{cases} \text{ценность ресурса}(t), t \in \text{ИЭЗ страны} \\ \frac{\text{ценность ресурса}(t)}{h(\text{расстояние}(t))}, t \in \text{межд. зоне.} \end{cases}$$

Неарктические страны заинтересованы в арктических морских путях, поскольку они являются гораздо более короткими по сравнению с путями, проходящими через Суэцкий канал или к югу от Африки. Таким образом, интерес неарктических стран никак не связан с удаленностью пути от их границ. Он полностью объясняется выгодой от использования пути, которая формируется в результате разности цен и расстояний и зависит от объема межконтинентального экспорта и импорта страны.

$$f_{HAC,M}(t) = \frac{\text{ценность маршрута } (t)}{C},$$

где  $C$  – константа, характеризующая важность трансатлантической транспортировки грузов для данной страны. Например, ее можно считать пропорциональной объему перевозок данной страны.

### 1.2.5 Агрегирование функций полезности

Функции полезности строились для анализа возможной напряженности в различных регионах Арктики. Для оценки напряженности было решено использовать двухэтапное правило порогового агрегирования. Оно построено аксиоматически, также показано, что оно является лексикографическим правилом, примененным определенным образом [36 – 38]. Сначала оно применялось для каждого ресурса в отдельности, чтобы понять, какие участки могут вызвать споры именно в связи с наличием там конкретного ресурса. Затем агрегирование производилось повторно с целью выявить зоны конфликта интересов всех стран.

Одним из принципиальных отличий правила порогового агрегирования является его пороговая некомпенсируемость. Пусть определенная территория была оценена двумя странами с точки зрения ее значимости по нескольким критериям в шкале от 1 до 5. Получены оценки  $O_1 = (o_{11}, o_{12}, \dots, o_{1n})$  и  $O_2 = (o_{21}, o_{22}, \dots, o_{2n})$ . Необходимо определить, для какой страны данная территория более значима. Если все оценки первой страны низкие, а одна из оценок второй страны высокая, можно сделать вывод, что интерес второй

страны к территории выше. Даже если страна сильно заинтересована во всего одном ресурсе, то ее суммарный интерес к территории считается высоким. Взяв в качестве итоговой меры заинтересованности сумму оценок, мы можем получить следующую ситуацию: по нашему мнению,  $O_1 = \underbrace{(1,1,\dots,1)}_6 < (5, \underbrace{0,\dots,0}_5) = O_2$ , но  $\sum_i o_{1i} = 6 > \sum_i o_{2i} = 5$ . Таким образом, слабый интерес первой страны к ресурсам 2-6 компенсировал отсутствие сильного интереса к первому ресурсу, что нежелательно. Пороговая некомпенсируемость не дает с помощью высоких оценок по одним критериям компенсировать низкие оценки по другим критериям.

Итак, все территории упорядочиваются согласно соответствующим им девятимерным векторам, которые предварительно сортируются по убыванию. После упорядочивания территориям присваивается ранг. Так, территории в начале списка получают ранг  $m$ , поскольку соответствуют зонам сильнейших разногласий. Чем дальше по списку находится территория, тем ниже ее ранг или «конфликтность».

Когда для каждого ресурса построена «карта конфликтности», то есть для каждой территории в Арктике определено, конфликт какой интенсивности в ней возможен, то правило агрегирования применяется повторно. На этот раз каждой категории соответствует четырехмерный вектор. В итоге все территории делятся на несколько классов «конфликтности» от высокой до несущественной. Схема данного алгоритма приведена на рисунке 1.7.

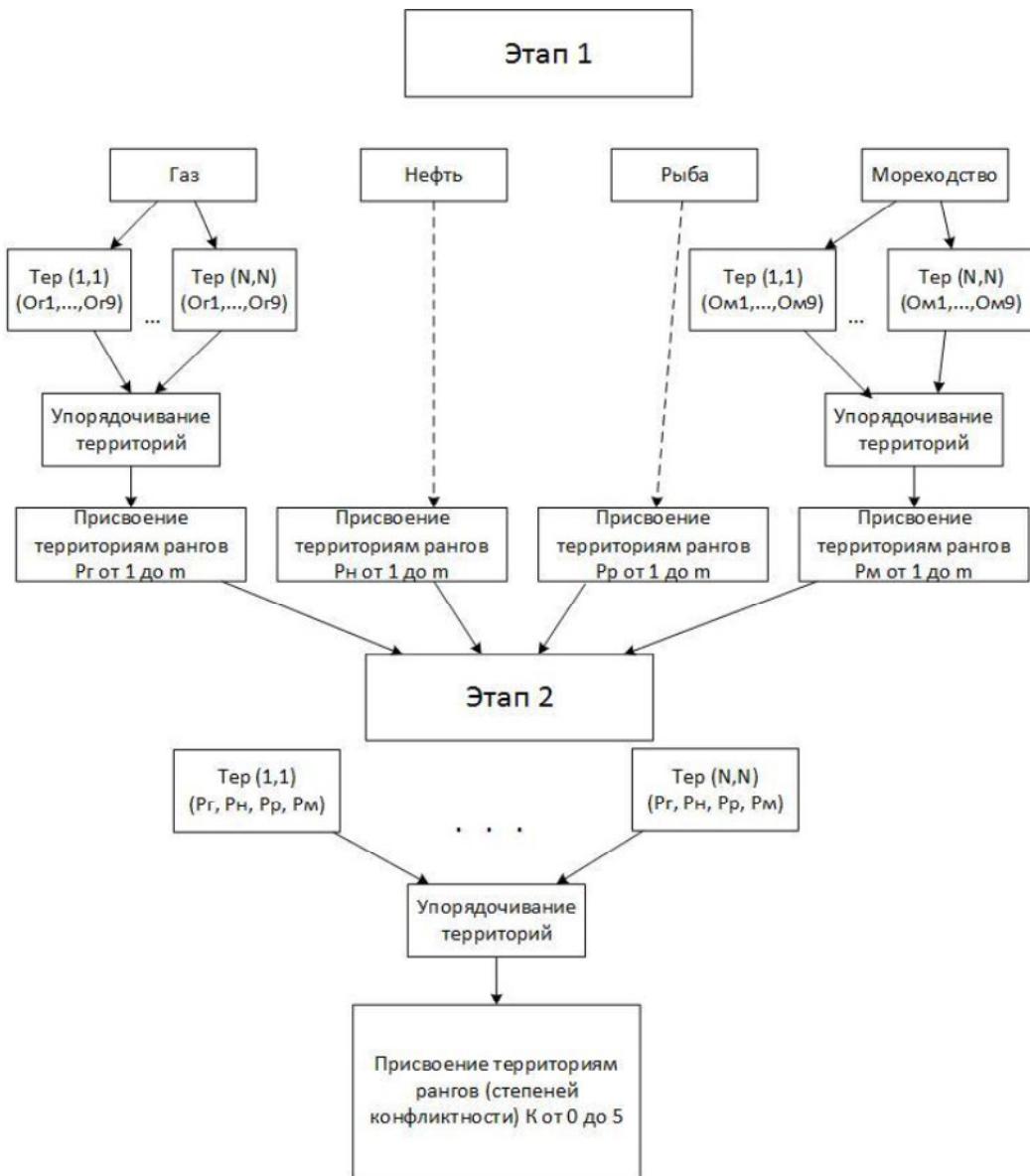


Рисунок 1.7 – Структура модели

### 1.2.6 Исследование зон конфликтов, сценарии

В пункте 1.2.4 функции полезности описаны в общем виде. Так, функция  $h(\cdot)$  и константа  $C$  могут отличаться для разных стран и ресурсов. Более того, можно применить взвешенное пороговое агрегирование, когда один из ресурсов считается более важным, чем другой.

Также, хотя формально иностранные государства не могут претендовать на ресурсы, принадлежащие ИЭЗ другой страны, их интересы могут быть очень сильными. Более того, страна, не заинтересованная в добыче какого-то ископаемого, может, тем не менее, отмечать данный ресурс как важный. На-

пример, Исландия – страна, сильно зависящая от экспорта рыбы, но не сильно заинтересованная в нефтяных ресурсах. Сравнительно недалеко от нее находится Норвегия, на континентальном шельфе которой ведется разработка нефтяных месторождений. Казалось бы, Исландию не должен никак волновать этот факт, однако в случае разлива нефти последствия для нее будут катастрофическими. Таким образом, Исландия заинтересована в норвежских запасах нефти, а, точнее, в том, чтобы иметь возможность влиять на стандарты добычи и безопасности.

Именно этим объясняется необходимость рассматривать сценарии, то есть различные ситуации, которые могут изменить итоговое распределение зон высокого или низкого уровня конфликта интересов.

Рассмотрим, например, результаты, которые отличаются только тем, что странам разрешено с некоторой интенсивностью интересоваться запасами нефти и газа в границах ИЭЗ других стран. При  $\alpha = 1$  страны воспринимают ИЭЗ как международные воды, а при  $\alpha = 0$  как неприкосновенную территорию иностранного государства. На рисунке 1.8 изображены полученные результаты.

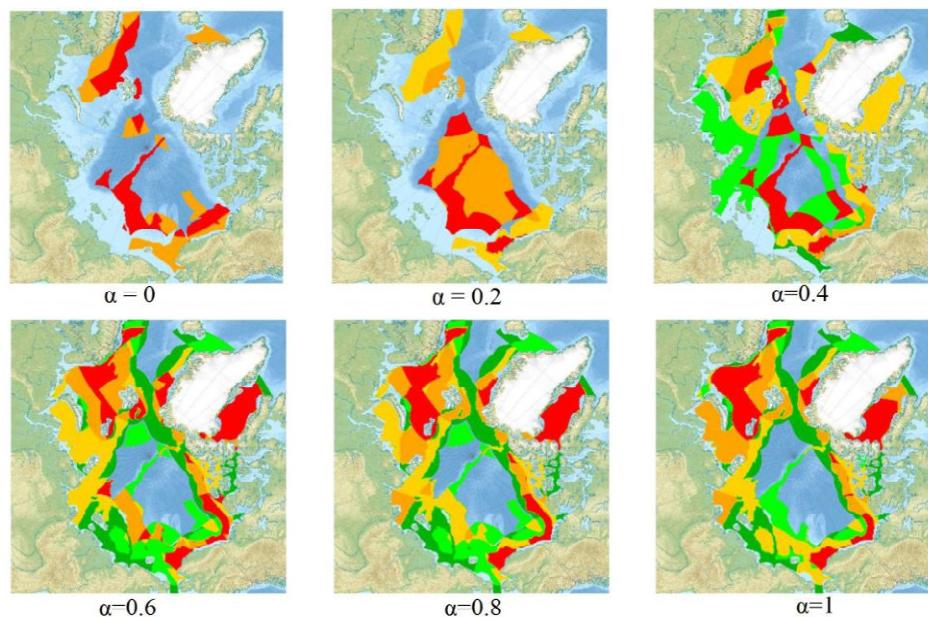


Рисунок 1.8 – Интенсивность конфликтов интересов при различных  $\alpha$

После итогового агрегирования территориям сопоставлялись ранги от 0 до 5. Красным выделены зоны с наибольшей интенсивностью конфликта интересов (интенсивность 5), затем, по убыванию, оранжевый (интенсивность 4), желтый (интенсивность 3), светло-зеленый (интенсивность 2) и зеленый (интенсивность 1). Зоны с еще меньшей «конфликтностью» не выделялись. К каждому рангу относилось одинаковое количество территорий. Если на изображении присутствуют не все цвета, это означает, что несколько категорий при применении правила оказались эквивалентными, то есть при присвоении, например, одному классу территорий интенсивность равна 1, а другой – 2, оказывалось, что конфликты в них имеют одинаковую интенсивность.

При увеличении параметра  $\alpha$  мы видим, как интересы смещаются из международной зоны в ИЭЗ арктических стран.

Теперь рассмотрим, как при значении параметра  $\alpha = 0.4$  будут изменяться зоны конфликтов интересов при увеличении веса того или иного ресурса в итоговом агрегировании. На рисунке 1.9 веса W одновременно изменяются для нефти и газа. Отметим, что при увеличении W с  $1/4$  до  $1/2$  количество «красных» зон немного уменьшается. Изменение веса с  $1/2$  до 1 не привносит изменений с точки зрения зон с интенсивностями конфликта интересов от 3 и выше. Однако «зеленые» зоны меняются сильно. При дальнейшем увеличении W слегка изменяются только «зеленые» зоны.

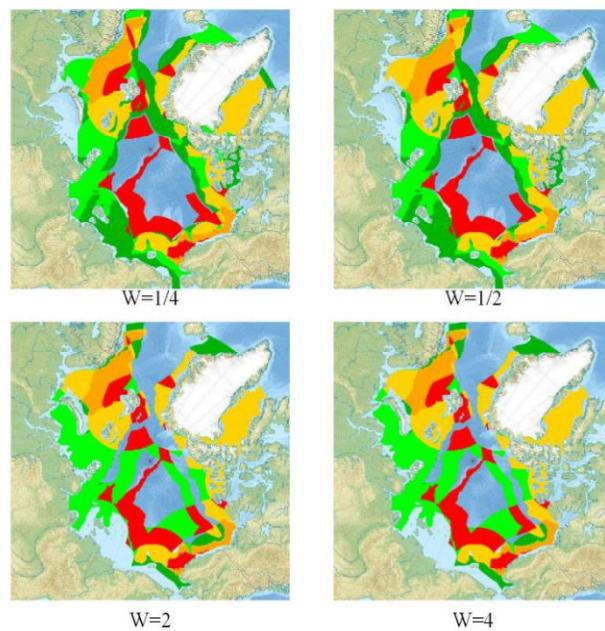


Рисунок 1.9 – Интенсивность конфликтов интересов при различных  $W$

Обозначим вес рыбных ресурсов через  $F(\text{fish})$ , вес мореходных ресурсов через  $R(\text{routes})$ . На рисунке 1.10 показано, какими будут интенсивности потенциальных конфликтов интересов при изменении весов  $F$  и  $R$ . При увеличении  $F$  с 1 до 2 некоторые зоны у берегов России становятся менее напряженными, в то же время обостряется ситуация у границ Аляски и Норвегии. При увеличении  $R$  с 1 до 2 напряженность перемещается из Карского моря в Восточно-Сибирское и море Лаптевых, от берегов Исландии на север, ближе к Гренландии, из международной зоны к России и Гренландии.

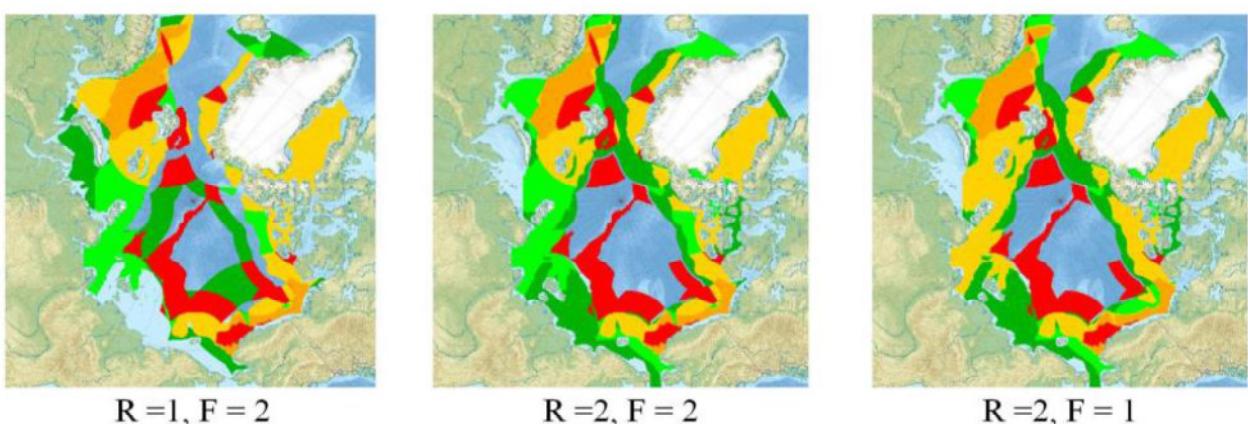


Рисунок 1.10 – Интенсивность конфликтов интересов при различных  $R, F$

### 1.2.7 Сценарий сильной напряженности

Рассмотрим сценарий, все страны очень сильно заинтересованы во всех ресурсах.  $\alpha = 0.8$ , поскольку страна не может вмешиваться в дела иностранного государства с тем же весом, что и оно само, но, тем не менее, оказывает сильное влияние. Результат моделирования такого сценария показан на рисунке 1.11.

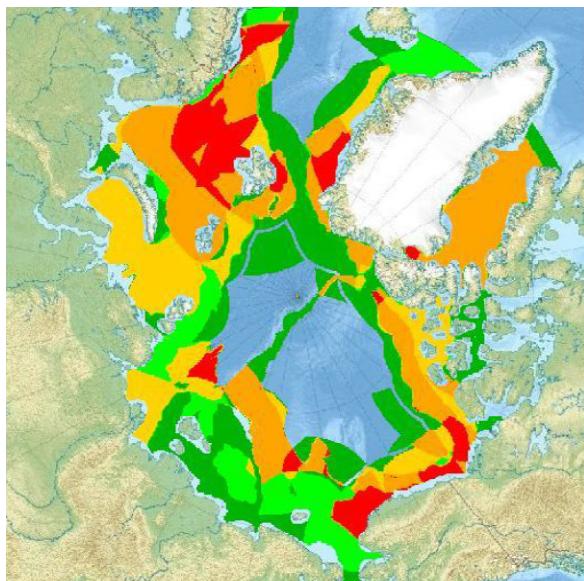


Рисунок 1.11 – Зоны конфликтов в случае сильной напряженности

Зонами с наибольшей интенсивностью конфликтов интересов стали Арктическая Аляска, участок над хребтом Ломоносова, часть Евразийского бассейна и восток Баренцева моря. «Оранжевые» зоны сконцентрированы в четырех провинциях: Восточно-Баренцевоморский бассейн, запад Гренландии – восток Канады, хребет Ломоносова и Свердупа. Наибольшая часть «желтой» области относится к Восточно-Сибирскому морю, остальные территории напряжены меньше.

Таким образом, в качестве наиболее напряженных зон можно выделить Баренцево море и северное побережье Аляски, богатые всеми четырьмя видами ресурсов.

### 1.2.8 Сценарий слабой напряженности

Теперь заинтересованность стран в ресурсах не высока. Поэтому страны согласуют действия по охране окружающей среды. Интерес неарктических государств к нефтяным и рыбным ресурсам также снизился (увеличилось значение константы C). Результат показан на рисунке 1.12.

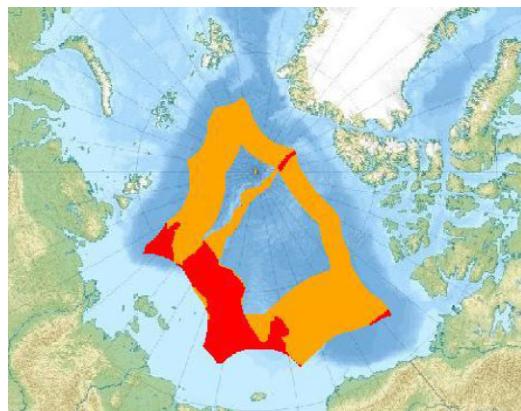


Рисунок 1.12 Зоны конфликтов в случае слабой напряженности

Мы видим, что конфликт интересов переместился из ИЭЗ арктических стран в международную зону. Самая высокая интенсивность приходится на провинции Ломоносов-Макаров и Евразийский бассейн.

### 1.2.9 Некоторые выводы

В рамках исследования пространственных моделей взаимодействия были проанализированы хозяйственно-экономические и политические интересы, а также возможны конфликты стран арктического региона. В результате проведенного исследования была получена оценка потенциальных зон конфликтов интересов стран и исследованы различные сценарии развития событий в зависимости от некоторых параметров. Однако оценка этих параметров является экспертной, что приводит к необходимости решения дополнительного ряда задач. Рассмотренная модель является прототипом. Ее усовершенствование предполагается проводить в двух направлениях. Во-первых, предполагается использование не только геологических данных (нефть газ), биологических данных (рыбные ресурсы) и экономических данных (мореходные

ресурсы), находящихся в открытых источниках, но и тех данных, которые находятся в специализированных (комерческих) базах. Во-вторых, в дальнейшем предполагается разработка методики уточнения параметров модели.

### 1.3 Исследование сетевых моделей взаимодействия

В этом подразделе отчета приведены результаты исследования сетевых моделей взаимодействия. Особое внимание в исследовании было уделено выявлению ключевых (наиболее важных и значимых) участников взаимодействия. Определение важности узла зависит от контекста и природы сети. В настоящее время существует большое число различных способов определения центральности узлов. Тем не менее, в связи с экспоненциальным ростом информации, а также высокой вычислительной сложностью большинства методов определения центральности, большинство существующих методов не может быть использовано при работе с сетями, состоящими из нескольких сотен тысяч и тем более миллионов вершин. В связи с этим возникает потребность в поиске более простых и в то же время достаточно точных методов определения центральности.

В исследовании рассматривались различные методы, которые обычно используются в задачах общественного выбора, а также адаптация этих методов к ориентированным сетям определенных типов. В частности, особое внимание было уделено правилам Копланда [39], интерес к которым может быть объяснен низкой вычислительной сложностью данных процедур. Кроме того, в исследовании также рассматривались различные версии непокрытых множеств (UC) [40, 41], а также другие правила, использующие мажоритарное отношение между элементами. На основе сетей различных размеров исследовалась взаимосвязь между существующими способами определения центральности и адаптируемыми методами.

Разработанный подход был применен к эмпирическим данным о денежных переводах среди 214 стран мира в 2014 году. Основной целью данного исследования было определение стран из верхнего и нижнего срезов полу-

ченного порядка. Помимо прочего, в рассматриваемой сети денежных переводов были найдены непокрытые множества, на основе которых можно произвести ранжирование существующих стран и определить наиболее устойчивые узлы в сети.

### 1.3.1 Исследование взаимосвязи между правилами Копланда и различными индексами центральности

#### 1.3.1.1 Описание данных

Для расчета индексов центральности в сетях были взяты две существующие сети размером около 10 тысяч вершин, а также сгенерированы две сети размером 10 и 15 тысяч вершин. Приведем краткое описание выбранных данных для тестирования.

##### 1. Сеть голосований в интернет-энциклопедии Wikipedia

Данная сеть содержит около 7 тысяч вершин и 104 тысяч ребер ([Wikipedia vote network, источник: http://snap.stanford.edu/data/wiki-Vote.html](http://snap.stanford.edu/data/wiki-Vote.html)).

Википедия (Wikipedia) представляет собой свободную общедоступную мультиязычную универсальную интернет-энциклопедию, статьи в которой создавать и редактировать может любой пользователь сети интернет, соблюдающий правила Википедии. Одним из важнейших элементов Википедии составляют ее администраторы, которые имеют дополнительный доступ к интернет-энциклопедии и отвечают за ее техническую поддержку, разрешение конфликтов и подведение итогов. Для того чтобы стать администратором Википедии, любой пользователь может отправить специальный запрос, который будет рассмотрен сообществом Википедии.

В рамках проекта SNAP (Стэнфордский проект по анализу сетей, SNAP – Stanford Network Analysis Project) была собрана последняя наиболее полная база о результатах голосования по заявкам на получение статуса администратора. Данные содержат результаты 2794 голосований, которые проводились с января 2001 года по январь 2008 года и в которых участвовало 7115 избира-

телей, половина из которых является администраторами Википедии, а остальные – обычными пользователями. Данные представляют собой ориентированный граф, вершинами которого являются участники голосования, а ребрами – факт голосования одного пользователя за другого. В Таблице 1.5 приведена краткая статистика по рассматриваемой сети.

Таблица 1.5 – Краткая статистика по сети голосований в Wikipedia

Характеристика	Значение
Общее число вершин (узлов)	7115
Общее число ребер (связей)	103689
Среднее число связей	14.57
Плотность графа	0.002
Диаметр сети	10
Средняя длина пути	3.34
Средний коэффициент кластеризации	0.081
Число слабо-связных компонент	25
Число сильно-связных компонент	5818

Распределение по количеству ребер для данной сети приведено на рисунке 1.13.

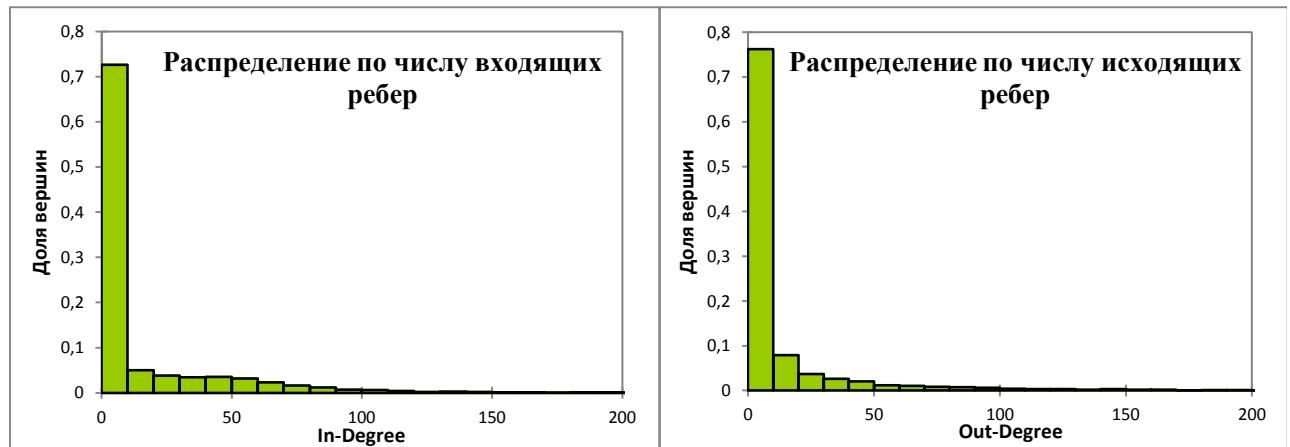


Рисунок 1.13 – Распределение по числу входящих и исходящих ребер в сети Wikipedia

## 2. Одноранговая компьютерная сеть Gnutella

Данная сеть содержит около 11 тысяч вершин и 40 тысяч ребер (Gnutella peer-to-peer network, источник: <https://snap.stanford.edu/data/p2p-Gnutella04.html>).

Рассматриваемая сеть представляет собой систему связи компьютеров и вычислительного оборудования между собой, основанная на равноправии участников. В рамках проекта SNAP была собрана информация об одноран-

говой сети “Gnutella” за август 2002 года. Данные представляют собой ориентированный граф, вершинами которого являются машины (компьютеры и вычислительное оборудование), а ребрами – наличие прямой связи между машинами.

В Таблице 1.6 приведена краткая статистика по рассматриваемой сети.

Таблица 1.6 Краткая статистика по сети

Характеристика	Значение
Общее число вершин (узлов)	10876
Общее число ребер (связей)	39994
Среднее число связей	3,68
Плотность графа	0
Диаметр сети	26
Средняя длина пути	6,77
Средний коэффициент кластеризации	0,0062
Число слабо-связных компонент	1
Число сильно-связных компонент	6560

Распределение по количеству ребер приведено на рисунке 1.14.

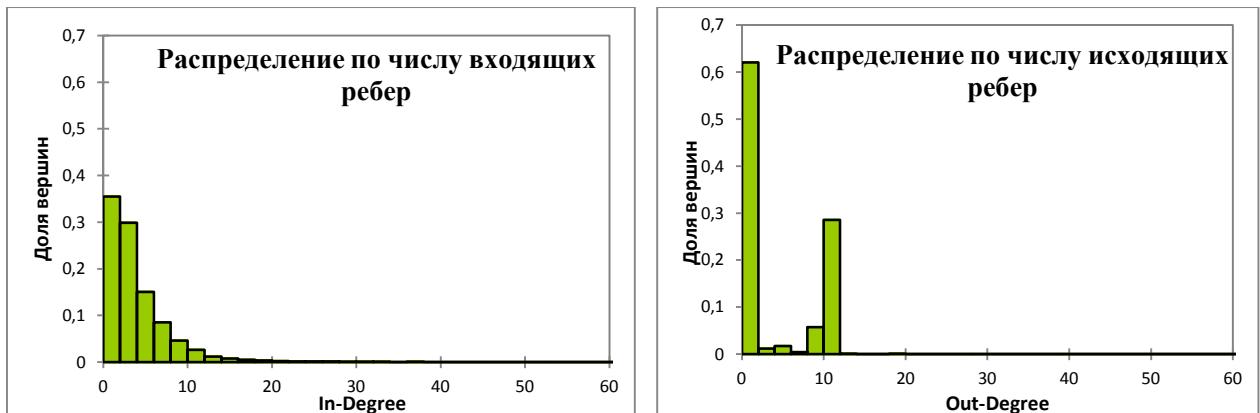


Рисунок 1.14 – Распределение по числу входящих и исходящих ребер в сети Gnutella

Для расчета индексов центральности помимо существующих сетей были также сгенерированы 2 сети размером 10 и 15 тысяч вершин. Данные были сгенерированы в пакете прикладных программ MATLAB так, чтобы число входящих вершин имело экспоненциальное распределение. Приведем ниже краткую характеристику сгенерированных сетей.

### 3. Сгенерированная случайная сеть размером 10 тыс. вершин

В Таблице 1.7 приведена краткая статистика по рассматриваемой сети.

Таблица 1.7 – Краткая статистика по данным

Характеристика	Значение
Общее число вершин (узлов)	10000
Общее число ребер (связей)	304541
Среднее число связей	30.454
Плотность графа	0.003
Диаметр сети	7
Средняя длина пути	3.18
Средний коэффициент кластеризации	0.005
Число слабо-связных компонент	1
Число сильно-связных компонент	1

Распределение по количеству ребер приведено на рисунке 1.15.

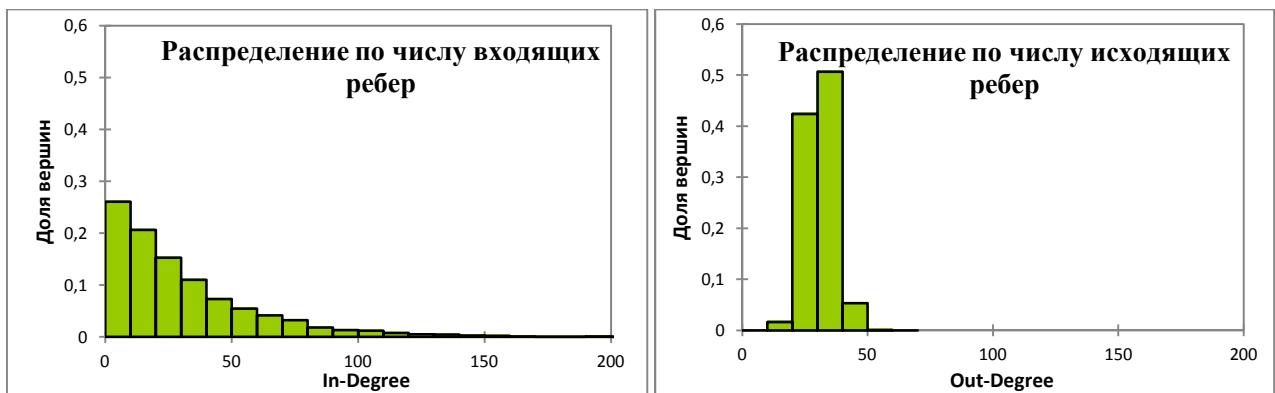


Рисунок 1.15 – Распределение по числу входящих и исходящих ребер в сети 10 тыс. вершин

#### 4. Сгенерированная случайная сеть размером 15 тыс. вершин

В Таблице 1.8 приведена краткая статистика по рассматриваемой сети.

Таблица 1.8 Краткая статистика по данным

Характеристика	Значение
Общее число вершин (узлов)	15000
Общее число ребер (связей)	384406
Среднее число связей	25.627
Плотность графа	0.002
Диаметр сети	7
Средняя длина пути	3.458
Средний коэффициент кластеризации	0.003
Число слабо-связных компонент	1
Число сильно-связных компонент	1

Распределение по количеству ребер приведено на рисунке 1.16.

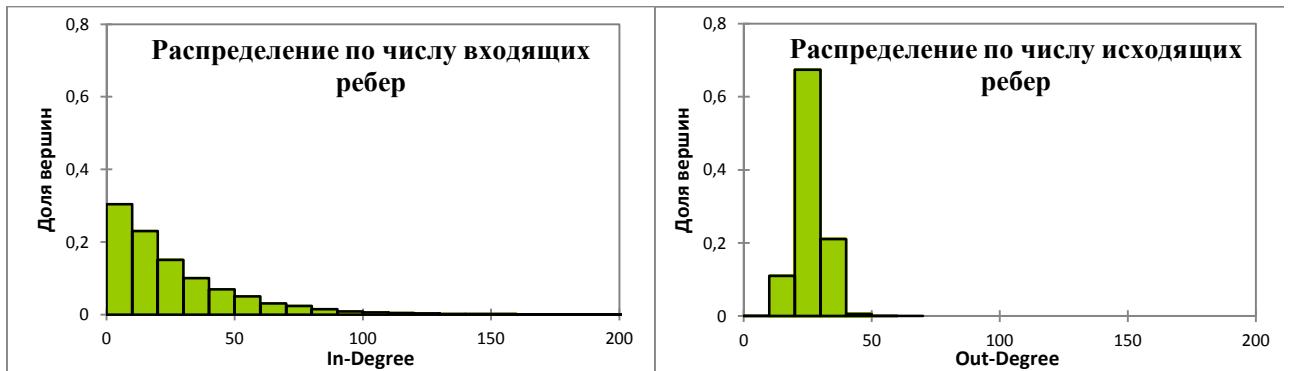


Рисунок 1.16 – Распределение по числу входящих и исходящих ребер в сети 15 тыс. вершин

### 1.3.1.2 Индексы центральности

В работе рассматриваются следующие индексы центральности:

- In-Degree – число связей, направленных к конкретному узлу;
- Out-Degree – число связей, направленных от конкретного узла;
- Degree – общее число связей (входящих и исходящих) конкретного узла;
- Eccentricity – максимальная дистанция от выбранного начального узла до наиболее удаленного узла;
- Closeness Centrality – средняя дистанция от выбранного начального узла до всех остальных узлов графа;
- Betweenness Centrality – рассчитывает, как часто узел лежит на кратчайшем пути между некоторыми двумя узлами графа;
- Closeness Centrality (normalized) – нормализованное от 0 до 1 значение closeness centrality (относительно максимального значения closeness centrality для конкретного графа);
- Betweenness Centrality (normalized) – нормализованное от 0 до 1 значение betweenness centrality (относительно максимального значения betweenness centrality для конкретного графа);
- PageRank – ранжирует узлы (как страницы), в соответствии с тем, как часто “пользователь”, переходя по ссылкам (ребрам), будет попадать в данный узел. Параметры “вероятность (р)” (вероятность того, что пользователь слу-

чайно обход графа с начала) и “epsilon” (параметр условия остановки) взяты по умолчанию и равны 0.85 и 0.001 соответственно;

- Eigenvector Centrality – рассчитывает вес узла в сети на основе связей данного узла.
- Clustering Coefficient – коэффициент кластеризации, который показывает, насколько плотно узлы упакованы в своем ближайшем окружении.

Расчет индексов центральности для рассматриваемых сетей был осуществлен с помощью инструмента визуализации и анализа социальных сетей Gephi (версия 0.8.2-beta). Помимо существующих индексов центральности в работе также рассматриваются 3 индекса центральности, основанные на правилах Копланда [42]:

- первое правило Копланда – узлы с максимальной разницей между числом исходящих (Out-degree) и входящих (In-degree) связей считаются наиболее центральными;
- второе правило Копланда – узлы с максимальным числом исходящих связей считаются наиболее центральными; данный индекс эквивалентен существующему показателю Out-degree;
- третье правило Копланда – узлы с минимальным числом входящих связей считаются наиболее центральными; данный индекс рассчитывается также, как и показатель In-degree, однако по правилу Копланда его значения нужно минимизировать.

### 1.3.1.3 Сравнение существующих индексов центральности с индексами, основанными на правилах Копланда

Проанализируем связь между существующими индексами центральности и индексами, основанными на правилах Копланда на рассматриваемых сетях. Одним из предположений является гипотеза о том, что вершины, имеющие высокие значения индексов центральности должны также иметь высокие значения центральности по правилам Копланда. Для проверки дан-

ной гипотезы было решено выделить по каждому индексу центральности некоторое множество лучших вершин (в зависимости от значения индексов центральности) и посмотреть, какие из данных вершин попадают в интервалы лучших значений для правил Копланда.

Замечание 1.4. В сети голосований Wikipedia ребра между вершинами имеют смысл “цитирования”, а не “доминирования”. Другими словами, чем выше значение показателя in-degree для конкретной вершины, тем лучше для нее. В таком случае значения по первому и второму правилу Копланда нужно минимизировать, по третьему – максимизировать.

В Таблицах А.1 – А.4 Приложения А приведены полученные результаты, которые показывают, какое число вершин попало в заданные интервалы лучших значений по правилам Копланда, а также, сколько вершин, входящих в список лучших по различным индексам центральности, попадает в заданные интервалы по правилам Копланда.

### 1.3.2 Расчет индексов на сети денежных переводов

В рамках данного направления было предложено скорректированное значение Копланда для взвешенных ориентированных сетей, которое выглядит следующим образом

$$CS_{adj}(i) = \sum_{j=1}^n w_{ij} - \sum_{j=1}^n w_{ji},$$

где  $w_{ij}$  – вес ребра, исходящего из узла  $i$  в узел  $j$ .

Расчет данного показателя был осуществлен для сети денежных переводов мигрантов между 214 странами мира в 2014 году (база данных ”The World Bank Prospects. Migration & Remittances Data”, источник – <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTDECPROSPECTS/0,,contentMDK:22759429~pagePK:64165401~piPK:64165026~theSitePK:476833,00.html>). Матрица двусторонних денежных переводов мигрантов была рассчитана Всемирным Банком. Методика проведения анализа двухстороннего коридора денежных переводов позволила учесть потоки через неофици-

альные каналы [43]. В оценках двусторонних переводов за 2014 год были использованы доходы принимающей страны мигрантов и их родной страны.

В сети денежных переводов мигрантов узлы представляют страны, а ребра – абсолютный поток денежных переводов между двумя странами (равен разности большего и меньшего денежных потоков между парой стран). Данная сеть также была приведена к виду невзвешенной направленной сети (путем бинаризации весов на ребрах сети). В Таблице 1.9 приведена краткая статистика по рассматриваемой сети.

Таблица 1.9 – Краткая статистика по сети денежных переводов

Характеристика	Значение
Общее число вершин (узлов)	214
Общее число ребер (связей)	3861
Среднее число связей	18.042
Плотность графа	0.085
Диаметр сети	7
Средняя длина пути	2.608
Средний коэффициент кластеризации	0.354

Распределение по количеству ребер приведено на рисунке 1.17.

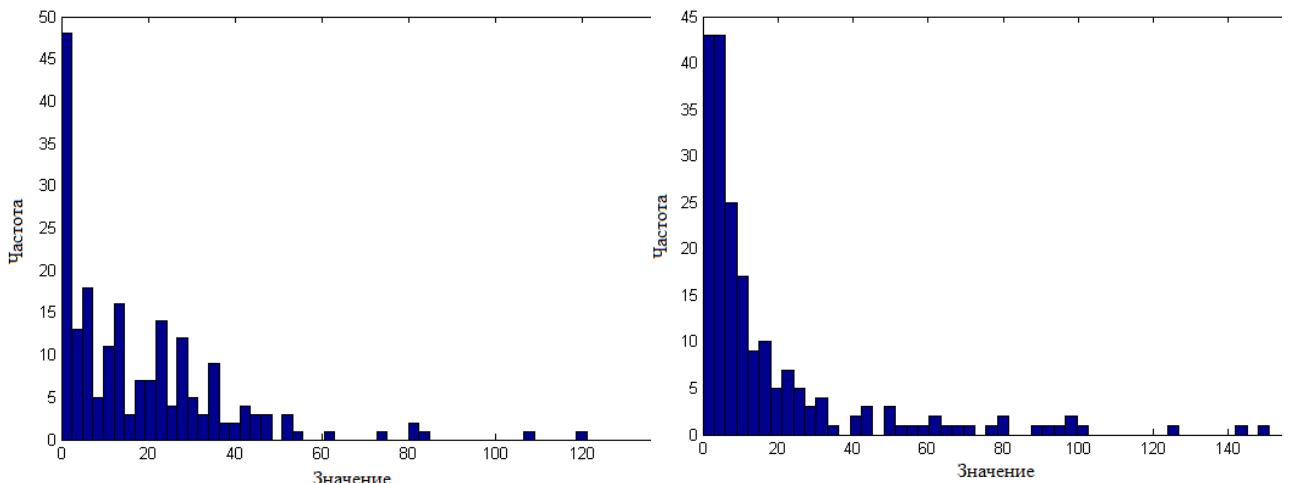


Рисунок 1.17 – Распределение входящих (слева) и исходящих (справа) ребер в сети денежных переводов

Была рассчитана степень престижа (мера центральности по собственному вектору) каждого узла рассматриваемой сети. На рисунке 1.18 показаны полученные результаты.

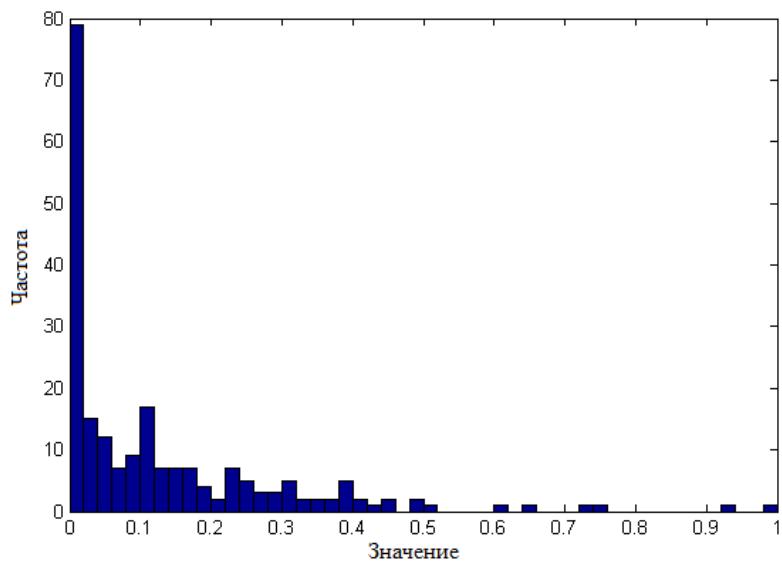


Рисунок 1.18 – Распределение узлов по мере престижа

Таким образом, для каждого узла сети денежных переводов было рассчитано скорректированное значение Копланда. Полученные результаты позволили выстроить на 214 странах мира слабый порядок. Были определены страны из верхнего и нижнего срезов полученного порядка. В Таблице 1.10 приведены по 10 стран из верхнего и нижнего срезов полученного слабого порядка.

Таблица 1.10 – Страны из верхнего и нижнего срезов слабого порядка, построенного по скорректированному значению Копланда

№	Страны в верхнем срезе порядка	Страны в нижнем срезе порядка
1	США	Ливан
2	Саудовская Аравия	Индонезия
3	Объединенные Арабские Эмираты	Вьетнам
4	Великобритания	Пакистан
5	Канада	Бангладеш
6	Гонконг	Нигерия
7	Россия	Мексика
8	Австралия	Филиппины
9	Кувейт	Китай
10	Катар	Индия

На рисунке 1.19 представлена сеть денежных переводов, размер надписей узлов соответствует их скорректированному значению Копланда.

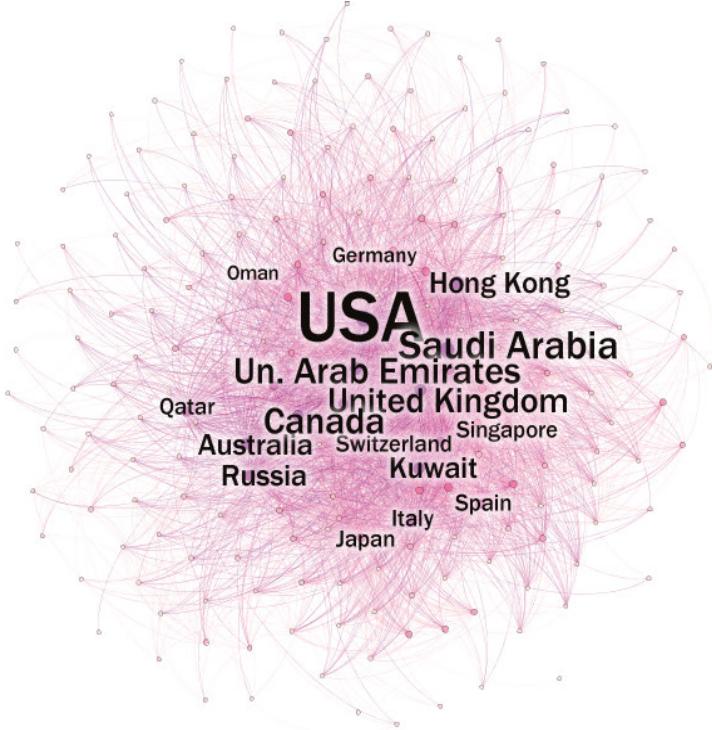


Рисунок 1.19 – Сеть денежных переводов мигрантов. Размер подпись узлов соответствует их скорректированному значению Копланда

Была дана интерпретация полученных результатов. Страны в верхнем срезе полученного слабого порядка – наиболее привлекательные направления для мигрантов. Это не только страны с высоким уровнем ВВП, доходов населения и традиционно высоким процентом мигрантов среди работающего населения (США, Канада, Великобритания, Австралия), но и богатые страны Персидского залива, привлекающие мигрантов близких стран, не имеющих нефтяных запасов (Саудовская Аравия, ОАЭ, Кувейт, Катар). Также в верхний срез входит Гонконг, привлекательный для мигрантов-выходцев из Юго-Восточной Азии, и Россия, в которой трудятся иммигранты из Средней Азии, Украины, Молдовы.

Страны в нижнем срезе полученного порядка – в основном, развивающиеся страны Юго-Восточной Азии. Также в этот срез входят Ливан, представляющий Ближний Восток, Нигерия, представляющая Африку, и Мексику, представляющая Латинскую Америку. Все страны их нижнего среза порядка характеризуются низким значением ВВП, доходов на душу населения

и большим количеством трудовых мигрантов, уезжающих на заработки (как правило) в соседние, более развитые и богатые, страны.

Определим теперь непокрытые множества [39, 40], одну из версий устойчивых множеств, для данной сети денежных переводов. Произведем ранжирование стран согласно отношению  $\delta$ , построенному на основании включения по их нижним контурам (множествам узлов, в которые входят исходящие ребра рассматриваемых узлов). Для любых двух узлов  $x, y$  верно следующее:

$$x\delta y \Leftrightarrow L(x) \supset L(y),$$

где  $L(x)$  и  $L(y)$  – нижние контуры узлов  $x$  и  $y$  соответственно.

В Таблице 1.11 приведены некоторые страны из верхнего и нижнего срезов полученного порядка.

Таблица 1.11 – Некоторые страны из верхнего и нижнего срезов порядка, заданного отношением  $\delta$  на множестве стран

№	Страны в верхнем срезе порядка	Страны в нижнем срезе порядка
1	Андорра	Непал
2	Каймановы острова	Бутан
3	ЦАР	Филиппины
4	Багамские острова	Вьетнам
5	Ангола	Китай
6	Сингапур	Шри-Ланка
7	Канада	Таджикистан
8	Катар	Индия
9	США	Пакистан
10	Великобритания	Афганистан

Страны из верхнего среза полученного порядка условно можно разделить на три группы. К первой группе можно отнести африканские страны с небольшим количеством исходящих ребер (ЦАР, Ангола). Ко второй группе можно отнести относительно изолированные страны, которые являются привлекательными направлениями не столько для трудовой миграции, сколько для перевода денег с целью вывода капитала, оффшоры (Андорра, Багамские острова, Каймановы острова, Сингапур). Третья группа – страны с большим количеством исходящих ребер, привлекательные для трудовых мигрантов (Канада, Катар, США, Великобритания).

В нижний срез данного порядка входят страны с большим количеством входящих потоков и очень маленьким количеством исходящих потоков. Все они – развивающиеся азиатские страны с низким уровнем дохода населения.

Итак, в данной работе инструменты, разработанные в теории коллективного выбора, были применены к эмпирическим данным о сети денежных переводов между 214 странами в 2014 году. Полученные результаты были проинтерпретированы и согласуются с основными экономическими и демографическими сведениями о данных странах. Результаты данной работы могут быть использованы для дальнейшего изучения экономического взаимодействия стран, популярных направлений миграции и денежных потоков между странами.

### 1.3.3 Некоторые выводы

В рамках исследования сетевых моделей был разработан новый подход к определению степени важности узлов в сложных сетях. В частности, были адаптированы правила Копланда к ориентированным сетям определенных типов и предложено скорректированное значение Копланда для взвешенных ориентированных сетей. Разработанный подход применен к эмпирическим данным о денежных переводах среди 214 стран мира в 2014 году. Результаты получены в виде слабого порядка. Определены страны из верхнего и нижнего срезов полученного порядка. Данна интерпретация полученных результатов. В рассматриваемой сети денежных переводов найдены непокрытые множества, на основе которых было осуществлено ранжирование стран. Данна интерпретация наиболее устойчивых узлов в сети.

Кроме того, была исследована взаимосвязь между значениями важности узлов по правилам Копланда и значениями классических мер центральности на примере четырех различных направленных сетей больших размеров (двух существующих и двух случайно сгенерированных размером 10-15 тысяч вершин). Полученные результаты показали, с какими из существующих индексов центральности лучше всего согласуются правила Копланда.

## 1.4 Прогнозирования структуры приема в вузы с учетом тенденций спроса на высшее образование

Тенденции развития мировой экономики демонстрируют, что спрос на специалистов высокой квалификации определяется поведением большого числа агентов, чьи интересы зачастую противоречат друг другу, а также множеством факторов, иногда слабо предсказуемых. В этих условиях традиционные методы долгосрочного прогнозирования не обеспечивают достаточно надежных результатов, позволяющих разрабатывать стратегии долгосрочного развития сферы образования, которые, с одной стороны, нацелены на развитие человеческого капитала для поддержания глобальной конкурентоспособности страны, а с другой стороны – воспринимаются и разделяются властью, бизнесом и обществом.

### 1.4.1 Анализ существующих подходов к прогнозированию контрольных цифр приема

Практика применения существующих подходов к прогнозированию контрольных цифр приема в вузы показывает, что проблема диспропорции между предложением и спросом на различные направления подготовки, выраженная в разрыве между количеством студентов, обучающихся за бюджетные и собственные средства, достигла существенных размеров. В СМИ обсуждается необходимость снижения КЦП по направлениям экономики, менеджмента и юриспруденции, как ответ на дефицит технических кадров в некоторых отраслях. Очевидно, что такие решения, которые чаще всего обосновывают прогнозами потребностей работодателей и отраслей, не устранит проблему.

Существует подход, основанный на отраслевом принципе в предположении, что структура человеческого капитала, необходимая для производства единицы продукции, не имеет значительных региональных различий, а

определяется уровнем технологий и организации труда для различных отраслей экономики (рисунок 1.20). При определенной корректировке, обеспечивающей учет региональной специфики, применимость этого подхода во многом определяется качеством параметрической настройки соответствующей модели прогноза. Как утверждают в [44], «эта методика расчета, называемая "технологической" методикой, обладает свойством универсальности для всех субъектов Федерации, в разрезе всех отраслей экономики и для всех уровней образования». Анализ показывает, что эта модель является вариантом нормативного подхода, применявшегося для решения аналогичной задачи Госпланином СССР.



Рисунок 1.20 – Существующая система прогнозирования контрольных цифр приема в вузы

#### 1.4.2 Моделирование поведения семей и абитуриентов, предъявляющих спрос на высшее образование

В этом подразделе отчета описан другой подход к решению задачи прогнозирования контрольных цифр приема, основанный на моделировании поведения семей и абитуриентов, непосредственно предъявляющих спрос на высшее образование, а не только на потребности работодателей и отраслей. В соответствии с глобальными процессами массификации высшего образования [45, 46] и все более активным использованием маркетинговых подхо-

дов вузами и системами высшего образования, представляется оправданным рассматривать спрос на высшее образование, как результат коммуникационного и маркетингового влияния на потребителей со стороны "продавцов" и собственных представлений потребителей о продукте, основанных на опыте и стереотипах, тиражируемых массовой культурой и СМИ (рисунок 1.21).

Спрос на бакалавриат, обладающий существенной инерцией, смещается вслед за мнением и поведением медийных персон (публичными представителями нового бизнеса, науки, государства и т.п.) и информационными возмущениями в поле масс медиа (рисунок 1.22).



Рисунок 1.21 – Предлагаемая система прогнозирования контрольных цифр приема в вузы

В не меньшей, а может быть и в большей степени, структура спроса на обучение в бакалавриате зависит от самооценки абитуриентами способностей и возможностей поступить в определенные вузы, на определенные направления подготовки в соответствие с полученными баллами ЕГЭ и конкурсом в вузах.



Рисунок 1.22 – Изменение спроса на высшее образование как результат взаимодействия экономической и социальной реальности с общественным сознанием

Предлагаемый прогноз опирается на модели многоступенчатого выбора. На первом этапе моделируется выбор абитуриентами направления подготовки с учетом распределения сданных баллов ЕГЭ по математике и русскому языку, средних и минимальных проходных баллов в вузы. Выбор осуществляется между естественнонаучными/техническими и гуманитарными направлениями подготовки. На втором этапе моделируется выбор предметных областей в рамках уже сделанного на первом этапе выбора. Он делается на основе индексов, характеризующих структуру информационного поля, влияющего на и отражающего общественное мнение.

#### 1.4.2.1 Моделирование выбора абитуриентом направления подготовки с учетом распределения баллов ЕГЭ

Предполагается, что у абитуриента есть следующие возможности – подать документы на специальности естественнонаучного или технического направления, которые будут обозначены одной буквой "Ф", или на специаль-

ности гуманитарного направления, которые будут обозначены одной буквой "Л", или вообще не подавать документы в высшее учебное заведение. Деление предметных областей условное и представлено в Таблице 1.12.

Таблица 1.12 – Разделение предметных областей по направлениям "Ф" и "Л"

	Укрупненные группы специальностей	Направление
1	Математические и естественные науки	Ф
2	Инженерное дело, технологии и технические науки	Ф
3	Здравоохранение и медицинские науки	Ф
4	Сельское хозяйство и сельскохозяйственные науки	Ф
5	Науки об обществе	Л
6	Образование и педагогические науки	Л
7	Гуманитарные науки	Л
8	Искусство и культура	Л

Очевидно, что группы Ф и Л неоднородны. В группе Ф представлены специальности/группы специальностей, требующих серьезной математической подготовки, развитого абстрактно-логического мышления и/или связанных с моделированием, проектированием, производством сложной техники и вооружения и т.п. Для качественной подготовки таких специалистов требуется мотивированные, хорошо подготовленные абитуриенты с, как правило, высокими баллами ЕГЭ по профильным предметам – назовем таких абитуриентов "чистыми Ф" (или "Фч").

В то же время, часть вузов принимает абитуриентов с экстремально низкими для таких профессий проходными баллами. Маловероятно, что такие студенты, даже закончив высшее учебное заведение, смогут быть высококвалифицированными специалистами и работать по специальности. Логично считать, что такие абитуриенты, даже сумев закончить вуз и получить диплом, будут работать не по специальности, а в смежных областях, связанных не с созданием высокотехнологического продукта, а, например, с его обслуживанием. Таких абитуриентов мы будем называть "сервисными Ф" (или "Фс"). Аналогичное деление введем и для "Л".

Обозначим за  $a_{\text{Ф},\text{ч}}$  и  $a_{\text{Л},\text{ч}}$  проходные баллы по обоим направлениям для групп чистых физиков и лириков,  $a_{\text{Ф},\text{min}}$  и  $a_{\text{Л},\text{min}}$  – минимальные проходные баллы для соответствующих направлений,  $a_{i,j}$  – способность абитуриента i

учиться по направлению  $j$ . Предположим, что способности оцениваются баллами ЕГЭ и абитуриенты знают свои способности. Для группы физиков основным будем считать ЕГЭ по математике, а для лириков – ЕГЭ по русскому.

Если абитуриенту известны все параметры  $a_{i,j}$ ,  $a_{\Phi,\chi}$ ,  $a_{L,\chi}$ ,  $a_{\Phi,min}$  и  $a_{L,min}$ , то схема выбора абитуриентом вуза выглядит следующим образом (рисунок 1.23):

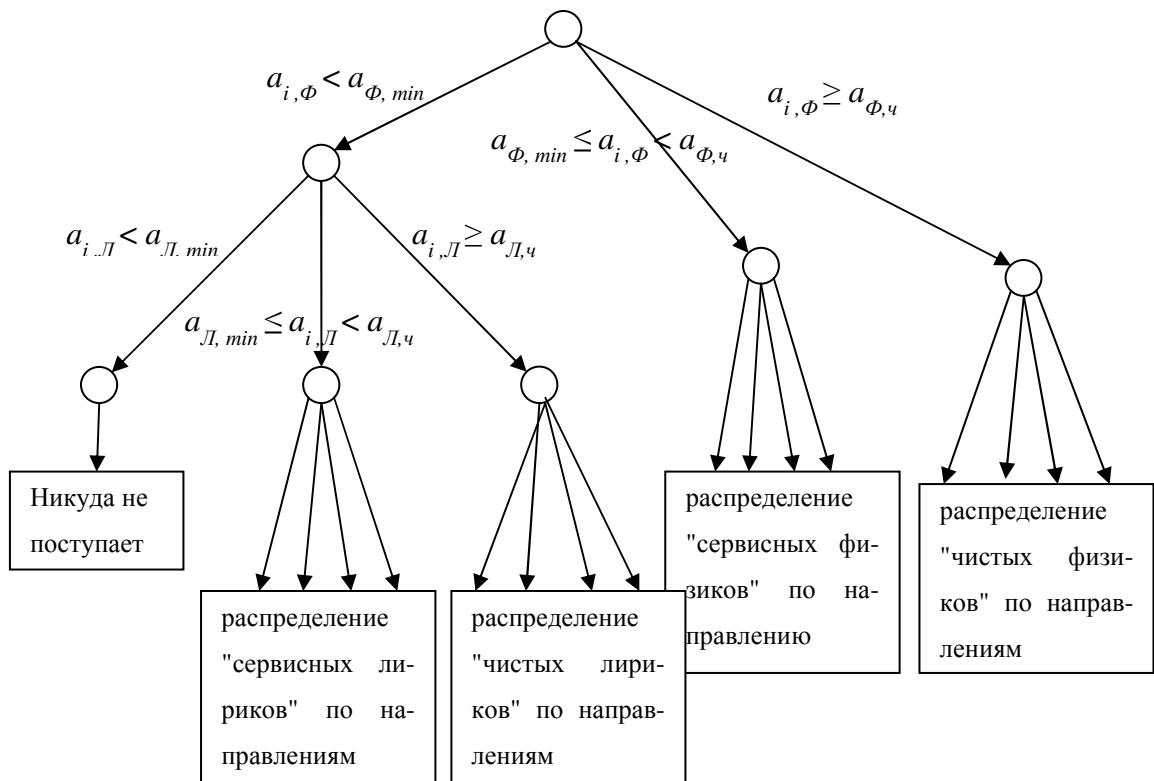


Рисунок 1.23 – Схема выбора абитуриента на основе самооценки способностей и возможности поступления в вуз по направлениям "Ф" и "Л"

Результаты моделирования выбора первого этапа представлены в Таблице 1.13. В первых столбцах таблицы даны реальные цифры КЦП 2013 г. предметных областей, входящих в направления Ф и Л, а в последних столбцах дана проекция выбора по модели на КЦП 2013 года (т.е. КЦП 2013 года были умножены на вероятность выбора направления).

По данным Таблицы 1.13 видно, что уже на первом этапе разница между реальными и моделируемыми КЦП почти 40 тыс. мест в пользу направлений

Л. Причем доля "чистых" Ф и Л, более чем в 2 раза меньше суммарных Ф и Л.

Таблица 1.13 – Результаты моделирования выбора абитуриентами направлений Ф и Л на КЦП 2013 года

Тип направления	КЦП 2013 г. по предметным областям этого направления	Доля типа направления в КЦП 2013 г.	Вероятность выбора абитуриентами направлений Ф и Л, рассчитанная по ЕГЭ на основе модели	Проекция выбора абитуриентами Ф и Л по ЕГЭ в соответствии с моделью на КЦП 2013 г.
Ф	221851	69,8%	57,70%	183343
в т.ч. "чистые Ф"			22,60%	71812
Л	95902	30,2%	42,30%	134410
в т.ч. "чистые Л"			12,30%	39084

#### 1.4.2.2 Моделирование выбора абитуриентом направления подготовки с учетом популярности предметных областей

На следующем этапе деление абитуриентов внутри каждой из первых четырех групп производится пропорционально индексам, характеризующим популярность каждой из предметных областей в информационном поле СМИ. Для нахождения индексов информационного поля были использованы инструменты семантического анализа. Сами индексы представляют собой частоты возникновения определенных тем в СМИ. Темы структурируются и задаются с помощью факторов, универсальных для всех предметных областей подготовки и представленных в Таблице 1.14.

Таблица 1.14 – Факторы, с помощью которых было структурировано информационное поле СМИ

1. Привлекательность образа (стиля) жизни и образа человека соответствующей профессии, транслируемый через СМИ
2. Перспективность данной профессии, транслируемая через СМИ
3. Качество обучения в вузах по направлению
4. Доступность высшего образования по направлению
5. Интерес общества и популярность профессии, отражаемая в СМИ
6. Внимание государства, корпораций, инвесторов к профессии, направлению подготовки
7. Научная активность в рассматриваемой области, внедрение инноваций, активность экспертов
8. Государственные стратегии и программы, инвестиции и оценки инвесторов, заключение контрактов, открытие объектов инновационной инфраструктуры

Результаты семантического анализа по одному из важнейших индексов популярности специальностей предметных областей показаны на графике (рисунок 1.24).

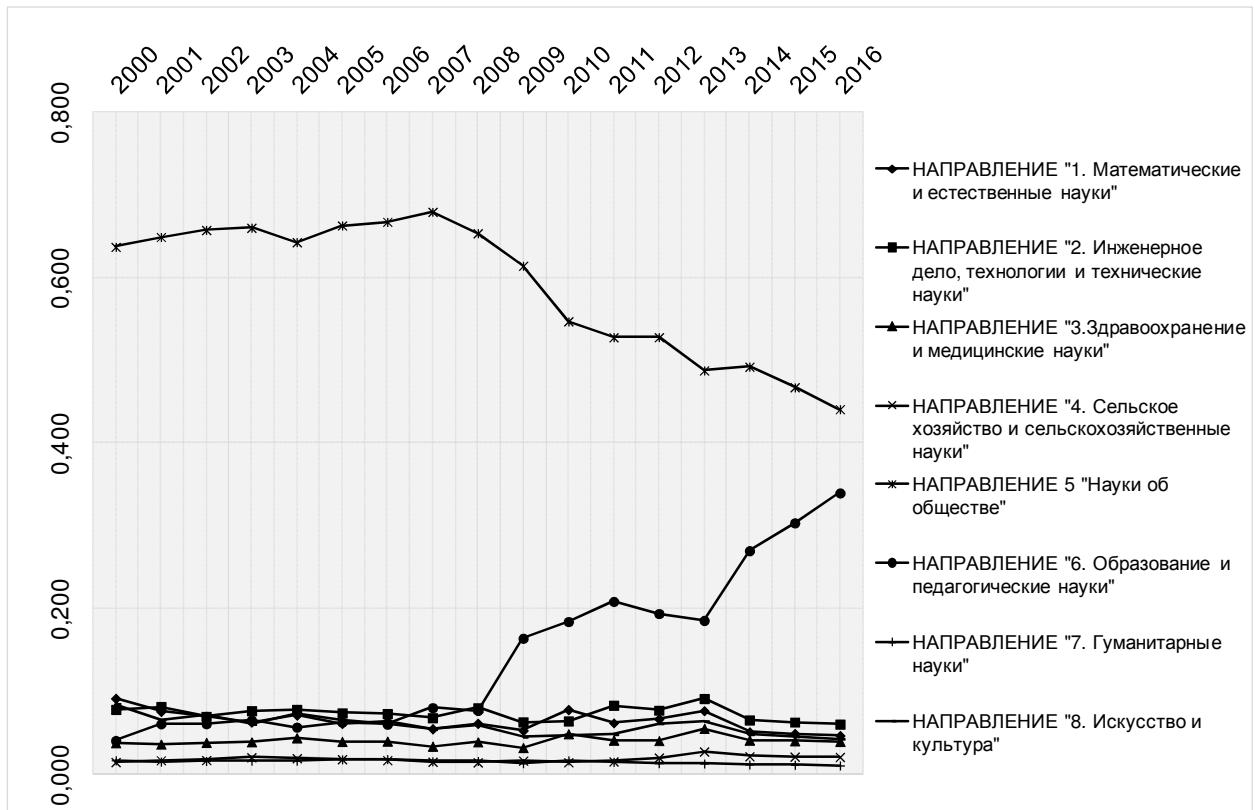


Рисунок 1.24 – Результаты семантического анализа: поведение индексов популярности специальностей предметных областей в относительных значениях (2000-2013 год – фактические данные, 2014-2016 - прогнозируемые)

На рисунке 1.24 видно поведение графиков индексов, отражающих популярность специальностей различных предметных областей. Прежде всего, видно заметное падение популярности предметной области «Науки об обществе». Оно начинается с 2008 года и продолжается до сегодняшнего момента, нет оснований считать, что падение популярности остановится в ближайшие 2-3 года. Однако даже в условиях падения эта предметная область по популярности опережает другие в разы. Второй момент, которые обращает на себя внимание, это поведение графика предметной области "Образование...". Он растет с того же момента и почти в той же степени, что падает график "Наук об обществе".

На следующем этапе работы стояла задача сравнения оценок популярности и престижности профессий, полученных методами семантического анализа, с результатами социологических опросов.

В обзорной социологической статье [47] делается экскурс в проблематику престижности профессий с точки зрения социологии. В ранних социологических работах прослеживается устойчивость представлений о престижности профессий во времени и в пространстве, а также для разных социальных групп и национальностей. В последнее время предприняты попытки объяснить устойчивость престижности профессий. Новые подходы критикуют традиционные методы измерения престижности профессий.

В другой статье [48] проводится анализ трех основных подходов к конструированию шкал, которые существуют в социологической литературе. Это шкалы профессионального престижа, социально-экономические индексы и реляционные (сетевые) шкалы.

Одним из наиболее представительных является исследование старшеклассников Новосибирской области, полученные коллективом авторов [49]. В нем зафиксировано резкое падение популярности профессий, требующих высшего образования и соответствующих «Наукам об обществе» (бизнесмен, банковский работник, юрист, журналист, бухгалтер, экономист) в промежутке с 2004 по 2013 год. Это связывается авторами со снижением популярности «перегретых» профессий после финансового кризиса 2008 года. По данным семантического анализа, это падение началось как раз кризисный 2008 год. Таким образом, можно зафиксировать соответствие популярности и престижности профессий, полученных разными методами, по этой группе специальностей в этот период.

Различия в оценке привлекательности «Наук об обществе», полученными разными методами, имеются в период 2000-2007 годы. Так как, согласно выводам коллектива авторов, пик популярности этих профессий приходится на 1994 год, в дальнейшем популярность незначительно, но ослабевает. Семантический анализ показывает увеличение популярности в этот период.

Вторая центральная тенденция, полученная методами семантического анализа – это увеличение популярности «Образования...». В списке профессий Константиновского к этой группе напрямую относится только профессия учителя. Увеличение популярности фиксируется с 2004 по 2013 год, но оно очень незначительное: 0,05 баллов по 10-балльной шкале. В то же время, методами семантического анализа зафиксировано гораздо более масштабное увеличение популярности: на 30 процентных пунктов.

Еще одна группа социологов [50] пришли к выводу, что «наиболее значимая тенденция в 2009 году по сравнению с 1997 годом – это значительное расширение спектра престижных профессий. Если в 1997 году 46% молодых среди наиболее престижных профессий называли в основном специальности юриста и финансиста (обе – более 40%), то через десять лет эти показатели уменьшились в среднем до 30%, а в 2009 году составили менее 20%. Да и в целом, по критерию престижности к 2009 году молодое поколение стало оценивать профессии с некоторым скепсисом. По-видимому, решающим при функциональной оценке молодежью тех или иных профессий является не этот критерий. Иная ситуация складывается, когда молодежь оценивает профессии по прибыльности. Наиболее прибыльными считаются профессии предпринимателя, финансиста, юриста. Однако в условиях нового кризиса прибыльность почти всех профессий оценивается низко». Таким образом, исследования Горшкова и Шереги подтверждают тенденцию снижения популярности профессий, которые могут быть соотнесены с «Науками об обществе». И так же, как и результаты Константиновского, отрезок резкого снижения популярности этой укрупненной группы специальностей с 2007 года по 2009 год согласуется с выводами, полученными методами семантического анализа. А отрезок медленного снижения популярности с 1997 года по 2007 – не согласуется.

Если в исследовании Горшкова и Шереги задавались два вопросы о престижности и прибыльности профессий и эти вопросы были адресованы молодежи, то Всероссийский центр изучения общественного мнения (ВЦИОМ)

[51, 52] выяснял отношения респондента уже по трем вопросам: о желательности профессии для своих детей или внуков, о престижности профессии и о доходности профессии. Кроме того, вопросы были адресованы индивидам всех возрастов. В 2013 году ВЦИОМ был проведен анализ еще одного дополнительного вопроса о доверии представителям разных профессий. «Учитель» оказался вторым в рейтинге после «ученого». Таким образом, мы видим, что по параметрам доходности и престижности «учитель» проигрывает остальным профессиям, то по параметру доверия – опережает их. Таки образом, социологические данные ВЦИОМ в некоторой степени согласуются с ростом престижности и перспективности направления «Образования...», полученных методами семантического анализа.

С помощью рассчитанных индексов информационного поля был смоделирован выбор специальностей предметных областей второго этапа и спрогнозированы соответствующие цифры КЦП (рисунок 1.25).

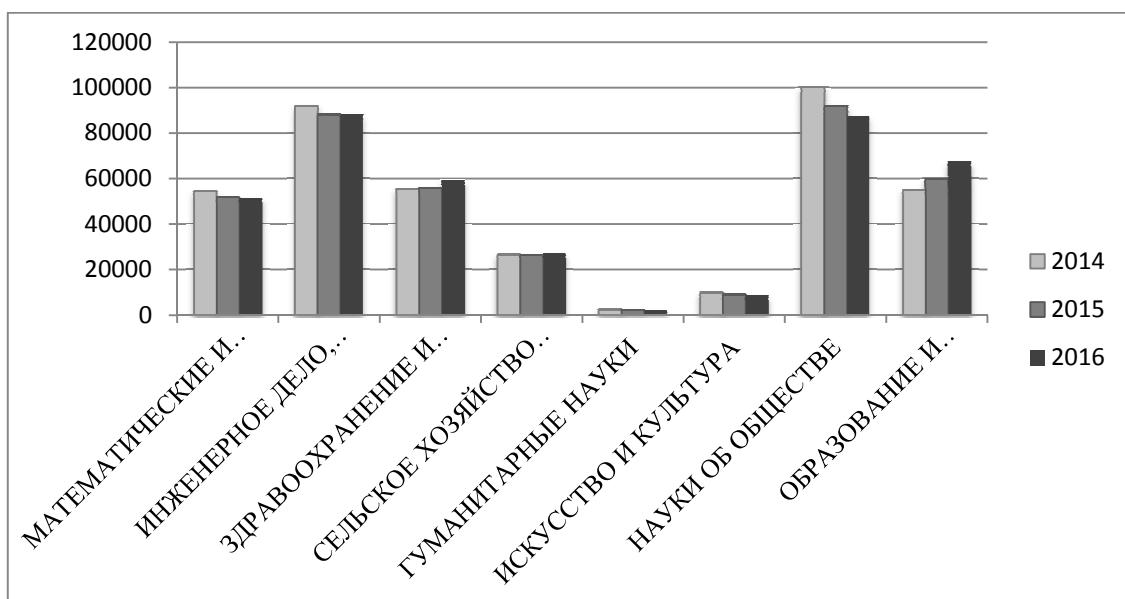


Рисунок 1.25 – Вариант расчета прогноза КЦП по предметным областям, выполненный с помощью ступенчатой модели выбора

Наибольший разрыв по сравнению с предшествующим этапом был спрогнозирован для направления «Инженерное дело...» – в 2 раза. Это связано с переоценкой доли абитуриентов, способных учиться по специальностям

направления «Ф» и недооценкой популярности специальностей других предметных областей этого направления. Следовательно, снижение выбора абитуриентами направления «Ф» по результатам сдачи ЕГЭ и снижение выбора данного УГС из-за сравнительно невысоких индексов популярности и перспективности обусловили снижение прогноза по этой предметной области.

Снижение КЦП по направлению «Инженерное дело...» в ходе расчетов было перераспределено в пользу других УГС направлений «Ф»: «Математики...» (105%) и «Здравоохранение...» (78%). Расчеты показывают, что направления "Математика..." и "Здравоохранение..." испытывают недостаток КЦП в пределах 80-100%. Высокая доля "коммерческих" студентов в "Здравоохранении..." и в отдельных сегментах "Математики..." позволяет судить о наличие потенциала для наращивания КЦП в этих предметных областях.

Несмотря на заметное падение популярности наук об обществе, реальные КЦП по этому направлению существенно ниже прогнозируемых, чтобы покрыть даже падающий спрос. Дефицит оценивается примерно в 60%. Этот недостаток компенсируется платными местами. В 2013 году доля бюджетных мест по направлению экономике и менеджменту составляла не более 12%, хотя в других социальных науках доля бюджетных мест была заметно выше, чем мест с полным возмещением затрат.

Одновременно со снижением индексов популярности и перспективности «Наук об обществе» наблюдается рост аналогичных индексов для «Образования...», прогнозные КЦП которого также получились выше ожидаемых цифр на 30%.

#### 1.4.3 Некоторые выводы

В этом подразделе отчета приведены результаты разработки модели многоступенчатого выбора вуза абитуриентом с учетом популярности и престижности профессий, самооценки абитуриентами своих способностей и возможностей поступить в определенные вузы в соответствии с полученными баллами ЕГЭ и конкурсом в вузах. В рамках этого исследования были

проведены серии вариативных расчетов на основе данных о баллах ЕГЭ выпускников 2013 года и индексов популярности профессий, рассчитанных на основе семантического анализа текстовых данных сети Интернет и СМИ. В частности, была спрогнозирована и проанализирована структура приема в вузы в 2014-2016гг. с учетом тенденций спроса на высшее образование на основе проведенных расчетов.

Разработанные модели и проведенные по ним серии вариативных расчетов показали устойчивость моделей, применимость их для прогнозирования КЦП по предметным областям с учетом трендов, определяющих спрос населения. Эти модели могут быть важной составной частью прогнозирования распределения КЦП принципиально дополняющей модели, ориентированные на оценку потребностей рынка труда.

Предварительные результаты этих исследований нашли отражение в работах [53].

## 2 Исследование методов многокритериального и коллективного принятия решений

В этом разделе отчета приведены результаты исследований методов многокритериального и коллективного принятия решений. Эти результаты получены в следующих теоретических и прикладных направлениях:

- исследование методов многокритериального принятия решений;
- сравнительное исследование и аксиоматическое описание концепций решений в задаче оптимального коллективного выбора;
- исследование нормативных свойств двухступенчатых процедур выбора;
- построение и применение теоретических моделей для расчета степени манипулируемости правил агрегирования предпочтений.

### 2.1 Исследование методов многокритериального принятия решений

Наиболее распространенными задачами принятия индивидуальных решений являются задачи выбора одной наилучшей альтернативы (варианта, плана, стратегии) из заданного (сформированного) множества вариантов. При анализе таких задач с использованием математических методов для моделирования предпочтений лица, принимающего решение (ЛПР), широко используются бинарные отношения предпочтения [54]. Если предпочтения описываются одним таким отношением и это отношение является частичным (т.е. не всякие два варианта сравнимы по этому отношению), то базовым является понятие *недоминируемости* [54].

Если предпочтения моделируются семейством бинарных отношений (т.е. заранее неизвестно, какое именно из входящих в него отношений описывает предпочтения ЛПР), и эти отношения являются полными (т.е. всякие два варианта сравнимы по этому отношению), то следует использовать понятие *потенциальной оптимальности* [55, 56]. Если же эти отношения являются лишь частичными, то базовым является понятие *потенциальной недоминируемости*, введенное в [57]. Свойства потенциально недоминируемых ва-

риантов подробно изучены в [58]. Далее в п.п. 2.1.1 – 2.1.2 изучается проблема проверки альтернативы на потенциальную недоминируемость в дискретных задачах выбора. Предварительно рассматриваются необходимые для ее решения свойства потенциально недоминируемых альтернатив.

В настоящее время меры риска, т.е. числовые характеристики, позволяющие судить о его величине и значимости («измерять риск»), широко используются в теории и практике принятия решений в условиях вероятностной неопределенности [59 – 63]. В п.п. 2.1.3 – 2.1.6 приводятся классификация и краткий обзор свойств основных известных мер риска как критериев выбора (оптимизации), причем подробно рассматриваются свойства относительно недавно введенной в практику меры риска – среднего полуотклонения.

### 2.1.1 Потенциально недоминируемые альтернативы и их свойства

Для удобства изложения вначале приведем необходимые сведения об отношениях предпочтения [54] и о потенциально недоминируемых альтернативах [57, 58]. Пусть  $X$  – множество альтернатив и  $R$  – заданное на нем отношение нестрогого предпочтения ЛПР:  $x'Rx''$  означает, что альтернатива  $x'$  не менее предпочтительна, чем  $x''$ . Предполагается, что множество  $X$  содержит не менее двух альтернатив, а отношение  $R$  является квазипорядком (оно рефлексивно и транзитивно). Если для любых альтернатив  $x'$  и  $x''$  верно  $x'Rx''$  или  $x''Rx'$ , то квазипорядок  $R$  называется *полным*, в противном случае – *частичным*. Отношение  $R$  порождает отношение безразличия  $I$  и отношение (строгого) предпочтения  $P$ : если верно  $x'Rx''$  и  $x''Rx'$ , то справедливо  $x'Ix''$  (альтернативы  $x'$  и  $x''$  безразличны, или одинаковы по предпочтительности); если же верно  $x'Rx''$ , но  $x''Rx'$  неверно, то справедливо  $x'Px''$  (альтернатива  $x'$  предпочтительнее альтернативы  $x''$ ). Для квазипорядка  $R$  отношение  $I$  есть эквивалентность (оно симметрично, рефлексивно и транзитивно), а отношение  $P$  – строгий частичный порядок (оно иррефлексивно и транзитивно). Альтерна-

тива  $x^*$  называется *недоминируемой* (*по*  $P$ ), если не существует альтернативы  $x$ , для которой верно  $xPx^*$ . В противном случае эта альтернатива называется *доминируемой* (*по*  $P$ ).

Пусть предпочтения ЛПР описываются семейством (непустым множеством)  $\mathbb{R}$  отношений нестрогого предпочтения (частичных квазипорядков) на  $X$ . Обозначим через  $\mathcal{P}$  семейство отношений предпочтения (строгих частичных порядков) на  $X$ , порождаемых квазипорядками  $R$  из  $\mathbb{R}$ . Альтернатива  $x^*$  называется *потенциально недоминируемой* (*для*  $\mathcal{P}$ ), если в  $\mathcal{P}$  найдется строгий частичный порядок  $P$  такой, что  $x^*$  является недоминируемой по  $P$ . В противном случае альтернатива  $x^*$  называется *заведомо доминируемой* (*для*  $\mathcal{P}$ ).

Важным для приложений является случай, когда семейство  $\mathbb{R}$  частичных квазипорядков представляет собой параметрическое семейство:  $\mathbb{R} = \mathbb{R}(A) = \{R(\alpha) \mid \alpha \in A\}$ , элементами которого являются частичные квазипорядки  $R(\alpha)$ , определенные на  $X$ . Множество  $A$  возможных значений параметра  $\alpha$  может состоять, например, из  $m$ -мерных векторов и определяться системой равенств и неравенств. Пусть  $\mathcal{P}(A)$  – множество отношений предпочтения (строгих частичных порядков)  $P(\alpha)$ , порождаемых квазипорядками  $R(\alpha)$  из  $\mathbb{R}(A)$ .

Будем полагать, что множество альтернатив конечно:  $X = \{x^1, \dots, x^n\}$ ,  $n \geq 2$ . Пусть  $x^t$  – альтернатива, которую мы хотим проверить на потенциальную недоминируемость. Решающее правило, задающее отношения  $P(\alpha)$ , для каждой альтернативы  $x^j$ ,  $t \neq j$ , в области  $A$  определяет подобласть  $A^{jt} \subseteq A$ , такую, что для каждого  $\alpha \in A^{jt}$  верно  $x^j P(\alpha) x^t$ . Справедлива

Теорема 2.1. *Альтернатива  $x^t$  заведомо доминуема тогда и только тогда, когда*

$$\bigcup_{j \neq t} A^{jt} = A. \quad (2.1)$$

Эта теорема указывает возможные пути построения вычислительных методов проверки фиксированной альтернативы  $x^t$  на потенциальную недоминируемость.

### 2.1.2 Проверка потенциальной недоминируемости альтернатив

Будем считать, что множество  $A$  возможных значений параметра  $\alpha$  является ограниченной областью в векторном числовом  $m$ -мерном пространстве  $\mathbb{R}^m$ . Пусть  $d(\alpha', \alpha'')$  – расстояние между точками  $\alpha'$  и  $\alpha''$  из  $A$ . Теорема 2.1 показывает, что если существует точка  $\alpha^0 \in A$ , для которой оказываются верными неравенства  $d(\alpha^0, \alpha^j) > 0, j \neq t$ , то она не входит во множество  $\bigcup_{j \neq t} A^{jt}$ , так что равенство в (2.1) не выполнено и альтернатива  $x^t$  потенциально недоминируема. Поэтому для проверки потенциальной недоминируемости можно пойти следующим путем.

Пусть  $J^t$  – множество номеров  $j \neq t$ , для которых множества  $A^{jt}$  не пусты, множества  $A$  и  $A^{jt}, j \in J^t$ , ограничены и определяются системами линейных равенств и неравенств. Обозначим через  $\bar{A}$  и  $\bar{A}^{jt}$  соответственно замыкания множеств  $A$  и  $A^{jt}$ , получаемые заменой всех строгих неравенств, участвующих в задании этих множеств, нестрогими. Пусть

$$d^* = \max_{\alpha \in \bar{A}} \min_{j \in J^t} d(\alpha, \bar{A}^{jt}), \quad (2.2)$$

где  $d(\alpha, \bar{A}^{jt})$  – расстояние от точки  $\alpha$  до замкнутого ограниченного множества  $\bar{A}^{jt}$ :

$$d(\alpha, \bar{A}^{jt}) = \min_{\alpha^{jt} \in \bar{A}^{jt}} d(\alpha, \alpha^{jt}). \quad (2.3)$$

Если оказывается, что  $d^* > 0$ , то существует точка  $\alpha^0 \notin \bigcup_{j \neq t} \bar{A}^{jt}$ . Следовательно, равенство в (2.2) не выполнено, и альтернатива  $x^t$  потенциально недоминируема. Если же оказывается, что  $d^* = 0$ , то требуется проведение дальнейших исследований.

Таким образом, для проверки потенциальной недоминируемости альтернативы  $x^t$  указанным путем нужно решить максиминную задачу в (2.2). Заметим, что если расстояние  $d(\alpha, \alpha^{jt})$  в (2.3) определяется формулой:

$$d(\alpha, \alpha^{jt}) = \sum_{i=1}^m |\alpha_i - \alpha_i^{jt}|, \quad (2.4)$$

то оптимизационная задача в (2.2) оказывается нелинейной и негладкой максиминной задачей.

Возможен, однако, подход, основанный на идее минимизации суммы расстояний до множеств  $\bar{A}^{jt}$ ,  $j \in J^t$ , исчисляемых по формуле (2.4). Он позволяет свести исходную максиминную задачу к билинейной задаче оптимизации, которую можно решить известными высокоеффективными методами квадратичного программирования, например, реализованными в программным пакете OOQP [64, 65]. А именно, пусть

$$\sigma^{jt*} = \max_{\alpha \in A} \sum_{j \in J^t} d(\alpha, \bar{A}^{jt}). \quad (2.5)$$

Пусть  $\alpha^{jt*}$  – точка минимума функции  $d(\alpha, \alpha^j)$  в (2.5). Если оказывается, что  $\sigma^{jt*} > 0$ , причем  $d(\alpha^*, \alpha^{jt*}) > 0$ ,  $j \in J^t$ , то точка  $\alpha^0 \notin \bigcup_{j \neq t} \bar{A}^{jt}$ . Следовательно, равенство в (2.2) не выполнено, и альтернатива  $x^t$  потенциально недоминируется. В противном случае требуется проведение дополнительного исследования. Если использовать формулу (2.4), то, учитывая свойства оператора  $\min$  [66] и применив известный прием [67], внутреннюю нелинейную задачу минимизации в (2.2) можно свести к линейной задаче минимизации. Действительно,

$$\sum_{j \in J^t} d(\alpha, \bar{A}^{jt}) = \sum_{j \in J^t} \min_{\alpha^j \in \bar{A}^{jt}} \sum_{i=1}^m |\alpha_i - \alpha_i^j| = \min_{\alpha^j \in \bar{A}^{jt}, j \in J^t} \sum_{j \in J^t} \sum_{i=1}^m |\alpha_i - \alpha_i^j|.$$

Поэтому, полагая

$$\alpha_i - \alpha_i^j = \alpha_i^{j+} - \alpha_i^{j-}, \quad \alpha_i^{j+} \geq 0, \quad \alpha_i^{j-} \geq 0, \quad |\alpha_i - \alpha_i^j| = \alpha_i^{j+} + \alpha_i^{j-}, \quad j \in J^t, i = 1, \dots, m, \quad (2.6)$$

вместо внутренней нелинейной задачи минимизации в (2.2) получим эквивалентную линейную задачу минимизации:

$$\sum_{j \in J^t} \sum_{i=1}^m (\alpha_i^{j+} + \alpha_i^{j-}) \rightarrow \min \quad (2.7)$$

при ограничениях (2.6) и  $\alpha^j \in \overline{A}^{jt}, j \in J^t$ .

Теперь, заменяя эту задачу линейного программирования двойственной ей, из максиминной задачи в (2.2) получим задачу с билинейной целевой функцией и линейными ограничениями.

Пример 2.1. Для иллюстрации «работы» предложенного метода рассмотрим расчетный пример, используя определения и решающие правила из теории важности критериев [68 – 71]. Пусть в трехкритериальной задаче критерии однородны и имеют порядковую шкалу с множеством градаций  $Z_0 = \{1, 2, 3\}$ . Множество  $X$  содержит три альтернативы  $x^1, x^2$  and  $x^3$  с векторными оценками  $y^1 = f(x^1) = (2, 1, 3), y^2 = f(x^2) = (3, 2, 1), y^3 = f(x^3) = (1, 2, 3)$ . Информация о важности критериев содержит только одно сообщение о том, что второй критерий важнее третьего. Тогда множество возможных значений коэффициентов важности  $A = \{\alpha \in \mathbb{R}^3 \mid \alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0, \alpha_1 + \alpha_2 + \alpha_3 = 1, \alpha_2 > \alpha_3\}$ .

Проверим на потенциальную недоминируемость альтернативу  $x^3$ , так что  $t = 3$ . Для множеств  $A^{13}$  и  $A^{23}$  имеем выражения, которые нетрудно получить при помощи решающего правила из [70, 71]:  $A^{13} = \{\alpha \in A \mid \alpha_1 > \alpha_2\}, A^{23} = \{\alpha \in A \mid \alpha_1 > \alpha_3\}$ . Эти множества не пусты, и поэтому  $J^t = \{1, 2\}$ . Задачу с целевой функцией (2.7) можно записать в таком виде:

$$\alpha_1^{1+} + \alpha_1^{1-} + \alpha_2^{1+} + \alpha_2^{1-} + \alpha_3^{1+} + \alpha_3^{1-} + \alpha_1^{2+} + \alpha_1^{2-} + \alpha_2^{2+} + \alpha_2^{2-} + \alpha_3^{2+} + \alpha_3^{2-} \rightarrow \min$$

при условиях:

$$\alpha_1^1 + \alpha_2^1 + \alpha_3^1 = 1, \alpha_2^1 - \alpha_3^1 \geq 0, \alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1,$$

$$\alpha_2^2 - \alpha_3^2 \geq 0, \alpha_1^1 - \alpha_2^1 \geq 0, \alpha_1^2 - \alpha_3^2 \geq 0,$$

$$\alpha_1^1 + \alpha_1^{1+} - \alpha_1^{1-} = \alpha_1, \alpha_2^1 + \alpha_2^{1+} - \alpha_2^{1-} = \alpha_2, \alpha_3^1 + \alpha_3^{1+} - \alpha_3^{1-} = \alpha_3,$$

$$\alpha_1^2 + \alpha_1^{2+} - \alpha_1^{2-} = \alpha_1, \quad \alpha_2^2 + \alpha_2^{2+} - \alpha_2^{2-} = \alpha_2, \quad \alpha_3^2 + \alpha_3^{2+} - \alpha_3^{2-} = \alpha_3,$$

все переменные неотрицательны.

Перейдя от этой задачи линейного программирования к двойственной ей задаче с переменными  $z_1, \dots, z_{12}$  и записав ее в качестве внутренней задачи в (2.2), получим задачу билинейного программирования:

$$z_1 + z_3 + \alpha_1 z_7 + \alpha_2 z_8 + \alpha_3 z_9 + \alpha_1 z_{10} + \alpha_2 z_{11} + \alpha_3 z_{12} \rightarrow \max$$

при условиях:

$$z_1 + z_5 + z_7 \leq 0, \quad z_1 + z_2 - z_5 + z_8 \leq 0, \quad z_1 - z_2 + z_9 \leq 0, \quad z_3 + z_6 + z_{10} \leq 0,$$

$$z_3 + z_4 + z_{11} \leq 0, \quad z_3 - z_4 - z_6 + z_{12} \leq 0, \quad z_7 \leq 1,$$

$$-z_7 \leq 1, \quad z_8 \leq 1, \quad -z_8 \leq 1, \quad z_9 \leq 1, \quad -z_9 \leq 1, \quad z_{10} \leq 1, \quad -z_{10} \leq 1, \quad z_{11} \leq 1, \quad -z_{11} \leq 1,$$

$$z_{12} \leq 1, \quad -z_{12} \leq 1,$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1, \quad -\alpha_2 + \alpha_3 \leq 0, \quad z_2 \geq 0, \quad z_4 \geq 0, \quad z_5 \geq 0, \quad z_6 \geq 0, \quad \alpha_1 \geq 0, \quad \alpha_2 \geq 0, \quad \alpha_3 \geq 0.$$

Решив эту задачу при помощи программного пакета OOQP, найдем максимальное значение целевой функции  $\sigma^* = 1,1666665$  при следующих значениях переменных:  $\alpha_1^* = 6,09250165 \cdot 10^{-13}$ ,  $\alpha_2^* = 5,00000000 \cdot 10^{-1}$ ,  $\alpha_3^* = 5,00000000 \cdot 10^{-1}$ , т.е. (с учетом точности расчетов)  $\alpha_1^* = 0$ ;  $\alpha_2^* = 0,5$ ;  $\alpha_3^* = 0,5$ .

Поскольку  $d(\alpha^*, \bar{A}^{13}) = 0,6666667 > 0$  и  $d(\alpha^*, \bar{A}^{23}) = 0,5 > 0$ , то альтернатива  $x^3$  потенциально недоминируема.

Для построения множества потенциально недоминируемых альтернатив достаточно поочередно проверить альтернативы из  $X$ , используя изложенный выше подход. При этом доминируемые по Парето альтернативы можно заранее отбросить.

### 2.1.3 Полезность и стохастическое доминирование

Далее рассматривается задача принятия решения, когда последствия реализации возможных альтернатив (вариантов, планов стратегий, ...) оцениваются при помощи одной числовой количественной характеристики (показателя, признака)  $x$ , которую в экономических задачах чаще всего (хотя и не всегда) рассматривают как денежную сумму. Ее большие значения могут быть более предпочтительными, чем меньшие (тогда в финансовых задачах говорят о доходе), или же менее предпочтительными (и тогда говорят об убытке). В технических задачах этой характеристикой может быть, например, время функционирования (в частности, при оценке надежности устройств), доля технологически обслуженных «потребителей» и т.д. Далее, если специально не оговорено иное, предполагается, что большие значения переменной  $x$  предпочтительнее меньших.

Последствия реализации варианта решения зависят также и от того, какие значения примут неопределенные факторы, распределения вероятностей которых считаются известными (именно это и означает, что решение принимается в условиях вероятностной неопределенности, или риска). Поэтому полагается, что каждый вариант  $v$  характеризуется случайной величиной  $X(v)$  с заданной функцией распределения  $F_{X(v)}(x) = \Pr\{X(v) < x\}$ . Таким образом, проблема сравнения по предпочтительности вариантов решений  $v$  и  $w$  сводится к сопоставлению соответствующих им случайных величин  $Y = X(v)$  и  $Z = X(w)$  с функциями распределения вероятностей  $F(x) = F_{X(v)}(x)$  и  $G(x) = F_{X(w)}(x)$ , или, короче говоря, к сопоставлению функций  $F$  и  $G$ . А исходную задачу выбора оптимального варианта  $v$  из множества всех вариантов  $V$  можно рассматривать как задачу выбора оптимальной случайной величины из множества случайных величин  $\mathbb{X} = \{X(v) \mid v \in V\}$ .

Для решения указанной проблемы требуется привлечение в той или иной форме информации о предпочтениях ЛПР и его отношении к риску. Такую информацию можно представить при помощи функции полезности  $u$  (условия ее существования и свойства см., например, в [72]). Эта функция

позволяет сравнить по предпочтительности любые две случайные величины  $Y$  и  $Z$  путем сопоставления математических ожиданий<sup>1)</sup>  $E[u(Y)]$  и  $E[u(Z)]$  функций случайных величин  $u(Y)$  и  $u(Z)$ :  $F$  предпочтительнее  $G$ , когда  $E[u(Y)] > E[u(Z)]$ . Число  $E[u(Y)]$  принято называть ожидаемой полезностью случайной величины  $Y$ .

Построение функции полезности – задача непростая, требующая получения от ЛПР достаточно сложной информации о его предпочтениях. Гораздо проще определить класс функций полезности  $U$ , к которому принадлежит неизвестная функция полезности ЛПР, и ввести отношение предпочтения  $\succ^U$  следующим образом: соотношение  $F \succ^U G$  верно тогда и только тогда, когда  $E[u(Y)] \geq E[u(Z)]$  для каждой функции полезности  $u \in U$ , причем это неравенство строгое хотя бы для одной функции  $u \in U$ . Заметим, что отношение  $\succ^U$  является частичным: оно не позволяет сравнить по предпочтительности любые две случайные величины, т.е. может не выполняться ни  $F \succ^U G$ , ни  $G \succ^U F$ .

Самым простым классом является множество  $U_1$  всех непрерывных возрастающих функций полезности. Отношение предпочтения  $\succ^{U_1}$  называют отношением *стохастического доминирования первого порядка*. Если имеет место  $F \succ^{U_1} G$ , то говорят, что  $F$  доминирует над  $G$  в смысле доминирования первого порядка, и пишут  $F \succ^{FSD} G$ . Оказывается, что  $F \succ^{FSD} G$  тогда и только тогда, когда верно  $F(x) \leq G(x)$  для каждого  $x \in (-\infty, +\infty)$ , причем это неравенство строгое хотя бы для одного значения  $x$ . Геометрически это условие означает, что график функции  $F(x)$  лежит не выше графика функции  $G(x)$ , причем эти графики не совпадают. Заметим, что введение отношения  $\succ^{FSD}$  можно обосновать и без использования функций полезности. Действительно, если

---

<sup>1)</sup> Через  $E$  обозначен символ операции нахождения математического ожидания. Здесь и далее предполагается, что все вводимые числовые характеристики случайных величин существуют.

верно  $F(x) \leq G(x)$  для каждого  $x \in (-\infty, +\infty)$ , причем это неравенство строгое хотя бы для одного значения  $x$ , то при любом уровне притязаний  $\tau$  вероятность  $\Pr\{Y \geq \tau\}$  получить значение  $x \geq \tau$  для случайной величины  $Y = X(v)$  не меньше, чем аналогичная вероятность  $\Pr\{Z \geq \tau\}$  для  $Z = X(w)$ , и, по крайней мере, для одного значения  $\tau$  такая вероятность будет больше, так как  $\Pr\{Y \geq \tau\} = 1 - F(\tau)$  и  $\Pr\{Z \geq \tau\} = 1 - G(\tau)$ .

Класс  $U_2$  составляют непрерывные возрастающие строго вогнутые функции полезности. Они описывают предпочтения ЛПР, который не склонен к риску, т.е. такого, который для любой случайной величины  $Y$  (не являющейся постоянной) предпочитет получить наверняка величину (сумму), равную ожидаемой полезности  $Y$ , чем участвовать в «лотерее», состоящей в том, что он получит величину (сумму)  $y$ , равную реализовавшемуся значению случайной величины  $Y$ . Это означает, что  $u(E[Y]) > E[u(Y)]$ . Отношение  $\succ^{U_2}$  называют отношением *стохастического доминирования второго порядка*. Если имеет место  $F \succ^{U_2} G$ , то говорят, что  $F$  доминирует над  $G$  в смысле доминирования второго порядка и пишут  $F \succ^{SSD} G$ . Оказывается, что  $F \succ^{SSD} G$  тогда и только тогда, когда  $F^2(x) \leq G^2(x)$  для каждого  $x \in (-\infty, +\infty)$ , причем это неравенство строгое хотя бы для одного значения  $x$ . Здесь  $F^2(x) = \int_{-\infty}^x F(t)dt$  для всех  $x \in (-\infty, +\infty)$  – площадь под графиком функции  $F(t)$  на промежутке от  $-\infty$  до  $x$ .

Отношение стохастического доминирования первого (и второго) порядка является частичным, т.е. для двух вариантов  $v$  и  $w$ , таких, что  $F \neq G$ , может не выполняться ни  $F \succ^{FSD} G$ , ни  $G \succ^{FSD} F$ , (соответственно, ни  $F \succ^{SSD} G$ , ни  $G \succ^{SSD} F$ ).

Отметим, что если верно  $F \succ^{FSD} G$ , то верно и  $F \succ^{SSD} G$ . Более подробно с различными отношениями стохастического доминирования и их практическими приложениями можно ознакомиться по работам [73, 74].

Таким образом, в любом случае выбираемая альтернатива  $v^*$  должен быть недоминируемой в смысле стохастического доминирования первого порядка, т.е. не должно существовать альтернативы  $v \in V$  такого, для которой верно  $F_{X(v)}(x) \succ^{FSD} F_{X(v^*)}(x)$ . А если ЛПР не склонно к риску, то выбираемая альтернатива  $v^*$  должна быть недоминируемой в смысле стохастического доминирования второго порядка, т.е. не должно существовать альтернативы  $v$  такой, для которой верно  $F_{X(v)}(x) \succ^{SSD} F_{X(v^*)}(x)$ .

#### 2.1.4 Меры риска как критерии выбора

Поскольку отношения стохастического доминирования являются частичными, то на практике широко применяются меры риска – такие числовые характеристики случайных величин, которые должны помочь ЛПР (или аналитику) сравнивать варианты решения по предпочтительности и, в итоге, выбрать из них (или рекомендовать для выбора) наилучший. Первые меры риска были предложены и начали использоваться задолго до разработки теории стохастического доминирования. Разумеется, желательно или даже обязательно, чтобы такие характеристики обладали теми или иными свойствами (удовлетворяли определенным требованиям), которые можно выявить (сформулировать) в результате анализа существа исходной задачи. При этом желательность наличия некоторых свойств меры риска может также зависеть от того, кто ей будет пользоваться – ЛПР или же аналитик (консультант). Особо следует подчеркнуть, что меры риска, которые предлагаются для использования ЛПР, должны иметь ясный содержательный смысл – допускать простую и понятную ему интерпретацию. Это требование назовем *интерпретируемостью* меры риска. Многие предлагавшиеся меры достаточно сложны и рассчитаны на аналитиков (финансовых, страховых, технических, ...), а не на ЛПР, которое, вообще говоря, не является специалистом по анализу решений.

Исторически первой и вполне естественной числовой характеристикой, которая в качестве критерия (но, разумеется, не меры риска) стала применяться при анализе решений в условиях вероятностной неопределенности, явилось математическое ожидание случайной величины (кратко говоря, её среднее). Она широко используется при анализе решений, связанных с событиями массового характера: среднее время работы оборудования до его отказа, среднее число потребителей, обслуживаемых за фиксированный промежуток времени, и т.д. Математическое ожидание издавна применяется при оценке эффективности (раньше говорили «действительности») стрельбы (см., например, [59]). Так, при планировании стрельбы по малым целям в качестве критерия эффективности используется математическое ожидание числа пораженных целей. (Цель считается малой, если она не накапливает ущерба и может находиться только в двух состояниях – поражена или не поражена.)

Математическое ожидание является, разумеется, интерпретируемой характеристикой: при достаточно большом числе реализаций случайной величины  $X$  среднее арифметическое ее наблюдённых значений будет «примерно равно»  $m_X$ . Оно обладает многими «хорошими» свойствами, однако не отражает возможных случайных колебаний около среднего. Поэтому был предложен целый ряд специальных числовых характеристик – мер риска. Меры риска бывают специальные (рассчитанные на задачи достаточно узких классов) и «общего назначения». Именно последние, в основном, и рассматриваются далее.

Классификация таких мер в зависимости от их конструкции представлена рисунке 2.1.



Рисунок 2.1 – Классификация мер риска

Характеристики (ожидаемого) *среднего отклонения* оценивают разброс, или рассеивание возможных значений случайной величины относительно центра группирования ее значений или же некоторого целевого (порогового) уровня  $\tau$ . Поскольку случайная величина  $X$  меньше всего отклоняется в смысле среднеквадратического от её математического ожидания  $m_X$ , а в смысле модуля – от её медианы  $\mu_X$ , то наиболее распространенными *двусторонними* мерами риска, входящими в эту группу, являются:

- дисперсия (англ. variance)  $D_X = E[(X - m_X)^2] = \int_{-\infty}^{+\infty} (x - m_X)^2 dF(x)$ ;
- среднее квадратическое, или стандартное отклонение (англ. mean root square deviation, или standard deviation)  $\sigma_X = \sqrt{D_X}$ ;
- средний модуль отклонения от медианы  $\mu_X$  (англ. mean absolute deviation from median)  $M_X = E[|X - \mu_X|] = \int_{-\infty}^{+\infty} |x - \mu_X| dF(x)$ .

Первые две из указанных характеристик распространены несравненно шире третьей.

Общим недостатком указанных мер риска является то, что они одинаково учитывают отклонения как в лучшую сторону от центра группирования, так и в худшую. Для устранения этого недостатка были предложены односторонние меры риска [75]:

- нижняя, или левая, и верхняя, или правая (полу) дисперсии (англ. lower and upper semivariance)  $D_X^- = E[(\max\{0, m_X - X\})^2] = \int_{-\infty}^{m_X} (x - m_X)^2 dF(x)$ ,

$$D_X^+ = E[(\max\{0, X - m_X\})^2] = \int_{m_X}^{+\infty} (x - m_X)^2 dF(x);$$

– нижнее, или левое, и верхнее, или правое средние квадратические (полу)

отклонения (*англ. standard lower and upper semideviations*)  $\sigma_X^- = \sqrt{D_X^-}$ ,

$\sigma_X^+ = \sqrt{D_X^+}$  (если  $X$  – случайный доход, то мерами риска являются  $D_X^-$  и  $\sigma_X^-$

; если же  $X$  – случайный убыток, то меры риска – это  $D_X^+$  и  $\sigma_X^+$ ).

Используется также такая характеристика риска как средний модуль отклонения от среднего (*англ. mean absolute deviation*)  $MAD_X = \int_{-\infty}^{+\infty} |x - m_X| dF(x)$ .

Отклонения значений случайной величины  $X$  влево и вправо от своего математического ожидания  $m_X$  в среднем можно оценивать следующими двумя числовыми характеристиками, связанными с  $MAD_X$ : средним левым, или нижним, и средним правым, или верхним, (полу)отклонениями (*англ. lower and upper mean absolute (semi)deviation*)

$$h_X^- = E[\max\{0, m_X - X\}] = \int_{-\infty}^{m_X} (m_X - x) dF(x),$$

$$h_X^+ = E[\max\{0, X - m_X\}] = \int_{m_X}^{+\infty} (x - m_X) dF(x). \quad (2.8)$$

Оказывается, что величины этих характеристик равны [76], так что можно говорить о такой характеристике, как среднее одностороннее отклонение, или, для краткости речи, среднее полуотклонение (*англ. mean absolute semideviation*):  $h_X = h_X^- = h_X^+$ . Полезно помнить, что среднее полуотклонение  $h_X$  равно половине среднего модуля отклонения от среднего:  $h_X = \frac{1}{2} MAD_X$ . Более подробно среднее полуотклонение, которое относительно недавно начало применяться как мера риска, будет рассмотрено ниже.

В качестве односторонних мер риска используются также средние нижние (полу)отклонения от целевого уровня (или порога)  $\tau$  (*англ. lower target moments*)  $M(\alpha, \tau) = E[(\max\{0, \tau - X\})^\alpha] = \int_{-\infty}^{\tau} (\tau - x)^\alpha dF(x)$ , где  $\tau$  – целевой

уровень (например, минимальный приемлемый доход, катастрофический уровень и т.п.),  $\alpha > 0$  – параметр, отражающий отношение ЛПР к риску (получению значений  $x$  меньших, чем  $\tau$ ) [77, 63].

*Пороговые* характеристики связаны с вероятностями получения значений, больших или же меньших некоторой величины (порога, уровня). Такие характеристики разбиваются на две подгруппы – *целевых* и *квантильных* характеристик, которые в определенном смысле взаимно обратны: для первых фиксируется некоторый «целевой» уровень  $\tau$  и затем исчисляется вероятность его достижения (или недостижения); для вторых же, наоборот, задается вероятность (надежность)  $\gamma$  и затем находится соответствующий ей уровень (квантиль).

В подгруппу целевых мер риска входят:

- вероятность того, что доходы будут меньше, чем уровень притязаний  $\tau$ :  
 $\Pr\{X < \tau\};$
- вероятность потерь:  $\Pr\{X < 0\};$
- вероятность того, что потери достигнут некоторого неприемлемого (в частности, катастрофического) уровня  $\tau$  или даже превзойдут его (если  $X$  – случайный убыток, то это  $\Pr\{X \geq \tau\}$ ).

К подгруппе квантильных мер риска относится такая характеристика, как сумма, или стоимость под риском (англ. Value at Risk). Если  $X$  имеет смысл случайного убытка, то  $\text{VaR}_X^\gamma$  – квантиль уровня  $\gamma$  случайной величины  $X$ , т.е. такое число, что вероятность получения равного ему или большего значения  $X$  равна  $\gamma$  (по своему смыслу вероятность  $\gamma$  должно быть весьма малой, скажем, 0,02 и меньше). Если функция распределения  $F(x)$  – непрерывная и монотонно возрастающая (как, например, в случае нормального распределения), то существует обратная ей функция  $F^{-1}(x)$ , и поэтому  $\text{VaR}_X^\gamma = F^{-1}(1 - \gamma)$ . В общем же случае  $\text{VaR}_X^\gamma = \inf\{x : F(x) \geq 1 - \gamma\}$ . Достоинст-

вом VaR является то, что она имеет ту же физическую размерность, что и  $X$  (например, выражается в денежных единицах).

Пороговые характеристики часто используются при анализе технических решений. Например, при заданной продолжительности времени функционирования невосстанавливаемой системы (аппаратуры космического объекта, ракеты, радиозонда)  $\tau$  ее надежность характеризуется вероятностью безотказной работы за это время. Для оценки боевой эффективности оружия при планировании ударов по площадным целям широко используются такие критерии (показатели эффективности), как вероятность (надежность) поражения не менее заданной части площади цели и часть площади цели, не менее которой будет поражено с заданной вероятностью (надежностью) [59].

Все меры риска второй группы (её обеих подгрупп) являются интерпретируемыми, хотя смысл квантильных характеристик более сложен, чем целевых.

*Комбинированные* характеристики формируются с использованием характеристик первых двух групп. Такими характеристиками являются:

- ожидаемый средний убыток при условии превышения им некоторого уровня  $\tau$  (если  $X$  – убыток, то это  $E[X | X > \tau]$ ); в частности, если  $\tau = 0$ , то оценивается ожидаемый средний убыток, если он случился;
- условное математическое ожидание хвоста [78], получаемое для убытка  $X$  из предыдущей характеристики при  $\tau = \text{VaR}_X^\gamma$ ;
- ожидаемый средний доход при условии, что вероятность того, что потери не превысят катастрофический уровень  $\tau$ , не больше критического уровня  $\gamma$  (если  $X$  – случайный доход, то это  $E[X | \Pr\{X \leq \tau\} \leq \gamma]$ ) (см. [79]);
- мера  $\rho_X = \Pr\{X < 0\} \cdot E[\min\{0, X\}]^\theta$ ; положительный параметр  $\theta$  предначен для описания изменения восприятия убытка с его увеличением [80, 81].

Первая из этих характеристик интерпретируема, вторую и третью можно считать трудно интерпретируемыми, а четвертая интерпретируемой не является.

Для денежных мер риска  $r$  важным считается наличие следующих свойств [81]:

- положительная однородность: если  $Y = kX$ , где  $k > 0$ , то  $r_Y = k r_X$ ;
- инвариантность к сдвигу: если  $Y = X + b$ , то  $r_Y = r_X$ ;
- субаддитивность: если  $Z = X + Y$ , то  $r_Z \leq r_X + r_Y$ .

Положительная однородность, по сути, есть свойство сохранения масштаба: при переходе к другим единицам измерения  $X$  (скажем, от рублей к евро) мера риска должна сохранить свою величину, но выраженную уже в новых единицах. Этим свойством обладают, например,  $\sigma_X$ ,  $\sigma_X^-$ ,  $\sigma_X^+$  и  $VaR_X^\gamma$ , но не  $D_X$ ,  $D_X^-$  и  $D_X^+$ .

Инвариантность к сдвигу для финансовых задач говорит о независимости оценки риска от объема наличных и подразумевает неявное принятие соглашения, что нет ограничения на ликвидность. Это свойство присуще  $D_X$ ,  $\sigma_X$ ,  $D_X^-$  и  $\sigma_X^-$ ,  $D_X^+$  и  $\sigma_X^+$ , а при нормальном распределении и  $VaR_X^\gamma$ .

Требование субаддитивности меры риска для ЛПР, не склонного к риску, выражает идею компенсации риска одних вложений за счет других. Этим свойством обладает  $\sigma_X$ ,  $\sigma_X^-$ ,  $\sigma_X^+$ , но не  $D_X$ ,  $D_X^-$  или  $D_X^+$ , а  $VaR_X^\gamma$  – для нормальных распределений. Отметим, что положительная однородность и субаддитивность обеспечивают мере риска наличие свойства *выпуклости*:  $r_{tX + (1 - t)Y} \leq tr_X + (1 - t)r_Y$ ,  $0 < t < 1$  (смешивание двух финансовых позиций сокращает риск в целом).

Более подробно различные свойства мер риска, в основном финансовых, рассматриваются в [60, 61, 78 – 83].

### 2.1.5 Модели выбора с мерами риска

Классификация математических моделей выбора в зависимости от количества критериев, в число которых входят меры риска, представлена на рисунке 2.2.

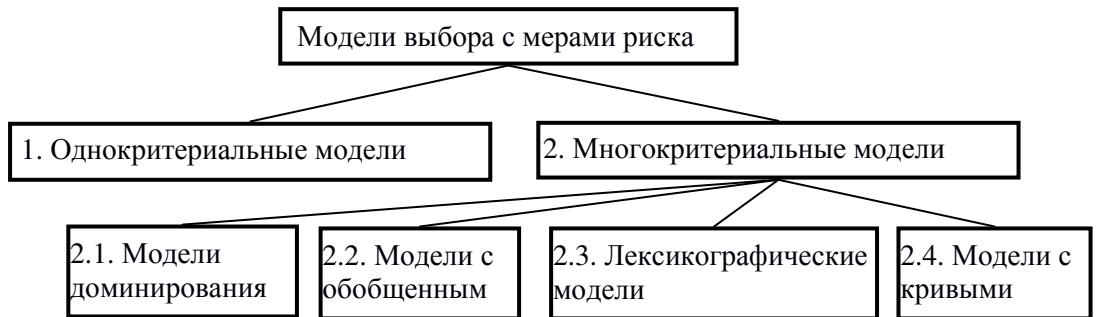


Рисунок 2.2 – Классификация математических моделей выбора с мерами риска

К однокритериальным относятся модели, в которых используется один критерий выбора, в роли которого выступает подходящая мера риска (чаще пороговая), например, вероятность потерь. Рекомендация использовать в качестве критерия выбора подходящую пороговую меру риска составляет содержание *принципа вероятностно гарантированного результата* [84]. В качестве единственного критерия можно выбрать также и среднее нижнее отклонение от целевого уровня  $M(\alpha, \tau)$  или же некоторую пороговую меру, скажем,  $VaR_X^\gamma$ . В роли критерия может выступать и подходящая комбинированная характеристика риска, например,  $E[X | \Pr\{X \leq \tau\} \leq \gamma]$  или же  $\rho_X$ ; частным случаем последнего критерия является мера с  $\theta = 1$  (см. [80, 82]).

Достоинством однокритериальных моделей является простота определения понятия наилучшей (оптимальной) альтернативы: она должна доставлять выбранному критерию, в зависимости от его смысла, наибольшее или же наименьшее значение. Недостаток их связан с тем, что найти один критерий – меру риска, достаточно полно характеризующую варианты решения в рассматриваемой конкретной задаче выбора, далеко не всегда возможно.

В многокритериальных моделях используются несколько числовых характеристик случайных величин, обычно это математическое ожидание (среднее) и одна или несколько мер риска. Если предпочтения возрастают при увеличении значения случайной величины (как, например, в случае дохода), то среднее желательно максимизировать, а меры риска – минимизировать.

В моделях доминирования участвуют обычно два критерия – математическое ожидание  $m$  (или иная характеристика среднего положения) и мера риска  $r$ ; такие модели называются моделями «среднее-риска» (англ. mean-risk model). В рассматриваемых моделях одна альтернатива считается предпочтительнее другой, если она не хуже ее по совокупности обоих критериев и, по крайней мере, по одному лучше. Это означает, что вводится отношение предпочтения  $\succ^{m,r}$ , которое определяется следующим образом:  $F \succ^{m,r} G$  в том и только том случае, когда  $m_Y \geq m_Z$  и  $r_Y \leq r_Z$ , причем хотя бы одно неравенство выполняется как строгое. Это отношение является частичным. Поэтому из множества  $V$  имеющихся в распоряжении ЛПР вариантов выделяется множество недоминируемых вариантов  $V^0$ : из этого множества ЛПР и должен эвристически (т.е. на основе не формализуемых представлений о своих предпочтениях и отношении к риску) выбрать один наилучший. Вариант  $v^*$  называется недоминируемым, если не существует другого варианта  $v \in V$ , который лучше его в смысле отношения по  $\succ^{m,r}$ , т.е. для которого выполнялось бы соотношение  $F_{X(v)} \succ^{m,r} F_{X(v^*)}$ . Заметим, что в теории принятия многокритериальных решений отношение  $\succ^{m,r}$  называется отношением Парето, а множество недоминируемых вариантов – множеством Парето, или множеством Эджворта-Парето [85].

Существует аксиоматическое обоснование применимости модели «среднее-отклонение» [86].

В многокритериальных моделях с обобщенным критерием на основе двух или более мер риска или же математического ожидания (или медианы) и подходящих мер риска (обычно одной) строится некоторый единый (обобщенный, агрегированный, глобальный и т.п.) критерий, который, в зависимости от его смысла, желательно максимизировать или же минимизировать. Предлагались, например, такие критерии, как [2, 75]:

$$m_X - \lambda D_X \text{ или же } m_X - \lambda \sigma_X;$$

$$m_X - \lambda D_X^- \text{ или же } m_X - \lambda \sigma_X^-.$$

Величина положительного коэффициента замещений  $\lambda$  должна отражать отношение ЛПР к риску. Использование в перечисленных обобщенных критериях  $\sigma_X$  и  $\sigma_X^-$  вместо  $D_X$  и  $D_X^-$  соответственно представляется более естественным (например,  $\sigma_X$  и  $\sigma_X^-$ , как и  $m_X$ , имеют ту же размерность, что и  $X$ ).

Исторически применение  $m_X - \lambda D_X$  было инициировано моделью выбора оптимального портфеля Марковица с парой среднее-дисперсия [2].

Следует отметить, что нередко существует тесная, хотя и не всегда сразу заметная взаимосвязь однокритериальных моделей и моделей с обобщенным критерием, особенно для специальных классов распределений. Предположим, что случайные величины, характеризующие варианты решения, подчинены нормальному закону, т.е. выбор осуществляется из заданного множества нормальных случайных величин  $X$ . Для таких величин – случайных убытков, можно записать:

$$1 - \gamma = \Pr\{X < \text{VaR}_X^\gamma\} = \Pr\left\{\frac{X - m_X}{\sigma_X} < \frac{\text{VaR}_X^\gamma - m_X}{\sigma_X}\right\}.$$

Но нормальная величина  $(X - m_X)/\sigma_X$  является стандартной (имеет математическое ожидание 0 и дисперсию 1), и поэтому

$$\frac{\text{VaR}_X^\gamma - m_X}{\sigma_X} = \alpha^{1-\gamma},$$

где  $\alpha^{1-\gamma}$  – квантиль уровня  $1-\gamma$  стандартной нормальной величины. Следовательно,  $\text{VaR}_X^\gamma = m_X + \alpha^{1-\gamma}\sigma_X$ , так что минимизация критерия  $\text{VaR}_X^\gamma$  эквивалентна минимизации обобщенного критерия  $m_X + \alpha^{1-\gamma}\sigma_X$ . Это проясняет возможности практического применения обобщенного критерия  $m_X + \lambda\sigma_X$  (см., например, [61]). Отметим, что аналогичным путем можно убедиться в том, что если выбор осуществляется из заданного множества показательно распределенных случайных величин, то минимизация критерия  $\text{VaR}_X^\gamma$  эквивалентна минимизации критерия  $\frac{\alpha^{1-\gamma}}{\lambda} (= \alpha^{1-\gamma}m_X = \alpha^{1-\gamma}\sigma_X)$ , где  $\alpha^{1-\gamma}$  – квантиль уровня  $1-\gamma$  «стандартной показательной» случайной величины (с  $\lambda=1$ ), что, в свою очередь, эквивалентно выбору случайной величины с наибольшим значением параметра  $\lambda$ . Это – вполне понятный результат.

Рассмотрим еще такой критерий, как  $\Pr\{X < \tau\}$  – вероятность получить случайный доход  $X$  меньше, чем уровень притязаний  $\tau$ . Так как

$$\Pr\{X < \tau\} = \Pr\left\{\frac{X - m_X}{\sigma_X} < \frac{\tau - m_X}{\sigma_X}\right\},$$

то для нормальных случайных величин минимизация критерия  $\Pr\{X < \tau\}$  эквивалентна минимизации обобщенного критерия  $(\tau - m_X)/\sigma_X$ .

Для логнормальной случайной величины  $Y = \exp X$ , где  $X$  – нормально распределенная случайная величина, можно записать:

$$\Pr\{Y < \tau\} = \Pr\{\exp X < \tau\} = \Pr\{X < \ln \tau\} = \Pr\left\{\frac{X - m_X}{\sigma_X} < \frac{\ln \tau - m_X}{\sigma_X}\right\}.$$

Поэтому для случайных логнормальных доходов минимизация критерия  $\Pr\{Y < \tau\}$  эквивалентна минимизации обобщенного критерия  $(\ln \tau - m_X)/\sigma_X$  (по заданным  $m_Y$  и  $\sigma_Y$  несложно рассчитать  $m_X(m_Y, \sigma_Y)$  и  $\sigma_X(m_Y, \sigma_Y)$ ).

*Многокритериальные лексикографические* модели опираются на лексикографический порядок. Например, если рассматривается случайный убыток, то при использовании двух критериев выбора – вероятности катастрофических потерь  $\text{Pr}^*$  и средних ожидаемых потерь, первый из которых несравнен-

но важнее второго, случайная величина  $Y$  будет считаться предпочтительнее, чем  $Z$ , если  $\Pr_Y^\bullet < \Pr_Z^\bullet$  или же ( $\Pr_Y^\bullet = \Pr_Z^\bullet$  и  $m_Y < m_Z$ ), где  $\Pr_Y^\bullet$  – вероятность катастрофических потерь для  $Y$ .

*Многокритериальные* модели с кривыми безразличия обычно являются двухкритериальными. В них предпочтения ЛПР описываются при помощи набора кривых безразличия на координатной плоскости (обычно в некоторой области первого квадранта), например,  $v(m_X, \sigma_X) = C$ ,  $C \in \{C_1, \dots, C_n\}$ . Любые два варианта решения, которым соответствуют точки, лежащие на одной и той же кривой безразличия, имеют равную ценность, т.е. являются одинаковыми по предпочтительности (безразличными).

Отметим, что в моделях всех рассмотренных типов, кроме многокритериальных моделей доминирования, любые два варианта (две случайные величины) сравнимы по предпочтительности. Такие модели в [79] названы моделями полного упорядочения.

Если для вариантов  $v, w$  соответствующие функции распределения  $F(x)$  и  $G(x)$  случайных величин  $Y = X(v)$  и  $Z = X(w)$  удовлетворяют условию  $F \succ^{FSD} G$  (или же условию  $F \succ^{SSD} G$ ), то вариант  $v$  должен считаться более предпочтительным, чем  $w$ . Поэтому при использовании той или иной модели выбора с мерами риска вариант  $v$  также должен оказаться более предпочтительным, чем  $w$ . Это требование называют условием *монотонности модели по стохастическому доминированию первого (соответственно, второго) порядка*. Отметим, что  $F \succ^{FSD} G$  (соответственно,  $F \succ^{SSD} G$ ) влечет выполнение неравенства  $m_Y > m_Z$  (соответственно,  $m_Y \geq m_Z$  (см. [76])).

В моделях доминирования «среднее – дисперсия», «среднее – нижняя дисперсия», «среднее – стандартное отклонение» и «среднее – нижнее стандартное отклонение» в случае нормальных случайных величин условия монотонности по стохастическому доминированию первого порядка выполняются. Это объясняется тем, что если обе случайные величины  $Y$  и  $Z$  нормаль-

ны, то  $F \succ^{FSS} G$ , если и только если  $m_Y > m_Z$  и  $\sigma_Y = \sigma_Z$  (рассматривая графики функций  $F$  и  $G$ , можно увидеть, что они не пересекаются). Однако в случае произвольных распределений оно может нарушаться. Несложный поучительный пример имеется в [61]. Приведем здесь совсем простой пример из [87]. Пусть случайная величина  $Y$  принимает значения 3 и 5 с равными вероятностями  $\frac{1}{2}$ , а величина  $Z$  является постоянной и равной 1. Разумеется,  $Y$  предпочтительнее, чем  $Z$ . Понятно также, что  $F \succ^{FSS} G$ . Однако эти случайные величины являются несравнимыми по отношению  $\succ^{m,\sigma}$ , так как  $m_Y = 4 > 1 = m_Z$  и  $\sigma_Y = 1 > 0 = \sigma_Z$ . Отметим также, что и  $\sigma_Y^- = \frac{1}{2} > 0 = \sigma_Z^-$ .

Для моделей доминирования «среднее – среднее нижнее отклонение от целевого уровня» справедливы следующие два утверждения [63, 77]:

- а) если в модели с мерой риска  $M(\alpha, \tau)$ , где  $\alpha > 0$ , верно  $F \succ^{FSD} G$ , то имеют место неравенства  $m_Y \geq m_Z$  и  $M_Y(\alpha, \tau) \leq M_Z(\alpha, \tau)$ , хотя бы одно из которых строгое;
- б) если в модели с обобщенным критерием  $M(\alpha, \tau)$ , где  $\alpha \geq 1$ , верно  $F \succ^{SSD} G$ , то имеют место неравенства  $m_Y \geq m_Z$  и  $M_Y(\alpha, \tau) \leq M_Z(\alpha, \tau)$ , хотя бы одно из которых строгое, исключая случай, когда  $m_Y = m_Z$  и  $M_Y(\alpha, \tau) = M_Z(\alpha, \tau)$ .

В случае нормальных случайных величин использование любого обобщенного критерия  $\varphi(m_X, \sigma_X)$ , построенного на основе функции  $\varphi$ , которая является возрастающей по первой переменной и убывающей по второй (в частности, критерия  $m_X - \lambda\sigma_X$ , где  $\lambda > 0$ ) условия монотонности по стохастическому доминированию первого порядка выполняется по той же причине, которая указана выше для моделей доминирования. Для общего случая произвольных случайных величин для моделей с обобщенными критериями вида  $m_X - \lambda\sigma_X^-$ , где  $\lambda > 0$ , можно указать простые условия выполнения монотонности по стохастическому доминированию [76, 88]:

- а) если  $\lambda \leq 1$  и верно  $F \succ^{FSD} G$ , то имеет место неравенство  $m_Y - \lambda\sigma_Y^- > m_Z - \lambda\sigma_Z^-$ ;
- б) если  $\lambda \leq 1$  и верно  $F \succ^{SSD} G$ , то имеет место неравенство  $m_Y - \lambda\sigma_Y^- \geq m_Z - \lambda\sigma_Z^-$ , причем это неравенство строгое, если и неравенство  $m_Y \geq m_Z$  строгое.

Приведенные утверждения позволяют при использование подхода «среднее – риск» относительно просто выделить среди эффективных те варианты (точки на эффективной границе), которые являются недоминируемыми по стохастическому доминированию обоих порядков: надо найти точку максимума на  $V$  функции  $m(v) - \sigma^-(v)$ , и она отделит эффективные по стохастическому доминированию точки от остальных эффективных точек.

Выбор модели с мерами риска, которая будет использоваться при анализе конкретной задачи, а также назначение входящих в нее параметров возлагается на ЛПР, которому может помочь аналитик. А это далеко не простая проблема. Указанная проблема осложняется тем, что результат может оказаться весьма чувствительным к величине назначаемого параметра, особенно для пороговых мер риска. Так, например, значение меры риска  $VaR_X^\gamma$  может сильно меняться даже при малом изменении вероятности  $\gamma$ . Вопрос об адекватности моделей и мер риска применительно к реальным задачам выбора в условиях неопределенности пока далек от своего разрешения [79].

Обычно та или иная мера риска предлагалась из соображений здравого смысла с учетом присущих ей свойств, а также особенностей задачи некоторого класса. Но возможен и аксиоматический подход к анализу мер риска как критериев выбора. Он был осуществлен, например, в [82, 89]. Однако, как справедливо отмечено в [82], система аксиом, определяющих меру риска (2.8) как единственный критерий для сравнения вариантов по предпочтительности, слишком многочисленна (двенадцать аксиом!) и среди этих аксиом есть весьма сильные.

## 2.1.6 Среднее полуотклонение и его применение как критерия выбора

Согласно (2.8), для среднего полуотклонения  $h_X$  случайной величины  $X$  с учетом равенства  $h_X^- = h_X^+$  справедливы следующие формулы, в которых, как и ранее,  $F(x)$  – функция распределения вероятностей:

$$h_X = \int_{-\infty}^{m_X} (m_X - x) dF(x), \quad h_X = \int_{m_X}^{+\infty} (x - m_X) dF(x). \quad (2.9)$$

Смысл равенства  $h_X^- = h_X^+$  проясняет, с учетом (2.9), следующая механическая интерпретация (см. рисунок 2.3).

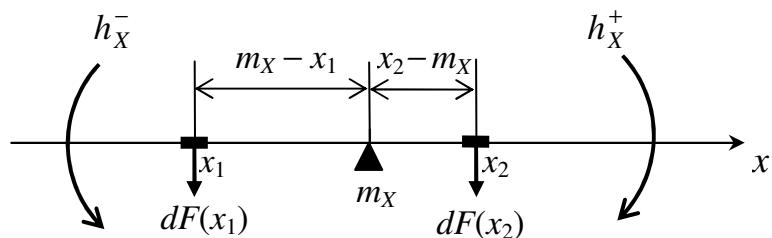


Рисунок 2.3 – Механическая интерпретация равенства левого и правого

Если полагать, что на невесомой оси  $Ox$  распределена «единичная масса» в соответствии с функцией  $F(x)$ , то  $m_X$  – координата центра тяжести такой системы, а  $h_X^-$  и  $h_X^+$  – моменты сил, направленные против и по часовой стрелке соответственно. Поскольку такая система находится в равновесии, то выполняется равенство  $h_X^- = h_X^+$ .

Среднее полуотклонение не только является хорошо интерпретируемой характеристикой, но и обладает целым рядом свойств, важных для приложений [76]:

- однородность по модулю (а потому и положительная однородность): если  $Y = kX$ , то  $h_Y = |k| h_X$ . Благодаря этому свойству при замене  $X$  на  $-X$  величина среднего полуотклонения не меняется. Это означает, например, что среднее полуотклонение остается неизменным при переходе от рассмотрения доходов к рассмотрению убытков, и наоборот;

- инвариантность к сдвигу: если  $Y = X + b$ , то  $h_Y = h_X$ ;
- субаддитивность: если  $Z = X + Y$ , то  $h_Z \leq h_X + h_Y$ ;
- выпуклость:  $h_{tX + (1-t)Y} \leq tR_X + (1-t)R_Y$ ,  $0 \leq t \leq 1$ .

Отметим, что, в силу разбираемых свойств, случайная величина  $Y = (X - m_X)/h_X$  является центрированной и нормированной в том смысле, что  $m_Y = 0$  и  $h_Y = 1$ .

Приведем еще несколько свойств среднего полуотклонения. Теоретический и практический интерес представляет вопрос о соотношении величин  $h_X$  и  $\sigma_X$ . Известно [76], что справедливо неравенство  $MAD_X \leq \sigma_X$ , и поэтому  $h_X \leq \sigma_X/2$ . Эту оценку можно уточнить.

Теорема 2.2. Справедлива неулучшаемая оценка:

$$h_X \leq \frac{1}{2} \sigma_X \sqrt{1 - p_X^0}. \quad (2.10)$$

Здесь  $p_X^0$  есть вероятность того, что значение случайной величины  $X$  будет равно ее математическому ожиданию:  $p_X^0 = \Pr\{X = m_X\}$ ; для непрерывной случайной величины эта вероятность равна нулю. Доказательство теоремы опускаем.

То, что оценка (2.10) является неулучшаемой, показывает следующий пример. Для случайной величины  $X^0$ , принимающей значение 0 с вероятностью  $p^0 \in [0, 1)$  и значения  $-1$  и  $1$  с равными вероятностями  $(1 - p^0)/2$ , имеем  $h_X = \sigma_X \sqrt{1 - p^0}/2$ .

Соотношение между средним полуотклонением и односторонними средними квадратическими отклонениями можно представить неравенством с использованием вероятностей  $p_X^- = \Pr\{X < m_X\}$  и  $p_X^+ = \Pr\{X > m_X\}$ .

Теорема 2.3. Справедлива неулучшаемая оценка:

$$h_X \leq \min\{\sigma_X^- \sqrt{p_X^-}, \sigma_X^+ \sqrt{p_X^+}\}. \quad (2.11)$$

Доказательство теоремы опускаем. То, что оценка (2.11) неулучшаема, показывает пример случайной величины  $X^0$ , введенной при рассмотрении теоремы 2.2: для  $X^0$  нестрогое неравенство (2.11) обращается в равенство.

Как уже подчеркивалось, важным требованием к мерам риска как критериям выбора является их интерпретируемость. Среднее полуотклонение является значительно проще интерпретируемой, чем любая из вышерассмотренных мер риска: при достаточно большом числе реализаций случайной величины  $X$  среднее арифметическое ее наблюдённых отклонений от  $m_X$  в меньшую сторону (как и среднее арифметическое ее наблюдённых отклонений в большую сторону) будет «примерно равно»  $h_X$ . Следует, однако, иметь в виду недостаток, присущий всем односторонним мерам риска: они учитывают распределение вероятностей только с одной стороны от выбранного уровня или центра группирования, и поэтому, например, могут оказаться равными для случайных величин с совершенной разными распределениями с другой стороны от него.

Поскольку  $h_X = \frac{1}{2} \text{MAD}_X$ , то значения среднего полуотклонения  $h_X$  или же расчетные формулы для того или иного распределения вероятностей можно сразу получить на основе соответствующих известных значений или формул для  $\text{MAD}_X$ . Например, для нормального распределения  $\text{MAD}_X = \sqrt{\frac{2}{\pi}} \sigma_X$ , и поэтому  $h_X = \sigma_X / \sqrt{2\pi}$ .

Поскольку среднее полуотклонение неразрывно связано с математическим ожиданием, то при анализе задач принятия решений в условиях вероятностной неопределенности целесообразно применять их совместно. Так как среднее полуотклонение имеет ряд полезных свойств, не присущих односторонним средним квадратическим отклонениям, то его как один из критериев выбора целесообразно использовать вместо них в известных многокритериальных моделях.

Будет ли модель доминирования с двумя критериями  $m_X$  и  $h_X$  удовлетворять условию монотонности по стохастическому доминированию первого порядка? К сожалению, нет. Это подтверждает рассмотренный выше пример

двух случайных величин: случайная величина  $Y$  принимает значения 3 и 5 с равными вероятностями  $\frac{1}{2}$ , а величина  $Z$  является постоянной и равной 1. Для них верно соотношение  $F \succ^{FSS} G$ . Однако эти случайные величины являются несравнимыми по отношению  $\succ^{m,h}$ , так как  $m_Y = 3 > 0 = m_Z$  и  $h_Y = \frac{1}{2} > 0 = h_Z$ .

Рассмотрим еще один пример с двумя случайными величинами, разбившийся в [61] для пары  $(m_X, \sigma_X)$ . Пусть две точки  $(m_1, h_1)$  и  $(m_2, h_2)$ , где  $m_1 > m_2$ ,  $h_1 > h_2$ , являются одинаковыми по предпочтительности (лежат на одной кривой безразличия). Пусть  $Z$  – случайная величина, которая может с вероятностями  $p \in (0, 1)$  и  $1 - p$  принимать значения  $a$  и  $b$ , а  $Y$  – случайная величина, которая может принимать с такими же вероятностями  $p$  и  $1 - p$  значения  $a$  и  $c$ . Если  $a < b < c$ , то из этих двух случайных величин вторая, очевидно, предпочтительнее первой. Подберем значения четырех параметров  $a, b, c$  и  $p$  так, чтобы математические ожидания и средние полуотклонения этих случайных величин были равны заданным числовым характеристикам, для чего запишем систему из четырех уравнений

$$m_Y = ap + c(1 - p) = m_1, \quad m_Z = ap + b(1 - p) = m_2;$$

$$h_Y = (m_2 - a)p = h_1, \quad h_Z = (m_1 - a)p = h_2.$$

Эта система имеет единственное решение:

$$p = \frac{h_1 - h_2}{m_1 - m_2}, \quad a = \frac{m_2 h_1 - m_1 h_2}{h_1 - h_2}, \quad b = a + \frac{m_2 - a}{1 - p}, \quad c = a + \frac{m_1 - a}{1 - p},$$

где, разумеется,  $a < b < c$ . Условие  $p < 1$  выполняется при  $m_1 - m_2 > h_1 - h_2$ . Таким образом, условие монотонности по стохастическому доминированию первого порядка при  $m_1 - h_1 > m_2 - h_2$  оказывается нарушенным: выполнено  $F \succ^{FSD} G$ , хотя, по предположению, точки  $(m_1, h_1)$  и  $(m_2, h_2)$  лежат на одной кривой безразличия.

Условие  $m_1 - h_1 > m_2 - h_2$  в разобранном примере появилось не случайно. Оно объясняется первым из следующих двух утверждений [61, 90]:

а) если  $\lambda \leq 1$  и верно  $F \succ^{FSD} G$ , то имеет место неравенство  $m_Y - \lambda h_Y > m_Z -$

$\lambda h_Z$ ;

б) если  $\lambda \leq 1$  и верно  $F \succ^{SSD} G$ , то имеет место неравенство  $m_Y - \lambda h_Y \geq m_Z -$

$\lambda h_Z$ , причем это неравенство строгое, если и неравенство  $m_Y \geq m_Z$  строгое.

Из этих утверждений следует, в частности, что при использовании  $MAD_X$  в качестве  $r_X$  в обобщенном критерии  $m_X - \lambda r_X$  монотонность по стохастическому доминированию первого и второго порядков обеспечивается при  $\lambda \leq \frac{1}{2}$ .

Рассмотренные свойства среднего полуотклонения показывают, что его целесообразно использовать в составе обобщенных критериев вместо  $\sigma^-$  и  $MAD$ . В частности, его рекомендуется применять и при анализе портфелей [91].

Однако нужно не забывать, что среднему полуотклонению присущ общий недостаток односторонних мер риска, указанный выше.

### 2.1.7 Некоторые выводы

В этом подразделе отчета приведены результаты по разработке метода проверки альтернатив на потенциальную недоминируемость. На его основе можно строить множества таких альтернатив в дискретных многокритериальных задачах выбора.

Дан аналитический обзор мер риска, применяемых в качестве критериев для оценки альтернатив в задачах принятия решений в условиях вероятностной неопределенности; составлена классификация таких мер риска; установлены несколько новых свойств среднего (абсолютного) полуотклонения и предложена его механическая интерпретация; дан аналитический обзор мно-

гокритериальных моделей выбора с мерами риска и разработана классификация таких

Поставленные перед началом исследований задачи выполнены полностью.

Результаты проведенных исследований нашли отражение в статье [92] и докладах [93] и [94].

## 2.2 Сравнительное исследование и аксиоматическое описание концепций решений в задаче оптимального коллективного выбора

### 2.2.1 Сравнительное исследование и аксиоматическое описание некоторых турнирных решений

Объектом данного исследования были три схожие функции коллективного выбора, зависящие от коллективных предпочтений, моделируемых бинарным отношением (так называемые турнирные решения): объединение минимальных Р-доминирующих множеств, объединение минимальных Р-внешнеустойчивых множеств и объединение минимальных Р-внешнеустойчивых множеств. Эти функции до сих не привлекали большого внимания теоретиков коллективного выбора. Задачей данного исследования было определить наличие у данных решений свойств, важных с точки зрения оптимальности совершаемого от имени коллектива выбора.

Напомним, что основной задачей теории коллективного выбора является описание способов определения альтернатив, которые или *будут* выбраны, или *должны быть* выбраны из числа имеющихся в наличии вариантов на основании мнения о них индивидуальных участников процесса принятия коллективных решений.

Пусть дано множество доступных для выбора альтернатив  $A$ ,  $|A|=m$ ,  $m \geq 3$ , и группа  $G$ ,  $|G|=n$ ,  $n \geq 2$ , лиц, участвующих в процессе принятия коллективного решения, которое сводится к выбору определенных альтернатив из  $A$ . Мнение отдельного участника  $i$ ,  $i \in G$ , об альтернативах из  $A$ , определяющее

его индивидуальный выбор, моделируется бинарным отношением  $P_i$  на  $A$ ,  $P_i \subseteq A \times A$ , фиксирующим результаты попарного сравнения альтернатив. Если при сравнении пары альтернатив  $x$  и  $y$  участник  $i$  отдает предпочтение альтернативе  $x$ , то говорят, что упорядоченная пара  $(x, y)$  принадлежит отношению  $P_i$ ,  $(x, y) \in P_i$ , или, что альтернатива  $x$   $P_i$ -доминирует над альтернативой  $y$ , т.е. доминирует по отношению  $P_i$ ,  $xP_iy$ . Если выбирающий не способен решить, какая из двух альтернатив лучше, или считает их равноценными, то будем полагать, что он не предпочитает ни одну из них другой, то есть что справедливо утверждение  $(x, y) \notin P_i \& (y, x) \notin P_i$ .

Если известны предпочтения того, кто делает выбор (т.е. бинарное отношение на  $A$ ), и если задано правило выбора, определенное как функция, отображающая множество бинарных отношений на  $A$  во множество непустых подмножеств  $A$ , то можно предсказать, какие альтернативы должны стать результатом выбора. Таким образом, зная индивидуальные предпочтения, задачу коллективного выбора можно решить, если, во-первых, определить бинарное отношение  $P$ ,  $P \subseteq A \times A$ , моделирующее строгие коллективные предпочтения (мнение коллектива об альтернативах из  $A$ ), а во-вторых, задать правило выбора  $S(\mu, A)$ :  $\{P\} \rightarrow 2^A \setminus \emptyset$ , называемое также решением. Часто отношение  $P$  строится из отношений предпочтений индивидуальных участников с помощью мажоритарного правила и поэтому называется мажоритарным отношением:  $x P$ -доминирует над  $y$ , если число тех членов группы, кто предпочитает альтернативу  $x$  альтернативе  $y$ , больше числа тех, кто предпочитает альтернативу  $y$  альтернативе  $x$ ,  $xPy \Leftrightarrow |G_1| > |G_2|$ , где  $G_1 = \{i \in G \mid xP_iy\}$ ,  $G_2 = \{i \in G \mid yP_ix\}$ .

Из определения следует, что отношение  $P$  асимметрично,  $(x, y) \in P \Rightarrow (y, x) \notin P$ . Если имеет место  $(x, y) \in P \& (y, x) \in P$ , то говорят, что альтернативы  $x$  и  $y$  находятся в отношении равенства голосов  $T$ ,  $T \subseteq A \times A$ ,  $(x, y) \in T \& (y, x) \in T$ . Очевидно, что отношение равенства голосов  $T$  - симметрично. Объединение

отношений  $P$  и  $T$  называется нестрогими коллективными предпочтениями и обозначается  $R$ ,  $R=P\cup T$ .

Для проведения компьютерных вычислений мажоритарное отношение  $P$  представляется с помощью мажоритарной матрицы  $\mathbf{M}=[m_{xy}]$ , определяемой так:

$$m_{xy}=1 \Leftrightarrow (x, y) \in P, \text{ или } m_{xy}=0 \Leftrightarrow (x, y) \notin P.$$

Аналогично мажоритарной матрице  $\mathbf{M}$  можно построить матрицу  $\mathbf{T}=[t_{ij}]$ , представляющую отношение равенства голосов  $T$ .

Для определения правил выбора необходимы также понятия нижнего среза, верхнего среза и горизонта альтернативы  $x$ . Нижним срезом альтернативы  $x$  называется множество  $L(x)$  альтернатив, над которыми  $x$   $P$ -доминирует,  $L(x)=\{y \mid xPy\}$ , верхним срезом – множество  $D(x)$  альтернатив,  $P$ -доминирующих над  $x$ ,  $D(x)=\{y \mid yPx\}$ , горизонтом  $x$  – множество  $H(x)$  альтернатив, находящихся с  $x$  в отношении равенства голосов,  $H(x)=\{y \mid yTx\}$ .

Исследуемые решения основаны на принципе выбора альтернатив, принадлежащих множеству с каким-либо хорошим свойством.

Подмножество  $B$  множества  $A$  называется  $P$ -доминирующим, если для любой альтернативы  $x$  из  $A$  во множестве  $B$  найдется альтернатива  $y$ , которая лучше (для большинства), чем  $x$ ,  $\forall x \in A, \exists y: y \in B \& yPx$  [95].

Подмножество  $B$  множества  $A$  называется  $P$ -внешнеустойчивым, если для любой альтернативы  $x$ , не принадлежащей  $B$ , в множестве  $B$  найдется альтернатива  $y$ , которая лучше (для большинства), чем  $x$ ,  $\forall x \notin B, \exists y: y \in B \& yPx$  [96].

Подмножество  $B$  множества  $A$  называется  $R$ -внешнеустойчивым, если для любой альтернативы  $x$ , не принадлежащей  $B$ , во множестве  $B$  найдется альтернатива  $y$ , которая не хуже (для большинства), чем  $x$ ,  $\forall x \notin B, \exists y: y \in B \& (yPx \vee yTx)$  [97].

Множество называется минимальным относительно некоторого свойства, если ни одно из его подмножеств, кроме него самого, этим свойством не

обладает. Наилучшей считается альтернатива, принадлежащая хотя бы одному минимальному множеству, поэтому решениями являются объединение минимальных  $P$ -доминирующих множеств  $MPD$ ; объединение минимальных  $P$ -внешнеустойчивых множеств  $MPES$  и объединение минимальных  $R$ -внешнеустойчивых множеств  $MPES]$ . Минимальное  $P$ -внешнеустойчивое множество было введено Субочевым [97] по аналогии с другим решением – минимальным слабоустойчивым множеством ( $MWS$ ), предложенным в [98]. Поэтому в работах [97] и [99] минимальное внешнеустойчивое множество называется второй версией минимального слабоустойчивого множества и обозначается  $MWS^II$ . Минимальное  $R$ -внешнеустойчивое множество было введено Алекскеровым и Субочевым [100] по аналогии с  $MPES$ . Минимальное  $P$ -доминирующее множество было введено Субочевым в [101] на основании определения Дуггана [95].

Две характеристические теоремы, доказанные ранее в [97] и [100], и одна теорема, доказанная в ходе настоящего исследования [101], связывают три вышеназванных решения с другим широко известным турнирным решением – непокрытым множеством.

Последнее решение основано на следующем принципе выбора: усилим условие доминирования, а затем выберем недоминируемые альтернативы. Если не усиливать условие доминирования, то выбор недоминируемой (по отношению  $P$ ) альтернативы в общем случае невозможен, так как таких альтернатив может не быть. Этот результат известен как «парадокс Кондорсе». См., например, [102].

Говорят, что альтернатива  $x$  строго покрывает альтернативу  $y$  (по Миллеру), если  $x$  доминирует и над  $y$ , и над всеми альтернативами, над которыми доминирует  $y$ ,  $xPy \& L(y) \subseteq L(x)$ ; если  $x$  доминирует над всеми альтернативами, над которыми доминирует  $y$ ,  $L(y) \subseteq L(x)$ , то говорят, что  $x$  покрывает  $y$  (по Миллеру) нестрого [41]. Т.е.  $x$  строго покрывает  $y$ , если для большинства голосующих  $x$  лучше, чем  $y$ , и не существует такой альтернативы  $z$ , которая была бы хуже (для большинства), чем  $y$ ,  $yPz$ , и одновременно не хуже чем  $x$ ,

$zPx \vee zTx$ . Наилучшими являются альтернативы, не покрытые строго/нестрого по Миллеру никакими другими альтернативами, их множество называется широким/узким непокрытым множеством Миллера  $UC_M$ . Существуют и другие версии определения отношения покрытия и, соответственно, непокрытого множества. Говорят, что альтернатива  $x$  строго покрывает альтернативу  $y$  (по Фишберну), если  $x$  доминирует над  $y$ , а все альтернативы, доминирующие над  $x$ , также доминируют и над  $y$ ,  $xPy \& D(x) \subseteq D(y)$ ; если все альтернативы, доминирующие над  $x$ , доминируют и над  $y$ ,  $D(x) \subseteq D(y)$ , то говорят, что  $x$  покрывает  $y$  (по Фишберну) нестрого [39]. Другие версии отношения покрытия можно посмотреть, например, в [100].

Для  $MPD$  в настоящем исследовании была доказана характеристическая теорема, аналогичная теоремам для  $MPES$  и  $MRES$ : альтернатива  $x$  принадлежит  $MPD$  тогда и только тогда, когда она  $P$ -доминирует альтернативу, принадлежащую узкому непокрытому множеству Фишбера. С помощью этого утверждения было показано, что, в отличие от  $MPES$  и  $MRES$ , объединение минимальных  $P$ -доминирующих множеств не связано отношением вложения с непокрытым множеством в случае конечного турнира, а, следовательно, ни с одной из его версий в общем случае.

В настоящем исследовании также была упрощена формулировка характеристической теоремы для  $MRES$ : альтернатива  $x$  принадлежит  $MRES$  тогда и только тогда, когда она  $R$ -доминирует альтернативу, принадлежащую узкому непокрытому множеству Миллера. С помощью характеристических теорем рассматриваемые решения могут быть вычислены с помощью матриц  $M$  и  $T$ , введенных выше. Формулы, выражающие  $MPES$  и  $MRES$  через  $M$  и  $T$ , приведены в [100].

В настоящем исследовании было доказано, что характеристическая теорема, устанавливающая условие принадлежности альтернативы одному из минимальных  $P$ -доминирующих множеств, может быть обобщена на случай произвольного топологического пространства альтернатив. Достаточным условием этого обобщения является компактность генерального множества

альтернатив  $A$  в топологии, порожденной предбазой, состоящей нижних срезов всех альтернатив по отношению  $P$ . В силу этого условия каждое  $P$ -доминирующее множество содержит минимальное, а любое минимальное  $P$ -доминирующее множество оказывается конечным. Поскольку любое конечное  $P$ -доминирующее множество является  $P$ - и  $R$ -внешнеустойчивым, а в любом конечном  $P$ - или  $R$ -внешнеустойчивом множестве есть минимальное  $P$ - или  $R$ -внешнеустойчивое множество, то вышеуказанное топологическое условие также оказывается достаточным условием непустоты всех трех исследуемых решений.

Также в общем случае топологического пространства из характеристических теорем получены необходимые условия принадлежности альтернативы одному из минимальных  $P$ -внешнеустойчивых множеств и одному из минимальных  $R$ -внешнеустойчивых множеств.

Важнейшей частью исследования был аксиоматический анализ свойств решений. В результате было установлено, что объединение минимальных  $P$ -внешнеустойчивых множеств и объединение минимальных  $R$ -внешнеустойчивых множеств обладают такими свойствами как монотонность по отношению к изменению строгих коллективных предпочтений  $P$  [103], независимость от посторонних альтернатив по Нэшу [104], идемпотентность [103], свойство Айзермана [105], но не удовлетворяют аксиоме согласия (аксиоме расширения) [103]; также установлено, что объединение минимальных  $P$ -доминирующих множеств не обладает ни одним из вышеперечисленных свойств.

Помимо этого было установлено, что такие решения как объединение минимальных  $P$ -внешнеустойчивых множеств и объединение минимальных  $R$ -внешнеустойчивых множеств обладают свойством монотонности по Санверу (монотонности покрытия) [106]. Из этого следует, что, несмотря на то, что  $MPES$  и  $MRES$ , как и все турнирные решения, немонотонны по Маскину, эти функции оптимального коллективного выбора являются имплементируемыми в случае, если предпочтения участников процесса принятия коллек-

тивных решений можно расширить с отдельных альтернатив на множества альтернатив. Также было показано, что объединение минимальных  $P$ -доминирующих множеств таким свойством не обладает. По результатам данного исследования подготовлен текст препринта [101].

### 2.2.2 Сравнительный анализ аксиоматических свойств процедур построения агрегированных ранжирований научных журналов по менеджменту, политологии и экономике

Работы прошлого года, связанные с построением агрегированных рейтингов научных журналов по менеджменту, политологии и экономике, результаты которых опубликованы в [107] были продолжены. Был проведен сравнительный анализ аксиоматических свойств процедур агрегирования, примененных для решения данной практической, - правила большинства, правила Коупланда [38] и сортировки с помощью двух турнирных решений (непокрытого множества и объединения минимальных  $P$ -внешнеустойчивых множеств). Было отмечено, что в силу нейтральности и анонимности правила большинства, все основанные на нем правила ранжирования являются нейтральными и анонимными, то есть относятся ко всем журналам и библиометрическим индексам одинаково. Также было отмечено, что для построения рейтингов журналов, число которых меняется от года к году и от задачи к задаче, особенно важным является свойство независимости от посторонних альтернатив. Правило большинства обеспечивает независимость от посторонних альтернатив по Эрроу [108]. Но ценой этого свойства является наличие циклов в мажоритарном отношении  $P$  (упоминавшийся выше парадокс Кондорсе). В силу теоремы Эрроу [108] никакое Парето-эффективное правило агрегирования ранжирований в единый рейтинг не может удовлетворять аксиоме независимости от посторонних альтернатив в форме Эрроу. Все использованные в работе методы ранжирования Парето-эффективны, поэтому они не удовлетворяют этому свойству. Однако правило Коупланда удовлетворяет слабому условию независимости от посторонних альтернатив по Эр-

роу, а объединение минимальных  $P$ -внешнеустойчивых множеств, как было доказано на предыдущем этапе исследования, удовлетворяет условию независимости от посторонних альтернатив по Нэшу [104]. Поскольку первое условие предполагает сохранения числа альтернатив и изменение коллективных предпочтений, а второе – изменение и предпочтений и множества альтернатив, из которых производится выбор, то второй тип независимости в нашей задаче выглядит более предпочтительным. Таким образом, аксиоматический анализ правил агрегирования показывает, что наилучшим по релевантным в данном случае свойствам способом агрегирования базовых рейтингов журналов является сортировка с помощью объединения минимальных  $P$ -внешнеустойчивых множеств. Результаты данного исследования отражены в статье [109].

### 2.3 Исследование нормативных свойств двухступенчатых процедур выбора

Многие задачи прикладной математики и теории управления могут быть сведены к задаче выбора, которая имеет следующую общую формулировку. Пусть задано конечное множество  $A$  альтернатив, мощность которого больше или равна 2 ( $\text{card}(A) \geq 2$ ). Любое подмножество  $X \subseteq 2^A$  может быть предъявлено для осуществления акта выбора и называется далее предъявлением. Пусть задана некоторая функция выбора  $C(\cdot)$ , которая является отображением  $C: 2^A \rightarrow 2^A$  с ограничением  $C(X) \subseteq X$  для любого  $X \subseteq 2^A$ . Выбор состоит в выделении из предъявления  $X$  по некоторому фиксированному правилу непустого подмножества альтернатив  $Y \subseteq X$  (непустого подмножества «лучших» альтернатив).

Существует большое число различных процедур выбора, которые позволяют производить выбор наилучших альтернатив, а также их ранжирование в условиях многокритериальности. Все процедуры выбора можно разделить на 5 основных групп:

- позиционные правила;
- правила, использующие мажоритарное отношение;
- правила, использующие вспомогательную числовую шкалу;
- правила, использующие турнирную матрицу;
- q-Паретовские правила большинства.

Изменение предъявляемого множества альтернатив, их оценок по критериям или наборам критериев, по которым осуществляется выбор, приводит к изменению полученного выбора. Отсюда следует необходимость более детального изучения существующих процедур выбора и определения того, какие из них могут применяться в конкретном случае.

В этом подразделе приведены результаты исследования более сложных процедур выбора, в основе которых лежит принцип суперпозиции. Напомним, что под суперпозицией двух функций выбора  $C_1(\cdot)$  и  $C_2(\cdot)$  подразумевается бинарная операция  $\odot$ , результат которой – новая функция  $C^*(\cdot) = C_2(\cdot) \odot C_1(\cdot)$ , имеющая вид  $\forall X \in 2^A \quad C^*(X) = C_2(C_1(X))$  [102]. Другими словами, суперпозиция заключается в последовательном применении функций выбора, что означает, что выбор на  $X$  по функции выбора  $C_1$  является предъявлением для функции  $C_2$ . При этом изменение порядка применения функций выбора может привести к различным результатам выбора, т.к. операция суперпозиции не обладает свойством коммутативности.

Интерес к подобным процедурам выбора возникает по нескольким причинам. Во-первых, процедуры выбора, основанные на идее суперпозиции, имеют достаточно низкую вычислительную сложность, что критично в том случае, когда число альтернатив и критериев очень велико. Дело в том, что при работе с большим объемом данных, когда каждая альтернатива характеризуется большим количеством различных признаков, большинство существующих одноступенчатых процедур выбора имеют высокую вычислительную сложность. Использование операции суперпозиции позволяет сокращать на каждом новом этапе общее число альтернатив, а также применять на раз-

ных этапах различные с точки зрения вычислительной сложности процедуры выбора, что позволяет получить результат за приемлемое время. Во-вторых, зачастую возникают ситуации, когда в результате применения некоторой процедуры полученный выбор содержит слишком большое число альтернатив. Использование операции суперпозиции позволяет избегать таких ситуаций за счет использования дополнительных процедур выбора.

В рамках данной работы рассматривались двухступенчатые процедуры выбора, у которых на первом этапе используются правила, использующие мажоритарное отношение, вспомогательную числовую шкалу и турнирную матрицу, а на втором этапе – позиционные правила, а также правила, использующие мажоритарное отношение, вспомогательную числовую шкалу и турнирную матрицу.

Целью работы было исследование данных процедур выбора на предмет удовлетворения существующим нормативным условиям, которые показывают, каким образом изменится выбор при добавлении в исходный набор данных новых альтернатив, при удалении из исходного набора не попавших в выбор альтернатив, при удалении из исходного набора попавших в выбор альтернатив, а также при изменении значения по критерию одной из альтернатив и т.д. Кроме того, исследовалась вычислительная сложность двухступенчатых процедур выбора с целью определения того, какие из них могут быть использованы при наличии большого числа альтернатив и критериев.

### 2.3.1 Двухступенчатые процедуры выбора

Будем рассматривать двухступенчатые процедуры выбора, у которых на первом этапе используются правила, использующие мажоритарное отношение, вспомогательную числовую шкалу и турнирную матрицу [42, 102, 110 – 112]. Более подробное описание используемых процедур выбора приведено в [113].

В Таблице 2.1 приведен список исследуемых двухступенчатых процедур выбора.

Таблица 2.1 – Двухступенчатые процедуры выбора

Этап 1	Этап 2
Правила, использующие мажоритарное отношение (10 процедур)	Позиционные правила (11 процедур) Правила, использующие мажоритарное отношение (10 процедур) Правила, использующие вспомогательную числовую шкалу (5 процедур) Правила, использующие турнирную матрицу (2 процедуры)
Правила, использующие вспомогательную числовую шкалу (5 процедур)	Позиционные правила (11 процедур) Правила, использующие мажоритарное отношение (10 процедур) Правила, использующие вспомогательную числовую шкалу (5 процедур) Правила, использующие турнирную матрицу (2 процедуры)
Правила, использующие турнирную матрицу (2 процедуры)	Позиционные правила (11 процедур) Правила, использующие мажоритарное отношение (10 процедур) Правила, использующие вспомогательную числовую шкалу (5 процедур) Правила, использующие турнирную матрицу (2 процедуры)

Таким образом, далее рассматриваются 476 двухступенчатых процедур выбора 12 различных типов. Перед тем как перейти к исследованию свойств данных процедур выбора, необходимо сделать несколько замечаний.

В Таблице 2.2 приведен список двухступенчатых процедур выбора, не имеющих смысла.

Таблица 2.2 – Двухступенчатые процедуры выбора, не имеющие смысла («...» – любая рассматриваемая процедура выбора)

Этап 1	Этап 2
Минимальное доминирующее множество	Минимальное доминирующее множество
Минимальное непокрытое множество	Минимальное доминирующее множество
Минимальное непокрытое множество	Минимальное непокрытое множество
Победитель Кондорсе	...
Ядро	Правило простого большинства
Ядро	Правило Борда
Ядро	Правило Блэка
Ядро	Обратное правило Борда
Ядро	Правило Нансона
Ядро	Минимальное доминирующее множество
Ядро	Минимальное непокрытое множество

## Продолжение таблицы 2.2

Ядро	Минимальное слабоустойчивое множество
Ядро	Правило Фишберна
Ядро	Непокрытое множество I
Ядро	Непокрытое множество II
Ядро	Правило Ричалсона
Ядро	Ядро
Ядро	$k$ -устойчивые множества ( $k > 1$ )
Ядро	Первое правило Копланда
Ядро	Второе правило Копланда
Ядро	Третье правило Копланда
Ядро	Минимаксная процедура
Ядро	Процедура Симпсона
Общее число процедур выбора - 50	

Двухступенчатые процедуры выбора, приведенные в Таблице 2.2, исключаются из дальнейшего рассмотрения.

Заметим, что свойства двухступенчатых правил выбора, использующих правило Блэка, полностью совпадают со свойствами двухступенчатых правил выбора, использующих правило Борда, в случае, когда победитель Кондорсе отсутствует.

В Таблице 2.3 приведен список двухступенчатых процедур выбора, эквивалентных существующим процедурам выбора.

Таблица 2.3 – Двухступенчатые процедуры выбора, эквивалентные существующим

Двухступенчатая процедура выбора		Какой процедуре эквивалентна?
Этап 1	Этап 2	
Минимальное доминирующее множество	Правило простого большинства	Победитель Кондорсе
Минимальное доминирующее множество	Минимальное непокрытое множество	Минимальное непокрытое множество
Минимальное доминирующее множество	Минимальное слабоустойчивое множество	Минимальное слабоустойчивое множество
Минимальное доминирующее множество	Правило Фишберна	Правило Фишберна
Минимальное доминирующее множество	Непокрытое множество I	Непокрытое множество I
Минимальное доминирующее множество	Непокрытое множество II	Непокрытое множество II
Минимальное доминирующее множество	Правило Ричалсона	Правило Ричалсона
Минимальное доминирующее множество	Победитель Кондорсе	Победитель Кондорсе
Минимальное доминирующее множество	Ядро	Ядро

### Продолжение таблицы 2.3

Минимальное доминирующее множество	Первое правило Копланда	Первое правило Копланда
Минимальное доминирующее множество	Второе правило Копланда	Второе правило Копланда
Минимальное доминирующее множество	Третье правило Копланда	Третье правило Копланда
Минимальное непокрытое множество	Правило простого большинства	Ядро, содержащее одну альтернативу
Минимальное непокрытое множество	Минимальное слабоустойчивое множество	Минимальное слабоустойчивое множество
Минимальное непокрытое множество	Победитель Кондорсе	Победитель Кондорсе
Минимальное непокрытое множество	Ядро	Ядро
Правило Фишберна	Правило простого большинства	Ядро, содержащее одну альтернативу
Правило Фишберна	Победитель Кондорсе	Ядро, содержащее одну альтернативу
Непокрытое множество I	Правило простого большинства	Ядро, содержащее одну альтернативу
Непокрытое множество I	Победитель Кондорсе	Ядро, содержащее одну альтернативу
Непокрытое множество II	Правило простого большинства	Ядро, содержащее одну альтернативу
Непокрытое множество II	Победитель Кондорсе	Ядро, содержащее одну альтернативу
Правило Ричалсона	Правило простого большинства	Ядро, содержащее одну альтернативу
Правило Ричалсона	Победитель Кондорсе	Ядро, содержащее одну альтернативу
Ядро	Победитель Кондорсе	Победитель Кондорсе
Общее число процедур выбора - 25		

Двухступенчатые процедуры выбора, приведенные выше, эквивалентны одной из существующих процедур выбора. Следовательно, свойства данных процедур выбора полностью совпадают со свойствами существующих процедур. Стоит отметить, что данные процедуры выбора не исключаются из дальнейшего рассмотрения, так как вычислительная сложность некоторых двухступенчатых процедур выбора в определенных ситуациях может быть ниже вычислительной сложности существующих процедур выбора.

Таким образом, остается исследовать свойства только 401 двухступенчатую процедуру выбора.

### 2.3.2 Исследование свойств двухступенчатых процедур выбора

В [113] приведен список исследуемых нормативных условий и их подробное описание.

Исследование свойств рассматриваемых процедур выбора осуществлялось следующим образом. В случае, когда рассматриваемое свойство не выполняется, приводится пример, который опровергает выполнение данного свойства. Напротив, в случае, когда процедура выбора удовлетворяет рассматриваемому нормативному условию, приводится необходимое доказательство. Необходимо отметить, что исследование двухступенчатых процедур выбора основано на исследовании свойств известных процедур выбора, которые было выполнено в [113].

Полученные результаты приведены в Таблице Б.1 Приложения Б. Доказательство полученных результатов приведено в [114].

### 2.3.3 Вычислительная сложность процедур выбора

В [114] приведены результаты о вычислительной сложности 28 существующих процедур выбора. В соответствии с [114] разделим все множество двухступенчатых процедур на несколько подгрупп в зависимости от их вычислительной сложности (см. Таблицу 2.4).

Таблица 2.4 – Теоретическая вычислительная сложность двухступенчатых процедур выбора («...» – любая рассматриваемая процедура выбора)

Двухступенчатые процедуры выбора	
Этап 1	Этап 2
Процедуры с низкой вычислительной сложностью	
Надпороговое правило выбора	Правило простого большинства Двухступенчатое правило относительного большинства Правило Хара (Процедура Уэйра) Правило Борда, Правило Блэка Победитель Кондорсе Правило относительного большинства Пороговое правило Обратное правило простого большинства Аналог правила одобряющего голосования ( $q>1$ ) Надпороговое правило выбора
Процедуры выбора с вычислительной сложностью, сильно зависящей от входных данных	

## Продолжение таблицы 2.4

Надпороговое правило выбора	Обратное правило Борда, Правило Нансона Ядро, Правила Копланда 1-3 Минимаксная процедура Процедура Симпсона, Процедура Кумбса
Процедуры со средней вычислительной сложностью	
Ядро Правила Копланда 1-3 Минимаксная процедура Процедура Симпсона	Правило простого большинства Двухступенчатое правило относительного большинства Правило Хара (Процедура Уэйра) Правило Борда, Правило Блэка Победитель Кондорсе Правило относительного большинства Пороговое правило Обратное правило простого большинства Аналог правила одобряющего голосования ( $q>1$ ) Надпороговое правило выбора Обратное правило Борда, Правило Нансона Ядро, Правила Копланда 1-3 Минимаксная процедура Процедура Симпсона, Процедура Кумбса
Процедуры с высокой вычислительной сложностью	
Надпороговое правило выбора Ядро Правила Копланда 1-3 Минимаксная процедура Процедура Симпсона	Минимальное доминирующее множество Минимальное непокрытое множество Минимальное слабоустойчивое множество Правило Фишберна, Непокрытое множество I Непокрытое множество II, Правило Ричалсона, k-устойчивые множества ( $k>1$ )
Минимальное доминирующее множество Минимальное непокрытое множество Минимальное слабоустойчивое множество Правило Фишберна Непокрытое множество I Непокрытое множество II Правило Ричалсона k-устойчивые множества ( $k>1$ )	...

Результаты, приведенные в Таблице 2.4, показывают, какие из предложенных двухступенчатых процедур выбора могут применяться на больших объемах данных, а какие процедуры выбора применять не рекомендуется, так как они не позволяют получить результаты за приемлемое время.

#### 2.3.4 Некоторые выводы

Приведены результаты исследования 401-й двухступенчатой процедуры многокритериального выбора на предмет удовлетворения существующим нормативным условиям. Нормативные условия показывают, каким образом

изменяется конечный выбор при изменении предъявляемого множества альтернатив, оценок альтернатив по критериям, набора критериев, по которым осуществляется выбор. Определение того, какие процедуры выбора наименее устойчивы, а какие наиболее уязвимы к изменениям этих параметров, приводит к более глубокому пониманию функционирования существующих процедур выбора, фундаментальных свойств выбора, а также того, насколько постоянен (устойчив) и осмыслен набор альтернатив, полученный в результате применения заданной процедуры выбора.

В результате исследования было выявлено, что большинство двухступенчатых процедур выбора не удовлетворяют ни одному из нормативных условий. Лишь некоторые из них удовлетворяют условию монотонности 1. Более подробная информация о том, каким нормативным условиям удовлетворяют рассматриваемые процедуры выбора, приведена в Таблице Б.1 Приложения Б.

В работе также оценена вычислительная сложность двухступенчатых процедур, в соответствии с которой все процедуры были поделены на четыре группы (см. Таблицу 2.4). Показано, что те процедуры, которые на первом этапе используют сложные с точки зрения вычислительной сложности процедуры выбора, требуют большого времени. Это означает, что такие двухступенчатые процедуры выбора не рекомендуется использовать на больших объемах данных. Процедуры, которые на первом этапе используют простые с точки зрения вычислительной сложности, а на втором – вычислительно сложные процедуры выбора, могут применяться на большом объеме данных, но их применение сильно зависит от количества альтернатив, оставшихся после первого этапа. Процедуры, которые используют на обеих стадиях простые с точки зрения вычислительной сложности процедуры выбора, могут применяться без ограничений на больших объемах данных.

## 2.4 Исследование степени манипулируемости правил агрегирования предпочтений в модели Impartial Anonymous Culture

В данном подразделе отчета приведены результаты продолжающегося исследования по расчетам степени манипулируемости правил агрегирования предпочтений.

Манипулирование – это ситуация, когда при голосовании участник предъявляет неискренние предпочтения, чтобы добиться лучшего для себя результата голосования. В работах [115] и [116] было показано, что любое недиктаторское правило принятия решений манипулируемо. В связи с этим возникает вопрос: если любое правило принятия решений манипулируемо, то можем ли мы определить наименее манипулируемое правило коллективного выбора?

Для измерения степени манипулируемости процедуры голосования в [117] и [118] был введен показатель, который впоследствии получил название индекса Нитцана-Келли. Индекс показывает долю манипулируемых профилей в общем количестве всех возможных профилей голосования.

В [119, 120] исследовалась манипулируемость правил коллективного выбора для случая Impartial Culture (IC), в котором подразумевалось, что все профили голосования одинокого вероятны. В данном подразделе отчета приведены результаты исследования манипулируемости правил коллективного выбора в случае Impartial Anonymous Culture (далее IAC), в котором одинаково вероятны не профили, а ситуации голосования, т.е. профили, отличающиеся перестановкой участников, составляют одну ситуацию голосования и при расчете индексов манипулируемости учитываются один раз.

Рассматривается ситуация голосования, в которой участвуют  $n$  участников и  $m$  кандидатов. У каждого участника есть предпочтение на множестве кандидатов, выраженное линейным порядком. Так как рассматривается ситуация множественного выбора, то необходимо ввести расширенные предпочтения для устранения несравнимости различных множественных выборов для участника голосования.

Рассматривались семь позиционных правил коллективного выбора. Цель исследования – найти наименее манипулируемые правила коллективного выбора.

#### 2.4.1 Описание исследуемых процедур и вычислительного эксперимента

Исследовалась манипулируемость следующих семи процедур коллективного выбора:

- 1) правило относительного большинства: выбирается альтернатива с наибольшим количеством голосов, т.е. первых мест в предпочтениях участников голосования;
- 2) одобряющее голосование  $q=2$ : выбирается альтернатива, занимающее наибольшее количество первых и вторых мест в предпочтениях участников голосования;
- 3) правило Борда: в предпочтении каждого участника вычисляется количество альтернатив (ранг Борда), худших, чем заданная альтернатива; сумма этих чисел для каждого из участников составляет ранг Борда для выбранной альтернативы; выбирается альтернатива с наибольшим рангом Борда;
- 4) процедура Блэка: выбирается победитель Кондорсе; если его не существует, то применяется правило Борда;
- 5) пороговое правило: для каждой альтернативы составляется вектор (количество последних мест, ... количество вторых мест, количество первых мест); выбирается альтернатива с лексикографически наименьшим вектором.
- 6) процедура Хара: если у какой-либо альтернативы большинство голосов ( $50\%+1$  голос), то она является выбором; если такой альтернативы нет, то из голосования выбывает альтернатива с наименьшим количеством голосов и процедура повторяется;

7) процедура Нансона: для каждой альтернативы вычисляется ранг Борда; далее из голосования выбывают альтернативы с рангом Борда ниже среднего по всем альтернативам; процедура повторяется, пока выбор непустой.

В процедурах голосования встречаются ситуации, когда выбираются сразу несколько альтернатив. Например, в правиле относительного большинства две альтернативы могут иметь одинаковое количество голосов. В данном исследовании была использована концепция множественного выбора, т.е. в ситуациях несравнимости или одинакового количества голосов все такие альтернативы входят в коллективный выбор.

В этом случае нам необходимо понять, как будет сравнивать для себя различные коллективные выборы участник голосования, который хочет произвести манипулирование. Для этого используется модель расширенных предпочтений, которая позволяет сравнивать не только альтернативы попарно, но и все возможные множественные выборы.

Для трех альтернатив используются 4 метода расширений предпочтений [119], среди них:

- Лексимин:  $\{a\} \succ \{a,b\} \succ \underline{\{b\}} \succ \{a,c\} \succ \{a,b,c\} \succ \{b,c\} \succ \{c\}$ ;
- Лексимакс:  $\{a\} \succ \{a,b\} \succ \underline{\{a,b,c\}} \succ \{a,c\} \succ \{b\} \succ \{b,c\} \succ \{c\}$ ,

которые и рассматривались в данном исследовании.

Для оценки степени манипулируемости процедуры голосования используется индекс Нитцана-Келли

$$NK = \frac{d_0}{C_{m! + n - 1}^n},$$

где  $d_0$  – количество манипулируемых профилей,  $C_{m! + n - 1}^n$  – общее количество всех возможных профилей.

Задача исследования – найти наименее манипулируемые процедуры голосования для ситуации с тремя кандидатами и количеством участников голосования от 3 до 100.

Для расчета индексов Нитцана-Келли для каждого правила коллективного выбора использовалось компьютерное моделирование.

Схема компьютерного моделирования:

- генерация профилей;
- для каждого профиля производилась проверка на манипулируемость;
- если профиль манипулируем, то увеличиваем счетчик ( $d_0$ ) манипулируемых профилей;
- вычисление индекса манипулируемости как доли манипулируемых профилей.

Далее полученные результаты для каждого правила коллективного выбора сравнивались друг с другом.

#### 2.4.2 Результаты компьютерного моделирования

Мы будем сравнивать значения индекса Нитцана-Келли для семи процедур голосования для ситуации с тремя альтернативами и различными количествами участников голосования.

Ниже, на рисунке 2.4 приведен график для расширенных предпочтений Лексимин (три альтернативы):

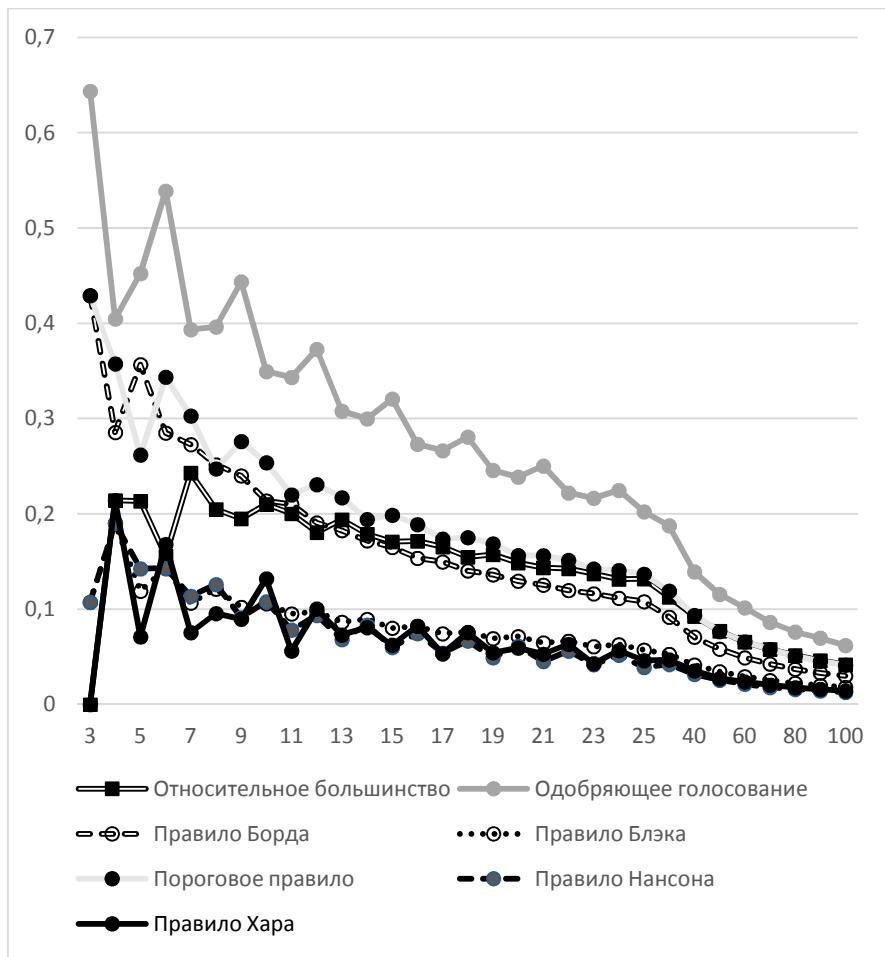


Рисунок 2.4 – Изменения индекса Нитцана-Келли для расширенных предпочтений Лексимин и различных процедур коллективного выбора в зависимости от числа участников голосования

Мы можем сделать следующие наблюдения:

- значения индекса Нитцана-Келли для IAC убывают с ростом числа участников голосования для всех правил;
- для большинства правил можно выделить периоды длины 2 или 3;
- процедура Нансона показывает наименьшее значение манипулируемости из всех семи рассмотренных правил;
- одобряющее голосование  $q=2$  наиболее манипулируемое правило из 7 рассмотренных.

На рисунке 2.5 приведен график для расширения предпочтений Лексимакс.

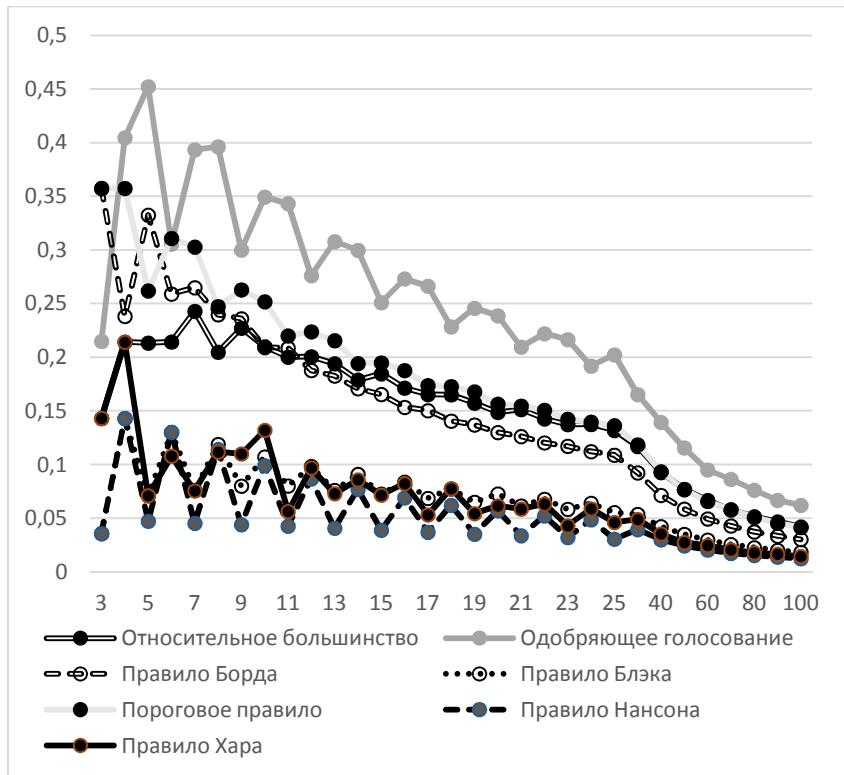


Рисунок 2.5 – Изменение индекса Нитцана-Келли для расширенных предпочтений Лексимакс и различных процедур коллективного выбора в зависимости от числа участников голосования

Мы можем заметить следующее:

- периодичность для расширенных предпочтений Лексимакс тоже имеет место и похожа на ситуацию с Лексимин;
- одобряющее голосование  $q=2$  вновь самое манипулируемое правило кол-lectивного выбора;
- процедура Нансона – наименее манипулируемое правило коллективного выбора для большинства случаев.

Теперь мы сравним результаты данного исследования в модели Impartial Anonymous Culture (IAC) с результатами работы по исследованию манипулирования в Impartial Culture (IC) [119, 120]. Рассмотрим три правила: правило относительного большинства как самое распространенное правило коллективного выбора, правила Нансона и Хара как правила показавшие на предыдущих графиках наименьшие значения индексов манипулируемости.

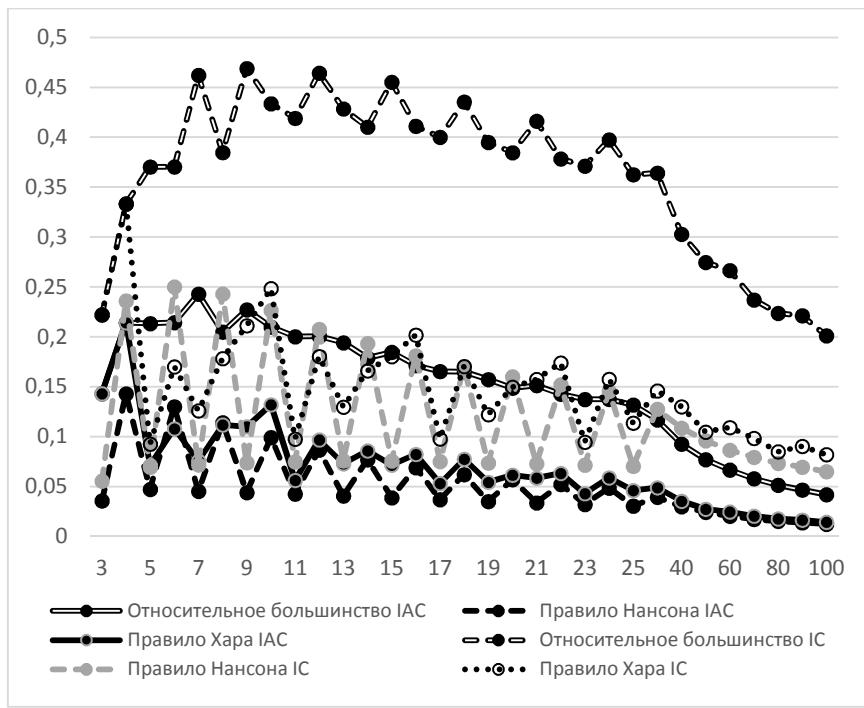


Рисунок 2.6 – Сравнение изменений индекса Нитцана-Келли для моделей IAC и IC

Из рисунка 2.6 можно сделать следующие выводы:

- значения индекса Нитцана-Келли для IAC в большинстве случаев ниже, чем для IC; начиная с 30 участников голосования это видно для всех трех правил на графике;
- значения индексов Нитцана-Келли уменьшаются с ростом количества участников голосования и для IAC, и для IC; все рассмотренные правила достигают пика в начале графика;
- лучшая процедура голосования с точки зрения манипулируемости – процедура Нансона – имеет значение индекса Нитцана-Келли для 100 участников равным 0.012, что соответствует 1% манипулируемых профилей;
- все рассмотренные правила имеют период в 2 или 3 участника голосования;
- наименее манипулируемые правила для одинаковых случаев в IC и IAC чаще всего совпадают (около 80% случаев), и лишь в небольшом числе (примерно 20%) случаев различаются.

### 2.4.3 Некоторые выводы

По результатам исследования можно сделать следующие выводы.

1. Все значения индекса Нитцана-Келли в Impartial Anonymous Culture уменьшаются с возрастанием количества участников голосования. Аналогичные результаты были получены в ходе предыдущего исследования манипулируемости в модели Impartial Culture. Это можно объяснить тем, что мы рассматриваем одиночное манипулирование, при котором с увеличением количества участников голосования удельный вес голоса манипулирующего агента снижается, поэтому индекс манипулируемости падает.

2. Для большинства рассмотренных правил коллективного выбора значения индекса Нитцана-Келли имеют период 2 или 3 по количеству участников голосования.

3. Процедура Нансона оказалась наименее манипулируемой из семи проанализированных процедур голосования в Impartial Anonymous Culture в рассматриваемом случае с тремя кандидатами. Значения индекса Нитцана-Келли для процедуры Нансона не превосходят 0.15, что соответствует 15% манипулируемых профилей в общем количестве всех профилей.

4. Одобряющее голосование при  $q=2$  показывает самые высокие значения индекса манипулируемости Нитцана-Келли. Максимальные значения достигают 0.45, что соответствует 45% манипулируемых профилей в общем количестве всех профилей.

Результаты данного исследования нашли отражения в работах [121] и [122].

### 3 Исследования распределения влияния и согласованности решений в политических и финансовых системах

В этом разделе отчета приведены результаты исследований распределения влияния и согласованности решений в политических и финансовых системах. Эти результаты получены в следующих теоретических и прикладных направлениях:

- анализ согласованности позиций групп в Европейском парламенте в 1979–2014 гг.;
- оценка распределения влияния в Европейском парламенте в 1979–2014 гг.;
- анализ структуры Верховной Рады Украины в 2007–2012 гг.;
- исследование конфликтности и точности рекомендаций российских финансовых аналитиков.

#### 3.1 Анализ согласованности позиций групп в Европейском парламенте в 1979–2014 гг.

В органах, принимающих коллективные решения, например, в парламентах, важно учитывать внутреннюю структуру, формальные и неформальные группы, действующие в общих интересах. Отношения между группами депутатов, а, порой, и отдельные индивидуальные связи и договоренности могут существенно влиять на ход голосования и на его результат.

Так, в парламенте, фракции, придерживающиеся похожих политических позиций, обладающие общими интересами, будут обладать высокой степенью согласованности при голосовании по множеству вопросов.

Практическое следствие таких «хороших» отношений в том, что коалиции, сформированные из этих групп, будут голосовать согласованно и будут стабильными. Если же группы придерживаются разных политических позиций, то с большей вероятностью и голосовать они будут по-разному, находясь в оппозиции друг к другу. Следовательно, когда в парламент проходят

партии, последовательно придерживающиеся некоторой идеологии, структура парламента будет более четкой, а решения — поддающимися прогнозу.

Голосования депутатов Европарламента и работа его комитетов — два основных направления исследования Европарламента. Ранние исследования часто содержали в себе описания организации работы в комитетах, распределение должностей вроде докладчиков (*rapporteur*) по вопросу голосования [123]. Анализ голосований и поведения депутатов часто проводился по небольшим выборкам в рамках одного-двух созывов, но исследования последних 10-20 лет активно используют сложные эконометрические и статистические методы, например, NOMINATE, для определения позиций парламентариев, исследования процедур принятия решений, анализа согласованности и конкуренции между депутатами и/или группами депутатов [124]. Такое развитие стало возможным благодаря большому объему данных, накопленных за последние семь созывов Европарламента и тому, что данные о поименных голосованиях, составу комитетов, именах и партийной принадлежности докладчиков находятся в открытом доступе на официальном сайте, а также агрегируются на отдельных ресурсах [125].

В данном подразделе отчета приведены результаты анализа согласованности по двум основных расколам, приводящих депутатов Европейского парламента к объединению в группы — идеологическому и национальному. В пункте 3.1.1 дается описание используемых данных. В пункте 3.1.2 проводится анализ согласованности в Европарламенте: рассчитывается индекс Райса, оценивающий степень согласованности депутатов в группе, и анализируется попарная согласованность между группами. В Приложении В приведены дополнительные рисунки.

### 3.1.1 Описание данных

Депутаты Европарламента могут голосовать «За» или «Против» некоторого решения и имеют право воздержаться от принятия решения. Они также могут присутствовать на сессии в целом, но по различным причинам не про-

голосовать по всем или некоторым вопросам — ‘Present-did-not-vote’ [126]. В первом созыве Европарламента в среднем не голосовала более чем половина парламентариев (до 22% отсутствовали и до 30% не голосовали). Но с четвертого по шестой созывы ситуация заметно исправилась: в голосовании не участвовали в среднем менее 40% — около 25% отсутствовали и около 14% не голосовали. В седьмом созыве отсутствия снизились до менее чем 13%.

Менее дисциплинированными в первых созывах Европарламента могут считаться национал-консерваторы, либералы, техническая (независимая) группа, националисты. В 5-6 созывах чаще отсутствовали евроскептики и либералы. Часто не принимали участие в голосовании депутаты от Бельгии, Франции, Ирландии, Люксембурга, и позже к ним добавились депутаты от Греции, Португалии, Дании. Тем не менее, нельзя сказать о том, что депутаты от некоторых стран-членов систематически не голосуют в Европарламенте. Так, например, во время шестого созыва чаще других не пользовались правом голоса парламентарии от Италии, Греции, Испании, Мальты, а во время пятого — парламентарии от Бельгии, Греции, Нидерландов, Финляндии. Стоит отметить и большую степень ротации депутатов, связанную, отчасти, с расширением Евросоюза. Так, например, в Европарламент четвертого созыва были избраны менее половины депутатов третьего созыва: 321 из 567 мандатов получили новые политики [126].

Таким образом, для исключения «шума» — неинформативных голосований и депутатов, кто регулярно пропускает сессии или не голосует во время процедуры — необходимо выбрать для анализа согласованности «полезные» голоса. Голосования с высокой согласованностью часто являются «затруднительно проходными» и способны маскировать расколы, по которым разделяются депутаты, их информативность для исследования невелика. С другой стороны, дисциплинированные депутаты более ответственно реализуют свой мандат и могут считаться более последовательным в своем поведении, поэтому в их поведении ожидается меньше «шума» и «случайности». Анализ устойчивости результатов показал, что в некотором смысле оптимальными

могут быть следующие пороги: не более 80% согласованности по вопросу голосования и не более 25% пропущенных голосований. Таким образом, в данной работе для оценки согласованности по семи созывам использовались почти 14 тыс. голосований по 74% депутатов (в среднем).

### 3.1.2 Анализ согласованности в Европарламенте

#### 3.1.2.1 Индекс согласованности Райса между группами

Индекс согласованности Райса, "*index of voting likeness*", рассчитывается по формуле

$$Rice_j = \frac{|q_j - k_j|}{q_j + k_j},$$

где  $q_j$  — доля проголосовавших «за» по  $j$ -ому голосованию в парламенте,  $k_j$  — доля проголосовавших «против» [127].

В целом по Европарламенту, на имеющихся отфильтрованных данных по всем голосованиям, согласованность между депутатами, без учета какой-либо принадлежности, за 25 лет увеличилась с 0,37-0,38 до 0,46, но была достаточно стабильной и не превышала 0,5 ни в одном созыве парламента (см. Таблицу 3.1).

Таблица 3.1 – Средний индекс согласованности Райса по всем голосованиям созыва по парламенту в целом, по политическим группам и странам-членам ЕС.

Созыв	Общий	По политическим группам	По странам-членам ЕС
1	0,382	0,862	0,626
2	0,370	0,841	0,569
3	0,403	0,873	0,574
4	0,416	0,841	0,548
5	0,381	0,420	0,441
6	0,434	0,763	0,617
7	0,458	0,762	0,579

Согласованность среди депутатов одной политической группы в среднем намного выше, чем среди депутатов от одного государства (*national groups of MEPs*), что также демонстрирует рисунок 3.1, на котором значения

индекса Райса по «политическим» группам и «национальным» группам нормированы на согласованность по парламенту в целом. Объяснение заключается в значимой выгоде национальных партий от присоединения к политической группе — так они тоже могут заручиться поддержкой идеологически близких групп, получить доступ к формированию повестки дня, каналам информации, делегируя при этом организационные издержки и проблемы управления на уровень крупных политических групп [128]. Но с конца 1990-х гг. это различие уменьшалось, что, по-видимому, обусловлено тем, что Европейский парламент стал играть куда большую роль в принятии решений, а национальные партии имеют в руках довольно мощный инструмент в виде контроля над «отбором» кандидатов в евродепутаты на национальном уровне. Однако заметим, что в 5 созыве внутригрупповая согласованность заметно снизилась, а согласованность по национальному признаку, в свою очередь, несколько подросла в шестом и седьмом созывах.

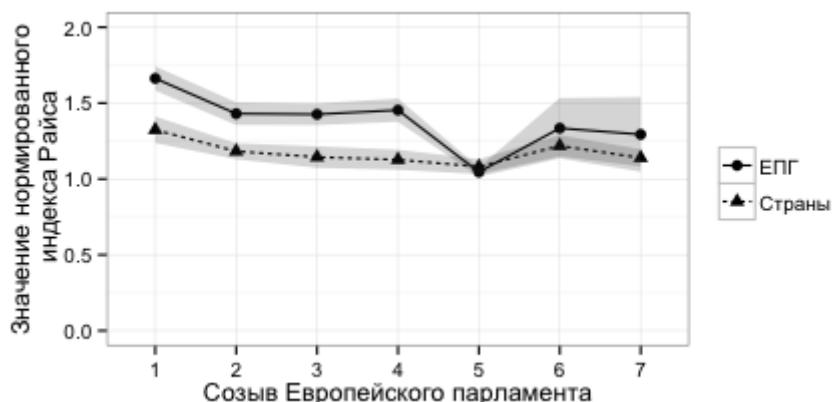


Рисунок 3.1 – Средний нормированный индекс согласованности Райса по европейским политическим группам (ЕПГ) и странам-членам ЕС, построенный по всем информативным голосованиям созыва

Стоит отметить, что более крупные политические группы в среднем более согласованы между собой, чем мелкие, и после вступления в силу Маастрихтского договора (и расширения полномочий Парламента) данная зависимость стала более выраженной [124]. На рисунке 3.2 показана средняя согласованность *внутри* политических групп по индексу Райса по созывам, а также размах — разница между максимумом и минимумом — значений ин-

декса согласованности, рассчитанный по имеющимся поименным голосованиям (пунктирной линией). «Внутригрупповой» индекс Райса также демонстрирует, что согласованность между депутатами одной политической группы выше, чем согласованность между депутатами от одной страны. Видно, что группы вроде евроскептиков, независимых, национал-консерваторов имеют более низкую среднюю согласованность и большой размах. Относительно высокий уровень согласованности можно отметить у христианских демократов, социал-демократов и радикальных левых. Последние, кстати, не являются многочисленной группой, в отличие от двух доминирующих, но индекс Райса свидетельствует о том, что они высоко дисциплинированы. Высоко согласованными также можно назвать консерваторов, националистов и зеленых/регионалистов, несмотря на то, что они и не всегда представлены в Европарламенте.

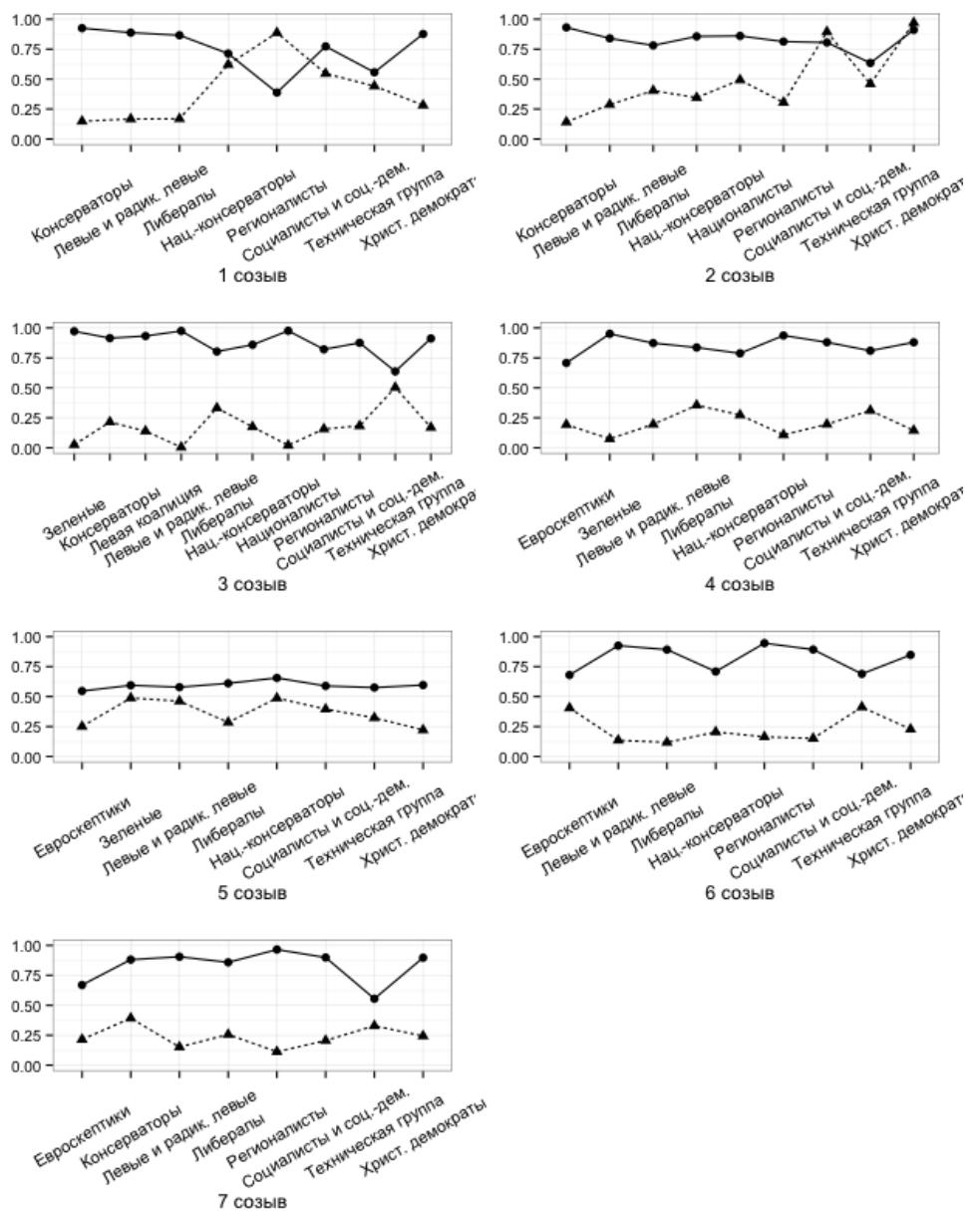


Рисунок 3.2 – Средний индекс согласованности Райса по всем голосованиям созыва по политическим группам (сплошная линия) и размах — *max – min* — (пунктирная линия).

Стоит заметить относительно более высокую согласованность между депутатами Великобритании и Ирландии, характерную для первых созывов Европарламента (рисунок 3.3). После расширения ЕС до 25 стран, заметна более высокая, чем в среднем, согласованность между депутатами из стран Центральной и Восточной Европы. Во многом это связано с тем, что немногочисленные «национальные» делегации присоединяются лишь к двум-четырем политическим группам из существующих семи-десяти.

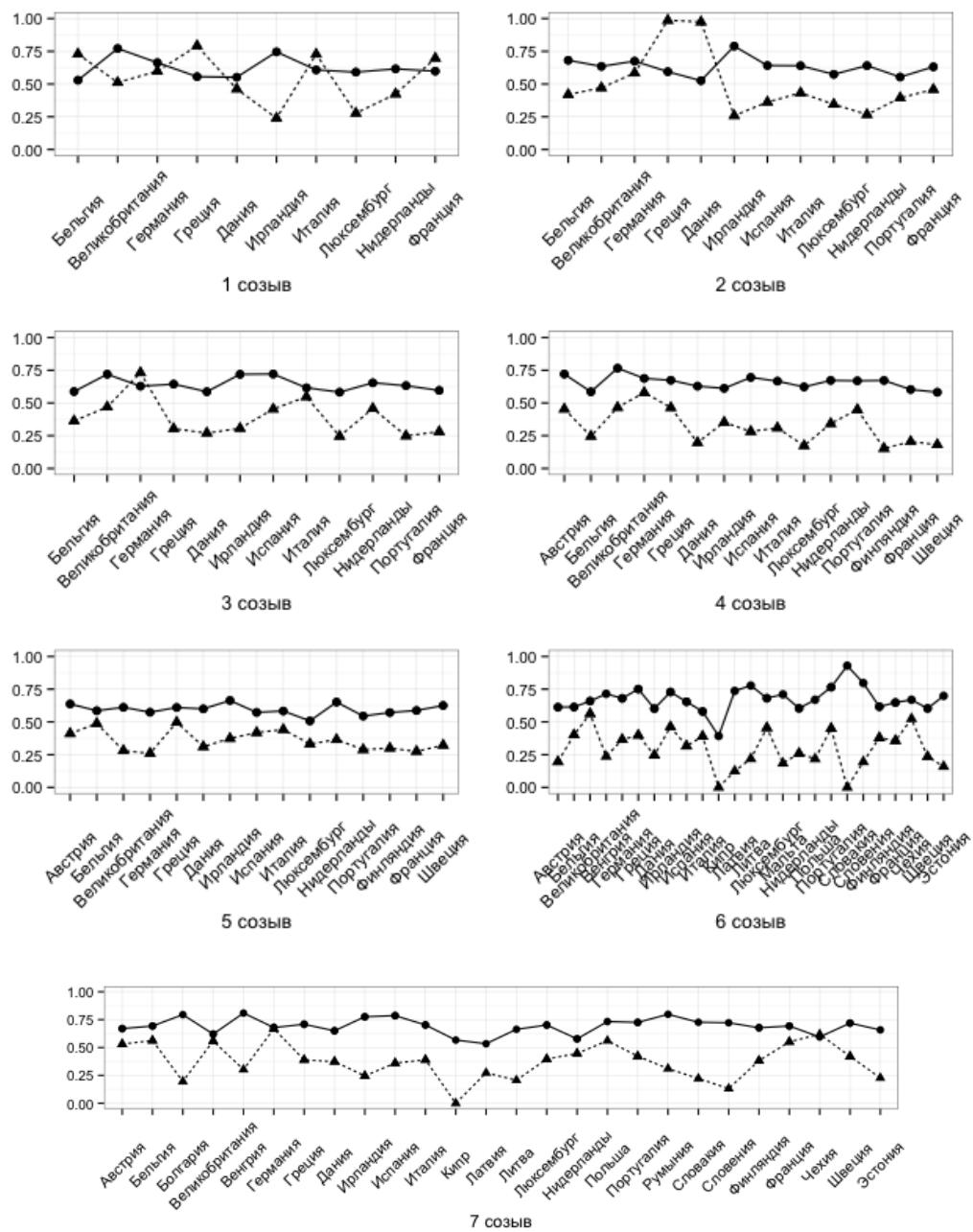


Рисунок 3.3 – Средний индекс согласованности Райса по всем голосованиям созыва по странам ЕС (сплошная линия) и размах —  $\max - \min$  — (пунктирная линия)

### 3.1.2.2 Индекс согласованности Сатарова

Индекс согласованности позиций двух групп законодателей (индекс согласованности) был построен на основе индекса конформизма, предложенного для определения того, насколько раскол в одной фракции по некоторому

голосованию отличается от раскола в другой фракции по тому же голосованию [129]. Формула расчета индекса имеет вид

$$c(q_1, q_2) = 1 - \frac{|q_1 - q_2|}{\max(q_1, 1-q_1, q_2, 1-q_2)},$$

где  $q_1$  — доля голосов за более распространенную позицию в первой группе,  $q_2$  — доля голосов за более распространенную позицию во второй группе. В данной работе для удобства примем в качестве  $q_j$  позицию «За».

Значения индекса согласованности позиций двух групп  $c(q_1, q_2)$  были рассчитаны для всех информативных голосований по активным депутатам и усреднены по каждому созыву.

Наиболее крупными политическими группами Европарламента являются фракции христианских демократов и социал-демократов, которые, для обеспечения большинства, часто вступают в коалицию с менее многочисленной группой либералов.

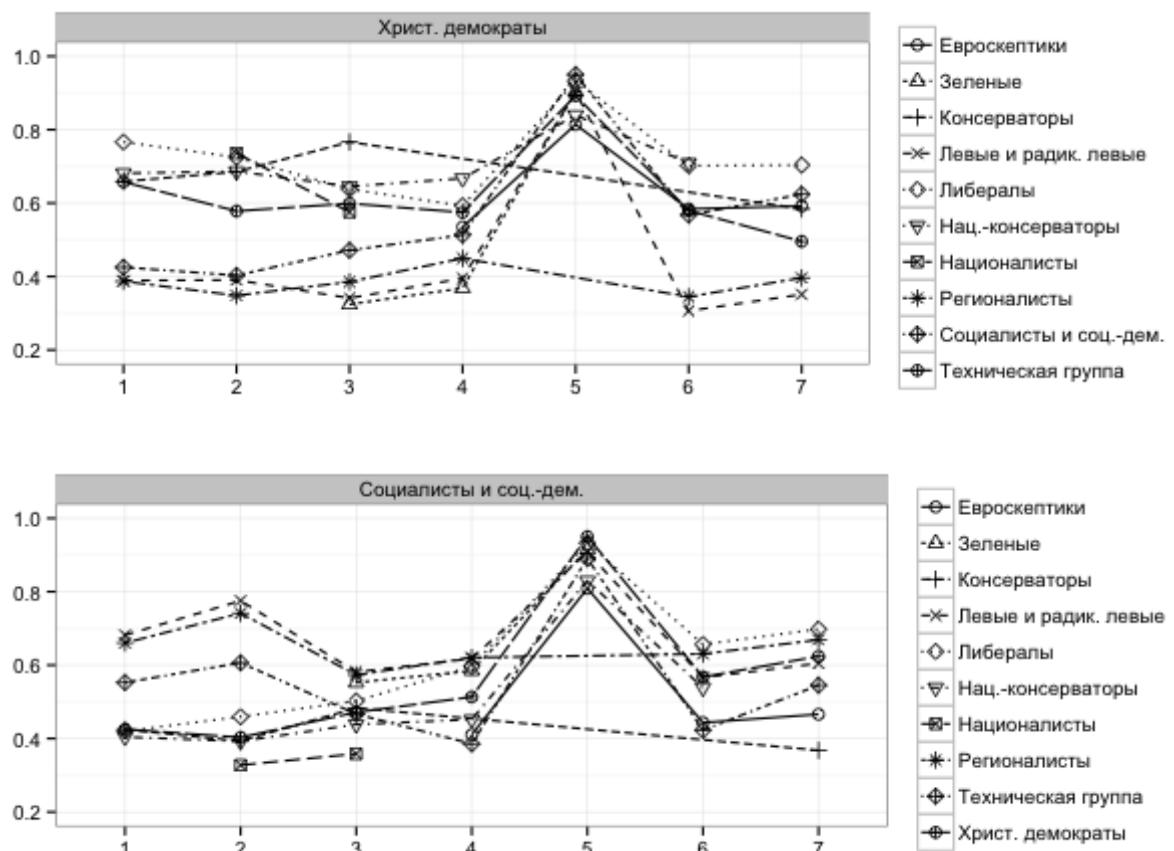


Рисунок 3.4 – Индекс согласованности политических групп по семи созывам

Наименее согласованными с христианскими демократами за весь рассматриваемый период были регионалисты и зеленые (степень согласованности при этом не превышала 0,5), а также левые и радикальные левые (рисунок 3.4). А наибольшей близостью позиций характеризовались либералы и национал-консерваторы, а также консерваторы (до третьего созыва включительно), националисты (во втором и третьем созыве) и независимые (техническая группа). Степень их согласованности с христианскими демократами в среднем превышала 0,5. Вместе с тем, в последних созывах выросла и согласованность между основными конкурентами: социал-демократами и христианскими демократами (значение индекса превысило 0,6).

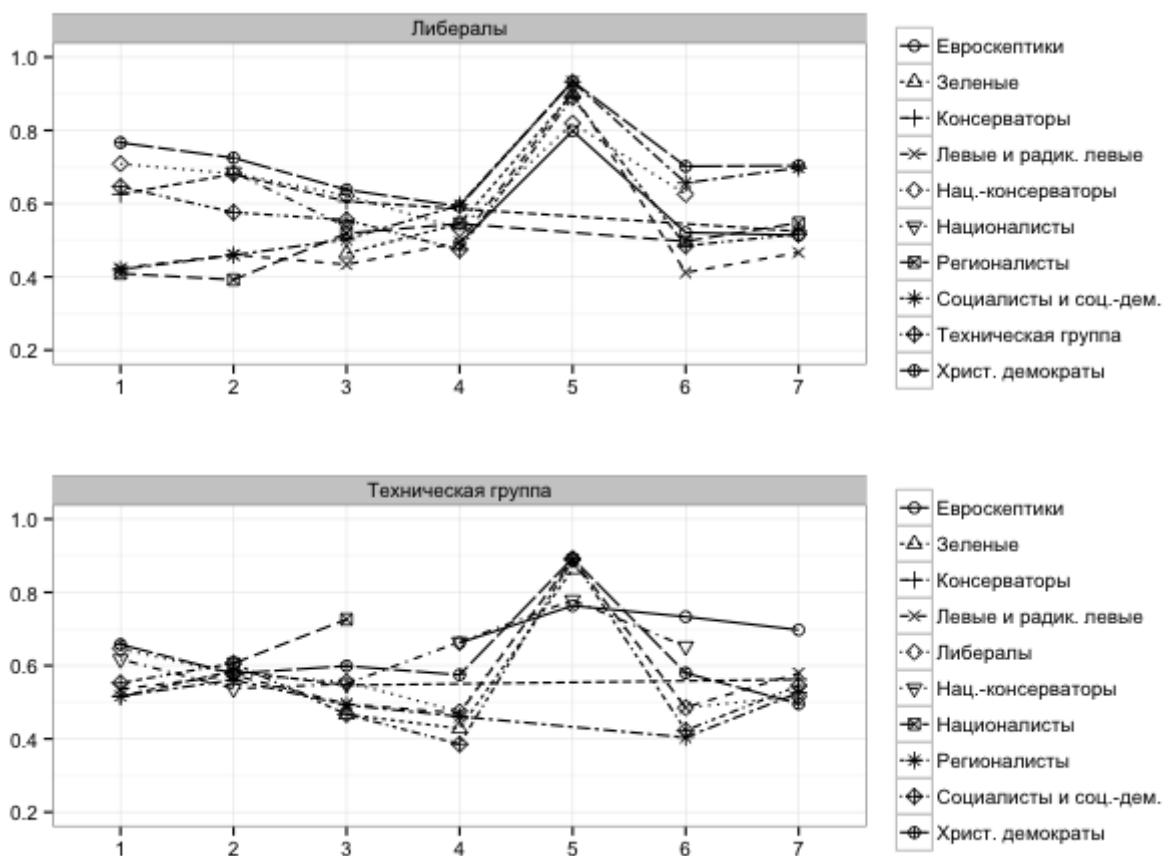


Рисунок 3.5 – Индекс согласованности политических групп по семи созывам

Вторую группу более согласованных часто формируют социал-демократы и примыкающие к ним более радикальные левые, а также зеленые и регионалисты (рисунок 3.4). (В среднем, по вопросам экономической политики зеленые, часто объединяющиеся с регионалистами, даже левее социал-

демократов [123].) В более поздних созывах либералы, как и христианские демократы, стали более лояльны социал-демократам (согласованность стала превышать 0,6). В целом, по вопросам экономического развития, либералы часто голосуют согласованно с христианскими демократами, а по вопросам гражданских прав, например, зачастую формируется лево-центристская коалиция либералов и социал-демократов. В «оппозиции» социал-демократам и «широкой» левой коалиции, особенно до пятого созыва, были депутаты из группы независимых, национал-консерваторы, а также в разное время — националисты и евроскептики.

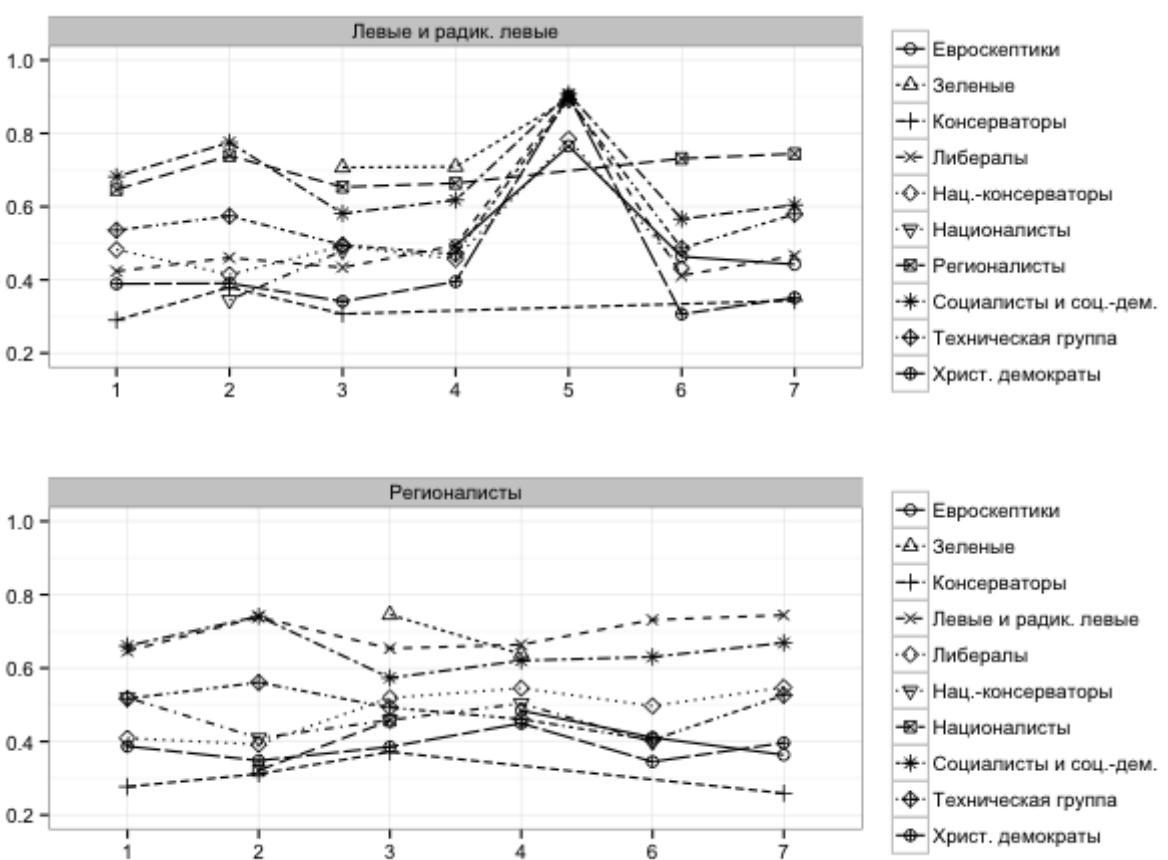


Рисунок 3.6 – Индекс согласованности политических групп по семи созывам

Помимо «широких» условных коалиций вокруг социал-демократов и христианских демократов, часто достаточно обособленно стоят группы евроскептиков, националистов, техническая группа, зеленые. Достаточно близко к ним примыкают радикальные левые. Все эти группы объединяет то, что на политической арене они являются относительно несистемными маргиналь-

ными игроками, чей политический вес мал, потому что позиция поддерживается достаточно узкими группами населения. При этом согласованность между этими группами депутатов не всегда является высокой. Так, например, степень согласованности регионалистов с консерваторами не превышала 0,4, как и согласованность регионалистов и националистов (не более 0,5) (рисунок 3.7), а с радикальными левыми и зелеными регионалисты достаточно согласованы — индекс в среднем равен 0,7 (рисунок 3.6).

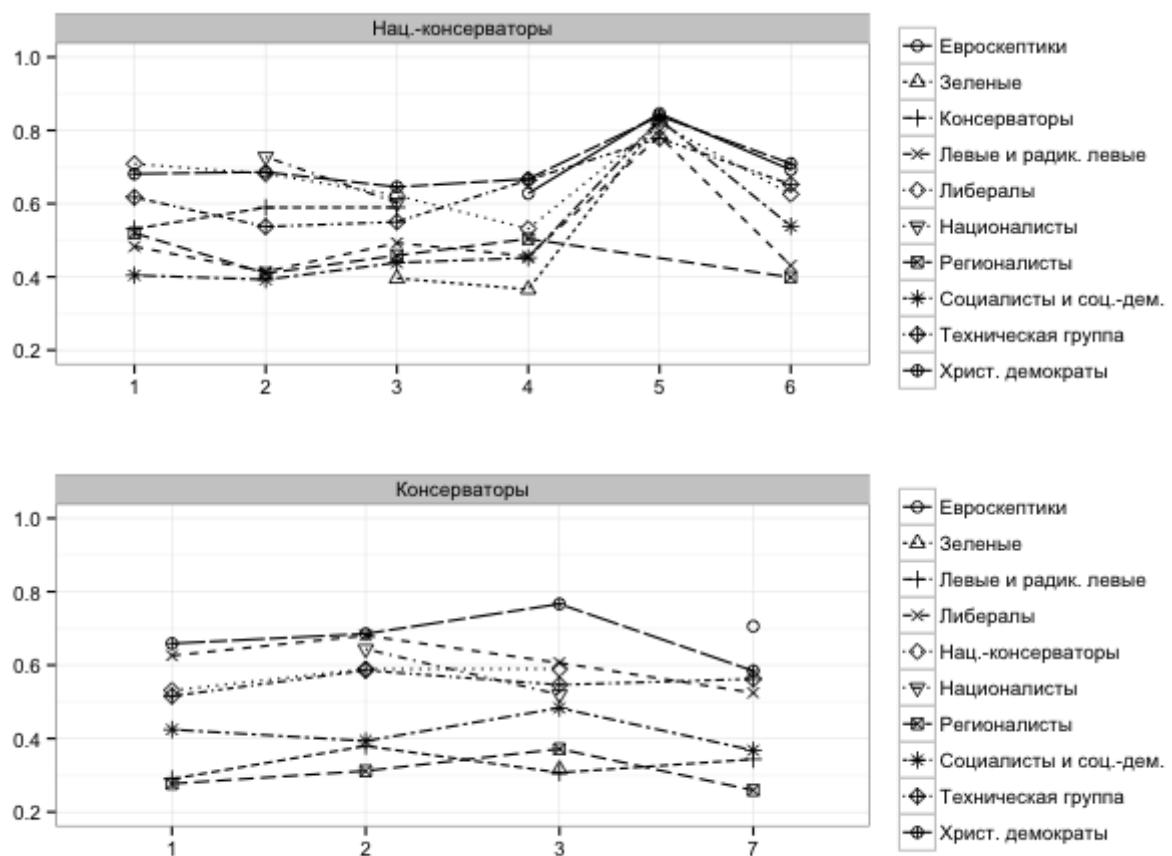


Рисунок 3.7 – Индекс согласованности политических групп по семи созывам

Графики выше были построены по матрицами попарной согласованности между политическими группами, и мы можем использовать их преобразование в качестве информации «на входе» для процедуры кластерного анализа, и увидеть, какие группы являлись наиболее близкими в каждом из для семи созывов Европарламента.

Являясь описательным методом, кластерный анализ позволяет получить группировку изучаемых объектов, не предъявляя каких-либо требований к

распределению признаков и априорной информации о числе групп. Все, что необходимо для реализации кластерного анализа — задать меру схожести объектов (метрику) и правило объединения в кластеры. Популярным методом агломерации является метод Варда, использующий идею минимизации внутригрупповой дисперсии при формировании кластеров [130]. Соответственно, метрикой для метода Варда является квадрат евклидова расстояния<sup>2)</sup>.

Исходя из визуального анализа дендрограмм — графического отражения последовательности появления кластеров, — можно говорить о следующих относительно устойчивых «коалициях» (рисунок В.1 в Приложении В):

- христианские демократы, консерваторы, национал-консерваторы, либералы (преимущественно до четвертого созыва);
- социал-демократы, левые, регионалисты, зеленые, либералы (с четвертого созыва);
- евросkeptики, техническая группа (независимые), националисты.

В основном в их основе лежит направление экономической политики, лево-правое разделение, но вторым важным критерием является позиция относительно европейской интеграции. Например, все крупные политические группы являются сторонниками общеевропейской интеграции, а мелкие партии, как правило, выступают против этого [126].

Другим «расколом», по которому в Европарламенте могут группироваться депутаты, является страна-член ЕС, откуда избирался депутат, т.е. национальная принадлежность. Стоит отметить, что значения индекса согласованности по странам-членам ЕС намного выше, чем по политическим группам, сформированным по идеологическому принципу, и имеет тенденцию к росту (см. рисунок 3.8).

---

<sup>2)</sup> Кластерный анализ позволяет получить относительно объективную группировку единиц наблюдения, т.к. является формальным методом, но в зависимости от метода агломерации и смены метрики он может выдавать различные по составу группы при одинаковом числе кластеров. Поэтому мы проверили устойчивость полученного результата, проведя кластерный анализ методом дальнего соседа с евклидовым расстоянием в качестве метрики. Отличия от ранее полученной группировки были минимальны.

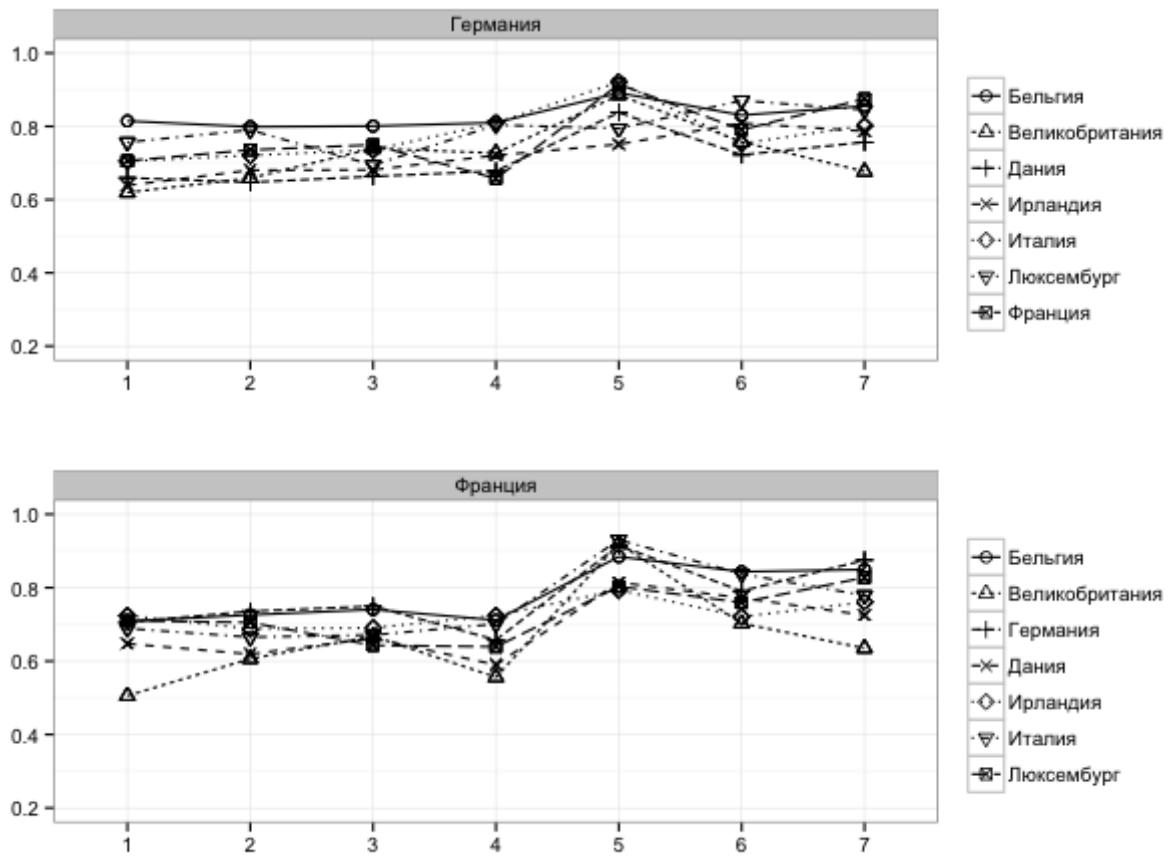


Рисунок 3.8 – Индекс согласованности стран по семи созывам

Для всех пар стран они редко опускались ниже 0.6. До четвертого созыва включительно более обособленно держались Великобритания и Ирландия, в том числе и друг от друга (рисунок 3.9).

Такой результат, не дающий возможности легко выделить четкие кластеры стран с согласованными паттернами голосования, не удивителен, потому что раскол по национальному признаку не является доминирующим в Европарламенте. Основной линией структурирования является идеологическая, лево-правая, принадлежность политических групп и отдельных депутатов [131 – 134]. Было показано, что внутригрупповая дисперсия голосований при группировке по идеологическим группам примерно вдвое ниже, чем межгрупповая. И ровно обратная картина получается, если сгруппировать депутатов по национальному признаку.

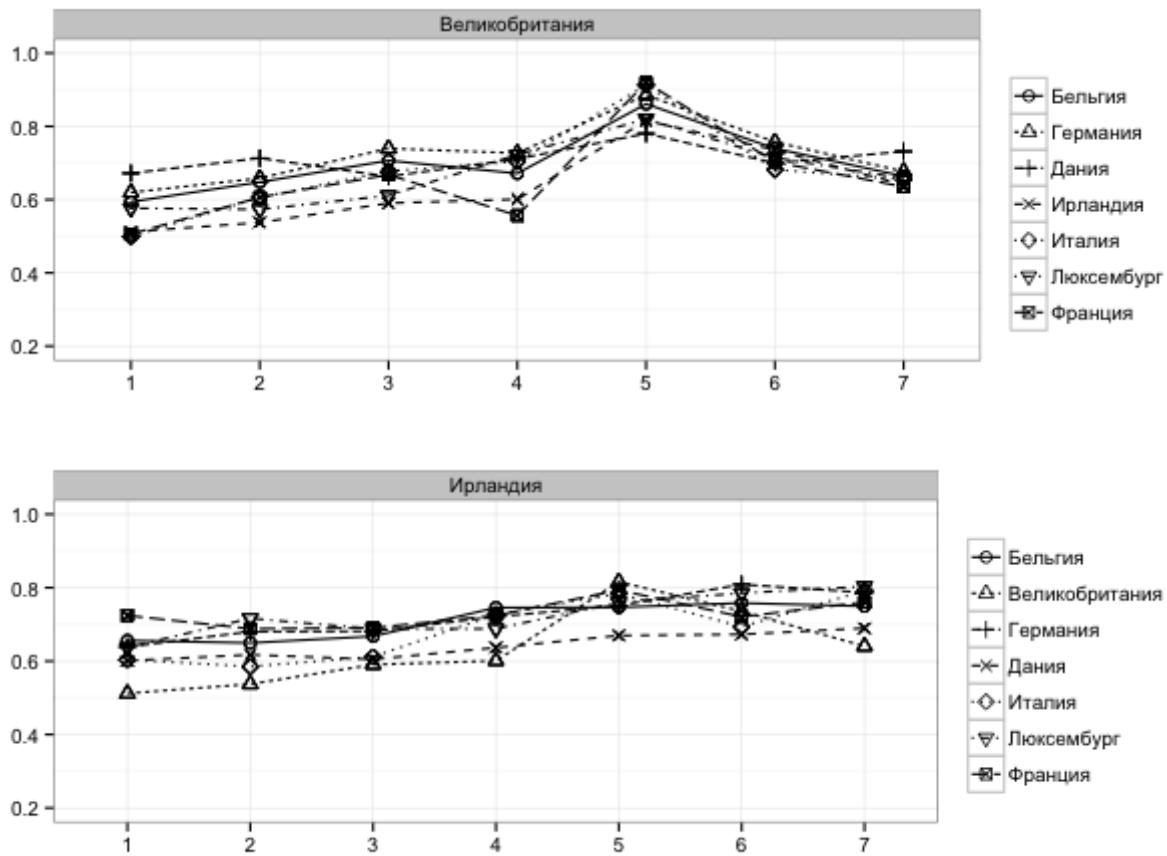


Рисунок 3.9 – Индекс согласованности стран по семи созывам

Тем не менее, национальная принадлежность депутата не является полностью бесполезной информацией — она отражает культурную разницу между Севером и Югом Европы, и частично — степень евроскептицизма. Кластерный анализ позволяет выделить группы стран, голосующих сходным образом (рисунок В.2 в Приложении В). К наиболее стабильным можно отнести следующие:

- Германия, Бельгия, Нидерланды
- Франция, Италия, Ирландия, Люксембург
- Великобритания, Дания
- Испания, Греция, Португалия
- Австрия, Финляндия, Швеция
- Венгрия, Словения, Словакия, Эстония, Литва
- Польша, Чехия, Латвия
- Кипр.

По-видимому, подобная группировка во многом является следствием «волн» расширения ЕС и географического и культурного соседства (рисунок 3.10).

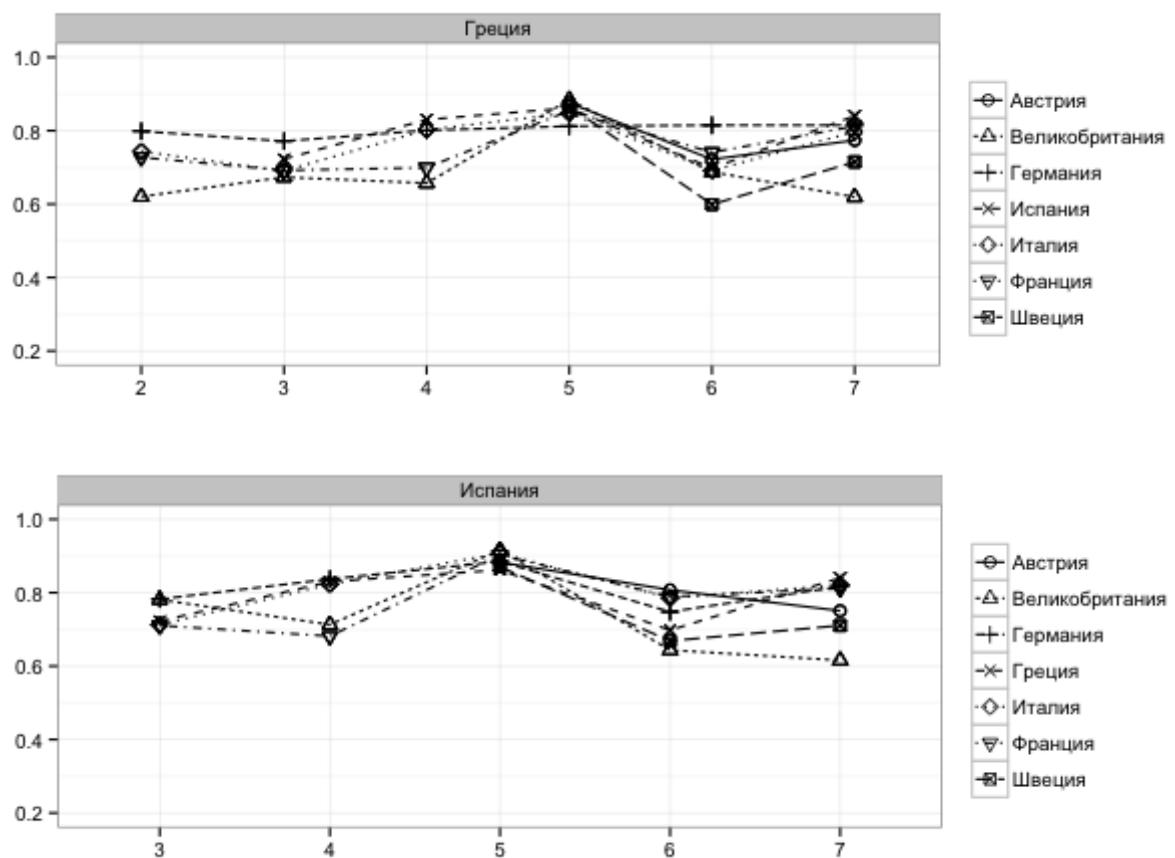


Рисунок 3.10 – Индекс согласованности стран по шести созывам

Другой канал влияния национальной принадлежности — через национальные партии. Национальные партии, хоть и являются теми организациями, которые борются за места в Европарламенте, контролируют деятельность своих депутатов преимущественно в период национальных выборов, когда ведется борьба за места в парламенте [135]. Тем не менее, систематическое игнорирование членом парламента позиции своей национальной партии, которая формирует партийные списки депутатов, чревато тем, что в следующий созыв такой депутат уже не попадет [126].

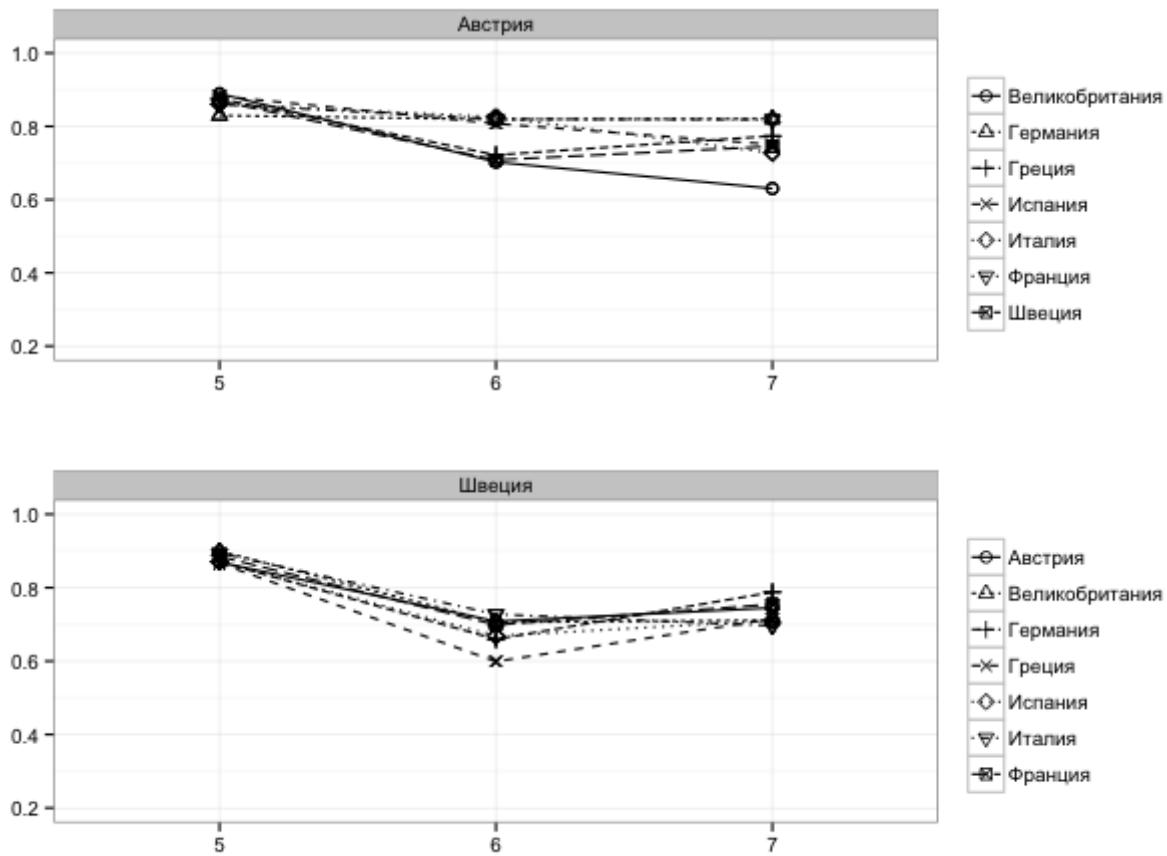


Рисунок 3.11 – Индекс согласованности стран по трем созывам

Например, не все независимые депутаты Европарламента считаются представителями экстремистских партий, однако часто их костяк формировали, например, крайние правые из Австрийской партии свободы, бельгийского «Фламандского блока» и французского «Национального фронта». А значит, индекс согласованности демонстрирует более высокие значения для этих стран ЕС (рисунок 3.11).

### 3.1.2.3 Фрагментированность и поляризованность Европейского парламента

Существуют два основных подхода к анализу культурного разнообразия обществ — с использованием мер фрагментированности общества [136] и мер поляризованности [137].

Фрагментированность представляет собой характеристику раздробленности стран или рынков на группы по некоторому признаку. Если страна  $A$  состоит из двух групп одинаковой численности, а страна  $B$  — из трех групп одинаковой численности, то фрагментированность в стране  $B$  больше.

Поляризованность, в свою очередь, учитывает еще степень сходства между группами, и чем меньше сходство, тем больше степень антагонизма в обществе. Поляризованность, как и фрагментированность, может измеряться по многим признакам: доходам, этнической, религиозной и лингвистической принадлежности, политическим симпатиям и пр., а также оцениваться как в масштабах рынка, населения страны, так и в рамках органа принимающего решения [138]. Предметом нашего интереса является оценка гетерогенности парламента с учетом близости групп, его составляющих, и без нее.

Пожалуй, наиболее популярной и доступной для расчета мерой фрагментированности является ИЭЛФ — индекс этнолингвистической фрагментированности. ИЭЛФ представляет собой преобразование индекса Херфиндаля—Хиршмана (ННІ), используемого в экономике для оценки степени концентрации рынка:

$$ИЭЛФ = 1 - HHI = 1 - \sum_{i=1}^N s_i^2,$$

где  $s_i$  — доля  $i$ -й группы на рынке,  $N$  — число групп. Полученную величину можно интерпретировать как вероятность, с которой два случайно выбранных в некотором обществе индивида будут принадлежать к разным группам. При полностью однородном обществе ИЭЛФ принимает значение ноль.

В основе индекса поляризованности  $ER$  Эстебана—Рея лежит стремление измерить, «насколько далеко распределение долей групп от биполярного, который представляет наивысший уровень поляризованности» [137]. Под биполярным распределением понимается ситуация, когда общество разделено на две одинаковые по численности группы, но придерживающиеся диаметрально противоположных позиций по важным вопросам. Индекс  $ER$  рассчитывается следующим образом:

$$ER = K \sum_{i=1}^N \sum_{j=1}^N s_i^{1+\alpha} s_j \tau_{ij},$$

где  $s_i$ ,  $s_j$  — доля  $i$ -й и  $j$ -й группы в общей численности населения,  $\tau_{ij}$  — дистанция между  $i$ -й и  $j$ -й группой,  $N$  — число групп,  $K$  и  $\alpha$  — масштабирующие коэффициенты, задающие необходимую исследователю область допустимых значений индекса. Зачастую исследователями используются  $\alpha = 1$  и  $K = 1$  [139]. В настоящей работе  $\tau_{ij}$  можно рассчитать как величину, обратную степени согласованности между парами групп, то есть  $1 - c(q_1, q_2)$ .

На рисунке 3.12 приведена динамика индексов фрагментированности и поляризованности по семи созывам Европарламента, рассчитанных как по идеологическому, так и по национальному признаку.

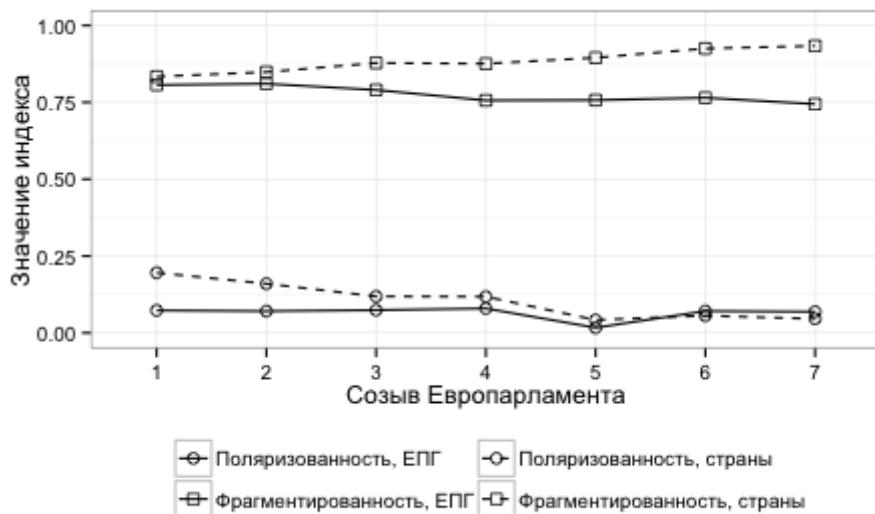


Рисунок 3.12 – Динамика фрагментированности и поляризованности по ЕПГ и странам-членам ЕС

Рост значений ИЭФЛ по странам-членам ЕС отражает тот факт, что почти в каждом созыве Европарламента мандаты получали депутаты от новых членов ЕС. Но при том, количество политических групп, которые формировались после выборов в парламент, значимо не менялось, удельный вес двух самых больших групп увеличился с 54% до 67%. Последний факт отражен в снижающейся фрагментированности, измеренной по ЕПГ.

Поляризованность по идеологическому признаку была примерно постоянной, что ранее демонстрировал и индекс согласованности Райса, а поляризованность по национальному признаку незначительно снизилась, что объясняется уменьшением относительной доли каждого из государств-членов и отсутствием нарастающего антагонизма между парламентариями от разных стран.

### 3.1.3 Некоторые выводы и заключение

В этой части отчета приведены результаты по оцениванию согласованности в Европейском парламенте. Для этого были получены оценки согласованности парламентариев как внутри каждой из групп, так и попарно между группами.

Было показано, что за первые семь созывов средняя внутригрупповая согласованность незначительно снизилась как по ЕПГ, так и по странам, причем согласованность внутри политических групп, относительно парламента в целом, падала быстрее. Относительная согласованность внутри стран-членов ЕС в последние два созыва даже выросла. Оценка фрагментированности и поляризованности по семи созывам показала, что степень поляризованности политических групп парламента практически не менялась, а степень поляризованности национальных групп заметно снизилась, в том числе за счет увеличения числа стран-членов ЕС.

Оценка попарной согласованности групп позволила выявить устойчивые множества из европейских политических групп и государств-членов ЕС, основанные на поименных голосованиях входящих в них депутатов.

Результаты исследований этого подраздела отражены в работе [140].

## 3.2 Оценка распределения влияния в Европейском парламенте в 1979–2014 гг.

В сегодняшнем мире объединенная Европа является значимым политическим игроком и принятие решений в рамках Европейского союза может иметь широкие последствия не только для стран-членов, но и далеко за ее

пределами. Политические институты Европейского союза уже давно являются предметом теоретических и эмпирических исследований. Активное расширение ЕС в последние 20 лет является предметом пристального внимания: год за годом происходила экономическая интеграция, менялась идентичность и ценности жителей Европы, реформировалось национальное и наднациональное законодательство, изменялись полномочия различных органов принятия решений, а вместе с ними — относительное влияние европейских политических групп и стран-членов ЕС.

С другой стороны, давно присутствует скепсис относительно того, является ли Европейский союз чем-то, кроме форума, на котором делегации от стран-членов и стран-кандидатов на вступление в ЕС могут отстаивать свои национальные интересы, а выборы в Европейский парламент называют «выборами второго порядка» и ставят под сомнение существование «европейской политической системы» [141 – 143].

Поэтому неудивительным является большое количество исследований на тему голосований членов Европейского парламента и распределения влияния в органах принятия решений ЕС [124, 126, 144 – 152].

В этом подразделе отчета приведены результаты анализа распределения влияния членов Европарламента. Для этого используются данные о поименных голосованиях депутатов, позволяющие ввести в анализ индивидуальные предпочтения по вступлению в коалиции: чем менее популярной является позиция депутата, тем в меньшем количестве коалиций он сможет состоять и менее влиятельным будет. Но для принятия решений необходимо большое число голосов, и вскоре после выборов идеологически близкие депутаты из разных стран формируют фракции — европейские политические группы, которые голосуют согласованно. В разных созывах группы являются разными по численности и составу, но численность не всегда отражает влияние на принятия решений, поэтому мы также оценивали влияние политических групп с учетом предпочтений их членов. С учетом постоянного расширения состава Евросоюза и связанных с ним изменений в численности Европейско-

го парламента, интерес представляет и распределение влияния на уровне стран-членов ЕС.

В пункте 3.2.1 описаны история Европейского парламента, правила его формирования и правила принятия решений. Пункт 3.2.2 включает описание исходных данных и методики оценки влияния депутатов. Пункт 3.2.3 содержит описание влияния на индивидуальном уровне, по европейским группам и по странам-членам Евросоюза. В Приложение Г вынесены некоторые громоздкие таблицы и рисунки.

### 3.2.1 Европейский парламент: история института и практика работы

#### 3.2.1.1 Основные задачи и история Европарламента

Европейский Союз представляет собой надгосударственное образование, для которого не характерно традиционное для национальных государств разделение на исполнительные, законодательные и судебные органы.

Основные институты системы управления в объединенной Европе, наряду с Европейским Советом, Европейской Комиссией, Советом Министров ЕС, Комитетом регионов и Судом ЕС, формирует и Европейский парламент (далее, для краткости, Европарламент или Парламент).

Европарламент является единственным институтом Европейского союза, который избирается непосредственно гражданами. Первые прямые выборы в Европарламент прошли в 1979 году, хотя объединение шести европейских стран (Бельгия, Германия, Италия, Люксембург, Нидерланды, Франция) в Европейское экономическое сообщество произошло еще в 1957 году.

Основными руководящими органами Европейского парламента являются председатель и бюро, включающие председателя, его 14 заместителей и 5 квесторов<sup>3)</sup>. Каждый парламентский цикл предполагает смену двух председателей.

Европарламент решает три важнейшие задачи:

---

<sup>3)</sup> Квестор отвечает за административные и финансовые вопросы, касающиеся депутатов, с правом совещательного голоса.

- законодательство (совместно с Советом министров ЕС);
- утверждение или отклонение бюджета в целом; контроль бюджета также совместно с Советом министров;
- контроль Европейской комиссии (Европейский парламент может выразить вотум недоверия Комиссии в целом, уполномочен одобрять или отклонять назначение Президента и членов Коллегии комиссаров).

Европейский парламент выполняет три функции:

- а) Европейский парламент делит с Советом министров ЕС законодательную функцию, но при этом не имеет права законодательной инициативы, а участвует лишь в процедуре принятия законов; право издания правовых норм принадлежит не Парламенту, а органам, представляющим правительства государств-членов, то есть Совету министров и в некоторых случаях — Комиссии Европейского сообщества;
- б) Парламент осуществляет демократический контроль над деятельностью всех институтов ЕС; он уполномочен одобрять или отклонять назначение Президента и членов Коллегии комиссаров, а также выражать вотум недоверия Комиссии в целом;
- в) Парламент ЕС влияет на расходы ЕС; Парламент утверждает или отклоняет бюджет в целом (статью «Расходы» в европейском бюджете устанавливают совместно Парламент и Совет, но лишь Совет определяет статью «Поступления» [153]).

Только с середины 1980-х гг. Европарламент перестал играть исключительно консультативную роль и получил полномочия контролировать, изменять и даже блокировать некоторые решения исполнительных органов ЕС. Амстердамский договор увеличил полномочия Европарламента путем значительного расширения Процедуры принятия совместных решений. Ниццкий договор также повысил его роль как соавтора законопроектов, расширив процедуру принятия совместных решений, а также наделив Европарламент правом предъявлять иски в Верховный Суд Европейского сообщества.

Тем не менее, эффективность работы Европарламента довольно долго оценивалась как сомнительная. В целом, он являлся довольно слабым в плане своей независимости и представительности: «Совет ЕС доминирует над Парламентом в законодательных вопросах и принятии бюджета, а граждане не связаны с членами Европарламента так, как они связаны с членами национальных парламентов» [154].

Лиссабонский договор, вступивший в силу 1 декабря 2009 года, предоставил Европарламенту право избрания председателя Еврокомиссии, из кандидатур, предложенных Европейским советом. Кроме того, он был наделен теми же правами, что и Совет ЕС. Так, когда эти положения впервые начали действовать, Европарламент получил более масштабные, чем ранее, полномочия и большее влияние на состояние и активность других ключевых органов Евросоюза — Еврокомиссии, Совета Европы.

В плане представительства интересов различных групп Евросоюз отличается большим разнообразием: группы потребителей, защитников окружающей среды, феминистские группы, бизнес-группы, профсоюзы активно лоббируют свои интересы в правительственные организациях и рабочих комитетах, система которых распространена в ЕС.

### 3.2.1.2 Правила формирования Европейского парламента

Выборы проводятся каждые пять лет во всех государствах-членах. В последних выборах 2014 года приняли участие более 168 миллионов жителей Европы (42.61%). Избирательная система на выборах в Европарламент в государствах-членах преимущественно пропорциональная. Императивный мандат запрещен, и депутаты не вправе следовать каким-либо инструкциям.

Последние выборы в Европейский парламент были проведены 22-25 мая 2014 года, хотя традиционно они проводились летом. С середины 1990-х гг. росли скептические настроения к ЕС, явка на общеевропейские выборы заметно падала и таким способом была предпринята попытка увеличить одновременно процент проголосовавших граждан и время до выборов

председателя Еврокомиссии [147]. Тем не менее, явка оказалась низкой — в целом всего 42%, а в Словакии, например, лишь 13% [155].

С 2014 года число членов Европарламента ограничено 751 человеком, включая председателя, а места распределяются по принципу «снижающейся пропорциональности»: минимум 6 представителей от государства (для Мальты, Люксембурга, Кипра и Эстонии), максимум — 96 (для Германии). Так, страны с большой численностью населения избирают больше депутатов по сравнению с меньшими государствами, но последние имеют больше мандатов, чем предполагала бы обычная пропорциональность. Порядок распределения мандатов закрепляется в Европейских договорах.

Впервые прямые выборы членов Европарламента прошли в 1979 году. В сентябре 1976 г. Совет принял Акт «Об избрании Европейской парламентской ассамблеи всеобщими прямыми выборами», в котором он установил лишь небольшое число правил, отдав остальное на усмотрение государств-членов. В Акте было прописано, что количество представителей, избираемых от каждого государства, зависит от численности населения каждой страны, но не напрямую. Неравенство в пользу малых стран теоретически рассматривается как способ обеспечения прав народов в условиях политической и экономической гегемонии крупных держав.

Члены парламента, чтобы эффективнее отстаивать свои позиции, организуются по политическим группам — европейским партиям, и такие фракции формируются по идеологическому принципу, а не по странам, от которых избирались депутаты. Регламент допускает создание группы при наличии 25 депутатов, представляющих 1/4 стран-членов. При нынешнем составе Евросоюза — 28 стран, речь идет о семи странах. Полный список европейских политических групп и используемых аббревиатур см. в Таблице Г.1 Приложения Г.

Традиционно среди политических групп выделяются социалисты, консерваторы, либералы, радикальные левые, крайние правые, «региональные», «зеленые» и евроскептики. Заметим, что распространена скептическая точка

зрения: «Европейские политические партии не функционируют кроме как на бюрократическом уровне» [156]. Тем не менее, принадлежность «идеологической семьи» на 75 процентов объясняет позицию национальных партий в «старых» странах ЕС по право-левой экономической оси и на 53 процента по про- и анти-европейской шкале. В «новых» странах эти проценты заметно ниже — 56 и 46, соответственно. Заметим, что поведение партий из «региональной» и анти-европейской группы является в меньшей степени согласованным [147].

В восьмом созыве Европейского парламента (2014-2019 гг.) выделяются 7 политических групп:

- европейская народная партия (христианские демократы), EPP — 221 место;
- прогрессивный альянс социалистов и демократов, S&D — 191 место;
- консерваторы и реформисты, ECR — 70 мест;
- альянс либералов и демократов, ALDE — 67 мест;
- европейские объединенные левые / левые и зеленые Северной Европы, GUE / NGL — 52 места;
- зеленые, GREENS/EFA — 50 мест;
- Европа свободы и демократии (евроскептики), EFP — 48 мест;
- вне партий / Независимые — 52 места.

В европейском и международном информационном пространстве итоги этих выборов характеризуются как революционные и неожиданные. Они изменили соотношение сил, считавшееся традиционным в последние десятилетия [157].

Ведущими стали две фракции: «Европейская народная партия», которая объединяет консервативные христианско-демократические партии Европы (Германии, Австрии, Греции и Испании), и «Прогрессивный альянс социалистов и демократов» во главе с нынешним председателем Европарламента Мартином Шульцем, сформированный из представителей социал-демократических, социалистических и рабочих партий Европы. Новизна со-

стоит в том, что до последнего времени евроскептики в Европарламенте воспринимались как докучливое меньшинство и в серьезный расчет при организации европарламентской деятельности не принимались. Теперь же обобщенное число публичных евроскептиков и причисляемых к ним депутатов измеряется, по разным оценкам, примерно на уровне 1/3 евродепутатов [157].

В научной литературе выборы в Европарламент часто рассматриваются как национальные выборы «второго порядка» [143, 147]. Долгое время ценность и эффективность Европейского парламента в смысле влияния на жизнь граждан Евросоюза была достаточно сомнительной. Так, в научной литературе сформировались две точки зрения. Первая состоит в том, что «европейское» измерение политики играет значимую роль на выборах (*“Europe matters”*). Вторая заключается в том, что избиратели склонны оценивать не работу своих партий и депутатов в Европарламенте, а рассматривать выборы в общеевропейский парламент как инструмент влияния на собственное национальное правительство, поэтому относятся к ним, как ко второстепенным выборам.

В пользу первой точки зрения говорит то, что во многих странах наибольший прирост голосов получают партии, которые четко позиционируют себя относительно ЕС. В 10 «новых» странах – тех, что вошли в состав ЕС в мае 2004 года, это преимущественно «региональные» партии, а в старых — антиевропейские партии и «зеленые».

Вторая позиция поддерживается тем, что наблюдается зависимость, при которой правящие партии теряют голоса в том случае, если выборы в Европарламент проходят спустя значительное время после национальных выборов. Граждане используют эту возможность, чтобы просигнализировать правительству о своем недовольстве и чаще голосуют за оппозицию. Более того, избиратели позволяют себе делать выбор искреннее, не ориентируясь на то, пройдет ли партия-фаворит в правительство, и это проявляется в том, что большие партии чаще теряют голоса, а мелкие, наоборот, получают больше голосов. Так, в [128] на данных по шести созывам Европарламента была

предпринята попытка оценить то, как выигрыш (*gain*) — разница в проценте голосов, полученным партией на выборах в Европейский парламент и на последних национальных выборах — зависит от времени прошедшего с момента проведения местных выборов, размера партии, ее правительственного статуса (в правительстве или в оппозиции) [128]. Было показано, что теория о выборах «второго порядка» имеет место, но выражена только в 15 «старых» странах ЕС, тогда как в 10 «новых» была обнаружена зависимость величины выигрыша только от того, входит ли партия в правительство (см. Таблицу Г.2 в Приложении Г). «Новые» члены Союза отчаются также тем, что в них наибольший прирост голосов получают «региональные» партии, а в старых — анти-европейские партии и «зеленые».

### 3.2.1.3 Правила принятия решений в Европарламенте

Принятие решений в Европарламенте может проходить в различном формате. Выделяются 4 процедуры:

- консультация — *consultation*;
- сотрудничество — *cooperation*;
- утверждение — *assent*;
- совместное решение — *codecision*.

В настоящий момент первые три носят наименование «специальные» процедуры, а процедура «совместного принятия решений» называется «обычной».

Первоначально учредительные договоры отвели Европарламенту только консультативную роль в принятии законодательства Сообщества. Принятие законодательства было в руках Совета. С развитием институционального устройства ЕС, в 1975 г. появился вариант согласительной процедуры — с целью дать Парламенту возможность блокировать реализацию законодательства, которое имело бюджетные последствия.

Процедура «сотрудничества» играла важную роль в 1987-1999 гг. до вступления в силу Амстердамского договора [158]. Суть ее в том, что после получения реакции Европарламента на решение Комиссии Совет может доработать текст, после чего Парламент может отклонить или принять «общее» решение во втором чтении. Но в случае отрицательного решения Парламент при единогласной позиции всех членов Совета все-таки может принять текст.

Процедура *«совместного принятия решений»* заключается в развитии последнего сценария: если после двух чтений, проведенных в каждом из институтов, Совет и Парламент не приняли единого текста, проект решения передается в согласительный комитет, который занимается согласованием компромиссного текста. Впоследствии он должен быть направлен для окончательного утверждения Парламенту и Совету. И если Парламент отклоняет общую позицию большинством своих членов, законопроект не принимается. Если согласительному комитету не удается прийти к соглашению в установленный период времени, текст автоматически считается не принятным и не может стать законом без того, чтобы не начать всю законодательную процедуру заново.

Процедура совместного принятия решений, значительно укрепившая позицию Парламента, была введена относительно недавно — Маастрихтским договором 1992 г. После вступления в силу Амстердамского договора в 1999 г. произошло значительное расширение случаев использования процедуры совместного принятия решений<sup>4)</sup> и сейчас она считается стандартной процедурой.

Разновидностью процедуры совместного принятия решений является процедура «утверждения» или «совпадающего положительного заключения»,

---

<sup>4)</sup> В самом начале применения процедуры Совет имел соблазн не изменять стиля своего поведения. Это означало разделение парламентских поправок на две группы, одна из которых включала поправки, принимавшиеся Советом, а другая состояла из отклоненных поправок. Парламент вынужден был согласиться на такое предложение по принципу справедливого деления пирога, но отказался следовать подобной практике и спровоцировал несколько серьезных конфликтов. Постепенно установка на конфронтацию изменилась, и появилось обоюдное желание найти третий путь.

когда Совет принимает решение после получения согласия Европарламента, рассматривающего проект решения в одном чтении.

Решение о принятии законопроекта или поправки к нему в Европарламенте принимается простым большинством. Кроме того, Парламент, большинством в 2/3, может выдвинуть вотум недоверия Комиссии, чем вызвать её отставку.

Помимо разных процедур, выделяются три регламента голосования:

- простое поднятие рук;
- электронное голосование (депутаты голосуют за, против или воздерживаются, но их голоса не записываются и оглашается только итоговый результат голосования);
- поименное голосование ('*roll-call votes*', депутаты голосуют за, против или воздерживаются, при этом их голоса записываются в протокол).

Протоколирование голосования проводится по некоторому списку вопросов, а также в случаях, когда на этом настаивает политическая группа или группа депутатов численностью более 20% от общего числа депутатов Европарламента. На практике около трети голосований проводятся с поименной записью голосов. Помимо решений каждого депутата, данные о поименных голосованиях содержат следующую важную информацию [159]:

- дата голосования;
- процедура:
  - а) законодательные процедуры (совместное решение, утверждение/санкция, сотрудничество, консультация);
  - б) бюджетные процедуры;
  - в) незаконодательные процедуры;
- правило принятия решения — абсолютное большинство или простое большинство;
- количество голосов «да», «нет», «воздержался»;
- количество депутатов, присутствовавших при голосовании;

– основной вопрос голосования.

Именно эти данные часто служат эмпирической базой для исследования поведения членов Европарламента, анализа партий, структурной устойчивости парламента и распределения влияния. Например, для определения размерности политического пространства в Европарламенте, а также позиционирования отдельных депутатов часто используется метод NOMINATE [160, 161]. В среднем поведение евродепутатов в большей степени согласовано с позицией политической группы, в которую входит их национальная партия, чем с позицией их национального правительства (позиции страны) [162]. По оценкам, знание только того, какую страну представляет член Европарламента, позволяет верно предсказать его голосование в 10% случаев, а знание только того, в какой группе он состоит — в 90% случаев. Так, можно говорить о том, что французские и немецкие социал-демократы более близки по поведению в Европарламенте, чем британские лейбористы и консерваторы.

Поэтому в данной работе мы предлагаем проанализировать распределение влияния на уровне отдельных депутатов, стран и общеевропейских политических групп.

### 3.2.2 Методика оценки распределения влияния и моделирования коалиционных предпочтений

#### 3.2.2.1 Методика оценки распределения влияния

Влияние фракций и депутатов на принятие решений в органах принятия коллективных решений далеко не всегда пропорционально доле мест, которой они обладают. Покажем это на простом примере.

Допустим, парламент состоит из 100 мест, и в нем представлены 3 партии. Партия *A* имеет 50 мест, партия *B* — 49, в партия *C* — только 1. В случае если для принятия решения необходимо набрать простое большинство, то коалиции *A+B*, *A+C*, *A+B+C* будут выигрывающими, т.е. они могут принять решение без учета мнения других партий. Партии *B* и *C* делают выигрываю-

щими по одной коалиции, а без участия партии  $A$  вообще невозможно принять решение. Иначе говоря, как партия  $B$ , так и партия  $C$  имеют одинаковое влияние на принятие решений, хотя одна из них имеет 49 голосов, а вторая — только 1 голос.

Индекс Банцафа основывается на подсчете доли коалиций, в которых участник является ключевым — без участия которого коалиция перестает быть выигрывающей, т.е. такой, которая может принять решение при голосовании по вопросу, преодолев квоту [163]. Формула для расчета значений индекса выглядит следующим образом

$$\beta_i = \frac{b_i}{\sum_j^n b_j},$$

где  $b_i$  — количество коалиций, в которых участник  $i$  является ключевым,  $i \in 1, \dots, n$ .

В приведенном примере влияние партии  $A$  составляет 1, т.е.  $\beta_A = 3/3 = 1$ , а партий  $B$  и  $C$ , соответственно,  $\beta_B = \beta_C = 1/3$ . При этом, все коалиции считаются одинаково возможными, т.е. партии с одинаковой готовностью формируют любые выигрывающие коалиции, вне зависимости от их состава. Но зачастую это не так. Некоторые выигрывающие коалиции не могут быть сформированы, к примеру, по идеологическим соображениям, или их созданию предшествуют сложные переговоры и коалиция постоянно сталкивается с риском распада. Способ расчета индексов влияния в органах принятия коллективных решений с учетом предпочтений членов по созданию коалиций предложен в [164].

Зная насколько близкими друг к другу являются депутаты или партии, т.е. силу связи  $p_{ij}$  участниками  $i$  и  $j$ , для каждой из них можно определить силу связи с той или иной выигрывающей коалицией  $S$ . Сила связи партии с коалицией задается функцией интенсивности связи  $f(i, S)$ .

В зависимости от специфики изучаемого органа принятия коллективных решений, могут использоваться различные виды функций связи [164]. В настоящей работе мы будем использовать функцию интенсивности связи, построенную на «прямых» предпочтениях каждой  $i$ -той группы вступить в коалицию с группой  $j$

$$f(i, S) = \frac{1}{|S|-1} \sum_{j \in S, j \neq i} p_{ij}.$$

Используя полученные значения функций связи, можно рассчитать влияние каждой партии,  $\chi_i$ , для чего суммируются все значения  $f(i, S)$  в таких выигрывающих коалициях, где  $i$  — ключевой участник,

$$\chi_i = \sum_{S, i \in S, |S|=q} f(i, S).$$

Ключевым называется такой участник коалиции, которых делает возможным принятие решения этой коалицией. Если же ключевой участник выходит из коалиции, то она не может преодолеть необходимую квоту.

Чтобы ограничить область возможных значений в границе от 0 до 1, производится нормировка:

$$\alpha_i = \chi_i / \sum_j^n \chi_j.$$

Для расчета значений индексов влияния, учитывающих предпочтения по созданию коалиций на уровне групп, введем индекс  $\alpha_{gr}^1$ . Функции связи в этом случае будет отражать силу связи между фракциями, например, степень согласованности между двумя группами или среднюю близость между парами депутатов. Стандартным предположением при его расчете является то, что группа голосует консолидировано.

Большинство широко используемых индексов рассчитываются для групп, например, парламентских фракций. Тем не менее, исследовательский интерес могут представлять и влияние отдельных депутатов. Для оценки индивидуального влияния членов Европарламента будем использовать индекс  $\alpha^1$  с функцией интенсивности связи

$$f(i, S) = \frac{1}{|S|-1} \sum_{j \in S, j \neq i} p_{ij},$$

где  $i$  и  $j$  — депутаты Европарламента.

В случае, когда влияние оценивается на уровне отдельных участников коалиций, а не консолидированных фракций, возникает очень большое количество выигрывающих коалиций, и задача оценки влияния становится чрезвычайно сложной с вычислительной точки зрения. Вычислительная схема для расчёта индексов влияния отдельных членов парламента была предложена в [165, 166].

### 3.2.2.2 Моделирование коалиционных предпочтений

Важным шагом при расчете значений индекса влияния с учетом предпочтений по созданию коалиций является описание силы связи между участниками коалиций, т.е. формирование матрицы предпочтений депутатов Европарламента. Чтобы задать матрицу  $P = \{p_{ij}\}$ , мы будем использовать данные о поименных голосованиях депутатов. Об отдельных членах Европарламента доступна информация о том, от какого государства депутат был избран в Европейский парламент, парламентская группа, в которой состоит депутат, и его национальная партия. По каждому протоколированному голосованию по законопроекту/решению про каждого депутата можно сказать, был он «за», «против», «воздержался», «присутствовал, но не голосовал», «не член парламента»<sup>5)</sup> или «отсутствовал».

---

<sup>5)</sup> Для стран, вступающих в союз, обычной практикой является отправление наблюдателей в парламент. Число наблюдателей и их назначение (обычно национальными парламентами) установлены для присоединяющихся стран в соглашениях по вступлению в союз.

Наблюдатели могут посещать заседания и участвовать в дебатах, но у них нет права голоса. Когда страна становится полноправным членом ЕС, наблюдатели становятся евродепутатами на период между вступлением государства и выборами в парламент. Например, ГДР имела 18 наблюдателей в 1990—1994 годах. Также с 26 сентября 2005 года по 31 декабря 2006 года у Болгарии было 18 наблюдателей в парламенте, а у Румынии — 35. Они были выбраны из правящей и оппозиционных партий и одобрены национальными парламентами этих стран. После вступления стран в союз наблюдатели стали европарламентариями (но с некоторыми изменениями в составе).

Мы предлагаем два подхода к моделированию коалиционных предпочтений. Первый заключается в том, что матрица предпочтений депутатов Европейского парламента  $P = \{p_{ij}\}$  определяется как вероятность того, что депутаты  $i$  и  $j$  проголосуют одинаково,  $p_{ij}$ . Для этого рассчитывается средняя величина  $p_{ij} = \frac{v_{ijt}}{n_t}$ , где  $v_{ijt}$  — количество голосований, когда депутаты  $i, j$  проголосовали одинаково,  $t$  — номер голосования,  $n_t$  — количество голосований, в которых приняли участие и депутат  $i$ , и депутат  $j$ . Полученная таким образом величина  $p_{ij}$  принимает значения от 0 до 1, где 1 означает, что депутаты вели себя максимально согласованно. Что касается распределения влияния на уровне групп депутатов (сформированных по идеологическому или национальному признаку), то их предпочтения по вступлению в коалиции будут рассчитываться как средняя согласованность.

Второй подход предполагает определение позиций евродепутатов в некотором политическом пространстве. NOMINATE — популярный метод многомерного шкалирования, разработанный К. Пулом и Г. Розенталем для оценки идеальных точек депутатов конгресса США на основании результатов поименных голосований [160, 161]. NOMINATE — это акроним от NOMINAL Three-Step Estimation, а W-NOMINATE — не-динамическая версия метода. С его помощью можно получить оценки расположения отдельных индивидов (депутатов) по нескольким измерениям — осям. Априорное знание количества осей не требуется, но содержательная интерпретация полученных измерений ложится на плечи исследователя. Было обнаружено, что первое измерение, ось NOM-D1, корректно объясняет 85% голосований в первом созыве Европейского парламента, около 90% в пятом созыве и 88% в шестом, и это измерение может интерпретироваться как расположение на лево-правой идеологической шкале [128]. Второе измерение, ось NOM-D2, может считаться оценкой позиции по «про- и анти-Европейской» шкале. Использование второй оси увеличивает долю объясненных голосований до 91% в первом созыве Парламента и 92% в пятом и 90% в шестом. Что касается распределе-

ния влияния на уровне групп депутатов (сформированных по идеологическому или национальному признаку), то их предпочтения по вступлению в коалиции будут рассчитываться как средние значения по группе.

Оба подхода дают схожие оценки в смысле распределения влияния относительно групп депутатов или отдельных парламентариев, но использование позиционирования в осях NOMINATE позволяет получить более выраженный результат и сделать интерпретацию более удобной.

Для семи созывов были зарегистрированы почти 23,5 тыс. поименных голосований по процедуре простого большинства. Из них 10,2% имели согласованность выше 0,95, 18% — согласованность выше 0,90, 38% — согласованность выше 0,75. Степень согласованности между членами парламента измерялась следующим образом:

$$m^t = |2q^t - 1|,$$

где  $q^t$  — доля проголосовавших «за» по  $t$ -ому голосованию в парламенте. (Более подробный анализ согласованности парламентариев и групп см. в [140].)

Таблица 3.2 – Процент голосований с согласованностью выше порога

Порог согласованности	Процент голосований с согласованностью выше порога
0.98	6,7
0.95	13,2
0.90	21,0
0.85	27,9
0.80	34,7
0.75	41,5
0.70	47,4

Полученные результаты также во многом зависят от того, как определить неинформативные голосования, подлежащие исключению из анализа, чтобы получить матрицу предпочтений без возможного влияния заведомо проходных вопросов, завышающих настоящую согласованность депутатов. В данной работе к неинформативным голосованиям были отнесены такие, для которых доля проголосовавших «за» была больше 0,80 или меньше 0,20. Так,

все голосования, для которых  $m^t > 0,80$ . В первом созыве после отбора осталось 506 информативных голосования (из 700). В последующих, соответственно, 1096, 1422, 2121, 3364, 2189 и 3078. В среднем процент информативных голосований составляет 65%, и для оценки распределения влияния по семи созывам использовались более 13,8 тыс. голосований.

Всего в анализе участвовали почти 2,7 тыс. активных депутата за пять созывов. Активным считаются депутаты, пропустившие не более четверти поименных голосований ( $\tau = 0,25$ ). Дисциплинированные депутаты более ответственно реализуют свой мандат и могут считаться более последовательным в своем поведении, поэтому в их поведении ожидается меньше «шума» и «случайности». Таким образом, в среднем были исключены 26% депутатов (в первом созыве — 37%, в седьмом — 23%).

Выбор таких пороговых значений был обусловлен анализом чувствительности значений индекса индивидуального влияния  $\alpha^1$  к изменениям значений двух параметров —  $m^t$  и  $\tau$  (см. рисунки Г.1 и Г.2 в Приложении Г).

### 3.2.3 Описание результатов оценки распределения влияния депутатов Европарламента, политических групп и стран-членов ЕС

На имеющихся данных были рассчитаны значения индексов индивидуального влияния и индексов влияния для групп депутатов (по европейским политическим группам и по странам-членам ЕС). В обоих подходах учитывались предпочтения членов Европейского парламента по созданию коалиций.

#### 3.2.3.1 Индивидуальное влияние членов Европарламента

Распределение значений линейного индекса  $\alpha^1$  индивидуального влияния в разные созывы довольно сильно различалось. Среднее (медианное) влияние отдельного депутата Европарламента в первом созыве было примерно вдвое выше, чем в седьмом, что не удивительно: в 1979 году в Европей-

ском парламенте было всего 410 депутатов, а в 2014 году (в начале 7 созыва) — 736. Хотя Евросоюз постоянно расширялся, тем не менее, количество политических групп в разных созывах не превосходило 10-12: радикальные левые, социал-демократы, либералы, христианские демократы, консерваторы, национал-консерваторы, националисты, регионалисты, зеленые, евроскептики. Такое деление отражает основные идеологические семьи, за исключением «технических» групп и евроскептиков, которые формируются по иным критериям, отличным от идеологической схожести взглядов депутатов.

В *первом созыве* Европейского парламента значения индекса индивидуального влияния менялись от 26.5 до 49 (умноженные на  $10^4$  для удобства). В него вошли депутаты от 44 партий из 10 европейских государств, сформировавшие 8 политических групп, всего 410 парламентариев (при этом, активных из них 253, подробнее см. Таблицу Г.3. в Приложении Г). Наименьшим числом активных депутатов были представлены Люксембург и Греция (по 4 активных депутата), наибольшим — Германия (59 активных депутата).

По итогам выборов многие депутаты вошли в группы христианских демократов (EPP), социалистов (SOC) и консерваторов (ED) (полные наименования политических групп Европейского парламента см. в Таблице Г.1 в Приложении Г). Малочисленные группы сформировали регионалисты (CDI) и национал-консерваторы (EPD), а также была образована «техническая» группа независимых депутатов<sup>6)</sup> — группа, объединявшая депутатов разных политических убеждений, не вошедших в политические группы, но желавших иметь преимущества и права тех, кто в группах состоит: места в комитетах и доступ к бюджетам, и потому сформировавших «номинальную» общеверопейскую группу.

Значения индекса индивидуального влияния не сильно различаются, но тем не менее, на рисунке 3.13 видно, что в среднем наибольшим влиянием обладали члены группы христианских демократов, либералов (LD) и «техни-

---

<sup>6)</sup> Такая «техническая группа» была расформирована в октябре 2001 года, с введением требования о наличии у политической группы Европарламента согласованной позиции [128].

ческой группы». Их активные депутаты имели влияние выше среднего. Наименьшим влияниям обладали левые и радикальные левые депутаты, регионалисты, а также большая часть социал-демократов.

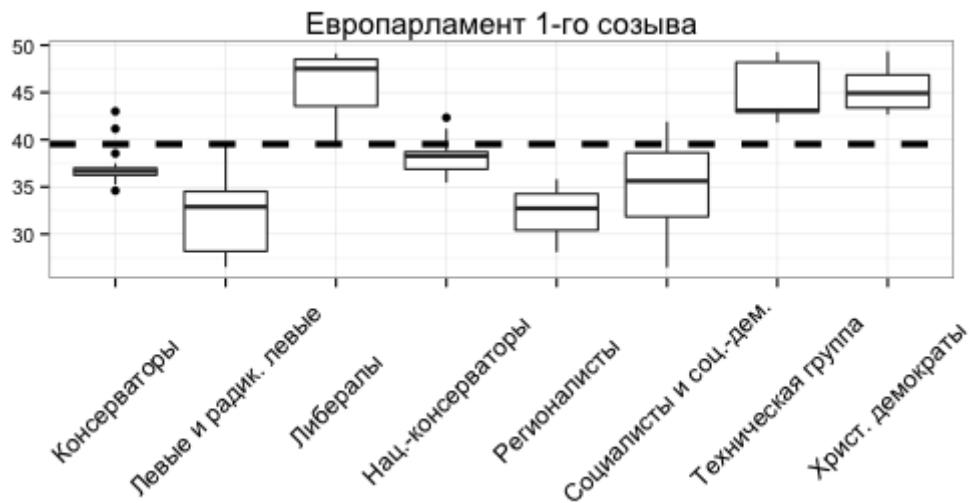


Рисунок 3.13 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по европейским политическим группам для 1-го созыва Европарламента

По странам, от которых баллотировались депутаты, наименьшим влиянием в среднем обладали представители Великобритании и Ирландии, а также многие депутаты от Италии, — у этих стран наименьшие медианные значения индекса индивидуального влияния. Наиболее влиятельными в среднем были представители Дании, Германии, Нидерландов и Люксембурга. Несмотря на то, что Люксембург, в силу небольшой численности населения, имеет маленькое количество мест в Европейском парламенте, его депутаты входят во влиятельные политические группы либералов, социалистов и христианских демократов. Подробнее см. рисунки 3.14 – 3.15.

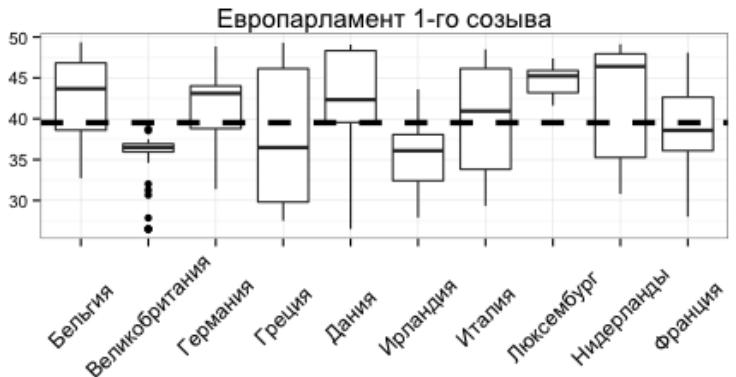


Рисунок 3.14 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС для 1-го созыва Европарламента

Максимальным влиянием во втором созыве Европейского парламента, согласно индексу  $\alpha^1$ , обладали христианские демократы: Д. Антониоцци, В. Гуммара, Дж. Джиавацци, Дж. Берсани из Италии от блока «Христианско-демократическая партия» / «Итальянская народная партия» и К. Малангра и Г. Л. Леммер из Германии от Христианско-демократического союза Германии.

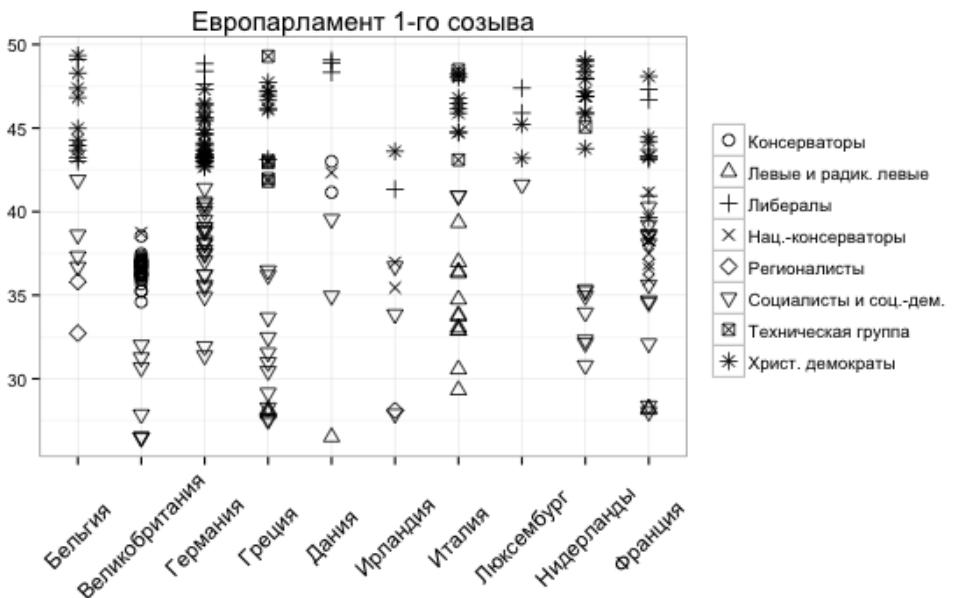


Рисунок 3.15 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС и европейским политическим группам для 1-го созыва Европарламента

Во *втором созыве* Парламента (1984 — 1989 гг.) отдельно выделилась еще одна общеевропейская группа — националисты (ER). Ее численность

была весьма невелика — всего депутатов 16, из которых 6 активных. Христианские демократы (EPP) и либералы (LD) по-прежнему имели в среднем наибольший относительный вес при принятии решений, и христианские демократы при этом были второй по численности группой депутатов (75 человек), в группе либералов, в свою очередь, состояли 15 человек. Достаточно влиятельной оппозицией им была группа социалистов (SOC), численностью 95 активных депутатов. В среднем наименее влиятельными в данном созыве вновь были группы консерваторов (ED) и национал-консерваторов (EDA), по 41 и 15 активных депутатов.

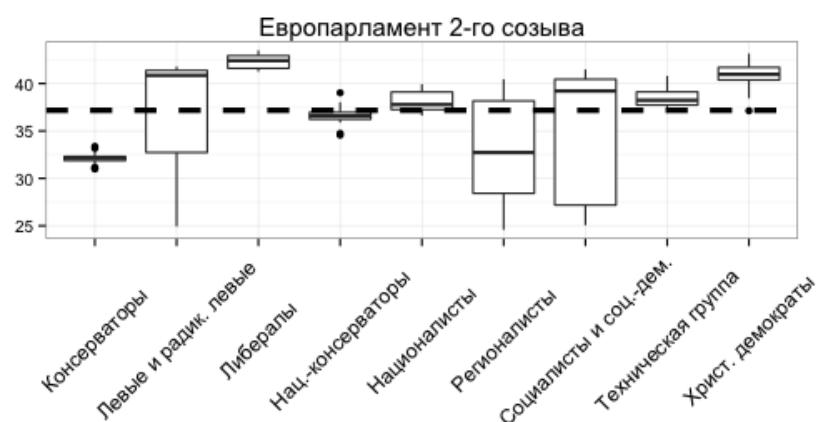


Рисунок 3.16 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по европейским политическим группам для 2-го созыва Европарламента.

В этом созыве Европарламента влияние британских представителей также было небольшим. В среднем более низким влияниям обладали и депутаты от Франции (40 депутатов состояли в радикальных правых — EDA, ER — и левой группе — СОМ). Для многих государств можно отметить представителей с нетипично маленькими значениями индекса индивидуального влияния.

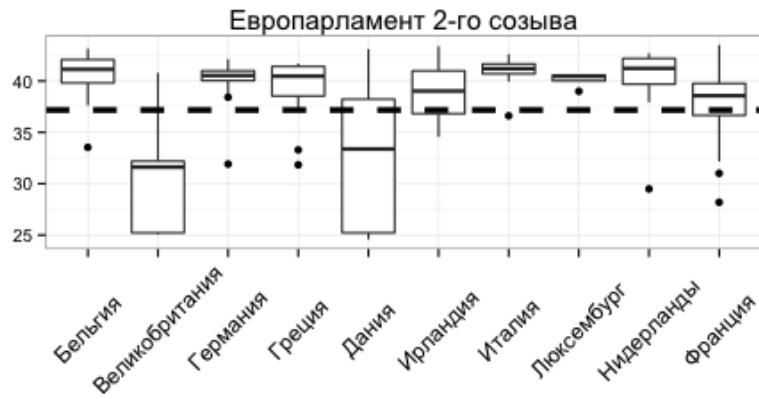


Рисунок 3.17 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС для 2-го созыва Европарламента

Максимальным влиянием во втором созыве Европейского парламента, согласно индексу  $\alpha^1$ , обладали христианские демократы: Р. Ластер, Э. Брок, Г. Ринш, К. Малангра из Германии от Христианско-демократического союза Германии и Ж. Пендрес и У.Дж. Вергер от голландской «Христианско-демократический призыв».

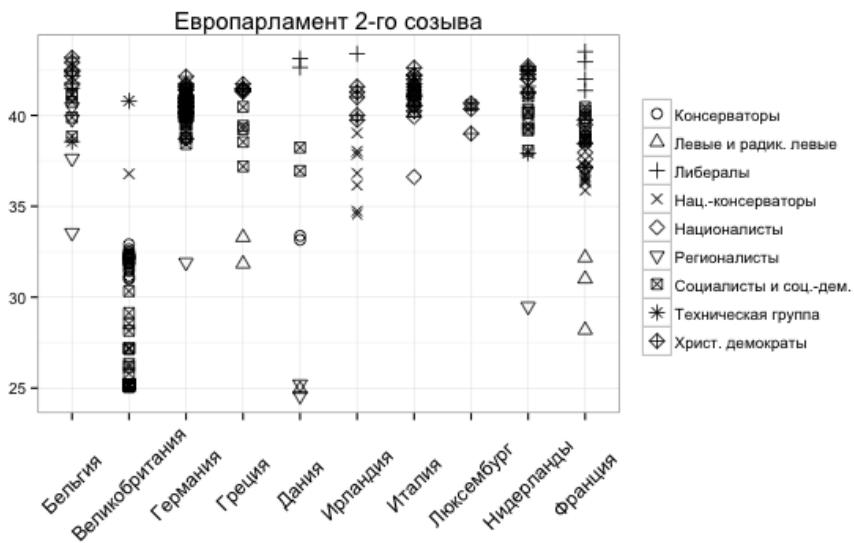


Рисунок 3.18 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС и европейским политическим группам для 2-го созыва Европарламента

В *третьем созыве* Европарламента (1989 — 1994 гг.) насчитывалось уже 518 депутатов, но активных среди них было только 432. Наиболее многочисленной группой стали социал-демократы (SOC), в эту группу вошли более 100 человек. При этом самой маленькой группой стали радикальные

левые EUL & LU (всего 3 активных депутата). В среднем наименее влиятельными в данном созыве были зеленые (G) и националисты (DR), имевшие по 12 и 7 активных депутатов. Максимальные значения индексов влияния в среднем были у коалиции левых и социал-демократов (по 13 и 132 человека соответственно). Либералы (LDR) и христианские демократы (EPP) составляли им менее влиятельную, но достаточно многочисленную оппозицию (более 110 активных депутатов).

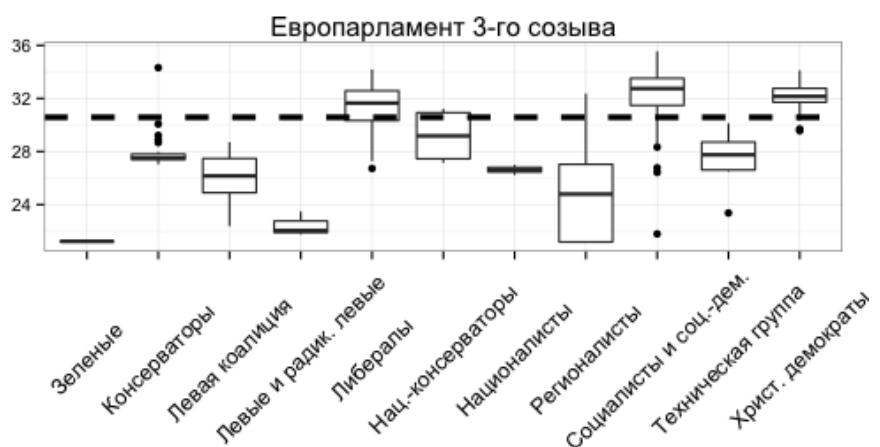


Рисунок 3.19 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по европейским политическим группам для 3-го созыва Европарламента

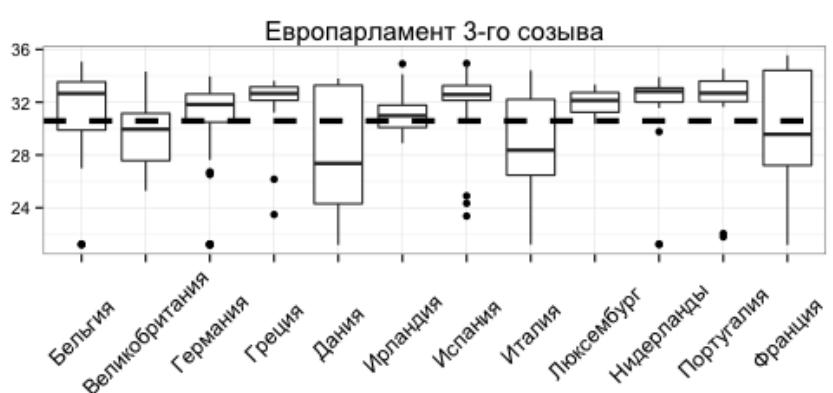


Рисунок 3.20 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС для 3-го созыва Европарламента

Депутаты из Германии, Италии, Португалии и Люксембурга в среднем имели относительно большое влияние. Несколько меньшим влиянием в среднем обладают парламентарии из Бельгии, Ирландии, а также отдельные

депутаты из Германии, Италии и Великобритании. Разброс значений индекса индивидуального влияния увеличился до 5.4 пункта — от 26.4 до 31.8. При этом максимальным относительным влиянием обладали депутаты-социалисты и социал-демократы из разных стран — Италии, Португалии, Бельгии и Франции (Н. Ларони, П. Карнити, Л. Марино, М. Галль и Ж-Ф. Ори).

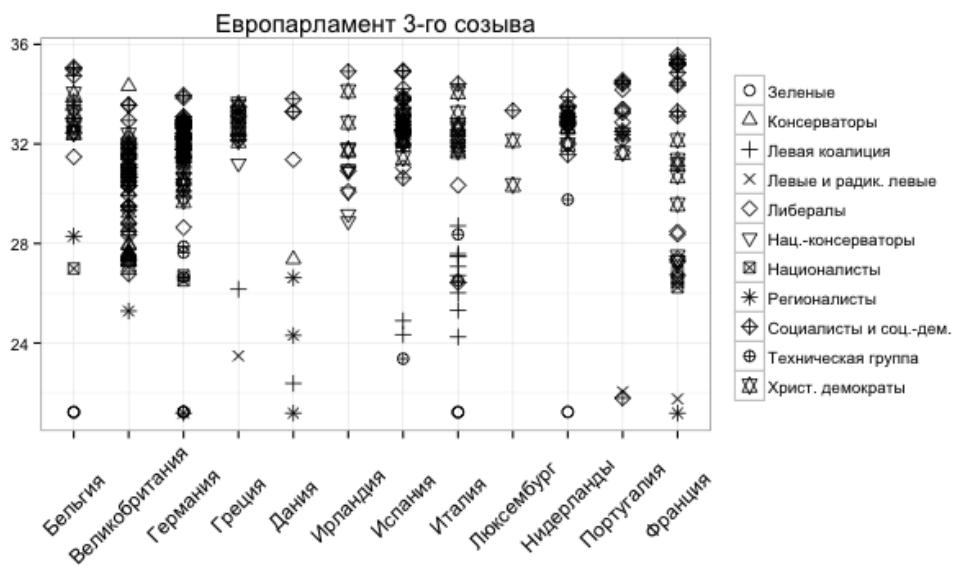


Рисунок 3.21 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС и европейским политическим группам для 3-го созыва Европарламента

В четвертый созыв Европарламента (1994 — 1999 гг.) вошли представители из 15 государств Европы: с 1995 года в состав Европейского сообщества вошли Австрия, Финляндия и Швеция. 432 депутата из 567 могут считаться активными.

Социал-демократы (PES) сохранили свое влияние на принятие решений, они также стали самой многочисленной группой. Более оформленной и многочисленной стала оппозиция, представленная христианскими демократами из Европейской народной партии (EPP) и Силы Европы (итальянская Forza Europa) и имевшая более 141 активного депутата. Наименее влиятельными в среднем стали евросkeptики и депутаты технической группы (NI). За исключением регионалистов с 9 активными депутатами, эти группы были самыми

малочисленными. Более заметную роль в этом созыве стала играть группа «Зеленых и регионалистов», и в ее работе принимали участие 19 депутатов.

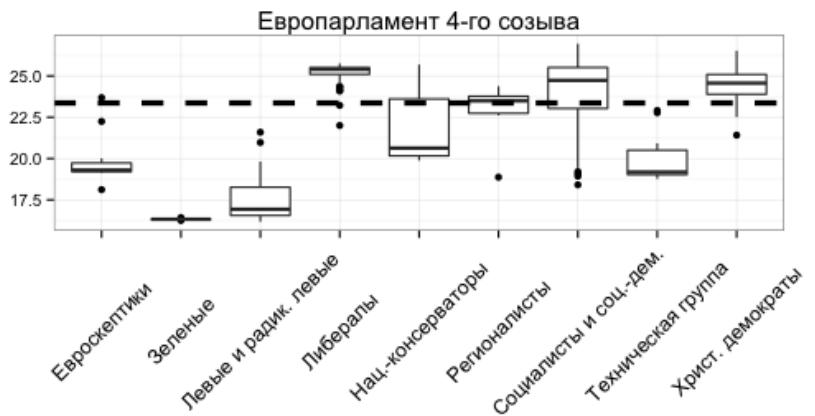


Рисунок 3.22 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по европейским политическим группам для 4-го созыва Европарламента

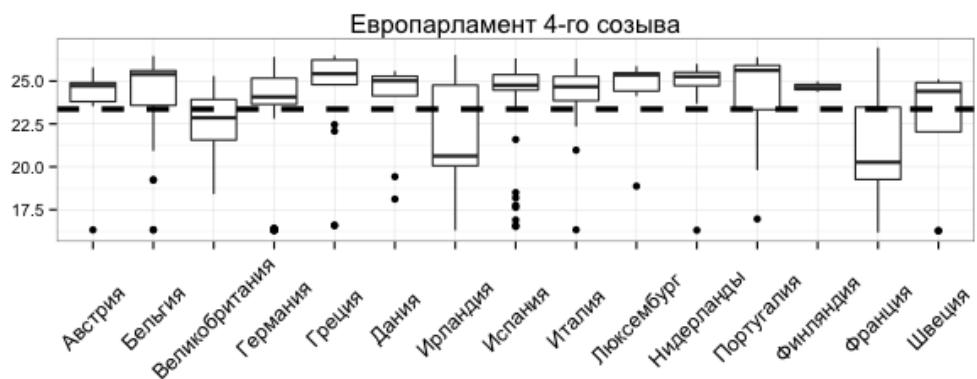


Рисунок 3.23 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС для 4-го созыва Европарламента

По национальному признаку в данном созыве парламента можно выделить британских депутатов, которые были в среднем достаточно влиятельными, это связано с тем, что из 63 депутата (60 из которых лейбористы) из 87 вошли в группу «Партию европейских социалистов» (PES). В среднем значительно более низким влиянием обладали депутаты от Франции и по-прежнему низкие значения индекса влияния были у Ирландии, что наглядно видно на рисунке 3.24. От этих стран в этот созыв прошло много консерваторов и евросkeptиков.

В среднем низкие значения индекса характерны также для депутатов из Дании (9 активных парламентариев) и Финляндии (3 активных парламентариев). Максимальные значения индекса индивидуального влияния вновь были у социалистов из разных стран — Германии, Италии, Нидерландов, Испании, Люксембурга и Великобритании.

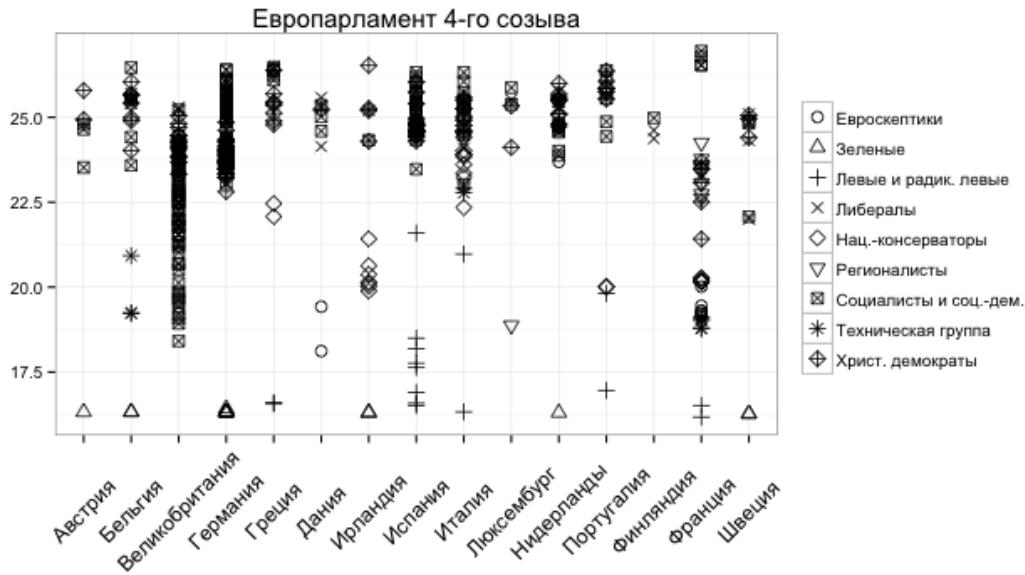


Рисунок 3.24 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС и европейским политическим группам для 4-го созыва Европарламента

*Пятый созыв* Европарламента (1999—2004 гг.) был сформирован из 626 депутатов. Если в первом созыве индексы влияния изменились от 43.09 до 50.85, то в пятом — от 20.85 до 25.39.

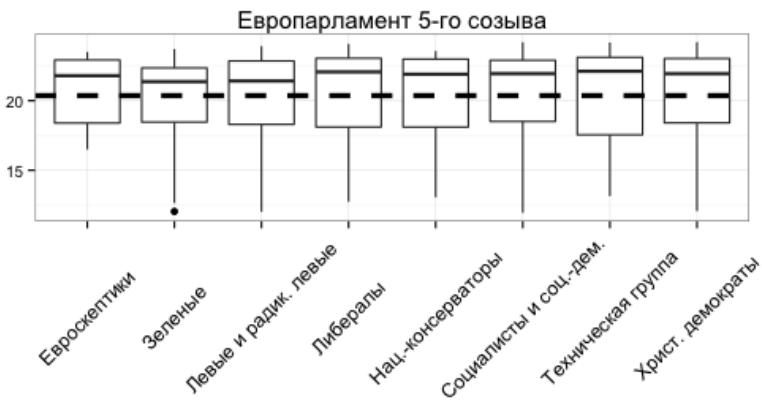


Рисунок 3.25 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по европейским политическим группам для 5-го созыва Европарламента

В этом созыве доминирующей по численности группой стала коалиция Европейской народной партии и Европейских демократов (EPP-ED) — 159 активных депутатов. Социал-демократы (PES), в свою очередь, составили достаточно сплочённую оппозицию. Среднее значение индекса индивидуального влияния у этих групп почти одинаково: 23.73 и 23.66 соответственно.

Достаточно влиятельными, помимо групп EPP-ED и PES, могли считаться зеленые и регионалисты G/EFA, наименее — малочисленные евроскептики EDD (7 активных депутатов) и национал-консерваторы UEN (20 активных депутатов). Небольшой по численности (21 депутатов) также была «техническая» группа NI & TGI.

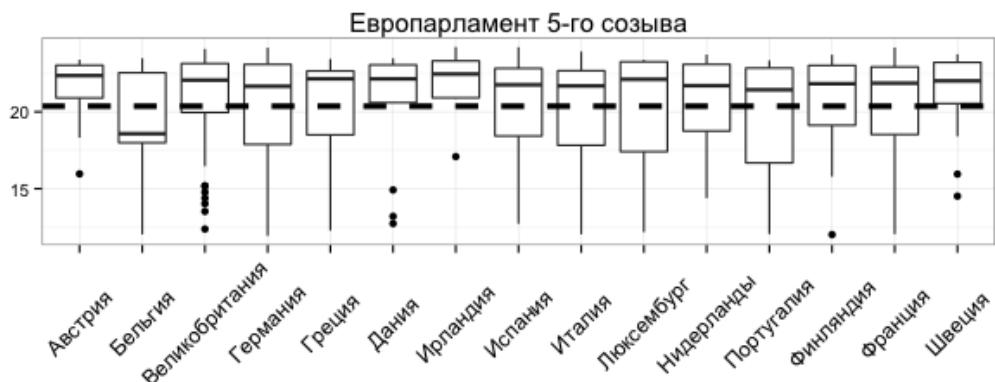


Рисунок 3.26 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС для 5-го созыва Европарламента

По национальному признаку малым влиянием обладали в основном депутаты от Люксембурга (6 активных депутатов) и Португалии (19 активных депутатов), а относительно более высоким — Швеции, Финляндии и Греции.

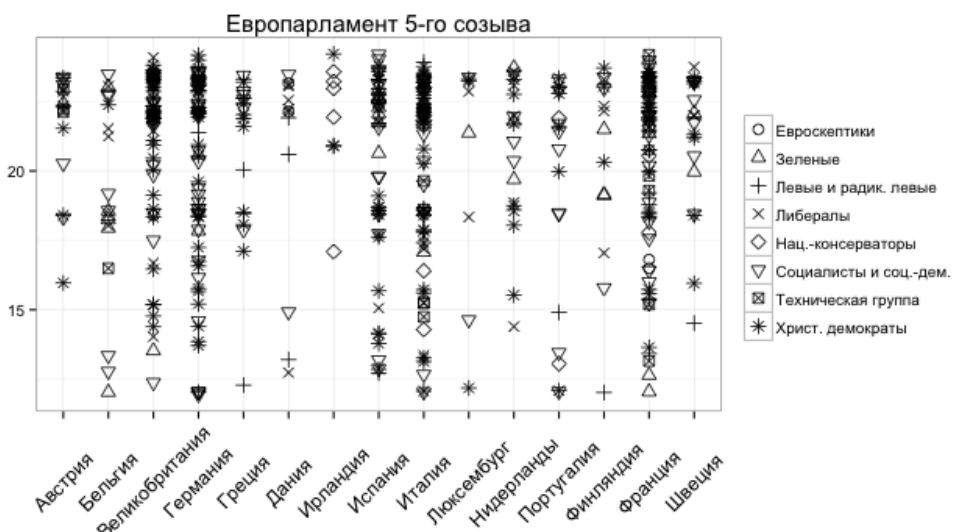


Рисунок 3.27 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС и европейским политическим группам для 5-го созыва Европарламента

Максимальные значения индекса индивидуального влияния в этом созыве имели представители разных политических сил: социал-демократы К. Хаш (Социал-демократическая партия Германии) и А. Балтас (Всегреческое социалистическое движение), И. Шорлинг (шведская «Партия зелёных»), национал-консерватор Ж.-Ш. Маршиани («Движение за Францию»), либерал Л. Кавери (итальянский «Союз Вальдостан»), христианский демократ Б. Пронк («Христианско-демократический призыв»).

*Шестой созыв* Европарламента (2004—2009 гг.) был сформирован из 732 депутатов. Но в 2007 году произошло расширение ЕС — в Союз вошли Болгария и Румыния, — и в Европейском парламенте появились еще 53 депутата.

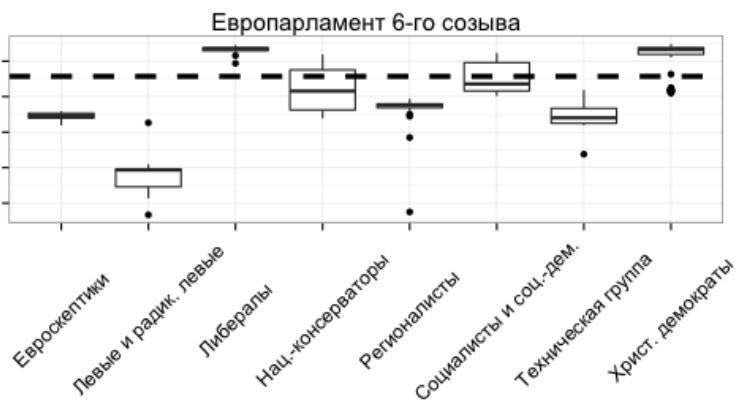


Рисунок 3.28 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по европейским политическим группам для 6-го созыва Европарламента

Стоит отметить, что разделение на относительно более и менее влиятельных депутатов в этом созыве Европарламента стало более выраженным, чем в предыдущем. В шестом созыве доминирующей по численности группой стала Европейская народная партия (EPP-ED) — у нее было 36% мест и 136 активных депутатов. Второй по численности и влиянию стала Прогрессивный альянс Социалистов и демократов (PES), у них число активных депутатов составило 94 человека. Среднее влияния депутатов из этих двух групп, а также группы либералов превышало среднее значение 27.5. Третьей группой с наиболее влиятельными депутатами был Альянс либералов и демократов. В свою очередь, наименьшим влиянием, а также небольшой численностью (17 активных депутатов) обладали левые и радикальные левые парламентарии. Четвертой по численности была группа «Зеленых/Регионалистов», и малое влияние, которым обладали депутаты, вероятно, было обусловлено низкой дисциплиной: из 33 депутатов только 8 соответствовали критерию «активных».

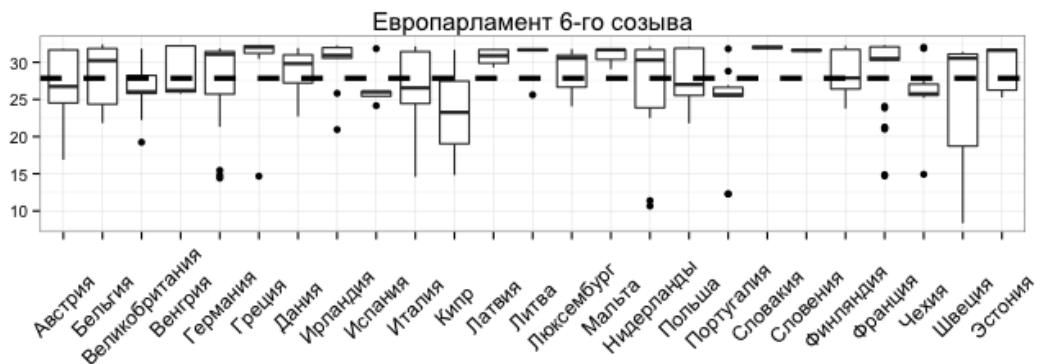


Рисунок 3.29 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС для 6-го созыва Европарламента

По национальному признаку, в среднем более высоким влиянием традиционно обладали парламентарии из Германии, Люксембурга. Недавно вступившие в ЕС страны — Венгрия, Латвия, Литва, Словакия, Словения, Эстония — тоже имели влиятельных депутатов более влиятельных, чем средний член парламента, однако число активных парламентариев у них не превышало шести. Наоборот, менее влиятельными в среднем были депутаты из Великобритании, Чехии, Швеции и Кипра (у последнего всего два активных депутата).

Для Великобритании типична большая доля членов Европарламента, состоящих в группе независимых (технической) и группе зеленых/регионалистов. В Чехии, несмотря на большую долю христианских демократов, сильны позиции крайних левых, относительно маргинальной группы в Европарламенте. Депутаты, обладающие самым слабым влиянием, баллотировались в Швеции: радикальная левая Э-Б. Свенссон, «зеленый» К. Шлитер и евроскептик Н. Лундгрен. На рисунке 3.30. видно, что во всех странах наиболее авторитетными являются христианские демократы, и особенно сильны их позиции в Германии, Польше, Бельгии, Франции.

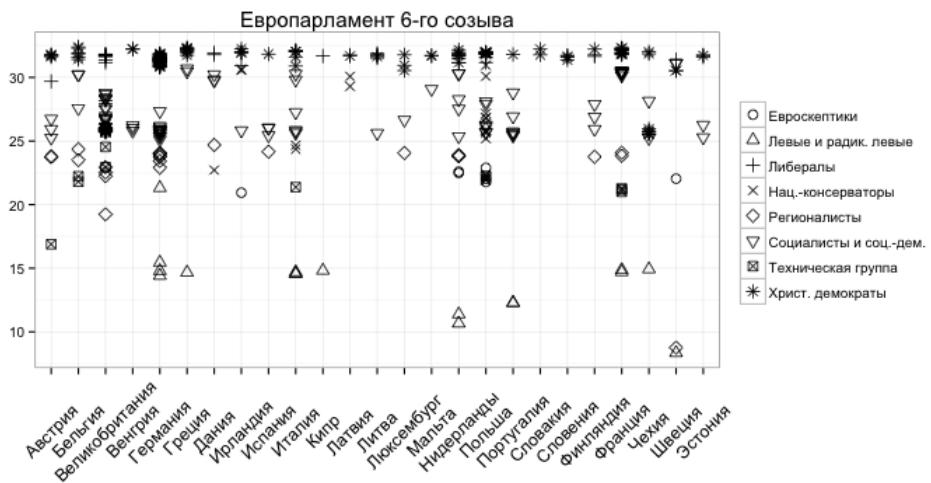


Рисунок 3.30 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС и европейским политическим группам для 6-го созыва Европарламента

*Седьмой созыв* Европарламента (2009—2014 гг.) был сформирован из 736 депутатов. Доминирующие позиции сохранили Европейская народная партия (EPP), Прогрессивный альянс Социалистов и демократов и Альянс либералов и демократов. Четвертыми по численности стали Зеленые. Такой порядок отражен и в распределении влияния. В седьмом созыве не были представлены национал-консерваторы, в отличие от шестого созыва, а в остальном распределение влияния во многом сохранилось в прежнем виде. В свою очередь, наименьшее влияние было характерно для радикальных левых, регионалистов и некоторых независимых депутатов.

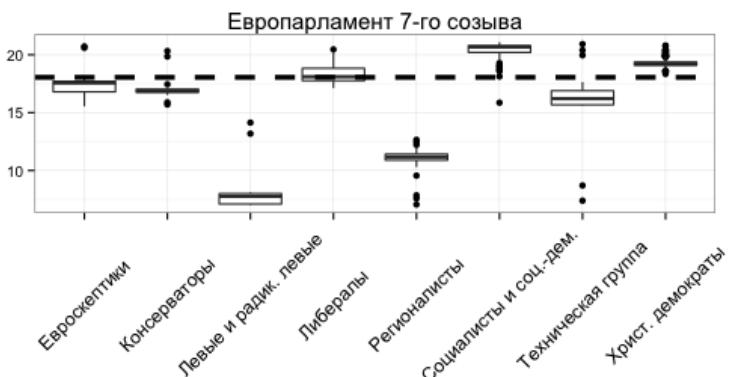


Рисунок 3.31 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по европейским политическим группам для 7-го созыва Европарламента

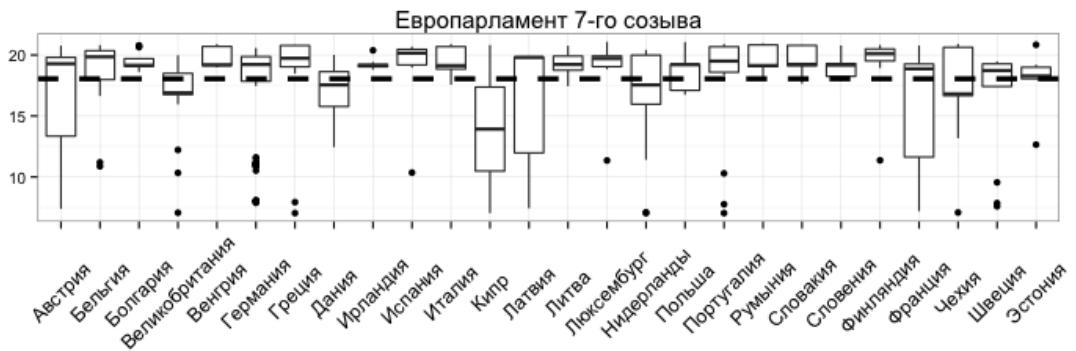


Рисунок 3.32 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС для 7-го созыва Европарламента

По национальному признаку в седьмом созыве, ввиду большого количества стран-членов ЕС и небольших различий в распределении мест и предпочтений, сохранился примерно такой же паттерн, как и в предыдущем. По-прежнему менее влиятельными в среднем были депутаты из Великобритании, Чехии и Кипра, а также Швеции и Дании. Более влиятельными стали депутаты из Италии, Бельгии, и наоборот, снизилось влияние парламентариев из Франции, Нидерландов и Латвии.

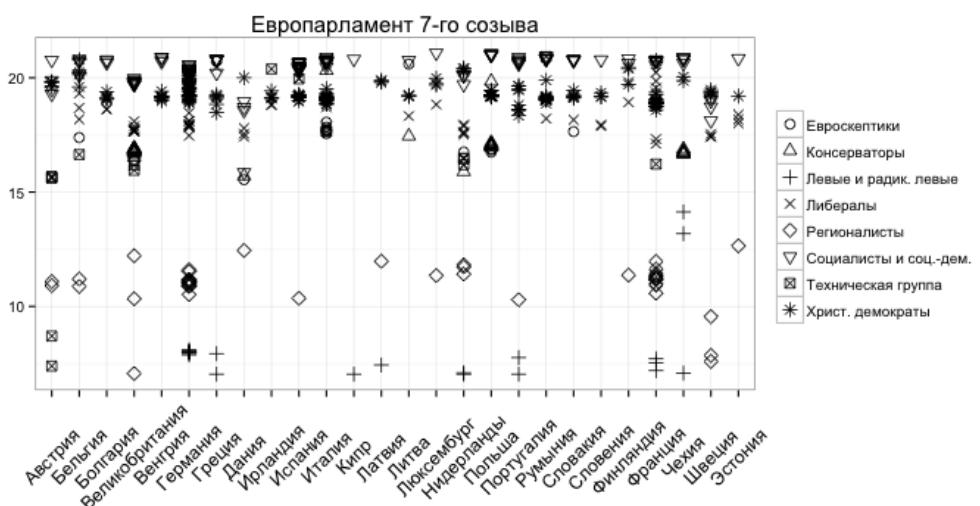


Рисунок 3.33 – Распределение значений индекса индивидуального влияния  $\alpha^1$  по странам-членам ЕС и европейским политическим группам для 7-го созыва Европарламента

На рисунке 3.33 видно, что христианские демократы, как и в шестом созыве, обладают наибольшим влиянием, но стоит отметить, что для некоторых

государств характерны достаточно сильные и многочисленные группы оппозиции. Например, в седьмом созыве наименьшим влиянием обладали депутаты-консерваторы из Великобритании. Европарламентарии из Соединенного Королевства также вошли в группы евросkeptиков и независимых (NI). В Германии и Франции характерны радикальные оппозиционные группы зеленых и крайних левых, в Италии многочисленна группа евросkeptиков, в Чехии — социал-демократов и консерваторов.

Наибольшим влиянием в этом созыве обладал венгерский депутат З. Баг, член группы христианских демократов. Наименьшим — евросkeptики Д.Р. Кларк и Дж.С. Агню, крайний правый А. Бронс.

### 3.2.3.2 Влияние политических групп

Сравним результаты, полученные на уровне отдельных депутатов, с оценками распределения влияния на уровне устойчивых фракций — политических групп Европейского парламента.

Для задания матрицы предпочтений по созданию коалиций используем данные о позициях депутатов на лево-правой шкале и про- и анти-Европейской. Для каждой политической группы была рассчитана средняя позиция депутата в каждой из осей, и для попарно всех групп (в рамках одного созыва) были рассчитаны расстояния — меры дальности  $\delta_{ij}$ . После чего, все элементов матрицы были преобразованы так, чтобы новые величины отражали идеологическую близость групп друг к другу, т.е. готовность вступать в коалицию:

$$p_{ij} = 1 - \delta_{ij}.$$

Распределение мест в Европейском парламенте между политическими группами и матрицы предпочтений для созывов приведены в Таблицах Г.4 – Г.11 в Приложении Г. Распределение мест между странами-членами ЕС и соответствующие матрицы предпочтений приведены в Таблицах Г.12 – Г.19 Приложении Г.

Из Таблицы Г.20 Приложении Г видно, что по индексу  $\alpha_{gr}^1$  в *первом созыве* Парламента наибольшим относительным весом обладали депутаты, состоявшие в группе христианских демократов. Заметно отличается от нуля влияние либералов и консерваторов, а также социал-демократов. Заметим, что согласно  $\alpha_{gr}^1$ , социал-демократы являются второй по влиятельности группой, при том, что индексы индивидуального влияния умеренных левых депутатов в среднем были относительно низкими. Такое различие может быть индикатором того, что группа социал-демократов имела взгляды, слабо совпадающие с мнением большинства, но величина группы делала ее достаточно весомой силой. Небольшие группы, наоборот, при малом значении  $\alpha_{gr}^1$  показывали в среднем относительно более высокие значения  $\alpha^1$  — как самостоятельная сила в Парламенте эти группы не были влиятельными, но их депутаты часто голосовали согласованно с депутатами из других, более многочисленных групп. С другой стороны, по  $\alpha_{gr}^1$  теряют социал-демократы и консерваторы: при достаточно большой численности групп, они реже были поддержаны прочими депутатами. Значения индекса влияния  $\beta$  для них выше, чем  $\alpha_{gr}^1$ , учитывающем предпочтения.

Похожий паттерн связи между значениями  $\alpha_{gr}^1$  и  $\beta$  характерен и для *второго созыва*. Тем не менее, влияние социал-демократов в Европейском парламенте в конце 1980-х гг. возросло, это демонстрируют индексы влияния групп и среднее значение индекса индивидуального влияния. Меньшим стало влияние христианских демократов, либералов и консерваторов. По итогам выборов выросла численность национал-консерваторов, влияние группы «Европейский демократический альянс», но в ходе работы парламента эти депутаты были скорее в меньшинстве. Во второй созыв прошли радикальные «Европейские правые», но представляли собой маргинальную силу, хотя и третьем созыве им удалось сформировать общеевропейскую группу.

В *третьем созыве* Европарламента доминирующей силой также были социал-демократы, что отражено в трех рассматриваемых подходах. Другие

левые силы на этом фоне тоже смогли увеличить влияние своих групп и отдельных депутатов. Либеральная группа, имея всего 9% мест, также смогла увеличить свое влияние за счет готовности сотрудничать с другими умеренными группами, например, христианскими демократами.

Обе группы имеют достаточно высокое среднее значение индекса индивидуального влияния. Специфичность взглядов националистов и зеленых, в совокупности с небольшой долей мест, отражается в низких значениях  $\alpha^1$ .

С середины 90-х гг., в *следующем созыве* Парламента, левые идеи становятся менее популярными, вес набирают либералы. У социал-демократов почти вдвое падает  $\alpha_{gr}^1$ , при том, что среднее влияние депутатов, измеренное как  $\alpha^1$ , осталось самым высоким. В четвертом созыве Парламента усиливаются национал-консерваторы и силы, выступающие за регионализацию, парламентские мандаты впервые получают евроскептики, при этом для них характерно  $\alpha_{gr}^1$  большее, чем  $\beta$ , при средних значениях индивидуальных индексов влияния.

*Пятый созыв* характеризуется схожим паттерном распределения влияния: влияние теряет группа умеренных социал-демократов и радикальные левые. В этот созыв также не прошли националисты, но все большим влиянием пользуются зелёные и регионалисты. Тем не менее, как основную силу себя заявили христианские демократы, это отражается в росте  $\alpha_{gr}^1$  и  $\beta$ , и в относительно высоком среднем значении индекса индивидуального влияния.

С 2004 года, в *шестом созыве* Европарламента, когда ЕС расширился до 25 стран, христианские демократы сохранили свое доминирующее положение, что демонстрируют индексы  $\alpha_{gr}^1$  и  $\beta$ , а также увеличилось влияние либеральных демократов. Националистические группы и регионалисты, вместе с «Зелеными», стали менее влиятельными (примерно на треть, по оценкам обоих индексов), но евроскептики, в свою очередь, смогли получить больше мест, отражая тем самым рост анти-интеграционных настроений, связанных с расширением Евросоюза, в старых и новых странах ЕС. «Техническая груп-

па» прекратила свое существование, уступив место группе «Независимых» депутатов.

В распределении влияния в *седьмом созыве* Европарламента произошли незначительные изменения. Немного ослабли позиции христианские демократов, но нарастили свое влияние либералы и социал-демократы. В парламент прошла небольшая группа консерваторов (преимущественно британских), но ее влияние практически нулевым, как у группы евроскептиков и независимых депутатов. Снизилось влияние радикальных групп: крайних левых и зеленых, но заметим, что «Зеленые» стали четвертой по численности партией в Европарламенте, обладая 55 местами (7.5%).

### 3.2.3.3 Влияние стран-членов ЕС

Относительно влияния стран-членов ЕС по индексу заметно  $\alpha_{gr}^1$  доминирующее положение Германии и Италии, с небольшим отставанием Франции и Великобритании. С третьего созыва в силу большой численности населения значительное влияние, сопоставимое с влиянием Соединённого Королевства, имела Испания. При этом, все страны уменьшали свое влияние, но быстрее его теряли немногочисленные группы представителей Бельгии, Дании, Ирландии, Нидерланды — 50% и более в седьмом созыве. Тогда как «потери» влияния Германии, Италии и Франции составляли менее 45%, а Великобритания, в свою очередь, даже увеличила свое влияние по индексу  $\alpha_{gr}^1$  в силу больше готовности ее депутатов сотрудничать с другими. Но расширение ЕС до 25 стран-членов отразилось на всех «старых» государства-членах: по 0.03-0.04 пункта потеряли Италия, Франция, Великобритания и Испания, по 0.01-0.02 пункта — менее населенные государства и Германия. При этом, вплоть до седьмого созыва Германии удалось сохранить максимальное влияние по индексу  $\alpha_{gr}^1$  — от 0.20 до 0.15.

Новые государства-члены получали достаточно ограниченное влияние, не превосходящее 0.03, как в 1994–1999 гг. с вхождением Австрии, Финляндии,

дии и Швеции, так и 2004–2009 гг. с вхождением стран Центральной и Восточной Европы, Кипра и Мальты. Исключение составляет Польша — ее индекс влияния составил 0,07 в шестом созыве и 0,06 в седьмом. Так, ее относительное влияние сопоставимо с влиянием Испании.

Наименьшее значение влияния в первых созывах было характерно для Люксембурга (например, 0,05 в первом). С шестого созыва компании ему составили Кипр, Латвия, Мальта, Словения и Эстония — государства, имевшие менее десятка мест в Парламенте. Их влияние, вне зависимости от предпочтений по созданию коалиций, не превосходило 0,01.

Значения влияния, рассчитанные по классическому индексу Банцафа  $\beta$ , мало отличаются от индекса, учитывающего предпочтения по созданию коалиций, и задают примерно тоже упорядочение стран по относительному влиянию. До седьмого созыва ЕП значения  $\alpha_{gr}^1$  крупных групп депутатов от Германии, Италии и Франции на 0.01-0.03 превосходили значения  $\beta$  (или во-все были равными), и с расширением ЕС различия становились все меньше. Например, для седьмого созыва коэффициент корреляции достиг 0.98. Стоит также отметить, что как раз для Великобритании характерны более низкие значения индекса индекса  $\alpha_{gr}^1$  по сравнению с  $\beta$ : 0,11 против 0,17 в первом созыве и 0,09 против 0,10 в седьмом созыве. Любопытно, что только Великобритании вплоть до третьего созыва удавалось увеличивать свое влияние по  $\alpha_{gr}^1$ , несмотря на то, что в состав Европарламента входили депутаты от новых государств-членов.

### 3.2.4 Некоторые выводы

За последние более чем 30 лет полномочия и влияние Европейского парламента существенно расширились, а наличие большого количества информации о его работе в открытом доступе делает возможным детальное изучение институтов Европарламента, национальных и европейских партий, вплоть до поведения отдельных парламентариев.

Полученные результаты позволяют заключить, что в семи созывах Европейского парламента (с 1979 по 2014 г.) относительно более высоким влиянием обладали депутаты из группы либералов: причиной тому является их «срединное» положение и возможность вступать в коалиции как с христианскими демократами, так и с социалистами и социал-демократами. Именно эти две коалиции составляли наиболее многочисленные группы депутатов Европарламента — их суммарный удельный вес никогда не был меньше 50%. В то время как доля мест либералов до шестого созыва не превышала 10%, только у группы либералов (и группы независимых) значение индекса влияния, учитывающего предпочтения, всегда превышало значение индекса влияния Банцафа — что демонстрирует идеологическую «гибкость» группы и готовность вступать в коалиции.

Сkeptически настроенные к объединенной Европе депутаты не являются доминирующей силой, что отражается в том, что среднее индивидуальное влияние леворадикальных депутатов, регионалистов и евроскептиков, ниже среднего по созыву. Однако их идеологическая сплоченность позволила группам быть не менее влиятельными, чем прогнозирует индекс Банцафа.

Основным структурирующим расколом в Европарламенте является идеологический, что неоднократно было отмечено в литературе, но для ряда стран характерно то, что депутаты от них традиционно вступают в те или иные политические группы. Так, стоит отметить, что систематически более низким влиянием обладают парламентарии-регионалисты из Великобритании и Швеции, радикальные левые из Франции, Испании, Италии, Греции, британские консерваторы, национал-консерваторы из Польши, Ирландии. В целом же, снижение влияния по национальному признаку заметно по всем используемым нами индексам — закономерное следствие увеличения численности парламента.

Результаты исследований этого подраздела отражены в работах [166, 167].

### 3.3 Анализ структуры Верховной Рады Украины в 2007–2012 гг.

Результаты голосований депутатов являются крайне ценным материалом для изучения законодательных органов. В частности, имеющиеся в открытом доступе в электронном виде результаты поименных голосований депутатов Верховной Рады (ВР) Украины, позволили значительно продвинуться в анализе структуры украинского парламента, начатом в [168].

Как правило, для оценки влияния членов законодательных органов используют так называемые индексы влияния. Большинство из них строятся в предположении, что все коалиции в парламенте равновероятны, т.е. участники не имеют предпочтений в создании коалиций. Тем не менее, часто это не соответствует действительности, и получаемые с помощью таких индексов результаты не согласуются с реальным положением дел. Однако, имеются инструменты, позволяющие оценить влияние фракций, учитывая при этом вероятности создания тех или иных коалиций. Для использования таких инструментов необходимо некоторым образом оценить предпочтения фракций в образовании коалиций.

Информация о поименных голосованиях депутатов дает возможность непредвзято определить занимаемые членами парламента позиции по тем или иным вопросам, оценить согласованность этих позиций, а значит – и предпочтения в формировании коалиций. Таким образом, становится возможным использовать более тонкие инструменты анализа распределения влияния, что и было проделано в данной работе.

Построение политических карт и вычисление индексов согласованности позволило глубже понять внутреннюю структуру ВР Украины 6-го созыва, и, в конечном счете, получить реалистичную картину распределения влияния между фракциями.

Первичный анализ распределения влияния в ВР Украины 6-го созыва проведен с помощью расчёта долей мест, занимаемых фракциями в парламенте, а также значений индекса Банцафа [163]. На данном этапе анализа предпочтения фракций в создании коалиций не учитывался. Отметим, что

индекс Банцафа является одним из классических инструментов изучения распределения влияния участников в органах коллективного принятия решений и широко используется в исследованиях, посвященных влиянию в законодательных органах [169 – 171].

Индекс Банцафа для некоторой фракции  $i$  есть доля выигрывающих коалиций, в которых данная фракция является ключевой (т.е. без неё у этой коалиции недостаточно голосов для принятия решения), среди всех возможных коалиций. Под выигрывающей коалицией понимается такое подмножество фракций среди всего множества фракций в парламенте, что суммарное количество их голосов превосходит некоторую заранее заданную квоту, что ведет к принятию решения.

Формально индекс Банцафа  $\beta(i)$  для фракции  $i$  представляется как дробь:

$$\beta(i) = \frac{b_i}{\sum_{j=1}^k b_j},$$

где  $b_i$  – число коалиций, в которых данная фракция  $i$  является ключевой,  $k$  – количество всех фракций в парламенте.

Динамика влияния фракций ВР Украины 6-го созыва, измеренного с помощью доли мест фракций и индекса Банцафа представлена на рисунках 3.34 и 3.35 соответственно.

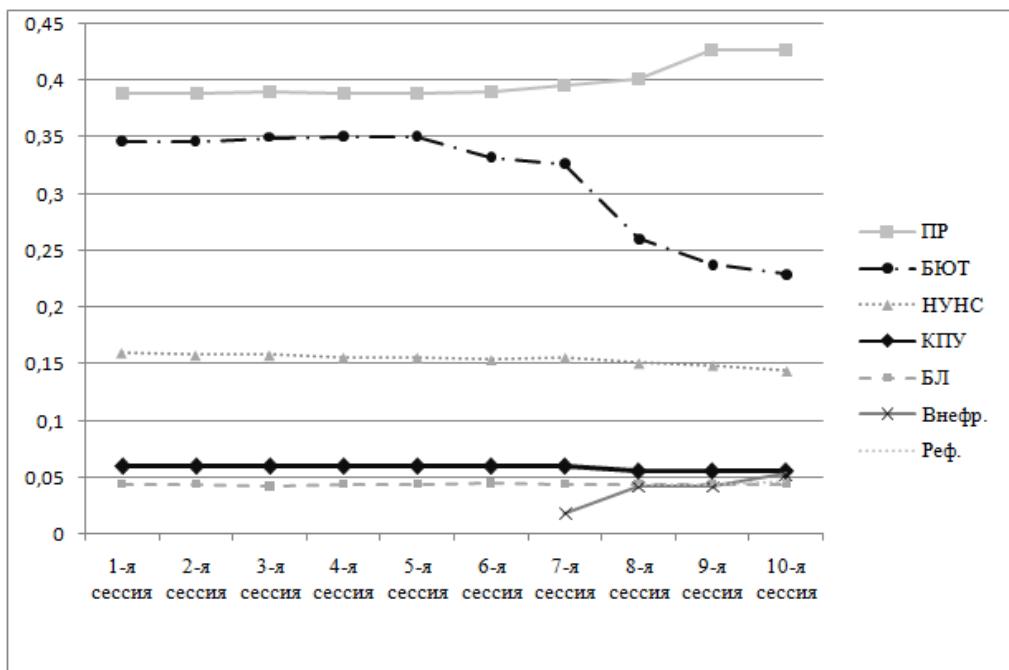


Рисунок 3.34 – Динамика долей мест фракций в ВР Украины 6-го созыва

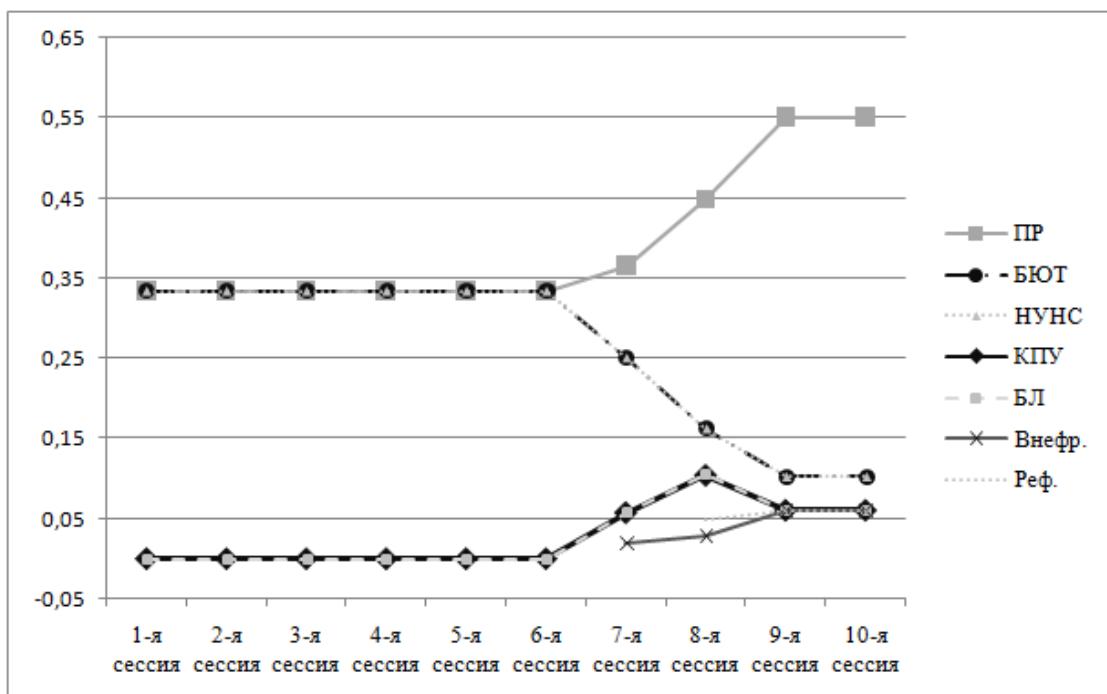


Рисунок 3.35 – Динамика значений индекса Банцафа фракций в ВР Украины 6-го созыва

Индекс Банцафа рассчитывался в предположении, что все коалиции равновероятные. Однако, если парламент, например, имеет непримиримых противников, которые придерживаются противоположных взглядов по важным вопросам, это допущение очень сильно искажает картину реального распределения сил.

деления влияния в законодательном органе. Таким образом, исследователи сталкиваются с необходимостью учесть предпочтения фракций в создании коалиций при оценке их влияния.

Предложенный Ф. Алескеровым в [164]  $\alpha$ -индекс позволяет решить эту непростую задачу. Для вычисления -индекса в первую очередь необходимо каким-то образом измерить предпочтения фракций в формировании коалиций и представить их в виде матрицы. Наименее субъективный подход при решении этой задачи предполагает использование результатов поименных голосований депутатов. В данной работе были использованы результаты более тысячи поименных голосований за весь период работы ВР Украины 6-го созыва, на основе которых была измерена согласованность позиций фракций и, следовательно, коалиционных предпочтений фракций.

Среди всех голосований были выделены наиболее информативные. Для этого был рассчитан показатель, названный в [172] показателем *фракционного раскола*:

$$r_{frac} = \max_{i=1,n} P_i^+ - \min_{i=1,n} P_i^+,$$

где  $P_i^+$  – доля голосов «за», отданных фракцией  $i$  в данном голосовании,  $n$  – количество фракций в парламенте. В данной работе в качестве порогового значения данного показателя было выбрано значение 0,6: все голосования с фракционным расколом, превышающим 0,6 были признаны информативными. В результате было отобрано 1063 информативных голосования, которые и были использованы для дальнейшего анализа.

Далее депутаты были представлены в виде точек метрического пространства с помощью метода неметрического многомерного шкалирования. Каждому депутату сначала была сопоставлена точка с 1063, принимающими значения -1, 0 или 1 в зависимости от того, как он проголосовал. Далее были построены политические карты – отображения множества депутатов в пространство малой размерности, – при которых взаимное расположение точек, соответствующих депутатам, отражало бы частоту совпадений их позиций в

процессе голосований. Для этого были использован алгоритм, предложенный в [129]. На рисунке 3.36 представлена промежуточная двумерная конфигурация депутатов-точек во время 2-й сессии, полученная с помощью неметрического многомерного шкалирования (НМШ).

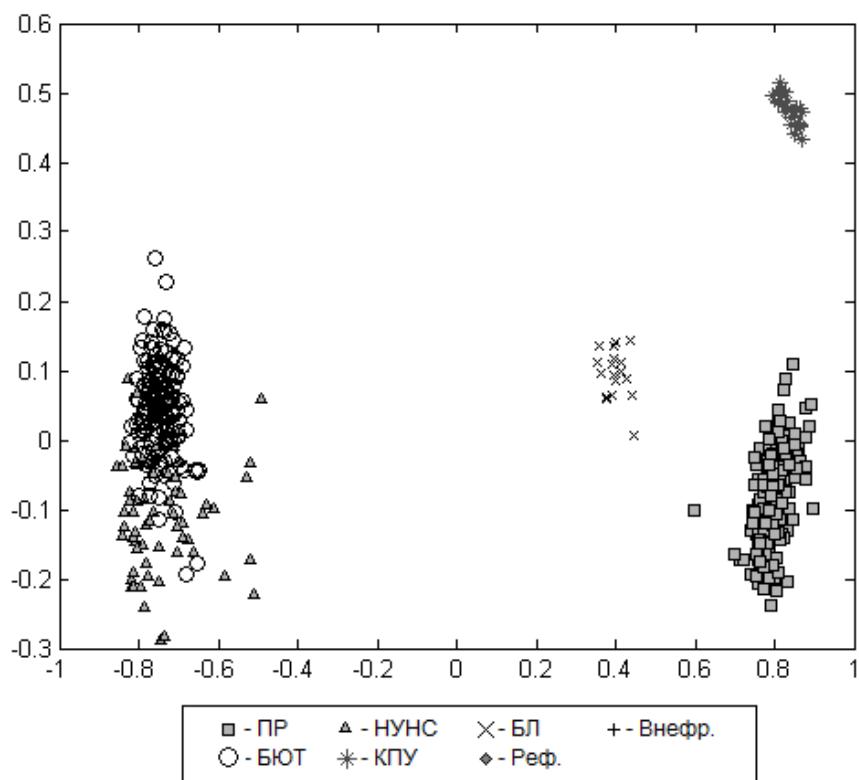


Рисунок 3.36 – Двумерная конфигурация точек-депутатов во 2-й сессии ВР Украины 6-го созыва

Выяснилось, что для описания позиций депутатов достаточно одного измерения. Т.е. в украинском парламенте имелся один самый важный фактор, по которому и происходило политическое размежевание депутатов на протяжении всего созыва.

На основании позиций депутатов на политических картах построим позиции фракций – среднее арифметическое позиций их участников. Результаты построения приведены на рисунке 3.37.

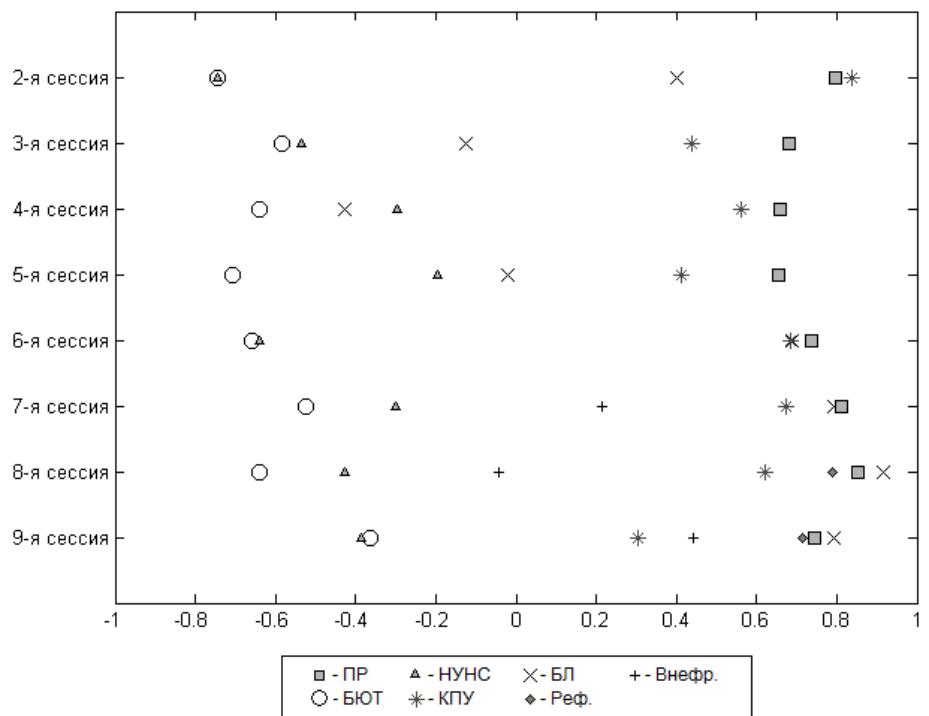


Рисунок 3.37 – Позиции фракций ВР Украины 6-го созыва со 2-й по 10-ю сессии на политических картах

В ВР Украины 6-го созыва в качестве латентного фактора политического размежевания разумным было бы считать степень лояльности к «оранжевым» силам, которая, в свою очередь подразумевает приверженность соответствующей культурно-этнической и внешнеполитической ориентации и лояльность к лидерам «оранжевых» сил.

К вычисленным координатам фракций было применено линейное преобразование  $L(x) = \frac{1}{2}(x + 1)$  и вычислен индекс согласованности  $\tau$ . На рисунке 3.38 отражена динамика согласованности наиболее крупных фракций в ВР Украины 6-го созыва.

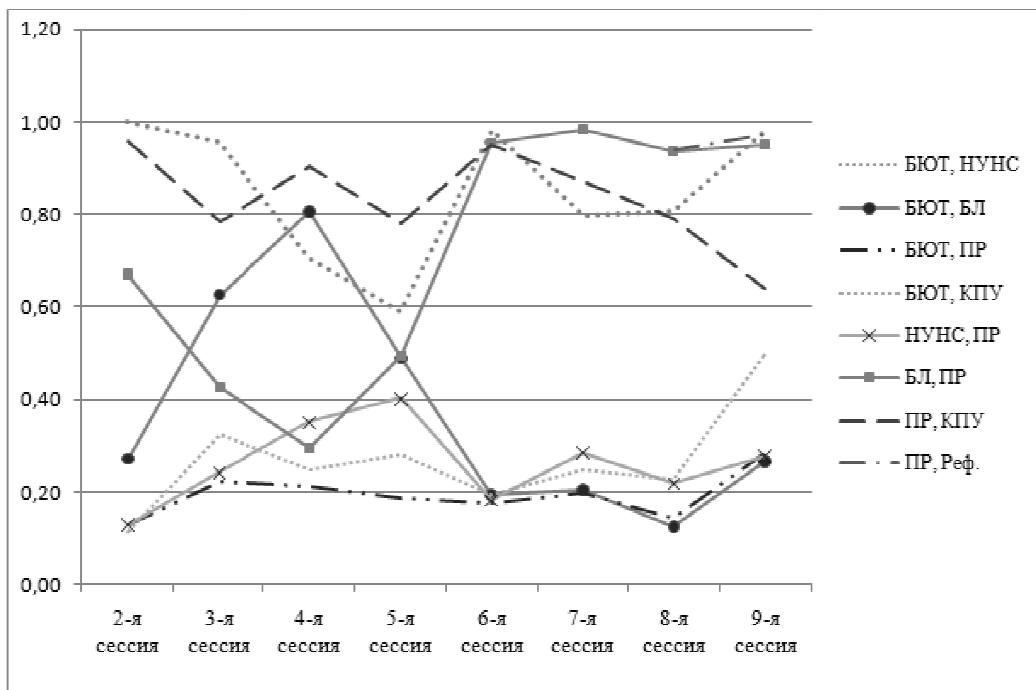


Рисунок 3.38 – Динамика индекса согласованности  $\tau$  наиболее крупных фракций в ВР 6-го созыва

Предложенные в [164]  $\alpha$ -индексы являются модификацией индекса Банцафа и позволяют учесть предпочтения игроков в формировании коалиций. Из доступных разновидностей  $\alpha$ -индексов в данной работе для анализа был выбран *кардиальный  $\alpha$ -индекс*.

Кардиальный  $\alpha$ -индекс определяется следующим образом:

$$\alpha^{card}(i) = \frac{\chi_i}{\sum_j \chi_j}, \text{ где } \chi_i = \sum_{\substack{\omega-\text{выигрывающая,} \\ i-\text{ключевая в } \omega}} f(i, \omega),$$

$$f(i, \omega) = \frac{\sum_{j \in \omega} p_{ij}}{|\omega|}.$$

Для построения кардиального  $\alpha$ -индекса в качестве значения интенсивности связи между фракциями был использован индекс согласованности  $\tau$ , построенный на основании координат фракций  $i, j$  на политической карте и расстояния  $d_{ij}$  между ними,  $p_{ij} = \tau(i, j) = \frac{2}{1+d_{ij}} - 1$ .

На рисунке 3.39 представлены результаты вычислений  $\alpha$ -индекса.

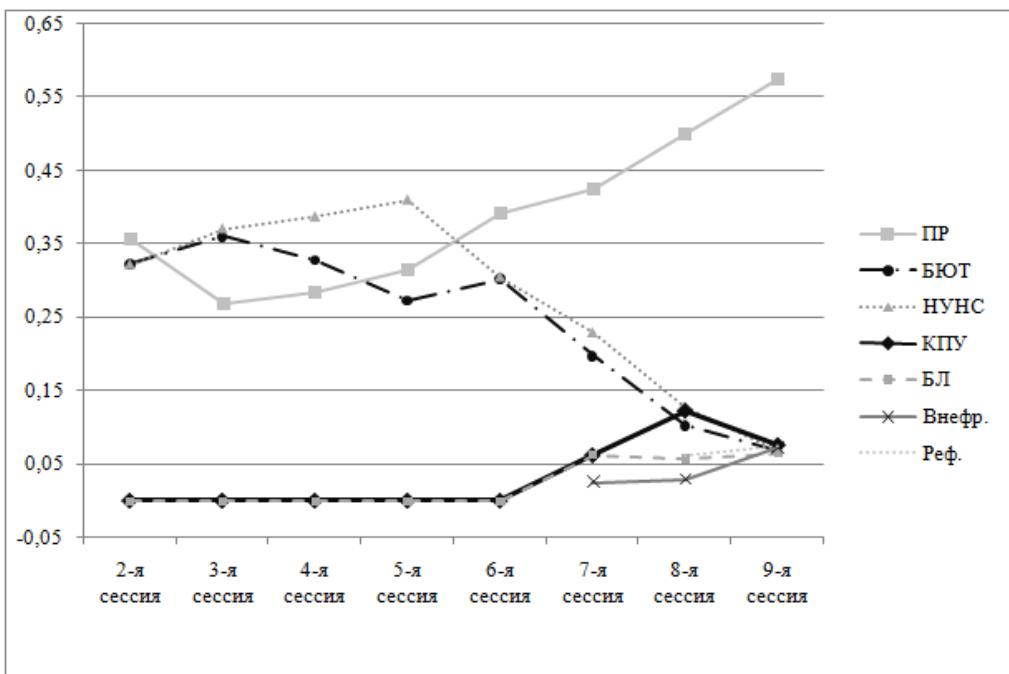


Рисунок 3.39 – Динамика значений а-индекса фракций в ВР Украины 6-го созыва

Согласно полученным результатам, в ВР Украины 6-го созыва, сложилась такая конфигурация политических сил, при которой даже небольшая фракция, занимающая удачную позицию, могла иметь сильное влияние в парламенте – стать обладателем «золотой акции» – тех самых нескольких голосов, которые могут предрешить результат голосований.

В работе был проведен анализ структуры Верховной Рады Украины 6-го созыва (2007–2012 гг.) на основании результатов более чем тысячи поименных голосований депутатов. Были рассчитаны значения индекса Банцафа для простого большинства. Было показано, что доля мест, занимаемых фракцией, может значительно отличаться от ее влияния, измеренного с помощью индекса Банцафа. Была проведена оценка согласованности позиций фракций с помощью различных индексов. Были построены одномерные политические карты Верховной Рады Украины 6-го созыва. Были рассчитаны значения учитывавшего коалиционные предпочтения фракций а-индекса, построенного с использованием различных индексов согласованности. Было показано, что при измерении влияния с помощью а-индекса небольшие фракции, зани-

мающие выгодные позиции, могут иметь большее влияние, чем крупные фракции с крайними позициями.

Результаты исследования этого подраздела отражены в работе [173].

### 3.4 Исследование конфликтности и точности рекомендаций российских финансовых аналитиков

Финансовый и, в частности, фондовый рынок – динамически развивающийся сегмент экономики. Кроме непосредственных участников фондового рынка, которые совершают сделки, существует институт финансовых аналитиков, которые анализируют рынок и стараются оценить будущую цену бумаг, дать прогноз по их стоимости и рекомендации участникам рынка (покупать, держать или продавать).

В данном подразделе приведены результаты исследования качества прогнозов и рекомендаций финансовых аналитиков на российском фондовом рынке. Под качеством прогнозов аналитиков будем понимать как точность, так и согласованность (несогласованность, конфликтность) этих прогнозов. Кроме того, проанализированы временной и отраслевой аспекты рекомендаций, структурные особенности рекомендаций, связь с волатильностью рынка и др.

#### 3.4.1 Обзор литературы

В экономической литературе большое число работ посвящено анализу прогнозов и рекомендаций зарубежных аналитиков (см. например, [174 – 178]).

Большой обзор современной (до 2010 года) литературы по исследованиям, связанным с прогнозами финансовых аналитиков можно найти в [174]. В этой работе также выявлена роль финансовых аналитиков в распределении ресурсов на рынке ценных бумаг, а именно оценено влияние прогнозов финансовых аналитиков на инвесторов и реакция рынка на эти

прогнозы. Кроме того, в работе выделен ряд возможных актуальных направлений исследований:

- анализ процессов принятия решений аналитиками;
- исследование различий между ролью аналитиков, которые являются интерпретаторами общей информации, и теми, кто вносит частную информацию, в определение цены акций;
- исследование эвристических правил в процессе создания цены;
- исследования взаимосвязи между экономическими стимулами аналитиков и их разногласиями;
- исследования факторов, которые влияют на точность прогнозов финансовых аналитиков;
- исследование долгосрочных прогнозов и рекомендаций;
- анализ влияния прогнозов прибыльности и текущих доходностей на размещение ресурсов на рынке ценных бумаг.

В [175] исследована взаимосвязь между известностью аналитиков и реакцией инвесторов на скорректированный прогноз. Например, при опубликовании положительного прогноза о ценах определенного актива на фондовом рынке инвесторы могут начать покупку данного вида актива. В работе доказана значимость статистической гипотезы о существовании положительной зависимости между известностью аналитиков и доверием инвесторов. В качестве основного фактора, определяющего известность аналитика, рассматривается его цитируемости в СМИ.

Согласованность рекомендаций и прогнозов финансовых аналитиков была проанализирована в [176] на основе данных 1985–1998гг. Под согласованной рекомендацией в работе понималась такая рекомендация, по которой аналитики в большинстве сходятся в одном мнении, будь то продать или купить. Уровень согласованности рекомендаций вычислялся как среднее значение всех рекомендаций по конкретной фирме. В работе методами регрессионного и многофакторного анализа была исследована зави-

симость согласованности прогнозов и изменение согласованности от ряда показателей – характеристик акций. Было выявлено, что изменение согласованности рекомендаций может достаточно точно предсказать краткосрочные (от 3 до 12 месяцев) доходности акций.

Согласованность рекомендаций частично может быть объяснена следованием «стадному чувству» при прогнозировании стоимости акций. Этот аспект был проанализирован в [177]. В этой работе на основе статистических данных была осуществлена проверка гипотезы о том, что аналитики склонны «следовать за толпой» к консенсусу, когда они делают пересмотр рекомендаций. Авторы, в частности, показали, что: 1) реакция рынка на пересмотр рекомендаций аналитиками сильнее, когда они отличаются от консенсуса, чем когда они ему соответствуют; 2) существует более сильный эффект стадности мнений, когда аналитик принимает решение «upgrade», чем когда принимается решение « downgrade»; 3) аналитики с более хорошей репутацией больше привержены «стадному поведению», чем менее известные новички, которые пытаются выделиться; 4) те аналитики, которые нечасто пересматривали свои рекомендации, более склонны к стадному чувству; 5) рынок ожидает «стадного поведения» аналитиков, и на дату пересмотра рекомендаций реакция рыночных цен ее учитывает.

Проблема «несимметричности» прогнозов и рекомендаций аналитиков (аналитики чаще дают позитивные рекомендации, чем негативные) и проблема манипулирования рекомендациями в целом исследовалась в [178]. В этой работе были выделены стратегические причины манипулирования, которые связаны с неправильными стимулами (например, аналитик ставит целью склонить инвесторов к покупке акций или хочет угодить менеджменту компании), и нестратегические причины, когда аналитики имеют слишком позитивные ожидания. Стратегические факторы манипулирования были оценены с помощью показателей: 1) аффилированности аналитика с компанией, по которой он дает рекомендацию; 2) финансового давления на инвестиционный банк со стороны компании, по которой делается прогноз;

3) банковского репутационного капитала; 4) приверженности банка своим клиентам; 5) институционального владения; 6) известности аналитика. В частности, было показано, что: а) аффилированные аналитики и инвестиционные банки, имеющие сильные связи с компаниями, по которым дается прогноз или рекомендация, искажают свои рекомендации, однако это несправедливо в отношении их прогнозов; б) у известных аналитиков рекомендации менее оптимистичны, чем прогнозы; в) аффилированные аналитики и аналитики, которые подвержены давлению инвестиционного банка, сначала «прячутся в толпе», когда делают рекомендации, а затем сохраняют позитивные рекомендации дольше, чем аналитики, инвестиционный банк которых не связан с компанией, по которой они делали рекомендации или прогнозы и не испытывающие давление со стороны инвестиционного банка; г) поведение, соответствующее стратегическим мотивам, присущее более чем половине аналитиков рассмотренной выборки; д) крупные инвесторы делают корректировку на искажение рекомендаций вверх, малые же инвесторы не делают корректировок на искажение.

Большинство исследований прогнозов финансовых аналитиков опираются на базы данных Investext<sup>7)</sup>, Zacks database<sup>8)</sup>, The Institutional Brokers' Estimate System (IBES)<sup>9)</sup>, Value Line<sup>10)</sup> и др. Однако данные, опубликованные в этих базах, относятся в основном к зарубежным компаниям и не предоставляют информацию о российском рынке.

Анализ качества прогнозов аналитиков на российском рынке ценных бумаг представлен значительно хуже. Так в [5] на основе анализа российских данных за 2012 г. было выявлено, что на прогнозном горизонте в 39 дней максимальная точность прогноза о направлении изменения цены составляет 56%. Этот результат согласуется с выводами [179] о том, что для устойчивости существования трейдеров на фондовом рынке им доста-

---

<sup>7)</sup> URL: <http://research.thomsonib.com/> (дата обращения: 03.08.2015)

<sup>8)</sup> URL: <http://www.zacks.com/> (дата обращения: 05.12.2013)

<sup>9)</sup> URL: <http://www.answers.com/topic/institutional-brokers-estimate-system> (дата обращения: 05.12.2013)

<sup>10)</sup> URL: <http://www.valueline.com/> (дата обращения: 03.08.2015)

точно прогнозировать регулярные «биржевые» события с точностью, незначительно превышающей 50%.

Исследование, результаты которого отражены в этом подразделе, было проведено в ключе нескольких актуальных направлений, упомянутых выше. Во-первых: был проведен анализ факторов, которые влияют на точность прогнозов аналитиков, во-вторых, было проведено исследование долгосрочных прогнозов и рекомендаций аналитиков на согласованность прогнозируемых цен и рекомендаций.

Можно выделить четыре категории факторов, влияющих на точность прогнозов финансовых аналитиков. Во-первых, это качество, объем и формы представления данных, которые используют финансовые аналитики для составления прогнозов.

Второй категорией, которая тесно связана с первой, являются внешние факторы. Прогнозы финансовых аналитиков зависят от экономической стабильности в мире, экономической конъюнктуры страны, в которой располагается исследуемая организация, от уровня экономического развития страны, который в свою очередь влияет на качество предоставленных данных для анализа. Кроме того, важно обращать внимание на отрасль, по которой аналитики составляют прогнозы; одни отрасли могут быть более стабильны и предсказуемы, другие – нет [180].

Третья категория – горизонты прогноза. Эта категория была проанализирована в [5].

Четвертым фактором, влияющим на точность прогнозов, являются индивидуальные особенности аналитиков и их стимулы для предоставления правильного прогноза. Точность прогноза часто зависит от того, является ли эксперт публичным (buy-side) или внутренним (sell-side)<sup>11)</sup> аналитиком, то есть, является ли он независимым экспертом или принадлежит к группе аналитиков, которые работают в определенной инвестиционной компании, а также, в какой степени вознаграждение аналитика зависит от его точности

---

<sup>11)</sup> Перевод дан Тепловой Т.В.

прогноза. Важными характеристиками аналитика являются его аналитические способности, образование, признанность (например, наличие сертификата Chartered Financial Analyst), известность [181], которые также могут быть детерминантами точности прогноза.

В данном исследовании не ставилась задача проанализировать все указанные категории точности прогнозов и рекомендаций. Основная цель состояла в исследовании взаимосвязи точности рекомендаций и их согласованности/несогласованности. Для оценки степени несогласованности рекомендаций в исследовании был использован аппарат теории функций доверия. Выбор этого математического аппарата обусловлен тем, что теория функций доверия, как развитие теории Депмстера-Шейфера [182, 183], хорошо приспособлена для анализа неточных данных, в том числе имеющих интервальный характер. В то же время рекомендации аналитиков неявно также носят интервальный характер. Кроме того, теория функций доверия (в отличие от чисто статистических методов) позволяет работать с выборочными данными небольшого объема (при соответствующем увеличении степени неопределенности результатов).

Содержание этого подраздела далее организовано следующим образом. В пункте 3.4.2 описаны использованные в исследовании данные о прогнозах и рекомендациях инвестиционных банков относительно стоимости акций российских компаний за период 2012-2014 гг. В пункте 3.4.3 описана методология исследования и используемый математический аппарат. В пункте 3.4.4 приведены полученные результаты, которые включают оценку точности рекомендаций и доходности, как по банкам, так и по отраслям; эконометрическое исследование рекомендаций (анализ детерминант рекомендаций); анализ конфликтности рекомендаций инвестбанков. Наконец, в последнем пункте сделаны некоторые выводы и намечены шаги дальнейших исследований.

### 3.4.2 Обзор данных

В работе были исследованы данные, которые отражают информацию о 1572 прогнозах 23 инвестиционных банков по 10 отраслям финансовых аналитиков за 2012 год, 1307 прогнозах за 2013 год и 1095 прогнозах 24 инвестбанков по 11 отраслям за 2014. Информация была взята из баз данных Bloomberg, Thomson Reuters и РБК. На портале QUOTE.RBC.RU<sup>12)</sup> «Прогнозы цен акций» содержатся прогнозы и рекомендации по акциям российских эмитентов, торгующихся как на российских, так и иностранных фондовых биржах. Прогнозы представлены авторитетными экспертами крупнейших мировых инвестбанков, в числе которых такие именитые компании, как Goldman Sachs, Credit Suisse, UBS, Deutsche Bank, Ренессанс Капитал и другие (всего более 30 компаний). Информация структурирована как по отраслям, так и по прогнозам одного инвестбанка. В базе данных указаны даты, опубликованных прогнозов, и динамика цен акции во времени. Доступна информация о последних прогнозах инвестиционных компаний, а также информация о динамике цен акций с 2004 года. Информация о количестве прогнозов по последним годам приведена в Таблице 3.3.

Таблица 3.3 – Обзор данных

Год	Количество прогнозов	Количество инвестиционных банков	Количество отраслей
2012	1572	23	10
2013	1307	23	10
2014	1095	24	11

В базе данных содержится банк, эмитент, отрасль, к которой принадлежит эмитент, а также тип ценной бумаги, по которой проводился прогноз (ADR – Американская депозитарная расписка, АО – акции обыкновенные), даты прогноза, цель<sup>13)</sup> в валюте прогноза, а также рекомендация аналитика. В данном исследовании будем рассматривать только рекомендации трех типов: «Покупать» (Buy, Outperform, Overweight, Выше рынка), «Держать» (Hold,

<sup>12)</sup> URL: <http://quote.rbc.ru/shares/> (дата обращения: 03.08.2015)

<sup>13)</sup> Цель – это значение цены, которое прогнозируется аналитиком, для определенного вида конкретного актива. Горизонт прогноза не указан.

Neutral, По рынку) и «Продавать» (Sell, Underweight, Underperform, Ниже рынка).

Можно заметить (см. рисунок 3.40), что количество рекомендаций стабильно снижается. Однако связь уменьшения количества рекомендаций с ростом волатильности фондового рынка Рисунок 3.40 – Динамика числа прогнозов и волатильности индекса РТС, 2012-2014 гг. явно не прослеживается. Кроме того, было проверено, что уменьшение числа рекомендаций не связано с объемом торгов и числом сделок на бирже.

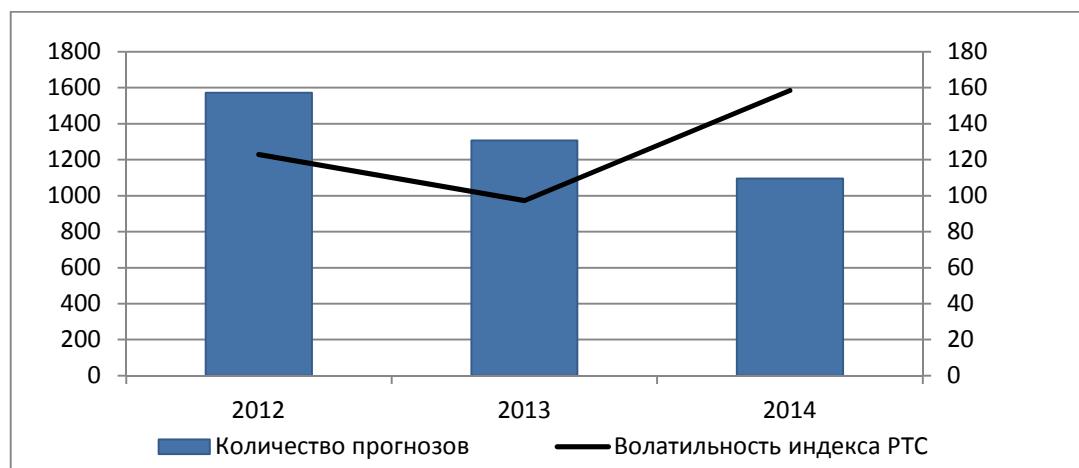


Рисунок 3.40 – Динамика числа прогнозов и волатильности индекса РТС, 2012-2014 гг.

За рассматриваемый период чаще всего была дана рекомендация «Покупать». Процентное отношение различных рекомендаций показано на рисунке 3.41.

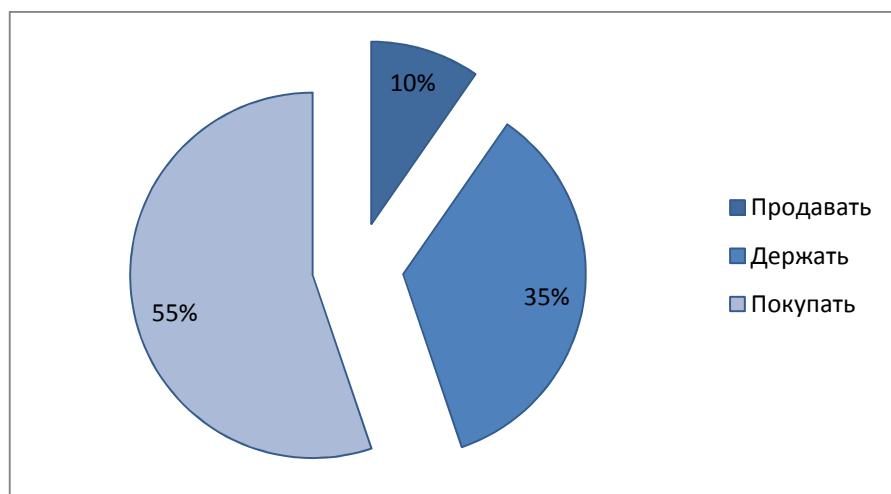


Рисунок 3.41 – Структура рекомендаций за период 2012–2014 гг.

### 3.4.3 Методология исследований

Данный пункт включает в себя несколько подпунктов, в каждом из которых описана методология или используемый математический аппарат определенной части исследования. В первом подпункте проанализированы основные возможные варианты определения точности рекомендаций и прогнозов. Во втором подпункте описана методология расчета точности рекомендаций и доходность, используемая в настоящем исследовании. Третий подпункт посвящен определению временного горизонта, при котором достигается наибольшая доля точных рекомендаций финансовых аналитиков. В четвертом подпункте анализируются пограничные значения относительных целевых цен между тремя типами рекомендаций. В последнем подпункте описан математический аппарат так называемой меры конфликта в рамках теории функций доверия, который вводится для оценивания степени несогласованности (конфликтности) рекомендаций и прогнозов финансовых аналитиков.

#### 3.4.3.1 Варианты определения точности рекомендаций и прогнозов

Сложность и неоднозначность определения точности рекомендаций финансовых аналитиков связана с тем, что границы рекомендаций точно не определены. В общем случае можно считать, что точность рекомендаций – это отношение числа успешных рекомендаций к общему числу рекомендаций. Но «успешность» рекомендации можно понимать по разному, и, в зависимости от этого, возможны, например, следующие варианты определения точности рекомендаций.

Слабая точность. В этом случае рекомендации «Покупать» или «Держать» будем считать успешными, а рекомендацию «Продавать» – неуспешной, если в результате по соответствующим акциям была получена прибыль. Если же была получена убыль, то успешными будем считать рекомендации

«Продавать» или «Держать», а неуспешной – рекомендацию «Покупать». Другими словами, в этом случае рекомендации «Держать» интерпретировались как равнозначные рекомендациям «Покупать» или «Продавать» в зависимости от того была получена прибыль или убыль соответственно. Недостаток такого подхода состоит в том, что если банк давал только рекомендации «Держать», то слабая точность будет равна единице безотносительно к тому, получена прибыль или нет.

Сильная точность. Эта точность вычисляется аналогично слабой, только из рассмотрения исключаются рекомендации «Держать». Недостаток такого определения состоит в том, что сильную точность нельзя вычислить, если банк давал только рекомендации «Держать». Компромиссным вариантом, лишенным указанных недостатков первых двух определений, является следующее понятие.

Симметричная точность вычисляется аналогично слабой точности, только рекомендации «Держать» считаются успешными при получении прибыли и неуспешными при получении убыли.

h-точность. В этом случае успешной будем считать следующие рекомендации: а) «Покупать», если в результате была получена прибыль, большая, чем  $h$  (в относительных ценах); б) «Держать», если прибыль/убыль оказалась в интервале от  $-h$  до  $h$ ; в) «Продавать», если убыль оказалась меньше  $-h$ . Во всех остальных случаях рекомендации считаются неуспешными. Недостаток такого подхода связан с зависимостью от параметра  $h>0$ .

Точность относительно типа рекомендации («Покупать» или «Продавать»). Например, для точности относительно рекомендации «Покупать» успешной будут рекомендации «Покупать» или «Держать», по которым была получена прибыль. Тогда точность относительно рекомендации «Покупать» равна отношению успешных рекомендаций к общему числу рекомендаций «Покупать» или «Держать».

Консенсус-точность. В этом случае рассматривается совокупность различных рекомендаций по одной бумаге (или по ряду бумаг), которые были

сделаны инвестбанками. Оптимальные границы интервалов относительных цен разных типов рекомендаций определяются как решение задачи максимизации суммарного числа относительных прогнозных цен, которые попадают в интервалы «своего типа» рекомендаций. После этого успешной будем считать ту рекомендацию, чья относительная цена попала в промежуток, который соответствует рекомендации. В остальных случаях рекомендацию считаем неуспешной.

Пример 3.1. Пусть инвестбанк дал: а) три рекомендации «Продавать» и по ним были получены относительные прибыли/убыли  $\{-0.2, -0.1, 0.2\}$ ; б) две рекомендации «Держать» и по ним были получены прибыли/убыли  $\{-0.3, 0.1\}$ ; в) четыре рекомендации «Покупать» и по ним были получены прибыли/убыли  $\{-0.1, 0.2, 0.3, 0.4\}$ .

Тогда:

$$\text{слабая точность: } t_{weak} = \frac{4+3}{3+2+4} = \frac{7}{9};$$

$$\text{сильная точность: } t_{strict} = \frac{3+2}{3+4} = \frac{5}{7};$$

$$\text{симметричная точность: } t_{sym} = \frac{2+1+3}{3+2+4} = \frac{2}{3};$$

$$0.05\text{-точность: } t_{0.05} = \frac{3+0+2}{9} = \frac{5}{9};$$

$$0.15\text{-точность: } t_{0.15} = \frac{3+0+1}{9} = \frac{4}{9};$$

$$0.25\text{-точность: } t_{0.25} = \frac{2+1+0}{9} = \frac{1}{3};$$

$$\text{консенсус-точность: } t_{cons} = \frac{2+1+3}{9} = \frac{2}{3};$$

$$\text{точность относительно «Покупать»: } t_{buy} = \frac{1+3}{2+4} = \frac{2}{3}.$$

Возможны, конечно, и другие варианты определения точности. В нашем исследовании в качестве определения точности рассматривалось понятие симметричной точности.

Точность прогноза целевой цены связана с абсолютным отклонением фактически полученной относительной цены от прогнозной.

### 3.4.3.2 Методология расчета точности прогнозов и доходности

При расчете точности прогнозов финансовых аналитиков и доходности их рекомендаций были сделаны следующие предпосылки:

- акции одного и того же эмитента, торгуемые на разных площадках, рассматриваются как отдельные бумаги;
- акции являются бесконечно делимыми – на 1000 единиц валюты можно купить любое дробное количество акций;
- рекомендация «Держать» интерпретировалась как равнозначная рекомендации «Покупать», если в результате получалась прибыль или «Продавать» в случае получения убыли. Возможны и другие схемы расчета точности (см. предыдущий подпункт).

Расчет точности прогнозов состоит из следующих шагов:

- 1) в момент выхода рекомендации по бумаге данная бумага покупалась на 1000 единиц валюты, в которой торгуется (рубли, фунты, доллары), в случае, если давалась рекомендация «Покупать», и продавалась при рекомендации «Продавать»;
- 2) через временной горизонт 50 дней все купленные бумаги продавались, либо закрывалась короткая позиция (т.е. бумаги, взятые ранее взаймы и проданные на открытом рынке, выкупались и возвращались обратно брокеру). Горизонт в 50 дней был обязательным для каждой рекомендации;
- 3) после совершения операции, описанной в пункте 2, фиксировалась прибыль или убыток по каждой рекомендации;
- 4) фиксировался факт успеха или неуспеха рекомендаций банка (аналитика): успешными считались рекомендации «Покупать» или «Держать» в случае фактического роста цены и рекомендация «Продавать» в случае фактическое снижение цены; неуспешными считались рекомендации «Покупать»

или «Держать» при фактическом снижении цены, а также рекомендация «Продавать» при фактическом росте цены;

5) точность прогноза для каждого банка вычислялась как сумма успешных рекомендаций, деленная на общее число рекомендаций данных банком.

Расчет доходности производился следующим образом:  
производились операции, перечисленные ранее в пунктах 1 – 3;

- 1) суммировалась сумма прибылей и убытков по всем рекомендациям, выданным определенным банком (аналитиком); итогом была сумма, заработанная (потерянная) при исполнении всех рекомендаций, данных каждым конкретным банком (аналитиком);
- 2) вычислялся объем изначально вложенных средств;
- 3) делением результата, полученного в пункте 3 на вложения, посчитанные в пункте 2, находилась доходность портфеля, сформированного по рекомендациям инвестбанка (аналитика).

#### 3.4.3.3 Выбор временного горизонта инвестирования

Для каждого временного горизонта в пределах от 1 до 91 дня был определен процент точных рекомендаций, данных финансовыми аналитиками всех банков и по всем отраслям. Результаты представлены на рисунке 3.42.

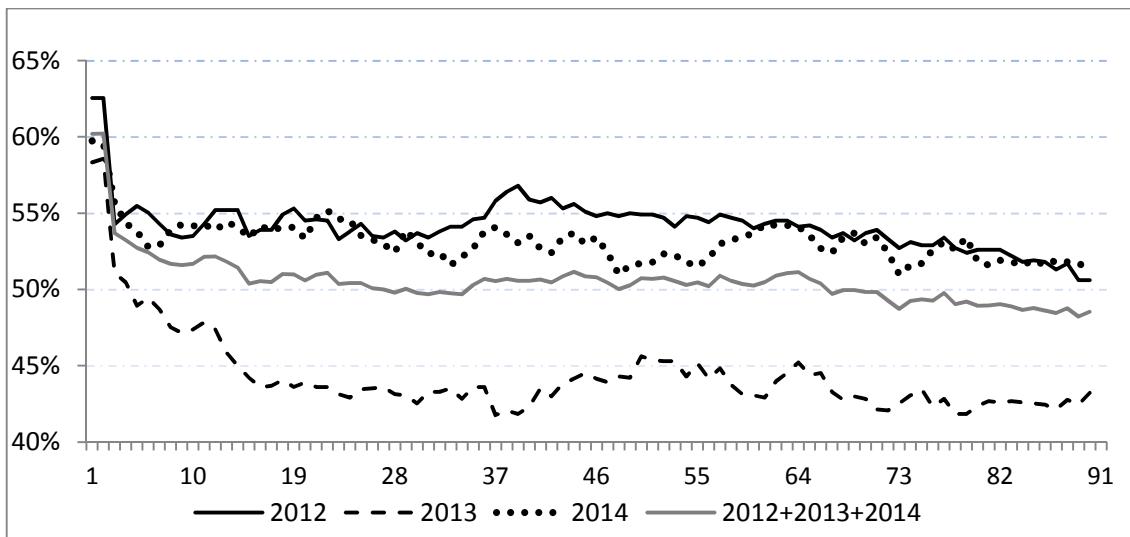


Рисунок 3.42 – Доля точных прогнозов в зависимости от временного горизонта

Максимальное количество успешных рекомендаций, данных аналитиками, среди горизонтов от 15 дней до 90 дней на основе данных 2012 года достигается на временном горизонте 39 дней, точность прогноза равна 56,8%. Доля успешных прогнозов 2013 года максимальна при временном горизонте в 50 дней и равна 45,6%. В 2014 же году это значение составляет 54%. Проанализировав средние значения долей успешных прогнозов 2012–2014 гг. было выявлено, что в среднем наиболее успешным временем горизонтом для трех лет является 2 дня с долей успешных прогнозов в 60% среди краткосрочных временных горизонтов (от 1 до 15 дней) и 44 дня среди временных горизонтов от 15 до 90 дней с долей успешных прогнозов в 51%.

С ростом временного горизонта (см. рисунок 3.43) доля верных прогнозов аналитиков снижается. Вероятно, данный факт связан с тем, что аналитики пересматривают прогнозные значение и тип рекомендации в связи с обновлением отчетности компании. Компании публикуют свои отчетности ежеквартально или каждые полгода.

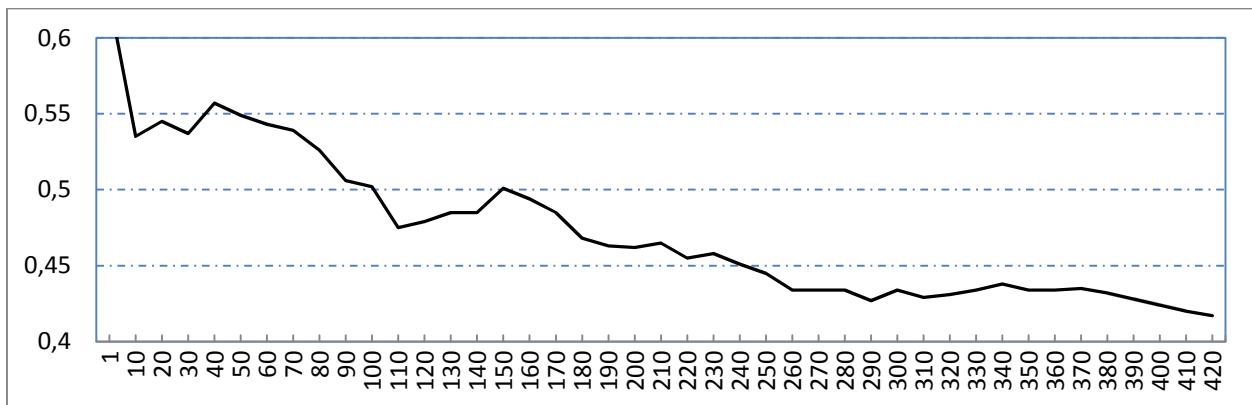


Рисунок 3.43 – Доля точных прогнозов за 2012 год в зависимости от временного горизонта

### 3.4.3.4 Оценка интервальных границ между тремя типами рекомендаций

Цель данного анализа заключается в выявлении границ относительных целевых цен<sup>14)</sup> между тремя типами рекомендаций.

Для этого по всем прогнозам была построена интервальная шкала относительных цен с шагом в 0,1 на промежутке [0,3; 2,9]. Кроме того, были рассмотрены интервалы [0; 0,3) и (2,9; +∞]<sup>15)</sup>. Для каждой из рекомендаций «Продавать», «Держать» и «Покупать» было вычислено процентное отношение количества прогнозов, попавших в тот или иной интервал. На основе полученных данных были построены гистограммы эмпирических плотностей распределения для каждого типа рекомендаций, представленные на рисунке 3.44.

Из рисунка 3.44 видно, что границы между рекомендациями «Продавать» и «Держать» (как точка пересечения двух соответствующих графиков) отдельно по каждому из трех рассматриваемых годов сосредоточены в интервале относительной цены от 0,89 до 0,92, в то время как границы между рекомендациями «Держать» и «Покупать» принадлежат диапазону от 1,17 до 1,46.

<sup>14)</sup> Относительное значение целевой цены – это отношение цены прогнозируемой аналитиком к значению котировки на датудачи прогноза.

<sup>15)</sup> Подробные данные могут быть предоставлены по запросу.



Рисунок 3.44 – Процентное соотношение количества рекомендаций для определенного уровня относительной цены по трем годам (2012–2014)

Кроме того, существует асимметрия интервалов, поскольку доли рекомендаций каждого типа в исследуемой выборке не равны друг другу – рекомендация «Покупать»дается значительно чаще остальных рекомендаций.

Визуальный анализ указывает на противоречие как в уровнях рекомендаций (относительные цены), так и их типах («Покупать»/«Держать»/«Продавать»). Так, при любых значениях относительной цены наблюдаются все три типа рекомендаций: «Продавать», «Держать» и «Покупать».

На следующем этапе были найдены оптимальные границы между различными типами рекомендаций, которые максимизируют суммарное число относительных прогнозных цен, которые попадают в интервалы «своего типа» рекомендаций. Значения границ для трех лет получились примерно одинаковыми, а именно: граница между рекомендациями «Продавать» и «Держать» равна 0,97 по всему массиву данных, а граница между «Держать» и «Покупать» равна 1,22 для трех лет. В Таблице 3.4 отражены средние значения относительных цен для каждого типа рекомендаций и пограничные значения относительных цен между рекомендациями.

Таблица 3.4 – Среднее значение относительной цены и границы цен между рекомендациями

	Среднее значение	Оценка	Среднее значение	Оценка	Среднее значение
Год	«Продавать»	Граница	«Держать»	Граница	«Покупать»
2012	0,86	0,98	1,12	1,21	1,52
2013	0,95	0,92	1,09	1,2	1,36
2014	0,84	0,97	1,16	1,24	1,49
2012–2014	0,87	0,97	1,12	1,22	1,46

### 3.4.3.5 Математический аппарат анализа конфликтности рекомендаций

Конфликтность характеризует то, насколько сильно не согласованы рекомендации некоторого множества экспертов. Конфликтность вместе с точностью рекомендаций является важной характеристикой качества прогнозирования. Высокая конфликтность рекомендаций некоторого банка вместе с высокой точностью могут быть связаны как с наличием некоторых эксклюзивных методик, так и с использованием, например, инсайдерской информации. Высокая конфликтность вместе с низкой точностью скорей всего свидетельствует о низком профессионализме аналитиков.

Для исследования конфликтности рекомендаций был использован аппарат теории функция доверия (теории свидетельств), как развитие теории Демпстера–Шейфера [182, 183]. Выбор этого математического аппарата был обусловлен тем, что он приспособлен именно для анализа интервальных данных и позволяет работать с выборками небольшого объема.

Принадлежность относительной цены рекомендации определенного типа («Покупать»–«Держать»–«Продавать») к котировке можно считать свидетельством данного аналитика (инвестбанка). Тогда можно найти тело свидетельств данного аналитика. Пусть для  $i$ -го аналитика и его рекомендации «Держать» оказалось, что  $c_k^{(i)}$  раз относительная цена прогноза попала в промежуток  $A_k$  (например,  $A_k = [1; 1,1)$ ), а всего он  $N_i$  раз давал прогноз «Держать». Тогда  $m^{(i)}(A_k) = c_k^{(i)} / N_i$  – частота попадания его прогнозов в промежуток  $A_k$ , а  $\mathcal{A}^{(i)} = (A_k, m^{(i)}(A_k))_k$  – тело свидетельств  $i$ -го аналитика

по данному типу рекомендаций. Функции множеств  $m^{(i)}$  называются в теории свидетельств основными вероятностными назначениями (the basic probability assingment (bpa) [183]). Эти функции удовлетворяют условиям:  $\sum_k m^{(i)}(A_k) = 1$  для всех  $i = 1, \dots, l$ . Зная все тела свидетельств, можно найти меру конфликта этих свидетельств, например, по формуле (мера конфликта в правиле Демпстера)

$$K_0 = K_0(\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(l)}) = \sum_{\{A_k\}: \bigcap_k A_k = \emptyset} \prod_i m^{(i)}(A_k) \quad (3.1)$$

или по другой формуле, характеризующей меру конфликта (см, например, [184, 185]). Величину  $K_0(\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(l)})$  ( $l$  – число аналитиков) будем называть канонической мерой конфликта. Аксиоматика и способы задания различных мер конфликта проанализированы в [185].

Для рассматриваемой структуры тел свидетельств меру конфликта  $K_0(\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(l)})$  можно вычислить по более простой формуле, чем (3.1). Для этого нам понадобится следующее утверждение.

Теорема 3.1. Если тела свидетельств  $\mathcal{A}^{(i)} = (A_k, m^{(i)}(A_k))_k$ ,  $i = 1, \dots, l$  такие, что  $A_s \cap A_k = \emptyset$  при  $s \neq k$ , то мера конфликта  $K_0(\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(l)})$  будет равна

$$K_0(\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(l)}) = 1 - \sum_k \prod_{i=1}^l m^{(i)}(A_k). \quad (3.2)$$

Чем ближе значение  $K_0$  к нулю, тем конфликтность между рекомендациями аналитиков ниже, и наоборот, чем ближе это значение к единице, тем конфликтность выше.

Пусть  $X = \{\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(l)}\}$  – множество всех свидетельств. Тогда меру конфликта  $K$  можно рассматривать как функцию множеств, определенную на  $2^X$  – множестве подмножеств из  $X$ . В частности, каноническая мера конфликта  $K_0$  определена на  $2^X$  и обладает свойствами:

- 1)  $K_0(B)=0$ , если  $|B|=1$ ;
- 2)  $K_0(B)\leq K_0(C)$ , если  $B\subseteq C$ .

Второе свойство говорит о том, что  $K_0$  – монотонная (неаддитивная) мера на  $2^X$  [186]. Положим по определению, что  $K_0(\emptyset)=0$ .

Условиям монотонной меры удовлетворяет и «близкая» к мере  $K_0$  функция

$$K_1(\mathcal{A}^{(1)}, \dots, \mathcal{A}^{(l)}) = 1 - \sum_k \min_{1 \leq i \leq l} m^{(i)}(A_k). \quad (3.3)$$

Формально, с точки зрения теории агрегирующих функций [187], функция множеств  $K_1$  получается из функции  $K_0$  заменой одной триангулярной нормы ( $t$ -нормы) – произведения, на другую  $t$ -норму –  $\min$ .

Мера (3.3) в отличие от меры (3.2) обладает свойством индифферентности к добавлению свидетельства-дубликата:  $K_1(B) = K_1(B \cup \{\mathcal{A}'\})$ , если существуют  $\mathcal{A}'' \in B$ ,  $\mathcal{A}' = \mathcal{A}''$ . Мера (3.2) в случае добавления свидетельства-дубликата не уменьшится.

Если известна монотонная мера конфликта  $K$ , определенная на всех подмножествах-наборах свидетельств из  $X$ , то можно определить «вклад»  $i$ -го аналитика в общий конфликт  $K(X)$  множества всех аналитиков из  $X$ , как разность  $K(X) - K(X \setminus \{\mathcal{A}^{(i)}\})$ . Более точно «вклад»  $i$ -го аналитика в конфликт можно определить как средний его вклад в конфликт группы (коалиции) аналитиков  $B$ :  $K(B) - K(B \setminus \{\mathcal{A}^{(i)}\})$ , где среднее вычисляется по всем группам (коалициям) аналитиков  $B$ ,  $B \subseteq X$ ,  $\mathcal{A}^{(i)} \in B$ . В этом случае мы получим так называемое значение (индекс) Шепли [188], которое широко используется в теории коалиционных (кооперативных) игр

$$v_i = \sum_{B \subseteq X, \mathcal{A}^{(i)} \in B} \frac{(l - |B|)!(|B| - 1)!}{l!} (K(B) - K(B \setminus \{\mathcal{A}^{(i)}\})). \quad (3.4)$$

Вектор  $\mathbf{v} = (v_i)_{i=1}^l$  называется вектором Шепли и выполняется равенство

$$\sum_{i=1}^l v_i = K(X).$$

Непосредственное вычисление индекса Шепли по формуле (3.4) имеет временную сложность вычисления  $O(2^l)$  и практические неосуществимо, если  $l$  равно нескольким десяткам и более. Существует ряд методов эффективного (за полиномиальное и даже линейное время) точного и приближенного вычисления значений Шепли. Обзор по этим методам можно найти в [189]. Ниже в своих вычислениях будем использовать самый простой из этих методов – метод случайной перестановки [190]. А именно, будем вычислять приближенное значение Шепли по формуле

$$\tilde{v}_i = \frac{1}{m} \sum_{\substack{\pi \in \Pi, |\Pi|=m, \\ \pi(j)=i}} (K(B_{\pi(j)}) - K(B_{\pi(j-1)})),$$

где  $\Pi$  – некоторое множество различных перестановок  $\pi: \{1, 2, \dots, l\} \rightarrow \{1, 2, \dots, l\}$ ,  $B_{\pi(j)} = \{\mathcal{A}^{(\pi(1))}, \dots, \mathcal{A}^{(\pi(j))}\}$ . Другими словами, будем генерировать множество из  $m$  различных коалиций и для каждой коалиции  $B_{\pi(j)}$  будем вычислять вклад  $i$ -го эксперта в конфликт этой коалиции. Если  $m = l!$  (т.е. мы переберём все различные коалиции), то мы получим точное значение Шепли. На практике  $m \ll l!$  и для «хорошей» аппроксимации  $\Pi$  должно быть равномерно сгенерированным множеством перестановок:  $\pi(1) = \text{rnd}(\{1, \dots, l\}), \dots, \pi(k) = \text{rnd}(\{1, \dots, l\} \setminus \{\pi(1), \dots, \pi(k-1)\}), k = 2, \dots, l$ , где  $\text{rnd}(B)$  – равномерно распределенное число на дискретном числовом множестве  $B$ .

### 3.4.4 Результаты исследований

#### 3.4.4.1 Точность рекомендаций и доходность в разрезе инвестиционных банков и в разрезе отраслей

Интересным представляется рассмотреть точность рекомендаций инвестиционных банков и доходность портфеля, сформированного по рекомендациям банка, в динамике за исследуемый период. Для исследования точности рекомендаций банки условно были поделены на три группы: 1) банки с точностью рекомендаций за три года ни разу не опускавшейся ниже 50%; 2) с точностью рекомендаций, никогда не превышавшей 50%; 3) с иной точностью рекомендаций. В Таблице 3.5 представлены результаты анализа точности рекомендаций (в последнем столбце указано изменение точности за три года).

Таблица 3.5 – Изменение точности рекомендаций банков за три года (2012–2014)

№	Инвестбанк	2012 (%)	2013 (%)	2014 (%)	Δ(2014–2012), %
1	Метрополь	63,79	52,63	53,33	-10,46
2	Ренессанс Капитал	63,78	51,56	51,67	-12,11
3	Goldman Sachs	54,14	52,63	61,40	7,26
4	J.P. Morgan	63,86	58,57	53,85	-10,01
5	Morgan Stanley	58,82	63,64	54,35	-4,48
6	Альфа-банк	56,57	36,11	66,67	10,10
7	БКС	60,00	36,92	59,42	-0,58
8	Велес Капитал	38,46	51,11	53,57	15,11
9	ВТБ Капитал	51,69	46,46	46,99	-4,70
10	Газпромбанк	55,00	28,21	59,09	4,09
11	Открытие	58,54	50,00	47,37	-11,17
12	Уралсиб Кэпитал	66,67	51,90	41,03	-25,64
13	Barclays	70,27	49,12	61,54	-8,73
14	Citi group	48,89	28,57	60,78	11,90
15	Credit suisse	58,73	43,75	63,27	4,54
16	Deutsche Bank	44,35	39,42	75,00	30,65
17	HSBC	56,00	44,74	52,63	-3,37
18	Raiffeisen	54,55	26,09	43,59	-10,96
19	Rye. Man&GorSecurities	61,70	33,33	88,89	27,19
20	Sberbank CIB	57,50	43,66	31,51	-25,99
21	UBS	56,52	45,16	63,16	6,64

Как можно заметить, пять банков, такие как Метрополь, Ренессанс Капитал, Goldman Sachs, J.P. Morgan и Morgan Stanley, демонстрировали точность выше 50%. Остальные банки достаточно волатильны в точности. Кроме того, в нашей выборке не оказалось банков, точность рекомендаций которых ни разу не превышала 50%.

Таким же образом была проанализирована доходность рекомендаций банков. Они были поделены на три группы: 1) банки с доходностью за три года всегда имеющей положительное значение; 2) банки с волатильной доходностью (как положительной, так и отрицательной); 3) с доходностью, всегда принимавшей отрицательное значение. В Таблице 3.6 представлены результаты анализа доходности (в последнем столбце указано изменение доходности за три года)

Таблица 3.6 – Изменение доходности за три года (2012–2014)

№	Инвестбанк	2012 (%)	2013 (%)	2014 (%)	$\Delta(2014-2012), \%$
1	Ренессанс Капитал	3,65	0,37	0,47	-3,18
2	Goldman Sachs	0,51	0,66	5,34	4,83
3	Morgan Stanley	0,60	1,38	1,46	0,85
4	Альфа-банк	1,60	-1,75	7,00	5,40
5	БКС	1,85	-2,08	3,03	1,18
6	Велес Капитал	0,58	2,56	-2,44	-3,02
7	ВТБ Капитал	1,12	1,11	-2,32	-3,44
8	Газпромбанк	1,34	-2,96	0,72	-0,62
9	Метрополь	2,54	-1,26	2,45	-0,10
10	Открытие	1,89	-1,62	1,17	-0,72
11	Уралсиб Кэпитал	2,82	-0,66	1,34	-1,48
12	Barclays	0,40	-0,43	2,21	1,81
13	Citi group	2,60	-3,24	3,53	0,93
14	Credit suisse	0,60	-1,44	6,48	5,88
15	Deutsche Bank	0,01	-1,35	3,29	3,28
16	HSBC	1,27	-0,88	0,58	-0,69
17	J.P. Morgan	1,93	-0,01	-0,95	-2,88
18	Raiffeisen	0,99	-1,93	-2,85	-3,85
19	Rye. Man&GorSecurities	-0,28	-3,66	12,89	13,16
20	Sberbank CIB	1,18	-0,40	-4,95	-6,12
21	UBS	-0,11	-0,49	-0,29	-0,18

Доходность портфелей, сформированных по рекомендациям банков Ренессанс Капитал, Goldman Sachs и Morgan Stanley, всегда была положитель-

ной. В то же время банк UBS всегда демонстрировал отрицательную доходность.

Кроме исследования точности рекомендаций и доходности в разрезе инвестбанков, были проанализированы эти показатели в разрезе отраслей, чтобы в дальнейшем сделать вывод о том, какие отрасли являются наиболее предпочтительными для инвестирования. Результаты представлены ниже в Таблицах 3.7 и 3.8.

Таблица 3.7 – Изменение точности рекомендаций инвестбанков по отраслям за три года (2012–2014)

№	Отрасль	2012 (%)	2013 (%)	2014 (%)
1	Потребительский сектор	59,49	30,88	50,00
2	Химия и нефтехимия	57,14	30,08	55,93
3	Строительство и стройматериалы	61,83	56,47	63,03
4	Энергетика	47,69	32,50	52,86
5	Финансовый сектор	61,17	41,18	52,48
6	Машиностроение	73,91	42,50	38,10
7	Металлургия	45,24	34,84	64,57
8	Нефть, газ и уголь	50,25	38,89	53,33
9	Телеком и медиа	56,07	54,42	45,71
10	Транспорт	51,56	52,94	37,78

Таблица 3.8 – Изменение доходности по отраслям за три года (2012–2014)

№	Отрасль	2012 (%)	2013 (%)	2014 (%)
1	Химия и нефтехимия	2,89	-3,58	6,28
2	Строительство и стройматериалы	3,24	-2,39	2,69
3	Потребительский сектор	3,85	-0,29	1,42
4	Финансовый сектор	2,86	-0,22	1,43
5	Машиностроение	8,99	-5,17	-3,49
6	Металлургия	-0,11	-1,26	3,56
7	Нефть, газ и уголь	-0,73	-0,25	0,38
8	Телеком и медиа	1,37	0,17	-0,12
9	Транспорт	3,48	0,88	-2,69
10	Энергетика	-0,41	-0,23	-0,36

Наиболее постоянен в точности рекомендаций потребительский сектор, в данной отрасли точность рекомендаций всегда была выше 50%. В то время как доходность непостоянна по всем отраслям, за исключением сектора энергетики с постоянной отрицательной доходностью.

### 3.4.4.2 Анализ конфликтности рекомендаций инвестбанков

Мера конфликта, характеризующая согласованность рекомендаций аналитиков, была рассмотрена в динамике за исследуемый период в разрезе, как банков, так и отраслей. В 2012 г. мера конфликта всех рекомендаций составила 79%, в 2013 г. – 73%, в 2014 г. – 85%. Мера конфликта рекомендаций по отдельным отраслям в 2012 г. была на уровне 64%, в 2013 г. – 67%, в 2014 г. – 54%. В целом же за весь период конфликтность всех рекомендаций по банкам составила 78%, а по отраслям – 62%.

Можно заключить, что конфликтность рекомендаций по всем банкам стабильно выше конфликтности прогнозов по отдельным отраслям.

Ниже на рисунке 3.45 представлен вклад каждого банка в общую конфликтность рекомендаций в 2012–2014 годах, оцененный с помощью значений вектора Шепли.

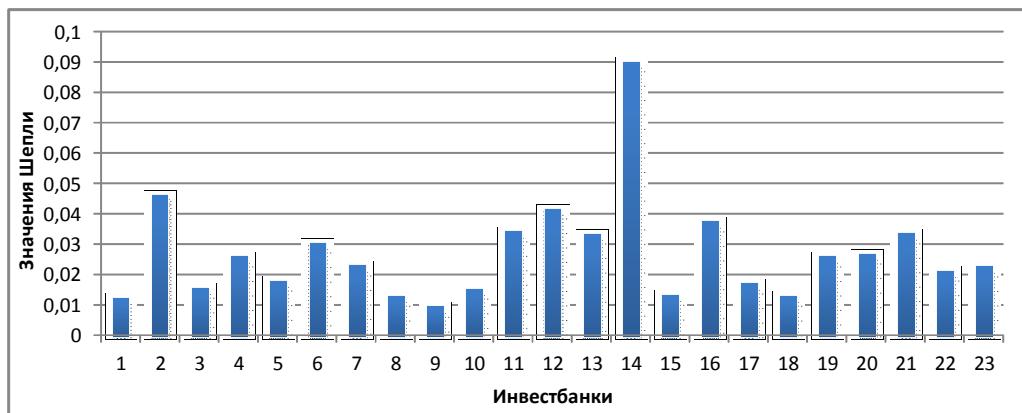


Рисунок 3.45 – Значения Шепли для каждого из инвестиционных банков

Примечание 3.1. На рисунке 3.45 и ниже на рисунке 3.46 использованы следующие обозначения инвестбанков: 1 – Альфа-банк, 2 – Атон, 3 – БКС, 4 – Велес Капитал, 5 – ВТБ Капитал, 6 – Газпромбанк, 7 – Метрополь, 8 – Открытие, 9 – Ренессанс Капитал, 10 – Уралсиб Кэпитал, 11 – Финам, 12 – Barclays, 13 – Citi group, 14 – Credit suisse, 15 – Deutsche Bank, 16 – Goldman Sachs, 17 – HSBC, 18 – J.P. Morgan, 19 – Morgan Stanley, 20 – Raiffeisen, 21 – Rye. Man&GorSecurities, 22 – Sberbank CIB, 23 – UBS.

Однако принятие во внимание лишь одного показателя конфликтности рекомендаций недостаточно для построения эффективной инвестиционной стратегии. Необходимо рассматривать его в совокупности с точностью прогнозов или доходностью. На рисунке 3.46 показано расположение инвестбанков в системе координат «точность рекомендаций – конфликтность» за три года.

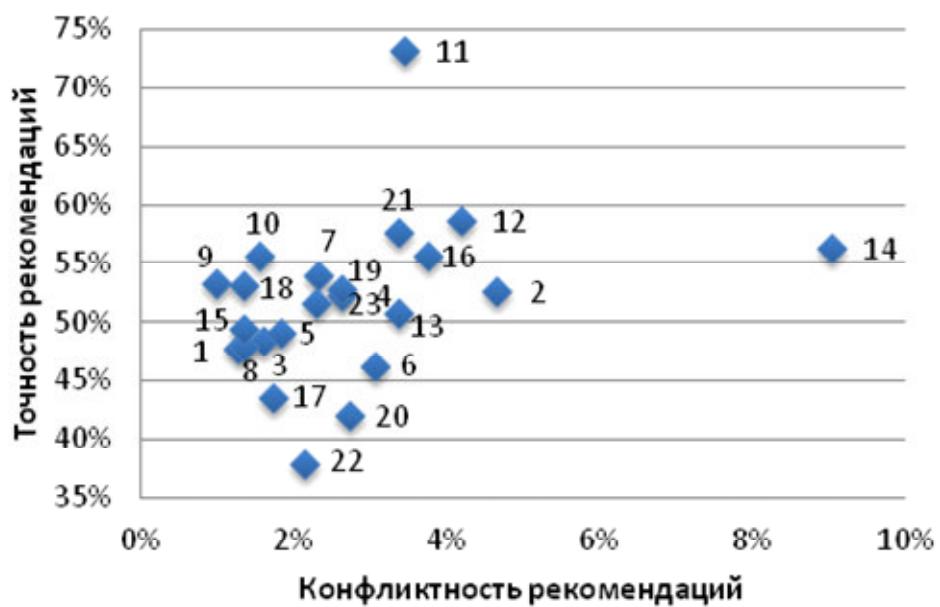


Рисунок 3.46 – Конфликтность рекомендаций инвестбанков – точность прогнозов

Наиболее предпочтительны для следования инвестиционной стратегии конфликтные рекомендации с наивысшей точностью, поскольку, это может говорить о том, что банк использует особые методики или особую информацию. К таким банкам относятся Credit Suisse (14), Barclays (12) и Финам (11).

На рисунке 3.47 показано расположение отраслей в координатах «доходность – конфликтность рекомендаций», вычисленной на основе рекомендаций банков в 2012–2014 годах.

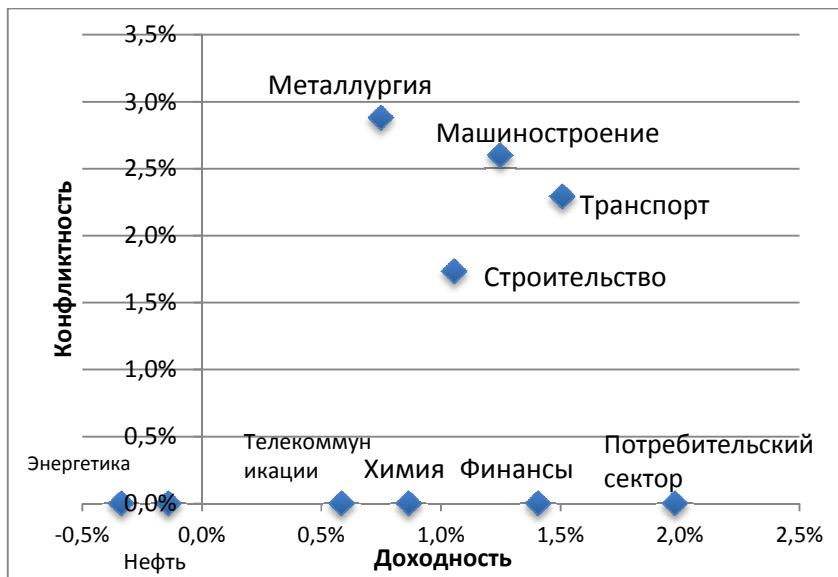


Рисунок 3.47 – Конфликтность рекомендаций – доходность инвестиционных банков по отраслям

Можно выделить четыре отрасли – металлургия, машиностроение, транспорт и строительство, для которых высокая конфликтность наблюдается вместе с высокой доходностью. Высокая конфликтность может говорить о рискованности инвестиционных вложений и о неопределенности перспектив развития отрасли. В тоже время для инвестиционной стратегии целесообразно выбирать данные отрасли, поскольку, вероятнее всего, в них есть банки с наибольшим конфликтом, чьи рекомендации более точны.

Если построить подобные графики в координатах «доходность – точность рекомендаций» (см. рисунок 3.48), то получим, что довольно успешным является банк Финам. Самый низкий уровень доходности и низкую точность рекомендаций демонстрировал Sberbank CIB.

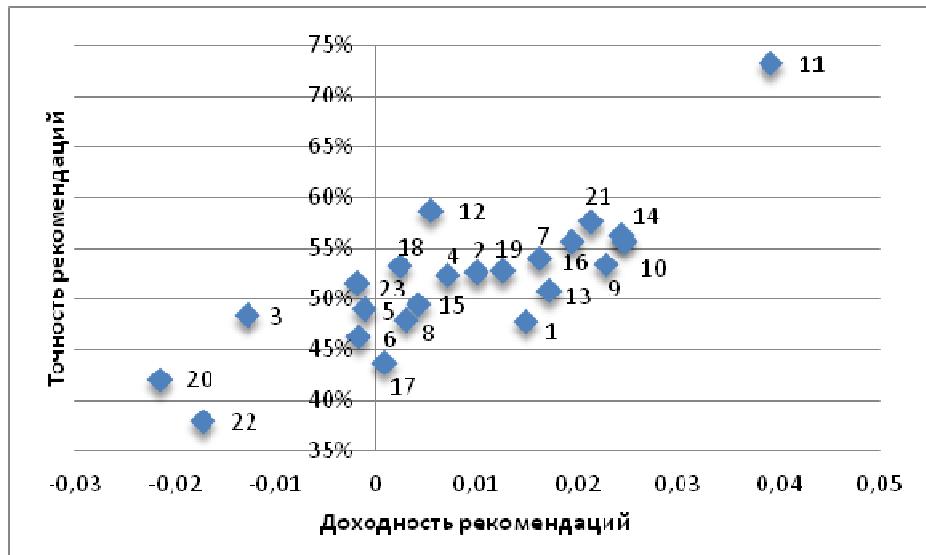


Рисунок 3.48 – Доходность рекомендаций – точность инвестбанков

Что касается отраслей и секторов, то наивысшая доходность рекомендаций при низкой конфликтности наблюдалась в потребительском секторе.

### 3.4.5 Некоторые выводы

В данном исследовании была предпринята попытка количественного анализа информации о качестве рекомендаций и прогнозов финансовых аналитиков (инвестбанков) о стоимости акций российских компаний за период 2012–2014 гг. Основное внимание было уделено анализу точности, доходности и согласованности (конфликтности) рекомендаций. В частности, было показано, что:

- число рекомендаций снижается на 20-30% в год и это не связано с волатильностью, объемом торгов и числом сделок на бирже;
- преобладает число положительных рекомендаций («Покупать»);
- существуют банки с устойчивой точностью прогноза более 50%.

Был проведен анализ границ относительных целевых цен между тремя типами рекомендаций. Показано, что большая часть рекомендаций «Продавать» дается при ожидании снижения цен на 3%, а «Покупать» – при ожидании роста цен на 22%. Кроме того, было выявлено, что если рекомендации

относятся к акциям из отрасли потребительского сектора, то они более точны.

Для анализа согласованности (конфликтности) рекомендаций было предложено использовать аппарат теории функций доверия и, в частности, понятие меры конфликта. «Вклад» отдельных аналитиков в общий конфликт было предложено оценивать с помощью значений Шепли. Конфликтность рекомендаций была исследована как по инвестбанкам, так и по отдельным отраслям. В частности:

- показано, что конфликтность по банкам за рассматриваемый период была выше конфликтности по отраслям;
- выделены группы банков, имеющих высокую конфликтность рекомендаций;
- выделены группы банков, имеющих и высокую конфликтность рекомендаций и высокую точность прогнозов; представляется, что следование рекомендациям банков именно этой группы может увеличить доходность;
- выделены отдельные отрасли, в которых рекомендации аналитиков имеют высокую конфликтность.

Одно из направлений дальнейшей работы может быть связано с учетом рекомендаций для выработки оптимальной инвестиционной стратегии. Ожидается, что большую доходность даст следование рекомендациям «конфликтных» банков, имеющих высокую точность рекомендаций и прогнозов. Аналогично, предпочтительными секторами для инвестирования могут быть сектора, где есть высокий конфликт рекомендаций и высокая доходность. Другое направление – исследование возможности комбинирования определенных прогнозов и рекомендаций для получения более точных прогнозов.

Результаты исследования этого подраздела отражены в работах [185, 191].

## 4 Исследования моделей и методов извлечения знаний из текстов

В этом разделе отчета приведены результаты исследований моделей и методов извлечения знаний из текстов.

### 4.1 Способы формального представления текстов и их приложения

Формальное представление текста – это математическая структура, построенная по неструктурированному тексту [192, 193]. Формальным представлением текста может быть матричная структура, графовая структура, комбинация распределений. Чаще всего говорят о формальном представлении большого числа – коллекции / корпуса – текстов, поскольку представление одного текста с помощью математической конструкции не представляет особого интереса. Напротив, представление каждого текста из коллекции с помощью одной и той же конструкции делает возможным использование математических методов для обработки, анализа, сравнения, определения сходства между текстами, классификации, кластеризации, генерации текстов и так далее. В этой главе будут рассмотрены четыре основных класса представления текстов: векторная модель, языковая модель, модели скрытых тем и модели суффиксных деревьев. Исторически первая векторная модель представления текста имеет наибольшее количество применений, однако некоторые ее недостатки (например, не учитывается порядок слов) делают ее невозможным использование в тех задачах, в которых необходимо генерировать фрагмент текста или оценить вероятность его появления. В таком случае используются генеративные модели представления текста, такие как языковая модель и некоторые модели скрытых тем, основанные на скрытом размещении Дирихле.

И векторная модель, и языковая модель, и модель скрытых тем основаны на общей идеи: текст является набором, так называемых термов – слов в исходном виде или их значимых фрагментов, например, основ. Отсюда следует общий недостаток всех перечисленных моделей: при обработке и анализе текстов учитывается только четкое совпадение между термами. Модель

суффиксных деревьев – менее популярная в силу невысокой вычислительной эффективности – до определенной степени позволяет учитывать нечеткие совпадения, что делает возможным ее использование в задачах интерпретации текстов.

#### 4.1.1 Векторная модель представления текста

Векторная модель – это одна из наиболее популярных моделей представления текста [193]. В основе этой модели лежит так называемый мешок слов – принцип максимального упрощения структуры текста [194]. Согласно этому принципу, текст является множеством или мультимножеством входящих в него слов. Очевидно, что использование этого принципа ведет к потере порядка слов, а, следовательно, и коротких, и длинных, в том числе, анафорических и кореферентных связей [192].

Согласно векторной модели текст представляется множеством слов (или каких-нибудь других элементов текста, так называемых, термов), причём каждому терму соответствует своя координата векторного пространства. В качестве значения используется частота терма в тексте. Если в общем пространстве слов представляют два или более текстов – так называемую коллекцию текстов – в качестве часто используют  $tf - idf$  кодировку значений вектора, равную количеству вхождений терма в данный текст, делённому на логарифм относительного количества текстов, содержащих это слово [195]:

$$tf - idf = tf_{w,d} \log \frac{|D|}{|\{d' \in D | w \in d'\}|}.$$

В этой формуле первый сомножитель  $tf_{w,d}$  – это локальный вес, то есть, частота терма  $w$  в тексте  $d$ , а второй сомножитель  $idf$  – это глобальный вес, равный логарифму от величины, обратной доле документов  $d'$ , содержащих терм  $w$  среди общего числа текстов  $|D|$ .  $tf - idf$  кодировка снижает вес часто встречающихся во всех текстах коллекции термов и повышает вес термов, характерных для данного текста.

Иногда формулу  $tf - idf$  весов меняют, сохраняя при этом общий смысл: первый множитель – локальный вес – отвечает за выбор частотных слов в данном тексте, второй множитель – глобальный вес – за отсеивание слов, одинаково частотных во всей коллекции. Таким образом, общая схема взвешивания устроена так:  $w_{ij} = l_{ij} \times g_i$  [195].

Некоторые другие возможные локальные веса представлены в работах [196]:

- бинарный вес  $l_{ij} = 0$ , если терм  $i$  не встречается в тексте  $j$ , и  $l_{ij} = 1$  в противном случае;
- частота  $l_{ij} = tf_{ij}$ ;
- логарифмический вес  $l_{ij} = \log(tf_{ij} + 1)$ ;
- скорректированный Гауссов вес  $l_{ij} = \frac{tf_{ij}}{2\max_i(tf_{ij})} + 0.5$ .

Некоторые глобальные веса:

- бинарный вес  $g_i = 1$ , если  $i$ -ый терм встретился в коллекции, и  $g_i = 0$ , в противном случае;
- гауссов вес  $g_i = \frac{1}{\sqrt{\sum_j tf_{ij}^2}}$ ;
- Gf Idf вес  $g_i = \frac{gf_i}{df_i}$ , где  $gf_i$  – сколько раз  $i$ -ый терм встретился во всей коллекции, а  $df_i$  – число текстов, в которых встретился  $i$ -ый терм;
- Idf вес  $g_i = \log \frac{N}{1 + df_i}$ ;
- энтропия  $g_i = 1 - \sum_j \frac{p_{ij} \log p_{ij}}{\log N}$ , где  $p_{ij} = \frac{tf_{ij}}{gf_i}$ .

В [197] рассматривалась следующая схема взвешивания

$$w_{ij} = \log(1 + tf_i) \log \left( \frac{N+1}{n(t_i)+1} \right),$$

которая получила название  $tf - icf$  (term frequency – inverse corpus frequency).

Основным достоинством векторной модели считается ее простота и тот факт, что векторное представление текстов делает возможным использование линейно-алгебраических операций для определения сходства между текстами и ранжирования текстов по соответству запросу [198]. Для этих целей используется косинусная мера релевантности, которая будет описана более подробно ниже. В общих чертах косинусная мера близости определяется как нормированная скалярное произведение [195]. Другим очевидным достоинством векторной модели является простота ее построения по заданному корпусу текстов [199]. Во многих современных библиотеках автоматической обработки текстов, таких как gensim [200] и NLTK [201] реализованы индексаторы коллекций текстов на основе векторной модели – функции, задающие как координаты векторного пространства (т.е. выделяющие термы), так и соответствующие каждому тексту.

Однако за простотой векторной модели кроются некоторые недостатки. Прежде всего, главная предпосылка векторной модели, а именно мешок слов, со статистической точки зрения означает гипотезу о независимости двух слов, что в корне не верно [199]. Использование нормированного скалярного произведения в качестве меры сходства приводит к тому, что более длинные документы всегда имеют низкую степень сходства с остальными документами из-за нормировки длинной документа [202]. Главным недостатком векторной модели является отсутствия учета синонимии между двумя словами [202]: в векторной модели словам «Голландия» и «Нидерланды» будут соответствовать различные координаты, поэтому векторная модель не учтет их синонимичность.

Тем не менее, векторная модель широко используется во многих задачах автоматической обработки текстов: категоризации, классификации и кластеризации текстов, а так же в задаче информационного поиска. Задача информационного поиска была исторически первой, в которой была использована векторная модель [195]. Задача категоризации заключается в распределении текстов по заранее заданному множеству категорий. Как правило, задача ка-

тегоризации текстов решается с помощью методов машинного обучения. Исчерпывающий обзор подходов к решению этой задачи приведен в [203]. В этом обзоре показано, в том числе, что представление коллекции текстов в виде общей матрицы терм–текст делает возможным использование любого метода машинного обучения. В статьях [204] и [205] впервые возникает задача классификации текстов по тональности, которая заключается в том, чтобы определить имеет текст положительную или отрицательную окраску. Чаще всего, речь идет об отзывах на что-либо [204]. В таком случае возникает потребность понять, остался ли пользователь доволен или нет. Эта задача тоже решается с помощью методов машинного обучения. На вход методу машинного обучения поступает стандартная матрица терм–текст, при этом, методам из обучающей выборки приписан либо положительный, либо отрицательный класс. В более поздних работах по классификации текстов по тональности используются вспомогательные ресурсы, такие как WordNet [206]. Согласно, [202] методы кластеризации текстов востребованы, в основном, в поисковых системах для улучшения результатов поиска или сжатого представления найденных по запросу текстов. Обзор методов кластеризации текстов [207] показывает, что чаще всего используются либо методы иерархического кластерного анализа, либо метод k-Means и его модификации. На вход этим методам подается стандартная матрица терм–текст, по которой и находятся кластеры.

Благодаря своей популярности векторная модель получила несколько направлений развития. К ним относится обобщенная векторная модель (generalized vector space model, GVSM) [208], векторные модели семантики [209 – 211] и, в некотором смысле, вероятностная модель релевантности [212] и модели скрытых тем [213, 214], которые заслуживают отдельного рассмотрения.

Обобщенная векторная модель [208] позволяет уйти от принципа попарной независимости термов и учсть попарную корреляцию между векторами, соответствующими термам, в новом пространстве большей размерности. В исходном случае в векторной модели пространство имеет размерность  $n$ ,

равную числу термов. В обобщенной векторной модели рассматривается пространство размерности  $2^n$ . В этом пространстве базис задается векторами  $m_k$ , каждый из которых соответствует конъюктивному одночлену

$$f(t_1^{\delta_1}) \cap f(t_2^{\delta_2}) \cap \dots \cap f(t_n^{\delta_n}), \text{ где } f(t_1^{\delta_1}) = \begin{cases} f(t_1) = D_{t_1}, & \text{если } \delta_t = 1 \\ f(\bar{t}_1) = \overline{D_{t_1}}, & \text{если } \delta_t = 0 \end{cases}$$

где  $D_t$  – множество текстов, содержащих терм  $t$ , а  $\delta_t$  определяет отрицание переменной  $t$ . В [208] предложен вычислительный алгоритм построения  $2^n$  таких конъюктивных одночленов по матрице смежности термов. Таким образом, в обобщенной векторной модели текст представляется вектором в пространстве, образованном  $2^n$  векторами, имеющими смысл связей между термами.

Векторные модели семантики основаны на гипотезе, сформулированной в [194]. Согласно этой гипотезе, слова, встречающиеся в одинаковом контексте, имеют одинаковый смысл. В основе большинства векторных моделей семантики лежит матрица терм-терм, построенная по аналогии с традиционной матрицей терм-текст [199]. В [210], например, матрица терм-терм строится исключительно для существительных, встречающихся рядом друг с другом в окне  $+/- 2$ . Для построения такой матрицы необходимо, во-первых, извлечь все существительные из текста в том порядке, в котором они встречаются в тексте, во-вторых, расположить их в этом же порядке по строкам и столбцам матрицы, в третьих, убрать все стоп-слова из текста (под стоп-словами понимаются предлоги, артикли и местоимения). К значению в клетке матрицы добавляется единица, если слов по строке встречается среди двух слов справа или слева от слова по столбцу в тексте, очищенном от стоп-слов. Далее такая матрица используется для поиска синонимов: согласно косинусной мере, которая будет описана ниже, определяются пары близких друг другу векторов – столбцов матрицы, соответствующих разным словам. В [210] этот алгоритм был протестирован на материалах экзамена TOEFL. В результате применения этого алгоритма синонимов было найдено порядок 90% пар синонимичных пар слов из заданий на поиск синонимов. Анализу таких мат-

риц посвящены работы [211] и [209]. В первой работе используются методы ансамбли классификаторов, а во второй – собственный метод кластеризации, названный «Комитеты кластеризации» (“Clustering By Committee”) для выделения групп синонимов. Таким образом, в этом случае векторная модель используется для представления смысла слова вектором других слов.

Некоторым упрощением векторной модели является бинарная модель независимости (“Binary Independent Model”), разработанная авторами векторной модели [215]. Ее основное отличие от исходной модели заключается в том, что значения матрицы терм-текст являются бинарными и показывают, встречается ли терм в документе или нет. Это отличие оказывается существенным и позволяет использовать Байесовский принцип для определения релевантности строки тексту, которая будет описана ниже.

На векторной модели представления текста основана и вероятностная модель релевантности, предложенная в [212]. Эта модель релевантности используется, в основном, в задаче поиска по запросу. Согласно этой модели слова в тексте не независимы, а распределены по смеси Пуассоновских распределений. Тем не менее, и запрос, и документ, следуя векторной модели, представляется вектором частот в пространстве слов.

#### 4.1.2 Языковая модель представления текста

Языковая модель (language model) [216] позволяет оценить вероятность появления последовательности слов в тексте. В отличие от векторной и вероятностных моделей, языковая модель является генеративной [202]. В этой модели текст представляется с помощью цепей Маркова, где каждому узлу соответствует одно слово, а на ребрах – вероятности того, что одно слово встретится после другого. Модель считается генеративной, поскольку позволяет сгенерировать искусственный текст. Обратимся к двум наиболее востребованным видам языковых моделей: модели униграмм (одиночных слов) и модели биграм (последовательных пар слов).

При использовании языковых моделей нет необходимости в формальном представлении всего текста. Говорят о вероятности текста, или о вероятности появления его фрагмента – последовательности слов. Так же как и векторная модель, модель униграм основана на предположении о независимости появления слова в тексте от предыдущего слова. Согласно модели униграмм, текст – это последовательный набор слов  $t_i$ , причем вероятность всего текста равна произведению

$$P(t_{1,n}) = P(t_1, t_2, \dots, t_n) = \prod_i P(t_i),$$

то есть, произведению вероятностей появления каждого слова по отдельности. В модели биграм вероятность появления слова зависит от вероятности появления предшествующего слова:  $P(t_i|t_1, t_2, \dots, t_{i-1}) \approx P(t_i|t_{i-1})$ . Таким образом, в модели биграм учитывается локальный контекст слова. Тогда вероятность всего текста

$$P(t_{1,n}) = P(t_1, t_2, \dots, t_n) = p(t_1) \prod_i P(t_i|t_{i-1}).$$

Следуя принципу максимального правдоподобия, такая вероятность может быть оценена как  $P(t_i|t_{i-1}) = \frac{n(t_{i-1}, t_i)}{n(t_{i-1})}$ , где  $n(t_{i-1}, t_i)$  – сколько раз пара слов  $t_{i-1}, t_i$  встретилась в тексте, а  $n(t_{i-1})$  – сколько раз слово  $t_{i-1}$  встретилось в тексте вообще. Языковые модели используются в тех случаях, когда важно сохранить длинные семантические связи: в задачах машинного перевода [217], распознавания речи [218], исправлении опечаток [219]. Данная работа посвящена задачам другого рода, поэтому мы не будем в дальнейшем заострять внимание на языковых моделях.

#### 4.1.3 Представление текста на основе моделей скрытых тем

Тематические модели – это класс моделей, объединенных общим предположением о существовании скрытых (латентных) тем. Допустим, есть коллекция текстов. В этих текстах отражено некоторое количество тем. Темы представляются набором слов, а документы – набором тем. Каждый доку-

мент характеризуется вектором, составленным из оценок степени принадлежности документа к различным темам. Каждая тема представляет собой вектор, состоящий из оценок степени принадлежности слова к данной теме [220]. Одна из первых тематических моделей – это латентно-семантический анализ (или латентно-семантическая индексация) [213]. Главное новшество латентно-семантического анализа заключается не в математических построениях, а в интерпретации получаемых результатов. Латентно-семантический анализ основан на следующем принципе: слова, похожие по смыслу, встречаются в похожих контекстах. Похожесть контекстов может быть установлена с помощью сингулярного разложения матриц. Пусть  $X$  – матрица слово (или любая его модификация, терм) – текст. Строки в этой матрице соответствует термам, столбцы – текстам. Значения матрицы показывают, как часто встречается терм в тексте. Заметим, что иногда в матрицу записывают не частоты, а  $tf-idf$  веса термов. Положим, что матрица  $X$  – это произведение трех матриц:

$$X_{t \times d} = U_{t \times n} \Sigma_{n \times n} (V_{n \times d})^T,$$

где  $t$  – число термов,  $d$  – число текстов,  $n = \min(t, d)$ ,  $\text{rank}(A) = r$ ,  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ , матрицы  $U, V$  – ортогональны,  $\sigma_i > 0$ , если  $1 \leq i \leq r$ ,  $\sigma_i > 0$ , если  $i \geq r + 1$ . Матрица  $U$  представляет термы, матрица  $V$  представляет тексты, диагональная матрица  $\Sigma$  – сингулярные значения. Сингулярное разложение матрицы позволяет получить приближенное представление ранга  $k$  исходной матрицы:

$$\hat{X}_{t \times d} = U_{t \times k} \Sigma_{k \times k} (V_{d \times k})^T.$$

В новой матрице  $\hat{X}$  слова и тексты представлены векторами размерности  $k$  значительно меньшей, чем исходная. Этот факт позволяет интерпретировать столбцы матрицы как группы близких по смыслу слов, т.е. скрытые темы, а строки матрицы – как представления текстов в новом семантическом пространстве [221]. Модель латентного семантического анализа относят к

классу векторных моделей, однако другие модели скрытых тем имеют более сложную структуру и уходят от векторного представления текстов [192].

К таким моделям относится модель вероятностного латентного семантического анализа [222]. Она основана на предположении о том, что каждый текст является смесью, так называемых тем, причем каждая тема задается собственным распределением слов.

В основе модели вероятностного латентного семантического анализа лежит вероятностная модель коллекции текстов:

$$p(d, w) = \sum_{t \in T} p(t)p(d|t)p(w|d, t), \quad (4.1)$$

где  $D$  – множество текстов,  $T$  – множество тем,  $d$  – один текст, состоящий из слов  $w$ . Для численного решения уравнения используется ЕМ-алгоритм, на каждом шаге которого оцениваются параметры модели  $p(t), p(d|t), p(w|d, t)$ .

Модель вероятностного латентного семантического анализа получила широкое распространение. Она используется в тех случаях, в которых требуется оценить скрытые переменные, связующие две явные. Например, в задаче колаборативной фильтрации в качестве скрытое переменной может выступать переменная, соответствующая классу пользователей, а через нее связаны пользовательские сообщества и модели поведения пользователей [223]. Аналогично, в задаче персонификации поиска в Интернете скрытые переменные, связывающие данные о пользователях и их запросы в поисковой системе, строятся на основе истории поведения пользователя [224].

Латентное размещение Дирихле является генеративной моделью, так же как и языковая модель. Так же как вероятностный латентный семантический анализ, латентное размещение Дирихле основано на уравнении (4.1). Каждый документ представляется смесью тем, причем вероятности тем распределены по закону Дирихле. Каждая тема состоит набора слов (термов) и вероятностей, что данное слово относится к этой теме. Вероятность слова принадлежать к теме описывается так же законом Дирихле. Генерация корпуса текстов  $D$ , состоящего из  $M$  текстов длиной  $N_i$ , устроена так. Пусть распределение тем в тексте  $i$ :  $\theta_i \sim Dir(\alpha), 1 \leq i \leq M$ ,  $Dir(\alpha)$  – распределение Дирихле с па-

раметром  $\alpha$ , а распределение слов в теме  $k$ :  $\varphi_k \sim Dir(\beta)$ ,  $1 \leq k \leq K$  – число тем. Для каждого позиции слова  $i, j$ ,  $1 \leq i \leq M, 1 \leq j \leq M$ :

- а) выбрать тему  $z_{ij} \sim Multinomial(\theta_i)$ ;
- б) выбрать слово  $w_{ij} \sim Multinomial(\varphi_{z_{ij}})$ ,

где  $Multinomial$  – мультиномиальное распределение с одним исходом.

Вероятностные тематические модели получили широкое распространение и используются в задачах информационного поиска (поиска по запросу) [225, 226], классификации текстов [227, 228], автоматического рефериования текстов [229, 230], фильтрации спама [231, 232], а так же в других областях, не связанных с автоматической обработкой текстов, таких как колаборативная фильтрация [223, 233], анализ изображений [234, 235].

В задаче поиска по запросу латентный семантический анализ может быть использован для снижения размерности. Допустим, исходная матрица терм-текст  $X$  имела размерность  $t \times d$ , а запрос был представлен вектором  $q$  из  $t$  компонент. После использования сингулярного матрицы  $X$ , вектор запроса может быть преобразован как  $\hat{q} = q^T U_k \Sigma_k^{-1}$ , после чего используется косинусная мера близости (которая будет описана ниже) для поиска ближайших  $\hat{q}$  столбцов матрицы  $X_k$ , соответствующих документам. Такой поиск по запросу дает результаты точнее, чем поиск по лексическому совпадению, поскольку учитывает скрытые отношения между термами и документами. В [226] формулируется вероятностная генеративная модель, позволяющая по аналогии с моделью языка оценить вероятность генерации одного текста и вероятность появления запроса в тексте.

В задаче классификации текстов латентно-семантический анализ так же может быть использован для снижения размерности. В [227] для классификации текстов на два класса используется метод ближайшего соседа и машины опорных векторов. Утверждается, что использование латентно-семантического анализа для аппроксимации исходной матрицы терм-текст матрицей меньшего ранга позволяет значительно сократить объем вычисле-

ний при незначительной (порядка 2-3%) потере в аккуратности. Адаптация метода латентного размещения Дирихле на случай заранее известных тем, предложенная в [228], носит название labeled LDA. Предполагается, что количество тем, существующих в зафиксированной коллекции текстов, известно заранее, при этом, известно, к какой теме или каким темам относится каждый текст. Примером такой коллекции может служить коллекция сообщений в блогах, помеченных различными тегами-метками. Предложенная в [228] генеративная модель такой коллекции текстов и основанный на ней классификатор превосходит машины опорных векторов, которые обычно используются для подобных задач классификации.

Сравнение векторной модели и модели скрытых тем на основе латентно-семантического анализа в задаче автоматического реферирования текстов проводится в [229]. Рефератом текста считается набор из фиксированного числа предложений из текста, наиболее полно отражающий его содержаний. Предложен следующий алгоритм суммаризации текста:

- 1) разбить исходный текст на множество предложений кандидатов  $S$ ;
- 2) в пространстве всех слов для каждого предложения составить свой вектор  $A_i$  и общий вектор  $D$  для всего текста (следуя принципам векторной модели);
- 3) найти близость каждого вектора  $A_i$  вектору  $D$  по косинусной мере близости, которая будет описана ниже;
- 4) выбрать предложение  $S_k$  соответствующее вектору  $A_k$ , наиболее близкому вектору  $D$ ;  $S_k$  будет входить в реферат; если достигнуто искомое число предложений в реферате, то алгоритм останавливается; иначе переходит на шаг 5;
- 5) исключить из рассмотрения все термы, входящие в  $S_k$ ; составить представления предложений и текста в новом пространстве термов; перейти на шаг 3.

Для использования латентно-семантического анализа предложена следующая модификация этого алгоритма:

- 1) разбить исходный текст на множество предложений кандидатов  $S$ , а предложения – на множество термов;
- 2) создать матрицу терм – предложение  $A$ ;
- 3) выполнить сингулярное разложение  $A = U\Sigma V^T$ , где столбцы правой сингулярной матрицы  $V^T$  отвечают предложениям:  $\psi_i = [\nu_{i1}, \dots, \nu_{ir}]^T$  – вектор-столбец, соответствующий предложению  $i$ ;
- 4) выбрать  $-t$ ый столбец правой матрицы сингулярной матрицы  $V^T$ ;
- 5) выбрать предложение, соответствующее максимальному значению выбранного  $-t$ ого столбца правой матрицы сингулярной матрицы  $V^T$ ; согласно гипотезе авторов статей, это предложение будет соответствовать  $k$ -той скрытой теме, т.е. его необходимо включить в реферат исходного текста; если достигнуто искомое число предложений в реферате, алгоритм останавливается, иначе переходит на шаг 4.

Показано, что вторая версия алгоритма незначительно превосходит первую.

В [230] предложено использовать латентное размещение Дирихле для автоматического рефериования текста. Согласно предложенному алгоритму, для построения реферата необходимо:

- 1) найти скрытые темы в тексте, используя латентное размещение Дирихле;
- 2) оценить вероятность порождения каждого предложения каждой темой;
- 3) выбрать наиболее вероятное предложение из каждой темы; если предложение уже входит в состав реферата, выбрать второе по вероятности.

Существующие методы фильтрации спама позволяют достичь высокой точности при сравнительно невысокой полноте [231]. В этой же статье [231] показано, что использование скрытых тем, полученных с помощью латентно-семантического анализа в качестве признаков для обучения трех разных классификаторов и ансамбля классификаторов, позволяет сохранить точность

на высоком уровне и повысить полноту. Однако автор отмечает важный недостаток предложенного метода, который затрудняет его использование в системах фильтрации спама: латентно-семантический анализ не является интерактивным методом, то есть, при появлении нового текста в коллекции необходимо заново формировать матрицу терм – документ и заново вычислять сингулярные матрицы и матрицу сингулярных значений. В [232] предложен метод разделения коллекции текстов на две части в соответствии с предположением о том, что одна часть коллекции является спамом, а вторая – нет. Авторы использовали размеченную на спам и не-спам коллекцию текстов UK2007-WEBSLAM. На обеих частях коллекции было использовано латентное размещение Дирихле для поиска скрытых тем. Распределения тем получаются разные, несмотря на то, что слова, формирующие темы присутствуют в обеих частях коллекции. Использование найденных скрытых тем в качестве признаков для классификации по признаку спам/не-спам позволяет получить результаты на 10% превосходящие по F-мере другие известные методы, примененные к этой же коллекции текстов.

#### 4.1.4 Представление текста на основе модели суффиксных деревьев

Суффиксные деревья были предложены в статье [236] в качестве средства для поиска нечетких совпадений между строками. Суффиксные деревья получили широкое распространение в биоинформатике, где они используются, в основном, для поиска закономерностей в ДНК или в белках, которые записаны длинной последовательностью символов [237]. Иногда суффиксные деревья используются для индексации текстов [238] и реже – в качестве моделей представления текста. Впервые суффиксные деревья в качестве модели представления текстовой коллекции были предложены в [239], где суффиксное дерево, построенное по коллекции текстов, послужило основой для поиска кластеров – т.е. групп близких по смыслу – текстов. Структура суффиксного дерева, использованного в [239] несколько отличалась от классического определения суффиксного дерева, и предполагала определенный способ ан-

анотирования суффиксного дерева порядковыми номерами текстов в коллекции, в результате чего стало возможным использование суффиксного дерева для поиска кластеров в коллекции текстов. Другие работы, в которых суффиксные деревья использованы в качестве модели представления коллекции текстов, так же предполагают тот или иной способ аннотирования дерева номерами текстов или частотами [240 – 242].

Согласно [243], суффиксное дерево для -символьной строки  $S$  представляет собой ориентированное дерево с корнем, имеющее ровно  $m$  листьев, за- нумерованных от 1 до  $m$ . Каждая внутренняя вершина, отличная от корня, имеет не меньше двух детей, а каждая строка помечена непустой подстрокой строки  $S$ . Никакие две дуги, выходящие из одной и той же вершины, не могут иметь пометок, начинающихся с одного и того же символа. Главная особенность суффиксного дерева заключается в том, что для каждого листа  $i$  конката- низация меток дуг на пути от корня к листу  $i$  в точности составляет / произ-носит / кодирует / прочитывает суффикс строки  $S$ , который начинается в по-зиции  $i$ , то есть,  $S[i:m]$ .

Там же приводится пример суффиксного дерева для строки  $S = "xabxac"$  (см. рисунок 4.1).

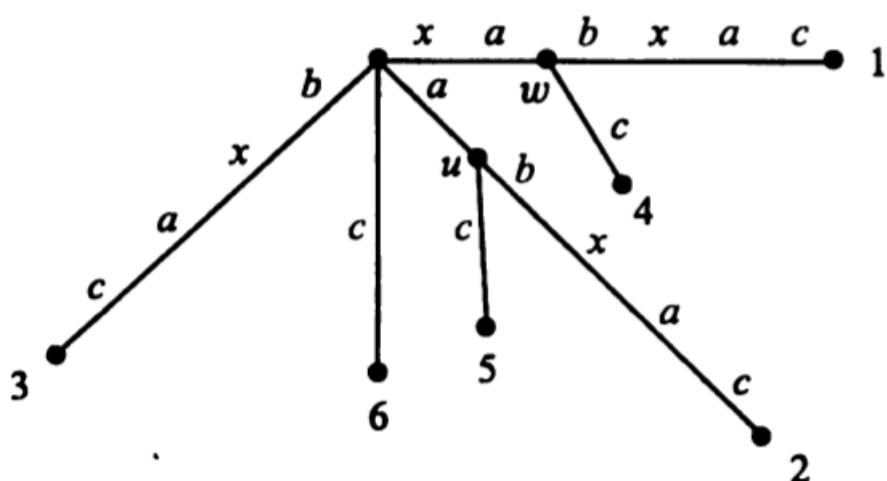


Рисунок 4.1 – Суффиксное дерево для строки  $S = "xabxac"$  [243]

Путь до листа 1 прочитывает первый суффикс строки  $S[1:6] = "xabxac"$ , путь до листа 2 – второй суффикс  $S[2:6] = "abxac"$ , путь до листа

3 – третий суффикс  $S[3:6] = "bxac"$  и т.д. Путь до листа 6 прочитывает последний шестой суффикс  $S[6:6] = "c"$ . В таком дереве всего две внутренние вершины: они помечены буквами  $u$  и  $w$ . Заметим, что с точки зрения частотного анализа, наличие внутренних вершин в таком суффиксом дереве означает, что путь от корня до внутренней вершины прочитывает повторяющийся фрагмент строки. Так, фрагмент " $xa$ " встречается в строке  $S = "habxas"$  дважды:  $S[1:2] = S[4:5] = "xa"$ , аналогично фрагмент " $a$ " встречается в строке  $S = "habxas"$  дважды:  $S[2] = S[5] = "a"$ . Тем не менее, такое представление суффиксного дерева не позволяет учитывать, сколько именно раз встречается во входной строке повторяющийся фрагмент.

Обобщенным суффиксным деревом для набора / коллекции строк называется суффиксное дерево, в котором представлено более одной строки.

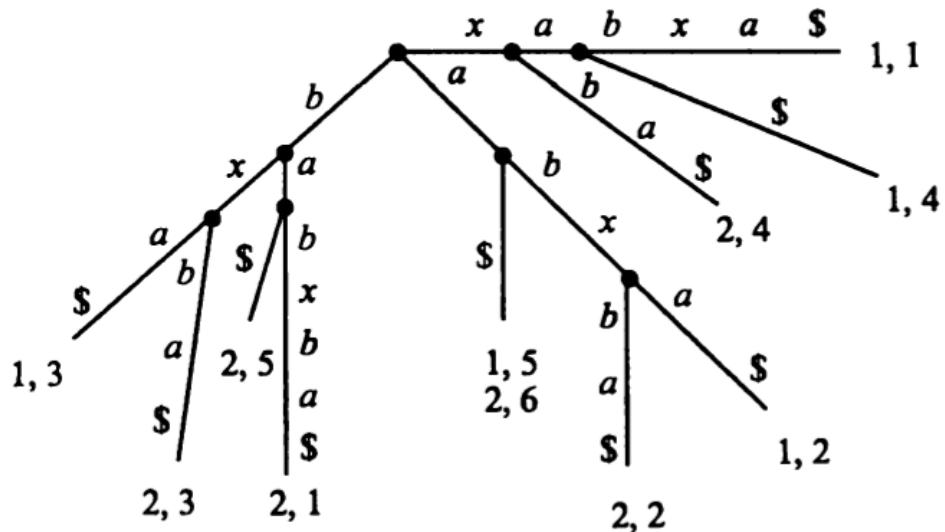


Рисунок 4.2 – Обобщенное суффиксное дерево для двух строк  
 $S_1 = "habxa", S_2 = "babxba"$  [243]

Представленное на рисунке 4.2 суффиксное дерево построено по двум строкам  $S_1 = "habxa", S_2 = "babxba"$ . Каждому листу в нем приписана своя метка – одна или две пары чисел: первое число в паре означает номер строки, второе – номер суффикса. Символ "\$" использован в качестве терминального и добавлен к концу каждой строки. Нетрудно убедиться, что пути от корня до внутренних вершин кодируют фрагменты, повторяющиеся во входных строках. Так, например, фрагмент " $a$ " встречается во входных строках четыре

раза: дважды в качестве последнего суффикса строк –  $S_1[5]$  и  $S_2[6]$  и дважды в качестве префикса суффиксов  $S_1[2:5]$  и  $S_2[2:6]$ . Заметим, что количество повторений фрагмента "a" можно установить по количеству меток у листьев, которые покрывает соответствующая внутренняя вершина.

Естественным развитием модели суффиксного дерева представляется модель аннотированного суффиксного дерева (АСД), в которой частоты всех фрагментов присутствуют явно [244]. В этой работе аннотированное суффиксное дерево определяется как суффиксное дерево, в котором:

- символы стоят не на ребрах, а в узлах;
- каждому узлу соответствует один символ;
- каждый узел помечен частотой фрагмента, который прочитывает путь от корня до этого узла;
- опущены терминальные символы и метки листьев, представляющие номер суффикса и входной строки.

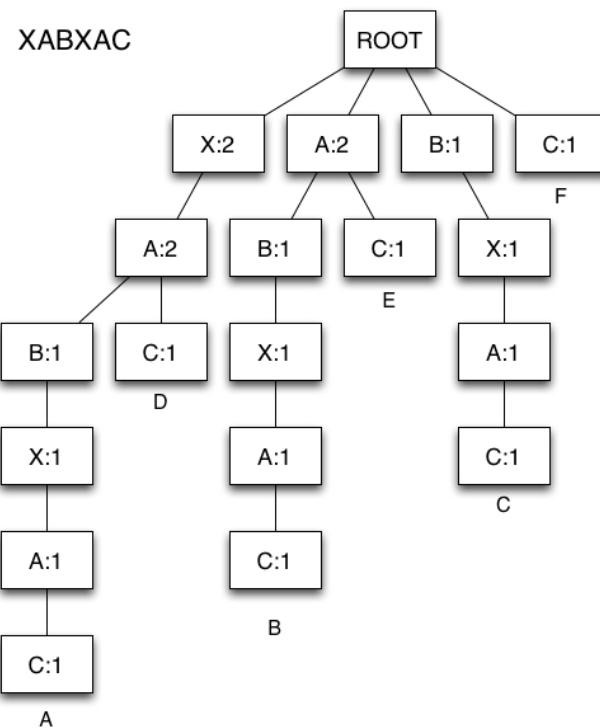


Рисунок 4.3 – Аннотированное суффиксное дерево для строки  $S$  – ки  $S = "xabxas"$ .

По аналогии с [243] построим АСД для строки  $S = "habxas"$  (рисунок 4.3). У этой строки шесть суффиксов:  $S[1: 6] = "habxas"$  – этому суффиксу соответствует цепочка узлов, помеченная буквой A,  $S[2: 6] = "abxas"$  (B),  $S[3: 6] = "bxas"$  (C),  $S[4: 6] = "xas"$  (D),  $S[5: 6] = "ac"$  (E),  $S[6] = "c"$  (F). Фрагменты " $xa$ ", " $a$ " входной строки  $S = "habxas"$  встречаются дважды:  $S[1: 2] = S[4: 5] = "xa"$ , поэтому частоты соответствующих им узлов равны 2, а частоты остальных узлов равны 1, поскольку они представляют уникальные фрагменты строки.

Продолжая аналогию с [243], построим обобщенное суффиксное дерево (рисунок 4.4) для двух строк  $S_1 = "xabxac"$ ,  $S_2 = "babxac"$ .

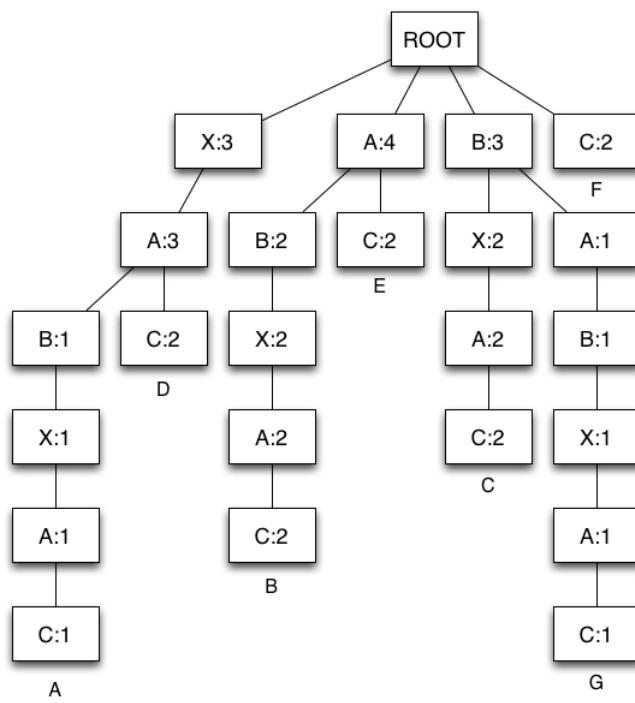


Рисунок 4.4 – Обобщенное аннотированное суффиксное дерево для строк  $S_1 = "xabxac"$ ,  $S_2 = "babxac"$

За исключением первого суффиксы этих строк совпадают:  $S_1[1:6] = "xabxac" \neq S_2 = "babxac"$ . Эти несовпадающие суффиксы представлены цепочками узлов A и G. Частоты почти всех узлов в этих строках составляют 1. От 1 отличаются только частоты первых узлов этих цепочек "xa", "b", поскольку они встречаются по 3 раза во входных строках. Остальные суффиксы

представлены такими же узлами, как на рисунке 4.3, но с удвоенными частотами.

В [244] показано два важных свойства АСД:

- свойство 1: частота любого узла равна сумме частот его узлов-детей, так как родительский узел соответствует префиксу нескольких суффиксов, и его частота складывается из частот этих суффиксов;
- свойство 2: частота родительского узла равна сумме частот листьев, которые он покрывает.

Рассмотрим алгоритм суффиксного дерева и его адаптацию на случай аннотированного суффиксного дерева.

### Наивный алгоритм построения суффиксного дерева

Пусть на вход поступает строка  $S$  из  $m$  символов, по которой требуется построить суффиксное дерево. Следуя [243], поместим сначала в дерево простую дугу для первого суффикса  $S[1:m]\$$ , затем последовательно добавим в растущее дерево суффиксы  $S[i:m]\$$ ,  $2 \leq i \leq m$ .

Построение суффиксного дерева для строки  $S$ .

1. Инициализация: создаем пустую структуру, в которой будет храниться АСД. Обозначим ее  $st$ . Далее итеративно будем добавлять в  $st$  суффиксы входной строки.
2. Добавим дерево простую дугу для первого суффикса  $S[1:m]\$$ .
3. На  $i$ -той итерации алгоритма,  $i = 2, \dots, m$ , для суффикса  $S[i:m]$  будем искать в  $st$  самый длинный путь от корня, метка которого совпадает с префиксом  $S[i:m]$ . Такой путь будет единственным, так как никакие две дуги, выходящие из одной вершины, не могут иметь одинаковых меток. Пусть совпадение найдено. Тогда
  - а) либо алгоритм останавливается в некоторой внутренней вершине  $w$ ;
  - б) либо алгоритм останавливается в середине дуги  $(u, v)$ ; тогда разобьем дугу  $(u, v)$  на две новые дуги и добавим новую вершину  $w$ .

4. Создадим новую дугу  $(w, i)$ , идущую из  $w$  в лист  $i$  и эту дугу пометим частью суффикса  $S[i:m]$ , не нашедшим совпадения.

### Наивный алгоритм построения аннотированного суффиксного дерева

В [244] использован наивный алгоритм построения аннотированных суффиксных деревьев. Предлагается разбить текст на последовательные фрагменты длины 2, 4, 6 и 8 и использовать эти фрагменты в качестве коллекции строк для построения АСД. Таким образом, можно построить АСД заведомо ограниченное по глубине.

Предварительная подготовка текста заключается в формировании фрагментов фиксированной длины, начинающихся с 1, 2, 3 и т.д. символа текста. Обозначим такие фрагменты через  $f_1, \dots, f_N$ . Длина всех фрагментов одинакова и составляет  $l$ , поэтому не возникает необходимости добавлять терминальный символ к концам фрагментов.

Построение АСД для коллекции фрагментов  $f_1, \dots, f_N$ .

1. Инициализация: создаем пустую структуру, в которой будет храниться АСД. Обозначим ее  $ast$ . Далее итеративно будем добавлять в  $ast$  фрагменты из входной коллекции.
2. На  $i$ -той итерации алгоритма,  $i = 1, 2, \dots, N$ , для фрагмента  $s = s_i$  длины  $l$ :
  - а) находим все суффиксы  $s[j:]$ , где  $j = 1, 2, \dots, l$ ;
  - б) для каждого суффикса  $s[j:]$  ищем в  $ast$  совпадение – путь от корня, совпадающий с максимальным префиксом суффикса  $s[j:]$ ;
  - в) пусть найдено совпадение  $m = s[j:k]$ , где  $k \leq l$ ; для узлов, попавших в совпадение, увеличиваем частоты на 1;
  - г) если  $k < l$ , то требуется создать новые узлы для фрагмента строки  $s[k+1:l]$ . Для этого создаем у последнего узла в найденном совпадении нового потомка, помечаем его символом  $s[k+1]$  и при-

писываем ему частоту 1. Таким же образом последовательно создаем узлы для всех оставшихся символов во фрагменте. В результате будет создана новая цепочка узлов, кодирующая текущий суффикс. Если  $k = 1$ , новые узлы не создаются.

Как пример, построим АСД для строки  $s = \text{'mining'}$  (рисунок 4.5). Для суффиксов первых трех суффиксов 'mining', 'ining', 'ning' и последнего суффикса 'g' совпадений не будет найдено. Поэтому в дерево будут добавлены соответствующие цепочки с частотами равными 1. При добавлении четвертого суффикса  $s[4:] = \text{'ing'}$  будет найдено непустое совпадение 'in', поэтому частота узлов из совпадения будет увеличена на 1, а у узла с меткой 'n' будет создан новый потомок с меткой 'g' и частотой 1. Аналогично, при добавлении суффикса  $s[5:] = \text{'ng'}$  будет найдено совпадение из одного узла с меткой 'n'. Следуя алгоритму, частота узла будет увеличена на 1 и у него будет создан новый потомок с меткой 'g' и частотой 1.

Если к уже построенному для строки  $s = \text{'mining'}$  АСД требуется добавить строку  $t = \text{'dining'}$ , то для первого суффикса  $t[1:] = \text{'dining'}$  не будет найдено совпадений, поэтому будет создана новая цепочка узлов с частотами 1. Для всех остальных суффиксов строки  $t$  будут найдены совпадения, полностью покрывающие суффиксы, поэтому у всех узлов в дереве частоты будут увеличены вдвое, но новых узлов создано не будет.

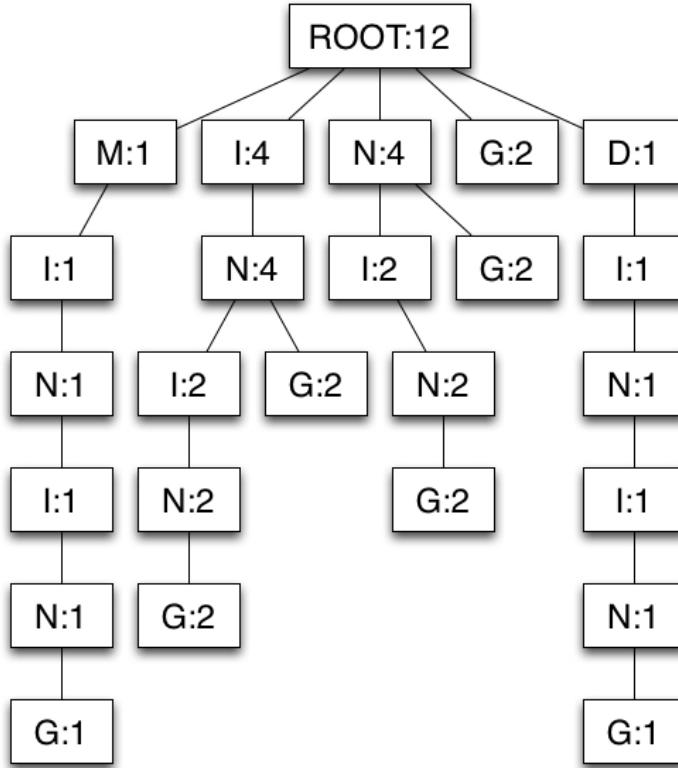


Рисунок 4.5 – АСД для двух строк 'mining' и 'dining'

Опишем наивный алгоритм построения АСД на псевдокоде и оценим его сложность по времени и по памяти, следуя [245].

Алгоритм NaïveConstruction( $C$ )

Вход. Коллекция фрагментов  $C = \{f_1, \dots, f_m\}$ .

Выход. АСД для  $C$ .

1. **for**  $i \leftarrow 1$  **to**  $m$
2.     **for**  $j \leftarrow 1$  **to**  $n_i = |s_i|$
3.         **do**  $k \leftarrow$  длина совпадения  $s_i[j:]$  с АСД
4.         **for** узел  $u$  из  $s_i[j:(j+k-1)]$
5.             **do** присвоить  $f(u) \leftarrow f(u) + 1$
6.         **for**  $l \leftarrow j+k$  **to**  $n_i$
7.             **do** вставить узел  $v$
8.                 присвоить  $f(v) \leftarrow 1$

Анализ сложности по времени и памяти данного алгоритма достаточно прост. Перебирая строки, мы посимвольно просматриваем все ее суффиксы,

затрачивая на  $i$ -ю строку длины  $n_i$  количество операций, пропорциональное  $(1 + 2 + \dots + n_i) = \Theta(n_i^2)$ . Общее время работы алгоритма для коллекции из строк, таким образом, может быть оценено как  $\Theta(n_1^2 + \dots + n_m^2)$ , или, если использовать более грубую оценку,  $O(mn_{max}^2)$ . Отметим, что определенное способом выше АСД невозможно построить с использованием меньшего числа операций, так как само оно занимает в памяти место, квадратично зависящее от длины закодированных в нем строк.

## 4.2 Оценивание релевантности строки тексту с использованием метода аннотированного суффиксного дерева (АСД)

### 4.2.1 Проблема оценивания релевантности строки тексту и основные подходы к ее решению

Проблема оценивания релевантности строки тексту формулируется следующим образом. Пусть дана некоторая коллекция текстов и некоторая строка. Под строкой понимается, как правило, одно слово или согласованное словосочетание. Требуется определить, насколько релевантна строка тексту из коллекции, то есть, встречается она или ее фрагменты в тексты, и, если да, насколько строка соответствует содержанию текста. Другими словами, необходимо ранжировать документы по степени релевантности им данной строки. Численные оценки релевантности тому или иному документу сами по себе не имеют смысла и интересны исключительно со сравнительной точки зрения, какому документу более релевантна строка. Понятие релевантности имеет двойственный характер. С одной стороны, чаще всего термин релевантность возникает в области информационного поиска. В задаче поиска требуется показать пользователю релевантные его запросу (строке) документы (тексты). Говорят, что релевантные документы – это такие документы, которые удовлетворяют информационные нужды пользователя [246, 247]. С другой стороны, в формальных моделях релевантность строки тексту определяется по близости между математическими представлениями строки и

текста или по вероятности появления строки в тексте. Как правило, в таких моделях отсутствует пользователь и его информационные нужды. Они появляются позже, в качестве надстройки над математическими моделями и учитывают не непосредственные характеристики строки или текста, а их контекст, время, место появления и другие свойства [248]. В этой работе мы будем обращаться исключительно к математическим мерам релевантности. Перечислим основные математические модели и меры релевантности.

#### 4.2.2 Релевантность в векторной модели

Релевантность строки вектору определяется в векторной модели так: строка представляется вектором в том же многомерном пространстве, а релевантность определяется по сходству двух построенных векторов.

Пусть дана строка *string* и коллекция текстов, в которой  $N$  текстов. Требуется определить релевантность  $\text{relevance}(\text{string}, \text{text})$  строки одному из текстов *text* из коллекции. Для начала зададим координаты векторного пространства:  $V$  – словарь, содержащий все слова коллекции,  $t \in V$  – слова. Каждому слову  $t$  соответствует своя координата. Тогда тексту соответствует вектор  $w = (w_1, \dots, w_{|V|})$ . Компоненты вектора – это либо частоты слов, либо  $tf - idf$  веса, которые вычисляются по формуле  $w_{ij} = tf \times idf = tf_{ij} \times \log \frac{N}{n(t_i) + 1}$ , где  $tf_{ij}$  – частота слова  $t_i$  в тексте  $j$ ,  $n(t_i)$  – число текстов, содержащих слово  $t_i$ ,  $N$  – количество текстов. Каждому тексту соответствует вектор в пространстве слов. Размерность этого вектора совпадает с количеством различных слов во всех текстах из коллекции. Составим аналогичный вектор для строки с использованием  $tf - idf$  весов для слов из строки. Релевантность строки тексту определяется через косинусную меру близости между соответствующими векторами:

$$\text{relevance}(\text{string}, \text{text}) = \cos(\overrightarrow{\text{string}}, \overrightarrow{\text{text}}) = \frac{\overrightarrow{\text{string}} \times \overrightarrow{\text{text}}}{\|\overrightarrow{\text{string}}\| \times \|\overrightarrow{\text{text}}\|}.$$

#### 4.2.3 Релевантность в бинарной модели независимости

Релевантность строки тексту в бинарной модели независимости определяется по следующему Байесовскому правилу:

$$P(R|\overrightarrow{string}, \overrightarrow{text}) = \frac{P(\overrightarrow{string}|R, \overrightarrow{text}) * P(R|\overrightarrow{text})}{P(\overrightarrow{string}|\overrightarrow{text})},$$

где  $R$  – бинарная переменная, принимающая два значения, 1, если строка релевантна тексту и 0 в обратном случае. Следовательно,  $P(\overrightarrow{string}|R = 1, \overrightarrow{text})$  и  $P(\overrightarrow{string}|R = 0, \overrightarrow{text})$  – вероятности того, что строка релевантна тексту и того, что строка нерелевантна тексту, соответственно. Заметим, что  $P(\overrightarrow{string}|R = 1, \overrightarrow{text}) + P(\overrightarrow{string}|R = 0, \overrightarrow{text}) = 1$ .

#### 4.2.4 Релевантность в вероятностной модели

Вероятностная модель представления текста используется, в основном, в задачах извлечения/поиска информации и сформулирована в терминах задачи поиска по запросу. Она основана на теоретическом принципе ранжирования вероятностей (“Probability Ranking Principle”, PRP) [212]: наиболее эффективная поисковая машина выдает тексты пользователю в соответствии с убыванием вероятности релевантности его запросу. Здесь под релевантностью понимается соответствие содержания текста запросу (более широкое понимание понятия релевантности будет изложено ниже). Предполагается, что релевантность  $Rel$  – это случайная величина, которая принимает бинарные значения:  $rel$  – если запрос релевантен документу,  $\overline{rel}$  – в обратном случае.

Она построена в предположениях теоретической модели, согласно которой каждый текстовый документ представляется как смесь двух Пуассоновских распределений [212]. Одно из них отвечает за распределение обычных слов, другое – за распределение «элитных» слов, то есть, тех, на которых лежит основная смысловая нагрузка в разрезе рассматриваемой тематики. Обычно тематика задаётся тем запросом на извлечение информации, относи-

тельно которого и оценивается релевантность. Релевантность строки тексту в этой модели определяется по вероятности того, что слова, принадлежащие строке, окажутся элитными в тексте.

Следуя векторной модели представления текстов, вероятностная модель релевантности предполагает, что строка и текст – два набора слов. Ставшая очень популярной в последнее время мера релевантности BM25 придаёт больший вес значимым словам и меньший – незначимым:

$$\text{relevance}(\text{string}, \text{text}) = \sum_{i=1}^N \text{IDF}(t_i) \frac{(k_1 + 1)t f_i}{t f_i + k_1(1 - b + b \frac{|A|}{avgdl})},$$

где  $avgdl$  – среднее количество слов в тексте, а  $b$ ,  $k_1$  – константы, равные, как правило, 1.5 и 0.75, соответственно, согласно [212].

В качестве нормализующего сомножителя используется функция IDF, имеющая следующий вид:  $\text{IDF}(t_i) = \log \frac{N - n(t_i) + 0.5}{n(t_i) + 0.5}$ , где  $N$  – количество текстов в коллекции, а  $n(t_i)$  – число текстов, содержащих слово  $t_i$ . Функция IDF имеет смысл обратной частоты: чем больше документов содержат данное слово, тем менее он значим.

#### 4.2.5 Релевантность в тематических моделях

Релевантность строки тексту в модели латентно-семантического анализа определяется следующим образом. Пусть для строки определен вектор частот слов  $string$  в исходном пространстве слов. Представим в его новом пространстве меньшей размерности:  $\widehat{string} = \Sigma_k^{-1} U_k^T string$ . Релевантность строки тексту определяется по косинусной мере близости между преобразованным вектором, соответствующим строке, и вектором, соответствующим тексту, т.е. столбцу в матрице  $\widehat{X}$ .

В генеративных моделях представления текста, таких как языковая модель или модель латентного размещения Дирихле, релевантность строки тексту составляет вероятность порождения текстом строки, то есть,

$P(string|text) = \prod_{t \in string} P(t|text)$ , где  $t$  – термы, из которых состоит строка  $string$ .

В [249] предложена схема оценивания вероятности одного терма в тексте для всего класса генеративных моделей:

$$P(t|text) = \frac{N}{N+\mu} P_{ML}(t|text) + \left(1 - \frac{N}{N+\mu}\right) P_{ML}(t|coll),$$

где  $N$  – количество текстов в коллекции,  $t$  – терм,  $text$  – текст,  $\mu$  – параметр распределения Дирихле, а  $P_{ML}(t|text)$  – вероятность терма в тексте, полученная по методу максимального правдоподобия, а  $P_{ML}(t|coll)$  – вероятность терма во всей коллекции текстов.

Для случая латентного размещения Дирихле [226] предложена следующая модификация этой формулы:

$$P(t|text) = \lambda \left( \frac{N}{N+\mu} P_{ML}(t|text) + \frac{\mu}{N+\mu} P_{ML}(t|coll) \right) + (1 - \lambda) P_{LDA}(t|text),$$

где  $P_{LDA}(t|text, \hat{\theta}, \hat{\varphi}) = \sum_{z=1}^K P(t|z, \hat{\varphi}) P(z|\hat{\theta}, d)$ .  $\hat{\theta}, \hat{\varphi}$  – апостериорные оценки параметров  $\theta, \varphi$  модели латентного размещения Дирихле,  $z$  – скрытые темы. Далее предлагается использовать сэмплирование Гиббса для численного оценивания необходимых параметров.

#### 4.2.6 Релевантность в модели аннотированного суффиксного дерева

Оценивание релевантности строки тексту с использованием АСД предполагает построение АСД для текста и последующее наложение строки на текст.

##### Оценивание релевантности строки тексту

Каждый текст представляется собственным АСД, с которым сличается строка для вычисления оценок релевантности. Оценка релевантности  $relevance(string, text)$  строки  $string$  тексту  $text$  вычисляется следующим образом.

1. Выделяются все суффиксы строки  $string$

2. Для каждого суффикса вычисляется оценка его совпадения (match) с АСД:

$$\text{score}(\text{match}(\text{suffix}, \text{ast})) = \sum_{\text{node} \in \text{match}} \varphi\left(\frac{\text{f}(\text{node})}{\text{f}(\text{parent}(\text{node}))}\right) \quad (4.2)$$

где совпадение – это путь от корня дерева, кодирующий совпадающий с ним префикс суффикса или суффикс целиком,  $\text{f}(\text{node})$  – частота, приписанная узлу АСД из совпадения,  $\text{f}(\text{parent}(\text{node}))$  – частота, приписанная родителю данного узла.

3. Оценка релевантности вычисляется как среднее всех оценок, приходящееся на один символ:

$$\begin{aligned} \text{relevance}(\text{string}, \text{text}) &= \text{SCORE}(\text{string}, \text{text}) = \\ &= \frac{1}{|\text{string}|} \sum_{\text{suffix}} \frac{\text{score}(\text{match}(\text{suffix}, \text{ast}))}{|\text{suffix}|}, \end{aligned} \quad (4.3)$$

где  $|\text{suffix}|$ ,  $|\text{string}|$  – количество символов в текущем суффиксе и в строке.

В этой формуле score – это шкалирующая функция, переводящая оценку совпадения в уровень релевантности. Рассмотрим три вида шкалирующей функции  $\varphi(x)$ , рекомендованных в [244] на основе экспериментов по категоризации электронной почты:

- линейная (обозначение – linear),  $\varphi(x) = x$ ;
- логистическая (обозначение – logit),  $\varphi(x) = \log \frac{x}{1-x} = \log(x) - \log(1-x)$ ;
- квадратный корень (обозначение – root),  $\varphi(x) = \sqrt{x}$ .

Из этих трёх только линейная, ничего не меняющая функция, имеет очевидный операциональный смысл – средней условной вероятности символа в совпадении (СУВСС); две нелинейные шкалы из [244] могут быть использованы для контроля.

Приведем пример вычисления оценки релевантности. Оценка релевантности строки “dine” при использовании линейной шкалирующей функции для АСД, изображенного на рисунке 4.5, равна, согласно вышеприведённому определению:

$$sim(dine, mining) = \frac{\left[ score(dine, ast)/_4 + score(ine, ast)/_3 + score(ne, ast)/_2 + score(e, ast) \right]}{4} = 0 +$$

$$\frac{\phi\left(\frac{2}{6}\right) + \phi\left(\frac{2}{2}\right)}{3*4} + \frac{\phi\left(\frac{2}{6}\right)}{2*4} + 0 = \frac{\frac{1}{3} + 1}{12} + \frac{\frac{1}{3}}{8} = \frac{11}{72}.$$

Обозначим очистку АСД от шума через  $\phi(X)$ , где  $\phi$  – вид шкалирующей функции, а  $X$  – уровень в АСД, до которого обнулялись частоты.

### 4.3 Метод nAST-k оценивания релевантности строки тексту с использованием нормированного АСД

Метод nAST-k используется для оценивания релевантности строки (или коллекции строк) тексту (коллекции текстов).

Метод nAST-k имеет несколько радикальных отличий от метода аннотированного суффиксного дерева, описанного в [244]. Во-первых, используется другой способ подготовки текстов: текст представляется набором строк нефиксированной длины, а не набором фрагментов. Во-вторых, используется нормированная оценка релевантности. В-третьих, метод nAST-k предусматривает параметризацию АСД, в том числе, выбор процедуры очистки АСД от шума. В-четвертых, для АСД используется алгоритм, имеющий линейную сложность по времени.

#### 4.3.1 Структура метода

##### 4.3.1.1 Подготовка текстов к обработке

Подготовка текстов к обработке проводится согласно стандартной схеме, представленной в [193]: удаление xml- и html-разметки, если она присутствует в тексте, токенизация, удаление знаков препинания и прочих символов, включая цифры и псевдографику, приведение всех слов к нижнему регистру. Под токенизацией мы понимаем процедуру последовательного разбиения текста на предложения и на слова.

Обработанный текст представляет собой последовательность строк. В [250] экспериментально показано, что глубина дерева, построенного по строкам из 2-4 слов, вполне достаточна в задаче интерпретации текстов. Таким образом, текст после обработки состоит из строк из 2-4 слов, соединенных через пробел. Строки строятся следующим образом: первая строка начинается с первого слова в тексте и заканчивается 2-4 словом в тексте, вторая начинается со второго и заканчивается соответственно на 3-5 слове. Например, если первая строка обработанного текста такова: “слово1 слово2 слово3”, то вторая строка текста будет такой: “слово2 слово3 слово4”. При этом учитываются границы предложений: в одну строку не должны попадать слова из разных предложений.

### Параметризация АСД

Рассмотрим три параметра АСД: глубину, уровень очистки от шума и размах. Эти параметры представляют некоторые свойства АСД, изменяя которые можно повысить эффективность метода оценивания релевантности строки тексту.

### Глубина АСД

Глубина АСД – это количество узлов в максимальной по длине цепочке. Очевидно, что глубина АСД определяется количеством символов в самой длинной цепочке. Глубина АСД не должна превосходить длину строк, которые на него накладываются (с поправкой на 1-10 символов) и не должна быть сильно меньше длины строк, которые на него накладываются. При разбиении текста на строки следует учитывать, что от длины строк зависит глубина АСД и выбирать длину строки, т.е. количество слов в строке разумным способом. Например, если заранее известно, что большая часть строк-словосочетаний состоят из трех слов, то и текст следует разбивать на строки из трех слов, как это сделано в [250].

## Уровень очистки АСД от шума

В больших АСД на первых уровнях располагается довольно большое число узлов с относительно высокими частотами, которые дают примерно одинаково большой вклад в оценку любой строки, накладываемой на АСД. Первый уровень характеризует частоты отдельных символов, второй – частоты упорядоченных пар символов, третий – частоты упорядоченных троек и т.д. Очевидно, что такие короткие элементы текста не могут нести особой тематической направленности. Поэтому возникает гипотеза, что вклады узлов начальных уровней АСД в оценки релевантности носят характер шума, и оценка релевантности станет более адекватной, если ее очистить от вклада узлов начальных уровней. Для проверки этой гипотезы мы обнуляли частоты узлов на первом, втором и т.д.

## Размах АСД

Под размахом АСД понимается среднее количество потомков у одного узла. Чем больше цепочек в АСД, не имеющих разветвлений, тем меньше размах АСД.

Определим размах следующим образом:

$$range = \frac{\sum_{node \in Parent} \text{количество потомков у узла } node}{|Parent|},$$

где *Parent* – множество узлов, у которых есть потомки (т.е. которые не являются листьями). Так, например, размах АСД, показанного на рисунке 4.5 составляет 1.18.

АСД, построенные по разным коллекциям строк, могут отличаться между собой по размаху. Чем больше и разнообразнее коллекция строк, тем больше размах соответствующего АСД.

#### 4.3.1.2 Нормирование оценки релевантности

Часто возникает потребность сравнить оценки сходства строк с двумя или более АСД. Получаемые оценки могут сильно зависеть от размаха АСД. Чем больше узлов в АСД, тем больше разброс оценок, получаемых при слиянии строк с этим деревом. Для того чтобы сделать оценки по разным деревьям сравнимыми между собой, модифицируем формулы (4.2) и (4.3) так, чтобы нормировать результаты по длине фактических совпадений:

$$\text{score}(\text{match}(\text{suffix}, \text{ast})) = \frac{\sum_{\text{node} \in \text{match}} \varphi\left(\frac{f(\text{node})}{f(\text{parent}(\text{node}))}\right)}{k}, \quad (4.2')$$

где  $k$  – количество символов в найденном совпадении  $\text{match}$ . Тогда общая оценка

$$\begin{aligned} \text{relevance}(\text{string}, \text{text}) &= \text{SCORE}(\text{string}, \text{text}) = \\ &= \frac{1}{|\text{string}|} \sum_{\text{suffix}} \frac{1}{|\text{suffix}|} \text{score}(\text{match}(\text{suffix}, \text{ast})) \end{aligned} \quad (4.3')$$

в случае линейной  $\varphi$  имеет смысл условной вероятности, приходящейся на одну букву суффикса в совпадении  $\text{match}$ . Это делает оценки сравнимыми как по документам, так и по словосочетаниям.

#### 4.3.1.3 Распространение линейных алгоритмов построения суффиксных деревьев на случай АСД

Определенное выше аннотированное суффиксное дерево, строго говоря, не является суффиксным деревом, поскольку не обладает одним из основных свойств суффиксных деревьев, приведенных в [243]. В суффиксном дереве у каждой внутренней вершины, отличной от корня, должно быть не менее двух детей, а в АСД в большом количестве присутствуют цепочки узлов, т.е. узлы с единственным потомком. В [245] показано преобразование АСД, необходимые для выполнения этого условия, а именно схлопывание вершин. Схлопывание вершин заключается в объединение каждой цепи узлов с единственным потомком в одну вершину и переносе пометки на входящее в нее ребро. Метка ребра получается конкатенацией символов, которыми были помечены

узлы в цепи. Частота самой вершины остается неизменной, так как у всех вершин в цепи она была одинаковой. Преобразовав таким образом исходное дерево, получим структуру данных, изображенную на рисунке 4.6 (АСД построено для строки ‘mining’).

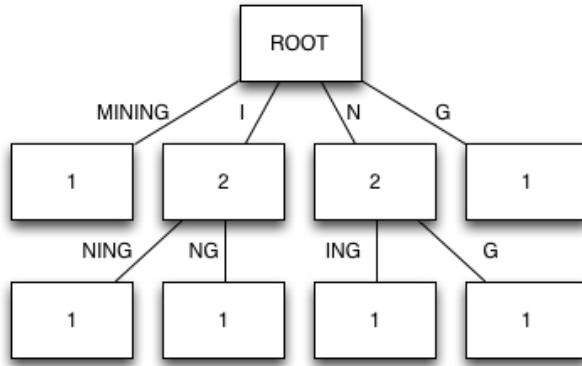


Рисунок 4.6 Оптимизация представления АСД. Схлопывание вершин

Реализованное таким образом АСД требует  $\Theta(mn_{max}^2)$  памяти из-за необходимости хранить все метки ребер в явном виде. Если использовать прием сжатия дуговых меток, описанный в [243], то получится снизить объем используемой деревом памяти до линейного относительно длин строк в коллекции. Прием сжатия дуговых меток заключается в том, чтобы хранить в каждом ребре только индексы начала и конца соответствующей подстроки, а не всю подстроку в явном виде. Окончательный вид АСД после всех описанных оптимизаций представлен на рисунке 4.7.

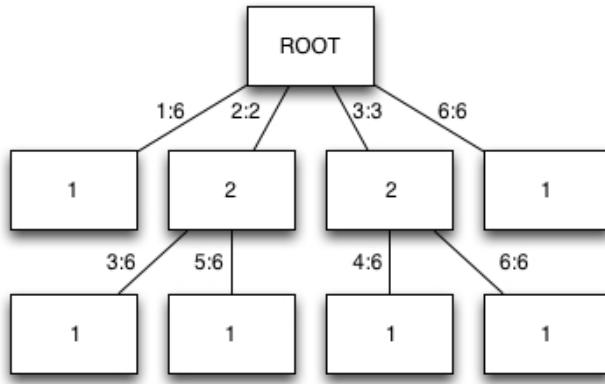


Рисунок 4.7 Оптимизация представления АСД. Сжатие меток

Значительное снижение объема используемой для хранения АСД памяти позволяет использовать асимптотически менее трудоемкие алгоритмы для построения АСД, чем наивный алгоритм, который был описан выше. Существует целый ряд линейных по времени алгоритмов построения обычных (неаннотированных) суффиксных деревьев. В [243] приведен обзор основных алгоритмов: алгоритмов П. Вайнера (1973), Э. МакКрейга (1976) и Э. Укконена (1995). Использование этих алгоритмов для построение АСД становится возможным благодаря Свойству 1 АСД: сначала построим обычное суффиксное дерево, затем, во время обхода снизу, аннотируем его частотами. После несложной обработки входных строк частоты листьев будут равны 1, а затем частоты узлов на уровнях выше получаются как сумма частот потомков. Предобработка строк заключается в добавлении уникального терминального символа в конец каждой строки. Будем обозначать терминальные символы через  $s_i$ . Таким образом, каждая подстрока, соответствующая одному из путей от корня до листа, встречается в исходном наборе строк только один раз. Приведем псевдокод такого алгоритма построения АСД и покажем, что построение АСД осуществляется за линейное время.

Алгоритм LinearConstruction( $C$ )

Вход. Коллекция строк  $C = \{s_1, \dots, s_m\}$ .

Выход. Обобщенное АСД для  $C$ .

1. Построить  $C' = \{s_1 \$_1, \dots, s_m \$_m\}$
2. Построить обобщенное суффиксное дерево  $T$  для коллекции  $C'$ , используя алгоритм с линейной сложностью.
3. **for**  $l$  **in**  $leaves(T)$
4. **do** присвоить  $f(l) \leftarrow 1$
5. Выполнить обход дерева  $T$  снизу вверх; в каждом внутреннем узле  $v$  присвоить  $f(v) \leftarrow \sum_{u \in T: parent(u)=v} f(u)$ .

Обход дерева требует времени, пропорционального числу вершин. Таким образом, все шаги алгоритма выполняются за линейное время, и общая оценка его трудоемкости составляет  $\Theta(n_1 + \dots + n_m)$  или  $O(mn_{max})$ .

Заметим, что при представлении АСД, показанном на рисунке 4.7, необходимо видоизменить формулы (4.2) или (4.2'):

$$\text{score}(\text{match}(\text{suffix}, \text{ast})) = \sum_{node \in \text{match}} \varphi\left(\frac{f(node)}{f(\text{parent}(node))}\right) + |\text{match}| \quad (4.2'')$$

или

$$\text{score}(\text{match}(\text{suffix}, \text{ast})) = \frac{\sum_{node \in \text{match}} \varphi\left(\frac{f(node)}{f(\text{parent}(node))}\right) + |\text{match}| - k}{|\text{match}|}, \quad (4.2'')$$

где  $|\text{match}|$  – длина фактического совпадения. Заметим, что при таком преобразовании смысл формул останется без изменения.

В [245] приведен эксперимент, в ходе которого сравнивается эффективность обоих алгоритмов на стандартных тестовых коллекциях текстов. Показано, что, действительно, линейный алгоритм работает быстрее, чем наивный. Сложность линейного алгоритма по памяти не отличается от сложности наивного алгоритма, поскольку тем или иным образом приходится хранить в памяти одинаковое количество символов и их частот.

#### 4.3.2 Построение таблицы релевантности Стока-Текст (рСТ)

С помощью метода nAST-k построим таблицу релевантности «строка-таблица» (рСТ таблица), в которой строки соответствуют отдельным строкам (ключевым словосочетаниям), столбцы – отдельным текстам, а элементы – оценки релевантности строк-словосочетаний соответствующим документам. Основное отличие рСТ таблицы от традиционного построения матрица терм–документ, многократно использованного в учебниках [192, 193], заключается в том, что рСТ таблица строится для зафиксированного множества строк-словосочетаний, а матрица терм–документ строится по термам, извлеченным из документов. Более того, элементами матрицы терм–документ, как правило, являются частоты, в то время как элементами рСТ таблицы являются оценки релевантности. Таким образом, рСТ таблицу, построенную для фиксированного множества строк-словосочетаний и фиксированной коллекции текстов, можно считать моделью данной коллекции текстов: каждый текст представляется вектором оценок релевантностей в пространстве строк-словосочетаний. Таблица 4.1 представляет фрагмент рСТ таблицы из [250].

Таблица 4.1 – Фрагмент рСТ таблицы: столбцы соответствуют публикациям, строки – словосочетаниям, а элементы – оценкам релевантности

	Доклад Всемирного Банка об экономике России	Международные стандарты финансовой отчетности	Если генеральный директор иностранец
Изменение организационно-правовой формы	0.3145	0.3616	0.3644
Изменение уровня концентрации собственности	0.5016	0.3148	0.2706
Повышение эффективности управления затратами	0.4433	0.2809	0.2445
Смена генерального директора	0.2264	0.2351	0.5947

Для построения рСТ таблицы с использованием метода nAST-k необходимо проделать следующие шаги.

1. Зафиксировать коллекцию текстов. Для того, что бы последующий анализ и интерпретация данной коллекции имел смысл, требуется сформировать однородную коллекцию текстов одинаковой стилистической и жанровой

специфики, принадлежащих к общей предметной области. Например, в [250] предметом анализа была коллекция новостных сообщений о бизнес-процессах в пост-кризисной России в 2009-2010 годах.

2. Зафиксировать множество строк–словосочетаний, описывающих основные события, явления и термины в той же предметной области. В [250] словосочетания были сформированы с помощью эксперта и описывали основные события в сфере бизнеса.
3. Для каждого текста построить собственное АСД, согласно методу описанному выше: каждый текст разбивают на строки из 2-4 слов, все множество строк подают на вход алгоритму построения АСД. В [250] были использованы строки из трех слов, поскольку большая часть строк–словосочетаний состояла из трех слов, так что глубина АСД получается близкой к длине словосочетаний, на него накладываемых .
4. На каждое АСД последовательно наложить все строки–словосочетания и получить оценки релевантности. Оценки релевантности сохранить в таблицу, которая и будет искомой рСТ таблицей.

Заметим, что, во-первых, строго говоря, рСТ таблица не должна храниться в памяти компьютера как таблица. Представление рСТ таблица в виде разреженной матрицы [251] вполне допустимо и оправдано с технической точки зрения: значения оценок релевантности часто не превосходят 0. Во-вторых, для построения рСТ таблицы может быть использована любая другая мера релевантности. Однако, мера релевантности, основанная на АСД, обладает некоторыми преимуществами. Она учитывает все нечеткие совпадения строки–словосочетания с текстом и дает им количественную оценку. Другие меры релевантности, в том числе, описанные выше, учитывают исключительно четкие совпадения между отдельными словами, составляющими словосочетание, и не могут дать оценку целой строке–словосочетанию.

#### 4.4 Некоторые выводы

Этот раздел отчета посвящен проблеме оценивания релевантности строки тексту. В подразделе 4.1 приведен обзор основных мер релевантности строки тексту и их теоретических обоснований: меры релевантности в векторной, вероятностной, языковой и тематических моделях. Все перечисленные методы обладают некоторыми общими свойствами: например, они учитывают только четкие совпадения строки с текстом и получаемые оценки релевантности не имеют арифметического или вероятностного смысла. Мера релевантности, основанная на аннотированных суффиксных деревьях [244] (АСД), преодолевает эти проблемы и позволяет учитывать все нечеткие совпадения строки с текстом. В подразделе 4.2 описан метод АСД, позволяющий получить оценки релевантности, которые имеют чёткий операциональный смысл – суммарной условной вероятности символа в совпавшем фрагменте. В подразделе 4.3 представлен метод оценивания релевантности строки тексту nAST-k, являющийся модификацией метода АСД. Этот метод учитывает такие параметры АСД, как глубина и разброс, что позволяет нормировать оценки.

Рассмотрены два алгоритма построения АСД: наивный алгоритм, имеющий квадратичную оценку сложности по времени, и линейный алгоритм, имеющий соответственно линейную оценку сложности по времени. Оба алгоритма не отличаются по сложности и по требуемой памяти.

Показано, что меру релевантности строки тексту, получаемую по методу nAST-k можно использовать для построения таблиц релевантности строка–текст (рСТ таблица) благодаря ее преимуществам перед другими мерами релевантности. Утверждается, что рСТ таблица является моделью коллекции текстов в фиксированном пространстве строк, где строки представляют основные понятия, термины, явления, темы и т.д., описанные в текстах коллекции. В комплексе программ EAST реализован полный функционал для построения рСТ таблиц с использованием меры релевантности, основанной на АСД.

Результаты представленных в этом разделе исследований отражены в работах [252 – 255].

## ЗАКЛЮЧЕНИЕ

При выполнении проекта получены следующие основные результаты, могут быть сделаны следующие выводы и рекомендации:

- предложены две математические модели для выбора стратегий инвестирования мелкими и средними участниками биржевых торгов, с использованием которых задачи поиска оптимальных стратегий этих игроков сформулированы в виде задач линейного программирования, смешанного математического программирования и задачи отыскания седловых точек в антагонистической игре на выпуклых многогранниках с платежной функцией, являющейся суммой линейной и билинейной функции векторных аргументов; доказана возможность отыскания верхней оценки значения указанной антагонистической игры на выпуклых многогранниках из решения задач линейного программирования, образующих двойственную пару;
- построена математическая модель для поиска оптимального плана модернизации грузовой транспортной системы посредством добавления в неё новых грузовых транспортных узлов, с использованием которой задача нахождения оптимального плана модернизации транспортной сети путем размещения грузовых транспортных узлов и добавления транспортных коммуникаций сформулирована как задача смешанного математического программирования;
- предложена новая математическая модель для отыскания взаимоприемлемых условий взаимодействия государства и частного сектора экономики по выполнению крупномасштабных проектов, которые государство не может финансировать в полном объеме; с использованием предложенной модели указанная задача формулируется в форме игры трех лиц на выпуклом многогранном множестве связанных стратегий игроков;
- разработана теоретико-игровая математическая модель для анализа экономической эффективности прямых контрактов между производителями и поставщиками электроэнергии на рынке мощности электроэнергии, с использо-

зованием которой перспективы инвестирования в строительство новых электростанций и хранилищ электроэнергии в регионе оцениваются в рамках бескоалиционной игры двух лиц, в которой одним из игроков является группа всех производителей региона, а другим – группа всех потребителей региона;

- в рамках исследования моделей пространственного взаимодействия проведен анализ интенсивности потенциальных конфликтов интересов стран в Арктике с использованием правила порогового агрегирования; исследованы различные сценарии развития событий в арктическом регионе;
- в рамках исследования моделей сетевого взаимодействия предложены новые методы определения степени важности узлов в сложных сетях, основанные на использовании правил коллективного выбора; разработанный подход применен к анализу эмпирических данных о денежных переводах среди 214 стран мира в 2014 году;
- разработана модель многоступенчатого выбора вуза абитуриентом с учетом популярности и престижности профессий, самооценки абитуриентами своих способностей и возможностей поступить в определенные вузы в соответствии с полученными баллами ЕГЭ и конкурсом в вузах;
- проведена серия вариативных расчетов по разработанным моделям на основе данных о баллах ЕГЭ выпускников 2013 года и индексов популярности профессий, рассчитанных на основе семантического анализа текстовых данных сети Интернет и СМИ;
- спрогнозирована и проанализирована структура приема в вузы в 2014–2016 гг. с учетом тенденций спроса на высшее образование на основе проведенных расчетов;
- разработан метод проверки альтернатив на потенциальную недоминируемость в дискретных многокритериальных задачах принятия решений, который позволяет построить множество потенциально недоминируемых альтернатив;

- в рамках сравнительного исследования и аксиоматического описания концепций решений в задаче оптимального коллективного выбора получены условия принадлежности альтернативы одному из минимальных а) Р-доминирующих множеств, б) Р-внешнеустойчивых множеств, в) R-внешнеустойчивых множеств;
- установлен ряд свойств для объединения минимальных Р-внешнеустойчивых множеств, объединения минимальных R-внешнеустойчивых множеств и объединения минимальных Р-доминирующих множеств;
- проведен сравнительный анализ аксиоматических свойств процедур агрегирования, использованных для решения задачи ранжирования научных журналов на основании значений нескольких библиометрических показателей; показано, что наилучшим по релевантным свойствам способом агрегирования базовых рейтингов является сортировка с помощью объединения минимальных Р-внешнеустойчивых множеств;
- исследованы новые двухступенчатые процедуры выбора на предмет удовлетворения существующим нормативным условиям (наследования, согласия, отбрасывания, аксиоме выбора Эрроу, монотонности, пороговой некомпенсируемости); исследована вычислительная сложность вновь рассмотренных двухступенчатых процедур выбора;
- исследована коалиционная манипулируемость в модели Impartial Anonymous Culture для трех альтернатив и семи различных правил коллективного выбора; выявлены наименее манипулируемая процедура (процедуры Нансона) и наиболее манипулируемая (процедура одобряющего голосование); проведено сравнение данных результатов с результатами предыдущих исследований (коалиционного манипулирования в Impartial Culture);
- проведен анализ межгрупповой и внутригрупповой согласованности Европарламента в 1979-2014гг. на данных о поименных голосованиях по семи

созывам с помощью индексов согласованности Райса и Сатарова, мер фрагментированности и поляризованности общества;

- проведен анализ распределения влияния в семи созывах Европарламента на данных о поименных голосованиях с использованием индексов влияния с учетом предпочтений по созданию коалиций; получены оценки влияния на уровне групп (европейских политических партий и стран), а также на уровне отдельных европарламентариев;
- исследована структура Верховной Рады Украины (2007–2012) с помощью анализа более тысяч поименных голосований депутатов; проведена оценка согласованности позиций фракций с помощью различных индексов;
- методами теории функций доверия проанализирована согласованность (конфликтность) и точность прогнозов инвестбанков о стоимости акций российских компаний;
- в рамках исследования моделей и методов извлечения знаний из текстов проанализированы различные модели представления текста и способы оценивания релевантности строк тексту в различных моделях; проанализирован метод оценивания релевантности строк тексту, основанный на применении нормированного аннотированного суффиксного дерева.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Отчет МНУЛ анализа и выбора решений, ТЗ-87 за 2014г.
- 2 Markowitz H. Portfolio Selection// Journal of Finance, 1952, VII(1), 77-91.
- 3 Odean T. Do investors trade too much?// American Economic Review, 1999, 89(5), 1279-1298.
- 4 Barber B., Odean T. Trading is hazardous to your wealth: The common stock investment performance of individual investors// Journal of Finance, 2000, 55(2), 773-806.
- 5 Malkiel B.G., Saha A. Hedge Funds: Risk and Return // Financial Analysts Journal, 2005, 61(6), 80-88.
- 6 Penikas H., Proskurin S. How Well Do Analysts Predict Stock Prices? Evidence from Russia// Working papers by NRU Higher School of Economics, series FE "Financial Economics", WP BRP 18/FE/2013.
- 7 Юдин Д.Б., Юдин А.Д. Экстремальные модели в экономике. – М.: URSS, 2009.
- 8 Асратьян А.С., Кузюрин Н.Н. Анализ точности вероятностного округления для задач целочисленного линейного программирования// Дискретная математика, 2004, 6(4), 3-13.
- 9 Беленький А.С. Минимаксные задачи планирования с линейными ограничениями и методы их решения// Автоматика и Телемеханика, 1981, 42(10), 1409-1419.
- 10 Belenky A.S., Egorova L.G. Optimization techniques for forming and managing portfolios of financial instruments traded in a stock exchange by small and medium price-taking traders// Optimization and Applications in Control and Data Science, Springer, 2016 (в печати).
- 11 Belenky A.S., Egorova L.G. Two approaches to modeling the interaction of small and medium price-taking traders with a stock exchange by mathematical programming techniques// препринт НИУ ВШЭ, 2016 (подготовлен к печати).

12 Беленький А.С., Болкунов Д.С. Теоретико-игровой подход к анализу перспектив развития системы электроснабжения региона// Управление большими системами, 2015, 57, 76-122.

13 Belenky A.S., Bolkunov D.S. A game-theoretic approach to analyzing potential investment opportunities in a regional electrical grid// Energy Systems, 2016 (сдана в печать).

14 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, 6(3), 389-415.

15 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, 6(2), 291-308.

16 Belenky A.S., Fedin G.G., Kornhauser A.L. Optimally locating a set of regional transport hubs with access roads to them// Transportation Systems, 2017 (подготовлена к публикации).

17 Belenky A.S., Fedin G.G., Kornhauser A.L. Optimization problems of finding optimally locations for freight transport hubs: a survey// Transportation or Transportation Research or Transportation Science, 2017 (подготовлена к публикации).

18 Belenky A.S. Game-theoretic models for analyzing two public-private partnership problems// Production and Operations Management, 2016 (подготовлена к публикации).

19 Belenky A.S. A Game-theoretic approach to solving a waste-management problem// Waste Management or Journal of Management, 2016 (подготовлена к публикации).

20 Беленький А.С., Егорова Л.Г. Две модели принятия решений участником торгов на фондовой бирже по формированию и изменению своего инвестиционного портфеля// Препринт ВШЭ, серия WP7, М.: ВШЭ, 2015, No.02.

- 21 Belenky A.S., Egorova L.G. An Approach to Forming and Managing a Portfolio of Financial Securities by Small and Medium Price-Taking Traders in a Stock Exchange// Advances in Intelligent Systems and Computing, 2015, 359, Springer, 257-268.
- 22 Koivurova T., Hossain K. Offshore Hydrocarbon: Current PolicyContext in the Marine Arctic// Arctic TRANSFORM, 2008.
- 23 Loeng H. Natural Resources in Polar Oceans// The European Polar Board and Horizont 2020, 2014.
- 24 Асмакова А.Т. Арктика как новый вектор во внешней политике Республики Корея// Материалы VI Международной студенческой электронной научной конференции «Студенческий научный форум», 2015.
- 25 National Capital Branch of the Canadian International Council. Interests and Roles of Non-Arctic States in The Arctic. – The Gordon Foundation, 2011.
- 26 House W. National strategy for the Arctic Region. – Washington: White House, 2013.
- 27 Zysk K. Russia's Arctic strategy: Ambitions and constraints//Joint Force Quarterly: JFQ, 2010, 57, 103.
- 28 Eurasia Group. Opportunities and Challenges for Arctic Oil and Gas Development, Washington, DC: The Wilson Center, 2013  
[http://www.wilsoncenter.org/sites/default/files/Artic%20Report\\_F2.pdf](http://www.wilsoncenter.org/sites/default/files/Artic%20Report_F2.pdf)
- 29 Arctic Wells. Timeline of the main historical milestones of oil and gas exploration in the Arctic region// Arctic Wells web-site, 2015 (<http://www.arctic-wells.com/history/>).
- 30 Clement J.P., Bengtson J.L., Kelly B.P. Managing for the future in a rapidly changing Arctic: A report to the President. – Interagency Working Group on Coordination of Domestic Energy Development and Permitting in Alaska, 2013.
- 31 Molenaar E.J, Corell R. Arctic Fisheries// Arctic TRANSFORM web-site, 2015. [Электронный ресурс]. – URL: <http://arctic-transform.org/download/FishBP.pdf> - 15.02.15

- 32 Humpert M., Rasputnik A. The Future of Arctic Shipping Port// Technology International, 2012, 55, 10–11.
- 33 Norwegian Shipowners' Association. High North – High Stakes. Maritime opportunities in the Arctic. Oslo, 2012.
- 34 Rainwater S. Racy to the North// Naval War College Review, 2013, 66(2), 62.
- 35 Smith L.C., Stephenson S.R. New Trans-Arctic shipping routes navigable by midcentury// Proceedings of the National Academy of Sciences, 2013, 110(13), E1191-E1195.
- 36 Aleskerov F.T., Chistyakov V.V. The threshold decision making effectuated by the enumerating preference function// International Journal of Information Technology & Decision Making, 2013, 12(6), 1201-1222.
- 37 Aleskerov F.T., Chistyakov V.V., Kalyagin V.A. Social threshold aggregations// Social Choice and Welfare, 2010, 35(4), 627-646.
- 38 Aleskerov F.T., Yakuba V.I. A method for threshold aggregation of three-grade rankings// Doklady Mathematics, 2007, 75(2), 322-324.
- 39 Copeland A.H. A ‘reasonable’ social welfare function// Notes from a seminar on applications of mathematics to the social sciences, University of Michigan, 1951.
- 40 Fishburn P. Condorcet social choice functions// SIAM J. Appl. Math., 1977, 33, 469-489.
- 41 Miller N. A new solution set for tournaments and majority voting: further graph-theoretical approaches to the theory of voting// Am. J. Pol. Sci., 1980, 24, 68-96.
- 42 Алескеров Ф.Т., Хабина Э.Л., Шварц Д.А. Бинарные отношения, графы и коллективные решения. – М.: Издательский дом ГУ-ВШЭ, 2006.
- 43 Ratha D., Shaw W. South-South migration and remittances// World Bank working paper, no. 102. 2007.

44 Балыхин Г.А., Суровов М.В., Маркова О.П., Коновалов В.В. Государственное задание на подготовку специалистов как механизм реализации государственной политики в области модернизации образования// Всероссийская научно-практическая интернет-конференция: Спрос и предложение на рынке труда и рынке образовательных услуг в регионах России: сб. докл. Кн. III. Петрозаводск: Петрозаводский гос. ун-т, 2005.

45 Карной М., Добрякова М.С., Фрумин И.Д. Массовое высшее образование. Триумф БРИК? / Пер. с англ. (под ред.: М.С. Добряковой). – М.: Издательский дом НИУ ВШЭ, 2014.

46 Abankina I.V., Scherbakova I. Russian Higher Education Reforms and the Bologna Process// Journal of the European Higher Education Area, 2013, 3, 3-25.

47 Руднев М.Г. Методология и основные результаты исследований престижа профессий в зарубежной социологии// Вопросы образования, 2008, 2, 217-239.

48 Бессуднов А.Р. Социально-профессиональный статус в современной России// Мир России, 2009, 89-115.

49 Новые смыслы в образовательных стратегиях молодежи: 50 лет исследования / Д.Л. Константиновский, М.А. Абрамова, Е.Д. Вознесенская и др. — М.: ЦСП и М, 2015.

50 Горшков М.К., Шереги Ф.Э. Молодежь России: социологический портрет. – М.: Центр социального прогнозирования и маркетинга, 2010.

51 ВЦИОМ. 2012. Пресс-выпуск №2124.

<http://wciom.ru/index.php?id=459&uid=113109>

52 ВЦИОМ. 2013. Пресс-выпуск №2420

<http://wciom.ru/index.php?id=459&uid=114530>

53 Абанкина И.В., Абанкина Т.В., Алескеров Ф.Т., Зиньковский К.В., Огороднийчук Д.Л., Филатова Л.М., Деркачев П.В., Николаенко Е.А., Се-роштан Э.С., Егорова Л.Г. Модель многоступенчатого выбора для прогнози-

рования поведения спроса на высшее образование// Университетское управление: практика и анализ, 2014, 4-5(92-93), 84-94.

54 Fishburn P.C. Decision and value theory. – New York: Wiley, 1964.

55 Подиновский В.В. О взаимосвязи понятий потенциальной оптимальности и недоминируемости// Автоматика и телемеханика, 2012, 1, 184-187.

56 Podinovski V.V. Non-dominance and potential optimality for partial preference relations// European journal of operational research, 2013, 229, 482-486.

57 Подиновский В.В., Нелюбин А.П. Потенциальная недоминируемость в задачах выбора// Материалы XLI Международной конференции «Информационные технологии в науке, социологии и бизнесе» (IT + SE`13) (Украина, Крым, Ялта – Гурзуф, 2 – 10 октября 2013 г.). Приложение к журналу «Открытое образование». Запорожье: Запорожский университет, 2013, 31-34.

58 Подиновский В.В., Нелюбин А.П. Потенциальная недоминируемость в задачах выбора при неточной информации о предпочтениях// Искусственный интеллект и принятие решений, 2014, 4, 83-95.

59 Гельвих П.А. Курс артиллерии для ускоренной подготовки в офицеры. Действительность стрельбы. – Петроград: Лит. Константиновского арт. училища. 1916.

60 Новосельцев А.А. Математическое моделирование финансовых рисков. – Новосибирск: Наука, 2001.

61 Шоломицкий А.Г. Теория риска. Выбор при неопределенности и моделирование риска. – М.: Издательский дом ГУ-ВШЭ. 2005.

62 Cheng S., Liu Y., Wang S. Progress in risk measurement// Advanced modeling and optimization, 2004, 6, 1-20.

63 Fishburn P.C. Mean risk analysis with risk associated with below target returns// American economic review, 1977, 67, 116-126.

- 64 Gertz E.M., Wright S. OOQP User guide// Mathematics and computer science division technical memorandum, 2001, 252, 1-62.
- 65 Gertz E.M., Wright S. Object-oriented software for quadratic programming// ACM Transactions on mathematical software, 2003, 29, 58-81.
- 66 Bourbaki N. Éléments de mathématique: topologie générale. Chapitres 1 à 4. Nouvelle édition. – Berlin: Springer, 2007.
- 67 Charnes A., Cooper W.W. Management models and industrial applications of linear programming. – New York: Wiley, 1961.
- 68 Подиновский В.В. Количественная важность критериев// Автоматика и телемеханика, 2000, 5, 110-123. *Английский перевод*: Podinovskii V.V. Quantitative importance of criteria// Automation and Remote Control, 2000, 61, 5(2), 817-828.
- 69 Подиновский В.В. Введение в теорию важности критериев в много-критериальных задачах принятия решений: Учебное пособие. – М.: Физматлит, 2007.
- 70 Podinovski V.V. Multicriteria optimization problems involving importance-ordered criteria// Modern Mathematical Methods of Optimization / Elster K.-H. (ed.). Berlin: Akademie Verlag, 1993, 254-267.
- 71 Podinovski V.V. The quantitative importance of criteria for MCDA// J. of Multi-Criteria Decision Analysis, 2002, 11, 1-15.
- 72 Кини Р., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения / Пер. с англ. – М.: Радио и связь, 1981.
- 73 Levy H. Stochastic dominance: investment decision making under uncertainty. – New York: Springer, 2006
- 74 Whitmore G.A., Findlay M.C. Stochastic dominance: an approach to decision making under risk. – Lexington: Heath, 1978.
- 75 Markowitz H.M. Portfolio selection. Efficient diversification of investments. – New York: Wiley, 1959.

- 76 Ogryczak W, Ruszczyński A. From stochastic dominance to mean-risk models: semideviations as risk measures // European journal of operational research, 1999, 116, 33-50.
- 77 Porter, R.B. Semivariance and stochastic dominance: a comparison. American economic review, 1974, 64, 200-204.
- 78 Artzner P. et al. Coherent measure of risk// Mathematical finance. 1999. V. 9. P. 223 – 228
- 79 Libby R, Fishburn P.C. Behavioral models of risk taking in business decisions: a survey and evaluation// Journal of accounting research. 1977, 12, 272-292.
- 80 Fishburn P.C. Foundation of risk measurement. II// Journal of mathematical psychology, 1984, 30, 396-406.
- 81 Kovacevic R.W., Pflug G.C., Pichler A. Measuring and managing risk. In: Baker H.K., Filbeck G. (Eds). Investment risk management. New York: Oxford university press, 2015, 17-41.
- 82 Fishburn P.C. Foundation of risk measurement. I// Journal of mathematical psychology, 1982, 25, 226-242.
- 83 Müller A., Stoyan D. Comparison methods for stochastic models and risks. – Chichester: Wiley, 2002.
- 84 Подиновский В.В. Математическая теория выработки решений в сложных ситуациях. – М.: МО СССР. 1981.
- 85 Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. – М.: Физматлит, 1982.
- 86 Grechuk B., Molyboha A., Zabarankin M. Mean-deviation analysis in the theory of choice// Risk analysis, 2012, 32, 1277-1292.
- 87 Ogryczak W., Opolska-Rutkowska M. On Mean-risk models consistent with stochastic dominance// Instytut automatyki i informatyki stosowanej politechniki Warszawskiej, 2004. Raport 04-08.

- 88 Konno H., Yamazaki H. Mean-absolute deviation portfolio optimization model and its applications in Tokyo stock market// Management science, 1991, 37, 519-531.
- 89 Luce R.D. Several possible measure or risk// Theory and decision, 1980, 12, 217-228, correction, 1981, 13, 381.
- 90 Ogryczak W, Ruszczyński A. On consistency of stochastic dominance and mean-semideviation models// Mathematical programming. Ser. B, 2001, 89, 217-232.
- 91 Mansini R., Ogryczak W., M. Speranza G. // Twenty years of linear programming based portfolio optimization. European journal of operational research, 2014, 234, 518-535.
- 92 Подиновский В.В. Меры риска как критерии выбора при вероятностной неопределенности// Искусственный интеллект и принятие решений, 2015, 2, 60-74.
- 93 Подиновский В.В., Нелюбин А.П. О построении множества потенциально недоминируемых вариантов// Информационные технологии в науке, образовании и управлении: Материалы XLII Международной конференции IT + SE`15 (Гурзуф, 22 мая – 01 июня 2015 г.). М.: ИНИТ, 2015, 67-71.
- 94 Подиновский В.В. Потенциальная оптимальность и потенциальная недоминируемость в задачах выбора// Proc. of the International Scientific Conference “Information and Computer Technologies, Modelling, Control – 2015” (ICT-MC 2015) (в печати).
- 95 Duggan J. General conditions for the existence of maximal elements via the uncovered set// Journal of Mathematical Economics, 2011, 47, 755-759.
- 96 von Neumann J., Morgenstern O. Theory of Games and Economic Behavior. – Princeton: Princeton University Press, 1944.
- 97 Subochev A. Dominant, Weakly Stable, Uncovered Sets: Properties and Extensions// Working paper WP7/2008/03. Moscow: SU – Higher School of Economics, 2008.

- 98 Aleskerov F., Kurbanov E. A Degree of Manipulability of Known Social Choice Procedures// Current Trends in Economics: Theory and Applications / Eds. Alkan A., Aliprantis Ch., Yannelis N. – N.Y.: Springer-Verlag, 1999, 13-27.
- 99 Aleskerov F., Subochev A. Matrix-vector representation of various solution concepts// Working paper WP7/2009/03. Moscow: SU – Higher School of Economics, 2009.
- 100 Aleskerov F., Subochev A. Modeling optimal social choice: matrix-vector representation of various solution concepts based on majority rule// Journal of Global Optimization, 2013, 56(2), 737-756.
- 101 Subochev, A. Axiomatic analysis of the union of minimal externally stable sets, concerning possibility of application and implementation of this novel optimal social choice correspondence, 2015. (unpublished)
- 102 Айзерман М.А., Алекскеров Ф.Т. Выбор вариантов: основы теории. – М.: Наука, 1990.
- 103 Laslier J.F. Tournament Solutions and Majority Voting. – Berlin: Springer, 1997.
- 104 Nash J. F. The Bargaining Problem// Econometrica, 1950,18(2), 155-162.
- 105 Aizerman M., Aleskerov F. Theory of Choice. Amsterdam: – North-Holland/Elsevier, 1995.
- 106 Özkal-Sanver I., Sanver M.R. A new monotonicity condition for tournament solutions// Theory and decision, 2010, 69(3), 439-452.
- 107 Aleskerov F.T., Pislyakov V., Subochev A. Ranking Journals In Economics, Management And Political Science By Social Choice Theory Methods// Working papers by NRU Higher School of Economics. Series WP BRP "Science, Technology and Innovation", 2014.
- 108 Arrow K.J. Social Choice and Individual Values. – New York: Wiley, 1951.

- 109 Aleskerov F., Pislyakov V., Subochev A. Ranking journals using social choice theory methods: a novel approach in bibliometrics// 2015 (forthcoming).
- 110 Алескеров Ф.Т., Курбанов Э. О степени манипулируемости правил коллективного выбора// Автоматика и телемеханика, 1998, 10, 134-146.
- 111 Вольский В.И. Процедуры голосования в малых группах с древнейших времен до начала XX века// Препринты. Издательский дом ВШЭ. Серия WP7 "Математические методы анализа решений в экономике, бизнесе и политике", 2014.
- 112 Алескеров Ф.Т., Юзбашев Д.В., Якуба В.И. Пороговое агрегирование трехградационных ранжировок// Автоматика и телемеханика, 2007, 1, 147-152.
- 113 Shvydun S. Normative properties of multi-criteria choice procedures and their superpositions: I// Working paper WP7/2015/07 (Part 1). Moscow: HSE Publishing House, 2015.
- 114 Shvydun S. Normative properties of multi-criteria choice procedures and their superpositions: II// Working paper WP7/2015/07 (Part 2). Moscow: HSE Publishing House, 2015.
- 115 Gibbard A. Manipulation of voting schemes// Econometrica, 1973, 41, 587-601.
- 116 Satterthwite M. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions// Journal of Economic Theory, 1975, 10, 187-217.
- 117 Nitzan S. The vulnerability of point-voting schemes to preference variation and strategic manipulation// Public Choice, 1985, 47, 349-370.
- 118 Kelly J. Almost all social choice rules are highly manipulable, but few aren't// Social Choice and Welfare, 1993, 10, 161-175.
- 119 Aleskerov F., Karabekyan D., Sanver R., Yakuba V. On manipulability of positional voting rules// SERIEs: Journal of the Spanish Economic Association, 2011, 2(4), 431-446.

120 Aleskerov F., Karabekyan D., Sanver R., Yakuba V. On the manipulability of voting rules: Case of 4 and 5 Alternatives// Mathematical Social Sciences, 2012, 64(1), 67-73.

121 Aleskerov F., Ivanov A., Karabekyan D., Yakuba V. Manipulability of Aggregation Procedures in Impartial Anonymous Culture// Procedia Computer Science, 2015, 55, 1250-1257.

122 Иванов А., Карабекян Д., Якуба В. Манипулируемость правил коллективного выбора в Impartial Anonymous Culture// сборник докладов XVI Апрельской международной научной конференции «Модернизация экономики и общества», 2015.

123 Hix S., Høyland B. Empowerment of the European Parliament// Annual Review of Political Science, 2013, 16(1), 171-189.

124 Hix S., Noury A., Roland. G. Power to the Parties: Cohesion and Competition in the European Parliament, 1979-2001// British Journal of Political Science, 2005, 35(2), 209-234.

125 Hoyland B., Sircar I., Hix S. Forum section: an automated database of the European Parliament// European Union Politics, 2009, 10(1), 143-152.

126 Noury A. Ideology, Nationality and Euro-Parliamentarians// European Union Politics, 2002, 3(1), 33-58.

127 Rice S.A. The Behavior of Legislative Groups: A Method of Measurement// Political Science Quarterly, 1925, 40(1), 60-72.

128 Hix S., Noury A., Roland G. Voting Patterns and Alliance Formation in the European Parliament// Philosophical Transactions of the Royal Society, 2009, 364(1518), 821-831.

129 Сатаров Г.А. 1993. Российские съезды: деюстификация политической системы// Российский мониторинг: Архив современной политики. №1.

130 Gore P.A., Jr. Cluster Analysis/ Tinsley H.E.A., Brown S.D. (eds.) Handbook of Applied Multivariate Statistics and Mathematical Modeling. – San Diego, 2000.

- 131 Kreppel A., Tsebelis G. Coalition Formation in the European Parliament// Comparative Political Studies, 1999, 32(8), 933-966.
- 132 Hix S., Noury A., Roland G. Dimensions of Politics in the European Parliament// American Journal of Political Science, 2006, 50(2), 494-511.
- 133 Han J-H. Analysing roll calls of the European Parliament// European Union Politics, 2007, 8, 479-507.
- 134 Hix S., Noury A. After enlargement: voting patterns in the sixth European Parliament// Legislative Studies Quarterly, 2009, 34(2), 159-174.
- 135 Lindstädt R., Slapin J.B., Vander Wielen R.J. Balancing competing demands: position taking and election proximity in the European Parliament// Legislative Studies Quarterly, 2011, 36(1), 37-70.
- 136 Alesina A., Devleeschauwer A., Easterly W., Kurlat S., Wacziarg R. Fractionalization// Journal of Economic Growth, 2003, 8(2), 155-194.
- 137 Esteban J., Ray D. On the Measurement of Polarization// Econometrica, 1994, 62(4), 819-851.
- 138 Алескеров Ф.Т., Бородин А.Д., Каспэ С.И., Маршаков В.А., Салмин А.М. Поляризованность избирательных предпочтений в России. Опыт выборов в Государственную думу Российской Федерации в 1993—2003 гг.// Препринты. Высшая школа экономики. Серия WP7 «Математические методы анализа решений в экономике, бизнесе и политике», №2, 2005.
- 139 Esteban J., Ray D. Linking Conflict to Inequality and Polarization// American Economic Review, 2011, 101(4), 1345-1374.
- 140 Камалова Р.У. Анализ согласованности позиций групп в Европейском парламенте в 1979-1024 гг.// WP7/2015. – М.: Изд. дом Высшей школы экономики, 2015.
- 141 Reif K., Schmitt H. Nine Second-Order National Elections: A Conceptual Framework for the Analysis of European Election Results// European Journal of Political Research, 1980, 8(1), 3-45.

142 Greven M.T. Political Parties between National Identity and Eurofication// in Brian Nelson, David Roberts and Walter Veit (eds) *The Idea of Europe, Problems of National and Transnational Identity*. New York: BERG Inc., 1992.

143 van der Eijk C., Franklin M.N. Choosing Europe? The European electorate and national politics in the face of union// Ann Arbor, MI: University of Michigan Press., 1996.

144 Aleskerov F., Avci G., Iakouba V., Tiirem Z.U. European Union enlargement: Power distribution implications of the new institutional arrangements// *European Journal of Political Research*, 2002, 41(3), 379-394.

145 Attina F. The Voting Behaviour of European Parliament members and the problem of Europarties// *European Journal of Political Research*, 1990, 5, 557-559.

146 Crombez C. Legislative procedures in the European Community// *British Journal of Political Science*, 1996, 26, 199-228.

147 Hix, S. and M. Marsh. Punishment or Protest? Understanding European Parliament Elections// *Journal of Politics*, 2007, 69, 495-510.

148 Lane J.-E., Maeland R., Berg S. Research Note: The EU Parliament: Seats, States and Political Parties// *Journal of Theoretical Politics*, 1995, 7, 395-400.

149 Nurmi H. The representation of voter groups in the European Parliament: A Penrose-Banzhaf Index analysis// *Electoral Studies*, 1997, 16, 317-339.

150 Nurmi H., Meskanen T. A priori power measures and institutions of the European Union// *European Journal of Political Research*, 1999, 35, 161-179.

151 Raunio T. The European Perspective: Transnational Party Groups in the 1989–1994 European Parliament. – London: Ashgate. 1997.

152 Tsebelis G., Garrett G. Legislative Politics in the European Union// *European Union Politics*, 2000, 1(1), 9-36.

- 153 Касте Ф. Европейскому парламенту не хватает представительности// Свободная мысль, 2007, 1, 53-56.
- 154 Hix S. The Political System of the European Union. – London: Palgrave, 2005.
- 155 Воронов К. Выборы в Европарламент – кризис доверия между элитой и избирателями углубляется [Электронный ресурс] / ИА REGNUM. 29.05.2014. URL: <http://www.regnum.ru/news/polit/1808323.html#ixzz3PqSPaaAe> (дата обращения: 01.05.2015).
- 156 Dahrendorf R. The Challenge for Democracy// Journal of Democracy, 2003, 14(4).
- 157 Оленченко В. Европарламент: усиление евроскептиков// Международная жизнь, 2014, 9, 42-59.
- 158 Зеленов Р.Ю. Эволюция полномочий европейского парламента// Государство и право, 2005, 1, 97-103.
- 159 Roll Call Votes in the European Parliament, 26 July 2011 [Электронный ресурс]. URL: <http://personal.lse.ac.uk/hix/HixNouryRolandEPdata.htm> (дата обращения: 01.05.2015).
- 160 Poole K.T., Rosenthal H. A Spatial Model for Legislative Roll Call Analysis// American Journal of Political Science, 1985, 29, 357-384.
- 161 Poole K., Lewis J., Lo J., Carroll R. Scaling Roll Call Votes with wnominate in R// Journal of Statistical Software, 2011, 42(14), 1-21.
- 162 Noury A., Roland G. More Power to the European Parliament?// Economic Policy, 2002, 34, 279-320.
- 163 Banzhaf J.F. Weighted Voting Doesn't Work: A Mathematical Analysis// Rutgers Law Review, 1965, 19, 317-343.
- 164 Aleskerov F. Power Indices Taking into Account Agents' Preferences// Mathematics and Democracy. Recent Advances in Voting Systems and Collective Choice. – Berlin-Heidelberg: Springer, 2006, 1-18.

- 165 Шварц Д.А. О вычислении индексов влияния, учитывающих предпочтения участников// Автоматика и телемеханика, 2009, 3, 152-159.
- 166 Ушаков М.А., Шварц Д.А. Вычисление индексов влияния, учитывающих предпочтения участников в анонимных играх// WP7/2015/04. – М.: Изд. дом Высшей школы экономики, 2015. – 16 с.
- 167 Камалова Р.У. Ушаков М.Н., Шварц Д.А. Оценка распределения влияния в Европейском Парламенте в 1979-2014 гг.// WP7/2015/08. – М.: Изд. дом Высшей школы экономики, 2015.
- 168 Олейник В.В., Лебедюк В.Н. Распределение влияния между фракциями и группами в Верховной Раде Украины (1990 – 2012 гг.)// препринт WP7/2013/08/ – М.: Изд. дом Высшей школы экономики, 2013. – 76 с.
- 169 Aleskerov F., Ersel H., Sabuncu Y. Power distribution, inequality of power and coalitional stability in the Turkish parliament// Turkish Studies, 2000, 1, 21-38.
- 170 Aleskerov F., Holler M., Kamalova R. Power Distribution in the Weimar Reichstag in 1919 – 1933// Working paper WP7/2010/08, Moscow: Publishing House of the University – Higher School of Economics, 2010, 54p.
- 171 Noury A., Mielcova E. Roll Call Voting in a Multi-Party Parliament: The Case of the Czech Parliament// Working Paper, Brussels: Free University of Brussels, Belgium, 1998.
- 172 Алескеров Ф.Т., Благовещенский Н., Сатаров Г., Соколова А.В., Якуба В.И. Влияние и структурная устойчивость в российском парламенте (1905-1917 и 1993-2005 гг.). – М.: Физматлит, 2007, 309 с.
- 173 Олейник В. Анализ структуры Верховной Рады Украины на основании поименных голосований депутатов (2007 – 2012)// Полития, 2016 (принята в печать)
- 174 Ramnath S., Rock S., Shane Ph. A Review of research related to financial analysts' forecasts and stock recommendations// Foundations and Trends in Finance, 2010, 2-4, 311-421.

- 175 Bonner S., Hugon A., Walther B. Investor reaction to celebrity analysts: the case of earnings forecast revisions// Journal of Accounting Research, 2007, 45(3), 481-513.
- 176 Jegadeesh N., Kim J., Krische S., Lee C. Analyzing the analysts: when do recommendations add value?// Journal of Finance, 2004, 59, 1083-124.
- 177 Jegadeesh N., Kim W. Do Analysts Herd? An analysis of recommendations and market reactions// Review of Financial Studies. 2010, 23(2), 901-937.
- 178 Malmendier U., Shanthikumar D. Do security analysts speak in two tongues?// Review of Financial Studies, 2014, 27(5), 1287-1322.
- 179 Aleskerov F.T., Egorova L.G. Is it so bad that we cannot recognize black swans?// Economics Letters, 2012, 117(3), 563-565.
- 180 Ahmad-Zaluki N., Wan-Hussin W. Corporate governance and earnings forecasts accuracy// Asian Review of Accounting, 2010, 18(1), 50-67.
- 181 Теплова Т.В. Инвестиции: учебник. – М.: ИД Юрайт, 2011.
- 182 Dempster A. Upper and lower probabilities induced by multivalued mapping// Ann. Math. Statist., 1967, 38, 325-339.
- 183 Shafer G. A mathematical theory of evidence. – Princeton, N.J.: Princeton University Press, 1976.
- 184 Lepskiy A. About relation between the measure of conflict and decreasing of ignorance in theory of evidence// Proceedings of the 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13). Amsterdam – Beijing – Paris: Atlantis Press, 2013, 355-362.
- 185 Bronevich A., Lepskiy A., Penikas H. The Application of Conflict Measure to Estimating Incoherence of Analyst's Forecasts about the Cost of Shares of Russian Companies// Procedia Computer Science, 2015, 55, 1113-1122.
- 186 Wang Z., Klir G.J. Generalized Measure Theory. – New York: Springer, 2009.
- 187 Grabisch M., Marichal J.-L., Mesiar R., Pap E. Aggregation functions. – Cambridge: Cambridge University Press, 2009.

- 188 Shapley L. A value for n-person games// Contributions to the Theory of Games. II (28) in Annals of Mathematics Studies, Princeton University Press, 1953, 307-317.
- 189 Fatima S., Wooldridge M., Jennings N. A linear approximation method for the Shapley value// Artificial Intelligence, 2008, 172, 1673-1699.
- 190 [Zlotkin & Rosenschein 1994] Zlotkin G., Rosenschein J. Coalition, cryptography, and stability: mechanisms for coalition formation in task oriented domains// Proceedings of the National Conference on Artificial Intelligence (AAAI-94), 1994, 432-437.
- 191 Броневич А.Г., Косюк Е.Д., Лепский А.Е., Пеникас Г.И. Исследование конфликтности и детерминант точности прогнозов в рекомендациях российских финансовых аналитиков// препринт WP7/2015/10. – М.: Изд. дом. Высшей школы экономики, 2015. – 32 с.
- 192 Jurafsky D., Martin J.H. Speech and Language Processing. – Prentice Hall: Englewood Cliffs, NJ, 2008.
- 193 Manning C.D., Schütze H. Foundations of Statistical Natural Language Processing. – Cambridge: MIT Press, 1999.
- 194 Harris Z. Distributional structure// Word, 1954, 23(10), 146-162.
- 195 Salton G., Buckley C. Term-weighting approaches in automatic text retrieval// Information Processing and Management, 1998, 25(5), 513-523.
- 196 Berry M.W., Browne M. Understanding Search Engines: Mathematical Modeling and Text Retrieval. – Philadelphia: SIAM, 2005.
- 197 Reed J.W. et al. TF-ICF: A New Term Weighting Scheme for Clustering Dynamic Data Streams// ICMLA '06 Proceedings of the 5th International Conference on Machine Learning and Applications. – Washington, DC, USA: IEEE Computer Society, 2006, 258-263.
- 198 Raghavan V.V., Won S.K.M. A critical analysis of vector space model for information retrieval// Journal of the American Society for Information Science, 1986, 37(5), 279-287.

- 199 Turney P.D., Pantel P. From frequency to meaning: vector space models of semantics// Journal of Artificial Intelligence Research, 2010, 37(1), 141-188.
- 200 Rehurek R., Sojka P. Software Framework for Topic Modelling with Large Corpora// Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 2010, 45-50.
- 201 Bird S., Klein E., Loper E. Natural Language Processing with Python. – O'Reilly Media, 2009.
- 202 Manning C.D., Raghavan P., Schütze H. An Introduction to Information Retrieval. – Cambridge: Cambridge University Press, 2009.
- 203 Sebastiani F. Machine learning in automated text categorization// Journal of ACM Computing Surveys, 2002, 34(1), 1-42.
- 204 Turney P.D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews// Proceedings of the Association for Computational Linguistics, 2002, 417-424.
- 205 Pang B., Lee L., Vaithyanathan S. Thumbs up? Sentiment Classification using Machine Learning Techniques// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2002, 79-86.
- 206 Strapparava C., Valitutti A. WordNet-Affect: An affective extension of WordNet// Proceedings of LREC, 2004, 1083-1086.
- 207 Andrews N., Fox E. Recent Developments in Document Clustering// [Online], 2007, <http://www.hse.ru/data/2012/12/20/1303727883/2.pdf>.
- 208 Wong S.K.M., Ziarko W., Wong P.C.N. Generalized vector spaces model in information retrieval// the 8th annual international ACM SIGIR conference on Research and development in information retrieval, New York, USA: ACM, 1985, 18-25.
- 209 Pantel P., Lin D. Discovering word senses from text// Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Canada, 2002, 613-619.

- 210 Rapp R. Word sense discovery based on sense descriptor dissimilarity// Proceedings of the Ninth Machine Translation Summit, 2003, 315-322.
- 211 Turney P.D. Littman M.L., Bigham J., Shnayder V. Combining independent modules to solve multiple-choice synonym and analogy problems// Proceedings of the International Conference on Recent Advances in Natural Language Processing, Borovets, Bulgaria, 2003, 482-489.
- 212 Robertson S., Zaragoza H. The Probabilistic Relevance Framework: BM25 and Beyond// Foundations and Trends in Information Retrieval, 2009, 3(4), 333-389.
- 213 Deerwester S. et al. Indexing by latent semantic analysis// Journal of the american society for information science, 1990, 4(6), 391-407.
- 214 Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet Allocation// Journal of Machine Learning Research, 2003, 3(4-5), 993-1022.
- 215 Yu C.T., Salton G. Precision Weighting - An Effective Automatic Indexing Method// Journal of ACM, 1976, 23(1), 76-88.
- 216 Ponte J.M., Croft W.B. A language modeling approach to information retrieval// ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA: ACM, 1998, 275-281.
- 217 Koehn P., Och F.J., Marcu D. Statistical phrase based translation// Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics, 2003, 48-54.
- 218 Katz S. Estimation of probabilities from sparse data for the language model component of a speech recognizer// IEEE Transactions on Acoustics, Speech and Signal Processing, 1987, 35(3), 400-401.
- 219 Байтин А. Исправление поисковых запросов в Яндексе. Вероятностная языковая модель// Российские Интернет-Технологии, 2008,  
<http://download.yandex.ru/company/experience/rit2008/baytin.pdf>

- 220 Воронцов К.В. Лекции по вероятностному тематическому моделированию// <http://www.machinelearning.ru/wiki/images/f/fb/Voron-ML-TopicModels.pdf>
- 221 Mirkin B.G. Core Concepts in Data Analysis: Summarization, Correlation and Visualization. – Heidelberg: Springer, 2011.
- 222 Hofmann T. Probabilistic Latent Semantic Indexing// Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval, 1999, 50-57.
- 223 Hofmann T. Latent semantic models for collaborative filtering// ACM Transactions on Information Systems, 2004, 22(1), 89-115.
- 224 Lin C. et al. Using Probabilistic Latent Semantic Analysis for Personalized Web Search// Web Technologies Research and Development-APWeb 2005 / ed. Springer, 2005.
- 225 Berry M. W., Dumais S.T., O'Brien G.W. Using Linear Algebra for Intelligent Information Retrieval// SIAM Review, 1995, 37, 573-595.
- 226 Wei X., Croft B.W. LDA-based document models for ad-hoc retrieval// Proceeding SIGIR '06 Proceedings of the 29th annual international ACM, New York, USA: ACM, 2006, 178-185.
- 227 Hyunsoo K., Howland P., Park H. Dimension reduction in text classification with support vector machines// Journal of Machine Learning Research, 2005, 37-53.
- 228 Ramage D. et al. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora// Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2009, 1, 248-256.
- 229 Gong Y., Liu X. Generic text summarization using relevance measure and latent semantic analysis// Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 2001, 19-25.

- 230 Arora R., Ravindran B. Latent dirichlet allocation based multi-document summarization// Proceedings of the second workshop on Analytics for noisy unstructured text data, 2008, 91-97.
- 231 Gee R.K. Using latent semantic indexing to filter spam// Proceedings of the 2003 ACM symposium on Applied computing, 2004, 460-464.
- 232 Bíró I., Szabó J., Benczúr A.A. Latent dirichlet allocation in web spam filtering// Proceedings of the 4th international workshop on Adversarial information retrieval on the web, 2008, 29-32.
- 233 Krestel R., Frankhauser P., Nejdl W. Latent dirichlet allocation for tag recommendation// Proceedings of the third ACM conference on Recommender systems, ACM, 2009, 61-68.
- 234 Monay F., Gatica-Perez D. On image auto-annotation with latent space models// Proceedings of the eleventh ACM international conference on Multimedia, 2003.
- 235 Wang X., Grimson E. Spatial latent dirichlet allocation// Advances in Neural Information Processing Systems, 2008.
- 236 Weiner P. Linear pattern matching algorithms// 14th Annual IEEE Symposium on Switching and Automata Theory, 1973, 1-11.
- 237 Barsky M. et al. Suffix trees for very large genomic sequences// Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009.
- 238 Grossi R., Vitter J.S. Compressed suffix arrays and suffix trees with applications to text indexing and string matching// SIAM Journal on Computing, 2005, 35(2), 378-407.
- 239 Zamir O. Etzioni O. Web document clustering: a feasibility demonstration// Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, 1998, 46-54.

- 240 Hu Z., Zhang Y., Zhou J. F. Method for extracting name entities and jargon terms using a suffix tree data structure [Patent]// 7,197,449, US, March 27, 2007.
- 241 Hung C., Deng X. A new suffix tree similarity measure for document clustering// Proceedings of the 16th international conference on World Wide Web, 2007.
- 242 Snowsill T. et al. Finding surprising patterns in textual data streams// IAPR Workshop on Cognitive Information Processing (CIP2010), Elba Island, Italy IEEE, 2010, 405-410.
- 243 Gusfield D. Algorithms on strings, trees and sequences: computer science and computational biology. – Cambridge: Cambridge University Press, 1997.
- 244 Pampapathi R., Mirkin B., Levene M. A suffix tree approach to anti-spam email filtering// Machine learning, 2006, 65(1), 309-338.
- 245 Дубов М.С. Черняк Е.Л. Аннотированные суффиксные деревья: особенности реализации// АИСТ, Екатеринбург: Национальный Открытый Университет «ИНТУИТ», 2013, 10-20.
- 246 Hjørland B. The foundation of the concept of relevance// Journal of the American Society for Information Science and Technology, 2010, 61(1), 217-237.
- 247 Сегалович И. Как работают поисковые системы?// Мир Интернет, 2002, 10, 24-32.
- 248 Taylor A. User relevance criteria choices and the information search process// Information Processing & Management, 2012, 135-153.
- 249 Zhai C., Lafferty J. A study of smoothing methods for language models applied to ad hoc information retrieval// Proceedings of ACM SIGIR, 2001, 334-342.
- 250 Миркин Б.Г., Черняк Е.Л., Чугунова О.Н. Метод аннотированного суффиксного дерева для оценки степени вхождения строк в текстовые документы// Бизнес–информатика, 2012, 21(3), 31-41.
- 251 Pissanetzky S. Sparse Matrix Technology. – Academic Press, 1984.

252 Chernyak E.L. Mirkin B.G. Refining a Taxonomy by Using Annotated Suffix Trees and Wikipedia Resources// Annals of Data Science, 2015, 2(1), pp.61-82.

253 Chernyak E. Some Thoughts on Using Annotated Suffix Trees for Natural Language Processing// 2nd Workshop on Interactions Between Data Mining and Natural Language Processing, DMNLP 2015, Porto, Portugal, 7 September 20, CEUR-WS, 2015, 1410, 5-18.

254 Яковлев М.С., Черняк Е.Л. Using annotated suffix tree similarity measure for text summarisation// Analysis of Large and Complex Data. Springer International Publishing, 2015. (в печати)

255 Долгалева И.И., Черняк Е.Л. Анализ и визуализация корпуса твитов// Сборник трудов XVII Международной конференции “Аналитика и управление данными в областях с интенсивным использованием данных”, 2015.

## ПРИЛОЖЕНИЕ А

### Таблицы к подразделу 1.3

**Таблица А.1 – Сеть голосований Wikipedia (7115 вершин)**

Индексы центральности (число взятых вершин)	Число вершин, попавших в интервалы											
	Правило Копланда 1 (верхний интервал)				Правило Копланда 2 (верхний интервал)				Правило Копланда 3 (нижний интервал)			
	40% (20)	50% (61)	70% (6954)	80% (7101)	60% (14)	70% (38)	80% (90)	90% (272)	5% (5618)	20% (6895)	30% (7037)	50% (7099)
In-Degree (23)	2	2	6	9	2	3	6	7	0	0	0	7
Out-Degree (27)	18	26	27	27	14	27	27	27	12	22	22	25
Degree (27)	11	13	24	24	11	19	24	24	4	9	10	19
Eccentricity (1956)	0	0	1874	1952	0	0	0	0	1479	1903	1933	1954
Closeness Centrality (1933)	10	11	1851	1929	10	11	11	11	1451	1877	1907	1929
Betweenness Centrality (27)	3	5	23	23	4	11	18	23	0	3	6	18
Closeness Centrality normalized (1222)	0	0	1141	1218	0	0	0	0	774	1169	1199	1220
Betweenness Centrality normalized (8)	3	3	7	7	3	4	6	7	0	0	0	3
PageRank (37)	1	1	8	23	1	2	5	8	0	0	4	22
Eigenvector Centrality (26)	2	2	7	14	2	3	6	8	0	0	0	10
Clustering Coefficient (41)	0	0	41	41	0	0	0	0	41	41	41	41

**Таблица А.2 – Одноранговая компьютерная сеть Gnutella (10876 вершин)**

Индексы центральности (число взятых вершин)	Число вершин, попавших в интервалы											
	Правило Копланда 1 (верхний интервал)				Правило Копланда 2 (верхний интервал)				Правило Копланда 3 (нижний интервал)			
	40% (7)	50% (21)	60% (4116)	80% (10860)	60% (6)	70% (12)	80% (21)	85% (31)	5% (7102)	20% (10624)	50% (10852)	70% (10870)
In-Degree (19)	0	0	0	4	0	0	0	0	0	0	0	13
Out-Degree (31)	7	21	29	31	6	12	21	31	10	28	31	31
Degree (34)	6	6	6	20	6	6	6	7	3	6	11	28
Eccentricity (6524)	0	0	285	6521	0	0	0	0	4569	6426	6522	6523
Closeness Centrality (6524)	0	0	285	6521	0	0	0	0	4569	6426	6522	6523
Betweenness Centrality (31)	2	2	11	25	2	2	3	4	7	11	21	28
Closeness Centrality normalized (10293)	7	21	3831	10278	6	12	21	31	6657	10051	1026 9	1028 7
Betweenness Centrality normalized (9)	1	1	3	7	1	1	2	3	1	3	6	7
PageRank (25)	0	0	11	25	0	0	0	0	10	22	24	25
Eigenvector Centrality (29)	0	0	0	14	0	0	0	0	0	0	9	23
Clustering Coefficient (18)	0	0	1	18	0	0	0	0	18	18	18	18

Таблица А.3 – Случайная сеть размером 10 тыс. вершин

Индексы центральности (число взятых вершин)	Число вершин, попавших в интервалы											
	Правило Копланда 1 (верхний интервал)				Правило Копланда 2 (верхний интервал)				Правило Копланда 3 (нижний интервал)			
	5% (851)	10% (5029)	40% (9914)	50% (9973)	10% (18)	20% (119)	70% (9008)	80% (9840)	5% (4669)	10% (7107)	30% (9744)	40% (9936)
In-Degree (28)	0	0	0	1	0	0	21	26	0	0	0	0
Out-Degree (46)	22	33	46	46	18	46	46	46	21	32	43	46
Degree (27)	0	0	0	2	0	0	22	26	0	0	0	0
Eccentricity (101)	21	65	101	101	0	8	101	101	60	79	101	101
Closeness Centrality (40)	21	28	40	40	17	38	40	40	20	28	38	40
Betweenness Centrality (78)	0	0	35	57	1	5	75	76	0	0	4	38
Closeness Centrality normalized (77)	0	15	77	77	0	0	0	0	39	58	76	77
Betweenness Centrality normalized (77)	0	0	35	57	1	5	75	76	0	0	4	38
PageRank (29)	0	0	0	4	0	0	22	27	0	0	0	0
Eigenvector Centrality (25)	0	0	0	2	0	0	18	23	0	0	0	0
Clustering Coefficient (26)	3	25	26	26	0	0	18	21	26	26	26	26

Таблица А.4 – Случайная сеть размером 15 тыс. вершин

Индексы центральности (число взятых вершин)	Число вершин, попавших в интервалы											
	Правило Копланда 1 (верхний интервал)				Правило Копланда 2 (верхний интервал)				Правило Копланда 3 (нижний интервал)			
	5% (285)	10% (4232)	50% (14837)	60% (14938)	30% (54)	35% (128)	85% (14773)	90% (14928)	5% (5736)	40% (14682)	50% (14878)	60% (14944)
In-Degree (44)	0	0	0	0	0	0	44	44	0	0	0	0
Out-Degree (82)	27	49	82	82	54	82	82	82	27	81	81	82
Degree (41)	0	0	0	1	0	0	41	41	0	0	0	1
Eccentricity (11424)	253	3422	11310	11379	51	121	11347	11409	4362	11182	11339	11384
Closeness Centrality (57)	19	35	57	57	44	55	57	57	18	55	56	57
Betweenness Centrality (56)	0	0	9	27	2	3	56	56	0	2	7	27
Closeness Centrality normalized (25)	0	0	24	25	0	0	0	0	13	24	25	25
Betweenness Centrality normalized (25)	0	0	1	8	1	1	25	25	0	0	0	8
PageRank (86)	0	0	1	26	0	0	86	86	0	0	2	30
Eigenvector Centrality (50)	0	0	0	3	0	0	50	50	0	0	0	5
Clustering Coefficient (25)	1	12	25	25	0	0	20	20	23	25	25	25

## ПРИЛОЖЕНИЕ Б

### Таблицы к подразделу 2.3

Таблица Б.1 – Свойства двухступенчатых процедур выбора («+» – процедура выбора удовлетворяет нормативному условию, «-» – процедура выбора не удовлетворяет нормативному условию)

Двухступенчатые процедуры выбора		Рациональность				Монотонность		Пороговая некомпенсируемость
Этап 1	Этап 2	Условие наследования (Н)	Условие согласия (С)	Условие отbrasывания (О)	Аксиома выбора Эрроу (ACA)	Условие монотонности 1	Условие монотонности 2	
Минимальное доминирующее множество Минимальное недоминируемое множество Правило Фишбэна, Непокрытое множество I, II Правило Ричалсона, Ядро, Правила Копланда 1-3 Надпороговое правило выбора (порог зависит от X) Минимаксная процедура, Процедура Симпсона	Двухступенчатое правило относительного большинства Правило Хара (Процедура Уэйра) Процедура Кумбса	-	-	-	-	+	-	-
Минимальное слабоустойчивое множество k-устойчивое множество( $k>1$ )	Правило простого большинства Двухступенчатое правило относительного большинства Правило Хара (Процедура Уэйра) Процедура Кумбса, Победитель Кондорсе	-	-	-	-	+	-	-
Правила Копланда 1-3 Надпороговое правило выбора (порог зависит от X) Минимаксная процедура, Процедура Симпсона	Правило простого большинства Победитель Кондорсе	-	-	-	-	+	+	-
Пороговое правило	Правило простого большинства Двухступенчатое правило относительного большинства Правило Хара (Процедура Уэйра), Победитель Кондорсе	-	-	-	-	+	+	-
Минимальное недоминируемое множество, Ядро	Правило простого большинства	+	+	-	-	+	+	-
Правило Фишбэна, Непокрытое множество I, II Правило Ричалсона	Правило простого большинства Победитель Кондорсе	+	+	-	-	+	+	-

Продолжение Таблицы Б.1

Ядро	Правило Борда, Правило Блэка Обратное правило Борда, Правило Нансона Минимальное доминирующее множество Минимальное недоминируемое множество Минимальное слабоустойчивое множество Правило Фишберна, Непокрытое множество I, II Правило Ричалсона, Правила Копланда 1-3 Минимаксная процедура, Процедура Симпсона	+	+	-	-	-	+	+	-	-
Непокрытое множество II	Минимальное доминирующее множество	-	+	-	-	-	+	-	-	-
Правила Копланда 1-3	k-устойчивое множество( $k>1$ )	-	-	-	-	-	+	-	-	-
Непокрытое множество I	Ядро	-	-	-	-	-	+	-	-	-
Пороговое правило Правила Копланда 1-3	Правило относительного большинства Обратное правило относительного большинства Аналог правила одобряющего голосования ( $q>1$ ) Правило Борда, Правило Блэка Минимальное доминирующее множество Минимальное недоминируемое множество Минимальное слабоустойчивое множество Правило Фишберна, Непокрытое множество I, II Правило Ричалсона, Ядро, Пороговое правило Правила Копланда 1-3 Надпороговое правило выбора (порог зависит от X) Минимаксная процедура, Процедура Симпсона	-	-	-	-	-	+	-	-	-
Пороговое правило	Обратное правило Борда, Правило Нансона k-устойчивое множество( $k>1$ )	-	-	-	-	-	+	-	-	-
Минимальное недоминируемое множество Непокрытое множество II	Непокрытое множество I Правило Ричалсона	-	-	-	-	-	+	-	-	-
Минимальное недоминируемое множество	Правило Фишберна, Непокрытое множество II k-устойчивое множество( $k>1$ ), Правила Копланда 1-3	-	-	-	-	-	+	-	-	-
Минимальное доминирующее множество Минимальное недоминируемое множество	Надпороговое правило выбора (порог зависит от X) Минимаксная процедура, Процедура Симпсона	-	-	-	-	-	+	-	-	-
Правило Фишберна, Правило Ричалсона Непокрытое множество I	Минимальное доминирующее множество Непокрытое множество I, Правило Ричалсона	-	-	-	-	-	+	-	-	-
Надпороговое правило выбора (порог зависит от X) Минимаксная процедура, Процедура Симпсона	Минимальное доминирующее множество Минимальное недоминируемое множество, Ядро	-	-	-	-	-	+	-	-	-
Минимальное доминирующее множество Минимальное недоминируемое множество Правила Копланда 1-3	Обратное правило Борда, Правило Нансона	-	-	-	-	-	-	-	-	-

## Продолжение Таблицы Б.1

Минимальное слабоустойчивое множество k-устойчивое множество( $k>1$ )	Минимальное доминирующее множество Минимальное недоминируемое множество Непокрытое множество I, Правило Ричалсона, Ядро	-	-	-	-	-	-	-	-
Минимальное слабоустойчивое множество Правило Фишбэра Непокрытое множество I, II Правило Ричалсона k-устойчивое множество( $k>1$ ) Надпороговое правило выбора (порог зависит от X) Минимаксная процедура Процедура Симпсона	Правило относительного большинства Аналог правила одобряющего голосования ( $q>1$ ) Правило Блэка Минимальное слабоустойчивое множество Правило Фишбэра, Непокрытое множество II k-устойчивое множество( $k>1$ ) Пороговое правило, Правила Копланда 1-3 Надпороговое правило выбора (порог зависит от X) Минимаксная процедура, Процедура Симпсона	-	-	-	-	-	-	-	-
Непокрытое множество I	Минимальное недоминируемое множество	-	-	-	-	-	-	-	-
Минимальное доминирующее множество Минимальное недоминируемое множество	Правило относительного большинства Обратное правило относительного большинства Аналог правила одобряющего голосования ( $q>1$ ) Правило Борда, Правило Блэка, Пороговое правило	-	-	-	-	-	-	-	-
Минимальное слабоустойчивое множество Правило Фишбэра, Непокрытое множество I, II Правило Ричалсона k-устойчивое множество( $k>1$ ) Надпороговое правило выбора (порог зависит от X) Минимаксная процедура, Процедура Симпсона	Обратное правило относительного большинства Правило Борда Обратное правило Борда Правило Нансона	-	-	-	-	-	-	-	-
Правило Фишбэра Непокрытое множество II, Правило Ричалсона	Минимальное недоминируемое множество Ядро	-	-	-	-	-	-	-	-
Ядро	Правило относительного большинства Обратное правило относительного большинства Аналог правила одобряющего голосования ( $q>1$ ) Пороговое правило Надпороговое правило выбора (порог зависит от X)	-	-	-	-	-	-	-	-
Надпороговое правило выбора (порог зависит от X) Минимаксная процедура, Процедура Симпсона	Непокрытое множество I Правило Ричалсона	-	-	-	-	-	-	-	-

## ПРИЛОЖЕНИЕ В

### Рисунки к подразделу 3.1

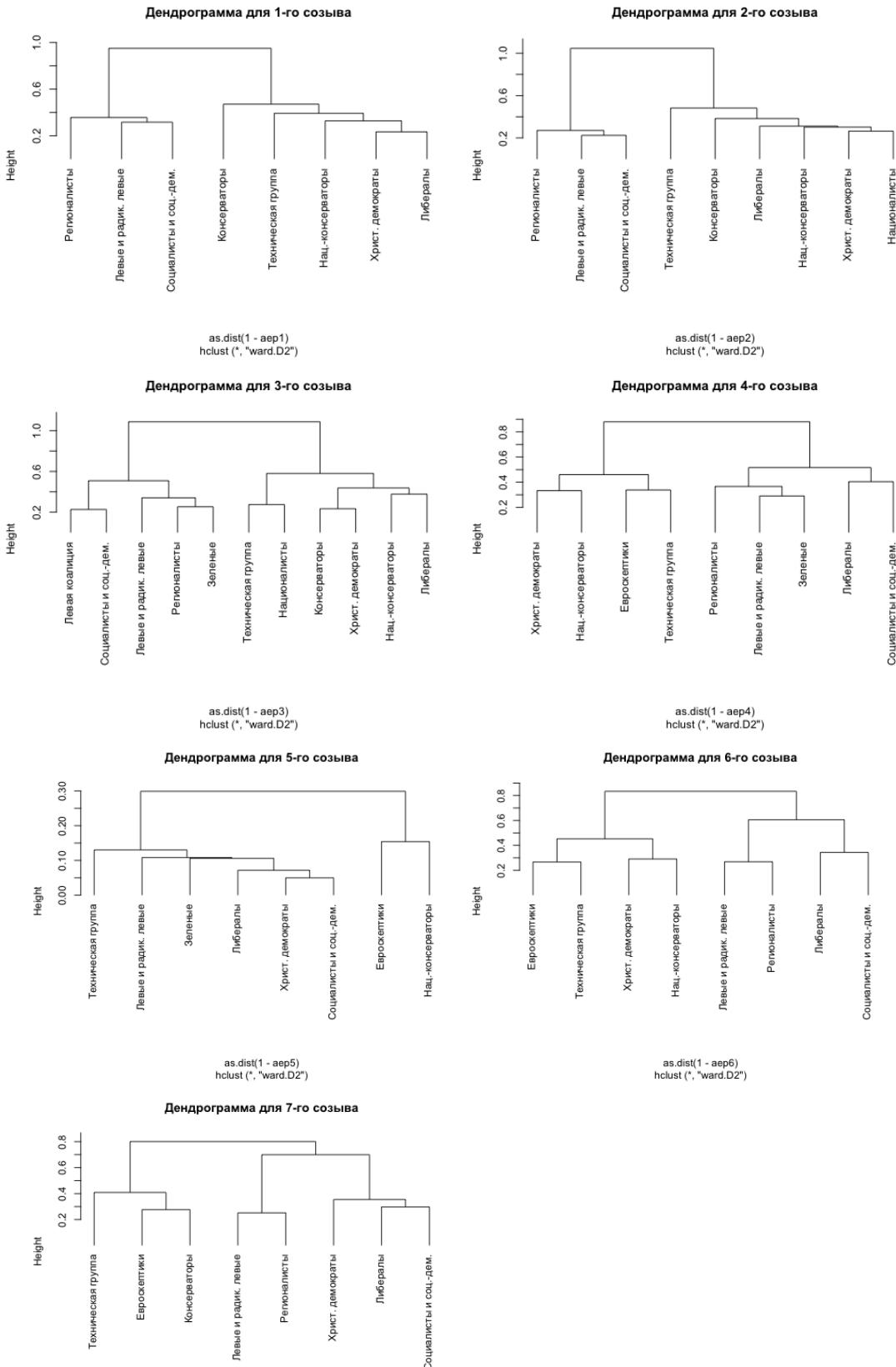


Рисунок В.1 – Кластеризация по индексу согласованности политических групп по семи созывам

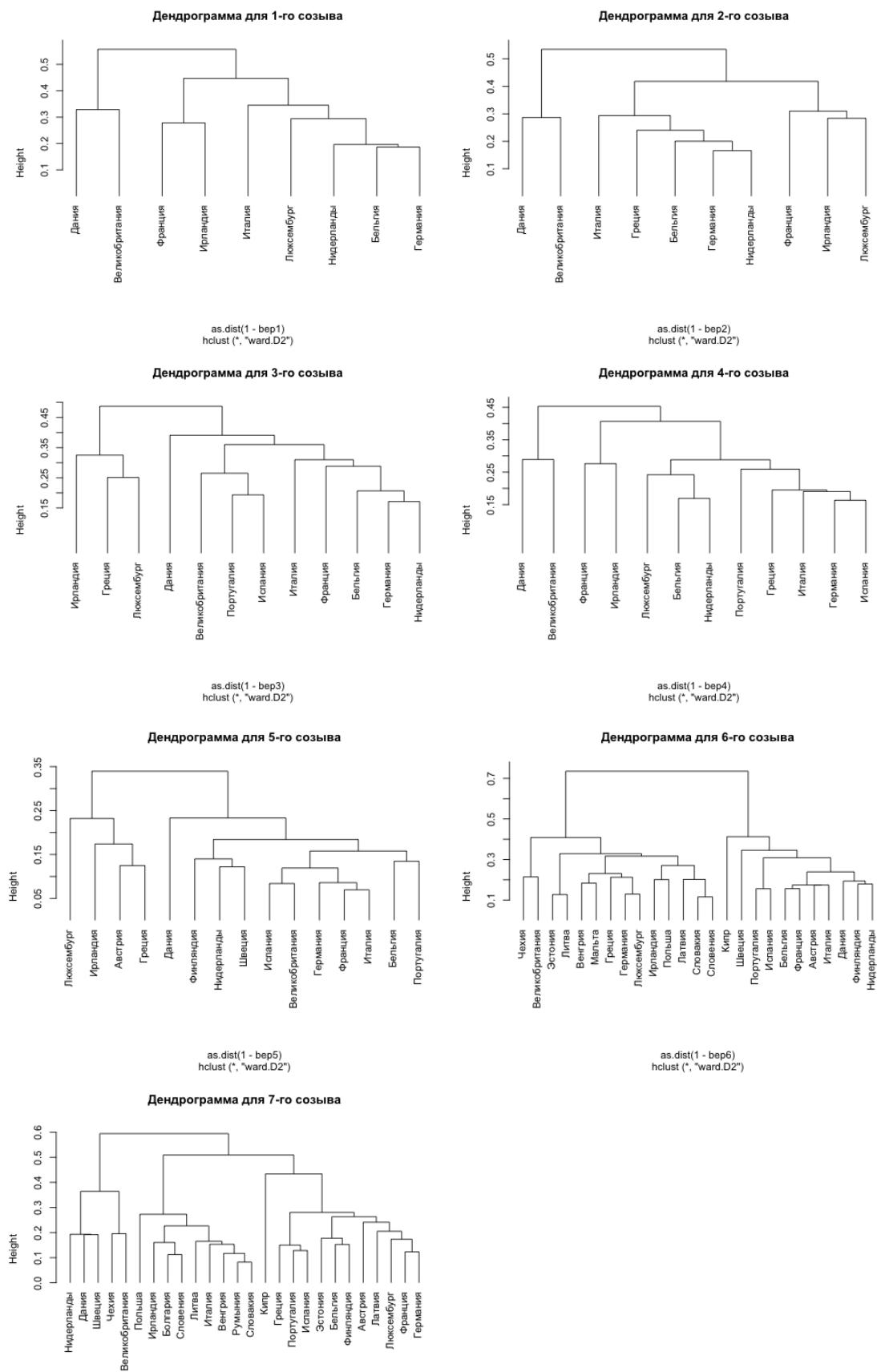


Рисунок В.2 – Кластеризация по индексу согласованности стран-членов ЕС по семи созывам

## ПРИЛОЖЕНИЕ Г

### Таблицы и рисунки к подразделу 3.2

**Таблица Г.1 – Политические группы Европейского парламента**

<b>Аббре- виатура</b>	<b>Название политической партии на анг- лийском</b>	<b>Название политической партии</b>
EDN	Europe of Nations	Европа наций
I-EDN	Independents for a Europe of Nations	Независимые за Европу наций
ED	European Democrats	Европейские демократы
EDA	European Democratic Alliance	Европейский демократический альянс
DEP	European Progressive Democrats	Европейские прогрессивные демократы
UPE	Union for a Europe of Nations	Союз за Европу наций
FE	Forza Europa	Сила Европы
EPP	European People's Party	Европейская народная партия
V	Greens	Зеленые
NI	Non-attached	Независимые / не вступившие
CDI	Technical Coordination and Defence of Independent Groups and Members	Техническая координация и защита независимых групп и членов
ELDR	European Liberal, Democrat and Reform Party	Европейская либеральная и демократическая партия реформ
LDR	Liberal Democrat and Reform Party	Либеральная демократическая партия реформ
COM	Communists	Коммунисты
EUL	European United Left	Европейские объединенные левые
GUE/NGL	European United Left/Nordic Green Left	Европейские объединенные левые / Зеленые Севера
LU	Left Unity	Левое объединение
ER	European Right	Европейские правые
ARE	European Radical Alliance	Европейский радикальный альянс
RB	Rainbow Group	Группа «Радуга»
PES	Party of European Socialists	Партия европейских социалистов
SOC	Socialists	Социалисты

Таблица Г.2 – Выигрыш партий (*gain*) в старых и новых странах ЕС (выигрыш рассчитывается как разница в проценте голосов, полученным партией на выборах в Европейский парламент и на последних национальных выборах)

	Старые члены ЕС, 15 государств			Новые члены ЕС, 10 государств		
	Средний выигрыш	Ст. ошибка	N	Средний выигрыш	Ст. ошибка	N
Правящая группа	-3	0,49	162	-5,2	2,25	24
Оппозиция	1,2	0,21	420	2,5	0,97	58
Еврискептики	9,5	1,39	22	3,3	2,45	3
Крайние правые	0,39	0,6	46	2,53	2,98	9
Радикальные левые	0,73	0,3	88	-2,53	1,66	6
Зеленые	2,75	0,34	67	1,63	4,2	3
Консерваторы	-1,9	0,87	65	0,58	1,25	25
Регионалисты	0,46	0,28	43	2,9	0,9	2
Социалисты и социал-демократы	-3,05	0,67	97	-1,23	5,29	12
Христианские демократы	-0,04	0,43	69	2,03	2,85	7
Либералы	-0,93	0,41	85	-1,65	2	15
Всего	-0,01	0,22	582	0,21	1,02	82

Таблица Г.3 – Средние, стандартные отклонения и число депутатов, по которым оценивалось влияние, для линейного индекса  $\alpha^1$  индивидуального влияния по семи созывам Европейского парламента

Созыв	N	Среднее	Ст. отклонение
1	253	39,5	5,9
2	269	37,2	5,3
3	327	30,6	3,4
4	428	23,4	2,7
5	491	20,4	3,3
6	359	27,9	4,7
7	554	18,1	3,3

Таблица Г.4 – Распределение мест между политическими группами в Европейском парламенте

Созыв	1	2	3	4	5	6	7
Бельгия	24	24	24	25	25	24	22
Великобритания	81	81	81	87	87	78	72
Германия	81	81	81	99	99	99	99
Дания	16	16	16	16	16	14	13
Ирландия	15	15	15	15	15	13	12
Италия	81	81	81	87	87	78	72
Люксембург	6	6	6	6	6	6	6
Нидерланды	25	25	25	31	31	27	25
Франция	81	81	81	87	87	78	72
Греция		24	24	25	25	24	22
Испания			60	64	64	54	50
Португалия			24	25	25	24	22
Австрия				21	21	18	17
Финляндия				16	16	14	13
Швеция				22	22	19	18
Венгрия						24	22
Кипр						6	6
Латвия						9	8
Литва						13	12
Мальта						5	5
Польша						54	50
Словакия						14	13
Словения						7	7
Чехия						24	22
Эстония						6	6
Болгария							18
Румыния							33

Таблица Г.5 – Матрица предпочтений политических групп по NOMINATE для 1 созыва ЕП

	Консерваторы	Христ. демократы	Нац.-консерваторы	Либералы	Левые и радик. левые	Регионалисты	Техническая группа	Социалисты и соц.-дем.
Консерваторы		0,420	0,194	0,383	0,894	0,378	0,537	0,367
Христ. демократы	0,420		0,687	0,838	0,214	0,839	0,237	0,265
Нац.-консерваторы	0,194	0,687		0,865	0,469	0,895	0,458	0,318
Либералы	0,383	0,838	0,865		0,372	0,995	0,399	0,387
Левые и радик. левые	0,894	0,214	0,469	0,372		0,369	0,912	0,659
Регионалисты	0,537	0,237	0,458	0,399	0,912	0,396		0,587
Техническая группа	0,378	0,839	0,895	0,995	0,369		0,396	0,382
Социалисты и соц.-дем.	0,367	0,265	0,318	0,387	0,659	0,382	0,587	

Таблица Г.6 – Матрица предпочтений политических групп по NOMINATE для 2 созыва ЕП

	Консерваторы	Христ. демократы	Нац.-консерваторы	Либералы	Левые и радик. левые	Техническая группа	Регионалисты	Социалисты и соц.-дем.	Националисты
Консерваторы		0,594	0,357	0,624	0,283	0,634	0,233	0,378	0,486
Христ. демократы	0,594		0,838	0,856	0,452	0,799	0,313	0,335	0,935
Нац.-консерваторы	0,357	0,838		0,727	0,425	0,693	0,328	0,329	0,819
Либералы	0,624	0,856	0,727		0,485	0,936	0,420	0,438	0,824
Левые и радик. левые	0,283	0,452	0,425	0,485		0,540	0,932	0,868	0,341
Техническая группа	0,634	0,799	0,693	0,936	0,540		0,459	0,499	0,742
Регионалисты	0,233	0,313	0,328	0,420	0,932	0,459		0,983	0,249
Социалисты и соц.-дем.	0,378	0,335	0,329	0,438	0,868	0,499	0,983		0,276
Националисты	0,486	0,935	0,819	0,824	0,341	0,742	0,249	0,276	

Таблица Г.7 – Матрица предпочтений политических групп по NOMINATE для 3 созыва ЕП

	Консерваторы	Христ. демократы	Нац.-консерваторы	Либералы	Левые и радик. левые	Техническая группа	Левая коалиция	Регионалисты	Социалисты и соц.-дем.	Зеленые	Националисты
Консерваторы		0,720	0,446	0,455	0,055	0,431	0,129	0,149	0,376	0,030	0,368
Христ. демократы	0,720		0,713	0,696	0,165	0,674	0,175	0,294	0,326	0,156	0,638
Нац.-консерваторы	0,446	0,713		0,910	0,311	0,909	0,249	0,466	0,279	0,321	0,923
Либералы	0,455	0,696	0,910		0,395	0,975	0,339	0,547	0,362	0,402	0,866
Левые и радик. левые	0,055	0,165	0,311	0,395		0,402	0,790	0,837	0,500	0,939	0,311
Техническая группа	0,431	0,674	0,909	0,975	0,402		0,338	0,556	0,349	0,411	0,880
Левая коалиция	0,129	0,175	0,249	0,339	0,790	0,338		0,686	0,675	0,729	0,228
Регионалисты	0,149	0,294	0,466	0,547	0,837	0,556	0,686		0,480	0,855	0,470
Социалисты и соц.-дем.	0,376	0,326	0,279	0,362	0,500	0,349	0,675	0,480		0,447	0,229
Зеленые	0,030	0,156	0,321	0,402	0,939	0,411	0,729	0,855	0,447		0,328
Националисты	0,368	0,638	0,923	0,866	0,311	0,880	0,228	0,470	0,229	0,328	

Таблица Г.8 – Матрица предпочтений политических групп по NOMINATE для 4 созыва ЕП

	Европсепти <sup>ики</sup>	Христ. демократы	Нац.- консервато <sup>ры</sup>	Либералы	Левые и радик. левые	Техничес <sup>кая</sup> группа	Регионали <sup>сты</sup>	Социалист <sup>ы и соц.- дем.</sup>	Зеленые
Европсептики	0,528	0,840	0,845	0,455	0,913	0,662	0,300	0,417	
Христ. демократы	0,528	0,652	0,588	0,164	0,545	0,385	0,350	0,116	
Нац.-консерваторы	0,840	0,652	0,776	0,320	0,890	0,542	0,264	0,279	
Либералы	0,845	0,588	0,776	0,533	0,778	0,761	0,453	0,489	
Левые и радик. левые	0,455	0,164	0,320	0,533	0,368	0,771	0,489	0,952	
Техническая группа	0,913	0,545	0,890	0,778	0,368	0,576	0,233	0,330	
Регионалисты	0,662	0,385	0,542	0,761	0,771	0,576	0,542	0,726	
Социалисты и соц.-дем.	0,300	0,350	0,264	0,453	0,489	0,233	0,542		0,453
Зеленые	0,417	0,116	0,279	0,489	0,952	0,330	0,726	0,453	

Таблица Г.9 – Матрица предпочтений политических групп по NOMINATE для 5 созыва ЕП

	Европсепти <sup>ики</sup>	Христ. демократы	Нац.- консервато <sup>ры</sup>	Либералы	Левые и радик. левые	Техничес <sup>кая</sup> группа	Зеленые	Социалист <sup>ы и соц.- дем.</sup>
Европсептики	0,923	0,954	0,885	0,865	0,902	0,859	0,921	
Христ. демократы	0,923	0,888	0,960	0,933	0,909	0,927	0,998	
Нац.-консерваторы	0,954	0,888	0,856	0,844	0,909	0,818	0,886	
Либералы	0,885	0,960	0,856	0,969	0,903	0,933	0,962	
Левые и радик. левые	0,865	0,933	0,844	0,969	0,909	0,908	0,934	
Техническая группа	0,902	0,909	0,909	0,903	0,909	0,841	0,907	
Зеленые	0,859	0,927	0,818	0,933	0,908	0,841		0,929
Социалисты и соц.-дем.	0,921	0,998	0,886	0,962	0,934	0,907	0,929	

Таблица Г.10 – Матрица предпочтений политических групп по NOMINATE для 6 созыва ЕП

	Европсепти <sup>ики</sup>	Христ. демократы	Нац.- консервато <sup>ры</sup>	Либералы	Левые и радик. левые	Техничес <sup>кая</sup> группа	Регионали <sup>сты</sup>	Социалист <sup>ы и соц.- дем.</sup>
Европсептики		0,489	0,866	0,626	0,595	0,978	0,635	0,226
Христ. демократы	0,489		0,622	0,834	0,240	0,505	0,465	0,667
Нац.-консерваторы	0,866	0,622		0,757	0,536	0,883	0,658	0,354
Либералы	0,626	0,834	0,757		0,406	0,644	0,626	0,596
Левые и радик. левые	0,595	0,240	0,536	0,406		0,602	0,744	0,109
Техническая группа	0,978	0,505	0,883	0,644	0,602		0,654	0,246
Регионалисты	0,635	0,465	0,658	0,626	0,744	0,654		0,365
Социалисты и соц.-дем.	0,226	0,667	0,354	0,596	0,109	0,246	0,365	

Таблица Г.11 – Матрица предпочтений политических групп по NOMINATE для 7 созыва ЕП

	Европекти <sup>ики</sup>	Консерват <sup>оры</sup>	Христ. демократы	Либералы	Левые и радик. левые	Техничес <sup>кая</sup> группа	Регионали <sup>сты</sup>	Социалист <sup>ы и соц.- дем.</sup>
Европекти <sup>ики</sup>		0,952	0,500	0,839	0,166	0,844	0,356	0,628
Консерваторы	0,952		0,496	0,791	0,118	0,796	0,308	0,594
Христ. демократы	0,500	0,496		0,466	0,131	0,462	0,132	0,679
Либералы	0,839	0,791	0,466		0,318	0,993	0,516	0,695
Левые и радик. левые	0,166	0,118	0,131	0,318		0,312	0,670	0,419
Техническая группа	0,844	0,796	0,462	0,993	0,312		0,512	0,688
Регионалисты	0,356	0,308	0,132	0,516	0,670	0,512		0,453
Социалисты и соц.-дем.	0,628	0,594	0,679	0,695	0,419	0,688	0,453	

Таблица Г.12 – Распределение мест между странами-членами ЕС в Европейском парламенте

Созыв	1	2	3	4	5	6	7
Европекти <sup>ики</sup>				19	16	37	23
Консерваторы	64	50	34				20
Христ. демократы	107	110	121	184	233	268	277
Нац.-консерваторы	22	29	20	26	31	27	
Либералы	40	31	49	43	50	88	106
Левые и радик. левые	44	41	28	28	42	41	40
Левая коалиция			14				
Техническая группа/ Независимые	9	7	12	27	26	29	14
Регионалисты / Регионалисты + Зеленые	11	20	13	19		42	42
Социалисты и соц.-дем.	113	130	180	198	180	200	218
Зеленые			30	23	48		
Националисты		16	17				

Таблица Г.13 – Матрица предпочтений стран-членов ЕС по NOMINATE для 1 созыва ЕП

	Бельгия	Великобр <sup>итания</sup>	Германия	Дания	Ирландия	Италия	Люксембу <sup>рг</sup>	Нидерлан <sup>ды</sup>	Франция
Бельгия		0,556	0,913	0,910	0,689	0,796	0,942	0,960	0,753
Великобритания	0,556		0,572	0,562	0,265	0,364	0,515	0,595	0,348
Германия	0,913	0,572		0,989	0,693	0,731	0,925	0,924	0,774
Дания	0,910	0,562	0,989		0,703	0,735	0,929	0,917	0,785
Ирландия	0,689	0,265	0,693	0,703		0,759	0,743	0,658	0,908
Италия	0,796	0,364	0,731	0,735	0,759		0,805	0,756	0,756
Люксембург	0,942	0,515	0,925	0,929	0,743	0,805		0,915	0,810
Нидерланды	0,960	0,595	0,924	0,917	0,658	0,756	0,915		0,727
Франция	0,753	0,348	0,774	0,785	0,908	0,756	0,810	0,727	

Таблица Г.14 – Матрица предпочтений стран-членов ЕС по NOMINATE для 2 созыва ЕП

	Бельгия	Великобритания	Германия	Дания	Ирландия	Италия	Люксембург	Нидерланды	Франция	Греция
Бельгия	0,645	0,900	0,756	0,691	0,969	0,762	0,927	0,817	0,939	
Великобритания	0,645		0,621	0,845	0,402	0,626	0,499	0,705	0,493	0,589
Германия	0,900	0,621		0,697	0,770	0,926	0,854	0,904	0,872	0,869
Дания	0,756	0,845	0,697		0,467	0,730	0,555	0,793	0,579	0,710
Ирландия	0,691	0,402	0,770	0,467		0,722	0,902	0,674	0,867	0,700
Италия	0,969	0,626	0,926	0,730	0,722		0,792	0,919	0,846	0,937
Люксембург	0,762	0,499	0,854	0,555	0,902	0,792		0,760	0,895	0,754
Нидерланды	0,927	0,705	0,904	0,793	0,674	0,919	0,760		0,783	0,867
Франция	0,817	0,493	0,872	0,579	0,867	0,846	0,895	0,783		0,832
Греция	0,939	0,589	0,869	0,710	0,700	0,937	0,754	0,867	0,832	

Таблица Г.15 – Матрица предпочтений стран-членов ЕС по NOMINATE для 3 созыва ЕП

	Бельгия	Великобритания	Германия	Дания	Ирландия	Италия	Люксембург	Нидерланды	Франция	Греция	Испания	Португалия
Бельгия	0,721	0,897	0,879	0,740	0,937	0,796	0,902	0,887	0,884	0,792	0,955	
Великобритания	0,721		0,781	0,699	0,517	0,714	0,711	0,760	0,621	0,776	0,928	0,750
Германия	0,897	0,781		0,794	0,735	0,846	0,879	0,979	0,832	0,984	0,842	0,886
Дания	0,879	0,699	0,794		0,635	0,943	0,680	0,791	0,794	0,778	0,767	0,908
Ирландия	0,740	0,517	0,735	0,635		0,687	0,778	0,755	0,840	0,741	0,576	0,695
Италия	0,937	0,714	0,846	0,943	0,687		0,735	0,845	0,844	0,830	0,786	0,957
Люксембург	0,796	0,711	0,879	0,680	0,778	0,735		0,888	0,791	0,895	0,752	0,770
Нидерланды	0,902	0,760	0,979	0,791	0,755	0,845	0,888		0,850	0,978	0,821	0,882
Франция	0,887	0,621	0,832	0,794	0,840	0,844	0,791	0,850		0,828	0,689	0,844
Греция	0,884	0,776	0,984	0,778	0,741	0,830	0,895	0,978	0,828		0,834	0,870
Испания	0,792	0,928	0,842	0,767	0,576	0,786	0,752	0,821	0,689	0,834		0,821
Португалия	0,955	0,750	0,886	0,908	0,695	0,957	0,770	0,882	0,844	0,870	0,821	

Таблица Г.16 – Матрица предпочтений стран-членов ЕС по NOMINATE для 4 созыва ЕП

	Бельгия	Великобритания	Германия	Дания	Ирландия	Италия	Люксембург	Нидерланды	Франция	Греция	Испания	Португалия	Австрия	Финляндия	Швеция
Бельгия		0,65	0,83	0,83	0,85	0,91	0,88	0,92	0,81	0,95	0,84	0,93	0,79	0,88	0,88
Великобритания	0,65		0,78	0,67	0,50	0,69	0,77	0,71	0,46	0,68	0,78	0,71	0,82	0,68	0,69
Германия	0,83	0,78		0,74	0,69	0,90	0,91	0,91	0,64	0,87	0,98	0,86	0,83	0,78	0,78
Дания	0,83	0,67	0,74		0,74	0,77	0,83	0,80	0,72	0,80	0,76	0,86	0,85	0,95	0,95
Ирландия	0,85	0,50	0,69	0,74		0,79	0,73	0,78	0,94	0,82	0,71	0,79	0,65	0,77	0,77
Италия	0,91	0,69	0,90	0,77	0,79		0,90	0,97	0,74	0,96	0,91	0,91	0,79	0,82	0,82
Люксембург	0,88	0,77	0,91	0,83	0,73	0,90		0,93	0,69	0,90	0,93	0,94	0,89	0,86	0,87
Нидерланды	0,92	0,71	0,91	0,80	0,78	0,97	0,93		0,73	0,96	0,92	0,94	0,82	0,84	0,85
Франция	0,81	0,46	0,64	0,72	0,94	0,74	0,69	0,73		0,77	0,66	0,75	0,62	0,75	0,75
Греция	0,95	0,68	0,87	0,80	0,82	0,96	0,90	0,96	0,77		0,89	0,93	0,80	0,85	0,85
Испания	0,84	0,78	0,98	0,76	0,71	0,91	0,93	0,92	0,66	0,89		0,88	0,84	0,80	0,80
Португалия	0,93	0,71	0,86	0,86	0,79	0,91	0,94	0,94	0,75	0,93	0,88		0,86	0,91	0,91
Австрия	0,79	0,82	0,83	0,85	0,65	0,79	0,89	0,82	0,62	0,80	0,84	0,86		0,86	0,87
Финляндия	0,88	0,68	0,78	0,95	0,77	0,82	0,86	0,84	0,75	0,85	0,80	0,91	0,86		1,00
Швеция	0,88	0,69	0,78	0,95	0,77	0,82	0,87	0,85	0,75	0,85	0,80	0,91	0,87	1,00	

Таблица Г.17 – Матрица предпочтений стран-членов ЕС по NOMINATE для 5 созыва ЕП

	Бельгия	Великобритания	Германия	Дания	Ирландия	Италия	Люксембург	Нидерланды	Франция	Греция	Испания	Португалия	Австрия	Финляндия	Швеция
Бельгия		0,83	0,84	0,75	0,95	0,80	0,68	0,90	0,85	0,95	0,82	0,71	0,73	0,92	0,93
Великобритания	0,83		0,92	0,78	0,85	0,93	0,84	0,92	0,91	0,86	0,96	0,85	0,91	0,79	0,89
Германия	0,84	0,92		0,86	0,84	0,96	0,84	0,94	0,98	0,85	0,96	0,88	0,87	0,83	0,88
Дания	0,75	0,78	0,86		0,73	0,85	0,78	0,82	0,87	0,74	0,82	0,84	0,76	0,80	0,77
Ирландия	0,95	0,85	0,84	0,73		0,81	0,69	0,90	0,84	0,99	0,83	0,72	0,76	0,87	0,95
Италия	0,80	0,93	0,96	0,85	0,81		0,88	0,90	0,94	0,81	0,96	0,91	0,91	0,79	0,85
Люксембург	0,68	0,84	0,84	0,78	0,69	0,88		0,78	0,82	0,70	0,86	0,94	0,92	0,67	0,74
Нидерланды	0,90	0,92	0,94	0,82	0,90	0,90	0,78		0,94	0,91	0,92	0,81	0,83	0,87	0,95
Франция	0,85	0,91	0,98	0,87	0,84	0,94	0,82	0,94		0,85	0,94	0,86	0,85	0,85	0,89
Греция	0,95	0,86	0,85	0,74	0,99	0,81	0,70	0,91	0,85		0,84	0,73	0,76	0,88	0,96
Испания	0,82	0,96	0,96	0,82	0,83	0,96	0,86	0,92	0,94	0,84		0,88	0,91	0,80	0,88
Португалия	0,71	0,85	0,88	0,84	0,72	0,91	0,94	0,81	0,86	0,73	0,88		0,90	0,71	0,77
Австрия	0,73	0,91	0,87	0,76	0,76	0,91	0,92	0,83	0,85	0,76	0,91	0,90		0,71	0,80
Финляндия	0,92	0,79	0,83	0,80	0,87	0,79	0,67	0,87	0,85	0,88	0,80	0,71	0,71		0,87
Швеция	0,93	0,89	0,88	0,77	0,95	0,85	0,74	0,95	0,89	0,96	0,88	0,77	0,80	0,87	

Таблица Г.18 – Матрица предпочтений стран-членов ЕС по NOMINATE для 6 созыва ЕП

	Бельгия	Великобритания	Германия	Дания	Ирландия	Италия	Люксембург	Нидерланды	Франция	Греция	Испания	Португалия	Австрия	Финляндия	Швеция	Венгрия	Кипр	Латвия	Литва	Мальта	Польша	Словакия	Словения	Чехия	Эстония	
Бельгия	0,88	0,88	0,96	0,92	0,96	0,85	0,95	0,98	0,90	0,77	0,79	0,93	0,88	0,78	0,72	0,67	0,89	0,93	0,78	0,88	0,80	0,84	0,85	0,83		
Великобритания	0,88		0,78	0,87	0,91	0,89	0,76	0,88	0,87	0,81	0,66	0,67	0,81	0,77	0,79	0,62	0,72	0,96	0,85	0,69	0,99	0,75	0,78	0,96	0,74	
Германия	0,88	0,78		0,85	0,86	0,84	0,97	0,83	0,87	0,97	0,87	0,84	0,91	0,94	0,67	0,84	0,54	0,81	0,93	0,90	0,79	0,90	0,92	0,75	0,95	
Дания	0,96	0,87	0,85		0,89	0,98	0,83	0,98	0,98	0,87	0,76	0,79	0,93	0,87	0,81	0,70	0,69	0,86	0,90	0,75	0,87	0,77	0,80	0,83	0,81	
Ирландия	0,92	0,91	0,86	0,89		0,89	0,84	0,88	0,90	0,89	0,73	0,73	0,86	0,84	0,73	0,70	0,64	0,94	0,93	0,78	0,92	0,84	0,87	0,88	0,82	
Италия	0,96	0,89	0,84	0,98	0,89		0,82	0,99	0,97	0,86	0,75	0,77	0,91	0,85	0,82	0,68	0,70	0,88	0,89	0,74	0,88	0,76	0,80	0,85	0,80	
Люксембург	0,85	0,76	0,97	0,83	0,84	0,82		0,81	0,85	0,95	0,89	0,85	0,89	0,94	0,65	0,86	0,52	0,78	0,91	0,92	0,76	0,90	0,91	0,72	0,98	
Нидерланды	0,95	0,88	0,83	0,98	0,88	0,99	0,81		0,96	0,85	0,74	0,77	0,90	0,85	0,84	0,68	0,71	0,86	0,88	0,73	0,87	0,75	0,78	0,84	0,79	
Франция	0,98	0,87	0,87	0,98	0,90	0,97	0,85	0,96		0,89	0,78	0,80	0,94	0,89	0,79	0,72	0,67	0,87	0,92	0,77	0,87	0,79	0,82	0,83	0,83	
Греция	0,90	0,81	0,97	0,87	0,89	0,86	0,95	0,85	0,89		0,84	0,82	0,91	0,93	0,68	0,81	0,56	0,84	0,96	0,88	0,82	0,90	0,92	0,78	0,93	
Испания	0,77	0,66	0,87	0,76	0,73	0,75	0,89	0,74	0,78	0,84		0,93	0,84	0,89	0,60	0,90	0,46	0,68	0,80	0,87	0,67	0,80	0,80	0,62	0,90	
Португалия	0,79	0,67	0,84	0,79	0,73	0,77	0,85	0,77	0,80	0,82	0,93		0,86	0,89	0,65	0,83	0,51	0,68	0,80	0,81	0,67	0,75	0,76	0,63	0,86	
Австрия	0,93	0,81	0,91	0,93	0,86	0,91	0,89	0,90	0,94	0,91	0,84	0,86		0,95	0,75	0,77	0,62	0,82	0,92	0,81	0,81	0,81	0,83	0,77	0,88	
Финляндия	0,88	0,77	0,94	0,87	0,84	0,85	0,94	0,85	0,89	0,93	0,89	0,89	0,95		0,70	0,83	0,56	0,79	0,91	0,86	0,77	0,84	0,86	0,73	0,93	
Швеция	0,78	0,79	0,67	0,81	0,73	0,82	0,65	0,84	0,79	0,68	0,60	0,65	0,75	0,70		0,52	0,86	0,75	0,72	0,57	0,77	0,59	0,62	0,77	0,63	
Венгрия	0,72	0,62	0,84	0,70	0,70	0,68	0,86	0,68	0,72	0,81	0,90	0,83	0,77	0,83	0,52		0,39	0,65	0,77	0,91	0,63	0,63	0,82	0,81	0,59	0,88
Кипр	0,67	0,72	0,54	0,69	0,64	0,70	0,52	0,71	0,67	0,56	0,46	0,51	0,62	0,56	0,86	0,39		0,67	0,60	0,44	0,70	0,48	0,51	0,72	0,50	
Латвия	0,89	0,96	0,81	0,86	0,94	0,88	0,78	0,86	0,87	0,84	0,68	0,68	0,82	0,79	0,75	0,65	0,67	0,88	0,73	0,97	0,79	0,82	0,94	0,76		
Литва	0,93	0,85	0,93	0,90	0,93	0,89	0,91	0,88	0,92	0,96	0,80	0,80	0,92	0,91	0,72	0,77	0,60	0,88		0,84	0,86	0,87	0,90	0,82	0,88	
Мальта	0,78	0,69	0,90	0,75	0,78	0,74	0,92	0,73	0,77	0,88	0,87	0,81	0,81	0,86	0,57	0,91	0,44	0,73	0,84		0,70	0,91	0,89	0,66	0,93	
Польша	0,88	0,99	0,79	0,87	0,92	0,88	0,76	0,87	0,87	0,82	0,67	0,67	0,81	0,77	0,77	0,63	0,70	0,97	0,86	0,70		0,77	0,80	0,96	0,74	
Словакия	0,80	0,75	0,90	0,77	0,84	0,76	0,90	0,75	0,79	0,90	0,80	0,75	0,81	0,84	0,59	0,82	0,48	0,79	0,87	0,91	0,77		0,97	0,73	0,89	
Словения	0,84	0,78	0,92	0,80	0,87	0,80	0,91	0,78	0,82	0,92	0,80	0,76	0,83	0,86	0,62	0,81	0,51	0,82	0,90	0,89	0,80	0,97		0,76	0,89	
Чехия	0,85	0,96	0,75	0,83	0,88	0,85	0,72	0,84	0,83	0,78	0,62	0,63	0,77	0,73	0,77	0,59	0,72	0,94	0,82	0,66	0,96	0,73	0,76		0,70	
Эстония	0,83	0,74	0,95	0,81	0,82	0,80	0,98	0,79	0,83	0,93	0,90	0,86	0,88	0,93	0,63	0,88	0,50	0,76	0,88	0,93	0,74	0,89	0,89	0,70		

Таблица Г.19 – Матрица предпочтений стран-членов ЕС по NOMINATE для 7 созыва ЕП

	Бельгия	Великобритания	Германия	Дания	Ирландия	Италия	Люксембург	Нидерланды	Франция	Греция	Испания	Португалия	Австрия	Финляндия	Швеция	Венгрия	Кипр	Латвия	Литва	Мальта	Польша	Словакия	Словения	Чехия	Эстония	Болгария	Румыния
Бельгия		0,81	0,92	0,80	0,84	0,81	0,90	0,88	0,88	0,80	0,77	0,76	0,90	0,98	0,82	0,68	0,62	0,82	0,82	0,78	0,75	0,79	0,84	0,92	0,94	0,88	0,76
Великобритания	0,81		0,72	0,91	0,70	0,67	0,72	0,86	0,69	0,61	0,60	0,57	0,78	0,83	0,85	0,53	0,48	0,63	0,70	0,69	0,66	0,63	0,70	0,87	0,85	0,73	0,60
Германия	0,92	0,72		0,73	0,85	0,83	0,92	0,81	0,97	0,87	0,82	0,84	0,87	0,90	0,76	0,73	0,68	0,91	0,82	0,90	0,75	0,82	0,85	0,84	0,86	0,88	0,80
Дания	0,80	0,91	0,73		0,66	0,63	0,70	0,91	0,70	0,60	0,57	0,57	0,83	0,82	0,94	0,49	0,54	0,65	0,65	0,75	0,60	0,60	0,66	0,84	0,86	0,69	0,56
Ирландия	0,84	0,70	0,85	0,66		0,97	0,93	0,73	0,83	0,85	0,89	0,83	0,75	0,83	0,67	0,83	0,54	0,80	0,96	0,78	0,89	0,93	1,00	0,82	0,78	0,97	0,90
Италия	0,81	0,67	0,83	0,63	0,97		0,90	0,69	0,81	0,85	0,91	0,84	0,72	0,80	0,63	0,86	0,52	0,78	0,96	0,74	0,91	0,96	0,97	0,79	0,75	0,93	0,93
Люксембург	0,90	0,72	0,92	0,70	0,93	0,90		0,78	0,90	0,88	0,87	0,85	0,81	0,88	0,72	0,78	0,60	0,85	0,90	0,92	0,82	0,89	0,93	0,85	0,84	0,96	0,86
Нидерланды	0,88	0,86	0,81	0,91	0,73	0,69	0,78		0,79	0,69	0,65	0,65	0,92	0,90	0,94	0,56	0,61	0,74	0,71	0,73	0,65	0,67	0,72	0,88	0,94	0,76	0,64
Франция	0,88	0,69	0,97	0,70	0,83	0,81	0,90	0,79		0,89	0,82	0,86	0,86	0,86	0,73	0,73	0,70	0,94	0,80	0,77	0,73	0,82	0,83	0,80	0,83	0,86	0,79
Греция	0,80	0,61	0,87	0,60	0,85	0,85	0,88	0,69	0,89		0,91	0,96	0,75	0,78	0,63	0,82	0,65	0,90	0,82	0,88	0,77	0,88	0,85	0,73	0,74	0,86	0,87
Испания	0,77	0,60	0,82	0,57	0,89	0,91	0,87	0,65	0,82	0,91		0,91	0,69	0,75	0,59	0,91	0,56	0,82	0,87	0,87	0,83	0,95	0,89	0,72	0,70	0,87	0,96
Португалия	0,76	0,57	0,84	0,57	0,83	0,84	0,85	0,65	0,86	0,96	0,91		0,71	0,74	0,60	0,84	0,64	0,88	0,80	0,81	0,75	0,87	0,83	0,70	0,70	0,83	0,87
Австрия	0,90	0,78	0,87	0,83	0,75	0,72	0,81	0,92	0,86	0,75	0,69	0,71		0,90	0,88	0,60	0,69	0,81	0,72	0,81	0,65	0,70	0,74	0,85	0,92	0,78	0,67
Финляндия	0,98	0,83	0,90	0,82	0,83	0,80	0,88	0,90	0,86	0,78	0,75	0,74	0,90		0,84	0,67	0,61	0,80	0,81	0,86	0,74	0,77	0,83	0,94	0,95	0,86	0,74
Швеция	0,82	0,85	0,76	0,94	0,67	0,63	0,72	0,94	0,73	0,63	0,59	0,60	0,88	0,84		0,50	0,60	0,69	0,65	0,57	0,59	0,61	0,66	0,83	0,88	0,70	0,58
Венгрия	0,68	0,53	0,73	0,49	0,83	0,86	0,78	0,56	0,73	0,82	0,91	0,84	0,60	0,67	0,50		0,48	0,73	0,83	0,91	0,83	0,89	0,83	0,65	0,62	0,80	0,93
Кипр	0,62	0,48	0,68	0,54	0,54	0,52	0,60	0,61	0,70	0,65	0,56	0,64	0,69	0,61	0,60	0,48		0,74	0,50	0,44	0,43	0,53	0,54	0,55	0,61	0,56	0,52
Латвия	0,82	0,63	0,91	0,65	0,80	0,78	0,85	0,74	0,94	0,90	0,82	0,88	0,81	0,80	0,69	0,73	0,74		0,76	0,73	0,69	0,79	0,79	0,74	0,77	0,81	0,78
Литва	0,82	0,70	0,82	0,65	0,96	0,96	0,90	0,71	0,80	0,82	0,87	0,80	0,72	0,81	0,65	0,83	0,50	0,76		0,84	0,93	0,92	0,97	0,81	0,77	0,94	0,89
Мальта	0,78	0,69	0,90	0,75	0,78	0,74	0,92	0,73	0,77	0,88	0,87	0,81	0,81	0,86	0,57	0,91	0,44	0,73	0,84		0,70	0,91	0,89	0,66	0,93	0,79	0,79
Польша	0,75	0,66	0,75	0,60	0,89	0,91	0,82	0,65	0,73	0,77	0,83	0,75	0,65	0,74	0,59	0,83	0,43	0,69	0,93	0,70		0,89	0,90	0,76	0,70	0,87	0,87
Словакия	0,79	0,63	0,82	0,60	0,93	0,96	0,89	0,67	0,82	0,88	0,95	0,87	0,70	0,77	0,61	0,89	0,53	0,79	0,92	0,91	0,89		0,94	0,76	0,73	0,91	0,97
Словения	0,84	0,70	0,85	0,66	1,00	0,97	0,93	0,72	0,83	0,85	0,89	0,83	0,74	0,83	0,66	0,83	0,54	0,79	0,97	0,89	0,90	0,94		0,82	0,78	0,96	0,90
Чехия	0,92	0,87	0,84	0,84	0,82	0,79	0,85	0,88	0,80	0,73	0,72	0,70	0,85	0,94	0,83	0,65	0,55	0,74	0,81	0,66	0,76	0,76	0,82	0,93	0,85	0,72	
Эстония	0,94	0,85	0,86	0,86	0,78	0,75	0,84	0,94	0,83	0,74	0,70	0,70	0,92	0,95	0,88	0,62	0,61	0,77	0,77	0,93	0,70	0,73	0,78	0,93		0,82	0,69
Болгария	0,88	0,73	0,88	0,69	0,97	0,93	0,96	0,76	0,86	0,86	0,87	0,83	0,78	0,86	0,70	0,80	0,56	0,81	0,94	0,79	0,87	0,91	0,96	0,85	0,82		0,87
Румыния	0,76	0,60	0,80	0,56	0,90	0,93	0,86	0,64	0,79	0,87	0,96	0,87	0,67	0,74	0,58	0,93	0,52	0,78	0,89	0,79	0,87	0,97	0,90	0,72	0,69	0,87	

Таблица Г.20 – Индекс Банцафа политических групп ( $\beta$ ) и индекс влияния с учетом предпочтений по созданию коалиций ( $\alpha_{gr}^1$ )

Созыв	1		2		3		4		5		6		7	
Группы	$\beta$	$\alpha_{gr}^1$												
Евросkeptики							0,05	0,07	0,04	0,04	0,07	0,08	0,03	0,03
Консерваторы	0,17	0,11	0,10	0,08	0,07	0,05						0,03	0,03	
Христ. демократы	0,26	0,29	0,27	0,30	0,13	0,13	0,20	0,17	0,39	0,39	0,39	0,39	0,37	0,30
Нац.-консерваторы	0,05	0,05	0,06	0,07	0,05	0,06	0,07	0,08	0,06	0,06	0,04	0,05		
Либералы	0,09	0,12	0,06	0,08	0,11	0,14	0,13	0,17	0,11	0,11	0,15	0,18	0,21	0,27
Левые и радик. левые	0,10	0,09	0,08	0,08	0,06	0,06	0,07	0,07	0,09	0,09	0,07	0,06	0,06	0,04
Левая коалиция					0,03	0,03								
Техническая группа/ Независимые	0,02	0,03	0,01	0,01	0,02	0,03	0,07	0,08	0,04	0,04	0,04	0,05	0,03	0,04
Регионалисты / Регионалисты + Зеленые	0,02	0,02	0,04	0,04	0,03	0,03	0,05	0,07			0,07	0,08	0,06	0,05
Социалисты и соц.-дем.	0,29	0,28	0,34	0,31	0,40	0,36	0,30	0,23	0,16	0,16	0,17	0,12	0,21	0,24
Зеленые					0,06	0,07	0,05	0,05	0,09	0,10				
Националисты			0,03	0,04	0,04	0,04								

Таблица Г.21 – Индекс Банцафа стран ЕС ( $\beta$ ) и индекс влияния стран с учетом предпочтений по созданию коалиций ( $\alpha_{gr}^1$ )

Созыв	1	$\alpha_{gr}^1$	2	$\alpha_{gr}^1$	3	$\alpha_{gr}^1$	4	$\alpha_{gr}^1$	5	$\alpha_{gr}^1$	6	$\alpha_{gr}^1$	7	$\alpha_{gr}^1$
Группы	$\beta$	$\alpha_{gr}^1$	$\beta$	$\alpha_{gr}^1$	$\beta$	$\alpha_{gr}^1$	$B$	$\alpha_{gr}^1$	$\beta$	$\alpha_{gr}^1$	$\beta$	$\alpha_{gr}^1$	$\beta$	$\alpha_{gr}^1$
Бельгия	0,06	0,07	0,06	0,07	0,04	0,04	0,04	0,04	0,04	0,04	0,03	0,03	0,03	0,03
Великобритания	0,17	0,12	0,17	0,13	0,17	0,15	0,14	0,12	0,14	0,14	0,11	0,11	0,10	0,09
Германия	0,17	0,19	0,17	0,18	0,17	0,18	0,17	0,17	0,17	0,17	0,14	0,15	0,14	0,15
Дания	0,05	0,06	0,05	0,05	0,03	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
Ирландия	0,05	0,05	0,05	0,05	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
Италия	0,17	0,18	0,17	0,18	0,17	0,17	0,14	0,15	0,14	0,15	0,11	0,11	0,10	0,10
Люксембург	0,01	0,01	0,01	0,01	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Нидерланды	0,07	0,08	0,07	0,08	0,04	0,04	0,04	0,05	0,04	0,04	0,04	0,04	0,03	0,03
Франция	0,17	0,17	0,17	0,17	0,17	0,17	0,14	0,13	0,14	0,15	0,11	0,11	0,10	0,10
Греция	0,06	0,07	0,06	0,07	0,04	0,04	0,04	0,04	0,04	0,04	0,03	0,03	0,03	0,03
Испания					0,13	0,13	0,11	0,11	0,11	0,11	0,07	0,07	0,07	0,07
Португалия					0,04	0,04	0,04	0,04	0,04	0,03	0,03	0,03	0,03	0,03
Австрия							0,03	0,03	0,03	0,03	0,02	0,02	0,02	0,02
Финляндия							0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
Швеция							0,03	0,03	0,03	0,03	0,02	0,02	0,02	0,02
Венгрия										0,03	0,03	0,03	0,03	0,03
Кипр										0,01	0,01	0,01	0,01	0,01
Латвия										0,01	0,01	0,01	0,01	0,01
Литва										0,02	0,02	0,02	0,02	0,02
Мальта										0,01	0,01	0,01	0,01	0,01
Польша										0,07	0,07	0,07	0,07	0,06
Словакия										0,02	0,02	0,02	0,02	0,02
Словения										0,01	0,01	0,01	0,01	0,01
Чехия										0,03	0,03	0,03	0,03	0,03
Эстония										0,01	0,01	0,01	0,01	0,01
Болгария												0,02	0,02	0,03
Румыния												0,04	0,04	0,04

На рисунке Г.1 показана динамика значений коэффициента ранговой корреляции Спирмена для индекса индивидуального влияния  $\alpha^1$ , рассчитанного без фильтрации по информативности, для расчетов используются все голосования ( $m^t = 100$ ), и индекса влияния, рассчитанного с некоторым порогом от 50 до 95% (по оси ОХ). Третьим измерением на данном рисунке является дисциплинированность депутатов: линиями отмечены разные значения порога активности парламентариев.

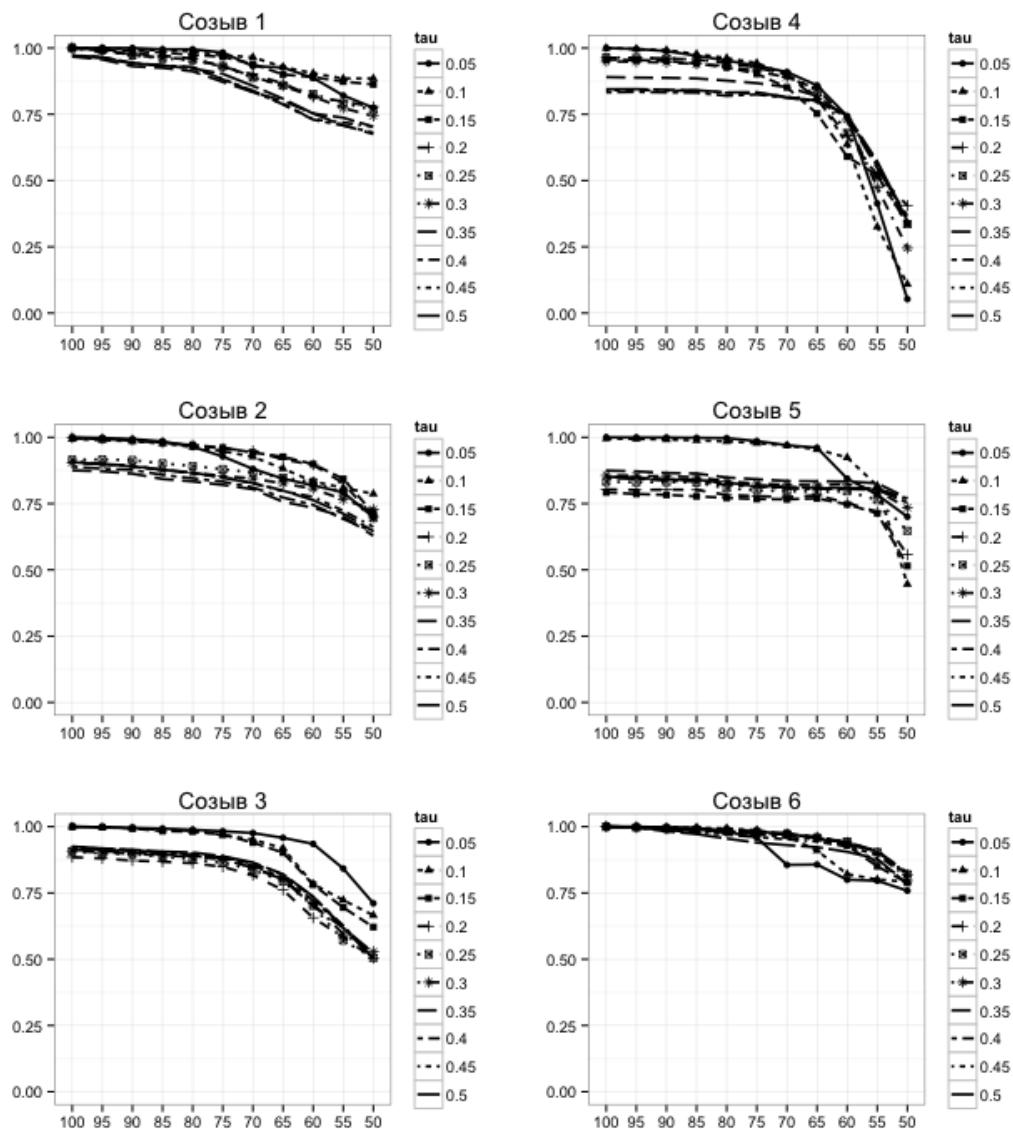


Рисунок Г.1 – Анализ корреляций индексов влияния в зависимости от пороговых значений  $m^t$  и  $\tau$ . Сравнение с базовым значением  $m^t = 100$

Видно, что если считать неинформативными голосования с согласованностью выше 85-100%, полученные значения индекса влияния почти не от-

личны друг от друга для разных значений порога «активности» депутатов. Наибольшая чувствительность результатов характерна для второго и четвертого созывов.

На рисунке Г.2 показана динамика коэффициента корреляции Спирмена для индекса влияния, рассчитанного по порогу активности 0,50, и индекса влияния, рассчитанного с другим порогом — от 0,45 до 0,05 (по оси ОХ). Разными линиями отмечены значения порога информативности голосований — от 50 до 100%.

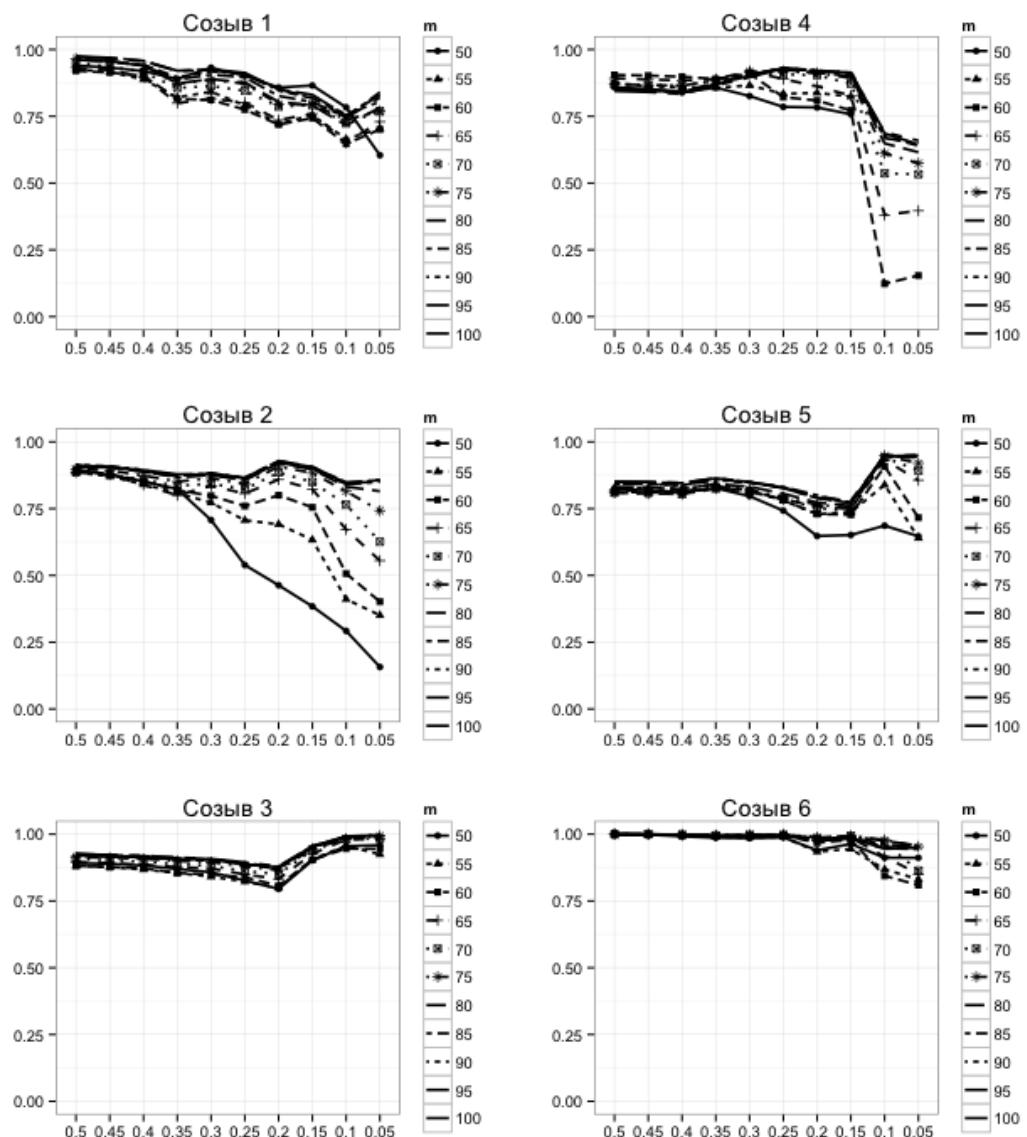


Рисунок Г.2 – Анализ корреляций индексов влияния в зависимости от пороговых значений  $m^t$  и  $\tau$ . Сравнение с базовым значением  $\tau = 0,50$

Заметим, что зачастую число депутатов, не пропустивших за созыв ни одного поименного голосования, равно нулю, поэтому стремление включить только самых дисциплинированных чревато тем, чем в анализе будут участвовать единицы депутатов. Чем ниже порог для включения депутата в число активных (например, 0,45), тем меньше отличаются результаты при оценке влияния для разных порогов информативности. Можно говорить, что в целом результаты более чувствительны, чем при изменении порога  $m^t$ . Разумным выглядит порог  $\tau = 0.25$  — результаты оценки влияния достаточно устойчивы к смене значения параметра  $m^t$ , в то же время депутаты, включенные в расчеты, могут считаться дисциплинированными и активными и, вероятнее, более последовательными. Таким образом, в среднем были исключены 26% депутатов (в первом созыве — 37%, в седьмом — 23%).