

НАУЧНАЯ МЫСЛЬ

СЕРИЯ ОСНОВАНА В 2008 ГОДУ

Министерство науки и высшего образования Российской Федерации
Сибирский федеральный университет

А.А. СТУПИНА
И.Р. РУЙГА
Л.А. КАЗАКОВЦЕВ

**МЕТОДЫ МОДЕЛИРОВАНИЯ
И ОПТИМИЗАЦИИ
ДЛЯ ОБЕСПЕЧЕНИЯ
ПОДДЕРЖКИ ПРИНЯТИЯ
РЕШЕНИЙ В СЛОЖНЫХ
СИСТЕМАХ**

МОНОГРАФИЯ

Москва
ИНФРА-М
2023

УДК 519.816(075.4)

ББК 22.171

C88

*Исследование выполнено при поддержке Министерства
науки и высшего образования Российской Федерации в рамках
гранта № 075-15-2022-1121*

Рецензенты:

Масич И.С., доктор технических наук, профессор кафедры системного анализа и исследования операций Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева (г. Красноярск);

Сташков Д.В., кандидат технических наук, ведущий специалист по автоматизированным системам управления технологическим процессом АО «СИНЕТИК» (г. Новосибирск)

Ступина А.А.

C88 Методы моделирования и оптимизации для обеспечения поддержки принятия решений в сложных системах : монография / А.А. Ступина, И.Р. Руйга, Л.А. Казаковцев. — Москва : ИНФРА-М, 2023. — 205 с. : ил. — (Научная мысль).

ISBN 978-5-16-019632-9

В монографии раскрыты теоретические и методические аспекты моделирования и оптимизации для обеспечения поддержки принятия решений в сложных системах. Представлена практическая реализация ряда методов и моделей в социально-экономических и производственных системах.

Предназначена для научных работников, специалистов, студентов и аспирантов, занимающихся вопросами моделирования и оптимизации для обеспечения поддержки принятия решений в сложных системах.

УДК 519.816(075.4)

ББК 22.171

© Ступина А.А., Руйга И.Р.,
Казаковцев Л.А., 2023

© Сибирский федеральный
университет, 2023

ISBN 978-5-16-019632-9

Введение

Актуальность исследования обусловлена необходимостью развития методической базы и математического аппарата управления сложными системами на основе методов моделирования и оптимизации для повышения обоснованности управленческих решений в условиях роста и бурного развития систем искусственного интеллекта, а также повышением общих требований к экономической эффективности, достигаемой, в том числе, за счет применения оптимизационных моделей.

Основные теоретические и прикладные результаты получены с применением методов системного анализа, исследования операций, теории оптимизации, теории вероятностей, параллельного программирования.

Теоретической и методологической базой исследования послужили работы зарубежных и российских авторов в аспекте накопленного опыта построения математических моделей из различных отраслей науки, техники, промышленности и экономики.

Информационная база исследования: нормативно-правовые акты Российской Федерации федерального и регионального уровня, данные Федеральной службы государственной статистики и ее территориальных отделений, программные и информационно-аналитические документы федеральных и региональных органов власти, конъюнктурные исследования, аналитические отчеты, прогнозы.

Монография предназначена для научных работников, специалистов, студентов и аспирантов, занимающихся вопросами моделирования и оптимизации для обеспечения поддержки принятия решений в сложных системах.

Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации в рамках гранта № 075-15-2022-1121.

Глава 1

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ НА ОСНОВЕ ПРИМЕНЕНИЯ МЕТОДА МОНТЕ-КАРЛО

1.1. ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ ПРИМЕНЕНИЯ МЕТОДА МОНТЕ-КАРЛО

Цель имитационного моделирования состоит в воспроизведении поведения исследуемой системы на основе результатов анализа наиболее существенных взаимосвязей между ее элементами или разработке симулятора исследуемой предметной области для проведения различных экспериментов.

Применение методов имитационного моделирования в исследовании сложных систем обусловлено, во-первых, отсутствием риска негативных последствий, во-вторых, необходимо для предупреждения подобного рода последствий, проведения в дальнейшем оценочных процедур в отношении альтернативных вариантов управленческих решений и направлений развития сложных систем. На основе имитационных процедур предоставляется возможность предугадать параметры системы, формируемые под воздействием факторов, их определяющих, и заранее предпринять предупредительные меры [1].

Метод Монте-Карло, один из методов имитационного моделирования, обладает определенными преимуществами его использования для исследования сложных систем (производственных и социально-экономических). Метод применяется с 1949 г. (дата опубликования статьи Метрополиса и С. Улама «Метод Монте-Карло» [2]). Исследование возможностей использования данного метода дает основание определить спо-

события устранения недостатков для повышения качества его применения.

Воздействие на изучаемый объект последовательности случайных событий в методе Монте-Карло заменяет комплекс сложных математических вычислений. Изначально метод Монте-Карло использовался применительно поиска решений уравнения диффузии, которые участвовали при разработке технологии обогащения урана. Широкое распространение данный метод получил с появлением быстродействующих вычислительных машин, так как его специфика заключается в построении серии вычислений с использованием в модели случайных чисел для получения результата [1]. Метод Монте-Карло — это численный метод, основанный на воспроизведении большого числа выполнения стохастического (случайного) процесса, специально созданного для условий решаемой задачи. Формируется случайный процесс таким образом, чтобы его вероятностные характеристики (например, математическое ожидание, вероятность события) соответствовали по количеству наблюдаемым характеристикам, либо посредством них предоставлялась возможность вычислить искомые величины исследуемой задачи [3].

Многие системы слишком сложны для исследования влияния неопределенности с использованием аналитических методов. Однако такие системы можно исследовать, если рассматривать входные данные в виде случайных переменных, повторяя большое количество вычислений N (итераций), для получения результата с необходимой точностью.

Модели систем могут быть разработаны с использованием таблиц и других традиционных методов. Однако существуют и более современные программные средства, удовлетворяющие высоким требованиям, многие из которых относительно недороги. Если модель разрабатывают и применяют впервые, то необходимое для метода Монте-Карло количество итераций может сделать получение результатов очень медленным и трудоемким. Современные достижения компью-

терной техники и разработка процедур генерации данных позволяют сделать продолжительность обработки незначительной во многих случаях.

Метод статистических испытаний в общем случае включает следующую последовательность этапов [4, 5].

1. Определение взаимосвязи между входными и выходными параметрами в формате математической модели. На данном этапе вводятся параметры модели, которые будут определены как случайные величины.

2. Определение для ключевых параметров модели законов распределения вероятностей (принять во внимание, что некоторые параметры модели коррелируют друг с другом). Законы распределения задаются экспертами (в основу могут быть заложены имеющиеся статистические данные или содержательное описание рассматриваемого процесса).

3. Реализация компьютерной имитации значений ключевых параметров модели. На этом этапе используются специальное программное обеспечение. Расчеты по методу Монте-Карло могут быть реализованы в собственном программном средстве, специализированных системах имитационного моделирования (Stella, AnyLogic и т.п.), а также с использованием встроенных функций и библиотек стандартных пакетов прикладных программ (MS Excel, MathCAD, MathLab, RStudio и др.)

4. Математическое вычисление основных характеристик распределений входных и выходных параметров. Верификация гипотезы на наличие характера распределения показателей (в том числе входных параметров, для верификации результатов третьего этапа).

5. Анализ полученных результатов и принятие управленческого решения. Построение оценок плотности распределения, а также функций распределения результативного индикатора. Наиболее часто в практической реализации метода используют следующие характеристики для выходных пара-

метров: коэффициент вариации; коэффициент асимметрии, эксцесса; математическое ожидание; стандартное отклонение и др.

Полученные результаты моделирования увеличением количества испытаний становятся статистически устойчивыми и могут приниматься в качестве оценок искомых показателей.

Наличие определенных преимуществ метода Монте-Карло позволяет его применять для широкого круга задач исследования сложных систем (производственных и социально-экономических). Как правило, к достоинствам метода относят следующие [1]:

1) адаптивность. Возможность использования метода с любым распределением входных данных (в том числе на основе использования эмпирического распределения), построенным путем наблюдения за сложной системой;

2) простота и наглядность моделей. Спроектированные модели учитывают только важные компоненты, состав полученных моделей может быть дополнен;

3) возможность применения моделей исследования многокомпонентных систем (в качестве примера можно привести сети Петри как более удобный аппарат моделирования процессов, протекающих независимо друг от друга, то есть параллельных процессов [6]);

4) учет различного рода внешних воздействий. Метод позволяет учитывать различного рода взаимосвязи, в том числе условные зависимости;

5) требуемая точность в получении результатов. Учитывая, что невозможно полное совпадение выходных данных модели и результатов функционирования системы под воздействием совокупности факторов, изначально не заложенных в модели, нельзя ожидать полного совпадения. Однако можно добиться высокой точности результатов для оценивания и прогнозирования параметров функционирования системы;

6) доступность программного обеспечения для реализации методического инструментария. Использовать метод

на практике можно с помощью относительно недорогих программных продуктов. Широкий круг задач для имитационного моделирования может быть реализован в собственном программном средстве, специализированных системах имитационного моделирования (Stella, AnyLogic и т.п.), а также с использованием встроенных функций и библиотек стандартных пакетов прикладных программ (MS Excel, MathCAD, MathLab, RStudio и др.).

Несмотря на указанные достоинства использования метода Монте-Карло для поддержки принятия управленческих решений имеются определенные ограничения для его применения. Выделяют следующие недостатки [1]:

1) точность результатов расчетов зависит от количества выполняемых итераций. С другой стороны, данное ограничение нивелируется специальными программными средствами;

2) неопределенность данных в системе расчетов не во всех случаях может быть нейтрализована введением установленных распределений;

3) усложнение моделей для повышения достоверности запланированного результата и, как следствие, последующие затруднения в использовании моделей. Как правило, данный недостаток нивелируется принятием в расчет влияния значимых факторов;

4) ограничения в учете всей совокупности факторов влияния на функционирование системы, как следствие, неточность и (или) ошибочность результатов расчетных процедур. Устранение недостатка возможно путем привлечения квалифицированных экспертов к верифицированному отбору значимых факторов;

5) наличие большого количества внутренних связей перегружает и усложняет модель и, как следствие, приводит к неустойчивому решению в виде выходного параметра. Ограничение может быть нивелировано за счет устранения малозначимых связей или установления уровня их приоритетности.

Указанные выше достоинства метода Монте-Карло, несмотря на наличие определенных ограничений, предопределили широту его использования при проведении имитационного моделирования сложных систем. Применение метода в решении задач, результат которых зависит от случайных процессов, дает основание к его использованию для составления различного рода социально-экономических прогнозов [7, 8], инвестиционного анализа [9, 10], финансового планирования, в риск-менеджменте [11], при оценке затрат [12].

1.2. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ МЕТОДА МОНТЕ-КАРЛО ДЛЯ РЕШЕНИЯ БИЗНЕС-ЗАДАЧИ

Метод Монте-Карло для решения бизнес-задачи целесообразно применять в интеграции с методами корреляционно-регрессионного анализа исходя из установленных целей.

Для реализации эксперимента регламентируется следующая постановка задачи: определить влияние факторов внешней среды на уровень экономической безопасности хозяйствующего субъекта (на примере предприятия химической промышленности).

На первом этапе оценки необходимо определить совокупность факторов влияния на объект исследования. Выбор факторов осуществляется на основе опыта сотрудников Компании и данных, полученных в ходе качественного анализа внешней среды, методом экспертной оценки. Первичный набор показателей (определяющих факторы внешнего воздействия) с учетом специфики деятельности промышленного предприятия представлен в табл. 1.1.

Следующим этапом после первичного выбора показателей для мониторинга угроз и рисков внешней среды предприятия является их проверка на наличие статистической взаимосвязи с интегральным показателем экономической безопасности предприятия с помощью корреляционного анализа.

Для определения тесноты связи между уровнем экономической безопасности и внешними факторами использован критерий корреляции Пирсона (метод параметрической статистики, позволяющий определить наличие или отсутствие линейной связи между двумя количественными показателями, а также оценить ее тесноту и статистическую значимость). Другими словами, критерий корреляции Пирсона позволяет определить, изменяется ли (возрастает или уменьшается) один показатель в ответ на изменения другого [13]. В статистических расчетах и выводах коэффициент корреляции обычно обозначается как r_{xy} или R_{xy} . Расчет данного критерия производится в соответствии с формулой

$$r_{XY} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}}, \quad (1.1)$$

где X — значение внешнего фактора влияния; \bar{X} — среднее значение внешнего фактора влияния в рассматриваемом периоде; Y — значение интегрального показателя экономической безопасности предприятия; \bar{Y} — среднее значение интегрального показателя экономической безопасности предприятия в рассматриваемом периоде.

По результатам расчетов корреляции отбираются факторы, связь которых с параметром «Уровень экономической безопасности хозяйствующего субъекта» подтвердится математически. Результаты расчетов критерия корреляции Пирсона представлены в табл. 1.2.

На основе произведенных вычислений можно сделать вывод о том, что высокая степень статистической связи присутствует только между показателем интегрального уровня экономической безопасности и факторами $X1$ (Цена апатитового концентрата, долл./т) и $X2$ (Средняя цена удобрений фосфорных, руб./т), коэффициент корреляции между ними составляет 0,83 и 0,98 пунктов соответственно.

Таблица 1.1

Первичный набор показателей для мониторинга угроз и рисков внешней среды предприятия с учетом специфики деятельности

Год	Интервальная оценка уровня экологической безопасности (Y)	Цена апатитового концентрата, долл./т (X1)	Средняя цена удобрений фосфорных, руб./т (X2)	Экспорт минеральных удобрений РФ, тыс. т (X3)	Цена DAP, долл./т (X4)	Средняя цена удобрений азотных, руб./т (X5)	Dutch TTF Natural Gas, euro (X6)
2017	0,24	80,00	12 336,00	16 561,20	355,50	10 329,00	19,32
2018	0,37	88,00	13 000,00	16 781,70	419,90	10 913,00	22,28
2019	0,54	89,00	14 086,00	16 517,50	348,30	9 937,00	12,44
2020	0,42	80,00	13 322,00	16 588,40	323,80	10 975,00	19,45
2021	0,63	101,00	15 125,00	18 236,91	569,90	14 350,00	80,43

Примечание: (X1) Цена апатитового концентрата на международном рынке (FOD Касабланка) включена в список первичных показателей, поскольку данный продукт является основным в товарной линейке предприятия химической промышленности (10675,5 тыс. тонн в 2021 году). Также значительная доля произведенной продукции экспортируется Компанией за границу. (X2) Средняя цена фосфорных удобрений на российском рынке включена в список показателей. Основание для включения: концентрат является основным сырьем для производства фосфорных удобрений. (X3) Экспорт минеральных удобрений производителями Российской Федерации является индикатором, который отражает как состояние отрасли минеральных удобрений в России, так и положение государства на международном рынке (поскольку наличие санкций, ограничений и т.д. находит свое отражение в данном показателе). (X4) Цена DAP (диаммоний фосфат) является одним из самых распространенных и востребованных видов фосфорных удобрений. Информация о цене ди-

1.2. Практическая реализация метода Монте-Карло для решения бизнес-задачи

аммофоса (FOB Мексиканский залив) отражает объемы спроса на удобрения на международном рынке. (X5) Среднегодовая цена на внутреннем рынке является показателем спроса среди российских аграриев. (X6) Цена фьючерсов на природный газ на нидерландском «хабе» TTF. Данный транспортный итоговый узел является ключевым для формирования цен на основное сырье для производства азотных удобрений.

Таблица 1.2

Рассчитанные коэффициенты корреляции интегральной оценки ЭБП и факторов внешнего влияния

Показатель	Y	X1	X2	X3	X4	X5	X6
Y		0,83	0,98	0,67	0,60	0,63	0,63
X1			-0,44	-0,72	0,79	-0,47	0,84
X2				0,84	-0,14	0,83	0,75
X3					-0,45	0,83	1,00
X4						-0,03	0,95
X5							0,99
X6							

Поскольку отобранные по результатам расчетов показатели (определяющие факторы влияния) будут использованы в процессе реализации регрессионного анализа (для дальнейшего получения регрессионного уравнения), в связи с этим необходимо исключить из модели факторы, дублирующие значения друг друга. Наличие корреляционной связи на уровне (-0,44) дает основание для включения X1 и X2 в последующие расчетные процедуры.

Регрессионный анализ представляет набор статистических методов исследования влияния одной или нескольких независимых переменных X1, X2, ..., Xn на зависимую переменную Y. Независимые переменные иначе называют регрессорами или предикторами, а зависимые переменные — критериальными или регрессантами [14]. Терминология зависимых и независимых переменных отражает лишь математическую зависимость переменных, а не причинно-следственные от-

ношения. Наиболее распространенный вид регрессионного анализа — линейная регрессия, когда находят линейную функцию, которая, согласно определенным математическим критериям, наиболее соответствует данным. Например, в методе наименьших квадратов вычисляется прямая (или гиперплоскость), сумма квадратов между которой и данными минимальна.

Для расчета коэффициентов из уравнения регрессии будет использован метод наименьших квадратов, математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от искомым переменных [15]. Он может использоваться для «решения» переопределенных систем уравнений (когда количество уравнений превышает количество неизвестных), для поиска решения в случае обычных (не переопределенных) нелинейных систем уравнений, для аппроксимации точечных значений некоторой функции. Метод наименьших квадратов является одним из базовых методов регрессионного анализа для оценки неизвестных параметров регрессионных моделей по выборочным данным. Результаты регрессионного анализа представлены в табл. 1.3.

Таблица 1.3

Результаты регрессионного анализа

Показатель	Значение
R-квадрат	0,974816568
Стандартная ошибка	0,034209112
Y-пересечение	-1,484449436
Переменная X1	-0,003001947
Переменная X2	0,000161282

На основе полученных данных формируется уравнение регрессии, отражающее влияние значений факторов внешней

среды на уровень экономической безопасности предприятия химической промышленности в соответствии с формулой

$$Y = -0,003001947 \cdot X_1 + 0,000161282 \cdot X_2 - 1,484449436, (1.2)$$

где Y – значение интегрального показателя экономической безопасности предприятия; X_1 – цена апатитового концентрата, долл./т; X_2 – средняя цена удобрений фосфорных, руб./т.

Также можно сделать вывод о высокой точности полученной модели, поскольку значение R-квадрата близко к единице, а стандартная ошибка составляет лишь 0,034 (3,4%).

На основе полученного уравнения становится возможным получать значения Y (значения интегрального показателя оценки экономической безопасности) при заданных значениях X_1 и X_2 . Для определения точности полученной математической модели произведем сравнение фактических значений интегрального показателя экономической безопасности предприятия с расчетными значениями Y , полученными с помощью многофакторного уравнения регрессии. Данные представлены в табл. 1.4.

Таблица 1.4

**Сравнение фактических и расчетных значений Y
(интегрального показателя экономической безопасности
предприятия), 2017–2021 гг.**

Год	X_1	X_2	Y	Y расчетный	$(Y - Y_{расчет})^2$
2017	80,00	12 336,00	0,24	0,264971242	0,00077
2018	88,00	13 000,00	0,37	0,348047005	0,00073
2019	89,00	14 086,00	0,54	0,520197458	0,00046
2020	80,00	13 322,00	0,42	0,423995428	0,00000
2021	101,00	15 125,00	0,63	0,651746234	0,00038

Квадрат отклонения фактических значений от расчетного находится в допустимых пределах. Однако для проверки аде-

кватности полученной математической модели необходимо использовать F -тест или критерий Фишера (статистический критерий, тестовая статистика которого при выполнении нулевой гипотезы имеет распределение Фишера (F -распределение)).

Статистика теста так или иначе сводится к отношению выборочных дисперсий (сумм квадратов, деленных на «степени свободы»). Чтобы статистика имела распределение Фишера, необходимо чтобы числитель и знаменатель были независимыми случайными величинами, при этом соответствующие суммы квадратов имели распределение Хи-квадрат. Для этого требуется, чтобы данные имели нормальное распределение. Кроме того, предполагается, что дисперсия случайных величин, квадраты которых суммируются, одинакова. Расчет критерия Фишера $F_{\text{расч}}$ осуществляется в соответствии с формулой

$$F_{\text{расч}} = \frac{\sum(Y_{i\text{расч}} - \bar{Y}_{\text{расч}})^2}{m} \times \frac{n-m-1}{\sum(Y_i - Y_{i\text{расч}})^2}, \quad (1.3)$$

где $Y_{i\text{расч}}$ — значение интегрального показателя экономической безопасности предприятия, полученное на основе уравнения регрессии; $\bar{Y}_{\text{расч}}$ — среднее значение интегрального показателя экономической безопасности предприятия за рассматриваемый период; Y_i — фактическое значение интегрального показателя экономической безопасности предприятия в отчетном периоде; n — количество исходных наблюдений; m — количество факторов X , влияющих на результирующий показатель.

В результате проведенных математических вычислений получены следующие значения: расчетное значение критерия Фишера — (38,71), тогда как табличное значение — (19,0). Сравнение указанных значений дает основание сделать вывод об адекватности построенной модели.

На следующем этапе оценочной процедуры воспользуемся методом имитационного моделирования Монте-Карло. В данном случае цель имитационного моделирования состоит в воспроизведении поведения исследуемой системы на основе результатов анализа наиболее существенных взаимосвязей между ее элементами. Точность метода Монте-Карло напрямую зависит от числа итераций. Для получения высокой точности было принято решение об использовании ста тысяч испытаний в представленной выше математической модели.

Решение поставленной задачи требует использования электронно-вычислительной техники, а также специального программного обеспечения для него (функционал MS Excel или AnyLogic).

В рамках данного эксперимента преимущество метода Монте-Карло заключается в возможности вероятностного прогнозирования значения уровня экономической безопасности предприятия при отсутствии точных прогнозных значений факторов X_1 и X_2 .

С помощью данного метода появляется возможность задать примерные промежутки значений, оценив вероятность «выпадения» определенного значения интегрального показателя экономической безопасности предприятия. На основе полученной информации появляется возможность определять такие значения внешних факторов, при которых значение Y будет находиться в приемлемых рамках.

На рис. 1.1 изображено распределение значений Y при X_1 (цена апатитового концентрата, долл./т), равном 90–100 долл., и X_2 (средняя цена удобрений фосфорных, руб./т), равном 13 000–14 000 руб.

При данных значениях X_1 и X_2 среднее арифметическое значение Y (интегральный показатель экономической безопасности предприятия) составляет (0,41). В связи с этим можно сделать вывод о том, что падение цен до указанного уровня негативно отразится на общей устойчивости предприятия.

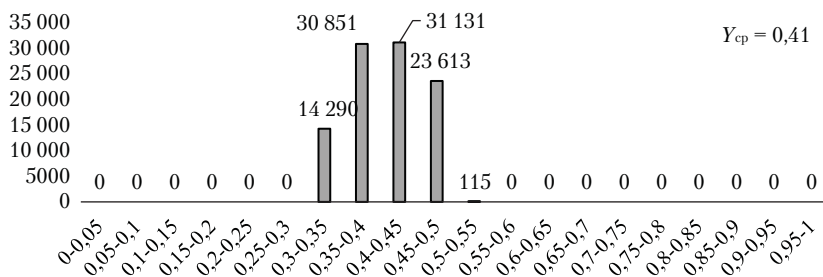


Рис. 1.1. Распределение значений Y при X_1 , равном 90–100 долл./т, и X_2 , равном 13 000–14 000 руб./т

На рис. 1.2 изображено распределение значений Y при X_1 (цена апатитового концентрата, долл./т), равном 80–85 долл., и X_2 (средняя цена удобрений фосфорных, руб./т), равном 11 000–12 000 руб.

При данных значениях X_1 и X_2 среднее арифметическое значение Y (интегральный показатель экономической безопасности предприятия) составляет (0,12). В связи с этим можно сделать вывод о том, что падение цен до указанного уровня может привести к полной потере общей устойчивости предприятия.

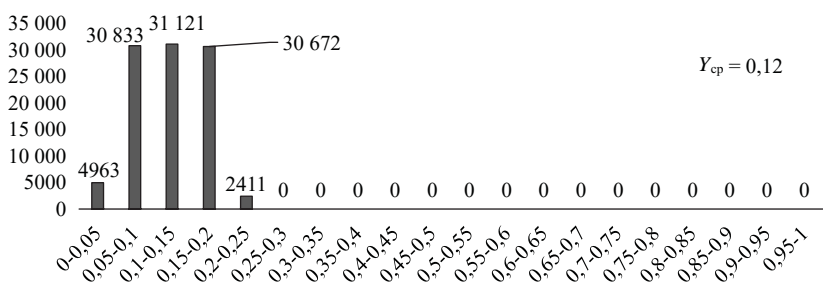


Рис. 1.2. Распределение значений Y при X_1 , равном 80–85 долл./т, и X_2 , равном 11 000–12 000 руб./т

На рис. 1.3 изображено распределение значений Y при X_1 (цена апатитового концентрата, долл./т), равном 100–

110 долл., и X_2 (средняя цена удобрений фосфорных, руб./т), равном 16 000–17 000 руб.

При данных значениях X_1 и X_2 среднее арифметическое значение Y (интегральный показатель экономической безопасности предприятия) составляет (0,86). Можно сделать вывод о том, что рост цен до указанного уровня благоприятно скажется на уровне экономической безопасности предприятия.

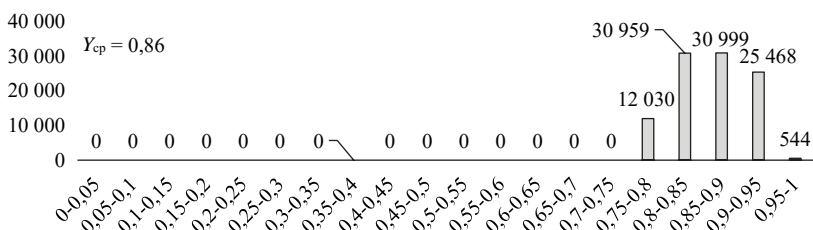


Рис. 1.3. Распределение значений Y при X_1 , равном 100–110 долл./т, и X_2 , равном 16 000–17 000 руб./т

Представленный способ анализа универсален, поскольку может применяться в рамках указанного эксперимента с данными показателями (идентифицирован риск изменения цены фосфоритового концентрата на мировом рынке и риск изменения цены фосфорных удобрений на внутреннем рынке), а также применим для анализа влияния различного рода факторов на уровень экономической безопасности хозяйствующего субъекта.

1.3. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ МЕТОДА МОНТЕ-КАРЛО ДЛЯ ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМАХ

Метод Монте-Карло может быть встроен в процесс оценки продовольственной безопасности для обеспечения поддержки принятия управленческих решений. Ориентируясь на установленную последовательность оценочной

процедуры экономической безопасности социально-экономических систем [8], включение метода может быть реализовано на этапе верификации частных индикаторов с позиции их влияния на интегральный уровень продовольственной безопасности. Возможной альтернативой для метода Монте-Карло в данном случае может быть метод построения модели векторной авторегрессии [16], метод параметрической коррекции обычными наименьшими квадратами [17, 18], метод нечеткой логики (теории нечетких множеств) [19, 20, 21].

На основе метода Монте-Карло осуществляется оценка будущих параметров с учетом предыдущих тенденций [22, 23, 24]. Так, в аспекте оценки продовольственной безопасности представляется возможным рассматривать динамику изменений индикаторов оценки в будущем как случайный процесс. В результате выполнения алгоритмов для каждого индикатора формируется набор значений по числу моделируемых реализаций в методе Монте-Карло, на основании которых осуществляется моделирование процентильных значений индикаторов в будущем [25].

Для реализации эксперимента регламентируется следующая постановка задачи: определить степень влияния факторов инфраструктурного развития на продовольственную безопасность территорий (на примере субъектов Арктической зоны Российской Федерации – далее АЗРФ).

Оценку влияния инфраструктурных факторов предполагается реализовать через комбинацию корреляционно-регрессионного инструментария с последующей проверкой результатов с использованием *метода Монте-Карло*. Учитывая специфику реализуемого исследования, методы корреляционно-регрессионного анализа на этапе первичной оценки влияния переменных позволяют осуществить процедуру отсеивания факторов слабого влияния для уравнения регрессии и последующего моделирования, поскольку наибольшую точность процедура моделирования на основании

метода Монте-Карло достигает при усеченном перечне влияющих переменных.

Для оценки влияния инфраструктурных факторов на продовольственную безопасность регионов Арктической зоны Российской Федерации выбран следующий перечень показателей, представленных в табл. 1.5.

Таблица 1.5

Показатели, учитываемые в контексте оценки влияния инфраструктурных факторов на продовольственную безопасность регионов АЗРФ

№ п/п	Наименование показателя	Условное обозначение
1	Фондонасыщенность территории с учетом степени износа основных фондов, тыс. руб. на км ²	X1
2	Величина ОФ в расчете на 1 жителя, тыс. руб. на 1 чел.	X2
3	Количество введенных сельскохозяйственных и промышленных зданий, ед.	X3
4	Коэффициент Гольца	X4
5	Доля автомобильных дорог, отвечающих нормативным требованиям, %	X5
6	Удельный вес сельских населенных пунктов, имеющих связь по дорогам с твердым покрытием с сетью дорог общего пользования в общем числе сельских населенных пунктов, %	X6
7	Удельный вес площади жилищного фонда, оборудованной водоотведением (канализацией), в общей площади всего жилищного фонда, %	X7
8	Доля затрат на внедрение и использование информационных и коммуникационных технологий организаций сельского хозяйства в общей совокупности затрат организаций на информационные и коммуникационные технологии, %	X8

Окончание табл. 1.5

№ п/п	Наименование показателя	Условное обозначение
9	Доля организаций, использовавших специальные программные средства, %	X9
10	Доля организаций, использовавших сеть Интернет для связи с поставщиками товаров (работ, услуг), %	X10
11	Доля организаций, использовавших сеть Интернет для связи с потребителями товаров (работ, услуг), %	X11

В результате проведенного анализа корреляционной зависимости факторов инфраструктурного влияния на интегральный показатель продовольственной безопасности можно сгруппировать регионы (объекты оценки) по частоте возникновения силы связи с целью определения наиболее важных факторов в разрезе макрорегиона Арктической зоны Российской Федерации (рис. 1.4).

В результате анализа частоты возникновения силы связи инфраструктурных факторов с интегральным показателем уровня продовольственной безопасности регионов АЗРФ, можно определить, что наибольшее влияние оказывает показатель фондонасыщенности территории с учетом степени износа основных фондов (частота возникновения связи — б) со средней и сильной силой прямой и обратной направленности.

В арктических регионах слабо развита производственная инфраструктура, существующие мощности имеют высокую степень износа и требуют модернизации, удельный вес полностью изношенных фондов в среднем по АЗРФ составляет 18%. Одной из традиционных отраслей регионов является рыболовство, однако на некоторых территориях прослеживается недостаточность приемно-транспортных судов, укомплектованных рефрижераторными трюмами, также в непосред-

Показатель	Число вхождений в корреляционную группу									
	(-1) – (-0,9)	(-0,9) – (-0,7)	(-0,7) – (-0,5)	(-0,5) – (-0,3)	(-0,3) – 0,0	0,0 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 1
Фондосыщенность территории с учетом степени износа основных фондов, тыс. руб. на км ²	0	1	4	0	1	2	0	1	0	0
Величина ОФ в расчете на 1 жителя, тыс. руб. на 1 чел.	0	1	4	0	1	2	0	1	0	0
Количество введенных сельскохозяйственных и промышленных зданий, ед.	0	0	0	1	4	1	1	2	0	0
Коэффициент Гольца	0	1	2	1	1	2	1	1	0	0
Доля автомобильных дорог, отвечающих нормативным требованиям, %	0	1	3	2	1	1	0	1	0	0
Удельный вес сельских населенных пунктов, имеющих связь по дорогам с твердым покрытием с сетью дорог общего пользования в общем числе сельских населенных пунктов, %	0	1	1	1	1	3	1	0	1	0
Удельный вес площади жилищного фонда, оборудованной водотведением (канализацией), в общей площади всего жилищного фонда %	0	2	2	2	3	0	0	0	0	0
Доля затрат на внедрение и использование информационных и коммуникационных технологий организаций сельского хозяйства в общей совокупности затрат организаций на информационные и коммуникационные технологии, %	0	1	0	3	2	1	1	0	1	0
Доля организаций, использовавших специальные программные средства, %	0	0	1	1	1	2	1	1	0	0
Доля организаций, использовавших сеть Интернет для связи с поставщиками товаров (работ, услуг). %	0	1	0	1	2	3	0	0	0	0
Доля организаций, использовавших сеть Интернет для связи с потребителями товаров (работ, услуг). %	0	0	0	1	2	3	0	0	0	0



Рис. 1.4. Диапазон корреляционной связи инфраструктурных факторов с интегральным показателем уровня продовольственной безопасности регионов АЗРФ

ственной близости к участкам массового вылова отсутствуют пункты приемки и сохранения уловов, места комплексной переработки, что оказывает негативное влияние на продовольственную безопасность и препятствует развитию собственного производства.

Обратная направленность силы связи обуславливается временным фактором, поскольку для создания прочной материально-технической базы необходимы крупные капитальные вложения, которые не могут принести положительный эффект за короткий временной промежуток. Для возведения новых объектов основных фондов в условиях суровых природно-климатических условий и вечной мерзлоты денежных средств требуется гораздо больше, нежели в регионах с умеренным климатом. Обслуживание данных мощностей также требует больших финансовых затрат. Именно поэтому развитие производственной инфраструктуры в АЗРФ с целью сохранения и развития сельского и промышленного хозяйства должно быть задачей государственного уровня.

Также корреляционная зависимость сильной и средней силы связи наблюдалась с показателями транспортной инфраструктуры, а именно: доля автомобильных дорог, отвечающих нормативным требованиям (частота возникновения связи — 5), коэффициент Гольца (частота возникновения связи — 4) и удельный вес сельских населенных пунктов, имеющих связь по дорогам с твердым покрытием с сетью дорог общего пользования (частота возникновения связи — 3).

Регионы Арктики являются зависимыми от ввоза продовольствия ввиду невозможности собственного производства в необходимом объеме продуктов питания в связи с суровыми природно-климатическими условиями, оставляя вопрос развития транспортной инфраструктуры крайне актуальным на протяжении всего периода исследования. По уровню развития транспортной инфраструктуры российская Арктика представлена двумя неравнозначными территориями:

1) в западных регионах (за исключением Ненецкого автономного округа) сформирована достаточно разветвленная сеть автомобильных и железных дорог, круглогодично связанная с общероссийской автодорожной и железнодорожной сетями. В частности, через территории Мурманской и Архангельской областей, Карелии и Ямала проходят участки Октябрьской и Северной железных дорог, а также автодороги федерального значения;

2) в арктических районах Сибири и Дальнего Востока наземные транспортные коммуникации практически отсутствуют, а сообщение с другими регионами страны обеспечивается воздушным и водным транспортом, что определяет безальтернативность схем транспортного обслуживания. В двух арктических регионах — Ненецком и Чукотском автономных округах — железные дороги и крупные магистрали отсутствуют совсем. Таким образом, во-первых, существует проблема отсутствия транспортных путей в некоторых регионах, а во-вторых, в настоящий момент около 50% существующих в АЗРФ автомобильных дорог общего пользования не отвечают нормативным требованиям технико-эксплуатационного состояния.

Такое положение обусловлено экстремальностью климата, уязвимостью природной среды, масштабами и удаленностью арктических территорий. В результате корреляционного анализа отмечено, что в большинстве случаев наблюдается обратная сила связи данных факторов с интегральным показателем продовольственной безопасности. Влияние всех перечисленных условий определяет сезонный характер использования отдельных видов транспорта, удорожание строительства и эксплуатации транспортной инфраструктуры, увеличение производственно-транспортных издержек. Следовательно, для развития транспортной инфраструктуры макрорегиону требуются значительные финансовые вложения и долгосрочный период, поскольку данная проблема является актуальной, так как неразвитость транспортной системы

усложняет задачи обеспечения арктических территорий жизненно важным продовольствием и снижает конкурентоспособность местных производителей товаров и услуг.

Таким образом, можно сделать вывод, что наибольшее влияние на продовольственную безопасность оказывают факторы производственной и транспортной инфраструктуры. Благодаря развитию в регионах АЗРФ производственных мощностей, улучшению состояния транспортной системы возможно развитие собственного агропромышленного комплекса, условий бесперебойных поставок жизненно важного продовольствия, что в конечном результате позволит повысить уровень продовольственной безопасности Арктической зоны Российской Федерации.

По результатам оценки корреляционной связи инфраструктурных факторов с показателями продовольственной безопасности регионов Арктической зоны Российской Федерации для процедуры моделирования на основе использования метода Монте-Карло отобраны следующие факторы, характеризующие высокую степень взаимозависимости (табл. 1.6). В основу моделирования закладываются уравнения регрессии, сформированные на основании значимых факторных переменных.

Таблица 1.6

Инфраструктурные факторы, включенные в формирование математической модели оценки влияния инфраструктурных факторов на продовольственную безопасность Арктической зоны Российской Федерации

Регион АЗРФ	Факторы инфраструктурного влияния, включенные в моделирование
Мурманская область	X1 — Фондонасыщенность территории с учетом степени износа основных фондов, тыс. руб. на км ² ; X4 — Коэффициент Гольца; X6 — Удельный вес сельских населенных пунктов, имеющих связь по дорогам с твердым покрытием с сетью дорог общего пользования в общем числе сельских населенных пунктов, %

Продолжение табл. 1.6

Регион АЗРФ	Факторы инфраструктурного влияния, включенные в моделирование
Ненецкий авто- номный округ	X1 – Фондонасыщенность территории с учетом степени износа основных фондов, тыс. руб. на км ² ; X5 – Доля автомобильных дорог, отвечающих нормативным требованиям, %; X9 – Доля организаций, использовавших специальные программные средства, %
Чукотский авто- номный округ	X1 – Фондонасыщенность территории с учетом степени износа основных фондов, тыс. руб. на км ² ; X4 – Коэффициент Гольца; X7 – Удельный вес площади жилищного фонда, оборудованной водоотведением (канализацией), в общей площади всего жилищного фонда, %
Ямало- Ненецкий авто- номный округ	X3 – Количество введенных сельскохозяйственных и промышленных зданий, ед.; X8 – Доля затрат на внедрение и использование информационных и коммуникационных технологий организаций сельского хозяйства в общей совокупности затрат организаций на информационные и коммуникационные технологии, %; X10 – Доля организаций, использовавших сеть Интернет для связи с поставщиками товаров (работ, услуг), %
Республика Ка- релия	X1 – Фондонасыщенность территории с учетом степени износа основных фондов, тыс. руб. на км ² ; X5 – Доля автомобильных дорог, отвечающих нормативным требованиям, %
Республика Коми	X1 – Фондонасыщенность территории с учетом степени износа основных фондов, тыс. руб. на км ² ; X6 – Удельный вес сельских населенных пунктов, имеющих связь по дорогам с твердым покрытием с сетью дорог общего пользования в общем числе сельских населенных пунктов, %; X8 – Доля затрат на внедрение и использование информационных и коммуникационных технологий

Окончание табл. 1.6

Регион АЗРФ	Факторы инфраструктурного влияния, включенные в моделирование
	организаций сельского хозяйства в общей совокупности затрат организаций на информационные и коммуникационные технологии, %
Республика Саха (Якутия)	X3 – Количество введенных сельскохозяйственных и промышленных зданий, ед.; X4 – Коэффициент Гольца
Красноярский край	X9 – Доля организаций, использовавших специальные программные средства, %
Архангельская область	X1 – Фондонасыщенность территории с учетом степени износа основных фондов, тыс. руб. на км ² ; X4 – Коэффициент Гольца; X5 – Доля автомобильных дорог, отвечающих нормативным требованиям, %

На основании использования метода регрессионного анализа сформированы уравнения регрессии, описывающие влияние факторных признаков на результирующий – интегральный уровень продовольственной безопасности регионов и макрорегиона АЗРФ в целом (табл. 1.7).

Таблица 1.7

Математические модели оценки влияния инфраструктурных факторов на продовольственную безопасность регионов АЗРФ

Регион АЗРФ	R^2	Уравнение регрессии
Мурманская область	0,78	$Y = -0,19 \cdot X1 + 1,2 \cdot X4 + 2,8 \cdot X6 - 4,52,$ где P -значения для модели: X1 0,11 > 0,05 X2 0,03 < 0,05 X6 0,01 < 0,05


Продолжение табл. 1.7

Регион АЗРФ	R^2	Уравнение регрессии
Ненецкий автономный округ	0,68	$Y = -0,14 \cdot X_1 - 0,01 \cdot X_5 + 0,1 \cdot X_9 + 0,68$, где P -значения для модели: X_1 0,17 > 0,05 X_5 0,65 < 0,05 X_9 0,03 < 0,05
Чукотский автономный округ	0,9	$Y = 1,89 \cdot X_1 - 5,7 \cdot X_4 - 2,18 \cdot X_7 + 5,06$, где P -значения для модели: X_1 0,04 < 0,05 X_4 0,006 < 0,05 X_7 0,07 < 0,05
Ямало-Ненецкий автономный округ	0,82	$Y = 0,016 \cdot X_3 - 0,01 \cdot X_8 - 0,16 \cdot X_{10} + 1,01$, где P -значения для модели: X_3 0,017 < 0,05 X_8 0,02 < 0,05 X_{10} 0,42 > 0,05
Республика Карелия	0,61	$Y = -0,59 \cdot X_1 - 1,09 \cdot X_5 + 2,14$, где P -значения для модели: X_1 0,6 > 0,05 X_5 0,6 > 0,05
Республика Коми	0,62	$Y = -0,05 \cdot X_1 - 4,4 \cdot X_6 + 0,65 \cdot X_8 + 7,4$, где P -значения для модели: X_1 0,94 > 0,05 X_6 0,72 > 0,05 X_8 0,03 < 0,05
Республика Саха (Якутия)	0,69	$Y = 0,18 \cdot X_3 + 0,69 \cdot X_4 + 0,49$, где P -значения для модели: X_3 0,03 < 0,05 X_4 0,02 < 0,05
Красноярский край	0,87	$Y = -1,3 \cdot X_9 + 2,74$, где P -значения для модели: X_9 0,047 < 0,05

Окончание табл. 1.7

Регион АЗРФ	R^2	Уравнение регрессии
Архангельская область	0,75	$Y = 1,79 \cdot X_1 + 1,04 \cdot X_4 - 4,89 \cdot X_5 + 1,17,$ где P -значения для модели: X_1 0,03 < 0,05 X_4 0,02 < 0,05 X_5 0,03 < 0,05
Арктическая зона Российской Федерации	0,92	$Y = -1,9 \cdot X_1 + 2,67 \cdot X_4 + 0,3 \cdot X_5 - 0,14,$ где P -значения для модели: X_1 0,017 < 0,05 X_4 0,021 < 0,05 X_5 0,023 < 0,05

Интерпретация значений:

 R^2 значения для моделиЕсли $R^2 > 0,95$ — высокая точность аппроксимации (модель хорошо описывает явление)Если R^2 от 0,6 до 0,95 — умеренная точность аппроксимации (модель хорошо описывает явление)Если $R^2 < 0,6$ — точность аппроксимации недостаточна (модель требует корректировки) P -значения для модели
 Коэффициент считается нулевым, не оказывающим влияния на Y
 Коэффициент считается ненулевым, оказывающим влияние на Y

По результатам построения уравнения регрессии по Мурманской области коэффициент детерминации составляет 0,78, что доказывает умеренную точность аппроксимации, модель имеет высокую значимость и достоверно описывает явление. Для определения важности показателей, включенных в модель, проанализированы P -значения коэффициентов. Выявлено, что фактор X_1 (фондонасыщенность территории с учетом степени износа основных фондов) не оказывает существенного влияния на модель (так как $P = 0,11$, что больше 0,05). Для остальных переменных P -значения приближены к нулю, подтверждая важность исследуемых факторов. Ана-

лизируя коэффициенты уравнения регрессии, можно сделать вывод, что наибольшее влияние оказывают показатель X_6 (удельный вес сельских населенных пунктов, имеющих связь по дорогам с твердым покрытием) и X_4 (коэффициент Гольца). Данные результаты подтверждают установленную связь переменных на основе корреляционного анализа. Существующие в регионе населенные пункты соединены лишь локальными дорогами, несоответствующими необходимым требованиям, в связи с чем физическая доступность продовольствия ряда сельских населенных пунктов находится под угрозой, следовательно, объясняется влияние на интегральный уровень продовольственной безопасности.

Коэффициент детерминации уравнения регрессии Ненецкого автономного округа ниже на 0,1, чем в Мурманской области, однако также отражает умеренную точность аппроксимации и достоверно описывает явление. Два показателя (X_1 и X_5) имеют высокие P -значения, что свидетельствует о их недостаточной значимости для модели. Таким образом, наибольшее влияние оказывает показатель X_9 (доля организаций, использовавших специальные программные средства). Результаты корреляционного анализа подтверждаются, поскольку с показателем доли организаций, использовавших специальные программные средства, уровень продовольственной безопасности имеет прямую связь. Ненецкий автономный округ на протяжении всего анализируемого периода является аутсайдером среди регионов АЗРФ по использованию организациями специальных программных средств, несмотря на развитие тенденции «цифрового сельского хозяйства» в Российской Федерации, предусматривающей внедрение новых технологий и их использование в производстве пищевой продукции.

Самое высокое значение коэффициента детерминации наблюдается в уравнении Чукотского автономного округа (0,9). Все включенные показатели оказывают существенное влияние по P -значению, следовательно, модель достоверна. Наибольшее

влияние оказывает коэффициент Гольца (X_4), как и в корреляционном анализе данный показатель имеет обратную связь. В регионе слабо развита транспортная система, основные грузоперевозки осуществляются авиационным и морским путем. В Чукотском автономном округе отсутствуют автотранспортные выходы в прилегающие субъекты. Обратная сильная связь объясняется тем, что в настоящее время, несмотря на реализуемые программы развития, улучшить существующие условия транспортной доступности невозможно. Меньший средний прирост интегрального уровня продовольственной безопасности создает фактор X_7 (удельный вес площади жилищного фонда, оборудованной водоотведением (канализацией)), так как показатели социальной инфраструктуры не несут прямого влияния на обеспечение населения продовольствием, а лишь способствуют повышению качества жизни населения и продовольственной безопасности. Показатель X_1 (фондонасыщенность территории с учетом степени износа основных фондов) также оказывает влияние на прирост уровня продовольственной безопасности. При корреляционном анализе с данным показателем выявлена обратная связь, что объясняется необходимостью крупных капитальных вложений, требующих большего количества времени для положительного эффекта. Коэффициент регрессии же принимает положительное значение, подтверждая важность модернизации и создания материально-технической базы для собственного сельскохозяйственного производства с целью получения прироста уровня продовольственной безопасности.

Коэффициент детерминации по уравнению регрессии Ямало-Ненецкого автономного округа принимает высокое значение (0,82). Недостаточную степень важности для модели имеет фактор X_{10} (доля организаций, использовавших сеть Интернет для связи с поставщиками товаров), так как R -значение превышает 0,05 и составляет 0,42. Практически равнозначное влияние оказывают факторы X_3 (количество введенных сельскохозяйственных и промышленных зданий)

и X_8 (доля затрат на внедрение и использование информационных и коммуникационных технологий организаций сельского хозяйства), только в противоположных направлениях. Увеличение количества новых зданий, предназначенных для развития сельского хозяйства, обеспечит прирост уровня продовольственной безопасности. Однако затраты на информационные технологии не позволяют за небольшой временной промежуток оказать положительный эффект.

Самое низкое значение коэффициента детерминации наблюдается в уравнении регрессии Республики Карелия и составляет 0,61. P -значения для коэффициентов, связанных с каждой независимой переменной, превышают 0,05, следовательно, включенные в уравнение факторы не являются существенно значимыми для модели, что объясняется спецификой данного региона и особенностями развития инфраструктуры.

Аналогичная ситуация отмечается в Республике Коми: коэффициент детерминации равен 0,62, данное значение находится на границе недостаточной и умеренной точности аппроксимации. Однако при анализе P -значений выявлено, что один факторов все же является значимым для модели. Положительный прирост к уровню продовольственной безопасности создает показатель X_8 (доля затрат на внедрение и использование информационных и коммуникационных технологий организаций сельского хозяйства). Данная зависимость обусловлена необходимостью перехода в современном мире на деятельность организаций с использованием новых информационных технологий, в то время как в Республике Коми недостаточно выделяется средств организациями на их внедрение. Таким образом, использование новых технологий позволит повысить уровень продовольственной безопасности региона.

Коэффициент детерминации в уравнении регрессии Республики Саха (Якутия) выше и составляет 0,69, что говорит об умеренной точности аппроксимации и достоверности модели. Оба показателя, в соответствии с P -значениями, явля-

ются значимыми. Наибольший прирост к значению уровня продовольственной безопасности оказывает фактор X_4 (коэффициент Гольца). Для возможности бесперебойных поставок продовольствия, обеспечения физической доступности продовольствия населению необходимо развитие транспортной системы, так как в настоящее время главным видом транспорта, обеспечивающим грузоперевозки в регионе, является водный, который, в свою очередь, является сезонным, в связи с чем для круглогодичного обеспечения продовольствием населения необходимо развитие альтернативных транспортных путей. Также положительный прирост для результирующего показателя обеспечивает X_3 (количество введенных сельскохозяйственных и промышленных зданий). Как уже было отмечено в ходе корреляционного анализа, в регионе существует проблема с хранением ввезенной продукции, следовательно, увеличение складских помещений и иных сельскохозяйственных зданий позволит повысить уровень продовольственной безопасности.

В связи с высоким уровнем продовольственной безопасности и достаточным уровнем развития инфраструктуры и, как результат, слабым влиянием большинства факторов в ходе корреляционного анализа в Красноярском крае при построении уравнения регрессии был включен только один показатель X_9 — доля организаций, использовавших специальные программные средства. Коэффициент детерминации в данном уравнении самый высокий среди всех регионов АЗРФ и составляет 0,9. В настоящее время развитие агропромышленного производства необходимо осуществлять на основе применения новых технологий как непосредственно в процессе производства, так и в процедурах мониторинга с последующим прогнозированием на базе специальных программных средств. Однако отрицательный прирост, который создает показатель, обусловлен необходимостью крупных финансовых вложений в развитие цифровых технологий, что за короткий срок не сможет обеспечить значительный положительный эф-

фekt. Реализация программ по внедрению цифровых технологий активно реализуется только с 2018 года, следовательно, для положительного прироста требуется большее количество времени.

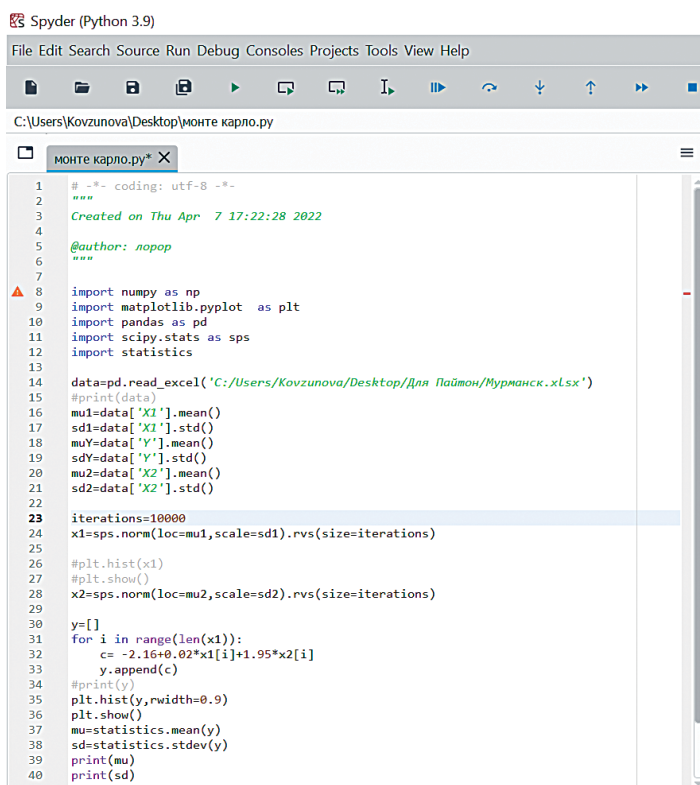
Коэффициент детерминации в уравнении регрессии Архангельской области составляет 0,75, что подтверждает умеренную точность аппроксимации. Большее влияние оказывает фактор X_5 (доля дорог, отвечающих нормативным требованиям). Эффект отмечается отрицательный ввиду того, что в Архангельской области около 80% дорог находятся в ненормативном по результатам оценки техническом состоянии, следовательно, крупные финансовые вложения не могут оказать положительный эффект за небольшой временной промежуток. Весомый положительный прирост оказывает также фактор X_1 — фондонасыщенность территории с учетом степени износа основных фондов; как ранее отмечалось, в регионе началась реализация пяти инвестиционных проектов, направленных на увеличение агропромышленных мощностей, что, в соответствии с уравнением регрессии, окажет прямое положительное влияние на уровень продовольственной безопасности. Также прирост к уровню продовольственной безопасности оказывает фактор X_4 (коэффициент Гольца): около 117 населенных пунктов Архангельской области не имеют автотранспортной связи с сетью автомобильных дорог общего пользования по автомобильным дорогам с твердым покрытием, водные пути сообщения также недостаточно развиты, в связи с чем развитие транспортной системы позволит повысить уровень физической доступности продовольствия.

В целом по Арктической зоне Российской Федерации отмечается самый высокий коэффициент детерминации, равный 0,92, что свидетельствует о высокой точности аппроксимации, модель достоверно описывает явление. P -значения для всех коэффициентов, связанных с каждой независимой переменной, приближены к нулю, что говорит о важности показателей для модели. Все факторы обеспечивают положительный при-

рост к уровню продовольственной безопасности. Наибольший прирост оказывает X_4 — коэффициент Гольца, что подтверждает все результаты проведенного исследования. В регионах АЗРФ слабо развита транспортная система, некоторые виды транспорта вовсе отсутствуют, следовательно, развитие транспортной инфраструктуры окажет положительное влияние на продовольственную безопасность как для каждого региона в частности, так и для АЗРФ как макрорегиона. Средний прирост оказывает фактор X_1 — фондонасыщенность территории с учетом степени износа основных фондов. Проблема высокого износа основных фондов и их недостатка характерна практически для всех регионов АЗРФ, развитие производственной инфраструктуры позволит обеспечить поддержание собственного растениеводства, животноводства и рыболовства, повысит коэффициент самообеспеченности регионов продовольствием. Также положительный прирост оказывает фактор X_5 — доля автомобильных дорог, отвечающих нормативным требованиям. Ранее отмечалось, что во многих регионах более половины существующих автомобильных дорог находятся в ненормативном, по результатам оценки, техническом состоянии. Производя качественные ремонтные работы на участках, предназначенных для перевозки грузов, снизятся временные и финансовые затраты на доставку продовольствия, замедлится процесс износа оборудования, что в конечном результате повысит уровень продовольственной безопасности в регионах Арктической зоны Российской Федерации.

Процедура моделирования завершается применением метода Монте-Карло, позволяющим проверить заявленную гипотезу о влиянии включенных переменных в уравнение регрессии. С помощью данного метода осуществляется процедура многократного перебора значений переменных (X_1, X_2, \dots, X_{11}) и определения значения результирующей переменной (Y). Реализация данного метода осуществляется на основе использования языка программирования Python,

интегрированного в кроссплатформенную интерактивную IDE-Spyder. Базовая вариация программного кода процедуры моделирования с использованием метода Монте-Карло представлена на рис. 1.5.



```
1  # -*- coding: utf-8 -*-
2  """
3  Created on Thu Apr  7 17:22:28 2022
4
5  @author: лорар
6  """
7
8  import numpy as np
9  import matplotlib.pyplot as plt
10 import pandas as pd
11 import scipy.stats as sps
12 import statistics
13
14 data=pd.read_excel('C:/Users/Kovzunova/Desktop/Для Питон/Мурманск.xlsx')
15 #print(data)
16 mu1=data['X1'].mean()
17 sd1=data['X1'].std()
18 muY=data['Y'].mean()
19 sdY=data['Y'].std()
20 mu2=data['X2'].mean()
21 sd2=data['X2'].std()
22
23 iterations=10000
24 x1=sps.norm(loc=mu1,scale=sd1).rvs(size=iterations)
25
26 #plt.hist(x1)
27 #plt.show()
28 x2=sps.norm(loc=mu2,scale=sd2).rvs(size=iterations)
29
30 y=[]
31 for i in range(len(x1)):
32     c= -2.16+0.02*x1[i]+1.95*x2[i]
33     y.append(c)
34 #print(y)
35 plt.hist(y,rwidth=0.9)
36 plt.show()
37 mu=statistics.mean(y)
38 sd=statistics.stdev(y)
39 print(mu)
40 print(sd)
```

Рис. 1.5. Базовый код на языке программирования Python для реализации метода Монте-Карло

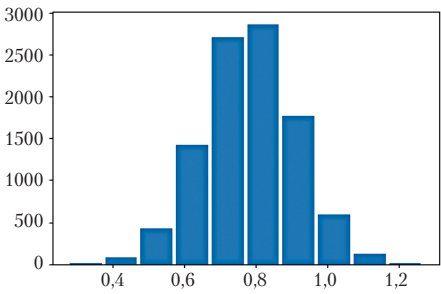
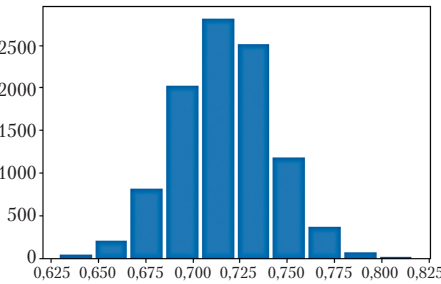
Под каждый регион программный код адаптируется в зависимости от уравнения регрессии и включенных факторов (X_1 , X_2 , ..., X_{11}).

Результаты моделирования с использованием IDE-Spyder для реализации метода Монте-Карло на языке программиро-

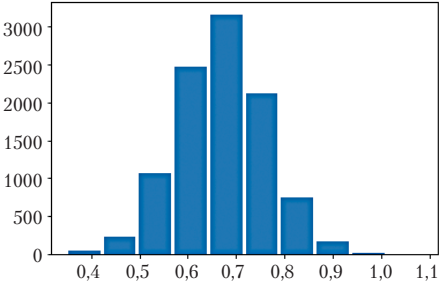
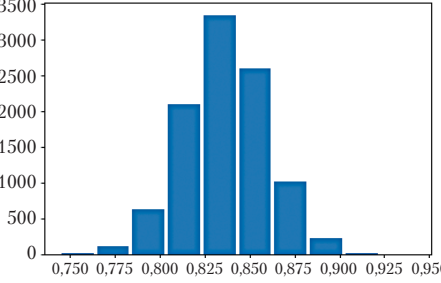
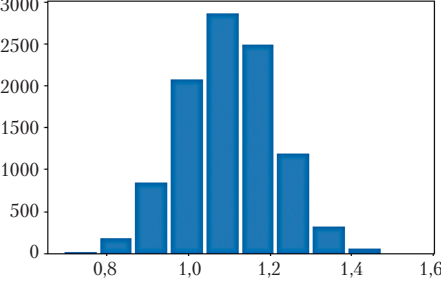
вания Python представлены в табл. 1.8. При моделировании с использованием метода Монте-Карло ключевыми показателями для анализа являются: среднее значение Y при 10 000 итераций (демонстрирует наиболее частый вариант Y при 10 000 изменениях, заявленных X -переменных); стандартное отклонение (характеризует допустимое отклонение от среднего значения; чем меньше отклонение, тем точнее моделирование).

Таблица 1.8

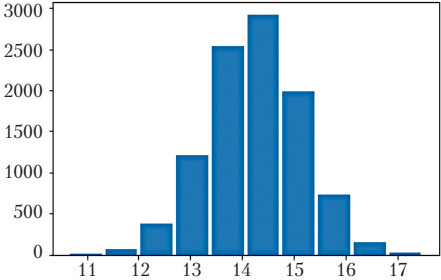
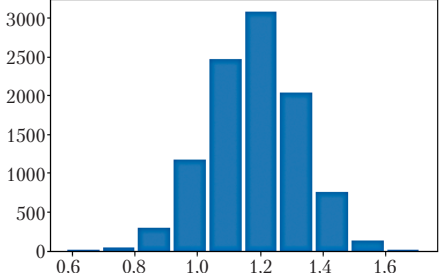
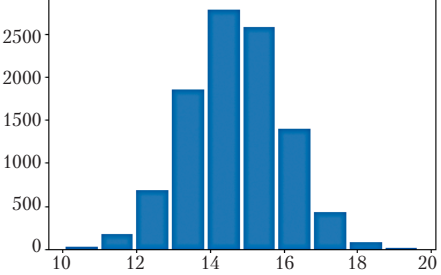
Результаты моделирования в контексте оценки влияния инфраструктурных факторов на продовольственную безопасность регионов АЗРФ на основе применения метода Монте-Карло

Регион АЗРФ	Результаты моделирования с использованием метода Монте-Карло
Мурманская область	 <p>Среднее значение Y при 10 000 итераций 0,81 Стандартное отклонение 0,18</p>
Ненецкий автономный округ	 <p>Среднее значение Y при 10 000 итераций 0,72 Стандартное отклонение 0,02</p>

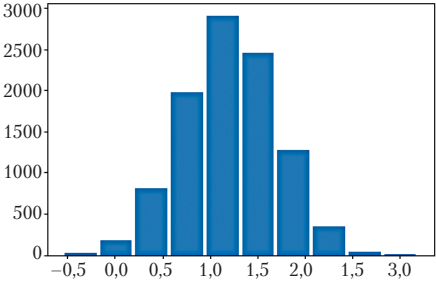
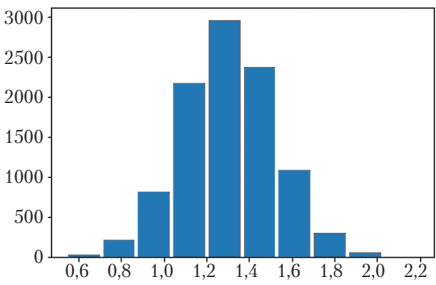
Продолжение табл. 1.8

Регион АЗРФ	Результаты моделирования с использованием метода Монте-Карло
Чукотский автономный округ	 <p data-bbox="322 639 958 701">Среднее значение Y при 10 000 итераций 0,67 Стандартное отклонение 0,09</p>
Ямало- Ненецкий автономный округ	 <p data-bbox="322 1016 958 1078">Среднее значение Y при 10 000 итераций 0,83 Стандартное отклонение 0,02</p>
Республика Карелия	 <p data-bbox="322 1393 958 1455">Среднее значение Y при 10 000 итераций 1,09 Стандартное отклонение 0,11</p>

Продолжение табл. 1.8

Регион АЗРФ	Результаты моделирования с использованием метода Монте-Карло	
Республика Саха (Якутия)		
	Среднее значение Y при 10 000 итераций	1,42
	Стандартное отклонение	0,08
Республика Коми		
	Среднее значение Y при 10 000 итераций	1,18
	Стандартное отклонение	0,14
Краснояр- ский край		
	Среднее значение Y при 10 000 итераций	1,46
	Стандартное отклонение	0,12

Окончание табл. 1.8

Регион АЗРФ	Результаты моделирования с использованием метода Монте-Карло
Архангель- ская область	 <p data-bbox="322 639 960 699">Среднее значение Y при 10 000 итераций 1,2 Стандартное отклонение 0,4</p>
Арктиче- ская зона Российской Федерации	 <p data-bbox="322 1011 960 1071">Среднее значение Y при 10 000 итераций 1,3 Стандартное отклонение 0,21</p>

В результате проведенной процедуры моделирования регионы АЗРФ можно классифицировать на три основные группы: 1) регионы с существенным влиянием инфраструктурных факторов на продовольственную безопасность; 2) регионы с умеренной зависимостью от влияния инфраструктурных факторов на продовольственную безопасность; 3) регионы, продовольственная безопасность которых слабо зависит от влияния инфраструктурных факторов.

К первой группе относится совокупность регионов, уровень продовольственной безопасности которых существенно повы-

шается благодаря развитию инфраструктуры. К данной группе относятся Республика Карелия и Архангельская область. В Республике Карелия на 2020 год уровень продовольственной безопасности составил 0,9, при моделировании с использованием метода Монте-Карло с учетом 10000 различных итераций переменных выявлено, что при учете влияющих инфраструктурных факторов значение интегрального уровня продовольственной безопасности составит в среднем 1,09 (+21%), что позволяет региону перейти границу с нормального уровня продовольственной безопасности на высокий. Таким образом, для Республики Карелия является актуальным вопрос развития производственной инфраструктуры в части повышения фондонасыщенности территории, а также транспортной инфраструктуры, уделяя внимание качеству существующих автомобильных дорог. Аналогичная ситуация отмечается в Архангельской области: в 2020 году в регионе уровень продовольственной безопасности составлял 1,08, в результате процесса моделирования с учетом развития производственной и транспортной инфраструктуры регион сможет укрепить свое положение и повысить уровень продовольственной безопасности до 1,2 (+11%).

Ко второй группе относятся регионы, продовольственная безопасность которых в результате моделирования повышается незначительно. В данную группу входят четыре субъекта – Ненецкий и Чукотский автономные округа (с низким уровнем продовольственной безопасности), Республика Коми и Саха (Якутия) (с высоким уровнем продовольственной безопасности). Чукотский и Ненецкий автономные округа являются аутсайдерами как по уровню продовольственной безопасности, так и по развитию инфраструктуры. Как отмечалось ранее, только в данных субъектах железные дороги и крупные магистрали отсутствуют полностью. Следовательно, для значительного улучшения положения субъектов необходимы глобальные изменения, что невозможно реализовать в короткие сроки. Несмотря на текущий уровень продовольственной

безопасности данных регионов, результаты моделирования прогнозируют положительную динамику. Так, Ненецкому автономному округу удастся перейти из кризисного уровня продовольственной безопасности в низкий с результатами 0,72 (+5,8%), Чукотскому автономному округу – повысить уровень продовольственной безопасности до 0,67 (+6,34%).

Что касается регионов с высоким уровнем продовольственной безопасности, в Республике Саха (Якутия) в результате корреляционного анализа выявлена слабая связь с инфраструктурными факторами, что подтверждается результатами моделирования. Продолжая постепенно развивать производственную и транспортную инфраструктуру на территории, регион сможет поддерживать существующий уровень продовольственной безопасности с поступательной тенденцией роста до 1,42 (+2,89%). В Республике Коми стоит акцентировать внимание не только на вышеуказанных видах инфраструктуры, но и на информационной составляющей: внедряя новые технологии в организациях сельского хозяйства, регион сможет поддерживать и развивать собственное производство, обеспечивая население продовольствием. Уровень продовольственной безопасности Республики Коми в результате моделирования увеличивается на 1,16%.

К третьей группе относятся регионы, продовольственная безопасность которых по результатам моделирования останется на прежнем уровне – Мурманская область и Ямало-Ненецкий автономный округ. Продовольственная безопасность данных регионов находится на низком уровне 0,84 и 0,88 соответственно. В Мурманской области и Ямало-Ненецком автономном округе ввиду географического положения особое внимание уделяется рыбохозяйственному комплексу, следовательно, регионам выгоднее обеспечивать себя и другие субъекты рыбной продукцией, импортируя большинство остальных продуктов питания, следовательно, развитие производственной инфраструктуры в части фондонасыщенности территории для развития сельского хозяйства незначительно

скажется на интегральном уровне продовольственной безопасности, однако позволит поддерживать ее.

В результате моделирования уровня продовольственной безопасности Красноярского края на основе применения метода Монте-Карло при 10 000 итераций факторов выявлено снижение уровня продовольственной безопасности. Регион является лидером среди всех субъектов по уровню продовольственной безопасности (на 2020 год значение показателя составило 1,83). В результате корреляционного анализа выявлена слабая связь практически со всеми инфраструктурными факторами. В регионе достаточно развита инфраструктура, которая позволяет не только самообеспечивать регион продовольствием, но также экспортировать продукты питания в иные регионы с менее благоприятными условиями для развития собственного производства, в связи с чем дополнительное развитие инфраструктуры в рамках текущих потребностей региона не является значимым.

* * *

Таким образом, учитывая наличие определенных достоинств метода Монте-Карло (адаптивность, простота и наглядность моделей, возможность применения моделей исследования многокомпонентных систем, учет различного рода внешних воздействий, доступность программного обеспечения для реализации методического инструментария и пр.), возможно его применение для широкого круга задач исследования сложных систем (производственных и социально-экономических).

Глава 2

МЕТОД МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ DATA ENVELOPMENT ANALYSIS

2.1. ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ ПРИМЕНЕНИЯ МЕТОДА DATA ENVELOPMENT ANALYSIS

Постановка математических задач, возникающих при организации проектирования и сравнении вариантов проектных решений, так или иначе связана с задачами моделирования, оптимизации, выбора вариантов [1].

Аппарат производственных функций хорошо известен. Производственная функция показывает максимальное количество выпуска (продукции), которое может быть произведено из данного количества входных факторов производства (входов) при использовании данной технологии. Если распространить рассуждения на случай, когда выпускается продукция не одного вида, а нескольких, то принято говорить не о производственной функции, а о производственной границе, или границе эффективности. В этом случае те, условно говоря, корпоративные объекты, представленные в выборке, которые производят максимальное количество выпуска из данного количества входов, считаются эффективными, и точки, соответствующие им в пространстве входов-выходов, находятся на этой самой границе эффективности. Точки, которые не находятся на границе эффективности, соответствуют объектам, функционирующим неэффективно. Степень неэффективности определяется степенью удаленности точки от границы эффективности: чем дальше точка от границы, тем объект, соответствующий этой точке, менее эффективен. Поскольку на практике граница эффективности неизвестна, ее

нужно каким-то образом оценить. Для оценивания применяются различные методы, которые позволяют составить комплексную интегральную оценку исследуемой модели на основе входящих в нее элементов и в дальнейшем дают возможность разработать конкретные управленческие решения, связанные с изменением составляющих элементов. Одним из таких методов является Data Envelopment Analysis (представитель группы непараметрических методов линейного программирования).

Данный метод продолжительное время применяется в зарубежной практике для оценки эффективности функционирования объектов со схожими свойствами в различных социальных, технических и экономических системах. Такими объектами могут быть корпоративные объединения, финансово-кредитные учреждения, объекты социальной сферы, органы государственного управления, регионы (территории) и т.д. [26–34].

Впервые идея анализа среды функционирования была сформулирована М.Ж. Farrell в 1957 году [35]. В дальнейшем она была реализована группой американских ученых: А. Charnes, W.W. Cooper, Е. Rhodes [36].

Метод Data Envelopment Analysis (DEA) анализирует некую выборку объектов со схожими свойствами (DMU — Decision Making Units). Суть метода DEA на примере хозяйственной системы представлена в работах российских и зарубежных ученых [26–34]. В методологии DEA используется термин «эффективность функционирования», отражающий эффективность, с которой исследуемые объекты преобразуют входы (input) в выходы (output).

Специфическая особенность методической основы анализа среды функционирования заключается в том, что исходные данные модели необходимо структурировать специальным образом: в виде входных и выходных параметров. К входным параметрам модели относят показатели, для которых предпочтительными являются меньшие значения. Противоположная

характеристика присуща выходным параметрам — к ним относят показатели, для которых предпочтительными являются большие значения [37]. Методика DEA не накладывает ограничения на максимальное количество входных и выходных параметров, однако присутствует минимальное ограничение на количество параметров — по крайней мере один входной и выходной параметр.

С точки зрения возврата к масштабу модели DEA можно разделить на модели с постоянным эффектом масштаба (constant returns to scale — CRS) и с переменным эффектом масштаба (variable returns to scale — VRS).

Первоначально группа американских ученых (Charnes, Cooper и Rhodes [36]) разработала метод измерения эффективности для постоянного возврата к шкале. Производственная граница в данной модели принимала вид прямой. Позже Banker, Charnes и Cooper [38] представили модель измерения эффективности с переменным возвратом к шкале, подразумевающую использование коэффициентов масштаба. Производственная граница в данной модели будет принимать вид кусочно-линейной прямой.

Поскольку методология использования метода DEA основана на построении границы эффективности в виде векторов ввода-вывода, следовательно, они (векторы) могут иметь различную ориентированность: 1) модели, ориентированные на вход; 2) модели, ориентированные на выход.

Целью модели, ориентированной на выход, является увеличение величины выходных параметров при постоянной величине входных параметров. Противоположная цель присуща моделям, которые ориентированы на вход: минимизация величины входных параметров при постоянной величине выходных параметров.

Для иллюстрации различий в работе CRS- и VRS-моделей, была использована модель с одним входом и одним выходом для 5 различных наблюдений, представленная на рис. 2.1.

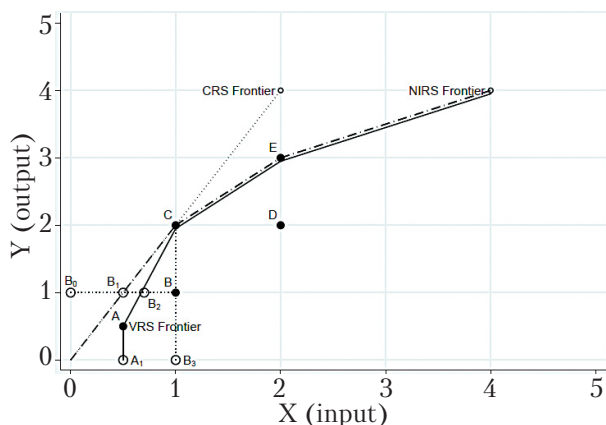


Рис. 2.1. Производственные границы для моделей постоянного и переменного масштаба [38]

На данном графике модель с постоянным масштабом представлена прямой $O-C$ (*CRS Frontier*). Эффективным наблюдением в данной модели является наблюдение C . В случае построения модели с переменным эффектом масштаба эффективными являются наблюдения A , C и E (*VRS Frontier*), поскольку они находятся на границе эффективности.

Для наблюдения B эффективность рассчитывается в соответствии с формулой

$$\theta_{B,input,CRS} = \frac{\overline{B_0B_1}}{\overline{B_0B}}, \quad (2.1)$$

где $\overline{B_0B_1}$ и $\overline{B_0B}$ — векторы расстояния до границы эффективности для модели с постоянным эффектом масштаба, ориентированной на вход.

В данной модели смещение происходит влево в сторону оси ординат. На данном примере видно, что изменение эффективности происходит при неизменном значении выходного параметра. Для *CRS*-модели, ориентированной на выход, изменение будет рассчитываться в соответствии с формулой

$$\theta_{B,output,CRS} = \frac{\overline{B_3B_1}}{\overline{B_3C}}, \quad (2.2)$$

где $\overline{B_3B_1}$ и $\overline{B_3C}$ — векторы расстояния до границы эффективности для модели, ориентированной на выход.

Эффективность для наблюдения в случае использования модели с переменным эффектом масштаба определяется в соответствии с формулой

$$\theta_{B,input,CRS} = \frac{\overline{B_0B_2}}{\overline{B_0B}}, \quad (2.3)$$

где $\overline{B_0B_2}$ и $\overline{B_0B}$ — векторы расстояния до границы эффективности, построенной с учетом переменного масштаба в модели, ориентированной на вход.

Также возможно разложение неэффективности в *CRS*-модели на «чистую» техническую неэффективность и масштабную неэффективность. На графике в данном случае вектор $\overline{B_2B}$ характеризует техническую эффективность наблюдения B относительно *VRS*-модели, а вектор $\overline{B_1B}$ характеризует техническую эффективность относительно *CRS*-модели. В таком случае вектор $\overline{B_1B_2}$ способствует масштабной эффективности.

Суть метода *DEA* представлена в работах Моргунова Е.П. и Моргуновой О.И. [37]. Пусть имеются данные для K входных параметров и M выходных параметров для каждого из N объектов (под термином «объект» могут подразумеваться регионы, отрасли хозяйства, предприятия, учебные заведения и т.д.). Для i -го объекта они представлены вектор-столбцами x_i и y_i соответственно. Тогда матрица X размерности $K \cdot N$ представляет матрицу входных параметров для всех N объектов, а матрица Y размерности $M \cdot N$ представляет матрицу выходных параметров для всех N объектов. Можно прийти к задаче математического программирования

и, используя теорию двойственности, сформулировать ее в виде формулы

$$\begin{aligned} & \min_{\theta, \lambda} (\theta), \\ & -y_i + Y\lambda \geq 0, \\ & \theta x_i - X\lambda \geq 0, \\ & \lambda \geq 0, \end{aligned} \tag{2.4}$$

где θ — скаляр; λ — вектор констант размерности $N \cdot 1$ [37].

Значение θ , полученное при решении задачи, и будет мерой эффективности i -го объекта. При этом эффективность не может превышать единицы. Важно помнить, что аналогичная задача решается N раз, т.е. для каждого объекта. Те объекты, для которых значение показателя эффективности оказалось равным единице, находятся на границе эффективности. В результате может быть сформирована кусочно-линейная граница эффективности. Точки, соответствующие тем объектам, у которых показатель эффективности оказался меньше единицы, можно спроецировать на границу эффективности таким образом, что каждая из этих точек будет равна линейной комбинации $(X\lambda, Y\lambda)$. Часть элементов вектора λ имеют «ненулевые значения». Эти элементы соответствуют тем объектам, которые являются эталонными для оцениваемого объекта. Линейная комбинация эталонных объектов и образует гипотетический объект, находящийся на границе эффективности. Гипотетический объект считался бы эффективным, если бы существовал в действительности. Но поскольку он не существует, то значения его переменных являются целью для реального (неэффективно) объекта. В результате для объектов с $\theta > 1$ могут быть установлены цели, которые заключаются в пропорциональном сокращении их входных факторов на величину θ при сохранении выходных значений на прежнем уровне. Чем ближе точка, соответствующая данному объекту, к границе эффективности, тем выше ее мера эффективности.

Приведенная модель называется моделью, ориентированной на вход и принимающей наличие постоянного эффекта масштаба. Для того чтобы учесть возможность переменного эффекта масштаба, нужно в данную модель добавить ограничение на сумму весовых коэффициентов λ [37, 39] и представить в соответствии с формулой

$$\sum \lambda_i = 1. \quad (2.5)$$

Следствием ввода этого ограничения является формирование выпуклой линейной комбинации эталонных объектов.

Метод *DEA*, обладает рядом существенных преимуществ, к которым можно отнести следующие [27, 37]:

- метод *DEA*, в отличие от большинства других популярных методов, позволяет анализировать модели с большим количеством входных и выходных параметров, каждый из которых может быть представлен в различных единицах измерения;
- метод *DEA* не налагает ограничения на характер зависимости между входными и выходными параметрами;
- использование анализа среды функционирования не требует проверки гипотезы о наличии функциональных связей между входными и выходными параметрами;
- метод *DEA* позволяет оценить влияние внешней среды на исследуемую систему;
- решение задач оптимизации с помощью метода *DEA* не требует обязательного указания значимости каждого из показателей, но при необходимости позволяет отразить мнения экспертов о важности каждого из используемых в модели факторов;
- данный метод позволяет сформировать единый показатель эффективности исследуемой системы;
- метод формирует Парето-оптимальное множество точек, которые соответствуют эффективным объектам;

- метод позволяет определить объем необходимых изменений в категориях входов и выходов, которые необходимо осуществить, чтобы неэффективные объекты стали эффективными.

При использовании статистических данных за несколько лет появляется возможность проследить перемещение границы эффективности во времени. На основании направления этих перемещений можно определить, имеет ли место прогресс в исследуемой группе объектов (отрасли) или же регресс. Данный метод также позволяет определить, за счет чего достигнут прогресс: за счет улучшения управления, за счет приведения масштаба объекта к оптимальному уровню либо за счет изменения технологии (например, внедрения нового оборудования). Кроме того, можно определить так называемую распределительную эффективность (*allocative efficiency*), т.е. эффективность использования ресурсов, если известны их стоимости [37].

Для иллюстрации концепции эффективности в целом и метода *DEA* в частности рассмотрим процесс производства, в котором задействованы два входных фактора производства x_1 и x_2 и производится один вид продукции y . При этом сделаем важное допущение о том, что в нашем случае будет иметь место постоянный эффект масштаба. Это допущение позволит нам для графического изображения применяемой технологии производства использовать двухмерный график. По осям координат этого графика будут откладываться удельные затраты входных факторов производства, т.е. затраты, приходящиеся на единицу выпускаемой продукции. Таким образом, мы получаем единичную изокванту, представленную на рис. 2.2 [38].

Если некоторая фирма использует входные факторы в количествах, представленной точкой H на рис. 2.2, то в этом случае ее техническая неэффективность будет выражаться длиной отрезка H_1H (где точка H_1 является проекцией точки H на границу эффективности). Эта длина представляет собой

величину, на которую могут быть пропорционально сокращены величины входов без уменьшения величины выпускаемой продукции (выпуска).

Техническая эффективность фирмы H , (TE), будет равна отношению длин отрезков OH_1 и OH (в соответствии с формулой):

$$TE = \frac{OH_1}{OH}. \quad (2.6)$$

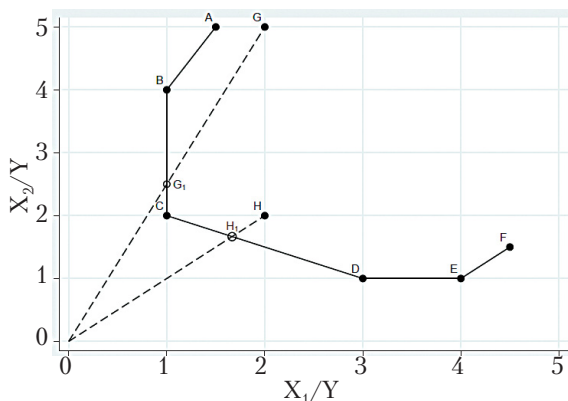


Рис. 2.2. Технология производства с двумя выходами и одним входом (составлено авторами на основании [38])

На рис. 2.2 точки A , B , C , D , E и F являются эффективными (они и формируют границу эффективности), а точки H и G — неэффективными (они не принадлежат данной границе).

Ясно, что значение технической эффективности не может превышать единицы. Важно заметить, что проецирование неэффективной точки на границу эффективности допустимо на основании одного из базовых положений метода *DEA*. Суть этого положения заключается в том, что если одна фирма может использовать входные факторы таким образом, что выпускает из них некоторое количество выпуска, то и другая

фирма — неэффективная — также должна быть в состоянии выпускать такое же количество продукции из такого же количества входных факторов производства. При проецировании на границу для каждого неэффективного объекта формируется эталонный гипотетический объект, который является эффективным и служит целью для неэффективного объекта в смысле достижения Парето-эффективности или оптимальности Парето-Кумпанса [37, 38]. Эталонный объект, как правило, является комбинацией двух или более реальных объектов, которые являются эффективными. Значения показателей этого эталонного объекта и служат целями для неэффективного объекта.

Аналогичные модели могут быть построены и с ориентацией на выход. В этом случае главной целью моделей будет увеличение выпуска продукции без увеличения затрат входных ресурсов. В результате расчетов по этим моделям будут получены не только значения показателя эффективности для каждого из объектов, но также указаны рекомендуемые значения выходных переменных, при достижении которых неэффективные объекты могут быть выведены на границу эффективности. В соответствии с формулой (2.7) представлена модель с постоянным эффектом масштаба:

$$\begin{aligned} & \max_{\varphi, \lambda} (\varphi), \\ & -\varphi y_i + Y\lambda \geq 0, \\ & x_i - X\lambda \geq 0, \\ & \lambda \geq 0. \end{aligned} \tag{2.7}$$

Следует обратить внимание на то, что в данном случае значение переменной φ , рассчитанное по этим моделям, будет больше или равно единице. Это объясняется тем, что целью моделей является пропорциональное увеличение значений выходных переменных. Для получения же традиционного

значения показателя эффективности, который находится в пределах от нуля до единицы, следует просто использовать величину, обратную к ϕ , что является общеприменимой практикой.

Метод анализа среды функционирования стал основой для создания индекса производительности Малмквиста. Его идея была выдвинута *S. Malmquist* [40] и в дальнейшем реализована *D.W Caves* и *L.R. Christensen* [41].

Исторически индекс Малмквиста (*Malmquist index*) используется для оценки изменения общей продуктивности всех факторов производства, когда рассматривается один и тот же объект в рамках двух различных периодов времени. Продуктивность определяется как отношение выпуска продукции (выхода) к затратам ресурсов (входу).

Для вычисления индекса используются различные методы, в том числе и метод *DEA*, поскольку данный метод может обрабатывать много выходов и входов одновременно. При использовании *DEA* нужно решить несколько оптимизационных задач для каждого из объектов в выборке (в соответствии с формулами (2.8–2.11):

$$\begin{aligned} & \max_{\phi, \lambda} (\phi), \\ & -\phi y_{it} + Y_t \lambda \geq 0, \\ & x_{it} - X_t \lambda \geq 0, \\ & \lambda \geq 0; \end{aligned} \tag{2.8}$$

$$\begin{aligned} & \max_{\phi, \lambda} (\phi), \\ & -\phi y_{is} + Y_s \lambda \geq 0, \\ & x_{is} - X_s \lambda \geq 0, \\ & \lambda \geq 0; \end{aligned} \tag{2.9}$$

$$\begin{aligned}
& \max_{\varphi, \lambda} (\varphi), \\
& -\varphi y_{is} + Y_t \lambda \geq 0, \\
& x_{is} - X_t \lambda \geq 0, \\
& \lambda \geq 0;
\end{aligned} \tag{2.10}$$

$$\begin{aligned}
& \max_{\varphi, \lambda} (\varphi), \\
& -\varphi y_{it} + Y_s \lambda \geq 0, \\
& x_{it} - X_s \lambda \geq 0, \\
& \lambda \geq 0,
\end{aligned} \tag{2.11}$$

где s — индекс, обозначающий начальный период времени; t — индекс, обозначающий конечный периоды времени; i — индекс, обозначающий объект, для которого производятся расчеты.

Как и в модели, представленной в формуле (2.4), Y и X — это матрицы выходов и входов для всех объектов в выборке, а λ — вектор весовых коэффициентов, которые образуют линейную комбинацию — гипотетический объект, являющийся целью для неэффективного объекта. В задачах, представленных в соответствии с формулами (2.10) и (2.11), есть важная особенность — показатели объекта и технология, относительно которой определяется его эффективность, принадлежат к различным временным периодам: в задаче, представленной в соответствии с формулой (2.10), объект из предыдущего периода сравнивается с технологией следующего периода. В задаче, представленной в соответствии с формулой (2.11), наоборот, объект из следующего периода сравнивается с технологией предыдущего периода. Важно отметить, что в задачах, представленных в соответствии с формулами (2.10) и (2.11), показатель эффективности, который в таких моделях определяется как величина, обратная к φ , может иметь значение, большее единицы. Это возможно в том случае, если в задаче, представленной в соответствии с формулой (2.10),

имел бы место технический регресс в данной группе объектов, а в задаче, представленной в соответствии с формулой (2.11), наоборот — технический прогресс.

Если рассчитать значения индекса Мальмквиста для всех объектов в выборке за целый ряд периодов (а не только за два периода), то, разложив эти значения на две составляющие, можно получить временной ряд показателей изменения технического прогресса. Эти сведения можно использовать для прогнозирования состояния технологии в данной отрасли (в данной группе объектов). Точнее говоря, можно прогнозировать значения показателей, которые могут быть достигнуты лучшими объектами в будущем. На рис. 2.3 продемонстрирована идея изменения технологий с течением времени.

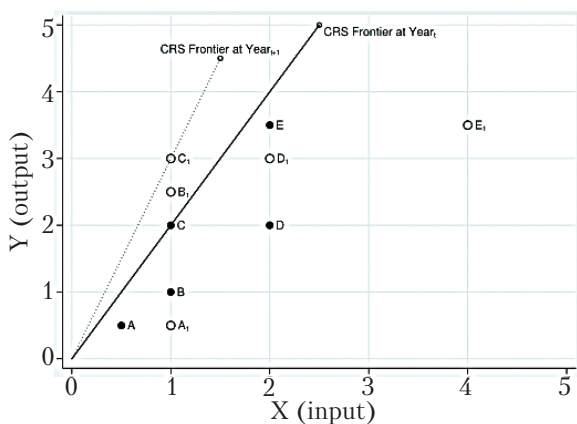


Рис. 2.3. Перемещение границы с течением времени (составлено авторами на основании [37, 38])

На рис. 2.3 представлено 5 наблюдений и границы, наблюдаемые за периодами t и $t+1$, определенные путем расчета модели с постоянным масштабом с одним входом и одним выходом. Рассматривая наблюдения в период t $A(0,5; 0,5)$, $B(1; 1)$, $C(1; 2)$, $D(2; 2)$, $E(2; 3,5)$ и соответствующие им наблюдения в пе-

риод $t+1$ $A_1(1; 0,5)$, $B_1(1; 2,5)$, $C_1(1; 3)$, $D_1(2; 3)$, $E_1(4; 3,5)$, представляется возможность для расчета индекса Малмквиста на основе применения метода *DEA*. Например, эффективность для наблюдения D может быть получена следующим образом. Во-первых, необходимо найти расстояние как от точки D , так и от границы до контрольной точки D_1 . Во-вторых, найти соотношение двух расстояний, которое рассчитывается как мера расстояния от наблюдения D до *CRS*-границы за период t , разделенное на меру расстояния от наблюдения D_1 до *CRS*-границы за период $t + 1$.

Для оценки динамики продуктивности исследуемых моделей используется индекс Малмквиста, включающий в себя 5 показателей [42].

Общий показатель продуктивности *TFPCH* (*total factor productivity change*) позволяет оценить абсолютное изменение показателя для наблюдения I . Он рассчитывается в соответствии с формулой

$$MPI = \left[\left(\frac{E_I^t(x^{t+1}, y^{t+1})}{E_I^t(x^t, y^t)} \right) \cdot \left(\frac{E_I^{t+1}(x^{t+1}, y^{t+1})}{E_I^{t+1}(x^t, y^t)} \right) \right]^{1/2}, \quad (2.12)$$

где E — расчет эффективности; t — базовый период; (x^t, y^t) — вектор наблюдения I за базовый период t .

Общий показатель продуктивности *TFPCH* можно представить в качестве произведения двух сомножителей:

- *EFFCH* — изменение эффективности наблюдения;
- *TECHCH* — изменение технического прогресса в отрасли.

Показатель изменения эффективности наблюдения рассчитывается в соответствии с формулой

$$EFFCH_I = \frac{E_I^{t+1}(x^{t+1}, y^{t+1})}{E_I^t(x^t, y^t)}. \quad (2.13)$$

Данный показатель характеризует изменение эффективности использования ресурсов в модели с постоянным эф-

фектом масштаба (характеризует изменения относительно внутренних факторов наблюдения).

Технический прогресс *TECHCH* (*technical change*) характеризует технический прогресс отрасли. Он служит для оценки внешних изменений в отрасли и рассчитывается в соответствии с формулой

$$TECHCH_I = \left[\left(\frac{E_I^t(x^t, y^t)}{E_I^{t+1}(x^t, y^t)} \right) \cdot \left(\frac{E_I^t(x^{t+1}, y^{t+1})}{E_I^{t+1}(x^{t+1}, y^{t+1})} \right) \right]^{1/2}. \quad (2.14)$$

При использовании модели с переменным эффектом масштаба показатель изменения эффективности *EFFCH* можно разделить на два сомножителя: *PECH* — «чистое» изменение эффективности, *SECH* — масштабное изменение эффективности.

«Чистая» техническая эффективность с использованием *VRS*-модели *PECH* (*pure technical efficiency change*) характеризует интенсивное изменение эффективности. Показатель рассчитывается в соответствии с формулой

$$PECH = \frac{E_{crs}^{t+1}(x^{t+1}, y^{t+1})}{E_{crs}^t(x^t, y^t)}, \quad (2.15)$$

где E_{crs}^{t+1} — показатель эффективности за отчетный период в модели с переменным масштабом; E_{crs}^t — показатель эффективности за базовый период в модели с постоянным масштабом.

Масштабная эффективность *SECH* (*scale efficiency change*) характеризует экстенсивное изменение эффективности. Данный показатель рассчитывается в соответствии с формулой

$$SECH = \left[\frac{E_{crs}^{t+1}(x^{t+1}, y^{t+1}) / E_{crs}^{t+1}(x^t, y^t)}{E_{crs}^{t+1}(x^t, y^t) / E_{crs}^{t+1}(x^t, y^t)} \times \frac{E_{crs}^t(x^{t+1}, y^{t+1}) / E_{crs}^t(x^t, y^t)}{E_{crs}^t(x^t, y^t) / E_{crs}^t(x^t, y^t)} \right]^{1/2}. \quad (2.16)$$

Таким образом, метод *Data Envelopment Analysis* может быть рекомендован в контексте формирования комплексной интегральной оценки исследуемой модели на основе входящих в нее элементов для последующего предоставления возможности разработать конкретные управленческие решения, связанные с изменением составляющих элементов.

2.2. ЭКСПЕРИМЕНТАЛЬНОЕ ПРИМЕНЕНИЕ МЕТОДА DATA ENVELOPMENT ANALYSIS ДЛЯ РЕШЕНИЯ ЗАДАЧИ ОЦЕНКИ ЭФФЕКТИВНОСТИ РЕГИОНАЛЬНОЙ СИСТЕМЫ ЗДРАВООХРАНЕНИЯ

В силу того, что задача при использовании метода DEA формулируется в терминах входов и выходов (inputs/outputs), то необходимо отнести одну часть показателей, характеризующих исследуемые объекты, к входам, а другую часть показателей — к выходам. Однако при использовании метода *DEA* в ряде предметных областей возникает проблема разделения показателей на входные и выходные. Это объясняется тем, что между показателями может не быть технологической связи, как это имеет место в процессе традиционного материального производства [37].

Одним из подходов к решению указанной проблемы может быть такой: показатели, для которых более предпочтительными считаются меньшие значения, следует условно относить к входным, а показатели, для которых, наоборот, предпочтительными являются большие значения, следует условно относить к выходным. В таком случае после выполнения вычислений по методу *DEA* мы получим для «неэффективных» объектов рекомендации по снижению значений входных и увеличению значений выходных показателей, что должно соответствовать логике конкретной предметной области.

Возможны и другие подходы к решению задачи разделения показателей на входные и выходные. Могут возникать ситуации, когда, исходя из специфики конкретной предметной

области, разделение показателей на входные и выходные произведено таким образом, что значения части входных показателей целесообразно увеличивать, а значения части выходных показателей — уменьшать (например, уровень загрязнения атмосферы при производстве металла). В таких случаях вместо фактических значений входных показателей следует использовать их отклонения от пороговых значений, установленных на уровне, заведомо превышающем значения соответствующих показателей для всех исследуемых объектов [43, 44].

Аналогично следует поступать и с выходными показателями. Если после преобразований, проведенных над исходными данными, окажется, что часть входных показателей для отдельных объектов имеет отрицательные значения, то это не будет являться препятствием для использования метода *DEA* при соблюдении определенных условий.

Как известно, модели *DEA* могут быть ориентированными на вход или на выход. В первом случае это означает, что модель фокусируется в первую очередь на снижении значений входных показателей неэффективных объектов (при неизменных значениях выходных показателей), а во втором случае главной целью ее работы является увеличение значений выходных показателей (при неизменных значениях входных показателей) неэффективных объектов. В том случае, когда используется модель, ориентированная на выход, и при этом среди значений входных показателей есть отрицательные, следует увеличить значения такого показателя для всех объектов на величину, равную по модулю наименьшему из отрицательных значений. При этом значение коэффициента эффективности не изменится. Для различных показателей это увеличение может быть различным. При использовании модели, ориентированной на вход, аналогичные действия можно произвести с выходными показателями [45].

Метод *DEA* позволяет определять относительную эффективность объектов. Это означает, что они сравниваются между собой. Однако может быть предложен такой подход: эксперты

формируют некоторое множество гипотетических объектов, имеющих значения показателей такие, что эти гипотетические объекты могут быть приняты в качестве эталонов [37, 46]. Конечно, значения показателей таких объектов должны выбираться с учетом реальной достижимости этих значений. Таким образом, гипотетические объекты будут образовывать границу эффективности, с которой можно сопоставлять реальные объекты.

Методический подход, заложенный в *Data Envelopment Analysis*, может быть адаптирован к проблемам регионального социально-экономического развития.

Для реализации эксперимента регламентируется следующая постановка задачи: определить уровень социально-экономической безопасности территорий и эффективность функционирования региональной системы здравоохранения в контексте ее влияния на социально-экономическое развитие на основе применения метода *Data Envelopment Analysis*.

Когда речь идет о конкретном регионе (субъекте Российской Федерации), нужно учитывать, что сам такой регион состоит из определенного числа районов, а каждый район включает в себя целый ряд муниципальных образований. С другой стороны, регион можно рассматривать и как совокупность отраслей экономики, функционирующих на его территории и на территориях его районов и муниципальных образований. Таким образом, можно прийти к матричной структуре, в которой одно измерение составляют регионы, а второе измерение образуют отрасли экономики. Следовательно, можно сформировать такие матрицы на различных уровнях иерархии. В нашем случае анализ проведен на региональном уровне в соответствии со спецификой выбранного исследования.

Применяя метод *DEA* на различных уровнях иерархии регионов и отраслей, можно определить не только уровень социально-экономической безопасности конкретного региона относительно совокупной выборки регионов России,

но также установить, внутреннюю структуру достигнутого уровня безопасности. В данном случае показателями структуры будут являться коэффициенты, формирующие гипотетический объект, являющийся целевым для конкретного объекта.

Важным фактором, влияющим на алгоритм формирования эталонных объектов в сфере экономической безопасности системы «Здравоохранение», является наличие пороговых значений показателей. К сожалению, в литературе эти пороговые значения приводятся без указания возможных (допустимых) отклонений в ту или иную сторону. По мнению некоторых авторов, такие допустимые отклонения имеют смысл, поскольку оценка уровня социально-экономической безопасности выводится на основе совокупного влияния всех факторов. Отрицательное влияние одного фактора может до некоторой степени компенсироваться положительным влиянием другого фактора [47]. Таким образом, отрицательное отклонение от порогового значения по одному показателю может нивелироваться превышением порогового значения по другому показателю. Более того, в рамках анализа сложных динамических систем уже упомянутые отклонения могут негативно сказаться на полученном результате.

Обоснованный вывод относительно уровня социально-экономической безопасности региона и развития всей сферы «Здравоохранение» региона можно делать только на основании детального анализа, проведенного на всех уровнях иерархии объектов (система здравоохранения субъекта Российской Федерации, муниципальных образований, учреждения здравоохранения, отделения). Таким образом, число анализируемых объектов увеличивается с понижением уровня иерархии.

Кроме собственно сферы «Здравоохранение» целесообразно ввести в рассмотрение еще и, условно говоря, социально-экономические сферы, которые формируются как результат взаимодействия нескольких отраслей. В качестве

примеров могут служить сфера промышленного развития, сфера экологии, сфера образования и т.п.

На первом этапе анализа проводится оценка уровня социально-экономической безопасности регионов России.

Поскольку на данном этапе каждый регион рассматривается как единое целое, то показатели должны иметь межотраслевой характер. В данной модели представлены следующие три показателя: 1) валовой региональный продукт (ВРП) на душу населения; 2) среднедушевые доходы населения; 3) ожидаемая продолжительность жизни.

Результатом этапа является ранжирование регионов по уровню относительного развития и установка целей для неэффективных регионов по достижению определенных значений показателей. Наличие данного этапа обусловлено необходимостью выявления уровня социально-экономической безопасности, выступающей фоном для дальнейшего рассмотрения сферы здравоохранения. К тому же сравнительные оценки результатов данного этапа с последующими этапами исследования поможет установить взаимосвязь между развитием сферы здравоохранения и уровнем социально-экономической безопасности.

На втором этапе проводится оценка здоровья населения регионов России.

На данном этапе модель представлена следующими тремя показателями: 1) уровень рождаемости; 2) уровень смертности; 3) уровень заболеваемости по различным группам заболеваний.

Результатом этапа является ранжирование регионов по уровню состояния здоровья населения и установка целей для неэффективных регионов по достижению определенных значений показателей.

На третьем этапе проводится оценка эффективности функционирования отрасли здравоохранения в регионах России.

В данной модели представлены следующие пять показателей: 1) число врачей; 2) число коек; 3) число посещений

в смену; 4) расходы регионального здравоохранения на душу населения; 5) хроническая заболеваемость на 1000 человек населения.

Результатом этапа является ранжирование регионов по уровню эффективности функционирования отрасли здравоохранения и установка целей для неэффективных региональных отраслей по достижению определенных значений показателей.

Реализация этапа «Оценка уровня социально-экономической безопасности регионов России».

Для исследования в рамках данной модели было взято три наиболее емких показателя социально-экономической безопасности: валовый региональный продукт на душу населения, среднедушевые доходы населения, ожидаемая продолжительность жизни. В отношении всех трех показателей в рамках исследуемой задачи по увеличению уровня безопасности предполагается повышение значений, поэтому все три показателя целесообразно было бы взять в качестве выходных. В данном случае входным показателем выступал бы унифицированный вход, равный единице. Однако специфика расчета модели не предполагает наличия унифицированных входов. В основном, это связано с тем, что был проведен расчет динамического индекса Мальмквиста, невозможный при использовании унифицированного входа.

Поэтому в качестве входного показателя был выбран показатель отклонения ожидаемого уровня продолжительности жизни от стратегически запланированного (78 лет). Использование данного показателя встраивается в логику решаемой задачи по оценке социально-экономической безопасности.

Спецификация модели *DEA* исходила из следующих соображений. Во-первых, представленные показатели в рамках логики решаемой задачи предполагают увеличение значений, следовательно, модель должна быть ориентирована на выход. Ориентация на выход совершенно не означает, что входные показатели не будут затронуты. В данном случае выбира-

ется логика расчета модели, предполагающая максимизацию функции при расчете интегрального показателя эффективности, которая, в соответствии с принципом Парето-Кумпанса, будет затрагивать как выходные, так и входные показатели. Во-вторых, поскольку рассматриваемые регионы структурно могут довольно сильно различаться, берется модель с переменным эффектом масштаба. В данном случае коэффициент эталонного объекта будет указывать на структурное различие или подобие объектов и то, в какой степени оно выражено.

Результатом данного этапа является ранжирование регионов по уровню социально-экономической безопасности и установка целей для неэффективных регионов по достижению определенных значений показателей.

Таким образом, следуя указанной логике, изначально был рассчитан интегральный показатель эффективности. В соответствии с его значением было проведено ранжирование регионов от наиболее эффективного к наименее эффективному. Результаты расчетов представлены в табл. 2.1 (выборка регионов на примере субъектов Сибирского федерального округа).

Таблица 2.1

Результат оценки эффективности социально-экономической безопасности (выборка регионов, 2019 г.)

№ п/п	Регионы РФ	Ранг	Эффективность	Коэффициенты, формирующие гипотетический объект
1	Республика Алтай	9	0,642168	0,632523
2	Республика Тыва	10	0,503335	0,503335
3	Республика Хакасия	8	0,750276	0,750276
4	Алтайский край	5	0,813576	0,739244
5	Красноярский край	1	1	1
6	Иркутская область	6	0,794293	0,794293
7	Кемеровская область	7	0,758445	0,758445

Окончание табл. 2.1

№ п/п	Регионы РФ	Ранг	Эффективность	Коэффициенты, формирующие гипотетический объект
8	Новосибирская область	3	0,999976	0,851064
9	Омская область	4	0,958475	0,866107
10	Томская область	2	0,999998	0,793492

По результатам расчетов только один регион (Красноярский Край) является эффективным и потому служит эталоном для остальных субъектов. В основном это связано с большим объемом ВРП на душу населения в сравнении с другими субъектами. Его значение превышает среднее по остальным регионам в два раза. Отсюда наблюдается сильная разрозненность в весовых коэффициентах, особенно значимая по сравнению с наименее эффективными регионами — Республикой Алтай и Республикой Тыва. Коэффициенты равны 0,63 и 0,50 соответственно, что свидетельствует о значительных региональных структурных различиях.

Упорядочение регионов по уровню их эффективности представлено в табл. 2.2.

Таблица 2.2

Регионы СФО, упорядоченные по значению эффективности (уровень социально-экономической безопасности)

Ранг	Регион РФ	Эффективность
1	Красноярский край	1
2	Томская область	0,999998
3	Новосибирская область	0,999976
4	Омская область	0,958475
5	Алтайский край	0,813576
6	Иркутская область	0,794293

Окончание табл. 2.2

Ранг	Регион РФ	Эффективность
7	Кемеровская область	0,758445
8	Республика Хакасия	0,750276
9	Республика Алтай	0,642168
10	Республика Тыва	0,503335

Среднее значение показателя эффективности равно 0,827893, минимальное значение составляет 0,503335 при стандартном отклонении, равном 0,154611, указывающем на сравнительно разнородную группу регионов. Несколько больше отстает Республика Алтай и сильно диссонирует Республика Тыва, что было отмечено ранее.

Учитывая, что метод *DEA* позволяет формировать рекомендуемые значения показателей для неэффективных объектов, в рамках данной задачи это регионы, у которых значение показателя эффективности меньше единицы. Если бы неэффективные регионы достигли рекомендуемых значений показателей, то они вышли бы на границу эффективности. Расчетные рекомендуемые значения показателей для достижения границы эффективности, представлены в табл. 2.3. У регионов, имеющих значение показателя эффективности, равное единице, рекомендуемые значения показателей совпадают с исходными значениями.

Следуя расчетным значениям, можно сформулировать следующие выводы. Большинство регионов выборки существенно отстают по показателю ВРП на душу населения (в сравнении с регионом, эффективность которого равна единице), в связи с этим для некоторых регионов требуется значительное увеличение объема указанного индикатора. Можно утверждать, что такие регионы, как Новосибирская область и Томская область, считаются эффективными, но в силу большой разницы между объемами ВРП на душу населения они не могут образовывать производственную границу. Для таких регионов изменение значений незначительно.

Таблица 2.3

Рекомендуемые значения показателей по уровню социально-экономической безопасности для выборки регионов

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Входы/Выходы	Значения показателей (исходные)	Значения показателей (реко-мендуемые)	Разность абсолютная	
1	Республика Алтай		0,642168			
	ВРП на душу населения		204849,50	270913,41	66063,91	32,25
	Среднедушевые доходы		13536	13536	0,00	0,00
	Ожидаемая продолжительность жизни		65,68	65,68	0,00	0,00
2	Республика Тыва		0,503335			
	ВРП на душу населения		166415,70	235776,95	69361,25	41,68
	Среднедушевые доходы		10160	10160	0,00	0,00
	Ожидаемая продолжительность жизни		60,45	63,36934	2,91934	4,83
3	Республика Хакасия		0,750276			
	ВРП на душу населения		365436,60	433556,63	68120,03	18,64
	Среднедушевые доходы		12806	12806	0,00	0,00
	Ожидаемая продолжительность жизни		67,1	67,610188	0,510188	0,76
4	Алтайский край		0,813576			0
	ВРП на душу населения		211658,10	273682,57	62024,47	29,30

Продолжение табл. 2.3

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Значения показателей (исходные)	Значения показателей (реконструируемые)	Разность абсолютная		
	Среднедушевые доходы	11029	11029	0,00	0,00	
	Ожидаемая продолжительность жизни	68,36	68,36	0,00	0,00	
5	Красноярский край	1				
	ВРП на душу населения	608083,30	608084,92	0,00	0,00	
	Среднедушевые доходы	18262	18262	0,00	0,00	
	Ожидаемая продолжительность жизни	67,57	67,57	0,00	0,00	
6	Иркутская область	0,794293				
	ВРП на душу населения	442342,20	474633,39	32291,19	7,30	
	Среднедушевые доходы	15110	15110	0,00	0,00	
	Ожидаемая продолжительность жизни	65,26	66,69767	1,43767	2,20	
7	Кемеровская область	0,758445				
	ВРП на душу населения	318927,50	353293,29	34365,79	10,78	
	Среднедушевые доходы	15341	15341	0,00	0,00	
	Ожидаемая продолжительность жизни	65,42	66,398394	0,978394	1,50	
8	Новосибирская область	0,999976				
	ВРП на душу населения	377812,40	398619,50	20807,10	5,51	

Окончание табл. 2.3

№ п/п	Регион	Показатели эффективности				
		Входы/Выходы	Значения показателей (исходные)	Значения показателей (реко-мендуемые)	Разность абсо-лютная	Разность (в %)
9	Среднедушевые доходы		16276	16276	0,00	0,00
	Ожидаемая продолжительность жизни		69,28	69,28	0,00	0,00
	Омская область		0,958475			
	ВРП на душу населения		314593,50	342988,32	28394,82	9,03
	Среднедушевые доходы		15199	15199	0,00	0,00
10	Ожидаемая продолжительность жизни		68,78	68,78	0,00	0,00
	Томская область		0,999998			
	ВРП на душу населения		445485,90	453131,43	7645,53	1,72
	Среднедушевые доходы		15070	15070	0	0,00
	Ожидаемая продолжительность жизни		68,83	68,83	0	0,00

В рамках данной задачи некоторым регионам необходима организация мероприятий, направленных на повышение текущего уровня продолжительности жизни. Однако для принятия решения необходимо принимать во внимание структурное различие между регионами. К примеру, учитывать, что Республика Тыва структурно развивается по азиатскому типу, когда в случае демографической ситуации при высокой рождаемости наблюдается и высокая смертность. Красноярский край, наоборот, принадлежит структурно к странам европейского типа, когда при низкой рождаемости наблюдается низкая смертность. С одной стороны, даже в случае применения модели с переменным эффектом масштаба можно сказать о некоторой «завышенной» требовательности к регионам азиатского типа. С другой стороны, из полученных результатов видно, что регионы, развивающиеся по европейскому типу, более эффективны с точки зрения обеспечения уровня социально-экономической безопасности.

Реализация этапа «Оценка состояния здоровья населения регионов».

Для оценки состояния здоровья населения использовано три показателя, характеризующие заболеваемость, рождаемость и смертность. В качестве входных показателей был принят показатель заболеваемости на 1000 человек и общий коэффициент смертности населения; в качестве выходного показателя — общий коэффициент рождаемости.

Так как результат модели предполагает как снижение входных показателей, так и рост выходных, расчеты были проведены в рамках двух моделей, ориентированных и на вход, и на выход. Предполагается, что модель также будет учитывать переменный эффект масштаба ввиду структурной разницы между объектами.

В первую очередь была рассчитана модель, ориентированная на вход. Ранжирование регионов в рамках данной модели представлено в табл. 2.4.

Таблица 2.4

Результаты оценки состояния здоровья населения (выборка регионов, 2019 г.) для *input*-модели

№ п/п	Регионы РФ	Ранг	Эффективность	Коэффициенты, формирующие гипотетический объект
1	Республика Алтай	2	0,989899	0,775862
2	Республика Тыва	1	1	1
3	Республика Хакасия	5	0,765625	0,607759
4	Алтайский край	10	0,695035	0,521552
5	Красноярский край	4	0,784	0,599138
6	Иркутская область	8	0,731343	0,637931
7	Кемеровская область	9	0,708647	0,521552
8	Новосибирская область	6	0,753846	0,594828
9	Омская область	7	0,742424	0,573276
10	Томская область	3	0,859649	0,568965

В данной модели эталонным регионом зафиксирована Республика Тыва, что неудивительно, поскольку в данном регионе наблюдается наименьший показатель смертности и заболеваемости одновременно с наибольшим значением показателя рождаемости. Весовые коэффициенты указывают на большую разницу в структуре между эталонным регионом и остальными регионами.

В табл. 2.5 приведено ранжирование регионов по уровню их эффективности.

Среднее значение показателя эффективности равно 0,808518, минимальное значение — 0,695035 при стандартном отклонении, равном 0,1024, указывающем на сравнительно однородную группу регионов. К наименее эффективным регионам относятся Кемеровская область и Алтайский край.

Таблица 2.5

**Регионы СФО, упорядоченные по значению эффективности
(критерий «состояние здоровья»), input-модель**

Ранг	Регион РФ	Эффективность
1	Республика Тыва	1
2	Республика Алтай	0,989899
3	Томская область	0,859649
4	Красноярский край	0,784
5	Республика Хакасия	0,765625
6	Новосибирская область	0,753846
7	Омская область	0,742424
8	Иркутская область	0,731343
9	Кемеровская область	0,708647
10	Алтайский край	0,695035

Рекомендуемые изменения показателей в рамках системы поддержки принятия решений по модели, ориентированной на вход, представлены в табл. 2.6.

Практически все регионы нуждаются в мероприятиях по уменьшению смертности в среднем на 10%. Таким регионам, как Иркутская область и Кемеровская область, в наибольшей степени необходимо снизить смертность (на 17% и 16% соответственно), поскольку за рассматриваемый период наблюдалось заметное отставание указанных субъектов.

Все регионы выборки нуждаются в мероприятиях по снижению заболеваемости в среднем на 19%. В большей степени необходимо снизить заболеваемость в таких регионах, как: Республика Алтай – на 25%, Иркутская область – на 25%, Алтайский край – на 29%.

Расчет результатов по модели, ориентированной на выход, представлен в табл. 2.7.

Таблица 2.6

Рекомендуемые значения показателей по состоянию здоровья населения, input-модель

№ п/п	Регион	Показатели эффективности				Разность абсолютная	Разность (в %)
		Входы/Выходы	Значения показателей (исходные)	Значения показателей (реконструируемые)	Значения показателей (исходные)		
1	Республика Алтай		0,989899				
	Смертность		9,9	9,8224142	-0,0775858	-0,78	
	Хроническая заболеваемость		1667,5	1251,25	-416,25	-24,96	
	Рождаемость		18	18	0,00	0,00	
2	Республика Тыва		1				
	Смертность		9,8	9,8	0,00	0,00	
	Хроническая заболеваемость		1131	1131	0,00	0,00	
	Рождаемость		23,2	23,2	0,00	0,00	
3	Республика Хакасия		0,765625				
	Смертность		12,8	10,97672	-1,82328	-14,24	
	Хроническая заболеваемость		1623,3	1324,1	-299,2	-18,43	
	Рождаемость		14,1	14,1	0,00	0,00	
4	Алтайский край		0,695035				
	Смертность		14,1	11,85733	-2,24267	-15,91	

Продолжение табл. 2.6

№ п/п	Регион	Показатели эффективности			
		Значения показателей (исходные)	Значения показателей (реконструируемые)	Разность абсолютная	Разность (в %)
	Хроническая заболеваемость	2539,4	1804,847	-734,553	-28,93
	Рождаемость	12,1	12,1	0,00	0,00
5	Красноярский край	0,784			
	Смертность	12,5	10,88233	-1,61767	-12,94
	Хроническая заболеваемость	1638,9	1334,598	-304,302	-18,57
	Рождаемость	13,9	13,9	0,00	0,00
6	Иркутская область	0,731343			
	Смертность	13,4	11,10345	-2,29655	-17,14
	Хроническая заболеваемость	1875,9	1400,705	-475,195	-25,33
	Рождаемость	14,8	14,8	0,00	0,00
7	Кемеровская область	0,708647			
	Смертность	14,3	11,95302	-2,34698	-16,41
	Хроническая заболеваемость	1596	1353,478	-242,522	-15,20
	Рождаемость	12,1	12,1	0,00	0,00
8	Новосибирская область	0,753846			
	Смертность	13	11,09655	-1,90345	-14,64

Окончание табл. 2.6

№ п/п	Регион	Показатели эффективности			
		Значения показателей (исходные)	Значения показателей (реконструируемые)	Разность абсолютная	Разность (в %)
	Хроническая заболеваемость	1674,6	1351,252	-323,348	-19,31
	Рождаемость	13,8	13,8	0,00	0,00
9	Омская область	0,742424			
	Смертность	13,2	11,25086	-1,94914	-14,77
	Хроническая заболеваемость	1850,9	1438,199	-412,701	-22,30
	Рождаемость	13,3	13,3	0,00	0,00
10	Томская область	0,859649			
	Смертность	11,4	10,489656	-0,910344	-7,99
	Хроническая заболеваемость	1442	1265,052	-176,948	-12,27
	Рождаемость	13,2	13,2	0,00	0,00

Так же как и при использовании модели, ориентированной на вход, только Республика Тыва является эффективным регионом. При этом следует напомнить, что при изменении ориентации модели множество эффективных объектов остаются теми же, однако изменяются значения показателей эффективности для неэффективных объектов.

Это объясняется тем, что граница эффективности остается одной и той же при ориентации модели как на вход, так и на выход. Ориентация влияет только на направление проецирования неэффективных точек на эту границу.

Таблица 2.7

**Результаты оценки состояния здоровья населения
для *output*-модели (выборка регионов, 2019 г.)**

№ п/п	Регионы РФ	Ранг	Эффективность	Коэффициенты, формирующие гипотетический объект
1	Республика Алтай	2	0,775862	0,775862
2	Республика Тыва	1	1	1
3	Республика Хакасия	4	0,607759	0,607759
4	Алтайский край	10	0,521552	0,521552
5	Красноярский край	5	0,599138	0,599138
6	Иркутская область	3	0,637931	0,637931
7	Кемеровская область	9	0,521552	0,521552
8	Новосибирская область	6	0,594828	0,594828
9	Омская область	7	0,573276	0,573276
10	Томская область	8	0,568965	0,568965

Расстояния от одной и той же неэффективной точки до границы эффективности, измеренные в направлении изменения значения входных показателей и выходных показателей, в общем случае различаются. В связи с этим наблюдаются различия в значениях показателя эффективности для одной

и той же точки в зависимости от выбранной ориентации модели *DEA*.

Данная модель (*output*) отличается уменьшением значений показателей эффективности, указывая на существенную разницу между эталонным регионом и остальными субъектами по показателю рождаемости.

В табл. 2.8 представлено ранжирование регионов по данной модели. Среднее значение показателя эффективности составляет 0,644755, минимальное значение — 0,521552 при стандартном отклонении, равном 0,133037, указывающем на сравнительно разнородную группу регионов. В категорию наименее эффективных регионов, как и в рамках предыдущей модели, вошли Кемеровская область и Алтайский край.

Таблица 2.8

Регионы СФО, упорядоченные по значению состояния здоровья, *output*-модель

Ранг	Регион РФ	Эффективность
1	Республика Тыва	1
2	Республика Алтай	0,775862
3	Иркутская область	0,637931
4	Республика Хакасия	0,607759
5	Красноярский край	0,599138
6	Новосибирская область	0,594828
7	Омская область	0,573276
8	Томская область	0,568965
9	Кемеровская область	0,521552
10	Алтайский край	0,521552

Следуя результатам, полученным в рамках первой и второй моделей, можно сказать, что, как и в случае оценки социально-экономической безопасности, эталонный регион работает на много эффективней остальных. В данной модели столь низкое

значение эффективности объясняется фактическими исходными данными. Республика Тыва имеет уровень рождаемости, превышающий средний практически в два раза, при этом самую низкую смертность и заболеваемость. Отсюда происходит столь сильная дифференциация по показателю структуры значений.

Рекомендуемые изменения показателей в рамках системы поддержки принятия решений по модели, ориентированной на выход, представлены в табл. 2.9.

В среднем регионах (в рамках выборки) требуется повышение уровня рождаемости на 63%. Такой высокий показатель изменения объясняется тем, что эталонный регион (Республика Тыва) имеет сравнительно высокий уровень рождаемости, превышающий среднее значение на 43%.

В данной модели (согласно принципу Парето-Кумпанса) практически все регионы нуждаются в снижении уровня заболеваемости (в среднем на 10%). Кроме этого, для Кемеровской области рекомендовано снижение показателя смертности.

Реализация этапа «Оценка эффективности функционирования сферы здравоохранения регионов».

Для оценки функционирования сферы здравоохранения планируется к расчету четыре показателя. Показатель заболеваемости, как и в предыдущей модели, был принят в качестве входного параметра. Также в качестве входных показателей будем использовать следующие: число больничных коек на 10 000 человек населения (характеризует текущее содержание и капитальные вложения в отрасль); численность врачей на 10 000 человек населения; численность среднего медицинского персонала на 10 000 человек населения.

В качестве выходного показателя будем использовать показатель ожидаемой продолжительности жизни при рождении. Этот показатель в интегральной форме характеризует эффективность работы системы здравоохранения. Следует добавить, что этот показатель зависит не только от эффективности работы данной сферы, но также от уровня развития экономики и экологических показателей.

Таблица 2.9

Рекомендуемые значения показателей по состоянию здоровья населения, *outrif*-модель

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Входы/Выходы	Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная	
1	Республика Алтай		0,775862			
	Смертность		9,9	9,9	0,00	0,00
	Хроническая заболеваемость		1667,5	1147,843	-519,657	-31,16
	Рождаемость		18	23,2	5,2	28,89
2	Республика Тыва		1			
	Смертность		9,8	9,8	0,00	0,00
	Хроническая заболеваемость		1131	1131	0,00	0,00
	Рождаемость		23,2	23,2	0,00	0,00
3	Республика Хакасия		0,607759			
	Смертность		12,8	12,8	0,00	0,00
	Хроническая заболеваемость		1623,3	1511,461	-111,839	-6,89
	Рождаемость		14,1	23,2	9,1	64,54
4	Алтайский край		0,521552			
	Смертность		14,1	14,1	0,00	0,00

Продолжение табл. 2.9

№ п/п	Регион	Показатели эффективности				
		Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная	Разность (в %)	
	Хроническая заболеваемость	2539,4	1905,427	-633,973	-24,97	
	Рождаемость	12,1	23,2	11,1	91,74	
5	Красноярский край	0,599138				
	Смертность	12,5	12,5	0,00	0,00	
	Хроническая заболеваемость	1638,9	1485,002	-153,898	-9,39	
	Рождаемость	13,9	23,2	9,3	66,91	
6	Иркутская область	0,637931				
	Смертность	13,4	13,4	0,00	0,00	
	Хроническая заболеваемость	1875,9	1634,973	-240,927	-12,84	
	Рождаемость	14,8	23,2	8,4	56,76	
7	Кемеровская область	0,521552				
	Смертность	14,3	13,966353	-0,333647	-2,33	
	Хроническая заболеваемость	1596	1596	0,00	0,00	
	Рождаемость	12,1	23,2	11,1	91,74	
8	Новосибирская область	0,594828				
	Смертность	13	13	0,00	0,00	

Окончание табл. 2.9

№ п/п	Регион	Показатели эффективности				
		Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная	Разность (в %)	
9	Хроническая заболеваемость	1674,6	1543,209	-131,391	-7,85	
	Рождаемость	13,8	23,2	9,4	68,12%	
	Томская область	0,568965				
	Смертность	11,4	11,4	0,00	0,00	
10	Хроническая заболеваемость	1442	1333,386	-108,614	-7,53	
	Рождаемость	13,2	23,2	10	75,76	
	Омская область	0,573276				
	Смертность	13,2	13,2	0,00	0,00	
	Хроническая заболеваемость	1850,9	1607,747	-243,153	-13,14	
	Рождаемость	13,3	23,2	9,9	74,44	

В рамках данного этапа реализация расчетов проведена на основе использования двух моделей: 1) модель, ориентированная на вход; 2) модель, ориентированная на выход.

В табл. 2.10 представлены результаты расчета по модели, ориентированной на вход.

В данной модели наблюдается несколько эффективных регионов, формирующих границу эффективности, а именно: Красноярский край, Томская область, Омская область, Республика Хакасия. Такая граница позволяет более корректно сформировать гипотетические объекты и выявить структурное сходство регионов. Так, для Красноярского края структурно схожей является Республика Алтай, в то время как для Новосибирской области эталонной будет являться Томская область. Следуя логике формирования данной модели, в табл. 2.11 представлено ранжирование регионов по степени эффективности функционирования сферы здравоохранения.

Среднее значение показателя эффективности составляет 0,924137, минимальное значение — 0,745614 при стандартном отклонении, равном 0,089533, указывающем на сравнительно однородную группу регионов. В группе наименее эффективных регионов — Республика Алтай и Республика Тыва.

Рекомендуемые изменения для показателей сферы здравоохранения представлены в табл. 2.12.

В ряде регионов наблюдается неэффективность капитальных вложений в здравоохранение, для таких регионов рекомендовано оптимизировать данный показатель. Такими регионами выступают: Республика Тыва — на 11%, Иркутская область — на 20%, Кемеровская область — на 11%.

В Иркутской области и Кемеровской области наблюдается неэффективная работа врачей. В соответствии с данной моделью предлагается оптимизация указанного показателя в среднем на 2%. Однако это не означает, что необходимо сокращение врачебного персонала, целесообразно выяснить причины неэффективной работы.

Таблица 2.10

**Результаты оценки функционирования сферы здравоохранения (выборка регионов, 2019 г.),
input-модель**

№ п/п	Регион	Ранг	Эффек- тивность	Коэффициенты, формирующие гипотетический объект (эталонные регионы)			
				Красноярский край	Томская область	Омская область	Республика Хакасия
1	Республика Алтай	9	0,938869	0,86407	0,21107	0,55547	—
2	Республика Тыва	10	0,745614		0,89604	—	0,00000116
3	Республика Хакасия	1	1		—	—	1
4	Алтайский край	1	1		0,60515	—	0,394852
5	Красноярский край	1	1	1		—	0
6	Иркутская область	8	0,980066		0,37164	—	0,599576
7	Кемеровская область	7	0,987306		0,16723	—	0,680557
8	Новосибирская об- ласть	1	1		0,85450	—	0,143752
9	Омская область	1	1			1	0
10	Томская область	1	1		1	—	0

Таблица 2.11

**Регионы СФО, упорядоченные по эффективности
функционирования сферы здравоохранения,
input-модель**

Ранг	Регион	Эффективность
1	Красноярский край	1
1	Томская область	1
1	Омская область	1
1	Республика Хакасия	1
5	Новосибирская область	1
6	Алтайский край	1
7	Кемеровская область	0,987306
8	Иркутская область	0,980066
9	Республика Алтай	0,938869
10	Республика Тыва	0,745614

Одна из возможных причин может заключаться в недостаточном техническом оснащении медицинских учреждений, особенно это касается фельдшерско-акушерских пунктов.

Для Республики Тыва и Республики Алтай предлагается также оптимизация среднего медицинского персонала. При недостатке эффективности (особенно в отношении Республики Тыва) указанные регионы демонстрируют наибольшее значение данного показателя: для Республики Алтай – 121,8, для Республики Тыва – 143,1. В данном контексте можно говорить о недостатке квалифицированного образования у персонала в медицинских учреждениях.

Результаты расчетов в рамках модели, ориентированной на выход, представлены в табл. 2.13.

Таблица 2.12

**Рекомендуемые значения показателей эффективности функционирования сферы
здравоохранения, *input-модель***

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Значения показателей (исходные)	Значения по- казателей (ре- комендуемые)	Разность абсо- лютная		
1	Республика Алтай	0,99063				
	Обеспеченность койками на 10 000 нас.	74,9				
	Обеспеченность врачами на 10 000 нас.	39,4	39,4	0,00		
	Обеспеченность средним медперсо- налом на 10 000 нас.	121,8	103,2849	-18,5151		-15,20
	Ожидаемая продолжительность жизни	70,13				
2	Республика Тыва	0,896037				
	Обеспеченность койками на 10 000 нас.	108,3	96,45	-11,85		-10,94
	Обеспеченность врачами на 10 000 нас.	45,6	45,6	0,00		
	Обеспеченность средним медперсо- налом на 10 000 нас.	143,1	129,3026	-13,7974		-9,64
	Ожидаемая продолжительность жизни	64,21	64,21			
3	Республика Хакасия	1				
	Обеспеченность койками на 10 000 нас.	68,9	68,9	0,00		
	Обеспеченность врачами на 10 000 нас.	34	34	0,00		

Продолжение табл. 2.12

№ п/п	Регион	Показатели эффективности				
		Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная	Разность (в %)	
	Обеспеченность средним медперсоналом на 10000 нас.	92,9	92,9	0,00		
	Ожидаемая продолжительность жизни	69,33				
4	Алтайский край	1				
	Обеспеченность койками на 10 000 нас.	87,4	87,4	0,00		
	Обеспеченность врачами на 10 000 нас.	37,1	37,1	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	98,3	98,3	0,00		
	Ожидаемая продолжительность жизни	70,74				
5	Красноярский край	1				
	Обеспеченность койками на 10 000 нас.	69,7	69,7	0,00		
	Обеспеченность врачами на 10 000 нас.	36,5	36,5	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	95,9	95,9	0,00		
	Ожидаемая продолжительность жизни	70,01				
6	Иркутская область	0,971211				
	Обеспеченность койками на 10 000 нас.	88,9	70,9755	-17,9245	-20,16	

Продолжение табл. 2.12

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная		
	Обеспеченность врачами на 10 000 нас.	36,4	35,15383	-1,24617	-3,42	
	Обеспеченность средним медперсоналом на 10000 нас.	92,3	92,3	0,00		
	Ожидаемая продолжительность жизни	68,2				
7	Кемеровская область	0,982835				
	Обеспеченность койками на 10 000 нас.	79,5	70,62127	-8,87873	-11,17	
	Обеспеченность врачами на 10 000 нас.	35,6	35,06073	-0,53927	-1,51	
	Обеспеченность средним медперсоналом на 10 000 нас.	92,4	92,4	0		
	Ожидаемая продолжительность жизни	68,72				
8	Новосибирская область	1				
	Обеспеченность койками на 10 000 нас.	85,5	85,5	0		
	Обеспеченность врачами на 10 000 нас.	39,6	39,6	0		
	Обеспеченность средним медперсоналом на 10 000 нас.	86,5	86,5	0		
	Ожидаемая продолжительность жизни	71,2				

Окончание табл. 2.12

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная		
9	Омская область	1				
	Обеспеченность койками на 10 000 нас.	72,4	72,4	0,00		
	Обеспеченность врачами на 10 000 нас.	40	40	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	96,9	96,9	0,00		
	Ожидаемая продолжительность жизни	70,78				
10	Томская область	1				
	Обеспеченность койками на 10 000 нас.	87	87	0,00		
	Обеспеченность врачами на 10 000 нас.	40,9	40,9	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	85,6	85,6	0,00		
	Ожидаемая продолжительность жизни	71,66				

Таблица 2.13

**Результаты оценки функционирования сферы здравоохранения (выборка регионов, 2019 г.),
outrup-модель**

№	Регион	Ранг	Эффек- тивность	Коэффициенты, формирующие гипотетический объект (эталонные регионы)			
				Красноярский край	Томская область	Омская область	Республика Хакасия
1	Республика Алтай	7	0,99063	0,22409	0,21107	0,55547	—
2	Республика Тыва	10	0,896037		0,89604	—	—
3	Республика Хакасия	1	1		—	—	1
4	Алтайский край	6	1		0,60515	—	0,39485
5	Красноярский край	1	1	1		—	
6	Иркутская область	9	0,971211		0,37164	—	0,59958
7	Кемеровская область	8	0,982835		0,16723	—	0,68056
8	Новосибирская область	5	1		0,85450	—	0,14375
9	Омская область	1	1			1	
10	Томская область	1	1		1	—	

Наблюдается чуть меньший разброс по показателям эффективности, при этом наблюдается незначительное изменение структуры. Так, для Республики Алтай эталонным регионом выступает уже Омская область.

Ранжирование регионов по уровню эффективности функционирования сферы здравоохранения представлено в табл. 2.14.

Среднее значение показателя эффективности равно 0,950475, минимальное значение — 0,84121 при стандартном отклонении, равном 0,057398, указывающем на сравнительно однородную группу регионов. В группе наименее эффективных регионов находятся Иркутская область и Красноярский край.

Рекомендуемые изменения для показателей сферы здравоохранения по модели *output* представлены в табл. 2.15.

Таблица 2.14

Регионы РФ, упорядоченные по эффективности функционирования сферы здравоохранения, *output*-модель

Ранг	Регион	Эффективность
1	Красноярский край	1
1	Томская область	1
1	Омская область	1
1	Республика Хакасия	1
5	Новосибирская область	1
6	Алтайский край	1
7	Республика Алтай	0,99063
8	Кемеровская область	0,982835
9	Иркутская область	0,971211
10	Республика Тыва	0,896037

Таблица 2.15

**Рекомендуемые значения показателей эффективности функционирования сферы
здравоохранения, *outrifit*-модель**

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Входы/Выходы	Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная	
1	Республика Алтай		0,99063			
	Обеспеченность койками на 10 000 нас.		74,9	74,9	0,00	
	Обеспеченность врачами на 10 000 нас.		39,4	39,4	0,00	
	Обеспеченность средним медперсоналом на 10 000 нас.		121,8	94,5241	-27,2759	-22
2	Ожидаемая продолжительность жизни		70,13			
	Республика Тыва		0,896037			
	Обеспеченность койками на 10 000 нас.		108,3	89,2144	-19,0856	-18
	Обеспеченность врачами на 10 000 нас.		45,6	41,38862	-4,21138	-9
	Обеспеченность средним медперсоналом на 10 000 нас.		143,1	91,5779	-51,5221	-36
3	Ожидаемая продолжительность жизни		64,21	69,33	5,12	8
	Республика Хакасия		1			
	Обеспеченность койками на 10 000 нас.		68,9	68,9	0,00	
	Обеспеченность врачами на 10 000 нас.		34	34	0,00	

Продолжение табл. 2.15

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная	Разность	
	Обеспеченность средним медперсоналом на 10 000 нас.	92,9	92,9	0,00		
	Ожидаемая продолжительность жизни	69,33				
4	Алтайский край	1				
	Обеспеченность койками на 10 000 нас.	87,4	79,85318	-7,54682		-9
	Обеспеченность врачами на 10 000 нас.	37,1	37,1	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	98,3	88,48242	-9,81758		-10
	Ожидаемая продолжительность жизни	70,74				
5	Красноярский край	1				
	Обеспеченность койками на 10 000 нас.	69,7	69,7	0,00		0
	Обеспеченность врачами на 10 000 нас.	36,5	36,5	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	95,9	95,9	0,00		
	Ожидаемая продолжительность жизни	70,01				
6	Иркутская область	0,971211				
	Обеспеченность койками на 10 000 нас.	88,9	76,2024	-12,6976		-14

Продолжение табл. 2.15

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная	Разность	
	Обеспеченность врачами на 10 000 нас.	36,4	36,4	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	92,3	90,16979	-2,13021		-2
	Ожидаемая продолжительность жизни	68,2				
7	Кемеровская область	0,982835				
	Обеспеченность койками на 10 000 нас.	79,5	74,60721	-4,89279		-6
	Обеспеченность врачами на 10 000 нас.	35,6	35,6	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	92,4	92,4			
	Ожидаемая продолжительность жизни	68,72				
8	Новосибирская область	1				
	Обеспеченность койками на 10 000 нас.	85,5	84,24625	-1,25375		-1
	Обеспеченность врачами на 10 000 нас.	39,6	39,6	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	86,5	86,5	0,00		
	Ожидаемая продолжительность жизни	71,2				

Окончание табл. 2.15

№ п/п	Регион	Показатели эффективности				Разность (в %)
		Значения показателей (исходные)	Значения показателей (рекомендуемые)	Разность абсолютная		
9	Омская область	1				
	Обеспеченность койками на 10 000 нас.	72,4	72,4	0,00		
	Обеспеченность врачами на 10 000 нас.	40	40	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	96,9	96,9	0,00		
	Ожидаемая продолжительность жизни	70,78				
10	Томская область	1				
	Обеспеченность койками на 10 000 нас.	87	87	0,00		
	Обеспеченность врачами на 10 000 нас.	40,9	40,9	0,00		
	Обеспеченность средним медперсоналом на 10 000 нас.	85,6	85,6	0,00		
	Ожидаемая продолжительность жизни	71,66				

Так же как и в модели сокращения издержек, наблюдается неэффективность капитальных вложений в сферу здравоохранения. В данном случае коррекция таких показателей необходима для выхода на производственную границу в соответствии с принципом Парето-Кумпанса. Корректировка необходима для следующих субъектов: Республика Тыва — на 18%, Алтайском край — на 9%, Иркутская область — на 14%, Кемеровская область — на 6%, Новосибирская область — на 1%.

В Республике Алтай, Республике Тыва и Алтайском крае необходима корректировка по количеству среднего медицинского персонала (на 22%, 36% и 10% соответственно). В Республике Тыва требуется небольшая корректировка по показателю обеспеченности врачами (+9%), а также необходимо увеличение качества оказываемой помощи (+8%). Данный регион нуждается в самой большой корректировке показателей, поскольку иллюстрирует наименьшее значение эффективности сферы здравоохранения.

По итогу результатов анализа, полученных в рамках реализации двух моделей, можно отметить, что в большинстве регионов сфера здравоохранения функционирует на высоком уровне. Республика Тыва относится к категории кризисного субъекта, эффективность которой на 18% ниже, чем по округу в целом. Однако угроза может заключаться в том, что статистические данные могут быть завышены и отличаться от реального положения дел в регионе. Поэтому некоторые выводы, возможно, могут быть некорректны ввиду некорректной исходной информации по показателям.

2.3. ЭКСПЕРИМЕНТАЛЬНОЕ ПРИМЕНЕНИЕ МЕТОДА DATA ENVELOPMENT ANALYSIS ДЛЯ РЕШЕНИЯ ЗАДАЧИ ОЦЕНКИ ЭФФЕКТИВНОСТИ РЕГИОНАЛЬНОЙ ИННОВАЦИОННОЙ ПОЛИТИКИ

Программная реализация процедуры оценки эффективности инновационной политики регионов России на основе

DEA-анализа возможна с использованием программного обеспечения *Open Source DEA (OS DEA)* (<https://opensourcedea.org/>). Данная программа предоставляет более 40 модификаций *DEA*-модели. Модификации зависят от ориентированности модели (вход / выход / неориентированная), вида эффективности и отдачи от масштаба производства.

Анализ данных с использованием программы *OS DEA* включает следующие этапы.

1. Подготовка датасета. Датасет должен представлять с собой таблицу в формате *comma-separated values (CSV)*.

DMU (объекты анализа) должны быть по строкам, а входные и выходные факторы — по столбцам. Лишь немногие разновидности *DEA*-модели допускают отрицательные значения в датасете. Также нежелательны нулевые значения. Суммарное количество входных и выходных переменных должно быть намного меньше количества *DMU*, иначе расчет эффективности будет некорректным. Некоторые модификации модели, такие как *SBM*-модели (*Slack based models*), устраняют это ограничение [48]. Так же как и при любом статистическом анализе, рекомендуется убирать пиковые значения в датасете.

2. Выбор входных и выходных переменных.

Помимо *input*- и *output*-переменных (входные и выходные), *OS DEA* предлагает *non-discretionary inputs* и *non-discretionary outputs* (неконтролируемые входы и неконтролируемые выходы). Это входящие или исходящие переменные, которые можно включить в модель, однако отсутствует возможность их изменить. Следовательно, влияние входных переменных на выходные сохраняется, но для достижения эффективной границы неконтролируемые входные (в случае ориентированности на вход) и неконтролируемые выходные (в случае ориентированности на выход) не изменяются. Таким образом, предоставляется возможность учитывать факторы влияния окружающей среды [37].

3. Выбор модели.

Выбор модели *DEA* в *Open Source DEA* производится по трем параметрам: 1) *ориентация*; 2) *эффективность*; 3) *отдача от масштаба*.

Первый параметр предполагает три варианта выбора: 1) модель, ориентированная на вход; 2) модель, ориентированная на выход; 3) неориентированная модель. При ориентации на вход для достижения эффективности изменяются входные переменные, на выход — выходные, а при отсутствии ориентации можно изменять как входные, так и выходные переменные (здесь обращаем внимание, что неконтролируемые переменные мы изменять не можем).

Второй параметр предполагает два варианта: техническая эффективность (*technical efficiency*) и смешанная эффективность (*mix efficiency*). Техническая эффективность — это центральное понятие *DEA*-анализа. Техническая эффективность подразумевает максимальный результат при неизменных затратах (модели, ориентированные на выход) или минимальные затраты при неизменном результате (модели, ориентированные на вход).

Смешанную эффективность рассчитывают только *SBM*-модели. Значения смешанной эффективности не ограничиваются нулевым и единичным значениями. *SBM*-модели изменяют не переменные, а прогибы (*Slacks*), возникающие не вне, а на границе эффективности [49].

Последний параметр модели, который следует задать, — это отдача от масштаба (*Returns to Scale*). Постоянная отдача от масштаба — это изменение выходных переменных на тот же коэффициент, что и входных, возрастающая отдача от масштаба — изменение выходных переменных с большим коэффициентом, чем выходных, убывающая — с меньшим коэффициентом. Когда закономерность между изменением входных и выходных переменных отсутствует, выбирают переменную (*variable*) отдачу от масштаба.

После того как экспортирован датасет, выбраны входные и выходные переменные и заданы параметры модели, *Open Source DEA* предоставляет решение задачи линейного программирования, которое можно импортировать в *Excel*. Решение задачи на отдельных листах отображает параметры модели, исходный датасет, переменные, объекты, проекции, лямбда, схожие группы, прогибы и веса.

Таким образом, технология *DEA*-моделирования позволяет сформировать конкретные оценки желательных изменений во входах/выходах, которые предоставляют возможность вывести неэффективные регионы на так называемую границу эффективности.

С другой стороны, проблема использования данного метода может заключаться в том, что статистические данные могут быть завышены и отличаться от реальной ситуации в регионе. В связи с этим некоторые выводы, возможно, могут быть некорректны ввиду некорректной исходной информации по статистическим показателям.

По мнению авторов, метод *DEA* целесообразно использовать в комплексе с другими методами оценки, например с методом кластеризации *k*-средних, алгоритм которого позволяет разбить множество объектов на группы, схожие по свойствам на основании вычисления расстояния до центра кластера.

Кластеризация позволяет разделить множество объектов на группы (кластеры) схожих по каким-либо признакам объектов [50]. В отличие от классификации, при кластеризации заранее не известно количество групп. Кластеризацию предлагается проводить одним из наиболее распространенных методов — методом *k*-средних. Его суть состоит в том, что вначале множество объектов разбивается на кластеры произвольным образом (объекты входят в кластер со случайно выбранным центром по принципу минимального евклидова расстояния), затем центры кластеров пересчитываются и осуществляется новое разбиение на кластеры. Изменение центров

кластеров и новые разбиения продолжаются до тех пор, пока центры кластеров не перестанут меняться.

Количество кластеров при использовании метода k -средних может определяться разными способами. Один из них — это реализация метода с несколькими вариантами количества кластеров и выбор наиболее отвечающего целям исследования варианта.

После того как выбрано количество кластеров, случайным образом определяются их центры. Центром может выступить как один из объектов, так и случайная точка в пространстве. Прочие объекты объединяются в кластеры путем выбора минимального евклидова расстояния до одного из центров в соответствии с формулой

$$d_{j-c} = \min\left(\sqrt{\sum_{i=1}^n (x_{ij} - x_{ic})^2}\right), \quad (2.17)$$

где d_{j-c} — расстояние от j -го объекта до центра кластера; n — количество признаков (измерений); x_{ij} — координата j -го объекта; x_{ic} — координата центра кластера.

После первой итерации координаты центры кластеров пересчитываются в соответствии с формулой

$$x_{ic} = \frac{\sum_{l=1}^L x_{il}}{L}, \quad (2.18)$$

где L — количество объектов в кластере; x_{ic} — i -я координата центра кластера; x_{il} — i -я координата l -го объекта, входящего в кластер.

Апробация предложенного оценочного инструментария произведена на примере регионов Российской Федерации (с последующей детализацией результатов расчетов на примере субъектов Сибирского федерального округа). Исследуемый период — 2019, 2020 годы; информационная база —

открытые статистические данные Федеральной службы государственной статистики Российской Федерации.

Для проведения процедуры анализа данных с использованием программы *Open Source DEA* последовательно реализуем следующие этапы.

Этап I. Отбор входных и выходных параметров, подготовка датасета.

Формирование входных и выходных параметров опирается на результаты теоретических исследований в аспекте формирования системы оценочных индикаторов эффективности реализации инновационной политики с учетом факторов их определяющих [51]. Исходя из анализа теоретических подходов к классификации условий инновационного развития, факторы влияния обобщены и классифицированы на 2 группы: 1) факторы, влияющие на инновационный потенциал региона; 2) факторы, влияющие на результаты инновационной деятельности. Исходя из этого, сформирована система показателей оценки инновационного развития с учетом установленных пороговых значений.

Используя данную систему оценивания [51], произведена выборка выходных переменных (результаты инновационной деятельности). Предварительно используем следующие показатели:

- объем инновационных товаров, работ, услуг (в процентах от общего объема отгруженных товаров, выполненных работ, услуг) – Y_1 ;
- число выданных патентов на 10 000 человек населения – Y_2 .

В качестве входных переменных (инновационный потенциал) предварительно будем использовать:

- степень износа основных фондов, % (с учетом установленного порогового значения не более 40) – X_1 ;
- доля персонала, занятого научными исследованиями и разработками в среднегодовой численности занятых в экономике, % (с учетом установленного порогового значения не менее 5) – X_2 ;

- доля внутренних затрат на научные исследования и разработки в объеме ВРП, % (с учетом установленного порогового значения не менее 2) – X_3 ;
- удельный вес предприятий и организаций, осуществляющих технологические инновации, в общем числе предприятий и организаций, % (с учетом установленного порогового значения не менее 30) – X_4 .

Поскольку все входные показатели, кроме первого, не должны уменьшаться для достижения оптимума, мы возьмем не их фактическое значение, а расстояние от фактического значения до порогового значения [51]. Если фактическое значение превышает пороговое, то будем считать показатель, равным нулевому значению.

Далее необходимо проверить входные и выходные переменные на взаимную корреляцию. В случае сильной парной корреляции можно будет снизить размерность датасета, убрав одну из связанных переменных. Также возможно убрать входные переменные, которые оказывают слабое влияние на выходные. Для расчета парных коэффициентов корреляции производится процедура нормирования каждой переменной. Визуализация корреляции входных и выходных переменных представлена на рис. 2.4.

На рис. 2.4 наблюдается сильная обратная зависимость между обеими выходными переменными и переменной X_4 . Более слабую обратную зависимость (около 0,5) переменная Y_1 показывает от переменных X_2 и X_3 . Обратная зависимость вполне логична – входные параметры направлены на уменьшение. Однако наблюдается слабая прямая зависимость переменной Y_1 от входной переменной X_1 , а переменная Y_2 демонстрирует слабую обратную зависимость. Отсюда можно сделать вывод, что показатель «Износ основных фондов» не оказывает влияния на выходные переменные. Следовательно, можно убрать данную переменную из датасета. Также наблюдается единичная корреляция между входными пе-

ременными $X3$ и $X2$, что позволяет исключить одну из них из датасета.

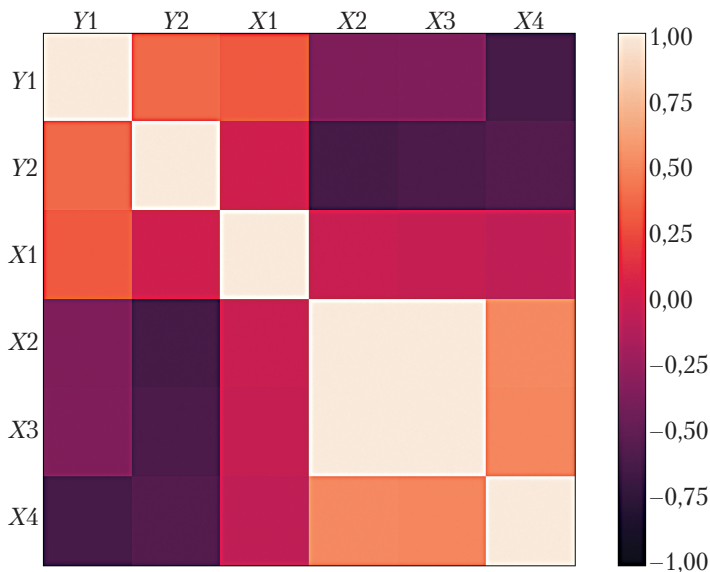


Рис. 2.4. Визуализация корреляционной матрицы

В подтверждение сформированных выводов целесообразно проверить коэффициенты парной корреляции на значимость. Для такой проверки необходима табличная t -статистика Стьюдента для 83 степеней свободы с уровнем значимости 95% ($t_{0,05;83} = 1,98861$). Расчетная t -статистика определяется в соответствии с формулой

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad (2.19)$$

где r — коэффициент корреляции; n — количество наблюдений.

Если расчетная t -статистика по модулю превышает табличную, то коэффициент корреляции значим. В табл. 2.16

представлены расчетные t -статистики по каждой паре переменных.

Таблица 2.16

Расчетные значения t -статистики по каждой паре переменных

	Y1	Y2	X1	X2	X3	X4
Y1						
Y2	3,66					
X1	2,96	0,26				
X2	-3,47	-7,30	-0,13			
X3	-3,40	-6,76	-0,27	8,40		
X4	-7,19	-6,14	-0,54	5,41	5,32	

Незначима корреляция между переменной $X1$ и тремя другими переменными датасета. Однако переменная $X1$ была исключена ранее. Следовательно, необходимо определить, какую из переменных, $X2$ или $X3$, мы исключим. Зависимость переменной $Y2$ от $X3$ чуть слабее, чем от переменной $X2$ (см. рис. 2.4). Также в исходном датасете переменная $X3$ показывает больше нулевых значений. Поэтому сделан вывод об исключении переменной $X3$.

В результате датасет для *DEA*-анализа представлен двумя входными и двумя выходными переменными.

Входные переменные: доля персонала, занятого научными исследованиями и разработками в среднегодовой численности занятых в экономике, % ($X2$); удельный вес организаций, осуществляющих технологические инновации, в общем числе предприятий и организаций, % ($X4$).

Выходные переменные: объем инновационных товаров, работ, услуг (в процентах от общего объема отгруженных товаров, выполненных работ, услуг) ($Y1$); число выданных патентов на 10 000 чел. населения ($Y2$).

В формируемой модели не будут присутствовать неконтролируемые переменные, так как поставленная исследова-

нием задача предполагает разработку рекомендаций для отдельных регионов (а именно для субъектов Сибирского федерального округа) по изменению входных факторов.

Этап II. Выбор модели *DEA*.

Ориентация модели и тип эффективности выбирается исходя из поставленной цели исследования. В нашем случае будет использована *DEA*-модель, ориентированная на вход и определяющая техническую эффективность.

Для выбора последнего параметра модели — типа отдачи от масштаба необходимо провести анализ темпов роста показателей, характеризующих входные и выходные переменные модели в период с 2010 по 2019 г. (допущение: используется среднероссийское значение). В результате сделан вывод об отсутствии взаимосвязи между коэффициентами изменения входных и выходных переменных. Отсутствие постоянной эластичности выходных показателей по отношению к входным показателям означает, что для выбора *DEA*-модели подходит переменная отдача от масштаба.

Таким образом, в нашем случае *DEA*-модель будет ориентирована на вход, считает техническую эффективность, предполагает переменную отдачу от масштаба. Такая модель без неконтролируемых переменных только одна: *The Banker Charnes and Cooper Model* — классическая *Charnes and Cooper Model (CCR_D)*, но с переменной отдачей от масштаба.

Этап III. Интерпретация решения модели.

Заявленная цель исследования ориентирована на оценку и анализ эффективности инновационной политики регионов Сибирского федерального округа. Однако осуществлять *DEA*-анализ по представленному выше алгоритму для отдельной выборки регионов (в нашем случае для субъектов СФО) не целесообразно. Данный тезис объясняется наличием в составе округа всего 10 регионов. Для получения адекватных результатов массив *DMU* должен быть объемным и количество *DMU* должно в несколько раз превышать количество факторов. В связи с этим *DEA*-анализ произведен для всей совокупности

регионов России (85 субъектов). В результате *DEA*-анализа рассчитаны значения показателей эффективности регионов за 2019 и 2020 год. Данные показатели с учетом эталонных регионов представлены в табл. 2.17 в порядке убывания.

Таблица 2.17

**Показатели эффективности регионов Российской Федерации
(фрагмент интерпретации решения *DEA*-модели)**

№ п/п	Субъект Российской Федерации	Ценность объекта (эффективность)	Регионы-эталоны
Результаты оценки за 2019 год			
1	г. Москва	1,00	г. Москва
2	Нижегородская область	1,00	Нижегородская область
3	Республика Мордовия	1,00	Республика Мордовия
4	г. Санкт-Петербург	1,00	г. Санкт-Петербург
5	Республика Татарстан	0,81	г. Санкт-Петербург, Республика Мордовия, Нижегородская область
6	Томская область	0,73	г. Санкт-Петербург, Нижегородская область
...
80	Костромская область	0,12	г. Санкт-Петербург, Нижегородская область
81	Республика Калмыкия	0,11	г. Санкт-Петербург, Нижегородская область
82	Ненецкий автономный округ	0,10	г. Санкт-Петербург, Нижегородская область
83	Республика Северная Осетия	0,10	г. Санкт-Петербург, Нижегородская область

Окончание табл. 2.17

№ п/п	Субъект Российской Федерации	Ценность объекта (эффективность)	Регионы-эталоны
84	Республика Дагестан	0,10	г. Санкт-Петербург, Нижегородская область
85	Чеченская Республика	0,09	г. Санкт-Петербург, Нижегородская область
Результаты оценки за 2020 год			
1	Республика Татарстан	1,00	Республика Татарстан
2	Нижегородская область	1,00	Нижегородская область
3	г. Санкт-Петербург	1,00	г. Санкт-Петербург
4	Республика Мордовия	1,00	Республика Мордовия
5	г. Москва	1,00	г. Москва
6	Хабаровский край	0,80	Республика Мордовия, Нижегородская область
...
80	Чеченская Республика	0,41	г. Москва
81	Республика Хакасия	0,41	г. Москва
82	Псковская область	0,41	г. Москва
83	Ямало-Ненецкий автономный округ	0,41	г. Москва
84	Чукотский автономный округ	0,40	г. Москва
85	Еврейская автономная область	0,40	г. Москва

В соответствии с информацией, представленной в табл. 2.17, только 5 регионов эффективны с точки зрения реализации инновационной политики (согласно выходным и входным пара-

метрам) — это г. Москва, Нижегородская область, Республика Мордовия, Республика Татарстан и г. Санкт-Петербург. 27 регионов показывают эффективность выше 0,5. Для большинства регионов эталонными объектами выступают город федерального значения Москва и Нижегородская область.

Опираясь на значения эффективности и эталонные объекты для регионов Сибирского федерального округа (табл. 2.18), продемонстрируем прочие показатели результатов *DEA*-анализа.

Таблица 2.18

Показатели эффективности регионов Сибирского федерального округа (интерпретация решения *DEA*-модели)

Субъект СФО	Ценность объекта	Регионы-эталоны	Ценность объекта	Регионы-эталоны
		2019 г.		2020 г.
Республика Алтай	0,12	г. Санкт-Петербург, Нижегородская область	0,41	г. Москва
Республика Тыва	0,12	г. Санкт-Петербург, Нижегородская область	0,43	г. Москва
Республика Хакасия	0,135	г. Санкт-Петербург, Нижегородская область	0,41	г. Москва
Алтайский край	0,17	г. Санкт-Петербург, Нижегородская область	0,43	г. Москва
Красноярский край	0,17	г. Санкт-Петербург, Нижегородская область	0,47	г. Москва, Нижегородская область
Иркутская область	0,15	г. Санкт-Петербург, Нижегородская область	0,44	г. Москва

Окончание табл. 2.18

Субъект СФО	Ценность объекта	Регионы-эталоны	Ценность объекта	Регионы-эталоны
		2019 г.		2020 г.
Кемеровская область	0,12	г. Санкт-Петербург, Нижегородская область	0,41	г. Москва
Новосибирская область	0,2	г. Санкт-Петербург, Нижегородская область	0,60	Город федерального значения Москва
Омская область	0,2	г. Санкт-Петербург, Нижегородская область	0,54	г. Москва, Нижегородская область
Томская область	0,73	г. Санкт-Петербург, Нижегородская область	0,68	г. Москва

Результаты *DEA*-анализа для регионов Сибирского федерального округа, а именно расстояние от фактического значения фактора до значения, соответствующего эффективной границе, представлены в табл. 2.19.

Основываясь на данных, представленных в табл. 2.19, возможна разработка рекомендаций для повышения уровня эффективности анализируемых регионов.

При использовании *Open Source DEA* на листе «Лямбда» (*Lambdas*) для каждого объекта показаны коэффициенты при эффективных объектах, которые формируют эталонный объект. Координаты эталона, то есть точки на границе эффективности, рассчитываются в соответствии с формулой

$$E = \lambda_1 x_1 + \lambda_2 x_2, \quad (2.20)$$

где λ_1, λ_2 — лямбды регионов-эталонов; x_1, x_2 — фактические значения показателей регионов-эталонов.

Таблица 2.19

Показатели результатов DEA-анализа для субъектов Федерального округа за 2020 год

Регион	Y1		Y2		X2			X4				
	Э	Ф	Р	Э	Ф	Р	Э	Ф	Р	Э	Ф	Р
Республика Алтай	3,6	0	3,6	5,2793	0,227	5,0523	2,02	4,8798	2,8598	0,00	17	17
Республика Тыва	3,6	0	3,6	5,2793	0,091	5,1883	2,02	4,6475	2,6275	0,00	17,7	17,7
Республика Хакасия	3,6	0,1	3,5	5,2793	0,338	4,9413	2,02	4,9522	2,9322	0,00	18,2	18,2
Алтайский край	3,6	2,6	1	5,2793	1,058	4,2213	2,02	4,7391	2,7191	0,00	3,7	3,7
Красноярский край	4,6	4,6	0	4,927	1,639	3,288	2,066	4,393	2,327	0,18	15,1	14,92
Иркутская область	3,6	0,7	2,9	5,2792	0,818	4,4612	2,02	4,6171	2,5971	0,00	11,7	11,7
Кемеровская область	3,6	2,3	1,3	5,2793	0,835	4,4443	2,02	4,9002	2,8802	0,00	15,9	15,9
Новосибирская область	3,6	3,5	0,1	5,2792	2,328	2,9512	2,02	3,3597	1,3396	0,00	12,2	12,2
Омская область	12,6	12,6	0	2,109	1,331	0,778	2,4335	4,5468	2,1132	1,64	4,9	3,264
Томская область	3,6	2,9	0,7	5,2793	3,098	2,1813	2,02	2,9888	0,9688	0,00	0	0

Примечание: Э – эталон; Ф – факт; Р – расстояние.

Рассмотрим в качестве примера Красноярский край. Лямбда по Москве для данного региона равна 0,91, а по Нижегородской области — 0,09. Рассмотрим показатель Y_1 , фактическое значение которого для Москвы составляет 3,6, а для Нижегородской области — 14,6. Следовательно, эталонное значение показателя Y_1 для Красноярского края равно 4,6.

Расстояние между фактическим и эталонным значением определяется как модуль разности между соответствующими показателями. Возвращаясь к примеру Красноярского края, мы видим, что это разница составит 0. Соответственно, переменную Y_1 не следует изменять для достижения эффективной границы. Аналогичная процедура была проделана для каждого показателя и каждого региона.

Отметим, что изначально модель не предполагала изменения выходных переменных, однако по причине «прогибов» вдоль эффективной границы эти изменения придется произвести.

Как было указано ранее, некоторые выводы могут быть некорректными, что обусловлено вероятностью искажения статистических данных. В связи с этим целесообразно использовать метод *DEA* в комплексе с методом кластеризации.

Следуя поставленным задачам, проведем кластеризацию методом *k*-средних с использованием среды программирования *R Studio*, зададим количество кластеров три (с учетом дифференциации на группы «красная зона», «желтая зона», «зеленая зона»). Объектами выступают 85 регионов России. В качестве входных и выходных параметров использованы те же показатели, что и в *DEA*-анализе (X_2 , X_4 , Y_1 , Y_2).

В результате получили разделение регионов по степени эффективности на три группы. Четырехмерное пространство не поддается визуализации, поэтому визуализацию итогов кластеризации произведем отдельно по входным (X_2 , X_4) и выходным (Y_1 , Y_2) параметрам (рис. 2.5).

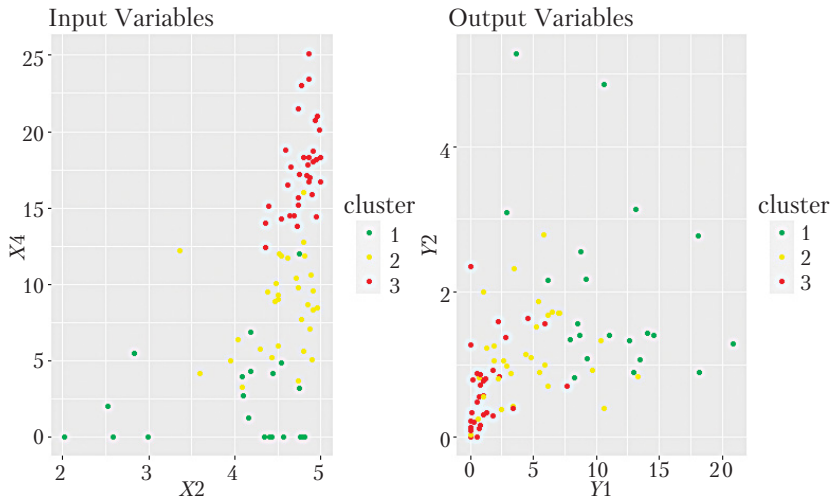


Рис. 2.5. Визуализация итогов кластеризации

По мнению авторов, применение технологии *DEA* целесообразно рекомендовать для оценки эффективности реализации инновационной политики регионов. Результаты оценки могут быть использованы для построения специальной системы рейтингования субъектов Российской Федерации по уровню эффективности.

Система показателей оценки может поддаваться корректировкам (как с позиции их качественной составляющей, так и с позиции количественного состава), исходя из целей стратегического развития регионов и технических возможностей технологии *DEA*. Кроме того, технология *DEA* позволяет сформировать конкретные оценки желательных изменений во входах/выходах, которые позволили бы вывести неэффективные регионы на так называемую границу эффективности.

С другой стороны, проблема использования данного метода может заключаться в том, что статистические данные могут быть завышены и отличаться от реальной ситуации в регионе. В связи с этим некоторые выводы, возможно, могут

быть некорректны ввиду некорректной исходной информации по статистическим показателям.

Таким образом, применение метода опирается на следующую последовательность действий [37]:

1) Формирование требуемых исходных данных. Для реализации метода *DEA* необходимы только лишь значения входных и выходных индикаторов исследуемых объектов. Не требуется задавать весовые значения для указания важности индикаторов (несмотря на то что ряд моделей предоставляют такую возможность). Важной особенностью данного метода является дифференциация индикаторов на входные и выходные, в то время как для других методов такое разделение не производится.

2) Формализация выходных результатов. Результатом работы метода *DEA* будет один интегральный показатель для каждого из изучаемых объектов. Объекты могут быть ранжированы по значениям этого индикатора. Кроме того, производятся конкретные оценки желательных изменений во входах/выходах, которые позволили бы вывести неэффективные объекты на так называемую границу эффективности.

3) Степень влияния человеческого фактора. При необходимости в методе *DEA* можно использовать знания экспертов (ряд моделей данного метода позволяют сделать это), например, для учета относительной важности показателей. Однако применение экспертного знания не является обязательным, что значительно снижает степень субъективизма при проведении исследования. Кроме того, метод *DEA* позволяет учесть влияние факторов окружающей среды.

Глава 3

ЗАДАЧИ КЛАСТЕРИЗАЦИИ И КЛАССИФИКАЦИИ В АНАЛИЗЕ ДАННЫХ

3.1. ОСНОВНЫЕ ПОДХОДЫ К РЕШЕНИЮ ЗАДАЧ ГРУППИРОВКИ ОБЪЕКТОВ И ДАННЫХ

Анализ и классификация подходов, предложенных для решения задач автоматической группировки объектов и данных и кластерного анализа, является трудной задачей ввиду чрезвычайного их разнообразия. В различных подходах используются всевозможные меры сходства, различные целевые функции (минимизация суммарного расстояния между объектами, минимизация суммарного расстояния до центров кластеров, минимизация максимальных расстояний в кластере и т.д.), различные эвристики и схемы локального и глобального поиска, необходимость применения которых обусловлена невыпуклостью целевых функций большинства таких задач [52].

Группы можно определить как области высокой плотности в пространстве признаков (характеристик), разделенные областями низкой плотности. Алгоритмы, непосредственно основанные на этом определении, занимаются поиском связанных областей высокой плотности в пространстве признаков, при этом различные алгоритмы используют различные определения связности. Алгоритм Джарвиса — Патрика и подобные ему алгоритмы определяют подобие между парой точек как число общих соседей, которых они разделяют, где соседи — точки, находящиеся внутри круга заданного радиуса вокруг точки, либо используют метод окна Парзена для определения плотности [53, 54]. Простейшими оценками плотностей можно считать гистограмму и полиграмму первого

и более высокого порядков [55, 56]. Вклад в методы непараметрического оценивания плотности внесли Е. Парзен [57, 58], В.А. Епанечников [59], А.В. Медведев [60]. В работах этих и других авторов вводятся новые классы оценок, обобщающие гистограмму. Так, один из этих классов, называемых «ядерными оценками», был предложен Розенблаттом и Парзенем. В [60–63] решается задача обеспечения несмещенности таких оценок, исследуется скорость сходимости отклонений, изучается влияние формы ядра на качество приближения оценки к функции плотности, а также рассматриваются вопросы определения параметров алгоритмов восстановления плотности — параметра размытости (сглаживания) непараметрических оценок плотности, например с гауссовым ядром. Разработаны принципы построения систем автоматической группировки, в том числе для задач с большим объемом входных данных, позволяющие обрабатывать данные за один проход [54].

В том или ином варианте алгоритмы, использующие методы восстановления плотности, зависят от выбранного масштаба, в котором измерены расстояния, от количества точек, попадающих в окрестность друг друга, достаточного для того, чтобы определить такое скопление точек как группу, и от максимального расстояния, на которое должны быть удалены точки в группе. Выбор перечисленных параметров — отдельная, весьма сложная задача, решением которой определяется точность метода и адекватность модели автоматической группировки. Весьма существенная проблема всех подобных методов — «проклятие размерности» [64], связанное с экспоненциальным увеличением количества данных, необходимых для эффективной работы алгоритмов восстановления плотности, с ростом размерности пространства. Если число признаков велико и все они учитываются при относительно небольшом объеме данных, то все объекты оказываются примерно на одинаковом расстоянии друг от друга, вследствие этого подобные подходы практически неприменимы к многомерным данным.

Тем не менее было предложено множество подходов, оперирующих вероятностными моделями плотности [65–67], таких как модель размещения Патинко [68] и ненаправленная графическая модель [69]. Хотя такие методы также чувствительны к выбору параметров, требовательны к вычислительным ресурсам и имеют ограничения на размерность данных, эти методы, в особенности основанные на непараметрических моделях плотности [67], достаточно популярны из-за их способности выделять группы (кластеры) произвольной формы. Если пространство признаков имеет очень высокую размерность, сравнимую с количеством группируемых объектов, данные (точки, объекты) в таком пространстве удалены друг от друга, как правило, весьма значительно, что делает практически невозможным построение более-менее универсальной процедуры выбора параметров этих алгоритмов, при которых они обеспечивали бы эффективную группировку. Существуют непараметрические алгоритмы, комбинированные с другими популярными методами, например с методом k -средних [70], применимые, однако, лишь для задач с небольшим объемом входных данных (в [70] рассматриваются примеры с 2 кластерами).

Метод k -средних (альтернативное наименование — MSSC, Minimum Sum of Squares Clustering — группировка по минимуму суммы квадратов [71]) — один из классических, хорошо исследованных алгоритмов автоматической группировки (обучения без учителя). Как и другие методы, данный алгоритм пытается найти возможные классы (группы) объектов данных, организованных групп, элементы которых подобны в некотором роде [70]. Группа соответствует коллекции объектов, которые до некоторой степени «эквивалентны» друг другу, а также некоторому виртуальному объекту — центру (центроиду) кластера (группы). Задача автоматической группировки может быть сформулирована как модель распределения объектов в метрическом d -мерном пространстве между k группами так, чтобы в каждой группе суммарное расстояние

от элементов группы до ее центра достигало минимума. Метод k -средних можно считать самым важным методом обучения без учителя [72, 70]. Метод обладает следующими потенциальными преимуществами: (1) позволяет учитывать признаки (характеристики) различных типов, (2), обнаруживать группы произвольной формы, (3) обрабатывать данные при минимальных требованиях к знанию предметной области, к которой относятся входные параметры (характеристики группируемых объектов), (4) использование данных с шумами и выбросами. Вследствие этого метод находит широчайшее применение практически во всех сферах автоматической группировки и кластерного анализа [73].

Алгоритм включает всего два чередующихся шага [74, 75]: разбиение на группы (кластеры) вокруг известных центров (объект относится к той группе, центр которой является к нему ближайшим) и переопределение центров групп. Аналогичная методика используется также в алгоритме *FOREL* [76] и алгоритме *H-means* (*H*-средних).

Таким образом, k -средних — простой и быстрый алгоритм, применимый к широчайшему классу задач. Алгоритм k -средних имеет некоторые ограничения, в частности нужно заранее задать число групп k , на которые разбиваются объекты. В главе 3 рассмотрены методы решения данной проблемы. Отметим, что метод жадных эвристик, которому в основном посвящена настоящая работа, позволяет решить одновременно серию задач с различными значениями k , частично данную проблему снимая. Отдельным направлением можно считать развитие методов разбиения объектов на две группы (более узкая по сравнению с темой настоящего исследования задача бикластеризации). Алгоритмы для решения этой задачи исследованы, в частности, А.В. Кельмановым [77, 78].

Группировка, основанная на теории графов, иногда называемая спектральным объединением в кластеры, представляет объекты (точки) данных как узлы во взвешенном графе. Ребра, соединяющие узлы, имеют длину (вес), равную или обратно

пропорциональную мере сходства между объектами (точками). Идея состоит в том, чтобы разделить узлы на два подмножества A и B таким образом, чтобы величина разреза, т.е. сумма весов ребер, соединяющих узлы, находящиеся в множествах A и B , была минимальной. Алгоритмы, изначально предлагавшиеся для решения таких задач, часто приводили к разбиению на группы сильно различающегося (несбалансированного) размера, из-за чего в дальнейшем было предложено использовать ограничение на число точек в группе и более сложные ограничения.

Хофманом и Буманом [79] предложен детерминированный алгоритм имитации отжига для группировки данных, представленных с использованием мер сходства между объектами данных. В [80] сформулирована задача попарного объединения в группы на основе максимальных доминирующих множеств [81], которые являются непрерывным обобщением клик в графе.

Советские и российские исследователи традиционно формулируют задачи автоматической группировки и размещения как на сетях, так и в непрерывном пространстве, в виде задач целочисленного линейного программирования [82], причем, несмотря на NP -трудность задач, разработан широчайший арсенал в основном точных эффективных методов решения, в том числе для двухуровневых задач, в которых оценка значения целевой функции предполагает решение вложенной оптимизационной задачи. Тем не менее для алгоритмов, эффективных при решении, например, экономических задач, NP -трудность решаемых задач становится проблемой при решении больших задач автоматической группировки с учетом взрывного роста объемов данных, собираемых и обрабатываемых в автоматизированных системах. Под NP -трудными задачами в теории вычислительной сложности алгоритмов понимают класс задач, которые «как минимум так же трудны, как самая трудная из задач класса NP ». Иными словами, NP -трудными называют задачи, к которым можно свести любую

задачу из класса NP за полиномиальное время (полиномиально зависящее от объема входных данных). В свою очередь, к классу NP относят задачи, решения которых могут быть проверены на недетерминированной машине Тьюринга за полиномиальное время. Такие задачи, как p -медианная на графе, кроме NP -трудности, обладают тем свойством, что за полиномиальное время в общем случае невозможно получить приближенное решение с константной оценкой точности [83, 84].

Методы иерархической группировки могут осуществлять группировку объектов в непрерывном пространстве характеристик, строя при этом древовидную модель взаимоотношений объектов, что сближает их в какой-то мере с методами автоматической группировки на графах [85]. Эти методы могут быть использованы для группировки при достаточно больших объемах данных. Иерархические алгоритмы автоматической группировки рекурсивно выделяют вложенные группы агломеративно (начиная с рассмотрения каждого объекта в качестве отдельной группы, затем последовательно объединяя эти группы попарно, формируя таким образом иерархию этих групп) или применяя аналитический (нисходящий, диссоциативный) способ, начинающийся со всех объектов данных в одной группе и рекурсивно делящий каждую группу на меньшие группы. По сравнению с иерархическими алгоритмами объединения в группы, алгоритмы, подобные методу k -средних, находят все группы одновременно, не налагая иерархическую структуру. Входные данные иерархического метода группировки должны быть представлены матрицей подобия размера $N \times N$, где N — число группируемых объектов. Для методов, подобных k -средних, объекты располагаются в d -мерном пространстве характеристик, хотя некоторые разновидности алгоритмов, такие как k -медоид, могут оперировать и матрицей подобия. Матрица подобия может быть легко получена по координатам объектов в пространстве характеристик, обратное преобразование требует применения сложнейших методов, таких как многомерное шкалирование,

которые сами по себе могут включать в себя многократное решение задач автоматической группировки.

Метод *X-means* [86] включает в себя агломеративные и диссоциативные приемы — попарное объединение или разъединение групп, позволяя таким образом находить оптимальное значение числа групп k на основе оценок такими критериями, как информационный критерий Акаике (*AIC* — *Akaike Information Criterion*) или Байесов информационный критерий (*BIC* — *Bayesian Information Criterion*). В методе k -медоид группировка осуществляется вокруг центров групп, в качестве которых могут выбираться исключительно группируемые объекты (не произвольные точки в пространстве характеристик).

Как и метод *X-means*, метод жадных эвристик, являющийся главной темой настоящей работы, включая в себя агломеративную стратегию последовательного (но не попарного) объединения групп, не строит при этом иерархической структуры, характерной для иерархических методов автоматической группировки. Метод *X-means*, позволяя одновременно с решением задачи k -средних определять оптимальное значение числа групп k для задач с относительно небольшим объемом входных данных, снижает точность результата и ни в коей мере не добавляет стабильности получаемым результатам. Использование таких критериев, как *BIC* и *AIC*, учитывающих расстояния между объектами внутри групп, но никак не учитывающих расстояния между объектами, принадлежащими разным группам, в практических задачах может давать неадекватную оценку числа групп. Метод жадных эвристик в то же время позволяет решить одновременно серию задач с различными значениями k , результаты которых могут быть оценены в дальнейшем с применением любых критериев.

Некоторые алгоритмы автоматической группировки, такие как метод минимальной энтропии [87], имеют формулировку, оперирующую понятиями информационной теории. Метод информационного узкого места или бутылочного горлышка (*IBC* — *Information Bottleneck Clustering*) [88] был предложен

как обобщение теории зависимости искажения информации от скорости передачи и оперирует понятиями, характерными для алгоритмов сжатия данных с потерями. Например, этот метод применяется при группировке документов по ключевым словам. *IBC*-методы начинают группировку, рассматривая каждый объект как отдельную группу (кластер), а затем исключают одну за другой группы таким образом, чтобы мера сходства, измеряемая в данном случае с применением понятий, характерных для информационной теории, оставалась максимальной для текущего числа групп. Такие методы чрезвычайно требовательны к вычислительным ресурсам.

Заметим, что методы, основанные на жадных агрегативных эвристических процедурах, имеют общие черты с методами информационного бутылочного горлышка и в предельных случаях вырождаются в детерминированные методы, аналогичные либо тождественные методам информационного бутылочного горлышка.

Кроме упомянутых выше вопросов о применяемой мере сходства, применяемой целевой функции (и связанного с ними вопроса о форме групп — кластеров) и вопроса о числе групп (кластеров), перед исследователем, сталкивающимся с задачей автоматической группировки, могут возникать и другие вопросы.

Так, представление данных — один из наиболее важных факторов, определяющих работу алгоритма группировки. Если представление данных (выбор обрабатываемых характеристик объектов) выбрано правильно, можно применить какой-либо из простейших алгоритмов, получив при этом компактные и изолированные друг от друга группы. Отсутствие важных характеристик делает разбиение на группы невозможным либо бессмысленным, присутствие «лишних» характеристик, несущественных с точки зрения конкретной задачи автоматической группировки, приводит к разбиению на более мелкие группы, не позволяя достичь требуемого уровня обобщения и структурирования данных. Если же используются

характеристики, каждая из которых, будучи информативной, является зависимой от других характеристик, это, не искажая результата автоматической группировки, может весьма существенно повысить вычислительную сложность большинства алгоритмов автоматической группировки.

Хотя сам по себе алгоритм, часто называемый стандартной процедурой k -средних, является достаточно высокопроизводительной процедурой, для дальнейшего улучшения производительности этого алгоритма предложено множество различных методик [89]. Алгоритм чувствителен к начальному выбору центров групп (кластеров), может сходиться к местным оптимумам. Кроме того, когда число группируемых объектов очень велико (десятки тысяч и более), требуется большое количество времени для нахождения глобального оптимума или близкого к нему решения.

Хотя существуют «глобальные» методы решения задачи [90], претендующие на нахождение глобально оптимального решения (практически неприменимые к очень большим задачам и фактически точное решение не гарантирующие), основное направление современных исследований — развитие эвристических алгоритмов, дающих субоптимальные решения, близкие к истинному оптимуму задачи.

В [91] задействованы возможности генетического алгоритма (ГА) для улучшения результата K -средних. Как правило, в генетическом алгоритме промежуточные решения («хромосомы») кодируются в виде множеств объектов данных, относящихся к той или иной группе (кластеру) или в виде множества объектов данных, выбранных в качестве начальных центров групп, которые затем уточняются алгоритмом k -средних (стандартной процедурой k -средних).

В [92] также представлен алгоритм автоматической группировки, основанный на ГА. В данной стратегии глобального поиска промежуточное решение кодируется вещественными числами (ГА с вещественным алфавитом). Промежуточное решение представляет собой множество центров

групп в d -мерном пространстве. В [93] представлен модифицированный алгоритм K -средних (*PSOKM*) с использованием метода оптимизации роя частиц. Результаты этой работы показали возможность построения высокоточных классификаторов с более низким уровнем ошибочной классификации для весьма небольших по объему наборов данных (несколько сотен объектов).

Багировым и др. [94] предложено развитие модифицированного глобального алгоритма k -средних Ликаса и др., позволяющее получать точное решение быстрее предыдущих алгоритмов глобального поиска. На тестовых примерах доказана применимость к задачам с несколькими тысячами группированных объектов, при этом для решения задачи требуется очень большое время счета и огромные объемы памяти. Тем не менее NP -трудность задачи не позволяет применять данный алгоритм к задачам с очень большим объемом данных. Кроме того, несмотря на присутствие слова «глобальный» в названии метода, он не гарантирует нахождение истинного глобального минимума, хотя и демонстрирует сходимость к «очень хорошему» результату.

Метод k -средних, как правило (и по умолчанию), используется с квадратом евклидовой метрики для вычисления расстояния между центрами группы и объектами. В результате k -средних находит круглые (шарообразные) группы в данных (т.е. имеющие форму круга с той метрикой или мерой расстояния, которая применяется [95]). Метрика Махаланобиса позволяет обнаружить гиперэллипсоидальные группы [96], что сказывается весьма существенно на вычислительных затратах. Вариант метода k -средних, использующий расстояние Итакуры — Саито [97], используется для векторной квантизации в речевой обработке [98]. Достаточно широко используется метод k -средних с расстоянием L_1 (иначе называемым прямоугольной или манхэттенской метрикой) [52], позволяющий весьма существенно снизить влияние шумов на результат по сравнению с использованием квадратичных

евклидовых расстояний практически без повышения вычислительной сложности. Группы (кластеры) при этом получаются менее компактными, что продемонстрировано на реальных данных в главе 4. Используются и многие другие специальные метрики и меры расстояния, например, расстояния Брегмана [99]. Отметим, что если в качестве меры расстояния используется метрика, то постановка задачи k -средних в полной мере соответствует постановке одной из классических задач теории размещения — непрерывной задаче о p -медиане. Отметим, что квадратичное евклидово расстояние метрикой не является: не выполняется правило треугольника — квадрат длины отрезка AC может быть больше суммы квадратов длин отрезков AB и BC . В то же время если считать, что в качестве меры расстояния в p -медианной задаче могут использоваться и другие меры расстояния (не метрики), то постановку задачи k -средних и p -медианной задачи следует считать эквивалентными. В случае, если выполняется группировка некоторых реальных объектов по мере близости друг к другу в реальном географическом/геометрическом пространстве, как правило, используются евклидовы расстояния [100] (в этом случае задача k -средних в полной мере соответствует классической непрерывной p -медианной задаче), хотя при группировке объектов в очень обширных географических регионах переходят к расстояниям на сфере. Метод k -средних предполагает получение относительно равновесных кластеров, хотя такое свойство не характерно для некоторых его разновидностей, иногда рассматриваемых как отдельные методы. Так, метод *ISODATA* [101] состоит в том, что, задавшись центрами K классов, размещают все объекты по группам (классам), используя вместо расстояния до центров групп внутриклассовый критерий минимальности. После отнесения всех элементов к группам центры пересматриваются, далее эти два шага чередуются. Иной способ отнесения объектов к группам предусмотрен в [102]. Здесь для определения ближайшего к объекту центра группы строится kd -дерево. Ме-

тоды *Kernel K-means* (k -средних с выделением «ядра» группы) [103] были предложены для обеспечения возможности обнаружить группы произвольной формы, при этом используется соответствующая «ядерная» функция подобия. Подобные модификации метода вносят изменения исключительно в способ распределения объектов между группами, вводя дополнительные внутрикласовые коэффициенты или более сложные модификации. Различные модификации метода k -средних, стратегии глобального поиска и иные его усовершенствования, разработанные для применения с задачами k -средних в «классической» постановке, могут применяться (возможно, после адаптации) и с модификациями, подобными *ISODATA* [101, 104].

Встречаются задачи автоматической группировки, в которых фигурируют большое число группируемых объектов (иногда — миллионы), каждый из которых характеризуется большим числом измерений (характеристик). Число измерений (размерность пространства характеристик) может достигать до нескольких тысяч (такова, например, задача автоматической группировки электрорадиоизделий, рассматриваемая [52]). К большим задачам автоматической группировки относится, например, задача контекстного поиска изображений (*Content Based Image Retrieval, CBIR*), цель которой состоит в том, чтобы найти визуально подобные изображения по заданному в запросе изображению. Хотя данная тематика изучается более 15 лет, достигнуты достаточно скромные успехи в данном направлении. В ранних версиях методы *CBIR* сосредотачивались на определении подобия изображений по вычислительным оценкам особенностей цвета, формы и текстуры изображения [105], более поздние версии [106] основываются на определении ключевых точек изображения. Тем не менее такие подходы малопригодны на более-менее крупных библиотеках изображений.

С другой стороны, поиск текстов требует гораздо меньше времени. Так, 0,1 секунды требуется для поиска в более чем

10 миллиардах документов *Google* [97]. К задаче поиска текстов может сводиться и задача поиска изображений: ключевые точки изображений сначала разбиваются на большое количество групп (решается задача автоматической группировки), называемых визуальными словами. Представляя каждое изображение гистограммой визуальных слов, задача превращается в простую задачу текстового поиска. Здесь также важной проблемой является определение числа выделяемых групп.

Для крупных наборов данных разработано большое число алгоритмов автоматической группировки, среди которых можно выделить основные категории:

А) Поиск эффективного ближайшего соседа. Одна из основных операций в любом алгоритме автоматической группировки состоит в определении членства в группе для каждого из группируемых объектов, которое может определяться в соответствии с отнесением ближайшего соседнего объекта к той или иной группе (кластеру) [107]. Метод ближайшего соседа сам по себе — достаточно грубый метод автоматической группировки, что справедливо и для методов, производных от него.

Б) Суммирование данных: цель подхода состоит в том, чтобы повысить эффективность алгоритма группировки, вначале выполнив подведение итогов на большом наборе данных, заменив близкие друг к другу объекты единственным объектом (первый этап группировки) и получив таким образом относительно небольшое подмножество, а затем применяя к этому подмножеству какие-либо алгоритмы автоматической группировки (второй этап группировки). К таким алгоритмам относятся *BIRCH* [108], алгоритмы Штайнбаха [109], методы огрубления [110]. Как следует из названия последнего из приведенных подходов, применение таких двухуровневых технологий неизбежно приводит к «огрублению» — снижению точности результата.

В) Нарастающее объединение в кластеры. Эти алгоритмы [111] направлены на обработку потоковых данных

таким образом, чтобы достичь максимальной эффективности (в плане вычислительной производительности) обработки массива данных за один проход (обработка в единственном цикле) в отличие от большинства подходов, требующих многократного обращения к данным. Как правило, используются методы иерархической классификации [112].

Г) Подходы, основанные на выборках (методы сэмплинга). Такие алгоритмы, как *CURE* [113], *CLARA/CLARANS* [114] и др., выбирают некое случайное подмножество из большого набора данных, выполняют автоматическую группировку этого набора, а результат ее используют как исходное решение при выполнении группировки исходного большого набора данных.

Нетрудно заметить, что все четыре подхода ставят своей главной целью повышение вычислительной производительности. В принципе, за исключением нарастающей кластеризации, все эти подходы могут применяться для решения задачи k -средних и p -медианной задачи. Однако и методы сэмплинга, вначале решающие упрощенную задачу, и методы суммирования данных, и методы, основанные на поиске ближайшего соседа, разбивают процесс автоматической группировки как минимум на два этапа — решение множества локальных задач автоматической группировки на первом этапе различными способами для получения упрощенной («огрубленной») задачи, затем — ее решение, а затем (не обязательно) — решение исходной задачи с использованием решения «огрубленной» задачи в качестве начального решения. Таким образом, все подобные методы «заточены» на получение быстрого решения упрощенной, «огрубленной» задачи. Соответственно, применение подобных подходов неизбежно ведет к снижению точности результата, а тот факт, что как минимум первый этап решения (получение «огрубленной» задачи) является рандомизированной процедурой, снижает стабильность получаемого решения.

Модели и методы автоматической группировки находят многочисленные успешные применения в области анализа

данных [52]. Тем не менее построение алгоритмов систем автоматической группировки связано с решением некоторых возникающих проблем, например [52]:

А) Желательно достигнуть более тесной интеграции между алгоритмами автоматической группировки и прикладными задачами. Например, в некоторых приложениях необходимо выделить несколько связанных групп (часть данных — менее связанные группы — могут быть проигнорированы), в то время как другие приложения могут требовать обязательного разделения всех данных. Часто важен не выбор наилучшего алгоритма разбиения на группы, а именно наиболее адекватной модели группировки.

Б) Независимо от принципа автоматической группировки и соответствующей целевой функции, большинство методов автоматической группировки в конечном счете сводятся к комбинаторным задачам оптимизации (требуется применение различных стратегий глобального поиска: задачи, как правило, *NP*-трудны). В результате при решении задач с большим объемом данных на первый план выходит проблема вычислительной сложности. Следовательно, важно выбрать принципы построения алгоритма автоматической группировки, которые обеспечивают эффективные в вычислительном отношении решения для широкого круга задач в той или иной предметной области.

В) Важнейшей проблемой, связанной с автоматической группировкой, является стабильность результата. Хорошим подходом к созданию алгоритмов и систем автоматической группировки следует считать тот, который стабильно приводит к разбиению на одни и те же группы (кластеры) либо к очень близким результатам и практически не зависит от малых колебаний в исходных данных. Хотя точность решения, безусловно, всегда важна, во многих практических приложениях на первый план выходит именно проблема создания стабильного алгоритма. Существующие детерминированные алгоритмы, во-первых, не достаточно точны (показано

в [52]), а во-вторых, требуют весьма высоких вычислительных затрат. Алгоритмы, предназначенные для обработки больших объемов данных [115], основаны на декомпозиции исходной задачи и решении ее в два этапа: создание «огрубленной» аналогичной задачи с использованием рандомизированной процедуры, ее решение и использование полученного решения в качестве начального решения исходной задачи. Такой двухуровневый подход неизбежно приводит к снижению стабильности результата (как и к снижению точности).

3.2. ЗАДАЧА КЛАССИФИКАЦИИ В АНАЛИЗЕ ДАННЫХ СЛОЖНЫХ СИСТЕМ

В качестве практической задачи прогнозирования и диагностики использовалась следующая задача [145, 146]. Существует выборка данных S , состоящая из 1700 больных острым инфарктом миокарда (ИМ), проходивших лечение в отделении реанимации и интенсивной терапии и в кардиологическом отделении городской клинической больницы [147]. Информация об анамнезе пациентов и течении ИМ получена из историй болезни и сконцентрирована в 117 полях электронной таблицы (базы данных). База данных содержит информацию о возрасте, поле пациента, локализации и глубине ИМ, данных анамнеза, изменениях ЭКГ, количестве калия, натрия, некоторых ферментах крови, особенностях клиники в первые часы заболевания.

За период пребывания больных в стационаре фибрилляция предсердий (ФП) наблюдалась у 170 (10,0%) больных, фибрилляция желудочков (ФЖ) у 71 (4,2%), отек легких (ОЛ) у 159 (9,4%), разрыв сердца (РС) у 54 (3,2%), летальный исход (ЛИ) у 271 (15,9%). Задача состоит в том, чтобы на основании имеющейся выборки данных (классифицированных ранее наблюдений) извлечь информацию о состоянии «нового» пациента, наблюдение которого не содержится в выборке. Пациент, имеющий осложнение, называется положительным объектом, пациент, не имеющий осложнения, — отрицательным объектом.

Для каждого осложнения инфаркта миокарда были сгенерированы следующие выборки:

1. Фибрилляция желудочков. В испытании участвовали 70 положительных и 70 отрицательных пациентов.

2. Фибрилляция предсердий. В испытании участвовали 169 положительных и 169 отрицательных пациентов.

3. Отек легких. В испытании участвовали 181 положительный и 157 отрицательных пациентов.

4. Разрыв сердца. В испытании участвовали 54 положительный и 54 отрицательных пациентов.

5. Летальный исход. В испытании участвовали 160 положительных и 160 отрицательных пациентов.

Оценка точности классификации проводилась с помощью разделения каждой из полученной выборки на обучающую и тестовую в процентном отношении 80/20.

Результат классификации характеризуется двумя параметрами — чувствительностью (количество верно классифицированных здоровых пациентов) и специфичностью (количеством верно классифицированных больных пациентов). Большее значение при анализе данных имеет специфичность.

В табл. 3.1–3.5 представлены результаты работы методов и алгоритма: наивной байесовской классификации (*NaiveBayes*), ближайших соседей (*IB1*), опорных векторов (*SMO*), деревьев решений (*J48*), построения 1-правил (*1-R*), которые были получены в системе анализа данных *WEKA*, и логического анализа данных (*LAD*), полученных с помощью программного приложения, реализованного авторами.

Таблица 3.1

Фибрилляция желудочков

Показатель \ Метод	NaiveBayes	IB1	SMO	1-R	J48	LAD
Количество верно классифицированных отрицательных объектов, %	89	89	78	78	78	94

Окончание табл. 3.1

Показатель \ Метод	NaiveBayes	IB1	SMO	1-R	J48	LAD
Количество верно классифицированных положительных объектов, %	69	46	77	92	92	100

Таблица 3.2

Фибрилляция предсердий

Показатель \ Метод	NaiveBayes	IB1	SMO	1-R	J48	LAD
Количество верно классифицированных отрицательных объектов (%)	72	72	66	50	75	71
Количество верно классифицированных положительных объектов (%)	58	61	67	94	58	79

Таблица 3.3

Отек легких

Показатель \ Метод	NaiveBayes	IB1	SMO	1-R	J48	LAD
Количество верно классифицированных отрицательных объектов (%)	72	59	69	52	90	85
Количество верно классифицированных положительных объектов (%)	76	61	76	90	66	76

Каждый из исследуемых методов показал достаточно высокие результаты точности в процессе исследования, однако метод ближайших соседей показал наихудшие результаты при прогнозировании осложнений, что говорит о слабой применимости метода к данным задачам. Метод *SMO* показал

сравнительно одинаковые результаты по параметрам чувствительности и специфичности, в отличие от метода ближайших соседей, 1-*R* и *J48*. Однако у алгоритма построения 1-*R* при классификации большее значение параметров специфичности для четырех первых задач, что является важным при прогнозировании. Следовательно, каждое осложнение характеризуется одним, наиболее важным признаком, который легко выявить. Метод *LAD* показал лучший результат классификации, кроме того, он дает ряд преимуществ при практическом использовании. Прежде всего, в явном виде известны правила, по которым принимается решение о принадлежности к какому-либо классу. Кроме того, при применении модели классификации к новому пациенту по тому, каким числом паттернов покрываются его данные, можно судить о вероятности возможной ошибки при распознавании.

Таблица 3.4

Разрыв сердца

Показатель \ Метод	NaiveBayes	IB1	SMO	1-R	J48	LAD
Количество верно классифицированных отрицательных объектов (%)	89	89	78	78	78	91
Количество верно классифицированных положительных объектов (%)	69	46	77	92	92	91

Таблица 3.5

Летальный исход

Показатель \ Метод	NaiveBayes	IB1	SMO	1-R	J48	LAD
Количество верно классифицированных отрицательных объектов (%)	90	81	88	97	84	84

Окончание табл. 3.5

Показатель \ Метод	NaiveBayes	IB1	SMO	1-R	J48	LAD
Количество верно классифицированных положительных объектов (%)	72	34	75	40	44	81

3.3. ЗАДАЧА КЛАСТЕРИЗАЦИИ В АНАЛИЗЕ ДАННЫХ СЛОЖНЫХ СИСТЕМ

Для реализации эксперимента регламентируется следующая постановка задачи: определить эффективность реализации инновационной политики в регионах Российской Федерации с включением процедуры кластеризации на основе метода k -средних.

Предлагается использовать следующий алгоритм оценки, включающий последовательное выполнение следующих этапов:

1) отбор, анализ и формирование комплекса показателей, необходимых для проведения оценки; определение пороговых значений; приведение показателей к сопоставимому виду; формирование интегрального индекса в разрезе двух групп показателей (инновационный потенциал и результаты инновационной деятельности);

2) кластеризация регионов по уровню инновационного потенциала (на основе значений интегрального индекса) и характеристика типологических групп;

3) кластеризация регионов по результатам инновационной деятельности (на основе значений интегрального индекса) и характеристика типологических групп;

4) сопоставление результатов кластеризации, проведенной на этапах (2) и (3), для определения результатов региональной инновационной политики.

На основе представленных выше факторов на первом этапе с учетом открытых к использованию статистических данных

сформирован комплекс показателей оценки инновационного потенциала региона, который в рамках данного исследования рассматривается как совокупность трех видов ресурсов: человеческих, финансовых и материально-технических (объединены в виду взаимозависимости инвестиций и технологий), а также организационных.

Для формирования показателей оценки результатов инновационной деятельности на региональном уровне проведен анализ нормативно-правовой и программно-целевой документации, а также программ развития инновационной сферы в разрезе субъектов Российской Федерации (85 субъектов). На основе синтеза информации анализируемой документации произведена выборка показателей, наиболее часто используемых (декларируемых) для оценки результатов инновационной деятельности. С учетом заключения проведенного анализа, а также исключая взаимозависимость показателей, выбраны и адаптированы для универсального применения следующие показатели:

- удельный вес объема инновационных товаров, работ, услуг в общем объеме отгруженных товаров, выполненных работ, услуг (в %);
- число выданных патентов на 10000 чел. населения региона (в ед.);
- соотношение объема отгруженной инновационной продукции и затрат на технологические инновации (в %);
- число патентных заявок на изобретения, поданных в регионе на 10 000 чел. населения (в ед.).

Далее для предложенной системы оценочных индикаторов целесообразно определить пороговые значения [116]. Для приведения показателей к сопоставимому виду целесообразно провести процедуру сложного нормирования в соответствии с формулами (3.1) и (3.2). Для нормирования индикаторов использована методика, предложенная исследователями Митяковым С.Н. и Сенчаговым В.К. [24]. В ее основу для ранжирования индикаторов заложена «теория зониро-

вания». Для соотношения типа «не менее» учеными предлагается функция вида

$$y = \begin{cases} 2^{\left(1 - \frac{a}{x}\right) / \ln \frac{10}{3}}, & \text{если } \left(\frac{x}{a}\right) > 1; \\ 2^{-\log_{10} \frac{a}{3x}}, & \text{если } \left(\frac{x}{a}\right) \leq 1, \end{cases} \quad (3.1)$$

где x — реальное значение индикатора; a — пороговое значение.

Для соотношения типа «не более» учеными предлагается функция вида

$$y = \begin{cases} 2^{\left(1 - \frac{x}{a}\right) / \ln \frac{10}{3}}, & \text{если } \left(\frac{x}{a}\right) < 1; \\ 2^{-\log_{10} \frac{x}{3a}}, & \text{если } \left(\frac{x}{a}\right) \geq 1, \end{cases} \quad (3.2)$$

где x — реальное значение индикатора; a — пороговое значение.

Использование сложных функций предоставляет возможность расширить диапазон визуализации результатов. Так, логарифмическая зависимость позволяет исследовать нелинейную структуру индикаторов внутри сектора; менее плавная степенная зависимость позволяет игнорировать несущественные детали при значительном превышении индикаторов своих пороговых значений [117].

После проведения процедуры нормирования на основе формулы (3.3) рассчитывается интегральный показатель в разрезе двух проекций: 1) инновационный потенциал; 2) результаты инновационного деятельности:

$$I = (\text{НП}_1 + \text{НП}_2 + \dots + \text{НП}_n) / n, \quad (3.3)$$

где НП — нормированный показатель, n — количество показателей в проекции.

Значения, характеризующие инновационный потенциал региона и результаты инновационной деятельности, предлагается вычислять как среднеарифметическое нормированных показателей соответствующей группы.

На втором и третьем этапах, исходя из полученных значений, проводится кластерный анализ (древовидная кластеризация, метод k -средних) с использованием пакета *Statistica* 6.0. Кластерный анализ является универсальным инструментом, получившим широкое распространение в региональном и территориальном моделировании [118–121].

На последнем этапе производится сопоставление результатов, полученных в рамках реализации второго и третьего этапов.

Апробация предложенного концептуального подхода произведена на примере 85 субъектов Российской Федерации, расчетный период 2010–2019 гг. (для Республики Крым и города федерального значения Севастополя расчеты производились с 2015 года).

На втором этапе с использованием статистической информации, представленной на официальном сайте Росстат [122], рассчитаны и нормированы по формулам (3.1), (3.2) фактические значения по всем показателям. После проведения процедуры нормирования на основе формулы (3.3) рассчитан интегральный индекс в разрезе двух проекций: инновационный потенциал и результаты инновационной деятельности.

Далее на основе полученных интегральных индексов реализована процедура кластерного анализа с использованием программы *Statistica* 6.0, а именно построены дендрограммы и проведен анализ методом k -средних для дифференциации регионов на группы (кластеры) по уровню инновационного потенциала.

Построение и дальнейший анализ полученной дендрограммы (рис. 3.1) показал целесообразность выделения 4 кластеров. Эксперименты по кластеризации с использованием метода k -средних также позволили выделить 4 кластера,

состав которых сопоставим с результатами древовидной кластеризации.

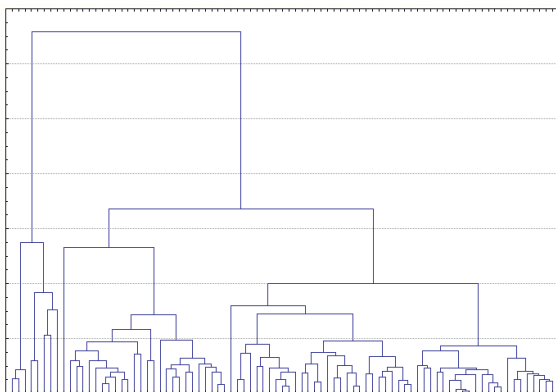


Рис. 3.1. Дендрограмма уровня инновационного потенциала (метод полной связи, евклидово расстояние)

В «Кластер 1» (8 субъектов, среди них Калужская, Московская, Новосибирская, Томская области и др.) входят регионы Российской Федерации с высоким уровнем инновационного потенциала — признанные лидеры в инновационной сфере. Интегральный индекс инновационного потенциала регионов первого кластера отличается устойчивостью и на протяжении всего исследуемого периода имеет значение «выше 0,7», а часть субъектов обладает индексом «свыше 0,9».

Регионы «Кластера 1» имеют наибольшие фактические значения по интенсивности исследовательских ресурсов, обеспечивая стабильно высокий уровень занятости в сфере научных исследований и разработок. Также отличительной особенностью регионов данного кластера является превышение пороговых значений по показателю «Фактические затраты на научные исследования и разработки».

«Кластер 2» (25 регионов, в том числе Республики Татарстан, Пермский и Красноярский края, Самарская и Свердловская области и др.) и «Кластер 3» (30 регионов, среди

них Алтайский и Хабаровский края, Иркутская область, Республика Башкортостан и др.) в совокупности объединили треть регионов страны и представляют субъекты, обладающие в целом более скромным инновационным потенциалом, в отличие от субъектов, входящих в «Кластер 1». При этом важно отметить, что если в целом у регионов «Кластера 2» интегральный индекс в течение исследуемого периода не опускается ниже значения в «0,5», а в динамике показывает значения «более 0,6», что характеризует уровень риска в инновационной сфере как «приемлемый», то субъекты «Кластера 3» по интегральному индексу инновационного потенциала характеризуются уже «критическим» риском.

Регионы кластера 4 (всего 22 субъекта РФ, среди которых Республики Дагестан, Калмыкия, Хакасия, ЯНАО, Вологодская область и др.) существенно уступают регионам первого, второго и третьего кластеров по уровню инновационного потенциала. Отметим, что регионы данного кластера имеют крайне низкие показатели интенсивности исследовательских ресурсов, что свидетельствует о наличии острой проблемы кадровой обеспеченности субъектов в сфере научных исследований и разработок. Кроме того, при наличии низких значений показателя «Износ основных фондов» в данных субъектах в начале исследуемого периода прослеживается негативная тенденция изменения данного показателя, что отрицательно характеризует уровень технологической оснащенности территорий.

На третьем этапе проведен кластерный анализ регионов по уровню интегрального индекса, характеризующего «Результаты инновационной деятельности» методом k -средних, также построена дендрограмма (рис. 3.2), что позволило распределить регионы на 2 кластера.

Выделение именно двух кластеров обусловлено сильной дифференциацией значений интегрального индекса в течение рассматриваемого периода для большинства субъектов. В этой связи показательным является среднее значение индекса.

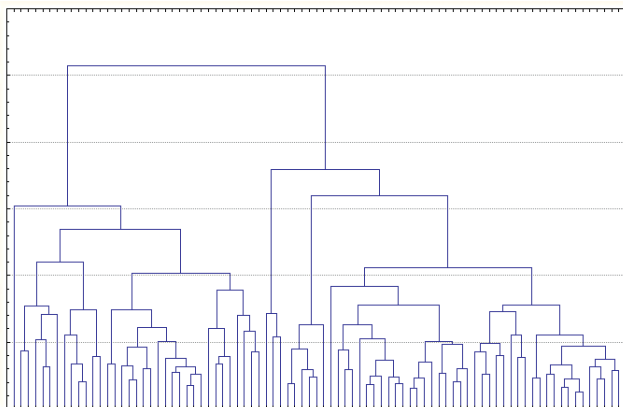


Рис. 3.2. Дендрограмма результатов инновационной деятельности (метод полной связи, евклидово расстояние)

«Кластер 1» объединил в себе 38 субъектов (в том числе Республика Татарстан, Нижегородская и Московская области, Хабаровский край и др.) с относительно высокими результатами инновационной деятельности за исследуемый период. Среднее значение интегрального индекса результатов инновационной деятельности превышает значение «0,5».

Отметим, что обособленное место в этом кластере занимает город федерального значения Москва как субъект, имеющий значение интегрального индекса в определенные периоды «выше 1,0», что свидетельствует о достижении пороговых значений по большинству показателей и положительно характеризует результаты инновационной деятельности на данной территории.

Регионы «Кластера 2» значительно отстают от регионов «Кластера 1» и характеризуются низкими результатами инновационной деятельности. Кроме того, отметим, что часть регионов (Республики Тыва, Хакасия, Алтай, Адыгея, Калмыкия, Ингушетия, Еврейская автономная область, ЯНАО, Чукотский автономный округ, Амурская область) по показателям, входящим в интегральный индекс, за определенные

периоды имеют нулевые значения, что свидетельствует об отсутствии инновационной деятельности.

В рамках заключительного этапа посредством проведения процедуры сопоставления результатов кластеризации, проведенной на этапах 2 и 3, были сделаны следующие выводы:

1. Все регионы с высоким уровнем инновационного потенциала (за исключением Калужской области) успешно используют имеющиеся ресурсы (исследовательские, материальные, технологические) и демонстрируют высокие результаты инновационной деятельности. К данной группе субъектов относятся регионы «Кластера 1» в разрезе проекции «Инновационный потенциал» и «Кластера 1» в разрезе проекции «Результаты инновационной деятельности».

2. Часть регионов, обладающих относительно высоким уровнем инновационного потенциала («Кластер 2»), не обеспечивают эффективность его использования (Калужская, Ленинградская, Новгородская области, Красноярский, Приморский края и др.). Так, Красноярский край, имея значение интегрального индекса инновационного потенциала «0,6», показывает значение индекса результатов инновационной деятельности лишь на уровне «0,4». Данную группу составляют регионы «Кластера 2» в разрезе проекции «Инновационный потенциал» и «Кластера 2» в разрезе проекции «Результаты инновационной деятельности». Такие результаты свидетельствуют о наличии дисбаланса между инновационным потенциалом территорий и отсутствием или несовершенством механизмов реализации инновационной деятельности, что, в свою очередь, может быть обусловлено, в частности, неэффективностью проводимой региональной инновационной политики.

3. Часть регионов, обладающих более скромным инновационным потенциалом («Кластер 4»), демонстрируют более высокие результаты инновационной деятельности (Брянская, Костромская области, Республики Адыгея, Республика Марий Эл, Ставропольский край). Например, Республика Марий Эл при значении интегрального индекса инноваци-

онного потенциала на уровне «0,38» имеет значение интегрального индекса результатов инновационной деятельности на уровне «0,66». Такая ситуация, наоборот, может свидетельствовать о наличии в регионе действенных механизмов активизации инновационной деятельности, формировании и реализации эффективной инновационной политики, несмотря на ограниченный потенциал.

В целом можно утверждать, что использования метода кластерного анализа, с одной стороны, предоставляет возможности дифференцировать субъекты с учетом их сходного сочетания по ряду критериев, исходя из этого, формировать «рамочные» механизмы, направленные на стимулирование и активизацию инновационной деятельности в регионах. С другой стороны, результаты оценки не предоставляют возможности сформировать количественные рекомендации для достижения границ эффективности для так называемых неэффективных регионов. Данная возможность предоставляется на основании использования *Data Envelopment Analyses (DEA)* (непараметрический метод исследования операций, основанный на линейном программировании).

В связи с этим становится очевидным необходимость комбинации метода кластеризации k -средних, алгоритм которого позволяет разбить множество объектов на группы, схожие по свойствам, на основании вычисления расстояния до центра кластера, с результатами метода *DEA*-анализа для формирования выборки выявленных «эталонных» регионов в качестве центров кластеров.

Методы автоматической группировки объектов, несмотря на сходство с методами решения задач теории размещения, развивались во многом отдельно. Популярные алгоритмы автоматической группировки объектов основаны в основном на эвристических методах локального поиска и стратегиях глобального поиска.

Отмечается некоторый дефицит компромиссных по качеству результата и времени счета методов автоматической группировки. Под качеством в данном случае понимается точность (близость значения целевой функции к глобальному оптимуму) и стабильность (близость получаемых значений друг к другу при множественных запусках алгоритма).

Практическая реализация задачи кластеризации доказывает, с одной стороны, наличие возможности дифференцировать субъекты с учетом их сходного сочетания по ряду критериев, исходя из этого, формировать «рамочные» механизмы, направленные на повышение эффективности функционирования сложных систем. С другой стороны, результаты оценки не предоставляют возможности сформировать количественные рекомендации для достижения границ эффективности для так называемых неэффективных объектов исследования.

Глава 4

МЕТОДЫ И МОДЕЛИ ПРОГНОЗИРОВАНИЯ СОСТОЯНИЯ СЛОЖНЫХ СИСТЕМ

4.1. ОБЗОР ОСНОВНЫХ МЕТОДОВ И МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ

Под методом прогнозирования понимают некую последовательность действий, необходимых к реализации для формирования модели прогнозирования. Модель прогнозирования будет представлять функциональную реализацию адекватного описания исследуемого процесса и основание для получения будущих прогнозных значений.

Методы прогнозирования классифицируются на интуитивные методы (основанные на экспертной оценке) и формализованные методы (основанные на некоей математической зависимости, формирующие базис для моделей прогнозирования). В свою очередь, модели прогнозирования подразделяются на модели предметной области и модели временных рядов (включают статистические и структурные модели) [123].

Ограничениями существующих моделей и методов многомерного анализа и прогнозирования состояния сложных систем, состоящих из разнородных и разнокачественных объектов, является сложность учета опосредованного влияния взаимозависимых компонентов в условиях неопределенности [124]. Для решения задач многомерного анализа и прогнозирования используются различные подходы, в большинстве случаев основанные на методах математической статистики, теории случайных процессов, распознавания образов. При этом, как правило, эти методы и модели базируются

на подходах, которые используются при анализе и прогнозировании одномерных временных рядов [124, 125, 126]. В настоящее время активно развиваются подходы к многомерному анализу и прогнозированию состояния сложных систем на основе временных рядов с использованием нечетких и нейросетевых методов и моделей [124, 127, 128].

Сравнительный анализ методических подходов к процедуре моделирования и прогнозирования представлен в табл. 4.1.

В результате проведенного сравнительного анализа методических подходов к процессам моделирования и прогнозирования, в том числе под влиянием факторов, можно сделать следующие выводы:

1) Наиболее распространенными являются методы прогнозирования, основанные на особенностях тренда временного ряда, которые можно отнести к «оптимистичным», поскольку, как правило, предполагается, что влияние негативных факторов не усилится. Среди таких методов оптимальным является модель Хольта – Винтерса. Модель может быть использована для прогнозирования индикаторов состояния сложных систем в неизменных условиях или без учета влияния факторов. В частности, модель может быть использована для пересчета и наблюдения динамики функциональных связей между плановыми показателями (которые используются в качестве факторов) и спрогнозированными индикаторами функционирования сложных систем.

2) Методы данной группы не учитывают влияние факторов и, соответственно, не могут выявить тенденцию (а также количественно выразить) изменение изучаемого показателя (-ей) под воздействием признаков. Из всех рассмотренных моделей оптимальным становится регрессионное прогнозирование (в частности метод наименьших квадратов), поскольку ключевым преимуществом выступает определение направления и количественное отражение динамики «зависимой» переменной.

Таблица 4.1

Сравнительный анализ методов и моделей прогнозирования

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
Метод Хольта — Винтерса (частный случай метода — Модель Брауна) [129, 130]	Метод Хольта — Винтерса является модификацией метода экспоненциального сглаживания для сезонных рядов. Этот метод учитывает экспоненциальный тренд (тенденция изменения показателей временного ряда) и аддитивную сезонность (периодические колебания, наблюдаемые во временных рядах)	Оценивает тренд и учитывает случайные колебания. Включает оценку точности прогнозов модели (определение вероятности ошибки). Отсутствует требование по длительности периода для расчета прогноза. Подходит для прогнозирования динамики количественных и качественных (главное, чтобы они были количественно измеримы) показателей	Исходный ряд анализируется обособленно от остальных явлений, оказывающих воздействие на него. Даже при наличии дополнительных данных они будут учтены только при настройке скорости адаптации модели. С течением времени стал использоваться как метод краткосрочного и среднего прогнозирования (для повышения надежности результатов в условиях высокой изменчивости влияния факторов на анализируемый показатель). Строится на основании динамики массива и допущении о наличии тенденций в рядах данных и общей идее о том,

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
Регрессионное прогнозирование (в частности, с применением МНК, а также классификационно-регрессионные деревья) [131, 132]	Основывается на результатах оценки влияния факторов на переменную (для отбора действительно воздействующих). Построенное регрессионное уравнение соединяет параметры развития показателя с воздействием факторов на него. Получение прогноза	Воздействие может быть выражено количественно. Наглядность и интерпретируемость результатов	что эти тенденции не должны сильно меняться во времени. На практике это допущение редко выполняется: плавные линейные тенденции в рядах данных сменяются резкими нелинейными тенденциями, а периодичность циклической составляющей не постоянна Оценивать лучше влияние ограниченного числа факторов (например, с наибольшим уровнем влияния). Для исключения влияния генероскедастичности массива применяется обобщенный метод МНК. Человеческий фактор при объяснении связей

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
Модель <i>ARIMA</i> (Модель Бокса – Дженкинса)	осуществляется путем подстановки переменных в уравнение. В соответствии с методом МНК аппроксимирующая эмпирическая функция имеет наименьшее среднее квадратическое отклонение от точек экспериментального временного ряда. Из-за простоты, доступности, удобства интерпретации, обеспечения приемлемой точности МНК нашел широчайшее применение в моделировании различных процессов	Подходит для анализа нестационарных (не имеющих четкой	Высокие требования к рядам данных: для построения адекватной модели <i>ARIMA</i>

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
Обновленная модель <i>ARIMA</i> [129, 133]	с помощью авторегрессии — скользящего среднего. Является развитием модели Хольта — Винтерса и включает разностические исследования ошибок, однако практические исследования показали, что оценка коэффициента при разности ошибок в большинстве случаев близка к нулю. Есть три различных числа (p, d, q), которые используются для параметризации моделей в <i>ARIMA</i> . Вместе эти параметры подчитывают сезонность, тенденцию и шум в наборах данных: (p) — это авторегрессионная	тенденции) временных рядов. Метод краткосрочного планирования. Модели имеют четкое математическое обоснование. Существует процедура проверки построенной модели на адекватность. Подходит для прогнозирования динамики количественных и качественных (главное, чтобы они были количественно измеримы) показателей	требуется не менее 40 наблюдений. Если тренд временного ряда далек от линейного, применение данного метода необходимо. <i>ARIMA</i> может учесть влияние не более 3 факторов. В обновленной модели <i>ARIMA</i> прогнозирование с учетом влияния факторов строится на основе будущих значений внешних факторов, которые спрогнозировали заранее или использовали плановые (целевые) показатели. Таким образом, каждый внешний фактор включает вероятность ошибки, которая может сильно отклоняться от реальности, что может сказаться на точности

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
	<p>модель. Модель временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда, что предопределяет возможность прогноза; (d) — является интегрированной частью модели. Например, следующее значение рассматриваемого показателя будет таким же, если разница между 3 предыдущими значениями была незначительной;</p> <p>(q) — это так называемая скользящая средняя модели.</p>		<p>прогнозирование искомого временного ряда</p>

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
	<p>Именно она позволяет установить погрешность модели как линейную комбинацию значений ошибок, которые произошли в прошлом. В настоящее время внедряется улучшенная модель <i>ARIMAX</i> – расширенная модель авторегрессии скользящего среднего (<i>ARIMA</i>), которая при построении прогноза учитывает не только значения самого процесса, но и внешние параметры, влияющие на него. При этом значения экзогенных (внешних) факторов должны соответствовать выбранному временному промежутку</p>		

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
Метод Монте-Карло [134]	Для моделирования генерируется определенное количество случайных величин, отвечающих установленным критериям, а затем на их основе вычисляются приближительное значение искомой величины	Позволяет учесть элемент случайности и сложность отдельных процессов. Метод является устойчивым по отношению к изменению различных параметров, таких как распределение случайной величины. Метод используется как альтернатива там, где не срабатывают привычные методы исследования, основанные на математических расчетах	Моделирование нельзя применить относительно заданной совокупности значений (можно только указать интервал разброса случайных величин). Неоднозначная точность результатов (зависит от размерности ряда данных: чем больше данных, тем выше точность)
Метод классификационных функций [135]	Прогнозирование на основе рассмотрения динамики факторов, влияющих на показатель (снижения	Исключение субъективного фактора. Простота расчетов	Надежность метода (в случае, если динамика показателей относится к нелинейной, скачкообразной, результаты прогноза могут быть недостоверными).

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
	или увеличения влияния). По временным рядам воздействующих показателей рассчитываются индексы роста (за исключением индикаторов, которые уже рассчитаны индексным методом)		Отсутствует возможность количественной оценки влияния факторов. Метод краткосрочного планирования (вероятность наступления угрозы рассчитывается только на следующий год), что в условиях высокой изменчивости факторов считается нормальным
Дисперсионный анализ ANOVA (<i>analysis of variance</i>) — анализ вариативности [136]	Применяется с целью анализа изменчивости признака под воздействием контролируемых переменных. Дисперсионный анализ — метод, направленный на поиск зависимостей в экспериментальных данных путем исследования значимости различий в средних значениях.	Позволяет проводить сравнение трех и более массивов. Подходит для изучения как количественных, так и качественных факторов.	При сравнении нескольких массивов высока вероятность возникновения эффекта множественных сравнений (увеличенная вероятность ошибки). Используется для установления влияния ряда факторов, но не позволяет количественно спрогнозировать изменение переменной под воздействием признаков

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
	Основная идея дисперсионного анализа состоит в сравнении «факторной дисперсии», порождаемой воздействием фактора, и «остаточной дисперсии», обусловленной случайными причинами		
Модель на базе цепей Маркова [137]	Цель Маркова — последовательность случайных событий с конечным или счетным числом исходов, где вероятность наступления каждого события зависит только от состояния, достигнутого в предыдущем событии.	Модель считается достаточно простой в использовании. Применима для генерации и предсказания случайностей	Больше подходит для прогнозирования динамики по дням (для коротких интервалов). Повышенное влияние человеческого фактора, поскольку предполагает принятие определенных решений и предпочтений. Требует множественной адаптации модели к конкретному случаю

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
	<p>Характеризуется тем свойством, что, говоря нестрого, при фиксированном настоящем будущее независимо от прошлого.</p> <p>Построение целей Маркова основывается на определении вероятности того, что, находясь в одном состоянии, система, в результате определенного числа испытаний, перейдет к другому состоянию</p>		
<p>Методы <i>PLS-PM</i> (<i>Partial Least Squares Path Modeling</i>) [131, 138, 139]</p>	<p>Применяется для оценки тесноты связей между массивами данных. Используется также для формирования структурных моделей, характери-</p>	<p>Подходит для оценки неявных связей.</p> <p>Применим к данным высокой размерности</p>	<p>Практика показывает, что не всегда выявленные связи можно считать надежными.</p> <p>Метод чаще используется как промежуточный для дальнейшего прогнозирования.</p>

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
	зующих силу влияния факторов на сложную систему. Формируемая эндогенная модель описывается стандартной системой линейных уравнений		Содержит только оценку влияния, без дальнейшего моделирования изменения зависимой переменной
Методы имитационного моделирования Метод с применением неординарного марковского процесса [140, 141]	Метод основывается на марковском свойстве — переход системы из одного состояния в другое осуществляется в случайный момент времени и будущее состояние зависит только от достигнутого на данный момент и не зависит от развития системы в прошлом. Применяется для оценки вероятности	Данный метод не учитывает тренд временного ряда. В условиях влияния многочисленных факторов тенденция не всегда сохраняется на протяжении всего периода	Метод достаточно объемный по количеству расчетов. Основывается на случайности. Требуется самостоятельно определить количество возможных переходов между состояниями сложной системы для дальнейших расчетов. Для достижения конечного результата привлекается множество других способов моделирования: Монте-Карло, регрессионный анализ, алгоритм Гиллеспи

4.1. Обзор основных методов и моделей прогнозирования

Продолжение табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
Нейронные сети (рекуррентное прогнозирование) [142, 143]	<p>перехода между состояниями объекта исследования (например, в соответствии с зонной теорией), зависящей от степени влияния конструктивных факторов. Предполагается, что при достижении определенного уровня влияния конструктивных факторов интуитивность и вероятность переходов между состояниями снижается</p> <p>Новейшие технологии искусственного интеллекта, включающие автоматическое программирование, настроенное на обучение и принятие самостоятельных</p>	<p>Способность к моделированию нелинейных процессов.</p> <p>Низкие трудозатраты при применении.</p> <p>Работает с любыми</p>	<p>Уровень изучения искусственного интеллекта на данный момент не может гарантировать надежность результатов.</p> <p>Малое число научных исследований применения</p>

Окончание табл. 4.1

Наименование метода	Краткая характеристика	Преимущества метода	Недостатки метода
	<p>решений. Данный инструмент имеет неограниченные возможности, за счет взаимодействия с интернетом доступ к огромному массиву данных и позволяет выявить даже самые неоднозначные зависимости. Рекуррентной нейронной сетью называют модели нейронных сетей, содержащие рекуррентные блоки, которые запоминают информацию в процессе прохождения сигнала и, обрабатывая ее, передают следующим слоям</p>	<p>возможными массивами данных</p>	<p>нейронных сетей для прогнозирования. Результат их работы зависит от выбора исходных данных. Занимают много места на сервере: чем сложнее задача, которую решает нейронная сеть, тем больше ее объем</p>

Процесс прогнозирования реализуется последовательностью выполнения следующих этапов [144]:

1. Предпрогнозная ориентация. На данном этапе формулируется и представляется к описанию объект прогнозирования; определяется предмет исследования; выполняется постановка проблемы, определяется цель и задачи прогнозирования.

2. Первичное моделирование. На данном этапе предполагается построение некоторой начальной модели предмета исследования как системы, характеристика которой может быть представлена совокупностью количественных, качественных и структурных показателей. Также на этом этапе выдвигаются рабочие гипотезы, определяется ретроспектива (время начала прогнозирования) и перспектива (время упреждения прогноза), выбирается метод, структура и способ реализации прогнозного исследования.

3. Информационный этап. На данном этапе осуществляется сбор данных об изменениях в прогнозном фоне, выявлении преимуществ и недостатков прогнозируемой системы, сборе и накоплении статистики по развитию изучаемой системы и систем-аналогов.

4. Аналитический этап. На данном этапе осуществляется диагностика текущего состояния системы, определяются перспективы и тенденции ее развития, анализ проблем и выявление основных противоречий внутри системы, оценку возможностей системы и ее чувствительности к колебаниям прогнозного фона, анализ альтернативных путей развития системы, оценку необходимых ресурсов и их источников (при необходимости).

5. Вторичное моделирование. Данный этап ориентирован на построение нормативной и/или поисковой модели прогнозируемой системы.

6. Этап контроллинга. На данном этапе реализуется процедура верификации (проверки надежности) результатов, их сопоставление с иными моделями и предположениями экспертов.

7. Заключительный этап. Этап ориентирован на разработку рекомендаций по управлению и планированию изучаемой системы, оформление результатов в виде отчета и т.д.

4.2. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ МЕТОДА ANOVA В КОМБИНАЦИИ С МЕТОДОМ ФЕХНЕРА

Методы и модели прогнозирования получили широкое распространение для оценки прогнозных индикаторов развития социально-экономических систем.

Для практической реализации методов и моделей прогнозирования регламентируется следующая постановка задачи: сформировать прогноз индикаторов экономической безопасности региона в условиях воздействия процессов теневизации.

Среди представленных методов и моделей оценки наиболее подходящей для реализации указанной задачи совокупность трех методов: коэффициент корреляции Фехнера, дисперсионный анализ, регрессионный анализ. При проведении экономических исследований обработка статистических данных с использованием данных методов позволяет выявить наличие существующей взаимосвязи и дать им количественную оценку. Стоит отметить, что для достоверности полученных данных все показатели были приведены к единому виду, а именно ВРП на душу населения был представлен как ВРП в среднероссийском ВРП на душу, %. Также показатель «Преступления экономической направленности» представлен в % от общего числа преступлений.

В качестве результативных признаков (Y_i), т.е. зависимых переменных, при проведении анализа будут выступать индикаторы экономической безопасности региона. Факторными признаками (X_i), влияющими на результативные признаки, будут следующие показатели:

- зарегистрированные преступления экономической направленности в % от общего числа преступлений (X_1);

4.2. Практическая реализация метода ANOVA в комбинации с методом Фехнера

- занятые в неформальном секторе в % к общей численности занятого населения (X_2);
- доля теневой экономики в ВРП, %, согласно Модели 1 (X_3);
- доля теневой экономики в ВРП, %, согласно Модели 2 (X_4).

На первом этапе оценки определена сила связи коэффициентов корреляции Фехнера. Результаты представлены в табл. 4.2.

Таблица 4.2

Значение коэффициента Фехнера (итого по Сибирскому федеральному округу)

Показатели	X_1	X_2	X_3	X_4
Y_1	-0,5	0,0	0,5	0,5
Y_2	0,2	0,0	-0,2	-0,2
Y_3	0,8	0,3	-0,5	-0,5
Y_4	0,2	0,3	-0,2	-0,2
Y_5	0,2	0,0	0,2	0,2
Y_6	-0,3	-0,5	0,7	0,7
Y_7	0,5	0,0	0,2	0,2
Y_8	-0,5	-0,3	0,2	0,2
Y_9	-0,2	0,3	-0,5	-0,5
Y_{10}	0,0	0,2	0,0	0,0
Y_{11}	-0,2	0,0	0,2	0,2
Y_{12}	0,7	0,2	-0,7	-0,7
Y_{13}	-0,5	0,0	0,5	0,5
Y_{14}	0,5	0,3	-0,2	-0,2
Y_{15}	0,3	-0,2	0,0	0,0
Y_{16}	0,5	0,0	-0,2	-0,2
Y_{17}	0,0	-0,2	0,3	0,3
Y_{18}	-0,5	-0,3	0,8	0,8
Y_{19}	0,3	-0,2	-0,3	-0,3
Y_{20}	0,3	0,5	-0,3	-0,3

Для более достоверной оценки влияния был проведен дисперсионный анализ, который с помощью сопоставления фактического и критического значения Фишера, учитывая, что величина P -значение $< 0,05$, позволяет выявить наличие влияния показателей теневого сектора на показатели экономической безопасности регионов Сибирского федерального округа.

Результаты проведения дисперсионного анализа, определенные с использованием статистических данных за период 2010–2021 гг., представлены в табл. 4.3.

Таблица 4.3

**Результаты дисперсионного анализа
(итого по Сибирскому федеральному округу)**

Показатели	X_1	X_2	X_3	X_4
Y_1	+	+	+	+
Y_2	+	+	+	+
Y_3	+	+	+	+
Y_4	+	+	—	+
Y_5	+	+	+	+
Y_6	+	—	+	+
Y_7	+	+	+	+
Y_8	—	+	+	+
Y_9	+	+	+	+
Y_{10}	+	+	+	+
Y_{11}	+	+	+	+
Y_{12}	+	—	+	+
Y_{13}	+	—	+	+
Y_{14}	+	+	+	+
Y_{15}	+	+	+	+
Y_{16}	+	+	+	+
Y_{17}	+	+	+	+

4.2. Практическая реализация метода ANOVA в комбинации с методом Фехнера

Окончание табл. 4.3

Показатели	X_1	X_2	X_3	X_4
Y_{18}	+	+	+	+
Y_{19}	+	+	+	+
Y_{20}	+	+	+	+

Примечание: (+) – F -факт $> F$ критическое, величина P -значение $< 0,05$, следовательно, критерий Фишера значим, обнаружено статистически достоверное влияние; (–) – влияние отсутствует.

Для проведения на следующем этапе оценки регрессионного анализа были исключены показатели, по которым отсутствует статистически достоверное влияние. Проведение регрессионного анализа сводится к построению уравнения регрессии, параметры которого позволяют проанализировать связь между несколькими независимыми переменными (X_i) и зависимой переменной (Y_i). Построенные уравнения представлены в табл. 4.4.

Таблица 4.4

**Уравнения регрессии
(итога по Сибирскому федеральному округу)**

Показатели	X_1	X_2	X_3	X_4
Y_1	$81,56 - 0,1X_1$	$109,79 - 1,23X_2$	$75,6 + 0,2X_3$	$75,6 + 0,2X_4$
Y_2	$97,7 - 1,001X_1$	$98,5 - 0,1X_2$	$113,6 - 0,3X_3$	$111 - 0,3X_4$
Y_3	$88,8 - 4,3X_1$	$108,9 - 0,2X_2$	$142,6 - 0,9X_3$	$135,8 - 0,9X_4$
Y_4	$56,2 - 3,3X_1$	$95,7 - 2,3X_2$	–	$9,3 + 1,04X_4$
Y_5	$-1,02 + 0,2X_1$	$3,1 - 0,1X_2$	$-0,1 - 0,004X_3$	$-0,7 + 0,01X_4$
Y_6	$22,3 + 0,4X_1$	–	$12,7 + 0,3X_3$	$11,9 + 0,4X_4$
Y_7	$6,1 - 0,03X_1$	$-0,02 + 0,3X_2$	$4,8 + 0,03X_3$	$4,7 + 0,04X_4$
Y_8	$4,65 - 0,5X_1$	$3,8 - 0,03X_2$	$-2,2 + 0,1X_3$	$-1,2 + 0,1X_4$
Y_9	$19,3 + 0,1X_1$	$5,9 + 0,6X_2$	$22,8 - 0,1X_3$	$22,8 - 0,1X_4$
Y_{10}	$0,6 - 0,001X_1$	$0,54 + 0,002X_2$	$0,57 + 0,001X_3$	$0,58 + 0,0004X_4$

Окончание табл. 4.4

Показатели	X_1	X_2	X_3	X_4
Y_{11}	$1,003 - 0,03X_1$	$1,3 - 0,02X_2$	$0,6 + 0,01X_3$	$0,7 + 0,01X_4$
Y_{12}	$16,3 + 1,6X_1$	–	$50,1 - 0,7X_3$	$48,07 - 0,8X_4$
Y_{13}	$48,4 - 6,7X_1$	–	$-52,7 + 1,9X_3$	$-42,9 + 2,01X_4$
Y_{14}	$83,3 + 5,6X_1$	$-3,8 + 4,7X_2$	$178,5 - 1,9X_3$	$166,4 - 1,9X_4$
Y_{15}	$5,6 + 0,5X_1$	$2,3 + 0,2X_2$	$9,9 - 0,1X_3$	$9,3 - 0,1X_4$
Y_{16}	$18,03 + 0,4X_1$	$62,1 + 0,2X_2$	$156,9 - 0,97X_3$	$142,6 - 0,5X_4$
Y_{17}	$19,8 - 0,3X_1$	$27,9 - 0,4X_2$	$11,9 + 0,2X_3$	$12,4 + 0,2X_4$
Y_{18}	$72,3 - 0,9X_1$	$90,1 - 0,9X_2$	$58,1 + 0,3X_3$	$59,2 + 0,3X_4$
Y_{19}	$58,3 + 0,5X_1$	$48,7 + 0,5X_2$	$62,2 - 0,1X_3$	$62,4 - 0,1X_4$
Y_{20}	$94,9 + 1,5X_1$	$74,3 + 1,2X_2$	$126,5 - 0,6X_3$	$123,5 - 0,7X_4$

Представленная интерпретация результатов проведенного регрессионного анализа, доказывает, что теневой сектор оказывает воздействие на показатели экономической безопасности.

Таким образом, в ходе проведения комплексной оценки на основе применения ряда методов (коэффициент корреляции Фехнера, дисперсионный анализ, регрессионный анализ) было доказано наличие влияния индикаторов теневой составляющей экономики на показатели экономической безопасности.

Для полноценной оценки процесса влияния необходимо реализовать процедуру прогнозирования на основе сформированных моделей.

Одной из наиболее распространенных моделей прогнозирования в условиях влияния факторов является линейная регрессия, которая использует статистические методы для определения связи между двумя переменными. Другие модели, такие как модели временных рядов и нейронные сети, также могут быть применимы. Однако, как и в случае с другими математическими моделями, прогнозы, полученные с помощью мо-

делей прогнозирования под воздействием влияния факторов, могут быть подвержены ошибочным расчетам и не всегда являются точными. Поэтому перед принятием решений на основе полученных прогнозов, необходимо проводить дополнительный анализ и оценку рисков. Кроме того, для достижения более точных прогнозов целесообразно использовать комбинацию различных моделей и методов прогнозирования.

Для реализации прогнозных процедур (формализация процессов влияния теневого сектора на показатели экономической безопасности) с применением инструментов программирования (*Python*) анализировались две модели, такие как *Wavelet ARIMA* и множественная линейная регрессия. При использовании модели *Wavelet ARIMA* необходимо учитывать большой массив данных, например, если это данные, представленные за месяц, то можно использовать данные за последние 2–3 года. Если это данные, представленные за год, то можно использовать данные за последние 10 лет. Но определение конкретного периода также зависит от того, как часто появляются новые данные и какой временной период заложен в прогноз. Следует учитывать, что чем больше данных будет использовано для обучения модели, тем более точные будут прогнозные значения [148].

В связи с большей вероятностью получения прогноза с недостоверными значениями из-за недостаточного массива данных целесообразным будет использование такой модели, как множественная линейная регрессия с использованием языка программирования *Python*. Во избежание ошибок при использовании данной модели необходимо доказать наличие зависимости между индикаторами экономической безопасности и индикаторами теневого сектора. Для этих целей реализован анализ коэффициентов Фехнера и дисперсионный анализ, с помощью которых были исключены данные, по которым отсутствовала взаимосвязь.

Алгоритм реализации моделирования и прогнозирования включает следующую последовательность этапов [149]:

1. Подготовка данных для импортирования необходимых библиотек, выполнение операторов импорта и загрузка статистических данных (рис. 4.1).

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

# Загрузка данных
data = pd.read_csv('data.csv')
```

Рис. 4.1. Операторы импорта данных в Python

2. Определение факторов влияния и целевых переменных для прогнозирования. В данном случае в качестве переменной, оказывающей влияние, будет выступать индикатор «доля теневой экономики согласно Модели 2», в качестве результирующего фактора будет выступать индикатор (интегральное значение показателя экономической безопасности), за исключением тех показателей, по которым было доказано отсутствие взаимосвязи (рис. 4.2).

```
# Преобразование даты в числовой формат
data['Date'] = pd.to_datetime(data['Date'])
data['Date'] = data['Date'].dt.strftime('%Y%m%d')

# Определение факторов и целевой переменной
X = data[['Date', 'Shadow']]
y = data['GRP']
```

Рис. 4.2. Определение факторов для обучения модели

3. Формирование и обучение модели множественной линейной регрессии по заданным параметрам на определение влияния процессов теневизации (рис. 4.3).

4. Формирование прогнозных значений показателей экономической безопасности региона, для каждого прогнозируемого года создается аналогичный код (рис. 4.4).

4.2. Практическая реализация метода ANOVA в комбинации с методом Фехнера

```
# Создание модели множественной линейной регрессии
model = LinearRegression()

# Обучение модели на всех данных
model.fit(X, y)
```

Рис. 4.3. Формирование модели множественной линейной регрессии

```
# Создание даты для прогноза
timeforprognoz22='2022-12-31'
timeforprognoz23='2023-12-31'
timeforprognoz24='2024-12-31'
timeforprognoz25='2025-12-31'
future_date = pd.to_datetime(timeforprognoz22)
future_date=int(future_date.strftime('%Y%m%d'))
# Создание массива с факторами для прогноза
future_factors = np.array([[future_date, 0]])
# Прогнозирование целевой переменной
futureprediction = model.predict (future_factors)
```

Рис. 4.4. Прогнозирование показателей экономической безопасности региона (на примере Красноярского края)

5. Вывод полученных результатов на график с отображением прогнозных значений (рис. 4.5).

```
# Вывод результата прогнозирования
print('Прогноз на 'timeforprognoz22+':',
future_prediction)
new_row = {'Date': '20221231', 'GRP':
int(future_prediction),
'Shadow': '0'}
data = data.append (new_row, ignore_index=True)
print(data)

X = data['Date']
y = data['GRP']
pit.figure(figsize=(9, 6))
pit.plot(X.str.slice(0, 4), y/1000, 'purple')
plt.xlabel("Год")
plt.ylabel("ВВП, руб.")
```

Рис. 4.5. Вывод результата прогнозирования

После воспроизведения модели множественной регрессии и вывода прогнозных значений формируется таблица, отображающая прогнозные значения показателей экономической безопасности (табл. 4.5).

Таблица 4.5

Прогнозные значения показателей экономической безопасности под влиянием доли теневого сектора, 2022–2025 гг. (на примере Красноярского края)

Показатели	Прогнозные значения			
	2022	2023	2024	2025
ВРП на душу населения, руб. (Y_1)	1245991	1315110	1384228	1453347
Темп роста ВРП к предыдущему году в сопоставимых ценах, % (Y_2)	104,5	104,2	103,9	103,6
Индекс промышленного производства, % (Y_3)	106,3	104,4	102,6	100,7
Степень износа ОФ, % (Y_4)	48,7	49,7	50,7	51,6
Дефицит/профицит консолидированного регионального бюджета, % к ВРП (Y_5)	3,9	4,1	4,4	4,6
Доля экспорта в ВРП, % (Y_6)	16,6	15,7	14,9	14,0
Доля импорта в ВРП, % (Y_7)	6,5	6,7	6,9	7,0
Доля инновационной продукции в процентах от общего объема отгруженных товаров, выполненных работ, услуг, % (Y_8)	1,8	1,9	2,1	2,2
Доля продукции высокотехнологичных и наукоемких отраслей в ВРП, % (Y_9)	9,8	9,5	9,2	8,9

4.2. Практическая реализация метода ANOVA в комбинации с методом Фехнера

Окончание табл. 4.5

Показатели	Прогнозные значения			
	2022	2023	2024	2025
Численность занятых научными исследованиями и разработками, в % к занятым в экономике (Y_{10})	0,55	0,56	0,57	0,58
Внутренние затраты на научные исследования и разработки, в % ВРП (Y_{11})	0,77	0,78	0,79	0,8
Доля иностранных инвестиций в инвестициях в основной капитал, в % (Y_{13})	105,2	113,5	121,9	130,2
Индекс физического объема инвестиций в основной капитал в совокупных ценах, % к предыдущему году (Y_{14})	119,1	118,9	118,7	118,6
Уровень безработицы, % (Y_{15})	4,9	4,8	4,8	4,7
Уровень преступлений на 100 тыс. человек населения (Y_{16})	1511	1440	1369	1298
Доля в населении людей, имеющих доходы ниже прожиточного минимума, % (Y_{17})	16,9	16,9	16,9	17,0
Продолжительность жизни населения, лет (Y_{18})	68,6	68,8	68,9	69,0
Уровень занятости населения, % (Y_{19})	56,7	56,1	55,5	54,8
Реальные доходы населения, % (Y_{20})	102,9	102,8	102,7	102,6

Исходя из прогнозных значений выявлено: 1) процессы теневизации оказывают деструктивное влияние на показа-

тели экономической безопасности; 2) 8 из 19 анализируемых показателей в прогнозном периоде показали отрицательную динамику; 3) в финансово-экономической сфере такие показатели, как «Темп роста ВРП к предыдущему году в сопоставимых ценах, %» и «Доля экспорта в ВРП, %», будут снижаться, «Степень износа ОФ, %» — увеличивается; 4) в инновационной сфере «Доля продукции высокотехнологичных и наукоемких отраслей в ВРП, %» характеризуется резким снижением (на 23,4%); 5) в инвестиционной сфере показатель «Доля иностранных инвестиций в инвестициях в основной капитал, в %» увеличится более чем в 2 раза, что приведет к отклонению от порогового значения на 85% в первый прогнозный период, и дальше продолжит наращивать процент отклонения; 6) по показателям социальной сферы произойдет увеличение «Доли в населении людей, имеющих доходы ниже прожиточного минимума», «Уровень занятости населения, %» будет снижаться и достигнет 54,8% на конец прогнозного периода, что на 8,67% меньше порогового значения.

Графическое представление прогнозных значений показателей экономической безопасности (на примере Красноярского края, выборка показателей) представлено на рис. 4.6–4.10.

Показатель ВРП на душу населения продолжает расти в прогнозном периоде, несмотря на то, что по результатам регрессионного анализа наблюдается снижение показателя на 0,9% при росте объемов теневого сектора на 1%. Однако стоит учитывать влияние прочих факторов, которые могут сказываться на изменениях показателя экономической безопасности. Предположительно, без влияния теневого сектора показатель ВРП мог быть на уровень выше. Следовательно, теневой сектор оказывает деструктивное влияние на показатель ВРП. По данному прогнозу, с 2021 по 2022 год ВРП вырос на 13,77%, в последующий период рост показателя составит 14,27%.

4.2. Практическая реализация метода ANOVA в комбинации с методом Фехнера

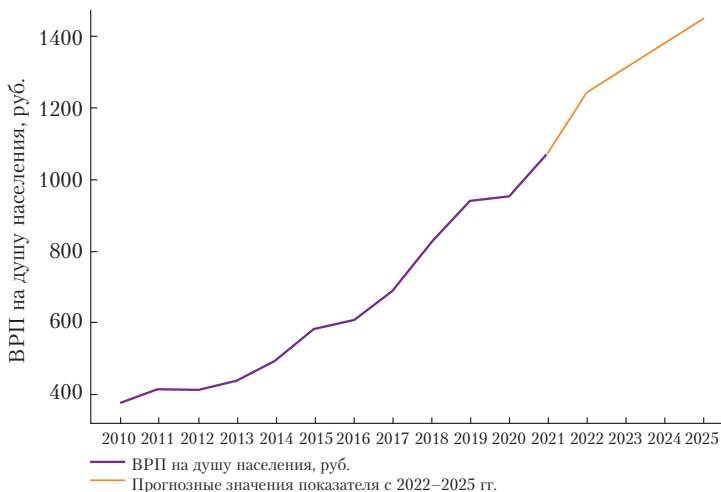


Рис. 4.6. Сценарий изменений индикатора «ВРП на душу населения, руб.» под влиянием теневого сектора (на примере Красноярского края)

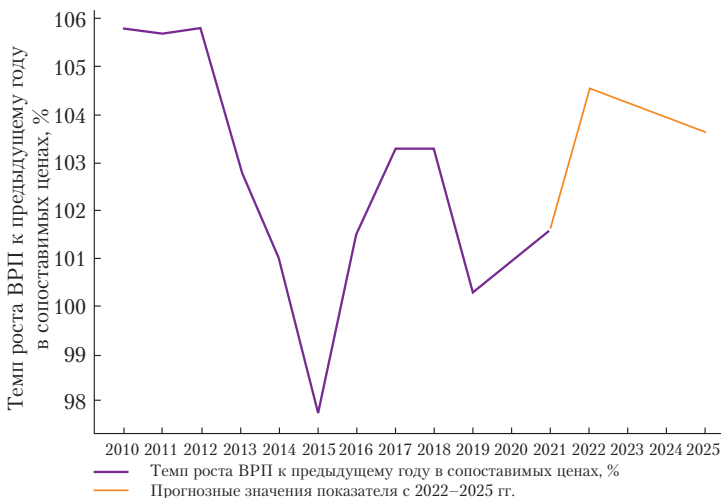


Рис. 4.7. Сценарий изменений индикатора «Темп роста ВРП к предыдущему году в сопоставимых ценах, %» под влиянием теневого сектора (на примере Красноярского края)

Результаты моделирования влияния процессов теневизации указывают на резкое увеличение показателя «Темп роста ВРП» в первом прогнозном году с 101,6% до 104,5%, однако далее показатель демонстрирует снижение, не достигнув нормативного значения.



Рис. 4.8. Сценарий изменений индикатора «Доля продукции высокотехнологичных и наукоемких отраслей в объеме ВРП, %» под влиянием теневого сектора (на примере Красноярского края)

В разрезе инновационной сферы наблюдается сокращение показателя по прогнозным данным. На начало прогнозного периода показатель «Доля продукции высокотехнологичных и наукоемких отраслей в объеме ВРП» снизилась на 23,4%. Снижение показателя продолжается в течение всего прогнозного периода, в целом к 2025 году изменение составит (-3,9%). В отношении иностранных инвестиций отмечается рост в целом на 84,6%; на конец прогнозного периода показатель будет отклоняться от порогового значения на 110,2%.

4.2. Практическая реализация метода ANOVA в комбинации с методом Фехнера

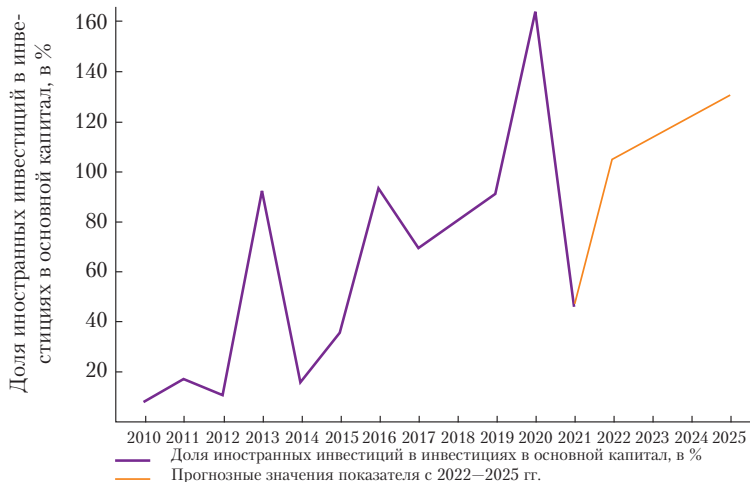


Рис. 4.9. Сценарий изменений индикатора «Доля иностранных инвестиций в общем объеме инвестиций в основной капитал, в %» под влиянием теневого сектора (на примере Красноярского края)

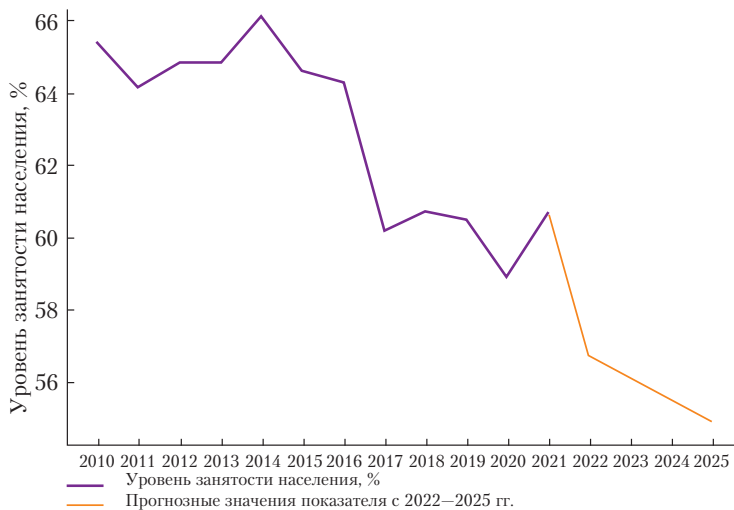


Рис. 4.10. Сценарий изменений индикатора «Уровень занятости населения, %» под влиянием теневого сектора (на примере Красноярского края)

Показатель уровня занятости населения в Красноярском крае за весь анализируемый период снижается, однако находится в пределах нормативного значения. По прогнозным данным, также происходит снижение, которое выходит за пределы порогового значения. На конец 2025 года фиксируется уменьшение показателя на 5,9%.

Подводя итог проведенной процедуры моделирования и прогнозирования, отмечается наличие зависимости индикаторов социально-экономического развития от параметров теневого сектора. Доказано, что показатели процессов теневизации воздействуют на показатели экономической безопасности; отражено количественное изменение данных показателей в аспекте влияния теневого сектора в прогнозном периоде. Ухудшение прогнозных значений наблюдается не по всем показателям, что обусловлено следующим: процедура анализа и прогноза охватывала достаточно небольшой массив данных. Однако по результатам моделирования и прогнозирования выявлено: теневой сектор в той или иной мере расстраивает стабильное функционирование всей системы экономических отношений, ухудшает социально-экономическое развитие не только в отдельно взятом регионе, но и в округе в целом.

4.3. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ МЕТОДА ХОЛЬТА — ВИНТЕРСА В КОМБИНАЦИИ С ИНСТРУМЕНТАМИ ПРОГРАММИРОВАНИЯ

Методы и модели прогнозирования получили широкое распространение для оценки прогнозных индикаторов развития социально-экономических систем.

Для практической реализации методов и моделей прогнозирования регламентируется следующая постановка задачи: сформировать прогноз индикаторов продовольственной безопасности региона и состояния продовольственных систем на основе применения метода Хольта — Винтерса в комби-

нации с инструментами программирования (использование языка программирования *Python*).

Отмечается, что автоматизированные системы расчета позволяют формализовать и упростить процедуру оценки параметров состояния продовольственных систем и продовольственной безопасности в условиях неопределенности, позволяя заблаговременно моделировать и регулировать проблемы, связанные с экономической, физической, транспортной, нутрициологической безопасностью продовольствия.

Проведенный анализ теоретических подходов к процедуре моделирования и прогнозирования временных рядов в теоретическом блоке данного исследования позволил сформировать вывод о том, что наиболее распространенными являются методы прогнозирования «оптимистически-сценарного» подхода, поскольку алгоритм данных методов исходит из того, что совокупность негативных факторов не изменится.

Среди таких методов оптимальным является модель Хольта — Уинтерса (Винтерса). Данная модель может быть использована для прогнозирования уровня продовольственной безопасности региона и состояния продовольственных систем (на основе прогнозирования оценочных индикаторов и пересчета интегральных показателей) в неизменных условиях или без учета влияния факторов. В частности, модель может быть использована для пересчета и наблюдения динамики функциональных связей между плановыми показателями (которые используются в качестве факторов) и спрогнозированными индикаторами функционирования продовольственных систем (или индикаторами продовольственной безопасности).

На основании данного метода был сформирован «Программный модуль идентификации и прогнозирования рисков и угроз продовольственной безопасности арктических регионов на основе модели Холта — Винтерса» [150]. Основной целью программного модуля является идентификация и прогнозирование рисков и угроз продовольственной безопасности по критерию «физическая доступность», «экономическая

доступность», «достаточность продовольствия» для арктических регионов. Математической базой для прогноза является модель анализа временных рядов, которая выявляет и прогнозирует трендовую компоненту нестационарного временного ряда на основе экспоненциального сглаживания. Интерфейс программы представлен на рис. 4.11.

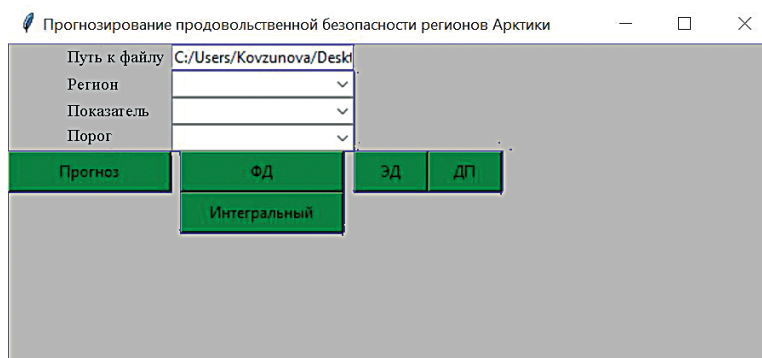


Рис. 4.11. Интерфейс программы для ЭВМ «Программный модуль идентификации и прогнозирования рисков и угроз продовольственной безопасности арктических регионов на основе модели Холта — Винтерса»

В результате использования данного модуля осуществляется процедура моделирования частных показателей, частных критериев и интегрального показателя продовольственной безопасности до 2025 года. Методология создания программы ЭВМ представлена функциональной моделью в виде контекстной диаграммы системных процессов (табл. 4.6).

«На входе» контекстной диаграммы аккумулируются исходные статистические данные, отражающие ключевые показатели регионов Арктической зоны Российской Федерации в разрезе трех критериев (физическая доступность продовольствия, экономическая доступность, достаточность потребления продовольствия), а также пороговые значения

для каждой группы критериев с учетом арктической специфики. «На выходе» формируются результаты оценки уровня продовольственной безопасности регионов Арктической зоны Российской Федерации в формате итогового интегрального показателя, а также частных критериев оценки. Логика взаимосвязи показателей и критериев представлена в иерархии связи (рис. 4.12).

Таблица 4.6

Содержание процессов контекстной диаграммы

Процесс	Содержание процесса
«Вход» (<i>Input</i>)	Исходная информация: статистические данные, отражающие ключевые показатели регионов Арктической зоны Российской Федерации в разрезе пяти критериев: физическая доступность продовольствия, экономическая доступность, достаточность потребления продовольствия
«Выход» (<i>Output</i>)	Оценка уровня продовольственной безопасности регионов Арктической зоны Российской Федерации в формате частных критериев оценки, а также итогового интегрального показателя
«Механизм» (<i>Mechanism</i>)	Вычислительная техника и операционная система — ресурс, необходимый для работы с информационной системой, т.к. все процессы являются компьютеризированными

Для формирования выводов и описания результатов моделирования необходимо конкретизировать условные обозначения, входящие в описательную часть графиков и обозначить пороговые значения для частных показателей (табл. 4.7).

В рамках проведения процедуры моделирования прогнозных значений были спрогнозированы частные и интегральные показатели продовольственной безопасности в целом по регионам, входящим в состав Арктической зоны Российской Федерации, результаты представлены на рис. 4.13 и 4.14.

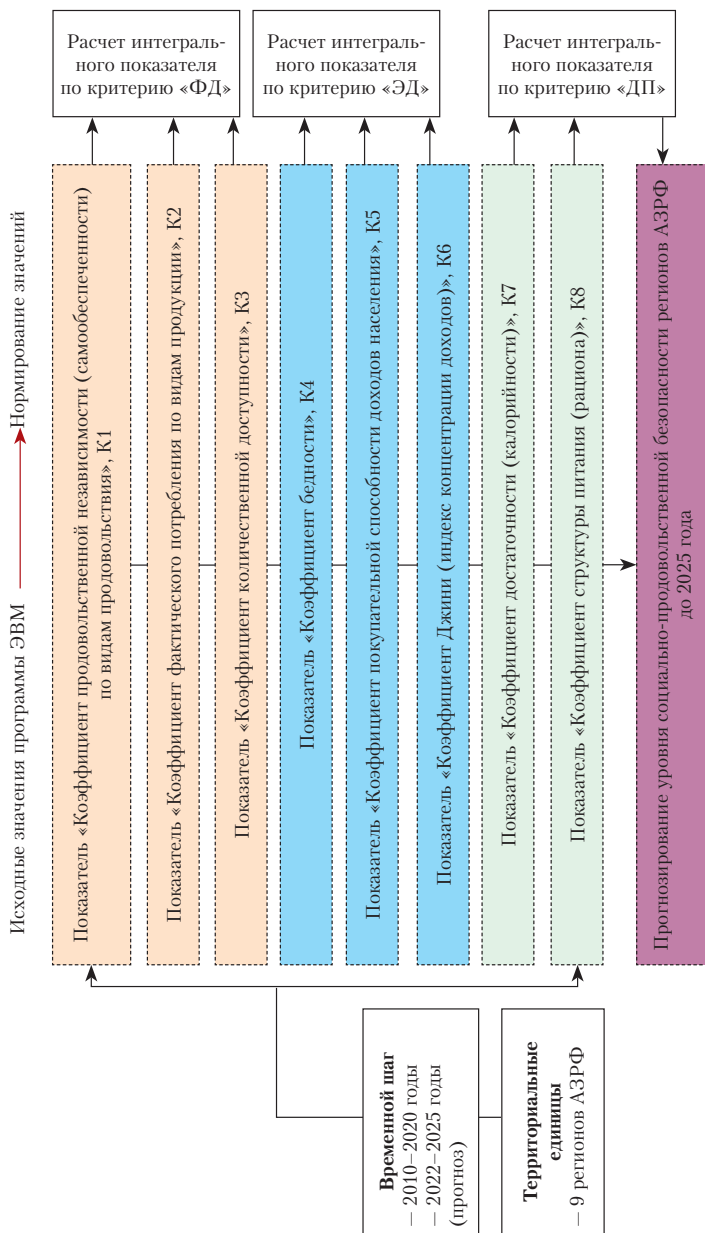


Рис. 4.12. Иерархия связи программы для ЭВМ «Программный модуль идентификации и прогнозирования рисков и угроз продовольственной безопасности арктических регионов на основе модели Холта — Винтерса»

Таблица 4.7

**Условные и цветовые обозначения графиков
моделирования**

Наименование	Интерпретация
К1	Коэффициент продовольственной независимости (самообеспеченности) по видам продовольствия
К2	Коэффициент фактического потребления по видам продукции
К3	Коэффициент количественной доступности
К4	Коэффициент бедности
К5	Коэффициент покупательной способности доходов населения
К6	Коэффициент Джини (индекс концентрации доходов)
К7	Коэффициент достаточности (калорийности)
К8	Коэффициент структуры питания (рациона)
ФД	Значения по критерию «Физическая доступность продовольствия»
ЭД	Значения по критерию «Экономическая доступность продовольствия»
ДП	Значения по критерию «Достаточность продовольствия»
	Значение показателя не достигает порогового параметра
	Значение показателя достигает порогового параметра

По результатам моделирования в прогнозируемых периодах три частных показателя не достигают пороговых значений: это К1 в 2022 году, К7 и К8 на протяжении всего периода. Низкие значения показателя продовольственной независимости (самообеспеченности) обусловлены территориальными и кли-

матическими условиями регионов Арктики, что ограничивает возможности местных сельскохозяйственных предприятий производить продукцию в необходимых и достаточных объемах. Следовательно, регионы Арктической зоны зависимы от ввоза продовольственных товаров (территориальная отдаленность, неразвитая транспортная инфраструктура препятствуют своевременным поставкам продовольствия и ограничивают ввоз скоропортящихся продуктов, например овощей и фруктов).

Как следствие, недостаточная физическая доступность пищевых продуктов негативно сказывается на коэффициентах достаточности питания, что проявляется в недоедании, низкой энергетической ценности фактического рациона и не соответствующей медицинским нормам структуре питания.

В свою очередь, значения интегрального показателя достаточности продовольствия не достигают порогового значения и негативно отражаются на уровне продовольственной безопасности регионов АЗРФ. В связи с этим необходима корректировка региональных политик, направленная на стимулирование сельскохозяйственных предприятий регионов к производству и ввозу необходимого продовольствия в достаточных объемах.

Совокупный показатель уровня продовольственной безопасности Арктической зоны Российской Федерации превышает пороговое значение и находится в безрисковой зоне за счет высоких показателей экономической и физической доступности продовольствия. В целом к 2025 году прогнозируется рост показателя продовольственной безопасности и критериев «ФД» и «ЭД». Данная динамика является результатом финансовой поддержки как агропредприятий, так и населения, за счет чего наращиваются объемы производства и ввоза продовольствия и улучшается благосостояние населения регионов в Арктической зоне Российской Федерации.

4.3. Практическая реализация метода Хольта — Винтерса...

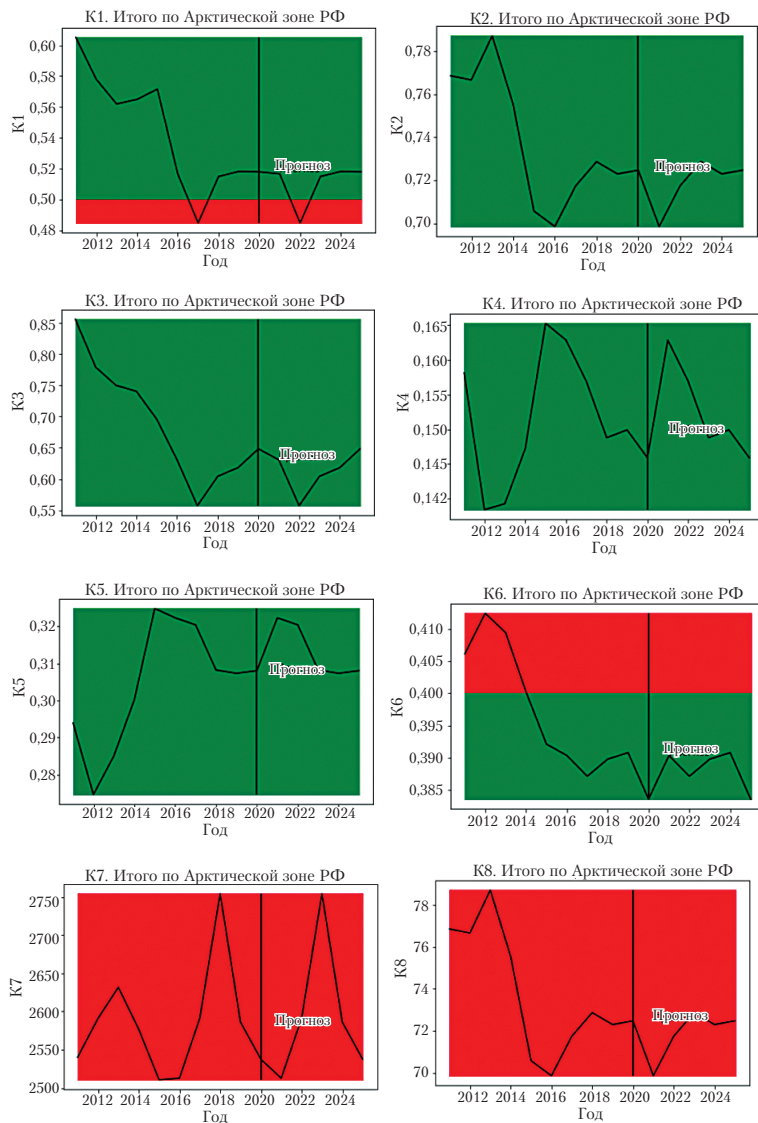


Рис. 4.13. Прогнозирование значений частных коэффициентов (K1–K8) оценки продовольственной безопасности регионов Арктической зоны РФ до 2025 года

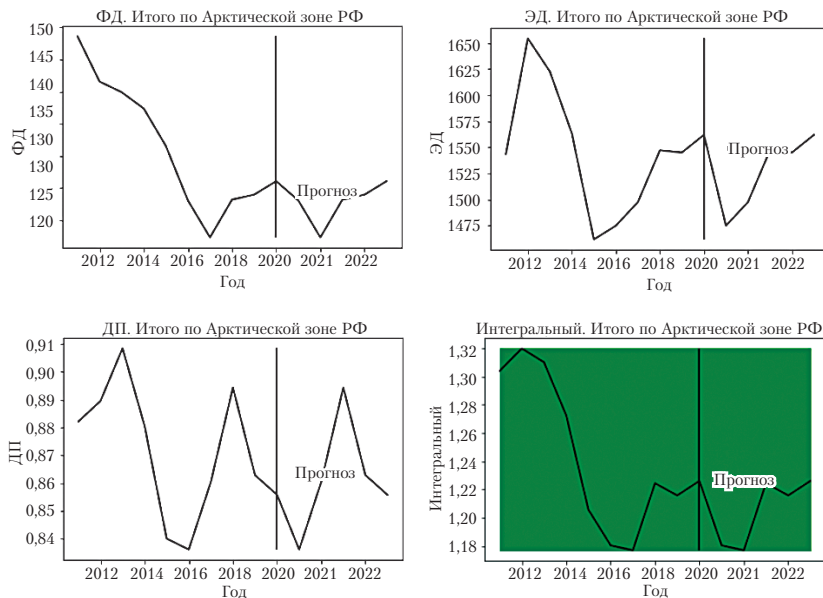


Рис. 4.14. Прогнозирование значений по критериям «ФД», «ЭД», «ДП» и интегрального показателя продовольственной безопасности регионов Арктической зоны РФ до 2025 года

Таким образом, модель Хольта – Уинтерса (Винтерса) доказала свою применимость для прогнозирования основных параметров социально-экономических систем в неизменных условиях или без учета влияния факторов. В частности, модель использована для пересчета и наблюдения динамики функциональных связей между плановыми показателями (которые используются в качестве факторов) и спрогнозированными индикаторами функционирования продовольственных систем (или индикаторами продовольственной безопасности).

Таким образом, прогнозирование позволяет определять некоторые тенденции в соответствии с установленными отклонениями от количественного баланса в сложных системах

(социально-экономических и производственных). Однако попытки прогнозирования долгосрочного развития сложных систем, где преобладает влияние не количественных, а качественных процессов, до сих пор остаются малорезультативными. Этим же фактом объясняются возникающие в долгосрочном прогнозировании трудности с использованием моделей. Без них невозможно дать количественную характеристику рассчитываемым соотношениям и зависимостям. Тем не менее при возникновении качественного аспекта прогнозные модельные рамки, разработанные для реализации сложных, но изначально заданных в определенном направлении расчетов, начинают проявлять свою несостоятельность, объясняемую статичностью математического инструментария.

Заключение

Учитывая актуальность исследования, которая обусловлена необходимостью развития методической базы и математического аппарата управления сложными системами в условиях роста и бурного развития систем искусственного интеллекта, а также повышением общих требований к экономической эффективности, достигаемой, в том числе, за счет применения оптимизационных моделей, в рамках работы получены основные теоретические и прикладные результаты на основе применения методов системного анализа, исследования операций, теории оптимизации, теории вероятностей, параллельного программирования.

Исследованы теоретические аспекты применения метода Монте-Карло в системе имитационного моделирования. Предложена практическая реализация метода Монте-Карло для решения бизнес-задач, а также для принятия управленческих решений в социально-экономических системах.

Исследованы теоретические и методические аспекты метода многокритериальной оптимизации *Data Envelopment Analysis*. Представлено экспериментальное применение метода *Data Envelopment Analysis* для решения задачи оценки эффективности региональной системы здравоохранения, а также для решения задачи оценки эффективности региональной инновационной системы.

Исследованы основные теоретические и методологические подходы к решению задач группировки объектов и данных. Представлено экспериментальное применение задачи кластеризации и задачи классификации в анализе данных сложных систем.

Представлен обзор основных методов и моделей прогнозирования. Представлена практическая реализация метода ANOVA в комбинации с методом Фехнера, а также практическая реализация метода Хольта — Винтерса в комбинации с инструментами программирования.

Список использованных источников

1. Божко, Л.М. Использование метода Монте-Карло в имитационном моделировании экономических систем // Интеллектуальные технологии на транспорте. — 2023. — № 1. — С. 38–41. — DOI: 10.24412/2413–2527-2023–133-38–42
2. Безруков, А.И. Математическое и имитационное моделирование: Учебное пособие / А.И. Безруков, О.Н. Алексенцева. — Москва: ИНФРА-М, 2019. — 227 с.
3. Кремер, Н.Ш. Теория вероятностей и математическая статистика: Учебник и практикум для вузов. — 5-е изд., перераб. и доп. — Москва: Юрайт, 2023. — 538 с.
4. Метод Монте-Карло и инструментальные средства его реализации: методические указания / А.В. Раменская, К.В. Пивоварова; Оренбургский гос. ун-т. — Оренбург: ОГУ, 2018. — 58 с.
5. Раменская, А.В. Анализ эффективности инвестиционного проекта создания комбикормового предприятия методом Монте-Карло / А.В. Раменская, Т.П. Негорожина // Институциональные и инфраструктурные аспекты развития различных экономических систем: сборник статей международной научнопрактической конференции: в 2 частях. — 2017. — С. 89–92.
6. Лычкина, Н.Н. Имитационное моделирование экономических процессов: Учебное пособие. — Москва: ИНФРА-М, 2022. — 254 с.
7. Фокин, И.В. Модель прогнозирования денежных потоков методом Монте-Карло для IT отрасли // Цифровая экономика и финансы: Материалы Международной научнопрактической конференции (Санкт-Петербург, Россия, 17–18 марта 2022 г.). — Санкт-Петербург: Астерион, 2022. — С. 284–289.

8. Ковзунова, Е.С. Концептуальный подход к оценке продовольственной безопасности макрорегиона на основе использования методов математического моделирования / Е.С. Ковзунова, И.Р. Руйга // Продовольственная политика и безопасность. — 2022. — Т. 9, № 2. — С. 177–196. — DOI 10.18334/ppib.9.2.114467.
9. Вардомацкая, Е.Ю. Имитационное моделирование инвестиционных рисков методом Монте-Карло / Е.Ю. Вардомацкая, П.С. Асоблева // Материалы и технологии. — 2022. — № 1 (9). — С. 50–57. — DOI: 10.24412/2617-149X-2022-1-50-57.
10. Голубенко, Е.В. Использование метода имитационного моделирования для диверсификации инвестиций в регионе / Е.В. Голубенко, В.В. Дегтярев, И.М. Магеррамов // Транспорт: наука, образование, производство (Транспорт-2022): Труды Международной научно-практической конференции (Ростов-на-Дону, Россия, 25–27 апреля 2022 г.). — Т. 1. — Ростов-на-Дону: Ростовский гос. ун-т путей сообщения, 2022. — С. 75–78.
11. Торопова, И.В. Оценка ожидаемых кредитных убытков методом Монте-Карло / И.В. Торопова, Е.А. Перминов, Н.И. Попова, О.В. Котова // Финансовый бизнес. 2022. № 3 (225). С. 94–98
12. Первун, О.Е. Технологии оценки затрат проекта методом Монте-Карло в среде программирования R // Ученые записки Крымского инженерно-педагогического университета. — 2022. — № 2 (76). — С. 149–154. — DOI: 10.34771/UZSERU.2022.2.76.030.
13. Саадалов, Т.Методика расчета коэффициента корреляции Фехнера и Пирсона, и их области применения / Т.Саадалов, Р.Мырзаibraимов, Ж.Д. Абдуллаева // Бюллетень науки и практики. — 2021. — Т. 7, № 10. — С. 270–276. — DOI 10.33619/2414-2948/71/31.
14. Борисов, В.В., Луферов В.С. Метод многомерного анализа и прогнозирования состояния сложных систем и про-

- цессов на основе нечетких когнитивных темпоральных моделей // Системы управления, связи и безопасности. — 2020. — № 2. — С. 1–23. DOI: 10.24411/2410–9916-2020–10201.
15. Звонарев, С.В. Основы математического моделирования: учебное пособие. — Екатеринбург: Изд-во Урал. ун-та, 2019. — 112 с. — URL: https://elar.urfu.ru/bitstream/10995/68494/1/978-5-7996-2576-4_2019.pdf (дата обращения: 12.11.2023).
 16. Каменский, Д.А. Применение моделей векторной авторегрессии при прогнозировании в финансах и экономике // Фундаментальные исследования. — 2019. — № 5. — С. 45–49.
 17. Bryon J.Parman & Allen M.Featherstone. A comparison of parametric and nonparametric estimation methods for cost frontiers and economic measures // Journal of Applied Economics. — 2019. — pp. 60–85
 18. Christopher F.Parmeter. Is it MOLS or COLS? Economic discussion papers. — 2021. — 26 p.
 19. Святохо, Н.В. Возможность использования метода нечеткой логики для оценки продовольственной безопасности региона // Материалы XII межрегиональной научно-практической конференции «интеграция республики Крым в систему экономических связей российской федерации: теория и практика управления». — 2016. — С. 296–298.
 20. Звягин Л.С. Теория нечетких множеств в современной экономике // Экономика и управление: проблемы, решения. — 2019. — № 6. — С. 29–35.
 21. Лисицына Л.С. Основы теории нечетких множеств // СПб: Университет ИТМО. — 2020. — 74 с.
 22. Momen, Mahyar and Shirinbakhsh, Mehrdad and Baniassadi, Amir and Behbahaninia, Ali. Application of Monte Carlo Method in Economic Optimization of Cogeneration

- Systems – Case Study of the CGAM System // Applied Thermal Engineering. – 2016. – pp. 1–23.
23. Paolo Brandimarte. Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics // Wiley. – 2014. – p. 688.
 24. Ковзунова, Е.С. Возможности использования методов линейного программирования для оценки продовольственной безопасности региона / Е.С. Ковзунова // В сборнике: Стратегия и тактика социально-экономических реформ: национальные приоритеты и проекты. Материалы IX Всероссийской научно-практической конференции с международным участием. – 2021. – С. 71–75.
 25. Масленников, Д.А. Выявление особенностей стратегического развития регионов на основе статистического анализа индикаторов / Д.А. Масленников, С. Н Митяков, Л.Ю. Катаева, Т.А. Федосеева // Экономика региона. – 2019. – № 3. – С. 707–719.
 26. Ковалев, Д.И. Обзор подходов и методов к оценке сравнительной эффективности технологических процессов и производств / Д.И. Ковалев, М.Ф. Козлова, О.И. Ольшевская, Т.П. Мансурова // Современные инновации, системы и технологии. – 2021. – № 1(3). – С. 5–25.
 27. Руйга, И.Р. Возможности использования метода Data Envelopment Analysis для оценки уровня инновационной устойчивости региона / И.Р. Руйга // Решетневские чтения. – 2017. – Т. 2. – С. 452–453.
 28. Туманянц, К.А. Оценка эффективности социальных расходов регионов РФ с помощью DEA-метода / К.А. Туманянц, Ю.Е. Сесина // Региональная экономика: теория и практика. – 2018. – № 1 (448).
 29. Fried H.O., Lovell C.A.K., Schmidt S.S. (Eds.) The Measurement of Productive Efficiency and Productivity Growth // Oxford University Press, 1993. 653 p.
 30. Pokushko M et al 2019 Application of data envelopment analysis method for assessment of performance of enterprises

- in fuel and energy complex *Journal of Physics: Conf. Ser.* 1351012140.
31. Pokushko M et al 2020 Applying the Data Envelopment Analysis method for evaluating the efficiency of the complex system operations in fuel and energy companies *Journal of Physics: Conf. Ser.* 1515052039.
 32. Stepan Zemtsov, Maxim Kotsemir. An assessment of regional innovation system efficiency in Russia: the application of the DEA approach. *Scientometrics* (2019) 120, 375–404.
 33. Ruiga I R et al 2019 Practical implementation of Data Envelopment Analysis technology to assess the innovative sustainability of resource-type regions *J.Phys.: Conf. Ser.* 1399033118.
 34. Руйга, И.Р. Сравнительная оценка эффективности функционирования системы здравоохранения сибирских регионов на основе метода Data Envelopment Analysis / И.Р. Руйга, И.А. Шкрадюк, А.О. Ступин // Решетневские чтения: Материалы XXIII Международной научно-практической конференции, посвященной памяти генерального конструктора ракетно-космических систем академика М.Ф. Решетнева. В 2-х частях, Красноярск, 11–15 ноября 2019 года / Под общей редакцией Ю.Ю. Логинова. Том Часть 2. — Красноярск: Федеральное государственное бюджетное образовательное учреждение высшего образования «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева», 2019. — С. 465–466.
 35. Farrell M.J. The Measurement of Productive Efficiency // *Journal of The Royal Statistical Society, Series A (General), Part III.* 1957. Vol. 120. P. 253–281
 36. Charnes A. Measuring the Efficiency of Decision-Making Units / Charnes A., Cooper W., Rhodes E. // *European Journal of Operational Research* 2: 429–444

37. Моргунов, Е.П., Моргунова О.И. Краткое описание метода Data Envelopment Analyses [Электронный ресурс] Режим доступа URL: DEA_intro.pdf (morgunov.org)
38. Banker, R. D., A.Charnes, and W.W. Cooper. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. / R.D. Banker, A.Charnes, and W.W. Cooper// *Management Science* 30: 1078–1092. 1984
39. Coelli, T.An Introduction to Efficiency and Productivity Analysis / T.Coelli, D.S. Prasada Rao, G.E. Battese. – Boston : Kluwer Academic Publishers, 1998. – 275 p.
40. Malmquist S.Index Numbers and Indifference Surfaces // *Trabajos de estadística*. 1953. Vol. 4, Iss. 2. P. 209–242.
41. Caves D.W. Multilateral Comparisons of Output, In-put, and Productivity Using Superlative Index Numbers / D.W. Caves, L.R. Christensen, W.E. Diewert // *The Economic Journal*. 1982. Vol. 92, Iss. 365. P. 73–86.
42. Camanho A.S. Data Envelopment Analysis and Malmquist Indices for Measuring Group Performance / A.S. Camanho, R.G. Dyson// *Journal of Productivity Analysis* 26(1), February 2006
43. Моргунов, Е.П. Измерение эффективности в сфере высшего образования с использованием метода «Анализ Среды Функционирования» / Е.П. Моргунов, О.Н. Моргунова // *Мягкие измерения и вычисления*. – 2019. – № 6(19). – С. 4–12.
44. Моргунов, Е.П. Продвижение метода оценки эффективности систем Data Envelopment Analysis в России / Е.П. Моргунов, О.Н. Моргунова // *Системный анализ в проектировании и управлении : сборник научных трудов XX Международной научно-практической конференции, Санкт-Петербург, 29 июня – 01 2016 года. Том Часть 2*. – Санкт-Петербург: Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого», 2016. – С. 390–398.

45. Cooper, W.W. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References, and DEA-Solver Software / W.W. Cooper, L.M. Seiford, K.Tone. — Boston : Kluwer Academic Publishers, 2000. — 318 p
46. Thanassoulis, E. Simulating Weights Restrictions in Data Envelopment Analysis by Means of Unobserved DMUs / E.Thanassoulis, R.Allen // *Management Science*. — 1998, April. — Vol. 44, No. 4. — P. 586–594.
47. Руйга, И.Р. Мониторинг угроз внешнеэкономической безопасности регионов сырьевой направленности: формирование системы критериев и показателей оценки, практическая реализация / И.Р. Руйга, А.С. Ямщиков, Ю.А. Тетерин, А.А. Чайка // *Российское предпринимательство*. — 2018. — Т. 19, № 12. — С. 3689–3704. — DOI 10.18334/rp.19.12.39554.
48. Stepan Zemtsov, Maxim Kotsemir. An assessment of regional innovation system efficiency in Russia: the application of the DEA approach. *Scientometrics* (2019) 120, 375–404.
49. Kaori Tone. Slacks-Based Measure Of Efficiency In Data Envelopment Analyses. *European Journal Of Operational Research* 130(2001), 498–509.
50. Руйга, И.Р. Оценка эффективности реализации инновационной политики в регионах Российской Федерации на основе использования метода кластеризации / И.Р. Руйга, Е.С. Ковзунова, Л.Н. Корпачева // *Региональная экономика: теория и практика*. — 2022. — Т. 20, № 2(497). — С. 259–288. — DOI 10.24891/re.20.2.259.
51. Zinenko A.V. Modeling and evaluating the effectiveness of innovation policy in the regions of Russia based on the data envelopment analysis method / A.V. Zinenko, I.R. Ruiga, L.N. Korpacheva [et al.] // *AIP Conference Proceedings* : 2, Krasnoyarsk, 29–31 июля 2021 года. — Krasnoyarsk, 2022. — P. 040021. — DOI 10.1063/5.0092763.
52. Казаковцев, Л.А. Метод жадных эвристик для систем автоматической группировки объектов: специальность 05.13.01

- «Системный анализ, управление и обработка информации (по отраслям)»: диссертация на соискание ученой степени доктора технических наук / Казаковцев Лев Александрович. — Красноярск, 2016. — 429 с.
53. Frank, I.E. Data Analysis Handbook / I.E. Frank, R.Todeschini.- [s.l.]: Elsevier Science Inc.- 1994.- P. 227–228..
54. Ester, M.Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise / M.Ester, H.-P. Kriegel, J.Sander, X.Xu // KDD-96 Proceedings. — [s.l.]:[s.n.]. — 1996. — P. 226–231.
55. Рубан, А.И. Идентификация и чувствительность сложных систем / А.И. Рубан. — Томск: Изд-во Томск, гос. Ун-та. — 1982. — С.302.
56. Справочник по прикладной статистике / Под ред. Э.Ллойда, У.Ледермана, С.А. Айвазяна и др.- М.: Финансы и статистика. — 1989. — Т. 1. — С. 510.
57. Parzen, E.On the Estimation of Probability Density Function and the Mode / E.Parzen / E.Parzen // Ann. Math. Statist. 1962. Vol. 33. P. 1065.
58. Parzen, E.On Estimation of a Probability Density, Function and Mode / E.Parzen // II IEEE Transactions on Information Theory. 1982. Vol. 4, No. 6. P. 663–666.
59. Епанечников, В.А. Непараметрическая оценка многомерной плотности вероятности / В.А. Епанечников // ТВиП. — 1969. — Т. 14. — С. 156–161.
60. Медведев, А.В. Непараметрические оценки плотности вероятности и ее производных / А.В. Медведев // Автоматизация промышленного эксперимента. — Фрунзе: Илим. — 1973. — с. 22–31.
61. Надарая Э.А. Об оценке регрессии / Э.А. Надарая // ТВиП. — 1964. — Т. 9. — № 1. — С. 157–159.
62. Sage, A.P. System Identification / A.P. Sage, J.L. Melsa. [s.l.]: Chu I. 1971. P. 221.
63. Иванова, Н.В. Определение параметров сглаживания в непараметрических оценках функции плотности по вы-

- борке / Н.В. Иванова, К.Т. Протасов // Математическая статистика и ее приложения. Томск: изд-во Томск, гос. Ун-та. — 1982. — Вып. 8. — С. 50–65.
64. Bellman, R.E. Dynamic Programming / R.E. Bellman.- NJ, Princeton: Princeton University Press. 1957. P. 392.
65. McLachlan, G.L. Mixture Models: Inference and Applications to Clustering / G.L. McLachlan, K.E. Basford. New York: Marcel Dekker. 1987. P. 253.
66. Blei, D.M. Latent dirichlet allocation / D.M. Blei, A.Y. Ng, M.I. Jordan // J.Machine Learn. Res. 2003. Vol.3. P. 993–1022.
67. Коплярова, Н.В. О непараметрических моделях в задаче диагностики электрорадиоизделий / Н.В. Коплярова, В.И. Орлов, Н.А. Сергеева, В.В. Федосов // Заводская лаборатория: диагностика материалов. — 2014. — № 80(7). — С.37–77.
68. Li, W.Pachinko allocation: Dag-structured mixture models of topic correlations / W.Li, A.McCallum // Proc. 23rd Internat. Conf. on Machine Learning. [s.l.]:[s.n.]. 2006. P. 577–584.
69. Welling, M.Exponential family harmoniums with an application to information retrieval / M.Welling, M.Rosen-Zvi, G.Hinton // Adv. Neural Inform. Process. Systems. 2005. Vol.17. P. 1481–1488.
70. Mohd, W.M.B.W. An Improved Parameter less Data Clustering Technique based on Maximum Distance of Data and Lioyd k-means Algorithm / W.M.B.W. Mohd, A.H. Beg, T.Herawan, K.F. Rabbi // First World Conference on Innovation and Computer Sciences (INSODE 2011). [s.l.]:[s.n.]. 2012. Vol.1. P. 367–371, DOI: 10.1016/j.protcy.2012.02.076
71. Долгушев, А.В. К вопросу об алгоритмической сложности одной задачи кластерного анализа /А.В. Долгушев, А.В. Кельманов // Дискретный анализ и исследование операций. — 2010. — Т. 17, № 2. — С. 39–45.
72. Bandyopadhyay, S.An evolutionary technique based on K-Means algorithm for optimal clustering / S.Bandyopadhyay,

- U.Maulik // Information Sciences. 2002. Vol. 146. P. 221–237.
73. Zhou, H. Accurate integration of multi-view range images using k-means clustering / H.Zhou, Y.Liu // Pattern Recognition. 2008. Vol. 41. P. 152–175.
74. Dunham, M.H. Data Mining: Introductory and advance topics / M.H. Dunham. N.J.: Prentice Hall. 2003. P. 315.
75. Lloyd, S.P. Least Squares Quantization in PCM / S.P. Lloyd // IEEE Transactions on Information Theory. 1982. Vol. 28. P. 129–137.
76. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. Новосибирск: ИМ СО РАН. — 1999. — С. 270.
77. Кельманов, А.В. Приближенный полиномиальный алгоритм для одной задачи бикластеризации последовательности / А.В. Кельманов, С.А. Хамидуллин // Журнал вычислительной математики и математической физики. — 2015. — Т. 55, № 6. — с. 1076–1085.
78. Кельманов, А.В. Квадратичные евклидовы задачи разбиения множества и последовательности: сложность и алгоритмы с оценками качества решения / А.В. Кельманов // Материалы VI международной конференции «Проблемы оптимизации и экономические приложения», Омск, 28 июня — 4 июля 2015, Омск: Омский государственный университет. — 2015. — С. 36–41.
79. Hofmann, T. Pairwise data clustering by deterministic annealing / T.Hofmann, J.M. Buhmann // IEEE Trans. Pattern Anal. Machine Intell. 1997. Vol.19, issue 1. P. 1–14.
80. Pavan, M. Dominant sets and pairwise Clustering / M.Pavan, M.Pelillo // IEEE Trans. Pattern Anal. Machine Intell. 2007. Vol.29, issue 1. P. 167–172.
81. Motzkin, T.S. Maxima for graphs and a new proof of a theorem of Turan / T.S. Motzkin, E.G. Straus // Canadian J.Math. 1965. Vol.17. P. 533–540.

82. Сергиенко, И.В. Математические модели и методы решения задач целочисленной оптимизации / И.В. Сергиенко, 2-е изд., доп. и перераб. — Киев: Наукова думка. — 1988. — С. 472.
83. Кочетов, Ю.А. Методы локального поиска для дискретных задач размещения: дис. ... доктора физ.-мат. Наук: 05.13.18: защищена 19.01.2010. — Новосибирск: Институт математики им.Соболева. — 2010. — С. 259.
84. Ausiello, G. Local Search, Reducibility and Approximability of NP-optimization Problems / G. Ausiello, M. Protasi // Information Processing Letters. 1995. Vol.54. P. 73–79.
85. Ganti, V. Clustering large datasets in arbitrary metric spaces / V. Ganti, R. Ramakrishnan, J. Gehrke, A. Powell, J. French. // Proc. 15th Int. Conf. Data Engineering. [s.l.]:[s.n.]. 1999. P. 502–511.
86. Pelleg, D. X-means: Extending k-means with efficient estimation of the number of clusters / D. Pelleg, A. Moore // Seventeenth Internat. Conf. on Machine Learning. 2000. P. 727–734.
87. Roberts, S.J. Minimum-entropy data clustering using reversible jump Markov chain Monte Carlo / S.J. Roberts, C. Holmes, D. Denison // Proc. Internat. Conf. Artificial Neural Networks. [s.l.]:[s.n.]. 2001. P. 103–110.
88. Still, S. Geometric Clustering using the Information Bottleneck method / S. Still, W. Bialek, L. Bottou // Advances In Neural Information Processing Systems 16 / Eds.: S. Thrun, L. Saul, and B. Scholkopf. — Cambridge: MIT Press. — 2004 [Электронный ресурс] Режим доступа URL <http://papers.nips.cc/paper/2361-geometric-clustering-using-the-information-bottleneck-method.pdf> (дата обращения 19.06.2023).
89. Kanungo, T. An efficient K-means clustering algorithm: analysis and implementation / T. Kanungo, D. Mount, N.S. Netanyahu, C. Piatko, R. Silverman, A. Wu // IEEE Trans. Pattern Anal. Mach. Intell. 2002. Vol.24, issue 7. P. 881–892.

90. Likas, A. The global k-means clustering algorithm / A. Likas, M. Vlassis, J. Verbeek // *Pattern Recognition*. 2003. Vol. 36. P. 451–461.
91. Chang, D. A genetic algorithm with gene rearrangement for K-means clustering / D. Chang, D. Xian, W. Chang // *Pattern Recognition*. 2009. Vol. 42. P. 1210–1222.
92. Maulik, U. Genetic Algorithm-Based Clustering Technique / U. Maulik, S. Bandyopadhyay // *Pattern Recognition*. 2000. Vol. 33. P. 1455–1465.
93. Kalyani, S. Particle swarm optimization based K-means clustering approach for security assessment in power systems / S. Kalyani, K. S. Swarup // *Expert Systems with Applications*. 2011. Vol. 38. P. 10839–10846.
94. Bagirov, A. M. Fast modified global k-means algorithm for incremental cluster construction / A. M. Bagirov, J. Ugon, D. Webb // *Pattern Recognition*. 2011. Vol. 44. P. 866–876.
95. Deza, M. M. *Encyclopedia of Distances* / M. M. Deza, E. Deza. Berlin, Heidelberg: Springer-Verlag. 2009. P. 590.
96. Mao, J. A self-organizing network for hyper-ellipsoidal clustering (HEC) / J. Mao, A. K. Jain // *IEEE Trans. Neural Networks*. 1996. Vol. 7 (January). P. 16–29.
97. Jain, A. K. Data clustering: 50 years beyond K-means / A. K. Jain // *Pattern Recognition Letters*. 2010. Vol. 31. P. 651–666.
98. Linde, Y. An algorithm for vector quantizer design / Y. Linde, A. Buzo, R. Gray // *IEEE Trans. Comm.* 1980. Vol. 28. P. 84–94.
99. Banerjee, A. Clustering with bregman divergences / A. Banerjee, S. Merugu, I. Dhillon, J. Ghosh // *J. Machine Learn. Res.* 2004. P. 234–245.
100. Capoyleas, V. Geometric Clusterings / V. Capoyleas, G. Rote, G. J. Woeginger // *J. Algorithms*. 1991. Vol. 12. P. 341–356.
101. Ball, G. H. ISODATA. An iterative method of multivariate analysis and pattern classification / G. H. Ball, D. J. Hall //

- 1966 IEEE Int. Commun. Conf., Philadelphia. New York: Digest of Techn. Pap. 1966. P. 116–117.
102. Pelleg, D. Accelerating exact k-means algorithms with geometric reasoning / D. Pelleg, A. Moore // Proc. Fifth Internat. Conf. on Knowledge Discovery in Databases. [s.l.]: AAAI Press. 1999. P. 277–281.
103. Scholkopf, B. Nonlinear component analysis as a kernel eigenvalue problem / B. Scholkopf, A. Smola, K.-R. Muller // Neural Comput. 1998. Vol. 10, issue 5. P. 1299–1319.
104. Миркин, В.Г. Шкалы упорядочения / В.Г. Миркин, Н.В. Высоцкая // Моделирование в экономических исследованиях. Новосибирск: Наука, Сибирское отделение, 1978. С. 109–120.
105. Datta, R. Image retrieval: Ideas, influences, and trends of the new age / R. Datta, D. Joshi, J. Li, J. Z. Wang // ACM Computing Surveys. 2008. Vol. 40, issue 2. Article 5. P. 5:1–5:60, DOI: 10.1145/1348246.1348248.
106. Lowe, D.G. Distinctive image features from scale-invariant keypoints / D.G. Lowe // Internat. J. Comput. Vision. 2004. Vol. 60, issue 2. P. 91–110.
107. Moore, A.W. Very fast EM-based mixture model clustering using multiresolution kd-trees / A.W. Moore // NIPS II. [s.l.]: MIT Press. 1998. P. 543–549.
108. Zhang T. BIRCH: An Efficient Data Clustering Method for Very Large Databases / T. Zhang, R. Ramakrishnan, M. Livny // Proceedings of the 1996 ACM SIGMOD international conference on Management of data (SIGMOD '96). New York: ACM. 1996, P. 103–114, DOI: 10.1145/233269.233324.
109. Steinbach, M. A comparison of document clustering techniques / M. Steinbach, G. Karypis, V. Kumar // KDD Workshop on Text Mining. 2000 [Электронный ресурс] Режим доступа URL http://www.cs.cmu.edu/~dunja/KDDpapers/Steinbach_IR.pdf (дата обращения 19.06.2023).

110. Karypis, G. A fast and high quality multilevel scheme for partitioning irregular graphs / G.Karypis, U.Fayyad, C.Reina // *SIAM J.Sci. Comput.* 1998. Vol.20, issue 1. P. 359–392, DOI: 10.1137/S1064827595287997.
111. Bradley, P.S. Scaling clustering algorithms to large databases / P.S. Bradley P.S., U.Fayyad, C.Reina // *Proceedings of the 4th Int. Conf. on Knowledge Discovery and Data Mining.* New York: [s.n.].1998. P. 9–15.
112. Fisher, D.H. Knowledge acquisition via incremental conceptual clustering / D.H. Fisher // *Machine Learn.* 1987. P. 139–172.
113. Kollios, G.Efficient biased sampling for approximate clustering and outlier detection in large data sets / G.Kollios, D.Gunopulos, N.Koudas, S.Berchtold // *IEEE Trans. Knowledge Data Eng.* 2003. Vol. 15, No.5. P. 1170–1187.
114. Ng, R. CLARANS: a method for clustering objects for spatial data mining / R.Ng, J.Han // *IEEE Trans. Knowl. Data Eng.* 2002. Vol.14. P. 1003–1016.
115. Bubeck, S.How the Initialization Affects the Stability of the k-Means Algorithm / S.Bubeck, M.Meila, U. von Luxurg // *Probability and Statistics.* 2012. Vol.16. P. 436–452.
116. Руйга, И.П. Эмпирическая оценка эффективности устойчивого инновационного развития регионов ресурсного типа / И.П. Руйга // *Журнал Сибирского федерального университета. Серия: Гуманитарные науки.* – 2021. – Т. 14, № 12. – С. 1863–1872. – DOI 10.17516/1997–1370-0865.
117. Сенчагов, В.К. Инновационные преобразования как императив экономической безопасности региона: система индикаторов / В.К. Сенчагов, Ю.М. Максимов, С.Н. Митяков, О.И. Митякова // *Инновации.* – 2011. – № . 5. – С. 56–61.
118. Жолудева, В.В., Применение кластерного анализа для оценки социально-экономического развития регионов на примере ЦФО и Ярославской области / В.В. Жо-

- лудева, Н.Ф. Мельниченко, Г.Е. Козлов // Статистика и математические методы в экономике. — 2014. — № 1. — С. 144–148.
119. Гренкин, Г.В. Методы вычислительной реализации рангового метода кластеризации // Информатика и системы управления. — 2012. — № 1 (31). — С. 71–79.
120. Махрусе, Н.Современные тенденции методов интеллектуального анализа данных: метод кластеризации // Московский экономический журнал. — 2019. — № 6. — С. 35.
121. Сотская, Т.В. Метод кластеризации субъектов предпринимательской деятельности как основа формирования институциональной политики государства в сфере экономических отношений // Общество и право. — 2010. — № 3 (30). — С. 298–303.
122. Федеральная служба государственной статистики [Электронный ресурс] — Режим доступа: <http://www.gks.ru/>.
123. Чучуева, И.А. Модель прогнозирования временных рядов по выборке максимального подобия, диссертация... канд. тех. наук / Московский государственный технический университет им. Н.Э. Баумана. Москва, 2012.
124. Борисов В.В., Луферов В.С. Метод многомерного анализа и прогнозирования состояния сложных систем и процессов на основе нечетких когнитивных темпоральных моделей // Системы управления, связи и безопасности. — 2020. — № 2. — С. 1–23. DOI: 10.24411/2410–9916-2020–10201.
125. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. — М.: ЮНИТИ-ДАНА, 2001. — 656 с.
126. Box E., Jenkins G.M., Reinsel G.C., Ljung G.M. Time Series Analysis: Forecasting and Control. — John Wiley & Sons, 2015. — 712 p.
127. Ярушкина Н.Г. Интеллектуальный анализ временных рядов: Учебное пособие — Ульяновск: УлГТУ, 2010–324 с.

128. Осовский С. Нейронные сети для обработки информации. — М.: Финансы и статистика, 2002. — 344 с.
129. Ладынин, А.И. Прогнозирование краткосрочных индикаторов экономической безопасности с использованием алгоритмов адаптивной фильтрации временных рядов / А.И. Ладынин, Е.С. Митяков // Развитие и безопасность. — 2021. — 1(9). — С. 42–54.
130. Светуных, И.С. Самообучающаяся модель краткосрочного прогнозирования социально-экономической динамики / И.С. Светуных // Национальный исследовательский университет «Высшая школа экономики»: официальный сайт. — 2011. — URL: https://www.hse.ru/data/2011/02/28/1211522815/2010_mk_article.pdf. (дата обращения 15.10.2023).
131. Баженов, О.В. Развитие методики прогнозирования при моделировании комплексных экономических систем / О.В. Баженов, А.Д. Галенкова // Экономический анализ: теория и практика. — 2017. — № 3. — С. 573–581.
132. Сазонов, А.А. Применение регрессионного анализа в прогнозировании / А.А. Сазонов // Научные исследования и разработки молодых ученых. — 2015. — URL: <https://cyberleninka.ru/article/n/primenenie-regressionnogo-analiza-v-prognozirovanii>. (дата обращения 15.04.2023).
133. Трофимов, П.Ю. Прогнозирование временных рядов методом ARIMA / П.Ю. Трофимов, В.Ю. Носков // 2017. — URL: https://elar.urfu.ru/bitstream/10995/52487/1/978-5-9908685-0-2_2017_72.pdf
134. Звонарев, С.В. Основы математического моделирования: учебное пособие. — Екатеринбург: Изд-во Урал. ун-та, 2019. — 112 с. — URL: https://elar.urfu.ru/bitstream/10995/68494/1/978-5-7996-2576-4_2019.pdf
135. Литвиненко, А.Н. Оценка вероятности роста угроз экономической безопасности на основе построения классификационных функций / А.Н. Литвиненко, А.В. Грачев, С.И. Тарашнина, И.И. Бритвина // Известия Юго-запад-

- ного Государственного университета. Серия: Экономика. Социология. Менеджмент. — 2019. — № 2(31). — С. 129–147.
136. Аликариев, Н.Ф. Моделирование социально-экономических факторов, влияющих на эффективность труда преподавателей / Н.Ф. Аликариев // Бюллетень науки и практики. — 2019. — Т. 5. — № 9. — С. 366–375.
137. Блем, А.Г. Имитационной модели социально-экономических процессов, представимых в виде цепей Маркова / А.Г. Блем, Д.Д. Никитин, Н.А. Таскин // Измерение, контроль, информатизация. — 2022. — С. 26–37.
138. Федорова, Е.А. Анализ влияния корпоративного управления на банкротство российских компаний на основе метода partial least Squares path modeling / Е.А. Федорова, Ю.А. Зеленков, Д.В. Чекризов, П.С. Добрянская // Корпоративные финансы. — 2016. — С. 108–123.
139. Хазова, Д.С. Применение методики PLS-PM для моделирования комплексных экономических систем / Д.С. Хазова // 2014. — URL: <https://izron.ru/articles/o-nekotorykh-voprosakh-i-problemakh-ekonomiki-i-menedzhmenta-sbornik-nauchnykh-trudov-po-itogam-mezh-sektsiya-5-matematicheskie-i-instrumentalnye-metody-ekonomiki-spetsialnost-08-00-13/primenenie-metodiki-pls-pm-dlya-modelirovaniya-kompleksnykh-ekonomicheskikh-sistem/>. (дата обращения 15.04.2023).
140. Логинов, К.К. Оценка уровня экономической безопасности с использованием неоднородного марковского процесса / К.К. Логинов, В.В. Карпов // Вестник Сибирского института бизнеса и информационных технологий. — 2022. — Т. 11. — 1. — С. 38–46.
141. Логинов, К.К. Оценка финального распределения вероятностей реализации угроз в сфере экономической безопасности с применением имитационного моделирования / К.К. Логинов, В.В. Карпов, А.А. Кораблева // Вестник

- Сибирского института бизнеса и информационных технологий. — 2021. — Т. 10. — № 1. — С. 67–75.
142. Астраханцева И.А. Рекуррентные нейронные сети для прогнозирования региональной инфляции / И.А. Астраханцева, А.С. Кутузова, Р.Г. Астраханцев // Научные труды Вольного экономического общества России. — 2020. — Т. 223. — С. 420–431.
143. Хамхоева, Ф.Я. Нейронные сети в экономическом анализе: плюсы и минусы / Ф.Я. Хамхоева // *Norwegian journal of development of the international science*. — 2020. — 51–4. — С. 72–75.
144. Капитанова, О.В. Прогнозирование социально-экономических процессов: Учебно-методическое пособие. // Нижний Новгород: Нижегородский госуниверситет, 2016. — 74 с.
145. Кузьмич, Р.И. Сравнительный анализ методов классификации данных на практических задачах прогнозирования и диагностики / И.С. Масич, Е.М. Краева, Р.И. Кузьмич, Т.К. Гулакова // Системы управления и информационные технологии. — 2011. — № 1(43). — С. 20–25.
146. Масич, И.С. Комбинаторная оптимизация в задаче классификации / И.С. Масич // Системы управления и информационные технологии. — 2009 — № 1.2(35). — С. 283–288.
147. Осложнения инфаркта миокарда: база данных для апробации систем распознавания и прогноза / А.Н. Горбань, В.А. Шульман, Д.А. Россиев и др. — Красноярск, Вычислительный центр СО РАН: Препринт № 6, 1997.
148. Дронов, В. Линейная регрессия с помощью Scikit-Learn в Python / В. Дронов // Обучение Python — URL: <https://tonais.ru/library/lineynaya-regressiya-s-pomoschyu-scikit-learn-v-python>
149. Чибирова М.Э. Анализ данных и регрессионное моделирование с применением языков программирования

- Python и R / М.Э. Чибирова // Научные записки молодых исследователей. — 2019. — № 2. — С. 37–45.
150. Свидетельство о государственной регистрации программы для ЭВМ № 2023619972 Российская Федерация. Программный модуль идентификации и прогнозирования рисков и угроз продовольственной безопасности арктических регионов на основе модели Холта-Винтерса: № 2023617833: заявл. 25.04.2023; опубл. 17.05.2023 / Е.С. Ковзунова, А.В. Зиненко, И.Р. Руйга; заявитель Федеральное государственное автономное образовательное учреждение высшего образования «Сибирский федеральный университет».

Оглавление

Введение	3
Глава 1. Имитационное моделирование на основе применения метода Монте-Карло	5
1.1. Теоретические аспекты применения метода Монте-Карло.....	5
1.2. Практическая реализация метода Монте-Карло для решения бизнес-задачи.....	10
1.3. Практическая реализация метода Монте-Карло для принятия управленческих решений в социально-экономических системах	19
Глава 2. Метод многокритериальной оптимизации Data Envelopment Analysis	45
2.1. Теоретические аспекты применения метода Data Envelopment Analysis	45
2.2. Экспериментальное применение метода Data Envelopment Analysis для решения задачи оценки эффективности региональной системы здравоохранения	60
2.3. Экспериментальное применение метода <i>Data Envelopment Analysis</i> для решения задачи оценки эффективности региональной инновационной политики	97
Глава 3. Задачи кластеризации и классификации в анализе данных	115
3.1. Основные подходы к решению задач группировки объектов и данных	115
3.2. Задача классификации в анализе данных сложных систем	130
3.3. Задача кластеризации в анализе данных сложных систем	134
Глава 4. Методы и модели прогнозирования состояния сложных систем	144
4.1. Обзор основных методов и моделей прогнозирования	144
4.2. Практическая реализация метода ANOVA в комбинации с методом Фехнера	160
4.3. Практическая реализация метода Хольта — Винтерса в комбинации с инструментами программирования.....	174
Заключение.....	184
Список использованных источников	185

ФЗ № 436-ФЗ	Издание не подлежит маркировке в соответствии с п. 1 ч. 2 ст. 1
----------------	--

Научное издание

**Ступина Алена Александровна,
Руйга Ирина Рудольфовна,
Казаковцев Лев Александрович**

МЕТОДЫ МОДЕЛИРОВАНИЯ И ОПТИМИЗАЦИИ ДЛЯ ОБЕСПЕЧЕНИЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В СЛОЖНЫХ СИСТЕМАХ

МОНОГРАФИЯ

Оригинал-макет подготовлен в НИЦ ИНФРА-М

ООО «Научно-издательский центр ИНФРА-М»
127214, Москва, ул. Полярная, д. 31В, стр. 1

Тел.: (495) 280-15-96, 280-33-86. Факс: (495) 280-36-29

E-mail: books@infra-m.ru <http://www.infra-m.ru>

Подписано в печать 29.12.2023. Формат 60×90/16. Бумага офсетная.

Гарнитура Newton. Печать цифровая. Усл. печ. л. 12,81.

Тираж 84 экз. Заказ № 00000

ТК 821666-2131543-291223

Отпечатано в типографии ООО «Научно-издательский центр ИНФРА-М»
127214, Москва, ул. Полярная, д. 31В, стр. 1

Тел.: (495) 280-15-96, 280-33-86. Факс: (495) 280-36-29