

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

*На правах рукописи*

Бачаев Умар Аптиевиц

НЕПАРАМЕТРИЧЕСКОЕ  
МОДЕЛИРОВАНИЕ НА ОСНОВЕ  
КОПУЛА-ФУНКЦИЙ В АНАЛИЗЕ  
ПРОСТРАНСТВЕННОГО РАЗВИТИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

5.2.2. Математические, статистические и инструментальные методы  
в экономике

ДИССЕРТАЦИЯ  
на соискание ученой степени  
кандидата экономических наук

Научный руководитель

Гринева Наталья Владимировна,  
кандидат экономических наук, доцент

Москва - 2026

## Оглавление

Введение.....	4
Глава 1 Копула-подход в анализе пространственного развития регионов Российской Федерации: постановка задачи и методы оценивания .....	13
1.1 Проблема пространственного социального-экономического и демографического развития регионов Российской Федерации: формулировка исследовательской задачи .....	14
1.2 Теория копула-функций: этапы развития, понятие, классификация и применимость в пространственном анализе .....	24
1.3 Методы статистического оценивания параметров копул.....	40
Глава 2 Анализ и оценка парных зависимостей социально-экономических и демографических показателей регионов Российской Федерации .....	52
2.1 Описание и анализ данных и переменных исследования .....	53
2.2 Меры зависимости случайных величин и их экономическая интерпретация.....	63
2.3 Кривые копула-зависимостей на примере парных моделей .....	71
2.4 Анализ хвостовой структуры зависимостей демографических и социально-экономических показателей .....	76
Глава 3 Непараметрическое многомерное моделирование и сценарный анализ демографического роста регионов Российской Федерации на основе Vine-копул.....	84
3.1 Vine-копулы как инструмент многомерного моделирования зависимостей .....	85
3.2 Vine-копула регрессия: предпосылки и ограничения .....	91
3.3 Множественная Vine-копула регрессия для оценки факторов демографического развития.....	103

3.4 Структура влияния социально-экономических факторов на демографический рост: региональные эффекты и сценарное моделирование .....	113
Заключение .....	129
Список литературы .....	133
Приложение А Исходный набор данных по Дальневосточному федеральному округу для 2023 года .....	141
Приложение Б Структура и параметры R-vine-копультной модели .....	142
Приложение В R-vine матрица структуры копультной модели.....	146
Приложение Г R-vine матрица графиков парных копул (дерево T1) .....	147

## Введение

**Актуальность темы исследования.** Пространственное развитие регионов Российской Федерации выступает одновременно заметным ограничением и важным ориентиром экономического и социального развития. Для принятия стратегических управленческих решений в рамках политики регионального развития необходимо количественно устанавливать, каким образом характеристики региональной экономики, демографии и социальной среды соотносятся с изменением ключевых показателей территориального развития. При этом принципиально важны не только усредненные зависимости, но и наблюдения, относящиеся к крайним значениям показателей: именно в регионах с наиболее выраженными отклонениями по уровню экономического и социального развития особенно остро встают вопросы адресных мер, выбора приоритетов и сценариев развития регионов. Региональные данные отличаются высокой неоднородностью, пространственной кластеризацией и асимметрией связей показателей экономического, социального и демографического развития, что ограничивает применимость классических подходов для их моделирования.

Классические регрессионные подходы и их расширения в основном базируются на оценивании усредненной линейной корреляции и предпосылках, которые являются ограничивающими моделирование неоднородных распределений и изменчивой структуры зависимостей. В качестве альтернативы этим подходам часто выбираются методы машинного обучения и более сложные нелинейные модели, однако они, как правило, не предоставляют вероятностное описание условной зависимости и затрудняют экономическую интерпретацию результатов моделирования.

Для рассматриваемой задачи нужен подход, позволяющий описывать многомерные связи не только через средние эффекты, но и через структуру совместного распределения показателей. В работе для этого используется аппарат копула-функций, который дает возможность отдельно моделировать

маргинальные распределения переменных и структуру зависимости между ними, включая нелинейность, асимметрию и хвостовые эффекты. В исследовании Vine-копула-регрессия используется как основной инструмент анализа пространственных региональных данных, поскольку она позволяет учитывать многомерный характер зависимостей между социально-экономическими и демографическими показателями.

**Степень разработанности темы исследования.** Проблематика пространственного развития регионов широко представлена в экономической географии, региональной экономике и пространственной эконометрике. В российской научной литературе вопросы территориальной организации хозяйства, межрегиональной дифференциации и необходимости целенаправленной государственной политики раскрыты в работах А.Г. Гранберга; роль агломераций, структурных факторов и пространственной неоднородности российских регионов отражена в работах Н.В. Зубаревич. Методы пространственно-эконометрического анализа регионального развития, включая учет пространственных эффектов, межрегиональных влияний и неоднородности связей между субъектами Российской Федерации, представлены в работах О.А. Демидовой и Д.С. Иванова. В зарубежной литературе исследуются механизмы территориальной поляризации, агломерационные эффекты, центрo-периферийные различия, инерционность территориальной структуры, а также роль технологических сдвигов и глобализации в усилении концентрации экономической активности. В этой части логика исследования опирается на подходы, представленные в работах Bathelt, Duranton, Puga, Iammarino и др., а также на результаты исследований по пространственным неравенствам в Европе (European Parliament, RegPol2).

Отдельный блок исследований посвящен пространственным аспектам демографической динамики: подчеркивается значимость географической неоднородности, пространственной взаимосвязанности территорий и миграционных взаимодействий, а также показывается, что демографические показатели формируются под влиянием локальных условий и окружения.

В диссертации используются выводы, представленные в работах Weeks, Tamura, Wilson, Baker, где обсуждаются пространственная зависимость, прогнозирование населения на малых территориях и эффекты пространственного сглаживания.

В математической статистике и прикладных областях сформирован развитый аппарат моделирования зависимости на основе копула-функций, позволяющий отдельно описывать маргинальные распределения и структуру зависимости, включая хвостовые эффекты. Теоретические основы копул связаны с работами Sklar, Deheuvels, Nelsen, Joe; методы оценивания и статистические свойства – с работами Genest, Rivest, Rémillard, Fan, Chen. Развитие многомерных конструкций парных копул (Vine-копул) связано с работами Bedford & Cooke и последующими исследованиями Aas, Czado, Nagler, Kurowicka & Joe, Brechmann, Schepsmeier и другие, где предложены алгоритмы спецификации структуры, выбора семейств и оценивания параметров в многомерной постановке.

В научной литературе аппарат копула-функций широко используется в анализе рисков, страховании и финансовой эконометрике, тогда как в исследованиях социально-экономических процессов применяется реже и преимущественно для анализа парных зависимостей. Анализ степени разработанности темы показывает, что теоретические основы копул и методы построения Vine-копул достаточно подробно представлены в литературе, однако их применение к моделированию зависимостей между показателями пространственного развития регионов Российской Федерации и построению сценарных расчетов на региональных данных остается недостаточно проработанным.

**Целью исследования** является оценка степени и направленности влияния ключевых социально-экономических факторов на демографическую составляющую пространственного развития регионов Российской Федерации, построение сценарных оценок региональной демографической динамики с помощью непараметрического моделирования на основе Vine-копул,

учитывающего нелинейность, асимметрию и хвостовое поведение распределений.

Для достижения поставленной цели в диссертации решаются следующие задачи:

– Систематизировать теоретические основы копула-моделирования зависимостей и подходов к оцениванию копула-функций для обоснования их применимости в анализе социально-экономических и демографических показателей регионов Российской Федерации.

– Провести статистический анализ распределений региональных социально-экономических и демографических показателей регионов Российской Федерации за 2010-2023 гг. на основе сформированной информационной базы, обеспечивающей сопоставимость наблюдений.

– Выполнить анализ парных зависимостей между темпами прироста численности населения и основными социально-экономическими показателями пространственного развития регионов Российской Федерации на основе расчета коэффициентов хвостовых зависимостей и построения кривых копула-функций.

– Определить предпосылки и ограничения применения Vine-копула-регрессии как непараметрического метода восстановления условного распределения демографического показателя при заданных значениях социально-экономических факторов.

– Построить и оценить множественную Vine-копула-регрессионную модель для анализа влияния социально-экономических факторов на демографическую динамику регионов Российской Федерации и сопоставить ее с линейными моделями с регуляризацией, чтобы оценить преимущества предлагаемого подхода при моделировании нелинейных и асимметричных зависимостей.

– Провести анализ структуры влияния факторов и выполнить сценарное моделирование демографического роста регионов Российской Федерации на основе значений восстановленных условных распределений, учитывая при

этом возможные нелинейности, взаимодействие факторов и региональную неоднородность.

**Объектом исследования** является пространственное развитие Российской Федерации на уровне ее регионов, рассматриваемое через систему социально-экономических и демографических показателей.

**Предметом исследования** являются методы непараметрического моделирования на основе копула-функций и их применение для анализа пространственной динамики российских регионов.

**Область исследования диссертации** соответствует п. 3. «Разработка и развитие математических и эконометрических моделей анализа экономических процессов (в т.ч. в исторической перспективе) и их прогнозирования»; п. 14. «Эконометрические и статистические методы анализа данных, формирования и тестирования гипотез в экономических исследованиях. Эконометрическое и экономико-статистическое моделирование» Паспорта научной специальности 5.2.2. Математические, статистические и инструментальные методы в экономике (экономические науки).

**Научная новизна исследования** заключается в разработке и апробации экономико-математического инструментария анализа пространственного развития регионов Российской Федерации на основе непараметрического моделирования с использованием копула-функций, включающего теоретическое обоснование применимости копула-подхода к исследованию региональных социально-экономических и демографических показателей, методику оценки парных и многомерных зависимостей с учетом нелинейности, асимметрии и хвостовых эффектов, Vine-копула-регрессионную модель анализа влияния социально-экономических факторов на региональную демографическую динамику, а также процедуру построения сценарных оценок развития регионов на основе значений восстановленных условных распределений.

**Положения, выносимые на защиту.** Результаты, полученные лично автором в ходе проведения исследования и отвечающие требованиям научной новизны, а также выносимые на защиту, состоят в следующем:

– в рамках задачи анализа пространственного развития регионов Российской Федерации теоретически обоснована применимость копула-подхода для совместного исследования социально-экономических и демографических показателей. В отличие от классических регрессионных подходов, ориентированных преимущественно на оценку усредненных эффектов, копула-подход позволяет рассматривать структуру зависимости между региональными показателями с учетом неоднородности распределений, нелинейности, асимметрии и поведения показателей в крайних областях распределений (С. 13-40);

– для региональных данных Российской Федерации выполнен комплексный анализ хвостовых зависимостей между демографическими и социально-экономическими показателями, позволивший выявить нелинейные и асимметричные связи и тем самым обосновать переход к копула-моделированию как инструменту учета зависимости в центральных и крайних областях распределений региональных показателей (С. 58-80);

– впервые разработан и реализован пошаговый алгоритм применения многомерной Vine-копула-регрессии для оценки влияния социально-экономических факторов на демографическую составляющую пространственного развития регионов Российской Федерации, учитывающий многомерную структуру условных зависимостей, взаимодействие факторов, нелинейность и асимметрию связей между региональными показателями (С. 91-113);

– разработан подход к сценарному анализу региональной динамики на основе Vine-копула-регрессии, позволяющий оценивать изменение демографических показателей при различных сочетаниях социально-экономических факторов пространственного развития. В отличие от расчетов, основанных на изолированных эффектах факторов, предложенный подход

позволяет учитывать их совместное воздействие и выявлять неаддитивные эффекты в сценариях регионального развития (С. 113-126).

**Теоретическая значимость работы.** Работа расширяет теоретический аппарат непараметрического моделирования в экономике, развивает направление применения копула-функций в анализе многомерных зависимостей пространственного регионального развития, а также вносит вклад в методику построения условных распределений на основе Vine-копул.

**Практическая значимость работы** заключается в том, что предложенный непараметрический копульный подход предлагает инструментарий для экономического анализа в условиях нелинейности, асимметрии и хвостовых эффектов зависимостей между региональными показателями: он позволяет оценивать влияние факторов не только через точечные прогнозы, но и через характеристики условных распределений (ожидания, квантили, хвостовые вероятности). В работе этот инструментарий реализован в рамках задачи пространственной демографической динамики регионов Российской Федерации: полученные модели и сценарные расчеты могут использоваться для обоснования и мониторинга мер пространственной и демографической политики, а также при подготовке прогнозов и аналитических материалов для региональных стратегий и программ.

**Методологию и методы исследования** составляют современные подходы к моделированию многомерных статистических зависимостей, теория копула-функций, методы непараметрической регрессии и инструменты анализа структуры распределений. В работе используется концепция разделения маргинальных распределений и структуры зависимости, лежащая в основе теоремы Скляра, а также подходы к построению условных распределений и условного математического ожидания в рамках копульного моделирования.

**Степень достоверности, апробации и внедрение результатов исследования.** Обоснованность и достоверность результатов исследования обеспечиваются использованием официальных статистических данных по

регионам Российской Федерации за 2010-2023 гг., применением аппарата копульного моделирования и воспроизводимостью вычислительной процедуры. Достоверность выводов подтверждается более высокой точностью и качеством Vine-копула-регрессии по сравнению с линейными моделями, а также сохранением аналитического представления совместной функции распределения, что позволяет интерпретировать структуру зависимостей между региональными показателями.

Основные положения и выводы диссертационной работы докладывались и были озвучены в рамках научно-практических мероприятий: на VIII Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 28 мая 2021 г.); на XI Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 19 июня 2024 года); на III Международной научно-практической конференции студентов и аспирантов «Цифровизация экономики и финансов: модели, методы и технологии» (Москва, Финансовый университет, 23-24 апреля 2025 года); на XII Международной научно-практической конференции «Современная математика и концепции инновационного математического образования» (Москва, Финансовый университет, 11 июня 2025 года).

Материалы диссертации были использованы при выполнении научно-исследовательской работы по теме: «Разработка методологии моделирования и прогнозирования динамики пространственного развития Российской Федерации на основе data-driven management» (Государственное задание, приказ Финуниверситета от 24.12.2025 № 3243/о) в части подготовки раздела «Выявление пространственных эффектов на основе индексов Морана и Джири».

Материалы диссертации используются в практической деятельности ООО «Яндекс Еда» при решении аналитических задач, связанных с оценкой

региональной неоднородности пользовательского спроса и различий между регионами присутствия компании. В практическую работу внедрен инструментарий непараметрического моделирования на основе Vine-копула-регрессии, предназначенный для оценки изолированных и совместных эффектов факторов на показатели спроса и пользовательской активности. Инструментарий был применен для оценки влияния коммуникационных каналов на показатели спроса, сопоставления совокупного эффекта нескольких каналов с суммой их изолированных эффектов, а также анализа различий в отклике пользователей по регионам присутствия компании. Полученные результаты были применены при подготовке медиастратегии компании на 2026 год и использованы в практической работе Группы по планированию и закупке медиа.

Материалы исследования используются Кафедрой информационных технологий Факультета информационных технологий и анализа больших данных Финансового университета в преподавании учебной дисциплины «Алгоритмы и структуры данных в языке Python» для обучающихся по направлению подготовки 01.03.02 Прикладная математика и информатика.

Апробация и внедрение результатов исследования подтверждены соответствующими документами.

**Публикации.** Основные положения и результаты исследования отражены в 7 научных работах общим объемом 5,8 п.л. (авторский объем – 4,9 п.л.), в том числе 4 работы общим объемом 3,5 п.л. (весь объем авторский) опубликованы в рецензируемых научных изданиях, рекомендованных ВАК при Минобрнауки России.

**Структура и объем диссертации.** Цель и задачи исследования определили структуру работы. Диссертация состоит из введения, трех глав, заключения, списка литературы, содержащего 65 наименований, 4 приложения. Текст диссертации изложен на 147 страницах, включает 21 рисунок, 11 таблиц и 86 формул.

## Глава 1

# **Копула-подход в анализе пространственного развития регионов Российской Федерации: постановка задачи и методы оценивания**

Пространственное развитие в экономическом смысле – это изменение во времени территориальной организации экономики и населения: где концентрируются люди и рабочие места, как распределяются производство, доходы, инфраструктура и доступ к базовым услугам между регионами и внутри них. Иначе говоря, это то, как устроена экономика на карте страны и почему одни территории ускоряются, а другие закрепляются в траектории отставания.

Для Российской Федерации проблема пространственного развития имеет принципиальное значение из-за масштабов территории и высокой неоднородности регионов. На практике это проявляется в концентрации экономической активности и населения в ограниченном числе центров роста, усилении разрывов по уровню доходов и качеству городской среды, различиях в инвестиционной динамике и доступности инфраструктуры. Такая поляризация создает долгосрочные ограничения для устойчивого роста: снижается связанность экономического пространства, усиливаются миграционные перетоки, повышается нагрузка на крупнейшие агломерации и одновременно сокращается потенциал периферийных территорий.

Именно поэтому государственная политика фиксирует пространственное развитие как самостоятельный приоритет: в Стратегии пространственного развития Российской Федерации на период до 2030 года с прогнозом до 2036 года акцент делается на снижении межрегиональной дифференциации, повышении устойчивости территорий и согласовании мер развития с демографическими ограничениями. В этой логике демография выступает не «социальным фоном», а ключевым фактором экономической динамики регионов: рождаемость, смертность и миграция определяют

предложение труда, структуру спроса, бюджетные возможности и перспективы человеческого капитала.

Отсюда возникает центральная исследовательская задача: количественно оценить, как социально-экономические условия и меры развития (доходы населения, инвестиции, рынок жилья, инфраструктура и качество среды) связаны с демографическими результатами в разрезе субъектов Российской Федерации. При этом для управленческих решений важны не только средние зависимости, но и различия между регионами с разными траекториями – территориями демографического роста и регионами устойчивой убыли. Такая неоднородность обычно сопровождается нелинейностью, асимметрией реакций и значимостью «хвостовых» эффектов.

В первой главе формулируется исследовательская проблема пространственного социально-экономического и демографического развития регионов Российской Федерации и обосновывается выбор копула-функций как инструмента анализа зависимостей, позволяющего выявлять нелинейные и асимметричные связи, в том числе в хвостах распределений. Далее последовательно раскрываются понятие копулы, ее ограничения и классификация, а также рассматриваются основные подходы к оцениванию копула-функций, что формирует методологическую базу для анализа парных зависимостей и последующего многомерного моделирования на основе копул.

### **1.1 Проблема пространственного социально-экономического и демографического развития регионов Российской Федерации: формулировка исследовательской задачи**

Пространственное развитие в современной экономической теории понимается как процесс, определяющий распределение хозяйственной активности, населения и социальных благ по географическому пространству [34]. Являясь многоаспектным феноменом, оно отражает взаимодействие экономических, демографических, институциональных и географических

факторов, которые в совокупности формируют долгосрочные траектории регионального развития.

Одной из центральных характеристик пространственного развития является территориальная поляризация – концентрация экономической активности, инвестиций и человеческого капитала в ограниченном числе регионов-центров при одновременной маргинализации периферийных территорий. Этот процесс не является случайным явлением, но результатом действия фундаментальных экономических механизмов. Duranton и Puga демонстрируют, что агломерационные эффекты, являющиеся положительными внешними эффектами, возникающими при концентрации производства в определенной локации, способствуют формированию самоподдерживающегося неравенства. Регионы, достигшие критического масштаба, как правило, привлекают больше инвестиций и квалифицированных специалистов, что усиливает их сравнительные конкурентные преимущества по отношению к другим территориям [43].

Исследования пространственных неравенств в Европе, включая Центральную и Восточную Европу [29; 31], выделяют факторы, связанные с ростом поляризации. Существенную роль играет историческая инерция: распределение экономической активности, сложившееся в период индустриализации, во многом сохраняется и при изменении внешней среды. Дополнительное влияние оказывают географические условия и ресурсная база, включая доступ к международным рынкам и портовой инфраструктуре, что предопределяет различия в исходных возможностях регионов. А также, технологические сдвиги и глобализация последних десятилетий усилили концентрацию высокотехнологичного производства и услуг в крупнейших городских агломерациях [49]. Наконец, избирательная миграция, при которой молодые и высокообразованные люди оставляют регионы с низкими перспективами, создает «порочный круг»: убыль человеческого капитала снижает инновационный потенциал и привлекательность региона для инвестиций, что еще больше стимулирует миграцию.

Эмпирические исследования последних лет подтверждают, что разрыв между богатыми и бедными регионами в развитых экономиках не только не сокращается, но и увеличивается [34]. Так, в США и в ряде европейских стран коэффициент вариации регионального ВВП на душу населения возрос с 1980-х годов, что указывает на стойкую тенденцию к поляризации несмотря на многолетние попытки государственного регулирования [30].

В ответ на нарастающие вызовы неравномерного пространственного развития Правительство Российской Федерации 28 декабря 2024 года утвердило Стратегию пространственного развития Российской Федерации на период до 2030 года с прогнозом до 2036 года (далее Стратегия-2030). Документ представляет собой комплексный политический и стратегический инструмент, направленный на переориентацию государственной политики с преобладавшего ранее подхода конкурентного развития отдельных регионов на более сбалансированную и взаимосвязанную модель пространственного устройства страны.

Стратегия-2030 содержит ряд фундаментальных положений. В частности, в ней фиксируется объективный характер и в определенной мере неизбежность процессов концентрации экономической активности; одновременно подчеркивается, что для снижения связанных с этим дисбалансов требуется их институциональное регулирование, прежде всего через развитие сети опорных населенных пунктов, которые должны обеспечивать инфраструктурный и экономический каркас страны. Далее документ выделяет укрепление межрегиональных связей как самостоятельный приоритет; для достижения этой цели предполагаются формирование логистических коридоров и меры по развитию цифровой связанности. Наконец, Стратегия-2030 исходит из необходимости дифференцированного подхода к макрорегионам, при котором принимаются во внимание траектории и ограничения развития таких территорий, как Арктическая зона, Дальний Восток, Сибирь, Центральная Россия и другие.

Особое внимание в работе уделяется демографическим процессам, рассматриваемым в качестве одного из ключевых факторов, определяющих долгосрочные тенденции пространственного развития. Согласно Стратегии-2030, сокращение численности населения в большинстве регионов, тенденция миграции молодежи в крупные агломерации и старение населения создают системные риски, осложняющие достижение прочих целей социально-экономического развития. Таким образом, анализ демографической и пространственной динамики становится необходимым элементом при обосновании эффективной долговременной государственной политики.

Российская и зарубежная академическая литература предлагает разнообразные интерпретации и аналитические подходы к проблематике пространственного развития регионов. Зубаревич Н.В. в своих исследованиях акцентирует внимание на роли агломерационных эффектов и структурных факторов в формировании территориальной поляризации. По ее оценке, концентрация ресурсов в крупнейших городских агломерациях – явление, характерное не только для России, но и для всего пространства бывшего Советского Союза, наследующего советскую модель территориального распределения хозяйства. Зубаревич подчеркивает, что традиционные подходы к региональному развитию, основанные на форме финансовой поддержки отстающих регионов, оказываются недостаточными, так как не устраняют фундаментальные причины неравенства – отсутствие экономических центров притяжения за пределами столицы и сырьевых центров [17].

Гранберг А.Г. рассматривал территориальную организацию хозяйства как сложную динамическую систему, находящуюся в состоянии постоянного неравновесия и требующую целенаправленного государственного регулирования. Его концепция подчеркивает, что без активной политики государства естественные рыночные силы будут только усиливать существующие территориальные диспропорции [1]. Эта позиция остается актуальной и для анализа Стратегии-2030, которая признает необходимость

государственного вмешательства в развитие системы опорных центров и сопутствующей инфраструктуры.

В международной литературе исследование пространственного развития связано с анализом процессов конвергенции и дивергенции региональных экономик [49]. Согласно этим исследованиям, в условиях глобализации и технологических сдвигов наблюдается парадокс: несмотря на теоретическое предположение о том, что дешевые коммуникации должны способствовать выравниванию региональных различий, на практике происходит противоположное – геолокация экономических преимуществ становится все более концентрированной. Регионы, добившиеся лидерства в инновационных секторах (ИТ, финансы, высокие технологии), создают своеобразные экосистемы, воспроизводящие и укрепляющие их конкурентное преимущество.

Исследования, проведенные в Европе относительно стран Центральной и Восточной Европы [29; 31], показали: хотя формально условия развития этих стран и ЕС стали более схожими, внутривластные различия между регионами продолжают увеличиваться. Это говорит о том, что сложности в пространственном развитии обусловлены не только разницей в доходах, но и глубинными различиями в способности регионов приспосабливаться к быстрым переменам и встраиваться в международные экономические процессы.

Стратегия-2030 отмечает, что демография – один из ключевых аспектов, который может серьезно повлиять на достижение долгосрочных целей в сфере пространственного развития. В документе обозначены несколько важных трендов, совместное воздействие которых способствует углублению демографического кризиса:

- 1) В существенной части регионов фиксируется сокращение населения по естественным причинам – смертность превышает рождаемость. По информации Росстата, в 2023 году убыль населения достигла 495,3 тысячи человек. Хотя этот показатель и снизился по сравнению с 2022 годом (тогда

он составил 599,6 тысячи человек) [4], тенденция к сокращению численности продолжает сохраняться.

2) Избирательная миграция молодежи и высокообразованных кадров в крупные агломерации, которая усиливает дефицит квалифицированной рабочей силы в периферийных регионах. По данным РА Эксперт [5], положительное миграционное сальдо в январе-сентябре 2024 года наблюдалось только в 32 субъектах Российской Федерации, тогда как остальные регионов испытывали отток населения.

3) Трансформация структуры расселения, при которой города-миллионники и крупнейшие городские агломерации растут, тогда как малые и средние города, а также сельские поселения теряют население. За период 2002-2021 годов население малых и средних городов сократилось на 7,4%, в то время как города-миллионники увеличились на 20,4% [2].

Связь между пространственным развитием и демографическими процессами носит взаимный характер. Экономическая поляризация усиливает миграционный отток в сторону крупнейших центров. В результате концентрация населения в них возрастает еще сильнее. Одновременно старение населения и его убыль в периферийных регионах сужают предложение труда и снижают их инвестиционную привлекательность.

Из содержания Стратегии-2030 следует, что демографическая динамика рассматривается не как вторичный результат экономического развития, а как самостоятельное направление государственной политики, связанное с целями пространственного развития.

Исследования демографических процессов в пространственном разрезе пересекают несколько дисциплин – демографию, экономическую географию, пространственный анализ и региональную экономику. Ключевые работы в этой области сосредоточены на следующих направлениях.

Первое направление связано с анализом роли пространственной автокорреляции в демографических процессах. Исследование показало, что показатели роста населения в одном регионе зависят не только от локальных

факторов, но и от демографических характеристик соседних регионов, причем эта зависимость может сохраняться на дистанциях до 20 километров. Авторы использовали модели гравитации для описания миграционных потоков, показав, что индивиды принимают решения о миграции, опираясь на агрегированную информацию о привлекательности соседних территорий, а не только на характеристики конкретного пункта назначения [61].

Второе направление сосредоточено на разработке методов прогнозирования населения на малых географических масштабах с высокой пространственной разрешенностью. Wilson в своем обзорном исследовании методов краткосрочного прогнозирования численности населения на уровне малых административно-территориальных единиц отметил, что исторически такие прогнозы базировались на когортно-компонентном методе, однако в последние годы все чаще используются пространственные подходы, включающие информацию о географической близости регионов [64]. Различные ученые в своих научных работах также показали, что применение пространственного сглаживания (*spatial smoothing*), при котором прогноз для каждого региона корректируется на основе прогнозов соседних регионов, повышает точность предсказаний на 20-30% [33; 64].

Третье направление включает исследования по разработке высокоразрешенных пространственных демографических моделей с использованием методов геоинформатики. Zoraghein и O'Neill разработали гравитационную модель пространственного распределения населения, которая позволяет проецировать не только общую численность населения, но и его пространственное размещение на сетке высокого разрешения. Авторы показали, что такие модели могут быть откалиброваны на исторических данных переписей и использованы для создания сценариев демографического развития, учитывающих различные варианты экономического развития и миграционной политики [60].

По мере усложнения пространственных демографических процессов количественный анализ должен учитывать, во-первых, территориальную

неоднородность и пространственную связанность регионов, а во-вторых, нелинейный и асимметричный характер связей между демографическими и социально-экономическими показателями, особенно в экстремальных режимах. В эконометрике пространственную структуру обычно диагностируют через показатели пространственной автокорреляции, включая индекс Морана и локальные меры. Для моделирования зависимости применяют пространственные регрессии SAR/SLM, SEM, SDM и их варианты, что позволяет количественно описывать межтерриториальные перетоки и учитывать пространственную коррелированность ошибок при анализе региональных данных.

В российских исследованиях применение такого класса моделей представлено, в частности, в работах О.А. Демидовой. Пространственно-эконометрический инструментарий используется для анализа социально-экономических процессов на региональных данных, оценки пространственных эффектов, межрегиональных влияний и эффективности государственных программ [15]. В работе О.А. Демидовой и Д.С. Иванова на примере российских регионов показано, что экономический рост территории связан не только с ее собственными характеристиками, но и с динамикой других регионов; при этом отдельно выделяется проблема неоднородности пространственных эффектов между группами субъектов Российской Федерации [16].

Несмотря на ряд достоинств, обсуждаемые методы, как правило, опираются на предпосылки, которые в отдельных прикладных задачах пространственного анализа оказываются чрезмерно жесткими. Во-первых, в большинстве моделей пространственные эффекты фиксируются как линейные зависимости, тогда как отклик регионов на социально-экономические воздействия нередко имеет нелинейный характер и может существенно различаться в зависимости от уровня развития территории. Во-вторых, традиционные подходы в основном нацелены на оценивание «среднего» эффекта; в результате их применимость снижается при изучении асимметрии

и крайних проявлений распределений, где зачастую сосредоточены ключевые риски: регионы-лидеры и регионы-аутсайдеры способны демонстрировать принципиально различающиеся траектории демографической динамики. В-третьих, при переходе к многомерному анализу требуется более гибкое описание структуры совместных зависимостей, не сводя ее к одной мере корреляции и одновременно сохраняя информацию о том, как характер зависимости меняется на различных уровнях рассматриваемых показателей.

Отдельно стоит отметить, что исследования в области пространственного анализа демографии показывают, что при моделировании демографических процессов очень важно учитывать географическую неоднородность и пространственную структуру. Согласно исследованиям Weeks, в традиционной демографии часто игнорируют территориальные различия, считая, что демографические процессы одинаково протекают везде. В результате модели демографических переходов и миграции обычно оказываются гипотетическими и очень общими. Но эти процессы в значительной степени определяются локальными экономическими, социальными и культурными условиями и в конечном итоге они идут по-разному во многих территориях [62]. Следовательно, для решения поставленной задачи нужен специальный методологический инструментарий. Он должен быть в состоянии учитывать пространственные различия, а также давать возможность для глубокого изучения связей между демографическими изменениями и социально-экономическими факторами.

Именно поэтому растет интерес к методам, которые позволяют моделировать общую структуру зависимостей, будучи не привязанными к таким ограничениям, как линейность и симметрия. Так, регрессионные модели с использованием копула-функций могут описывать нелинейные и несимметричные взаимосвязи между большим числом переменных. Кроме того, они позволяют изучать поведение переменных при экстремальных значениях и исследовать хвостовую зависимость. Это важно при анализе и прогнозах динамики регионов с очень высокими или низкими темпами

миграции/естественного прироста. В зарубежной литературе примеров использования копула-методов в демографии совсем немного, но имеются успешные подходы применения в близких областях. К примеру, копула-методы применялись для анализа многомерной зависимости между различными аспектами бедности, где показали, что связь между такими измерениями, как доход, образование и занятость, различаются по странам и имеют асимметричные характеристики, которые стандартные корреляции не могут адекватно описывать. Применение копула-регрессии в этом контексте позволяет более четко выявлять такие эффекты [60].

Таким образом, учет пространственной структуры и территориальной неоднородности в сочетании с необходимостью моделирования нелинейных и хвостовых эффектов формирует методологическую рамку настоящего исследования и служит основанием для обращения к копула-подходу как к инструменту, позволяющему перейти от усредняющих зависимостей к более гибкому описанию региональных режимов демографической динамики.

Тем самым исследовательская проблема в рамках работы сводится к следующему: необходимо перейти от общего описания пространственной поляризации и демографических диспропорций к количественной оценке того, как социально-экономические условия развития регионов связаны с их демографической динамикой. Практически это означает выявление устойчивых зависимостей между демографическими показателями и ключевыми характеристиками региональной экономики с учетом того, что эти связи могут быть неодинаковыми для разных типов регионов, проявляться нелинейно и особенно заметно в экстремальных ситуациях.

В соответствии с этим дальнейшее изложение в главе 1 направлено на обоснование применимости копула-функций для решения поставленной задачи, а также на рассмотрение их классификации, ограничений и базовых подходов к оцениванию как методологической основы последующего эмпирического анализа.

## **1.2 Теория копула-функций: этапы развития, понятие, классификация и применимость в пространственном анализе**

Теория копула-функций, зародившаяся во второй половине XX века, претерпела значительную эволюцию, отражающую общие тенденции развития математической статистики и ее приложений. Настоящий обзор охватывает ключевые этапы этого развития, выделяя основные методологические прорывы и их практическую реализацию.

**Теоретические истоки и ранний период (1950-1990 годы).** Появление копула-функций в математической статистике связано с теоремой Склера. Эта теорема выявила ключевую связь между совместным распределением и его маргинальными распределениями посредством связующей функции – копулы. В своем труде «Fonctions de répartition à n dimensions et leurs marges» Скляр А. показал: для любого многомерного распределения, у которого есть определенные маргинальные функции распределения компонент, можно найти копулу. Она будет описывать, как маргинальные функции связаны с многомерной функцией распределения [59].

Этот вывод лег в основу теории копул: стало возможным отдельно моделировать маргинальные распределения и структуру зависимости между случайными величинами. Но в первые десятилетия после того, как работа Склера была опубликована, ее редко использовали в прикладных исследованиях – в основном она оставалась в рамках теоретической вероятности [26].

В 1970-х годах ученые стали активно изучать свойства копул в рамках многомерного анализа. Значительный вклад в развитие этой области внес Дехевельс П. [40] – он занимался исследованием эмпирических копул и тем, как они приближаются к теоретическим. Его работа «Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes» заложила основы для непараметрического оценивания копул, что позже стало ключевым в прикладных исследованиях.

Параллельно в экономике и финансах возникла потребность в более гибких моделях зависимости, чем стандартный корреляционный анализ. Однако из-за вычислительных ограничений того времени практическое использование копул оставалось редким. Лишь в конце 1980-х – начале 1990-х годов, с развитием вычислительной техники, появились первые значимые приложения, например, в актуарной математике [45] и гидрологии [27].

При всей теоретической обоснованности ранние копульные подходы ограничивались рядом факторов. Основной проблемой оставалась трудоемкая оценка параметров в многомерных постановках, поскольку прикладных алгоритмов тогда не хватало. Кроме того, доступные параметрические семейства, архимедовы и гауссовы копулы, задавали узкую область применимости. Эмпирическая верификация также была редкой: публикации чаще оставались на уровне теории без проверки на данных. Ситуация начала меняться в 1990-х годах по мере появления вычислительных процедур и расширения класса копула-моделей [23].

Таким образом, ранний этап становления теории копул (1950-1990 гг.) уместно рассматривать как период выработки ключевых теоретических предпосылок, которые в то время лишь в ограниченной степени сопровождались масштабными прикладными разработками. В последующие десятилетия, по мере роста вычислительных ресурсов и накопления эмпирических данных, копулы получили существенно более широкое распространение в экономике, финансах и ряде смежных дисциплин.

**Переход к прикладным исследованиям (1990-2000 годы).** 1990-е годы стали переломным моментом в развитии копула-методов, что было связано с несколькими ключевыми факторами:

- 1) **Развитие вычислительных мощностей.** Появление более мощных компьютеров позволило реализовать численные методы оценки параметров копул. Работа Генеста и Ривеста [47] «Statistical inference procedures for bivariate Archimedean copulas» предложила практические

методы оценки параметров архимедовых копул, что значительно упростило их применение.

2) **Расширение класса копула-моделей.** В этот период были разработаны новые семейства копул:

– эллиптические копулы (гауссова и Стьюдента) – подробно изучены в работах Фанга и др. [28] «Symmetric Multivariate and Related Distributions»;

– экстремальные копулы – развитие теории в исследованиях Гудендорфа [48];

– параметрические семейства с тяжелыми хвостами.

3) **Применение в финансах и риск-менеджменте.** Особенно значимым стало использование копул в:

– моделировании кредитных рисков [51];

– анализе рыночных зависимостей [44].

4) **Развитие многомерных подходов.** В указанный период впервые было введено понятие Vine-копул, параллельно с чем были разработаны различные методы их оценки. В числе ключевых исследований следует отметить работу Бедфорда и Кука, которая заложила основы построения гибких многомерных зависимостей посредством композиции парных копул [35]. Кроме того, Чен и Фан предложили полупараметрический подход к оценке копула-функций [38], а Генест и Ривест разработали критерий согласия для проверки гипотезы о принадлежности многомерной выборки к классу копула-функций [47].

Таким образом, для периода 1990-2000 годов характерен перенос исследовательского внимания с преимущественно теоретических построений на задачи прикладного анализа. Разработанные в это время методы позволили в определенной степени продвинуться в решении вопросов, связанных с «проклятием размерности», идентифицируемостью параметров копула-функций и процедурой выбора адекватной модели копула-функции.

**Современный этап (2010-2024 годы).** Современный этап развития характеризуется значительными достижениями в теории и применении копула-методов. Одним из ключевых результатов стало активное развитие Vine-копул, что подробно освещено в фундаментальном труде «Dependence Modeling with Copulas» [24]. В настоящей работе автор систематизировал существующие методы моделирования многомерных зависимостей, уделив особое внимание вопросам спецификации моделей и оценке параметров.

Значительное внимание уделяется исследованиям динамических копула-моделей. В статье «Copula methods for forecasting multivariate time series» [58] предложен оригинальный метод, учитывающий временную изменчивость структуры зависимостей, что стало основой для дальнейших разработок в финансовой эконометрике. За последние годы публикаций, посвященных интеграции копула-функций с GARCH-моделями для оценки волатильности активов в портфелях ценных бумаг, значительно прибавилось. В числе таких исследований выделяется работа «Measuring risk of portfolio: GARCH-copula model» [55].

В наши дни копула-методы начинают использовать и в машинном обучении. Например, в обзоре «Evading the curse of dimensionality in nonparametric density estimation» [56] показано, как можно объединить традиционные статистические методы с моделями на основе нейронных сетей. Кроме того, копула-функции оказываются полезными в задачах классификации и кластеризации: они позволяют добиться высокой точности за счет применения смесей многомерных распределений для анализа структуры данных. А в статье «Copula-based clustering methods» [41] подробно описан метод кластеризации, который как раз базируется на копулах.

Итак, эволюция теории копул – это пример того, как абстрактная математическая идея превращается в практический инструмент для анализа данных. Копулы были созданы для описания многомерных зависимостей, но из-за ограничений в вычислениях и ограниченного числа параметрических семейств долгое время применялись преимущественно в теоретических

исследованиях. Однако с ростом вычислительных возможностей и появлением запросов со стороны прикладных наук методы, основанные на копулах, стали широко использоваться на практике.

После описания этапов развития целесообразно перейти к строгим определениям и базовым положениям теории копула-функций, которые используются в последующем анализе. В частности, вводится формальное определение копулы и рассматриваются ключевые свойства и классы.

### **Определение 1 [20]**

*Пусть некоторая функция  $C(u_1, u_2, \dots, u_n)$ , зависящая от  $n$  переменных, определена на множестве  $I^n = [0; 1]^n$ , то есть на единичном кубе (то есть  $u_i \in [0; 1], i = \overline{0, n}$ ). Такая функция является **копулой**, если для нее выполняются свойства:*

- а) область значений функции расположена в интервале  $[0; 1]$ ;*
- б) если хотя бы одно  $u_i = 0, i = \overline{0, n}$ , то  $C(u_1, u_2, \dots, u_n) = 0$ ;*
- в)  $C(1, 1, \dots, u_i, \dots, 1) = 1 \forall u_i$  кроме 0;*
- г) функция  $C(u_1, u_2, \dots, u_n)$  является  $n$ -возрастающей.*

### **Теорема 1 (теорема Склера) [59]**

*Пусть  $H(x_1, \dots, x_n)$  –  $n$ -мерная функция распределения с некоторыми маргинальными распределениями  $F_1(x_1), \dots, F_n(x_n)$ . Тогда существует копула функция  $C(\cdot)$  размерности  $n$ , такая, что  $\forall x_i \in \mathbb{R}$  имеет вид формулы (1)*

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (1)$$

*В случае, когда все маргинальные распределения являются непрерывными, копула определяется однозначно единственным образом, в противном случае она может быть определена однозначно лишь на области определения  $R(F_1(x_1)) \times \dots \times R(F_n(x_n))$ ,*

*где  $R(F_i(x_i))$  – область определения маргинальных распределений.*

*Тогда можно однозначно определить и обратное, если  $C(\cdot)$  – некоторая копула-функция, то функция  $H(x_1, \dots, x_n)$  (определяется по правилу (1)) –*

многомерная функция распределения с маргинальными функциями  $F_1(x_1), \dots, F_n(x_n)$ .

В моделировании многомерных распределений с помощью копул наибольшее распространение получило следствие из теоремы Скляра, опубликованное Нельсеном позже, в 2006 году.

**Следствие [26]**

Пусть  $F_1^{-1}(y_1), \dots, F_n^{-1}(y_n)$  – обратные маргинальные функции распределения, тогда для каждого  $(y_1, \dots, y_n)$  из  $n$ -мерного единичного куба существует копула-функция и она единственная, причем представляется по формуле (2)

$$C(y_1, \dots, y_n) = H(F_1^{-1}(y_1), \dots, F_n^{-1}(y_n)). \quad (2)$$

Хоть изначально копула была определена выражением (1), основным теоретическим инструментом для построения копула-функции является соотношение (2).

Следует подчеркнуть, что копула-функции для дискретных распределений имеют значительно более сложную структуру по сравнению с непрерывными, поэтому в дальнейшем анализе мы ограничимся рассмотрением копул непрерывных статистических законов.

Ключевым свойством, которое помогает идентифицировать копула-функции, является их инвариантность относительно монотонно возрастающих преобразований случайных величин. Это свойство универсально и применимо ко всем копулам.

**Теорема 2 (свойство инвариантности) [20]**

Рассмотрим  $n$  случайных величин  $\xi_1, \dots, \xi_n$ . Пусть зависимость между этими случайными величинами определяется некоторой копула-функцией  $C(\cdot)$ . Если преобразования  $L_k: \mathbb{R} \rightarrow \mathbb{R}$ ,  $k = 1, \dots, n$  определяются строго возрастающими функциями, то структура зависимости случайных величин  $L_1(\xi_1), \dots, L_n(\xi_n)$  определяется той же копулой  $C(\cdot)$ .

Еще одно важное свойство копула-функций – их ограниченность. А именно, любая копула имеет верхнюю и нижнюю границы, известные как границы Фреше-Хеффдинга из формул (3)-(5)

$$W(y_1, \dots, y_n) \leq C(y_1, \dots, y_n) \leq M(y_1, \dots, y_n), \quad (3)$$

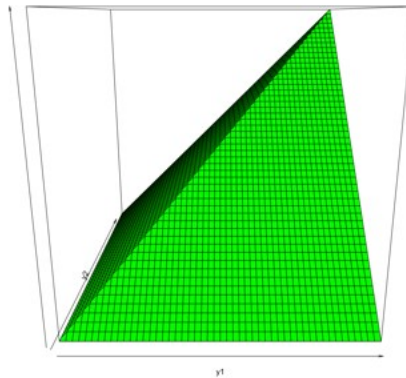
$$W(y_1, \dots, y_n) = \max(y_1 + \dots + y_n - n + 1, 0); \quad (4)$$

$$M(y_1, \dots, y_n) = \min(y_1, \dots, y_n). \quad (5)$$

Рассмотрим границы Фреше-Хеффдинга на примере стандартного равномерного распределения. Пусть  $Y_1 \sim Y_2 \sim U(0,1)$ , если  $Y_1 = Y_2$ . Тогда их копула-функция имеет вид формулы (6)

$$C(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2) = \min(y_1, y_2) = M(y_1, y_2). \quad (6)$$

Это означает, что в случае полной зависимости двух стандартных равномерных величин их копула совпадает с верхней границей Фреше-Хеффдинга  $M(y_1, y_2)$  на рисунке 1.



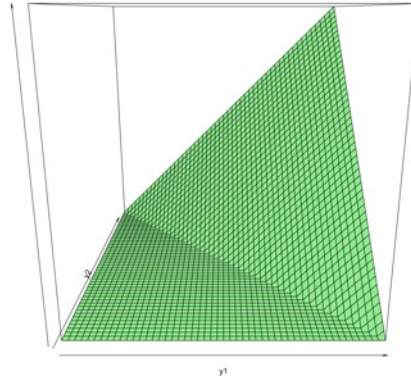
Источник: составлено автором по материалам [20].

Рисунок 1 – Верхняя граница Фреше-Хеффдинга стандартного равномерного распределения

Далее, предположим, что  $Y_2 = 1 - Y_1$ , тогда копула-функция будет иметь вид из формулы (7)

$$C(y_1, y_2) = P(Y_1 < y_1, Y_2 < y_2) = P(1 - y_2 \leq Y_1 \leq y_1) = \begin{cases} y_1 + y_2 - 1, & y_1 + y_2 > 1 \\ 0 & \text{иначе} \end{cases} = \max(y_1 + y_2 - 1, 0) = W(y_1, y_2). \quad (7)$$

В случае, когда  $Y_2 = 1 - Y_1$ , копула стандартных равномерных распределений достигает нижней границы Фреше-Хефдинга, что иллюстрируется на рисунке 2.



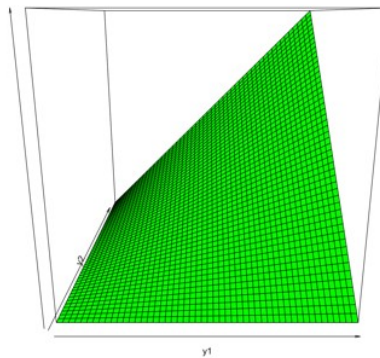
Источник: составлено автором по материалам [20].

Рисунок 2 – Нижняя граница Фреше-Хефдинга стандартного равномерного распределения

И наконец рассмотрим копулу для независимых равномерных случайных величин  $Y_1$  и  $Y_2$ , представленную формулой (8)

$$C(y_1, y_2) = P(Y_1 < y_1, Y_2 < y_2) = y_1 \times y_2 = P(y_1, y_2). \quad (8)$$

Третий случай соответствует независимости случайных величин и описывается мультипликативной копулой, изображенной на рисунке 3.

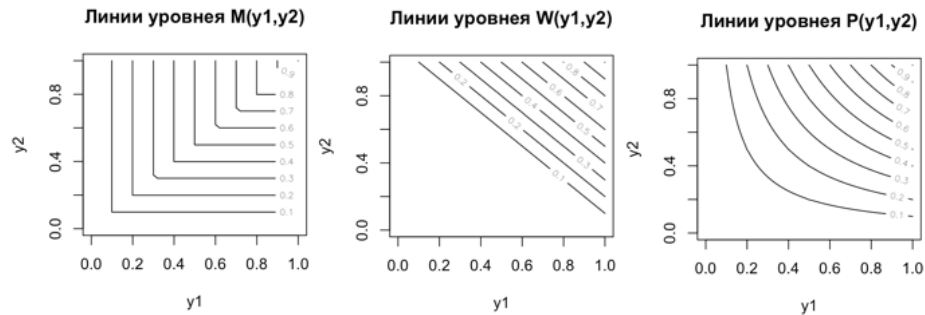


Источник: составлено автором по материалам [20].

Рисунок 3 – Копула независимых стандартных равномерных распределений

Для визуализации на рисунке 4 построим линии уровня для  $M(y_1, y_2), W(y_1, y_2), P(y_1, y_2)$ .

Важно отметить принципиальное отличие общего случая от приведенного примера: границы Фреше-Хефдинга обычно не удовлетворяют всем свойствам копула-функций.



Источник: составлено автором по материалам [20].

Рисунок 4 – Линии уровня для границ Фреше-Хефдинга и мультипликативной копулы

Если копула-функция является дифференцируемой, то, аналогично функциям распределения, для нее существует функция плотности – формула (9)

$$c(y_1, \dots, y_n) = \frac{\partial^n C(y_1, \dots, y_n)}{\partial y_1 \dots y_n}, \quad (9)$$

где  $c(y_1, \dots, y_n)$  – плотность копула-функции  $C(y_1, \dots, y_n)$ .

Плотность копула-функции  $c(F_1(x_1), \dots, F_n(x_n))$  может быть представлена через плотность многомерного распределения, рассчитанную по формулам (10) и (11)

$$\begin{aligned} f_H(x_1, \dots, x_n) &= \frac{\partial^n (C(F_1(x_1), \dots, F_n(x_n)))}{\partial F_1(x_1) \dots \partial F_n(x_n)} \prod_{i=1}^n f(x_i) = \\ &= c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f(x_i), \end{aligned} \quad (10)$$

$$c(F_1(x_1), \dots, F_n(x_n)) = \frac{f_H(x_1, \dots, x_n)}{\prod_{i=1}^n f(x_i)}. \quad (11)$$

**Определение 2 (распределения эллиптического типа) [28]**

Пусть  $\mathbf{Y}$  – случайный вектор некоторой размерности  $n$ , а  $\sigma \in \mathbb{R}^{n \times n}$  – положительно определенная и симметричная матрица. Если существует некоторое  $\mu \in \mathbb{R}^n$  и функция  $\zeta: \mathbb{R} \rightarrow \mathbb{R}^+$  такие, что характеристическая функция случайного вектора  $\mathbf{Y} - \mu$  имеет следующий вид из формулы (12)

$$\phi_{\mathbf{Y}-\mu}(s) = e^{iz^T \mu \zeta \left( -\frac{s^T \sigma s}{2} \right)} \quad \forall s \in \mathbb{R}^n, \quad (12)$$

то  $\mathbf{Y}$  называется случайным вектором, имеющим эллиптическое распределения с параметрами  $\mu$  и  $\sigma$ .

Если многомерное распределение относится к эллиптическому классу, то и соответствующая копула имеет эллиптическую структуру. К таким моделям относят симметричные семейства, среди которых можно выделить нормальное распределение, распределение Коши и распределение Стьюдента, последние два чаще всего используют для описания тяжелых хвостов. На практике эллиптические копулы применяют в актуарных задачах, при моделировании страховых и кредитных рисков. Их роль особенно велика при исследовании экстремальных явлений, где большое значение имеют совместные реализации редких, но существенных событий.

Плотность многомерного эллиптического распределения имеет следующий вид, представленный в формулах (13) и (14) [28]

$$f(\mathbf{y}) = \frac{\gamma}{\sqrt{|\sigma|}} g((\mathbf{y} - \mu)^T \sigma^{-1} (\mathbf{y} - \mu)), \quad \mathbf{y} \in \mathbb{R}^n, \quad (13)$$

$$\gamma = \frac{1}{2\pi \int_0^{+\infty} g(u) du}, \quad (14)$$

для функции  $g(\cdot)$  такой, что  $g: \mathbb{R} \rightarrow \mathbb{R}^+$  и  $\int_0^{+\infty} g(u) du$  сходится. Здесь функция  $g(\cdot)$  является оператором плотности эллиптического распределения, а  $\gamma$  нормирующей константой. Ниже, в таблице 1 представлен список операторов

плотности и нормирующих констант для некоторых эллиптических распределений.

Рассмотрим копула-функцию гауссовского распределения. Для наглядности и упрощения расчетов рассмотрим двумерный вариант задачи, поскольку в многомерном случае применяется аналогичный подход. Копула двумерного гауссовского распределения имеет вид, как показано в формуле (15)

$$C_N(y_1, y_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(y_1)} \int_{-\infty}^{\Phi^{-1}(y_2)} e^{-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}} du dv, \quad (15)$$

также можно записать через функцию распределения нормального закона, как показано в формуле (16)

$$C(y_1, y_2) = \Phi_2(\Phi^{-1}(y_1), \Phi^{-1}(y_2)), \quad (16)$$

где  $\Phi_k(\cdot)$  - функция  $k$ -мерного нормального распределения.

Таблица 1 – Функция  $g(t)$  и  $\gamma$  для распределений эллиптического типа

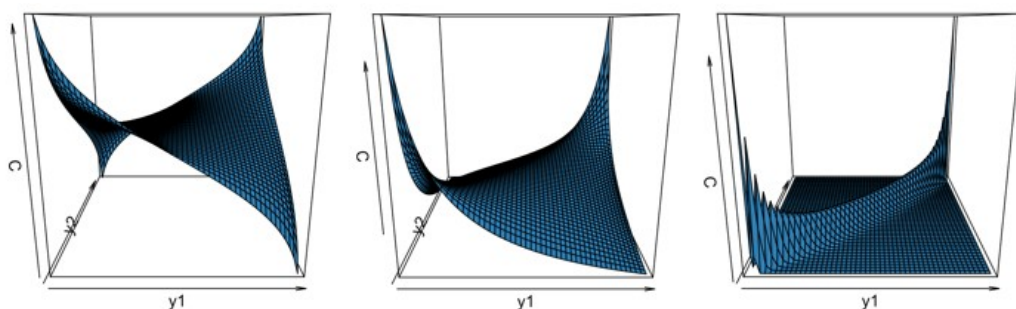
Распределение	Генератор $g(t)$	Нормирующая константа $\gamma$
Нормальное	$e^{-\frac{t}{2}}$	$(2\pi)^{-1}$
Коши	$(1+2t)^{-\frac{3}{2}}$	$(2\pi)^{-1}$
Стьюдента	$\left(1+\frac{2t}{v}\right)^{-\frac{v+2}{2}}$	$(2\pi)^{-1}$
Логистическое	$e^t(1+e^{-t})^{-2}$	$\pi^{-1}$
Лапласа	$e^{-\sqrt{t}}$	$(2\pi)^{-1}$
Котца	$e^{-\frac{rt}{2}}$	$r(2\pi)^{-1}$
Экспоненциальное	$e^{-r\left(\frac{t}{2}\right)^s}$	$sr^{\frac{1}{s}}\left(2\pi\Gamma\left(\frac{1}{s}\right)\right)^{-1}$

Источник: [20].

В общем случае копула многомерного гауссовского распределения определяется как указано в формуле (17)

$$C(y_1, \dots, y_n) = \Phi_n(\Phi^{-1}(y_1), \dots, \Phi^{-1}(y_n)). \quad (17)$$

Гауссовская копула находит широкое применение в финансовом анализе, особенно в условиях стабильных рынков, где распределения обычно не имеют тяжелых хвостов. Ее популярность обусловлена способностью эффективно моделировать высокоразмерные портфели и относительной простотой оценки параметров. На рисунке 5 представлены графики плотности двумерной нормальной копулы при различных значениях коэффициента корреляции, демонстрирующие ее свойства.



Источник: составлено автором по материалам [20].

Рисунок 5 – Плотность копулы нормального распределения при  $\rho = 0,01; 0,5; 0,99$  соответственно

Архимедовы копулы представляют собой еще один важный класс копула-функций. Вначале рассмотрим их определение для двумерного случая, а затем распространим на многомерную ситуацию.

### Определение 3 (архимедова копула) [54]

Рассмотрим строго убывающую, непрерывную и выпуклую функцию  $\varphi(u)$ , которая принимает неотрицательные значения и определена на  $u \in [0; 1]$  и  $\varphi(1) = 0$ . Пусть  $\varphi^{[-1]}(u)$  – псевдообратная функция как в формуле (18)

$$\varphi^{[-1]}(u) = \begin{cases} \varphi^{-1}(u), & 0 \leq u \leq \varphi(0) \\ 0, & u > \varphi(0) \end{cases}, \quad (18)$$

где  $\varphi^{-1}(u)$  - функция, обратная к  $\varphi(u)$ .

Функция  $C: [0; 1]^2 \rightarrow [0; 1]$ , определенная по формуле (19)

$$C(y_1, y_2) = \varphi^{[-1]}(\varphi(y_1) + \varphi(y_2)), \quad (19)$$

называется **архимедовой копулой** с функцией-генератором  $\varphi(u)$ . Кроме того, если  $\varphi(0) \rightarrow \infty$ , то  $\varphi^{[-1]}(u) = \varphi^{-1}(u)$ . В таком случае  $\varphi(u)$  называется **строгим генератором**, а  $C(y_1, y_2)$  **строгой копула-функцией**.

Плотность архимедовой копулы определяется по общему правилу, в соответствии с формулой (9), однако она может быть представлена в более развернутом виде, если функция-генератор дважды дифференцируема, тогда она имеет вид формулы (20)

$$c_\varphi(y_1, y_2) = \frac{(\varphi^{[-1]}(\varphi(y_1) + \varphi(y_2)))''}{(\varphi^{[-1]}(\varphi(y_1)))' \times (\varphi^{[-1]}(\varphi(y_2)))'}. \quad (20)$$

Рассмотрим несколько примеров архимедовых копул, которые имеют наиболее широкое применение на практике.

#### **Определение 4 (копула-функция Клейтона) [20]**

Пусть генерирующая функция  $\varphi(u)$  и обратная  $\varphi^{-1}(u)$  имеют вид, представленный в формуле (21)

$$\varphi(u) = \frac{u^{-\alpha} - 1}{\alpha}, \alpha \in [-1; +\infty) \setminus \{0\} \Rightarrow \varphi^{-1}(u) = (1 + u)^{-\frac{1}{\alpha}}. \quad (21)$$

Тогда используя соотношение (19), получаем формулу (22)

$$C_{clayton}(y_1, y_2) = \max\left((y_1^{-\alpha} + y_2^{-\alpha} - 1)^{-\frac{1}{\alpha}}, 0\right). \quad (22)$$

Функция вида (22) называется **копула-функцией Клейтона**

Стоит отметить, что при  $\alpha = -1$  копула функция Клейтона соответствует нижней границе Фреше-Хеффдинга, при  $\alpha \rightarrow 0$  случаю независимости и при  $\alpha \rightarrow +\infty$  случаю комонотонности (абсолютная зависимость).

**Определение 5 (копула-функция Гумбеля) [20]**

*Копула функция относится к классу копул Гумбеля, если функция-генератор и обратная функция имеют вид, представленный формулой (23)*

$$\varphi(u) = (-\ln u)^\alpha, \alpha \geq 1 \Rightarrow \varphi^{-1}(u) = e^{-u^{\frac{1}{\alpha}}}, \quad (23)$$

*а сама копула-функция формулой (24)*

$$C_{Gumbel}(y_1, y_2) = e^{-((- \ln y_1)^\alpha + (- \ln y_2)^\alpha)^{\frac{1}{\alpha}}}. \quad (24)$$

При  $\alpha = 1$  получаем копулу функцию двух независимых случайных величин  $Y_1$  и  $Y_2$ , а при  $\alpha \rightarrow +\infty$  для комонотонных случайных величин (абсолютно зависимых).

**Определение 6 (копула-функция Франка) [20]**

*Пусть функция-генератор и обратная функция имеют вид, представленный формулой (25)*

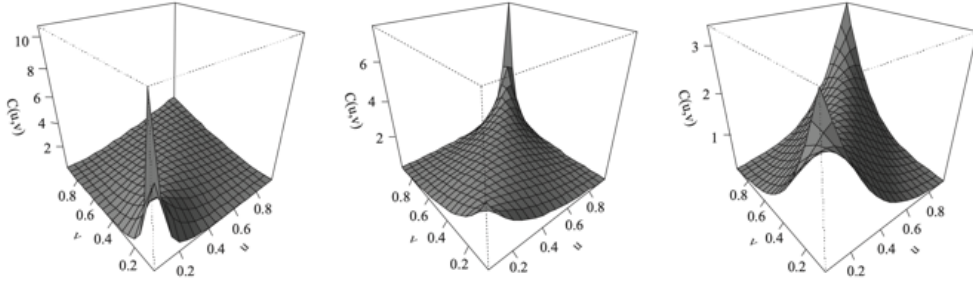
$$\varphi(u) = \ln\left(\frac{e^{-\alpha} - 1}{e^{-\alpha u} - 1}\right) \Rightarrow \varphi^{-1}(u) = -\frac{1}{\alpha} \ln(1 + e^u(e^{-\alpha} - 1)). \quad (25)$$

*Архимедова копула с таким генератором называется копулой Франка и имеет вид формулы (26)*

$$C_{Frank}(y_1, y_2) = -\frac{1}{\alpha} \ln\left(1 + \frac{(e^{-\alpha y_1} - 1)(e^{-\alpha y_2} - 1)}{e^{-\alpha} - 1}\right). \quad (26)$$

При значении  $\alpha = 0$  копула Франка совпадает с копулой независимых случайных величин, а при  $\alpha \rightarrow \pm\infty$  с верхней и нижней границей Фреше-Хеффдинга соответственно.

На рисунке 6 приведены графики копул Клейтона, Гумбеля и Франка. Их сопоставление позволяет увидеть, как различается описываемая ими хвостовая зависимость.



Источник: [20].

Рисунок 6 – Плотности копула-функций Клейтона, Гумбеля и Франка соответственно

Как видно из рисунка 6, копулы Клейтона и Гумбеля хорошо отражают хвостовую зависимость. Для первой характерна концентрация плотности в нижнем хвосте, для второй – в верхнем. Копула Франка, напротив, выраженной хвостовой зависимости не показывает.

Архимедовы копулы можно обобщить и для  $n$ -мерного случая. Тогда в общем виде копула будет иметь вид, представленный формулой (27)

$$C(y_1, \dots, y_n) = \varphi^{[-1]}(\varphi(y_1) + \dots + \varphi(y_n)). \quad (27)$$

Функция вида (27) называется серийно-итеративной. Для нее выполняется еще одно важное свойство, описанной формулой (28)

$$C(y_1, \dots, y_n) = C(C(y_1, \dots, y_{n-1}), y_n). \quad (28)$$

Формула (28) отражает ассоциативное свойство архимедовой копулы. Это означает, что многомерная зависимость может задаваться последовательно, через поэтапное добавление новых компонент. Такое представление удобно как с теоретической точки зрения, так и при численной реализации модели.

Применение копула-функций в пространственном анализе обусловлено тем, что региональные социально-экономические и демографические

показатели обычно характеризуются неоднородными распределениями и измеряются в несопоставимых шкалах. Для таких данных нередко наблюдаются асимметрия, выбросы и «тяжелые хвосты», а также различия в динамике и профиле регионов-лидеров по сравнению с регионами-аутсайдерами. В подобных ситуациях описание зависимости единственной мерой линейной корреляции фиксирует усредненную связь и потому не позволяет адекватно отразить вариации зависимости в различных областях распределения, включая его экстремальные участки.

Ключевым теоретическим основанием применения копул в пространственном анализе служит теорема Скляра, в рамках которой маргинальные распределения отделяются от структуры зависимости. В контексте межрегиональных сопоставлений это означает, что зависимость между показателями допускает анализ независимо от вида распределения каждой переменной и различий в шкалах измерения. Иначе говоря, копула задает совместную динамику ранговых позиций регионов по рассматриваемым индикаторам, благодаря чему результаты остаются сопоставимыми даже при заметной неоднородности распределений самих переменных.

Значимое методологическое преимущество связано с инвариантностью копулы к строго монотонно возрастающим преобразованиям. Для пространственных исследований это свойство принципиально, поскольку показатели нередко приводятся к сопоставимому виду посредством монотонных преобразований (например, логарифмирования, перехода к значениям на душу населения, дефлирования). При таких преобразованиях конфигурация зависимости сохраняется, следовательно, выводы о характере связей между региональными показателями не определяются выбором конкретной монотонной шкалы.

Интерпретационный каркас задают границы Фреше-Хефдинга, которые очерчивают допустимый диапазон структур зависимости: от независимости до предельных режимов совместного поведения. В задачах

пространственного развития это удобно тем, что наблюдаемую зависимость можно соотнести с понятной шкалой «слабая связанность – высокая согласованность», не редуцируя описание к единственному коэффициенту корреляции. При этом в пространственном анализе нередко информативнее не усредненная степень зависимости, а то, как она проявляется в крайних областях распределений.

Соответственно, ключевое свойство копул состоит в том, что они описывают хвостовую зависимость и позволяют различать связь в нижнем и верхнем хвостах распределений. В региональном анализе это интерпретируется как совместная динамика территорий в экстремальных состояниях: «совместная уязвимость» в нижнем хвосте и «совместное лидерство» в верхнем. Разные классы копул по-разному отражают такую асимметрию: эллиптические копулы обычно соответствуют близко симметричным структурам, тогда как архимедовы, в частности копулы Клейтона и Гумбеля, формируют преимущественно нижнюю и верхнюю хвостовую зависимость.

Эти теоретические положения задают корректный аппарат для работы с региональными данными, где важны сопоставимость показателей при различии их распределений, устойчивость выводов к монотонным преобразованиям и анализ зависимости в крайних значениях. В дальнейшем этот аппарат используется для эмпирической диагностики парных зависимостей и хвостовых эффектов, а также для непараметрического регрессионного подхода, которые раскрываются во второй и третьей главах.

### **1.3 Методы статистического оценивания параметров копул**

Переход от теоретического описания копула-функций к эмпирическому анализу требует выбора процедуры статистического оценивания параметров копулы, то есть параметров, описывающих структуру зависимости между переменными. В зависимости от того, каким образом учитываются

маргинальные распределения при построении правдоподобия, в литературе выделяют параметрические, полупараметрические и непараметрические подходы к оцениванию копула-функций.

К **параметрическим методам** оценки относятся процедуры, основанные на методе моментов и методе максимального правдоподобия (далее – ММП). В настоящем параграфе исследования рассмотрим подробнее метод максимального правдоподобия, как самый часто используемый метод с хорошими асимптотическими свойствами.

Пусть  $z_{1t}, \dots, z_{nt}$ ,  $t = 1, \dots, k$  – некоторые многомерные наблюдения,  $n$  – размерность наблюдаемых значений, а  $k$  – число наблюдений. Пусть  $f_H(\cdot)$  – плотность многомерного распределения  $(Z_1, \dots, Z_n)$ , формула (29)

$$f_H(\tilde{z}_1, \dots, \tilde{z}_n, \alpha_1, \dots, \alpha_n, \gamma) = c(F_1(\tilde{z}_1, \alpha_1), \dots, F_n(\tilde{z}_n, \alpha_n), \gamma) \times \prod_{j=1}^n f_j(\tilde{z}_j, \alpha_j), \quad (29)$$

где  $f_j$  – плотность маргинального распределения  $Z_j$ ;

$\tilde{z}_1, \dots, \tilde{z}_n$  – текущие значения случайных величин  $Z_1, \dots, Z_n$ ;

$c(\cdot)$  – плотность копула-функции.

Пусть  $\theta = (\alpha_1, \dots, \alpha_n, \gamma)$  – вектор параметров, подлежащих оцениванию, где  $\alpha_j$  – параметры маргинальных распределений случайного вектора,  $\gamma$  – вектор параметров совместной копула-функции случайного вектора. Тогда логарифм функции правдоподобия будет иметь вид, представленный формулой (30)

$$l(\theta) = \ln L(\theta) = \sum_{t=1}^k \ln(c(F_1(z_{1t}, \alpha_1), \dots, F_n(z_{nt}, \alpha_n), \gamma)) + \sum_{t=1}^k \sum_{j=1}^n \ln f_j(z_{jt}, \alpha_j). \quad (30)$$

**Оценка максимального правдоподобия** для вектора  $\theta$  будет считаться по классическому правилу  $\hat{\theta}_{\text{ОМП}} = \arg \max_{\theta} l(\theta)$ .

Сформулируем теорему о статистических свойствах оценки максимального правдоподобия.

**Теорема 3 [21]**

Пусть  $\hat{\theta}_{\text{ОМП}}$  – оценка максимального правдоподобия вектора параметров  $\theta$ . Тогда при выполнении условия регулярности оценка обладает свойствами:

- $\hat{\theta}_{\text{ОМП}} \xrightarrow{P} \theta$  – состоятельность оценок
- $\sqrt{k}(\hat{\theta}_{\text{ОМП}} - \theta) \xrightarrow{d} N(0, I^{-1}(\theta))$  – асимптотическая нормальность ( $I(\theta)$  – информационная матрица по Фишеру).

Другим параметрическим методом оценки параметров является модифицированный ММП, так называемый **двухшаговый метод максимального правдоподобия** (далее – ДММП). В этом методе по отдельности оценивается  $\alpha_j$  и  $\gamma$ .

Первый шаг метода заключается в оценивании  $\alpha_j$  по стандартному методу максимального правдоподобия одномерных случайных величин в соответствии с формулой (31)

$$\hat{\alpha}_j = \arg \max_{\alpha} l_j(\alpha) = \arg \max_{\alpha} \sum_{t=1}^k \ln f_j(z_{jt}, \alpha), \quad (31)$$

где  $l_j(\cdot)$  – логарифм функции правдоподобия маргинального распределения.

Второй шаг заключается в оценивании параметров копула-функции  $\gamma$ , используя оценки  $\hat{\alpha}_j$ , полученные на предыдущем шаге с помощью формулы (32)

$$\hat{\gamma}_{\text{О2МП}} = \arg \max_{\gamma} l_c(\gamma) = \arg \max_{\gamma} \sum_{t=1}^k \ln(c(F_1(z_{1t}, \hat{\alpha}_1), \dots, F_n(z_{nt}, \hat{\alpha}_n), \gamma)), \quad (32)$$

где  $l_c(\cdot)$  – логарифм функции правдоподобия для копулы.

Для двухшаговой процедуры так же выполняется теорема 3 и сохраняются свойства состоятельности оценок и асимптотической нормальности.

В качестве **полупараметрического метода** оценивания копула-функции можно рассмотреть канонический метод максимального правдоподобия (далее – КММП). В отличие от вышеописанных методов, здесь не делаются никакие предположения о маргинальных распределениях. КММП состоит из двух шагов.

1) Наблюдения  $z_{1t}, \dots, z_{nt}$  преобразуем  $\hat{y}_{it} = F_{ik}(z_{it})$ , здесь  $F_{ik}(z) = \frac{1}{k+1} \sum_{t=1}^k I_{\{x_{it} \leq z\}}$ . Иначе говоря,  $F_{ik}(z)$  – непараметрическая оценка маргинальной функции распределения.

2) Далее, оцениваем параметры копулы, максимизирую логарифм функции правдоподобия по формуле (33)

$$\hat{\gamma}_{\text{КММП}} = \arg \max_{\gamma} \sum_{t=1}^k \ln(c(F_{1k}(z_{1t}), \dots, F_{nk}(z_{nt}), \gamma)). \quad (33)$$

Опишем некоторые асимптотические свойства метода для двумерного случая. Введем обозначения:  $l(y_1, y_2, \gamma) = \ln c(y_1, y_2, \gamma)$ . Пусть формулы (34) и (35) имеют вид

$$S_k = \frac{1}{k} \sum_{t=1}^k l'_y(F_{1k}(z_{1t}), F_{2k}(z_{2t}), \gamma), \quad (34)$$

$$H_k = \frac{1}{k} \sum_{t=1}^k l''_{yy}(F_{1k}(z_{1t}), F_{2k}(z_{2t}), \gamma). \quad (35)$$

Так как оценка  $\hat{\gamma}_{\text{КММП}}$  является решением уравнения  $S_k = 0$ , разложив в ряд Тейлора в окрестности точки  $\hat{\gamma}_{\text{КММП}}$  до двух слагаемых  $S_k$  получим уравнение, описанное формулой (36)

$$0 = \frac{1}{k} \sum_{t=1}^k l'_y(F_{1k}(z_{1t}), F_{2k}(z_{2t}), \gamma) \cong S_k + H_k(\hat{\gamma}_{\text{КММП}} - \gamma). \quad (36)$$

Таким образом, получаем свойство оценок КММП по формуле (37)

$$\sqrt{k}(\hat{\gamma}_{\text{КММП}} - \gamma) \cong -\sqrt{k} \frac{S_k}{H_k}. \quad (37)$$

Выделим некоторые асимптотические свойства оценок КММП, представленные формулами (38), (39) и (40)

$$\begin{aligned} \hat{\gamma}_{\text{КММП}} &\rightarrow \gamma, \text{ при } k \rightarrow +\infty; \\ \sqrt{k}(\hat{\gamma}_{\text{КММП}} - \gamma) &\xrightarrow{d} N(0, \sigma^2 h^{-2}), \end{aligned}$$

где

$$\sigma^2 = \text{Var} \left( l'_y(F_{1k}(z_{1t}), F_{2k}(z_{2t}), \gamma) + W_1(z_{1t}) + W_2(z_{2t}) \right), \quad (38)$$

$$h = \mathbb{E} \left( l''_{yy}(F_{1k}(z_{1t}), F_{2k}(z_{2t}), \gamma) \right), \quad (39)$$

$$W_i(z_{it}) = - \int I_{\{F_i(z_{it}) \leq y_i\}} l''_{y_i \gamma}(y_{1t}, y_{2t}, \gamma) c(y_{1t}, y_{2t}, \gamma) dy_1 dy_2. \quad (40)$$

Также стоит отметить, что КММП хорошо работает тогда, когда случайные величины являются независимыми. В остальных случаях параметрические методы являются более эффективными.

Наконец, **непараметрические методы** оценки. В качестве простейшего примера можно рассмотреть эмпирическую копулу. Эмпирическая копула является непараметрической оценкой функции копулы, которая строится на основе рангов наблюдений. Данный инструмент используется для приближенного описания структур зависимости между случайными величинами без введения параметрических предположений о ее форме.

Для построения эмпирической копулы необходимо реализовать следующие шаги:

## 1) Преобразование данных в ранги

Для выборки  $\{(X_{11}, X_{12}, \dots, X_{1d}), \dots, (X_{n1}, X_{n2}, \dots, X_{nd})\}$  вычисляются ранги по каждой компоненте в соответствии с формулой (41)

$$R_{ij} = \text{rank}(X_{ij}), i = 1, \dots, n; j = 1, \dots, d. \quad (41)$$

Ранги нормируются на интервал  $[0,1]$ :  $U_{ij} = \frac{R_{ij}}{n+1}$ . Деление на  $n+1$  вместо  $n$  исключает вырожденность на границах.

## 2) Расчет значений копулы

Для произвольной точки  $(u_1, \dots, u_d) \in [0,1]^d$  значение эмпирической копулы вычисляется как доля наблюдений, попавших в гиперкуб  $[0, u_1] \times \dots \times [0, u_d]$ :  $\hat{C}_n(u_1, \dots, u_d) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(U_{i1} < u_1, \dots, U_{id} < u_d)$ , где  $\mathbb{I}$  – индикаторная функция (равна 1, если все условия выполнены, и 0 иначе).

Таким образом, эмпирическая копула – это ступенчатая функция, возрастающая от 0 до 1. В точке  $(u_1, \dots, u_d)$  она показывает, какая доля выборки имеет нормированные ранги, не превышающие  $u_j$  по всем координатам.

При построении моделей с многомерной зависимостью нередко бывает так, что несколько различных копула-функций обеспечивают достаточно качественное описание наблюдаемых данных. В подобных случаях требуется выбрать из множества кандидатов оптимальную модель. Для решения этой задачи применяются следующие наиболее распространенные методы:

- 1) информационный критерий Акаике (AIC) – позволяет оценить компромисс между качеством подгонки модели и ее сложностью;
- 2) тесты отношения псевдоправдоподобия – сравнивает результаты применения метода максимального правдоподобия;
- 3) байесовские методы выбора – учитывают априорные распределения параметров и обеспечивают вероятностную интерпретацию результатов.

Эти подходы позволяют систематически сравнивать конкурирующие копула-модели и выбирать наиболее адекватную для конкретного набора данных.

**Критерий Акаике** – это критерий, с помощью которого можно выбрать копулу, отвечающую минимальному из значений формулы (42)

$$AIC_c(\hat{\alpha}_l) = -2 \sum_{i=1}^k \ln c_l(y_{1i}, \dots, y_{ni}, \hat{\alpha}_l) + 2q_l, l = 1, \dots, L, \quad (42)$$

здесь  $q_l$  – штраф за количество параметров в модели. Критерий хорош тем, что он достаточно простой, однако для стандартных (не вложенных) копула-функций теряется смысл ввода штрафной части в формуле (42) [22].

Рассмотрим другой метод сравнения копула-функций – **тесты отношения псевдоправдоподобия**. Идея тестов – сравнение исследуемых моделей с точки зрения правдоподобия. Лучше та модель, которая оказалась наиболее правдоподобной. Рассмотрим нулевую гипотезу теста в соответствии с формулой (43)

$$H_0: \max_{j=2, \dots, L} \mathbb{E} \left( \ln \frac{c_j(F_1^0(\delta_{1t}), \dots, F_n^0(\delta_{nt}), \alpha_j^*)}{c_1(F_1^0(\delta_{1t}), \dots, F_n^0(\delta_{nt}), \alpha_1^*)} \right) \leq 0. \quad (43)$$

Пусть  $C_1(\cdot)$  – наилучшая из рассматриваемых моделей, модели  $C_l(\cdot)$ ,  $l = 2, \dots, L$  – остальные, а  $C_0(\cdot)$  – истинная, не наблюдаемая копула, описывающая зависимости на самом деле. Тогда при постановке гипотезы (43) используется расстояние Кульбака-Лейблера, а затем проверяется является ли модель  $C_1(\cdot)$  наилучшей среди всех остальных. Тогда очевидно, что вид альтернативной гипотезы будет иметь вид, представленный формулой (44)

$$H_1: \max_{j=2, \dots, L} \mathbb{E} \left( \ln \frac{c_j(F_1^0(\delta_{1t}), \dots, F_n^0(\delta_{nt}), \alpha_j')}{c_1(F_1^0(\delta_{1t}), \dots, F_n^0(\delta_{nt}), \alpha_1')} \right) > 0. \quad (44)$$

При принятии альтернативной гипотезы делается вывод, что среди прочих моделей есть копула, описывающая исходные данные лучше, чем  $C_1(\cdot)$ . Статистика теста рассчитывается по правилу в соответствии с формулой (45)

$$LR_n(F_{1k}, \dots, F_{nk}, \tilde{\alpha}_j, \tilde{\alpha}_1) = \frac{1}{k} \sum_{t=1}^k \left( \ln \frac{c_j(F_{1k}(\delta_{1t}), \dots, F_{nk}(\delta_{nt}), \tilde{\alpha}_j, )}{c_1(F_{1k}(\delta_{1t}), \dots, F_{nk}(\delta_{nt}), \tilde{\alpha}_1)} \right), \quad (45)$$

Здесь знак « $\sim$ » над  $\alpha$  означает, что параметры  $\tilde{\alpha}_j$  и  $\tilde{\alpha}_1$  оценены каноническим методом максимального правдоподобия.

Для принятия или отклонения основной гипотезы, в отличие от других статистических тестов, недостаточно сравнить статистику с некоторой критической точкой. Здесь необходимо проделать нетривиальную работу с использованием метода Монте-Карло. Однако, при всей сложности реализации теста, он дает хорошие результаты.

Как было сказано выше, существует еще **байесовский подход** для сравнения копул. В этом подходе параметры копулы рассматриваются как случайные величины, а значит они не подлежат оцениванию. Параметризация плотности копула-функций в этом методе рассматривается в терминах  $\tau$ -Кендалла. Априорные распределения параметров заменяются априорными распределениями  $\tau$ -Кендалла, и они берутся одними и теми же для всех тестируемых копул и используется как база при сравнении.

Пусть  $C_q(\cdot)$ ,  $q = 1, \dots, Q$  – множество копул, подлежащих сравнению. Тогда на первом шаге необходимо протестировать следующие  $Q$  гипотез:

$H_q$ : данные извлечены из копулы  $C_q(\cdot)$ ,  $q = 1, \dots, Q$ .

Для проверки гипотез необходимо вычислить вероятности реализации гипотезы  $H_q$  для заданных наблюдений  $Y$ , которые представляют собой  $k$  независимых пар  $(y_{1t}, y_{2t})$ ,  $t = 1, \dots, k$  с компонентами, имеющими равномерное распределение. Применяя теорему Байеса, получаем вероятность по формуле (46)

$$P(H_q|X, Inf) = \frac{P(X|H_q, Inf)P(H_q|Inf)}{P(X|Inf)}, \quad (46)$$

где  $P(X|H_q, Inf)$  – функция правдоподобия;

$P(H_q|Inf)$  – априорное распределение для копулы;

$P(X|Inf)$  – нормирующая константа;

$Inf$  – дополнительная информация.

Наилучшей копулой признается та, у которой максимальная апостериорная вероятность  $P(H_q|X, Inf)$ . Если записать это выражение в терминах  $\tau$ -Кендалла получим формулу (47)

$$\begin{aligned} P(H_q|X, Inf) &= \int_{-1}^1 P(H_q, \tau|X, Inf) d\tau = \\ &= \int_{-1}^1 \frac{P(X|H_q, \tau, Inf)P(H_q|\tau, Inf)P(\tau|Inf)}{P(X|Inf)} d\tau, \end{aligned} \quad (47)$$

где  $P(H_q|\tau, Inf)$  – априорная вероятность гипотезы  $H_q$  для копулы;

$P(\tau|Inf)$  – априорная плотность  $\tau$ -Кендалла;

правдоподобие  $P(X|H_q, \tau, Inf)$ , зависящее теперь еще от  $\tau$ , вычисляется по формуле (48)

$$P(X|H_q, \tau, Inf) = \prod_{t=1}^k P(y_{1t}, y_{2t}|\tau, q, Inf) = \prod_{t=1}^k c_q(y_{1t}, y_{2t}|g_q^{-1}(\tau)), \quad (48)$$

где  $c_q(y_{1t}, y_{2t}|g_q^{-1}(\tau))$  – плотность  $q$ -ой копулы, параметризованной в терминах  $\tau$ -Кендалла.

Вероятность  $P(\tau|Inf)$  можно вычислить по правилу в соответствии с формулой (49)

$$P(\tau|Inf) = \begin{cases} \frac{1}{\lambda(\Delta)}, \tau \in \Delta, \\ 0, \tau \notin \Delta \end{cases}, \quad (49)$$

где  $\Delta$  – множество, которое хранит информация о  $\tau$  (то есть, если известно, что  $\tau \geq 0$ , то  $\Delta = [0,1]$ , если о  $\tau$  ничего не известно, то  $\Delta = [-1,1]$ );  $\lambda(\Delta)$  – длина интервала (норма Лебега).

Для вероятности  $P(H_q|\tau, Inf)$  выполняется соотношение, представленное формулой (50)

$$P(H_q|\tau, Inf) \propto I_{\{\tau \in \Omega_q\}}, \quad (50)$$

где  $\Omega$  – область значений  $\tau$  в условиях выполнения гипотезы  $H_q$ .

Тогда, если подставить формулы (48), (49) и (50) в формулу (47) получим формулу (51)

$$\begin{aligned} P(H_q|X, Inf) &= \frac{1}{P(X|Inf)} \int_{-1}^1 \prod_{t=1}^k c_q(y_{1t}, y_{2t}|g_q^{-1}(\tau)) \times \frac{I_{\{\tau \in \Omega_q \cap \Delta\}}}{\lambda(\Delta)} d\tau = \\ &= \frac{1}{P(X|Inf)\lambda(\Delta)} \int_{\Omega_q \cap \Delta} \prod_{t=1}^k c_q(y_{1t}, y_{2t}|g_q^{-1}(\tau)), \end{aligned} \quad (51)$$

где нормирующую константу  $P(X|Inf)$  можно рассчитать по правилу, в соответствии с формулой (52)

$$P(X|Inf) = \sum_{q=1}^Q P(X|H_q, Inf)P(H_q|Inf). \quad (52)$$

Предлагаемый метод представляет научный интерес, поскольку, снимая необходимость предварительного параметрического оценивания, позволяет проводить сравнительный анализ копула-функций с высокой практической результативностью.

В настоящем исследовании параметры копула-функций оцениваются методом максимального правдоподобия, так как в задачах пространственного анализа важно корректно учитывать свойства маргинальных распределений региональных показателей, в том числе асимметрию и особенности поведения в хвостах. При наличии оценок маргинальных распределений ММП обеспечивает состоятельность и асимптотическую нормальность оценок и, как правило, оказывается более эффективным по сравнению с процедурами, не опирающимися на параметризацию маргиналей. Канонический ММП и непараметрические подходы можно рассматривать как альтернативы в случаях, когда получение надежных оценок маргинальных распределений затруднено; однако в рамках настоящей постановки учет маргиналей является содержательно значимым элементом анализа, поэтому в качестве базового метода оценивания параметров копула-функций используется ММП.

### **Выводы по главе 1**

В первой главе сформулированы исследуемая проблема и теоретико-методические положения, необходимые для дальнейшего анализа пространственного развития регионов Российской Федерации с учетом демографической динамики. Показано, что межрегиональная дифференциация вместе с демографическими диспропорциями проявляется в выраженной территориальной неоднородности, нелинейности взаимосвязей и асимметричности региональных реакций на социально-экономические факторы; отдельно отмечены особенности поведения показателей в крайних областях распределений. При таких свойствах данных подходы, опирающиеся на усредненные линейные зависимости, оказываются ограниченно применимыми, что делает необходимым более гибкое представление структуры зависимости.

Обоснована целесообразность использования копула-функций при анализе зависимостей между пространственными социально-экономическими

и демографическими показателями. Копульный подход разделяет описание маргинальных распределений и структуры зависимости, благодаря чему результаты сохраняют сопоставимость при различиях шкал измерения и формах распределений региональных индикаторов. Дополнительное методологическое преимущество связано с инвариантностью копул к монотонно возрастающим преобразованиям, что, в частности, повышает устойчивость выводов при нормировках и иных преобразованиях исходных данных.

Рассмотрены ключевые классы копула-функций и их интерпретация в пространственном анализе, включая средства диагностики хвостовой зависимости. Показано, что ее учет важен для изучения совместной динамики регионов в экстремальных состояниях и, в частности, для анализа территорий со стойким демографическим спадом и зон концентрации демографического и социально-экономического роста.

В заключение описаны подходы к оцениванию параметров копул, в качестве базовой процедуры выбран метод максимального правдоподобия, обеспечивающий необходимые статистические свойства оценок. Тем самым первая глава задает методологическую основу последующей эмпирической части исследования.

## Глава 2

### **Анализ и оценка парных зависимостей социально-экономических и демографических показателей регионов Российской Федерации**

Во второй главе представлена эмпирическая оценка парных зависимостей между демографическими и социально-экономическими показателями регионов Российской Федерации. Такой подход позволяет перейти от общей постановки проблемы к количественному описанию связей на региональном массиве данных и, кроме того, выделить особенности выявляемых зависимостей, которые целесообразно учитывать на этапе последующего многомерного моделирования.

Региональные показатели заметно различаются как по масштабу, так и по форме распределений; в этой ситуации опора исключительно на усредненные линейные характеристики способна приводить к искаженному представлению о характере взаимосвязей. В главе применяется набор диагностических и аналитических процедур, позволяющих рассматривать зависимость более содержательно, в том числе с учетом возможной асимметрии и поведения в крайних областях распределений.

В начале главы даются описание данных и первичная диагностика распределений и вариативности показателей. Затем излагаются теоретические положения, необходимые для дальнейшего анализа: обсуждаются меры зависимости, их экономическая интерпретация, а также вводятся кривые копула-зависимостей как инструмент изучения парных взаимосвязей. В заключительной части выполняется эмпирическая оценка структуры зависимостей между демографическими и социально-экономическими показателями с акцентом на хвостовую зависимость.

## 2.1 Описание и анализ данных и переменных исследования

Эмпирический анализ пространственной демографической и социально-экономической динамики регионов Российской Федерации основан на панельных данных, объединяющих временной и территориальный разрезы. Исследование охватывает период 2010-2023 годов, в течение которого была обеспечена полнота и сопоставимость статистической информации по всем интересующим показателям.

Географический охват включает все регионы Российской Федерации, за исключением Республики Крым, Севастополя и регионов, присоединенных к Российской Федерации в 2022 году (Донецкая Народная Республика, Луганская Народная Республика, Запорожская область и Херсонская область). Исключение этих территорий обосновано недостаточностью исторических данных для построения статистически надежных оценок: указанные регионы не имеют полных временных рядов, начиная с 2010 года, что нарушило бы требование сбалансированной панели, необходимой для корректного применения копула-регрессии.

Для целей моделирования панель вида «год × регион» была преобразована в единую совокупность наблюдений. Такое объединение допустимо, поскольку используемые показатели заданы либо в относительной форме, либо в расчете на душу населения. За счет этого различия, обусловленные масштабом региона, в значительной степени снимаются, а данные становятся сопоставимыми между собой.

В рамках принятой постановки каждое наблюдение вида «год-регион» рассматривается как отдельная реализация изучаемых социально-экономических и демографических характеристик. Это позволяет работать с объединенной выборкой объемом около 1000 наблюдений.

**Выбор эндогенной переменной.** Ключевой переменной исследования являются годовые темпы прироста численности населения по регионам

Российской Федерации, рассчитанные как процентное изменение между соседними годами по формуле (53)

$$\Delta P_t = \frac{P_t - P_{t-1}}{P_{t-1}}, t = 2010, \dots, 2023, \quad (53)$$

где  $P_t$  – численность населения в регионе в году  $t$ .

Демографический рост – центральный показатель, интегрирующий воздействие множества социально-экономических факторов. Согласно Стратегии-2030, изменения численности населения рассматриваются как ключевой показатель, отражающий качество условий жизни, доступность занятости и ожидаемые траектории экономического развития регионов. Использование годовых темпов приростов имеет несколько преимуществ:

- **детрендирование**: применение разностей первого порядка позволяет ослабить влияние долгосрочного тренда убыли или прироста, связанного с историческими и географическими условиями, а также с устойчивыми демографическими факторами, которые, как правило, не изменяются в краткосрочном периоде;

- **ориентация на краткосрочные изменения**: приросты между смежными годами позволяют судить о том, как население реагирует на актуальную ситуацию в местной экономике, параметры занятости, состояние жилищного фонда и уровень доходов – то есть на факторы, которые в той или иной степени могут быть объектом государственной политики;

- **сопоставимость масштабов**: применение темпов прироста делает корректным сравнение как крупных регионов (Московская область, Санкт-Петербург), так и небольших (Магаданская область, Сахалин), поскольку в абсолютных значениях приросты и сокращения нередко различаются на несколько порядков.

В рассматриваемый период годовые темпы прироста по субъектам Российской Федерации находились в диапазоне от  $-2,3\%$  до  $+2,4\%$  в годовом

выражении, что указывает на выраженную асимметрию распределения. Для большинства регионов были характерны умеренный прирост либо сокращение численности, тогда как ограниченное число территорий, отличавшихся высокой миграционной привлекательностью, демонстрировало заметно более быстрый рост. Эта асимметрия и нелинейность являются основной причиной применения копула-подхода.

Набор экзогенных переменных составлен на основе экономической теории пространственного развития и ключевых социально-экономических показателей, отраженных в Стратегии-2030. Все данные достоверны и получены с сайта Федеральной службы государственной статистики Росстат [4], пример данных для 2023 года можно увидеть в приложении А.

1) **Валовой региональный продукт на душу населения (ВРП on capita).** ВРП на душу населения служит агрегированной характеристикой экономической производительности региона. Согласно моделям миграции, люди переселяются в регионы с более высокой предельной производительностью труда, которая коррелирует с ВРП на душу. Кроме того, регионы с высоким ВРП обладают большей фискальной способностью инвестировать в социальную инфраструктуру, образование и здравоохранение, что повышает привлекательность для молодого трудоспособного населения.

2) **Среднедушевые месячные располагаемые доходы населения (далее - СДМРД).** В отличие от ВРП, которая отражает экономическую мощь региона, располагаемые доходы характеризуют реальное материальное благосостояние домохозяйств. Именно эта переменная непосредственно влияет на потребительское поведение, способность оплачивать жилье (ипотека, аренда) и решение о миграции. Локальная кластеризация доходов, проявляющаяся в их концентрации в крупных городах, прежде всего в Москве и Санкт-Петербурге, выступает значимым фактором притяжения для миграции молодежи и высокообразованных работников. В результате

принимающие территории в определенной мере демографически «омолаживаются», а их демографический прирост, как правило, усиливается.

3) **Индекс потребительских цен на товары и услуги к прошлому году** (далее - **ИПЦ**). Инфляция неоднозначно соотносится с демографической динамикой. С одной стороны, повышенная региональная инфляция ведет к снижению реальных доходов и покупательной способности населения, что, как правило, уменьшает привлекательность территории для мигрантов. С другой стороны, при высоких темпах инфляции домохозяйства нередко ускоряют решение о покупке жилья, стремясь опередить дальнейшее повышение цен, вследствие чего предложение на рынке недвижимости может кратковременно расти. Вместе с тем стойко высокая инфляция обычно воспринимается как симптом макроэкономической нестабильности и часто сопровождается спадом инвестиционной и миграционной активности. Учет ИПЦ в модели позволяет учесть влияние регионально специфической макроэкономической нестабильности.

4) **Уровень безработицы (в процентах в отчетном году)**. Безработица является ключевым показателем состояния рабочего рынка. Обычно ее повышение снижает привлекательность миграции, так как ухудшает перспективы поиска работы, а иногда и приводит к оттоку мигрантов из региона. В то же время этот показатель говорит о циклических макроэкономических дисбалансах, в том числе о рецессии и снижении спроса на рабочую силу в регионе, что важно учитывать в моделировании.

5) **Число браков на 1000 населения в год**. Показатель, характеризующий социально-демографическую устойчивость и в определенной мере ожидания населения относительно будущего. Рост брачности обычно соотносится с повышением рождаемости, однако в России в последние годы эта зависимость ослабевает. Интенсивность заключения браков также может служить косвенным индикатором социального благополучия территории: чаще вступают в брак более молодые, социально активные и экономически обеспеченные группы. Регионы с увеличением

брачности нередко демонстрируют более благоприятные демографические тенденции и поступательное развитие социальной инфраструктуры.

**6) Индекс изменения стоимости жилья к прошлому году.**

Стоимость жилья – ключевой фактор, влияющий на миграционные решения молодых семей и работников, вступающих на рынок труда. Резкий рост цен на жилье может отпугнуть потенциальных мигрантов (особенно в регионы типа Москвы и Санкт-Петербурга), но также может быть сигналом растущего спроса и улучшающихся перспектив региона. В России структура рынка жилья сильно варьируется между регионами: в крупных агломерациях жилье дорожает, привлекая инвесторов и способствуя миграции, в то время как в периферийных регионах падение цен на жилье отражает демографический упадок.

**7) Инвестиции в основной капитал на душу населения.**

Уровень инвестиций в основной капитал отражает ожидания бизнеса относительно будущих экономических перспектив региона. Высокие инвестиции сигнализируют о создании новых рабочих мест, развитии инфраструктуры и долгосрочного экономического роста. Это, в свою очередь, привлекает трудовую миграцию, особенно молодых специалистов. Падение инвестиций может быть признаком экономического неблагополучия и сопровождается оттоком населения в более процветающие регионы.

**8) Доля городского населения в регионе (в процентах).**

Доля городского населения отражает степень урбанизации региона. Высокоурбанизированные регионы (большинство населения в городах) обычно имеют лучше развитую инфраструктуру (транспорт, здравоохранение, образование), более разнообразный рынок труда и, как следствие, более высокий уровень миграционной привлекательности. Кроме того, урбанизированные регионы характеризуются другими демографическими паттернами: более низкой рождаемостью (из-за альтернативных затрат на образование и карьеру) и более высокой смертностью (старение населения).

Включение этой переменной позволяет контролировать структурные различия между типами регионов.

9) **Расстояние от административного центра региона до Москвы (в километрах).** Расстояние до столицы – фундаментальная пространственная переменная, капсулирующая несколько экономических эффектов. Во-первых, близость к Москве снижает транспортные издержки для деловых контактов, что привлекает инвестиции и компании, генерирующие занятость. Во-вторых, близость облегчает маятниковую миграцию (работа в столице с проживанием в соседних регионах). В-третьих, согласно теории пространственной экономики, столица как центр притяжения оказывает «гравитационное» воздействие, которое убывает с расстоянием. Для России Москва и в меньшей степени Санкт-Петербург являются доминирующими полюсами притяжения, создавая мощный центр-периферийный градиент в экономическом развитии и демографической динамике.

Далее выполняется предварительный статистический анализ выбранных переменных. На данном этапе представляются описательные характеристики и распределения показателей демографической динамики, а также основных социально-экономических переменных (ВРП на душу населения, доходы, инвестиции, безработица, ИПЦ, параметры рынка жилья, брачность, урбанизация и удаленность от Москвы). Дополнительно рассчитываются характеристики формы распределений, включая асимметрию и эксцесс, и приводятся сводные статистики.

Результаты первичного анализа сведены в таблицу 2. По каждой переменной вычислены показатели центральной тенденции (среднее и медиана), диапазон значений (минимум и максимум), а также параметры, отражающие форму распределения, а именно асимметрия и эксцесс. Кроме того, нормальность распределений проверена с использованием критерия Jarque-Bera.

Таблица 2 – Описательная статистика исходных данных

Переменная	Среднее	Медиана	Стандартное отклонение	Минимальное	Максимальное	Асимметрия	Эксцесс	p(JB)
Y - темпы прироста населения	-0,0030	-0,0043	0,0086	-0,0228	0,0234	0,5580	-0,0449	0,000
X1 - ВРП на душу	721038,8	425801,3	1144607,9	91610,2	11995390,0	5,8040	41,7518	0,000
X2 - среднедушевые доходы	33188,1	28006,0	17133,0	11700,0	156988,0	2,7054	10,3607	0,000
X3 - безработица	8,0258	7,0070	4,5682	1,4530	44,2080	2,5145	10,1358	0,000
X4 - инвестиции	177804,6	88084,9	333948,4	23297,4	2866067,0	5,1687	29,3069	0,000
X5 - браки	6,7845	6,7000	1,4356	1,8000	11,0000	0,0337	0,0224	0,911
X6 - стоимость жилья	105,6524	103,8800	9,6276	85,4600	166,8200	1,7439	5,5811	0,000
X7 - ИПЦ	107,8147	106,9700	4,4277	100,0700	122,0900	0,3646	-0,4793	0,000
X8 - доля городского населения	70,5641	71,5900	12,7145	29,3000	100,0000	-0,5035	1,0097	0,000
X9 - расстояние до Москвы	1776,471	1160,000	1807,987	0,0000	6800,0000	1,4806	1,0925	0,000

Источник: составлено автором.

Анализ таблицы 2 показывает, что переменная (Y), характеризующая динамику изменения численности населения, изменяется в промежутке от -0,0228 до 0,0234. Отрицательное среднее (-0,003) и медиана (-0,0043) указывают на то, что в большинстве случаев население сокращается, а увеличение численности встречается не так часто. Положительная асимметрия (0,558) говорит о том, что имеются отдельные случаи, когда темпы прироста населения были заметно выше среднего. Значение эксцесса близко к 0 (-0,045), однако тест Jarque-Bera показывает (p-value  $\approx$  0), что распределение данной выборки существенно отличается от нормального.

Ярко выраженные неоднородность и отклонение от нормального распределения присутствует у показателей, связанных с экономическим развитием и инвестициями. Так ВРП на душу населения (X1) распределено очень несимметрично (показатель асимметрии, 5,804, показатель эксцесса,

41,752), и тест Jarque-Bera однозначно отвергает гипотезу о нормальности распределения ( $p\text{-value} \approx 0$ ). Это означает, что высокий ВРП характерен только для немногих регионов, на которые приходится основной объем экономической активности. Похожая ситуация складывается с инвестициями в основной капитал на душу населения (X4): распределение крайне несимметричное, асимметрия 5,169, эксцесс 29,307,  $p\text{-value} \approx 0$ . Получается, что ключевые экономические показатели во многом определяются успехами нескольких регионов-лидеров.

Среднедушевые доходы (X2) и уровень безработицы (X3) оказались сдвинутыми вправо и имеют положительный эксцесс. Асимметрия выборки (X2) равна 2,705, эксцесс 10,361 ( $p\text{-value} \approx 0$ ), что говорит о наличии регионов с очень высокими доходами. Для (X3) значения асимметрии 2,515, а эксцесса 10,136 ( $p\text{-value} \approx 0$ ) указывает на неравномерность рынка труда и наличие районов с высоким уровнем безработицы. Изменение стоимости жилья (X6) варьируется в более узких диапазонах, однако распределение выборки также не является нормальным: асимметрия, 1,744, эксцесс 5,581 ( $p\text{-value} \approx 0$ ), что означает, что распределение сдвинуто в сторону более высоких темпов роста цен.

Для индекса потребительских цен (X7) показатели отклонения от нормальности выражены не так ярко, как для большинства остальных переменных: асимметрия 0,365, эксцесс -0,479, то есть распределение обладает более «плоским» видом. Однако, результаты теста Jarque-Bera также свидетельствуют о присутствии статистически значимых отклонений от нормального распределения ( $p\text{-value} \approx 0$ ).

Доля городского населения (X8) имеет скошенность влево (0,504) и положительный эксцесс (1,01), что указывает на неоднородность по степени урбанизации: в большинстве регионов удельный вес городского населения достаточно высок, однако встречаются и такие, где уровень урбанизации крайне низок. Расстояние до Москвы (X9) изменяется в диапазоне от 0 до

6800 километров, а само распределение имеет положительную асимметрию (1,481) и эксцесс (1,093).

Уровень брачности (X5) примечателен тем, что среди всех переменных, рассматриваемых в исследовании, он наиболее близок к нормальному распределению. Значение коэффициента асимметрии для него близко к нулю (0,034), эксцесс также незначителен (0,022), а критерий нормальности Jarque-Bera не отклоняет нулевую гипотезу о нормальности ( $p$ -value = 0,911). Это означает, что показатель распределен относительно равномерно, без экстремальных значений. В этом смысле, показатель брачности отличается от других социально-экономических факторов.

Для дополнительной характеристики пространственной организации данных был рассчитан индекс Морана на срезе за 2023 год. Матрицу пространственных весов сформировали по геометрии субъектов Российской Федерации (GADM, уровень ADM1), задав структуру  $k$  ближайших соседей на основе центроидов ( $k = 4$ ) с последующей симметризацией. Статистическую значимость оценивали с помощью пермутационного теста (999 перестановок). Расчеты выполнены на согласованной выборке, включающей 83 региона за 2023 год. Полученные результаты представлены в таблице 3.

Таблица 3 – Индекс Морана для исходных переменных

Переменная	Moran's I	p-value
Y - темпы прироста населения	0,2490	0,002
X1 - ВРП на душу населения	0,3330	0,001
X2 - среднедушевые доходы	0,4161	0,001
X3 - безработица	0,4219	0,001
X4 - инвестиции	0,3153	0,002
X5 - браки	0,5449	0,001
X6 - стоимость жилья	0,0429	0,534
X7 - ИПЦ	0,1300	0,043
X8 - доля городского населения	0,4024	0,001
X9 - расстояние до Москвы	0,9839	0,001

Источник: составлено автором.

Оценки индекса Морана свидетельствуют о статистически значимой положительной пространственной автокорреляции демографического прироста (Y) и большинства рассматриваемых факторов. Так, для Y значение Moran's I составляет 0,249 ( $p\text{-value}=0,002$ ), что указывает на пространственную кластеризацию: соседние регионы, как правило, характеризуются близкими темпами изменения численности населения. Для ВРП на душу населения, доходов, безработицы, инвестиций и доли городского населения также получены устойчиво положительные значения Moran's I (в интервале 0,315-0,422) при статистической значимости на уровне  $p\text{-value} \approx 0,001\text{-}0,002$ ; данный результат согласуется с территориальной концентрацией экономической активности и пространственной неоднородностью рынка труда. Показатель брачности демонстрирует наиболее выраженную кластеризацию среди «социальных» переменных (Moran's I=0,545;  $p\text{-value}=0,001$ ), что может отражать устойчивые различия в демографическом поведении на смежных территориях. Для индекса потребительских цен пространственная автокорреляция менее выражена, однако сохраняет статистическую значимость (Moran's I=0,130;  $p\text{-value}=0,043$ ). В то же время для индекса изменения стоимости жилья статистически значимая пространственная автокорреляция не выявляется (Moran's I=0,043;  $p\text{-value}=0,534$ ), что согласуется с более локальным и сегментированным характером региональных рынков жилья. Переменная расстояния до Москвы, как и ожидается, в значительной степени задается географией и демонстрирует близкую к максимальной пространственную упорядоченность (Moran's I = 0,984;  $p\text{-value} = 0,001$ ).

Анализ показал, что многие из исследуемых социально-экономических показателей отличаются значительной асимметрией и существенно отклоняются от нормального распределения. Также у ряда переменных наблюдается скопление значений в «хвостах» распределений – особенно это заметно в данных, которые отражают экономическую мощь и инвестиционную активность. В таких обстоятельствах описание взаимосвязей

между демографическими изменениями и соответствующими факторами только с помощью линейной корреляции и усредненных показателей зависимости может оказаться недостаточным. Есть риск потерять важную информацию о том, как переменные варьируются совместно.

Обнаруженная пространственная автокорреляция большинства показателей также указывает на то, что они имеют тенденцию группироваться в определенные территориальные кластеры. Следовательно, изучая зависимости, нам необходимо учитывать не только общие эффекты, но и различия между группами соседних регионов, обладающих схожими характеристиками.

Как раз по этой причине с методологической точки зрения оправдано применение копул. Этот инструмент позволяет весьма гибко выстраивать структуру взаимосвязей, практически не накладывая строгих требований к нормальности и линейности, тем самым давая возможность эффективно работать с неоднородными распределениями в региональных данных.

## **2.2 Меры зависимости случайных величин и их экономическая интерпретация**

Предварительный анализ показал, что распределение основных социально-экономических показателей далеко не всегда близко к нормальному. Часто наблюдаются асимметричные распределения с тяжелыми хвостами, что особенно заметно на примере данных ВРП на душу населения и инвестиций в основной капитал. В таких случаях коэффициент корреляции Пирсона не всегда является эффективной мерой для описания зависимостей, потому что он оценивает их средний линейный характер. Для региональных данных это важно, поскольку в крайних значениях зачастую содержится важная информация о ситуации как в регионах-лидерах, так и в отстающих. Поэтому далее используются методы, учитывающие ранги и согласованность экстремальных значений.

В этом разделе последовательно рассмотрим ранговые меры зависимости (коэффициенты Спирмена  $\rho$  и Кендалла  $\tau$ ), а также коэффициенты верхней и нижней хвостовой зависимости. Ранговые показатели позволяют определить направление и силу монотонной связи, а хвостовые коэффициенты количественно характеризуют, насколько согласованно ведут себя переменные в экстремальных областях распределения. В дальнейшем эти инструменты будут использованы двумя способами: во-первых, для интерпретации парных зависимостей в региональных данных, во-вторых, для сравнения различных семейств копул и анализа хвостовых эффектов.

Меры зависимости применяются для описания характера и структуры связей между случайными величинами. Среди наиболее широко используемых показателей выделяют коэффициент корреляции: он, как правило, адекватно фиксирует линейную зависимость. Вместе с тем принципиальное ограничение этого подхода состоит в том, что он чувствителен преимущественно к линейным отношениям и, следовательно, может не выявлять более сложные, в том числе нелинейные, формы взаимосвязи.

В региональном разрезе это означает, что влияние различных факторов на демографическую динамику способно проявляться неодинаково на разных уровнях показателей. Так, увеличение доходов населения или рост инвестиционной активности могут ассоциироваться с более быстрым демографическим приростом главным образом среди регионов, которые уже характеризуются высокой миграционной привлекательностью; при этом для значительной части территорий соответствующая связь может быть менее выраженной либо принимать иной вид. По этой причине для корректной интерпретации целесообразно использовать меры зависимости, которые отражают монотонные соотношения и не сводят их к предпосылке линейности.

В таких случаях в исследовании для анализа нелинейных зависимостей применяются используются ранговые меры зависимости, которые допускают представление через копулу. К ним относятся коэффициент ранговой корреляции Спирмена ( $\rho$ ) и коэффициент Кендалла ( $\tau$ ), которые позволяют оценивать силу и направление нелинейных связей. Коэффициенты  $\rho$ -Спирмена и  $\tau$ -Кендалла – метрики, основанные на согласованности и рассогласованности случайных величин.

**Определение 7 (Коэффициент  $\rho$ -Спирмена) [21]**

*Для двух непрерывных случайных величин  $Y_1$  и  $Y_2$  коэффициент  $\rho$ -Спирмена определяется формулой (54)*

$$\begin{aligned}\rho_S(Y_1, Y_2) &= 3Q(C, P) = 12 \int \int y_1 y_2 dC(y_1, y_2) - 3 = \\ &= 12 \int \int C(y_1, y_2) dy_1 dy_2 - 3,\end{aligned}\quad (54)$$

где  $y_1 = F_{Y_1}(x_1)$ ,  $y_2 = F_{Y_2}(x_2)$ ;

$Q(\cdot)$  – разность вероятности согласованности и рассогласованности случайных величин.

Иначе говоря, если выразить коэффициент корреляции  $\rho$ -Спирмена через классическую корреляцию Пирсона, получим формулу (55)

$$\rho_S(Y_1, Y_2) = \rho(F(Y_1), F(Y_2)), \quad (55)$$

где  $F(Y_1)$  и  $F(Y_2)$  – маргинальные функции распределения.

Несмещенная оценка коэффициента корреляции  $\rho$ -Спирмена для выборки  $(Y_1, Y_2)$  объема  $k$  с элементами  $Y_{1i}, Y_{2i}$  выражается формулой (56)

$$\hat{\rho}_S = \frac{6}{k(k-1)(k-2)} \sum_{1 \leq i < j < l \leq k} 3 \operatorname{sign}(Y_{1i} - Y_{1j}) \operatorname{sign}(Y_{2i} - Y_{2l}). \quad (56)$$

**Определение 8 (Коэффициент  $\tau$ -Кендалла) [21]**

Пусть  $(Y_1, Z_1)$  – некоторый случайный вектор, имеющий совместную функцию распределения  $H(y, z)$ , а  $(Y_2, Z_2)$  – случайный вектор, где  $Y_2$  и  $Z_2$  независимы и распределения случайных величин  $Y_2$  и  $Y_1$ ,  $Z_2$  и  $Z_1$  совпадают. Коэффициент  $\tau$ -Кендалла для вектора  $(Y, Z)$  определяется как разница между вероятностью согласованности и рассогласованности согласно формуле (57)

$$\begin{aligned} \tau(Y, Z) &= P((Y_1 - Y_2)(Z_1 - Z_2) > 0) - P((Y_1 - Y_2)(Z_1 - Z_2) < 0) = \\ &= P(Y_1 > Y_2, Z_1 > Z_2) + P(Y_2 > Y_1, Z_2 > Z_1) - \\ &\quad - P(Y_1 < Y_2, Z_1 > Z_2) - P(Y_1 > Y_2, Z_1 < Z_2) \end{aligned} \quad (57)$$

Коэффициент  $\tau$ -Кендалла так же может быть выражен через копулу пары случайных величин формулой (58)

$$\tau(Y_1, Y_2) = 4 \int_0^1 \int_0^1 C(y_1, y_2) dC(y_1, y_2) - 1. \quad (58)$$

Несмещенная оценка коэффициента  $\tau$ -Кендалла для пары выборок  $(\mathbf{Y}, \mathbf{Z})$  определяется формулой (59)

$$\hat{\tau} = \frac{2}{k(k-1)} \sum_{1 \leq i < j \leq k} \text{sign}(Y_i - Y_j) \text{sign}(Z_i - Z_j). \quad (59)$$

В контексте исследуемой проблемы положительные значения коэффициентов Спирмена ( $\rho$ ) и Кендалла ( $\tau$ ) между демографическим ростом и фактором означают, что более высокие значения этого показателя соответствуют чаще всего и более высоким темпам прироста населения в регионах. При отрицательных значениях коэффициентов наблюдается обратная зависимость: увеличение показателя сопровождается снижением темпов демографического развития. Ранговые коэффициенты выявляют монотонную связь и сохраняют интерпретируемость вне зависимости от

формы распределений, что ценно при сравнении регионов, так как позволяет учитывать различия в распределениях данных.

Помимо коэффициентов ранговой корреляции, для анализа зависимостей применяются коэффициенты верхней и нижней хвостовой зависимости. Коэффициенты хвостовых зависимостей дают информацию о тесноте связи двух случайных величин в экстремальных областях распределения, при крайне высоких (верхний хвост) и крайне низких (нижний хвост) значениях.

Такие характеристики довольно часто применяются при управлении рисками, в финансах, страховании и актуарных науках, где знание связей между переменными в экстремальных ситуациях имеет решающее значение. Иными словами, хвостовая зависимость – это своеобразный ключ к пониманию того, как ведет себя нелинейная и асимметричная корреляция в области тяжелых хвостов распределения.

В контексте региональной экономики хвостовую зависимость можно рассматривать содержательно. Верхний хвост отражает синхронность экстремально высоких значений у «лидирующих» регионов, то есть вероятность совместного появления высоких доходов, инвестиций или цен на жилье и ускоренного прироста населения. Нижний хвост фиксирует сопряженность неблагоприятных экстремумов: вероятность того, что очень низкие доходы или инвестиции, а также повышенная безработица сопутствуют наиболее выраженной демографической убыли. Таким образом, коэффициенты  $\lambda^U$  и  $\lambda^L$  дают возможность различать режимы зависимости для лидеров и аутсайдеров, что напрямую соотносится с задачами пространственной политики и обоснованием дифференцированных мер развития.

**Определение 9 (верхняя хвостовая зависимость) [21]**

*Пусть  $(\xi_1, \xi_2)$  – некоторый случайный вектор, компоненты которого непрерывные случайные величины с частными функциями распределения  $F_{\xi_1}$  и*

$F_{\xi_2}$ . Тогда коэффициент верхней хвостовой зависимости  $\lambda^U$  случайных  $\xi_1$  и  $\xi_2$  определяется формулой (60)

$$\lambda^U = \lim_{s \rightarrow 1} P \left( \xi_2 > F_{\xi_2}^{-1}(s) \mid \xi_1 > F_{\xi_1}^{-1}(s) \right) = \lim_{s \rightarrow 1} \frac{1 - 2s + C(s, s)}{1 - s}, \quad (60)$$

при условии, что пределы существуют.

Если коэффициент верхней хвостовой зависимости удовлетворяет условию  $\lambda^U \in (0, 1]$ , то случайные величины  $\xi_1$  и  $\xi_2$  называются асимптотически зависимыми в верхнем хвосте. Это говорит о том, что в экстремальных ситуациях между переменными наблюдается стабильная взаимосвязь: если одна из них достигает аномально высоких значений, то с определенной вероятностью вторая переменная тоже будет иметь экстремальные значения.

Если значение  $\lambda^U$  равно нулю, тогда переменные считаются асимптотически независимыми в верхней части распределения. Иными словами, одна переменная может принимать экстремальные значения, и это ни коим образом не влияет на вероятность того, что с ней окажется в экстремальной зоне и вторая.  $\lambda^U$  можно рассматривать как условную вероятность. К примеру, если уровень доходов или объем инвестиций в регионе достигает высоких квантилей, то  $\lambda^U$  показывает, насколько вероятно, что и темпы роста населения в этом регионе тоже окажутся в верхней, экстремальной зоне.

Эта характеристика важна, когда речь идет о редких, но чрезвычайно опасных явлениях, таких, как финансовые кризисы, природные катастрофы или системные сбои. В подобных сценариях важно понимать, насколько велик риск того, что несколько экстремальных событий возникнут одновременно.

#### **Определение 10 (нижняя хвостовая зависимость) [21]**

Пусть  $(\xi_1, \xi_2)$  – некоторый случайный вектор, компоненты которого непрерывные случайные величины с частными функциями распределения  $F_{\xi_1}$  и

$F_{\xi_2}$ . Тогда коэффициент нижней хвостовой зависимости  $\lambda^L$  случайных  $\xi_1$  и  $\xi_2$  определяется формулой (61)

$$\lambda^L = \lim_{s \rightarrow 0} P \left( \xi_2 \leq F_{\xi_2}^{-1}(s) \mid \xi_1 \leq F_{\xi_1}^{-1}(s) \right) = \lim_{s \rightarrow 0} \frac{C(s, s)}{s}. \quad (61)$$

Случайные величины  $\xi_1$  и  $\xi_2$  называются асимптотически зависимыми в нижнем хвосте, если коэффициент нижней хвостовой зависимости  $\lambda^L$  принимает значения из интервала  $(0, 1]$ . Это говорит о том, что переменные тесно связаны в ситуации экстремально низких значений: если одна из них достигает аномально низкого уровня, то есть вероятность, что и вторая величина тоже окажется в экстремально низкой зоне.

Если  $\lambda^L$  равна нулю, переменные считаются асимптотически независимыми в нижней части распределения. Это значит, что даже если одна из них достигает экстремально низких значений, это не влияет на вероятность того, что вторая переменная также окажется в экстремально низкой зоне.

Содержательно  $\lambda^L$  отражает вероятность: «При условии, что  $\xi_2$  достигает экстремально низкого уровня, с вероятностью  $\lambda^L$  величина  $\xi_1$  также окажется в области экстремального минимума».

Так, коэффициенты хвостовой зависимости верхней и нижней особо важны при анализе редких, катастрофических ситуаций, например, финансовые кризисы, стихийные бедствия, системные сбои, когда возникает необходимость определять риск одновременных экстремальных отклонений.

Для эллиптических распределений коэффициенты хвостовой зависимости для нижней и верхней частей совпадают ( $\lambda^L = \lambda^U$ ). Это обусловлено их симметричностью: экстремальное поведение в области малых значений соответствует поведению в области больших значений.

Следует отметить, что нормальное распределение представляет собой единственное устойчивое эллиптическое распределение с легкими хвостами. Это говорит о том, что хвосты нормального распределения убывают

экспоненциально, в следствии чего значение коэффициента хвостовой зависимости равно нулю (62).

$$\lambda^L = \lambda^U = 2 \lim_{s \rightarrow \infty} (P(\xi_1 > s | \xi_2 = s)) = 2 \lim_{s \rightarrow \infty} \left( 1 - \Phi \left( s \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) \right) = 0 \quad (62)$$

В качестве эллиптического распределения с тяжелыми хвостами рассмотрим распределение Стьюдента, представленное формулой (63)

$$\lambda^L = \lambda^U = 2 \lim_{s \rightarrow \infty} (P(\xi_1 > s | \xi_2 = s)) = 2 - 2t_{\nu+1} \left( \sqrt{\nu+1} \times \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right). \quad (63)$$

Распределение Стьюдента демонстрирует ненулевую хвостовую зависимость ( $\lambda^L = \lambda^U > 0$ ) при малом числе степеней свободы  $\nu$ , что отражает его способность моделировать совместные экстремальные события. Однако с ростом  $\nu$  хвосты распределения становятся легче, и коэффициенты  $\lambda^L$  и  $\lambda^U$  стремятся к нулю, так как при  $\nu \rightarrow +\infty$  распределение Стьюдента сходится к стандартному нормальному.

Ранговые меры зависимости дают возможность оценивать наличие и силу монотонной зависимости между переменными, в то время как хвостовые коэффициенты зависимости дают возможность понять, сохраняется ли согласованность переменных при экстремальных значениях и является ли она одинаковой с обеих сторон. Эти показатели тесно связаны с подбором семейства копул: различных классы копул по-разному описывают поведение в хвостах распределения и могут учитывать асимметрию. В следующем разделе рассматриваются кривые копула-зависимости. Они позволяют наглядно сопоставлять парные структуры и лучше понимать характер взаимосвязей. После этого проводится эмпирический анализ: оцениваются хвостовые эффекты между динамикой населения и ключевыми социально-экономическими показателями регионов Российской Федерации.

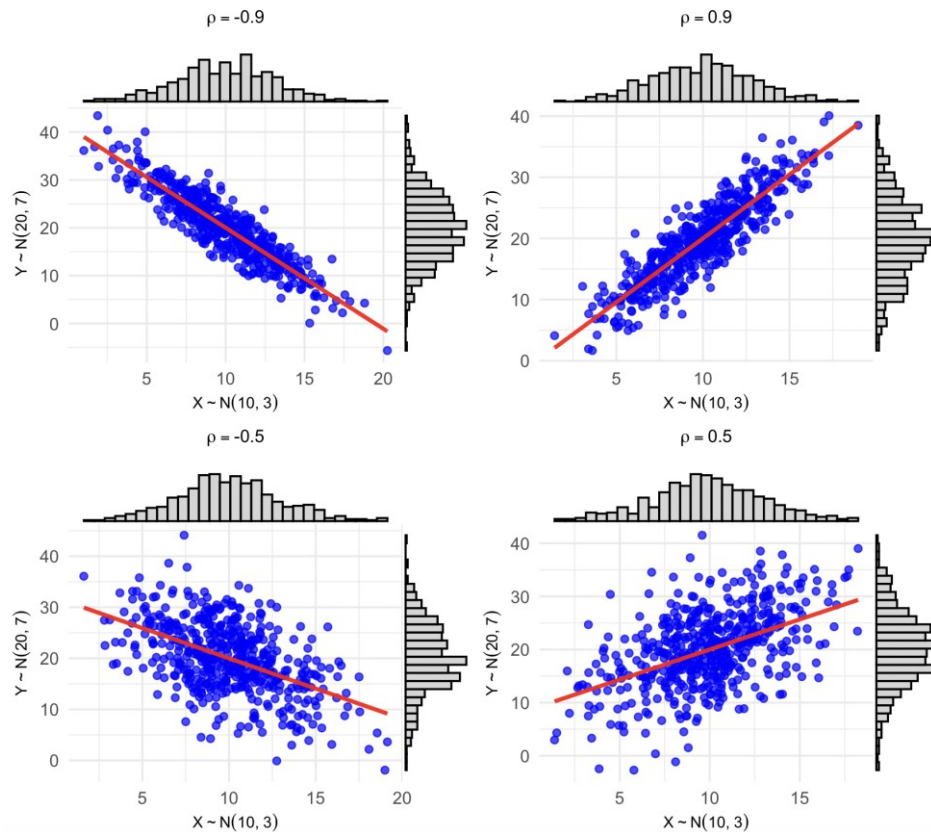
### 2.3 Кривые копула-зависимостей на примере парных моделей

После того как были рассмотрены ранговые меры зависимости и коэффициенты хвостовой зависимости, логично перейти от числовых показателей к визуальному представлению структуры парных связей. Одной и той же величине  $\tau$ -Кендалла или  $\rho$ -Спирмена могут соответствовать разные формы совместного поведения переменных, особенно в крайних областях распределения. Поэтому далее рассматриваются кривые копула-зависимостей, которые позволяют визуально сопоставить различные семейства копул и понять, какие типы регрессионных форм они порождают при одинаковой постановке задачи.

В контексте копульного подхода регрессионная кривая возникает из совместного распределения, заданного копулой и маргиналиями, а значит ее форма определяется не только «силой» зависимости, но и ее типом (симметрия/асимметрия, наличие хвостовой зависимости). Для сопоставимости далее используются одинаковые маргинальные распределения и меняется только семейство копулы и ее параметр.

Сгенерируем пары выборок из соответствующих копул при разных значениях параметров, а маргинальные распределения зададим нормальными:  $Y \sim N(20, 7)$ , а  $X \sim N(10, 7)$ . Такой выбор маргиналий носит иллюстративный характер и нужен для того, чтобы выделить вклад именно копулы в форму условного ожидания. В качестве базовой линии сравнения дополнительно приводится оценка парной линейной регрессии.

Начнем с нормальной копулы для различных значений корреляции  $\rho = -0,9; -0,5; 0,5; 0,9$ , представленных на рисунке 7. Для каждого значения параметр  $\rho$  строится выборка и оцениваются две кривые: линейная регрессия и копульная регрессия  $E(Y | X = x)$ . Это позволяет увидеть, в каких случаях копульная постановка воспроизводит линейную зависимость, а в каких – дает иную форму условного ожидания за счет структуры совместного распределения.



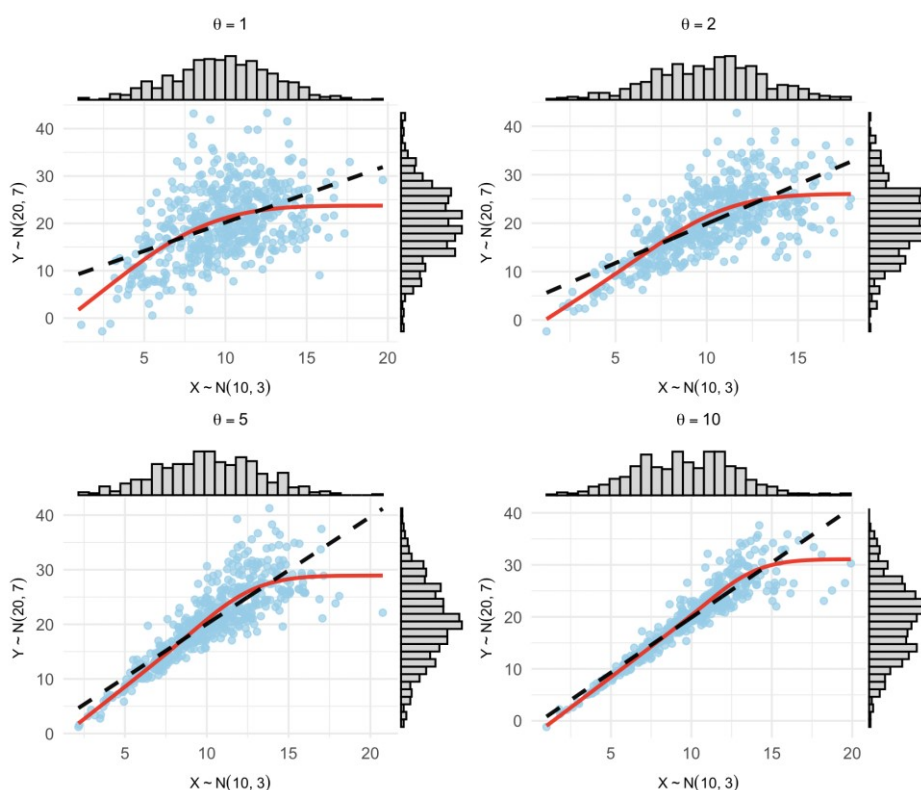
Источник: составлено автором.

Рисунок 7 – Парная копульная регрессия на примере нормальной копулы

Регрессия на основе нормальной копулы идентична оценке классической парной линейной регрессии, что является логичным. Нормальная копула предполагает наличие линейной зависимости между случайными величинами, отражаемой коэффициентом корреляции. В линейной регрессии основное внимание сосредоточено на условном среднем, а сама зависимость обычно задается в линейной форме. Копульный подход шире по своим возможностям, поскольку позволяет работать не только с условным математическим ожиданием, но и с совместным распределением переменных. За счет этого можно анализировать как изолированное влияние отдельных факторов, так и их совместное действие.

В случае копулы Стьюдента форма регрессионной зависимости остается близкой к линейной, поскольку она относится к тому же классу эллиптических распределений, что и нормальная копула. Отличие проявляется прежде всего в более корректном учете тяжелых хвостов.

Эллиптические копулы удобны как базовый ориентир, однако они в основном описывают симметричную зависимость. Для иллюстрации асимметричных режимов и различий между верхним и нижним хвостами далее рассматриваются архимедовы копулы, для которых хвостовое поведение и форма зависимости существенно меняются при варьировании параметра  $\theta$ . Парная копульная регрессия на примере копулы Клейтона с параметрами  $\theta = 1; 2; 5; 10$  представлена на рисунке 8.

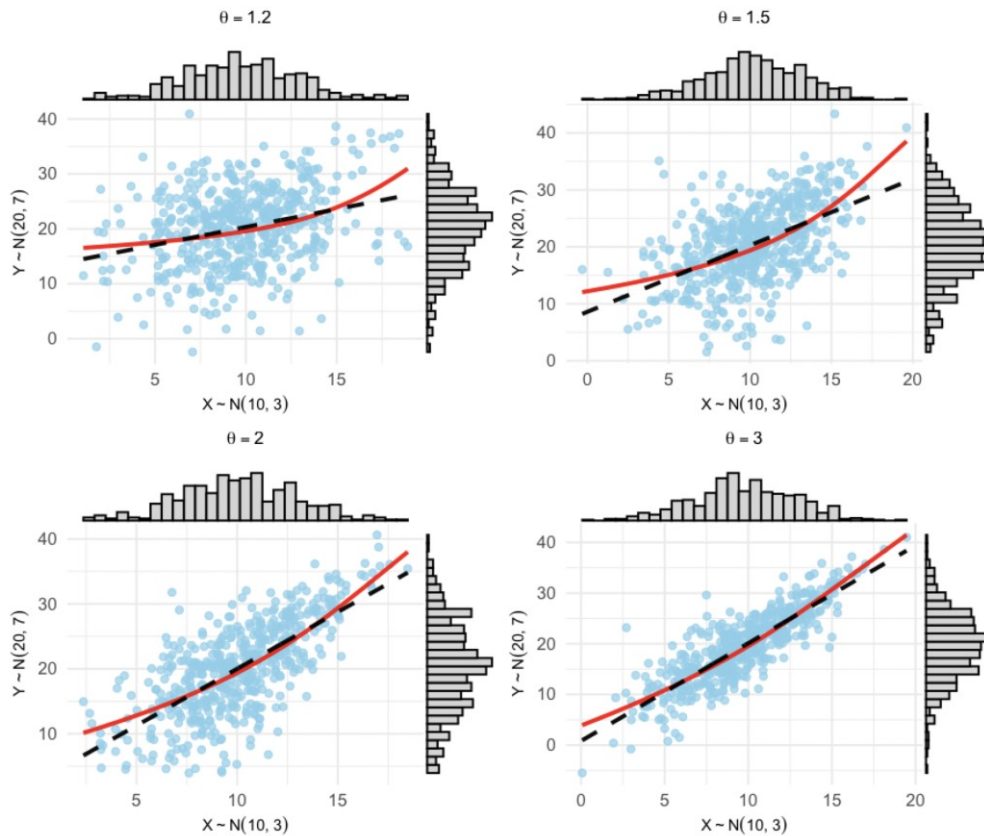


Источник: составлено автором.

Рисунок 8 – Парная копульная регрессия на примере копулы Клейтона

Анализ рисунка 8 показывает, что копула Клейтона хорошо подходит для описания эффекта затухания после определенного значения регрессора, до начала затухания эффект линейный. Копула Клейтона может быть использована для моделирования цензурированных данных. В отличие от моделей Тобита и Хэкмана, в которых отсечение данных происходит в регрессионном уравнении, копула Клейтона позволяет учесть «потолок» распределения при задаче маргинальных функций распределения.

Копула Гумбеля – противоположная по хвостовой зависимости, рассмотренной ранее копуле Клейтона. Рассмотрим ее для параметров  $\theta = 1.2; 1.5; 2; 3$  на рисунке 9.



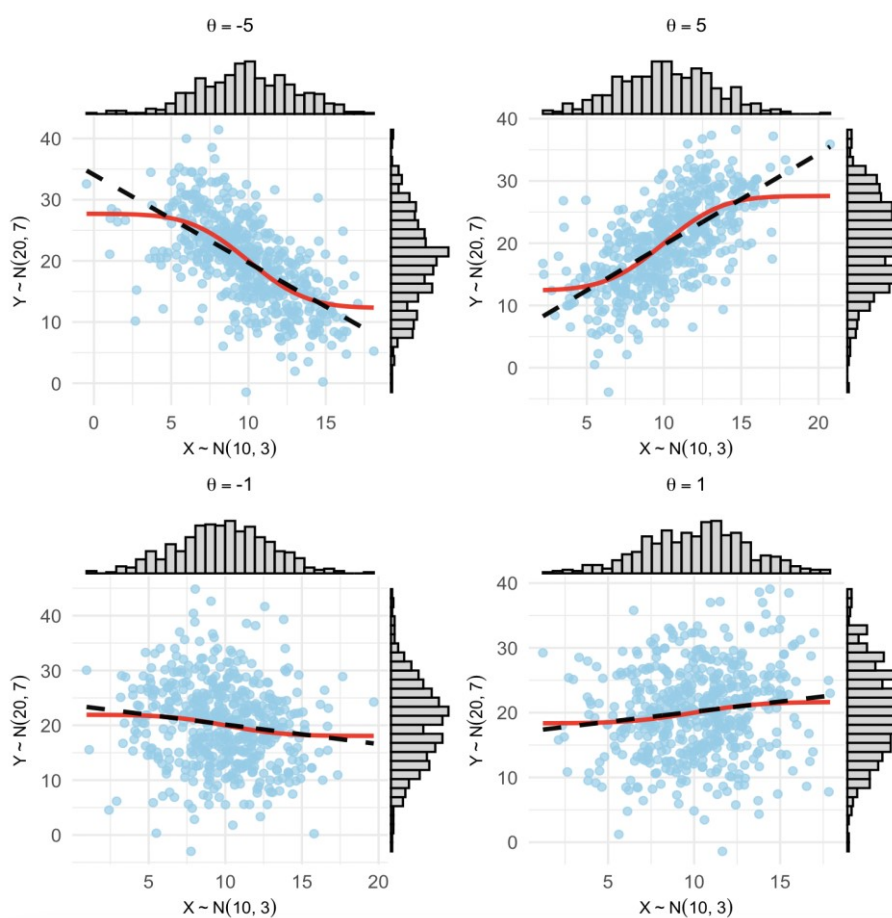
Источник: составлено автором.

Рисунок 9 – Парная копульная регрессия на примере копулы Гумбеля

Копула Гумбеля объясняет зависимость, схожую с показательной при высоких значениях, а при низких – схоже с усеченной снизу выборкой и также может быть использована вместо модели Тобита и Хэкмана в случае, обратном копуле Клейтона.

Наконец, копула Франка, описывающая и верхнюю, и нижнюю хвостовую зависимости. для параметров  $\theta = -5; -1; 1; 5$  копула Франка представлена на рисунке 10. Копула Франка описывает S-образную зависимость в данных и может иметь широкий спектр для применения. Ее можно использовать для моделирования цензурированных и усеченных выборок, для моделирования маркетинговых коммуникаций с учетом их эффекта накопления и насыщения. А моделируя зависимости на интервале от

0 до 1, по аналогии с логистической регрессией, можно решать задачу классификации с оценкой вероятности попадания в класс.



Источник: составлено автором.

Рисунок 10 – Парная копульная регрессия на примере копулы Франка

Рассмотренные примеры показывают, что выбор копула-функции заметно определяет вид регрессионной зависимости между переменными. Нормальная копула и копула Стюдента, как правило, не приводят к существенному отклонению от линейной структуры, тогда как архимедовы копулы дают возможность описывать более сложные, в том числе асимметричные, типы связи; кроме того, в рамках регрессии они позволяют корректно воспроизводить эффекты насыщения факторов. В результате копульный подход не сводится к формальному обобщению классической регрессии: он предоставляет набор средств для более гибкого описания эмпирических данных с учетом цензурирования, насыщения, нелинейности и иных характерных свойств. По этой причине копульные модели представляют

особый интерес в задачах, где важно не только оценить наличие и силу зависимости, но и адекватно учесть ее поведение в хвостах распределения либо в отдельных областях пространства признаков.

Синтетические примеры показывают, что при фиксированных маргинальных распределениях именно выбор семейства копулы определяет геометрию зависимости и то, где эта зависимость наиболее выражена – в центральной части распределения или в хвостах. В прикладной задаче пространственного демографического анализа это означает, что регионы-лидеры и регионы-аутсайдеры могут демонстрировать разные режимы совместного поведения показателей, и эти различия не сводятся к одной усредненной корреляции.

#### **2.4 Анализ хвостовой структуры зависимостей демографических и социально-экономических показателей**

В предыдущем параграфе на синтетических примерах было показано, что при фиксированных маргинальных распределениях различия между семействами копул проявляются прежде всего в геометрии парной зависимости и в том, где именно она концентрируется: в центральной области или в хвостах распределений. Для региональных данных такой акцент принципиален, поскольку содержательно значимыми являются именно крайние режимы: регионы с наиболее выраженной демографической убылью и регионы с максимально быстрым ростом населения. Поэтому далее проводится количественная оценка хвостовой зависимости между темпами прироста населения и ключевыми социально-экономическими факторами. Это позволяет перейти от общей констатации нелинейности к измеримым характеристикам совместного поведения переменных в экстремальных состояниях.

На основе формул (60) и (61) проводится оценка верхней и нижней хвостовой зависимости между эндогенной переменной (темпами прироста

населения) и экзогенными факторами. Коэффициенты хвостовой зависимости  $\lambda^U(q)$  и  $\lambda^L(q)$  рассчитываются для квантилей  $q = 0,90$  и  $q = 0,95$  и трактуются как условные вероятности того, что переменные одновременно окажутся соответственно в верхних или нижних экстремальных областях распределения. Иначе говоря,  $\lambda^U(q)$  отражает вероятность попадания темпов прироста населения в верхний хвост при условии, что соответствующий фактор также принимает экстремально высокие значения; величина  $\lambda^L(q)$  задает сопоставимую характеристику для нижнего хвоста и тем самым описывает степень согласованности в ситуациях одновременной реализации неблагоприятных режимов.

Оценка хвостовой зависимости дает возможность установить, какие факторы наиболее существенно связаны с демографической динамикой в экстремальных сценариях, то есть в регионах с наибольшими темпами прироста населения или, напротив, с наиболее выраженной убылью. Подобная информация имеет принципиальное значение при разработке адресных мер государственной политики для территорий, которые отстают по ключевым показателям или находятся в стадии развития. В таблице 4 ниже приведены оценки коэффициентов хвостовой зависимости между темпами прироста населения и социально-экономическими факторами.

При интерпретации полученных оценок следует исходить из того, что хвостовая зависимость характеризует не «средний» вклад фактора в демографическую динамику, а согласованность поведения переменных в крайних состояниях. Ненулевые значения  $\lambda$  при  $q = 0,90$  указывают на наличие связи в относительно широкой области повышенных или пониженных значений, тогда как при  $q = 0,95$  рассмотрению подвергается более строгий режим экстремумов. Уменьшение коэффициента при переходе от  $0,90$  к  $0,95$  означает, что зависимость проявляется в умеренно крайних состояниях, однако ослабевает на наиболее экстремальных уровнях; напротив, увеличение либо устойчивости  $\lambda$  при  $q = 0,95$  позволяет заключить, что взаимосвязь в основном сосредоточена на наиболее редких и критических наблюдениях.

Таблица 4 – Оценки коэффициентов хвостовой зависимости между темпами прироста населения и экзогенными факторами

Переменная	$\lambda_{0,9}^U$	$\lambda_{0,9}^L$	$\lambda_{0,95}^U$	$\lambda_{0,95}^L$
X1: ВРП на душу населения	0,0108	0,1720	0,0000	0,0000
X2: Среднедушевые доходы	0,0108	0,1075	0,0000	0,0213
X3: Безработица	0,0000	0,2688	0,0000	0,3617
X4: Инвестиции	0,0645	0,1398	0,0000	0,0000
X5: Браки	0,0108	0,1613	0,0000	0,1064
X6: Стоимость жилья	0,0538	0,0323	0,0213	0,0000
X7: ИПЦ (инфляция)	0,0860	0,1290	0,0426	0,0213
X8: Доля городского населения	0,0000	0,2043	0,0000	0,1489
X9: Расстояние до Москвы	0,0430	0,0000	0,0000	0,0000

Источник: составлено автором.

Отдельно отметим, что нулевые значения в таблице не следует трактовать как отсутствие зависимости вообще: они лишь указывают на то, что в рассматриваемой хвостовой области, при фиксированном уровне квантиля, статистически значимая асимптотическая согласованность не выявляется. В прикладном смысле это позволяет отделить факторы, которые «работают» преимущественно в экстремальных режимах демографической динамики, от факторов, связь которых с демографией проявляется в центральной части распределения и описывается другими мерами зависимости ( $\tau$ ,  $\rho$ ) и копула-кривыми.

Сделаем выводы из исследования парной зависимости для трех групп факторов.

**Группа 1: рынок труда и структурные факторы спроса (безработица, ВРП, доходы).** Анализ таблицы 4 выявляет комплементарный эффект показателей рынка труда и экономической производительности в

объяснении демографической динамики. Безработица демонстрирует абсолютно доминирующую верхнюю хвостовую зависимость ( $\lambda_{0,95}^U = 0,3617$ ), указывая на то, что состояние рынка труда – единственный фактор, чья связь с демографическим ростом усиливается в экстремальных сценариях. Напротив, показатели экономической производительности (ВРП на душу, среднедушевые доходы) демонстрируют размывание хвостовой зависимости при переходе от квантиля 0,90 к 0,95. Это указывает на важное различие: если безработица – это стрессовая переменная (ее влияние максимально в условиях кризиса), то доходы и ВРП – это благоприятные сигналы, теряющие релевантность при экстремальном усилении (парадокс высокого ВРП в анклавной экономике добывающих регионов). В целом совокупность факторов первой группы указывает на то, что в предельных ситуациях демографическая динамика в большей степени обусловлена степенью напряженности на рынке труда, нежели абсолютным уровнем экономической производительности.

**Группа 2: социально-демографические индикаторы (браки, ИПЦ, инвестиции).** Вторая группа переменных характеризуется более умеренными и асимметричными структурами зависимости в хвостах распределений, а также демонстрирует определенную устойчивость при анализе экстремальных ситуаций. Такие показатели, как брачность ( $\lambda_{0,90}^U = 0,1613$ ;  $\lambda_{0,95}^U = 0,1064$ ) и инвестиции ( $\lambda_{0,90}^U = 0,1398$ ) оказывают относительно стабильное влияние на исследуемый процесс, причем их значимость не исчезает при движении к более экстремальным значениям квантилей. В то же время инфляция, измеряемая индексом потребительских цен, демонстрирует двухстороннюю зависимость: величины коэффициентов нижней и верхней хвостовой зависимости оказываются близкими друг к другу. Можно рассматривать ее как показатель макроэкономической нестабильности: как правило, рост цен и периоды дефляции сопровождаются снижением миграционной активности.

Именно поэтому вторая группа переменных в основном отражает среднесрочные, опосредованные влияния на демографическую динамику.

**Группа 3: структурные и географические факторы (урбанизация, стоимость жилья, расстояние до Москвы).** Третья группа фиксирует структурные и долгосрочные характеристики пространственной дифференциации. Доля городского населения ( $\lambda_{0.90}^U = 0,2043$ ;  $\lambda_{0.95}^U = 0,1489$ ) демонстрирует устойчивую верхнехвостовую зависимость, близкую к безработице на квантиле 0,90, и сохраняет ее при переходе к более редким экстремумам. Урбанизацию в этой связи целесообразно трактовать как структурный фактор, ассоциированный с улучшениями на рынке труда и в доходах, что согласуется с эффектами городской агломерации. Стоимость жилья связана с динамикой неоднозначно: выраженная нижнехвостовая зависимость при слабой верхнехвостовой поддерживает гипотезу «проклятия высокой стоимости жилья». Удешевление жилья чаще сопряжено с демографическим спадом, тогда как рост цен не ведет к максимальному приросту населения из-за ограничения спроса. Расстояние до Москвы характеризуется асимметрией и, по-видимому, не является определяющим для экстремальных траекторий роста; высокие темпы в крупных регионах в большей степени объясняются внутренней динамикой, чем близостью к столице.

В настоящем параграфе сформирована эмпирическая база, необходимая для последующего копула-регрессионного анализа пространственной демографической динамики. Оценка зависимости в верхних и нижних хвостах распределения показала, что характер связей в экстремальных режимах неоднороден и асимметричен: для части факторов зависимость становится более выраженной преимущественно в нижнем хвосте, что соответствует неблагоприятным сценариям демографической динамики, тогда как для других переменных она обнаруживается главным образом в верхнем хвосте либо, напротив, ослабевает по мере перехода к более строгим квантилям. Полученные результаты хвостового анализа тем самым задают

содержательные ориентиры для дальнейшего многомерного моделирования, в рамках которого важно учитывать различие режимов зависимости для регионов-лидеров и регионов-аутсайдеров, а также возможную вариативность парных связей при контроле других факторов.

Результаты исследования имеют большое значение для разработки государственной политики в рамках Стратегии-2030. Тот факт, что в экстремальных сценариях роста доминирует проблема безработицы, говорит о том, что для регионов важнейшей задачей остается контроль уровня безработицы. Потенциальный недостаток рабочих рук может стимулировать приток рабочей силы, а урбанизация и инвестиции в инфраструктуру будут выполнять вспомогательную функцию. В этом разделе заложена база для перехода к многомерной копула-регрессии в следующей главе. Учет выявленных нелинейных и условных связей между факторами позволит углубить анализ и получить более точные результаты.

## **Выводы по главе 2**

В рамках второй главы было выполнено исследование, цель которого – выявить взаимосвязи между демографическими трансформациями и социально-экономическими характеристиками российских регионов. Изучение парных зависимостей дало возможность сфокусироваться на детальном анализе структуры связей, опираясь на данные по регионам. Полученные выводы лягут в основу создания многомерных моделей.

Первичный анализ данных показал, что распределение множества ключевых показателей существенно отличается от нормального: заметны асимметричные тенденции и «тяжелые» хвосты, особенно когда речь идет об индикаторах, связанных с экономической активностью и инвестициями. Что касается динамики численности населения, то в большинстве случаев изменения незначительны, однако встречаются примеры с существенно более высокими или низкими темпами прироста, что подчеркивает необходимость

уделять внимание крайним значениям. Из-за этих особенностей стандартные методы, которые опираются на усреднение, не дают должного эффекта – необходимо использовать более изощренные методики, учитывающие неоднородность распределений.

Также в исследовании обнаружена сильная пространственная автокорреляция между рассчитанным темпом прироста населения и множеством социально-экономических показателей. Это говорит о том, что данные сгруппированы в отдельные региональные кластеры. Поэтому важно рассматривать не только усредненные по региону показатели, а также сопоставлять группы территорий, которые имеют схожие характеристики, но развиваются по разным траекториям.

Для анализа связей между переменными были применены методы копульного анализа. В частности, использовались ранговые коэффициенты корреляции, а также показатели хвостовой зависимости. Ранговые коэффициенты позволяют выявлять монотонные зависимости без ограничений на их линейность и меньшей чувствительностью к изменениям маргинальных распределений. А меры хвостовой зависимости позволяют сфокусироваться на экстремальных значениях и узнать, насколько согласованно ведут себя переменные в этих областях. Это важно в исследовании регионов-лидеров и аутсайдеров.

Анализ парных зависимостей на синтетических данных продемонстрировал, что форма условного ожидания варьируется в зависимости от используемого семейства копул и их параметров при фиксированных маргинальных распределениях. Эллиптические копулы, как правило, характеризуют симметричные зависимости, в то время как архимедовы могут выявлять асимметрию и усиливать зависимость в хвостах распределений. Таким образом, один и тот же уровень зависимости может отражать разные совместные сценарии, в том числе и в экстремальных областях, в зависимости от выбранного семейства копул.

Изучение хвостовых зависимостей между демографическими и социально-экономическими переменными показало, что в экстремальных условиях они ведут себя по-разному: для некоторых показателей сильная связь прослеживается лишь в одном хвосте и меняется при переходе к более жестким квантилям. Следовательно, невозможно охарактеризовать взаимосвязь между демографическими изменениями и факторами регионального развития одним средним показателем, необходим метод, позволяющий учесть разные режимы взаимодействия и асимметрию связей.

Итак, вторая глава заложила основу для многомерного моделирования: были выявлены ключевые особенности распределений и пространственной структуры данных, а также разработан инструментарий для анализа парных зависимостей (ранговые меры, копула-кривые, хвостовые коэффициенты). Это позволит в дальнейшем корректно специфицировать многомерные модели и интерпретировать результаты с точки зрения пространственной демографической динамики российских регионов.

### Глава 3

## **Непараметрическое многомерное моделирование и сценарный анализ демографического роста регионов Российской Федерации на основе Vine-копул**

В третьей главе осуществляется переход от парной диагностики зависимостей к многомерному анализу факторов демографической динамики регионов Российской Федерации. Если в предыдущих разделах были показаны выраженная неоднородность распределений социально-экономических показателей, наличие пространственной кластеризации и значимость хвостовых режимов зависимости, то в многомерной постановке ключевым становится вопрос о совместном учете нескольких факторов одновременно и о том, как меняется связь с демографическим ростом при контроле остальных переменных. Такая постановка особенно важна для региональных данных, где одни и те же территории могут одновременно отличаться по уровню доходов, инвестиционной активности, состоянию рынка труда и параметрам городской среды, а наблюдаемые эффекты в хвостах распределений часто формируются именно сочетанием факторов, а не одним показателем.

Актуальность многомерного подхода обусловлена задачами государственной пространственной политики, сформулированными в Стратегии пространственного развития Российской Федерации до 2030 года с прогнозом до 2036 года, где демографическая динамика рассматривается как системное ограничение и одновременно как целевой ориентир развития территорий. Для обоснования дифференцированных мер требуется не только фиксировать наличие связи между отдельными социально-экономическими показателями и изменением численности населения, но и количественно оценивать структуру влияний в условиях многомерности, возможной асимметрии и концентрации эффектов в крайних режимах. В связи с этим в главе рассматривается непараметрический многомерный инструментарий на основе Vine-копул, позволяющий гибко задавать структуру совместной

зависимости через композицию парных связей и тем самым перейти от усредняющих оценок к более детальному описанию региональных режимов демографического роста и убыли.

В рамках главы последовательно вводится и обосновывается многомерный аппарат Vine-копул для моделирования зависимостей региональных показателей; формулируются предпосылки и ограничения Vine-копула регрессии как непараметрического подхода к оценке условного распределения; строится и оценивается множественная Vine-копула регрессионная модель для количественной идентификации факторов демографического развития; на заключительном этапе анализируется структура влияния социально-экономических переменных с учетом региональной неоднородности и выполняется сценарное моделирование демографического роста регионов Российской Федерации.

### **3.1 Vine-копулы как инструмент многомерного моделирования зависимостей**

В современных исследованиях, связанных с многомерным статистическим анализом, копула-функции зарекомендовали себя как мощный инструмент для моделирования зависимостей между случайными величинами. Однако при росте размерности применение стандартных параметрических семейств многомерных копул сталкивается с ограничениями гибкости: одно и то же семейство задает единую форму зависимости для всех компонент и не позволяет отдельно настраивать парные и условные зависимости, что нередко приводит к недостаточной согласованности модели с эмпирическими данными.

Одним из ключевых недостатков традиционных многомерных копул является их неспособность адекватно отражать сложные, **гетерогенные структуры зависимости**, которые могут существенно различаться для разных пар переменных. В задачах финансовой аналитики зависимость между

доходностями различных финансовых активов варьирует в широких пределах: для одних пар активов зависимость ярко выражена и носит положительный характер, для других она близка к нулю, а порой даже отрицательна. Различия такого рода можно наблюдать и во многих других областях, например, в метеорологии, страховании, медицине, где ключевым моментом является точное описание многомерных распределений.

Чтобы справиться с упомянутыми ограничениями, были разработаны **Vine-копулы**. Они позволяют описать многомерную зависимость, разбив ее на композицию парных копул. Этот метод отличается большей гибкостью, так как:

- 1) дает возможность создавать неоднородную структуру взаимосвязей между парами переменных – можно подбирать семейства копул, исходя из их особенностей;
- 2) позволяет целенаправленно определять условные зависимости, что особенно важно при работе с объемными наборами данных;
- 3) упрощает процесс оценки параметров, поскольку разбивает сложную многомерную задачу на ряд более простых двумерных подзадач.

Для построения Vine-копул ключевую роль играет декомпозиция условных распределений через парные копулы. Рассмотрим многомерную плотность случайного вектора  $(X_1, \dots, X_d)$ , которую можно представить в виде произведения маргинальных плотностей и копулы, задающей зависимость между компонентами. Однако для гибкого моделирования сложных структур удобно использовать условные плотности.

Пусть  $f(x_1, \dots, x_d)$  – это совместная плотность вектора  $(X_1, \dots, X_d)$ , запишем ее через условные плотности формулой (64)

$$\begin{aligned} f(x_1, \dots, x_d) &= f_{d|1, \dots, d-1}(x_d|x_1, \dots, x_{d-1}) \times f(x_1, \dots, x_{d-1}) = \dots = \\ &= f_1(x_1) \times f_{2|1}(x_2|x_1) \times f_{3|1,2}(x_3|x_1, x_2) \times \dots \times f_{d|1, \dots, d-1}(x_d|x_1, \dots, x_{d-1}), \end{aligned} \quad (64)$$

где  $f_{i|1,\dots,i-1}(x_i|x_1, \dots, x_{i-1})$  – это условная плотность  $X_i$  при заданных  $X_1, \dots, X_{i-1}$ .

Каждую условную плотность в свою очередь можно выразить через плотности парных копул. Например, для  $f_{2|1}(x_2|x_1)$  будет определяться формулой (65)

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) \times f_2(x_2). \quad (65)$$

А для тройки случайных величин  $X_1, X_2, X_3$  формулой (66)

$$f_{3|1,2}(x_3|x_1, x_2) = c_{31|2}(F_{3|2}(x_3|x_2), F_{1|2}(x_1|x_2)) \times f_{3|2}(x_3|x_2), \quad (66)$$

кроме того, эту же условную плотность можно выразить по-другому – формулой (67)

$$f_{3|1,2}(x_3|x_1, x_2) = c_{32|1}(F_{3|1}(x_3|x_1), F_{2|1}(x_2|x_1)) \times f_{3|1}(x_3|x_1). \quad (67)$$

Аналогично, более высокоразмерные условные плотности могут быть рекурсивно выражены через вложенные парные копулы, что приводит к идее Vine-копул [32].

#### **Определение 11 (Vine-копулы) [35]**

*Vine-копула* – это структура, декомпозирующая многомерную зависимость на иерархическую последовательность парных копул (би-копул), организованных в виде упорядоченного графа. Формально, Vine-копула представляет собой:

- 1) Иерархию деревьев  $T_1, \dots, T_{d-1}$ , где каждое дерево  $T_i$  описывает зависимости между переменными или их условными представлениями.
- 2) Набор парных копул, соответствующих ребрам деревьев, где каждая копула моделирует зависимость между парой переменных (или условными переменными).

Выделяют три основных класса Vine-копупл на основе их структуры: **Regular Vine (R-Vine)**, **Canonical Vine (C-Vine)** и **Drawable Vine (D-Vine)**. Эта классификация основана на способе организации иерархии деревьев и выбора условных зависимостей между переменными [25]. Критерием разделения служит:

- степень централизации зависимостей (наличие «главных» переменных в C-Vine);
- гибкость структуры (R-Vine допускает произвольные связи, D-Vine – последовательные);
- условная независимость (разные стратегии декомпозиции многомерной плотности).

**C-Vine** предполагает наличие **центральной переменной** на каждом уровне дерева, через которую выражаются все условные зависимости. В первом дереве одна переменная (например,  $X_1$ ) соединена со всеми остальными, во втором – условные зависимости строятся относительно  $X_1$  и так далее. Плотность для C-Vine записывается формулой (68)

$$f(x_1, \dots, x_d) = \prod_{i=1}^d f_i(x_i) \times \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,\dots,j-1}, \quad (68)$$

где  $c_{j,j+i|1,\dots,j-1}$  – парная копула между  $X_j$  и  $X_{j+i}$  при условии  $X_1, \dots, X_{j-1}$  [32].

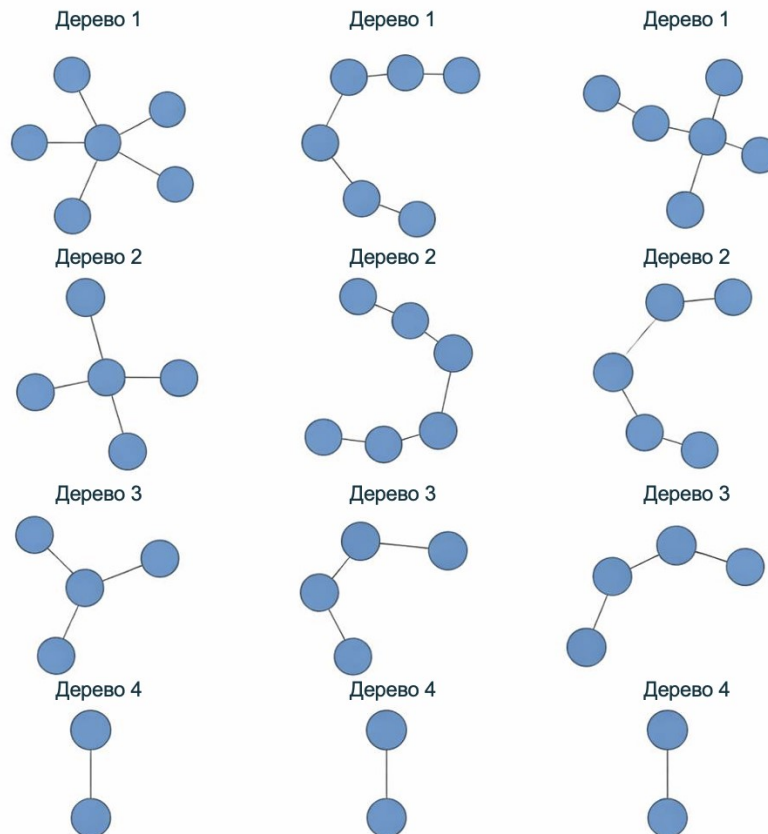
**D-Vine** имеет **линейную структуру**, где зависимости строятся последовательно, без выделенных центральных переменных. В первом дереве переменные соединяются попарно ( $X_1 - X_2, X_2 - X_3$  и т.д.), во втором – моделируются условные зависимости при фиксированных промежуточных переменных (например,  $X_1$  и  $X_3$  при заданном  $X_2$ ). Плотность для D-Vine имеет место формула (69)

$$f(x_1, \dots, x_d) = \prod_{i=1}^d f_i(x_i) \times \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\dots,i+j-1}, \quad (69)$$

где  $c_{i,i+j|i+1,\dots,i+j-1}$  – парная копула между  $X_i$  и  $X_{i+j}$  при условии  $X_{i+1}, \dots, X_{i+j-1}$  [37].

**R-Vine** – наиболее общий случай, объединяющий C-Vine и D-Vine как частные варианты. Структура деревьев не накладывает жестких ограничений на связи, что позволяет адаптировать модель к данным. Критерием выбора ребер на каждом уровне часто служит максимизация информационного критерия или правдоподобия. Математически R-Vine задается набором деревьев  $\{T_1, \dots, T_{d-1}\}$ , где каждое  $T_i$  соответствует уровню условной зависимости [42].

Рассмотрим примеры деревьев для всех трех классов Vine-копул. На рисунке 11 проиллюстрированы примеры C-Vine, D-Vine и R-Vine деревьев, где четко отражается изложенное выше определение.



слева – C-Vine, по центру – D-Vine, справа – R-Vine

Источник: составлено автором.

Рисунок 11 – Пример деревьев Vine-копул различных классов

Так как Vine-копула – композиция из классических парных копул, методы оценки параметров сохраняются. Однако из-за иерархической структуры Vine процесс разбивается на последовательные шаги, соответствующие уровням деревьев.

**Выбор структуры Vine.** На этом этапе в зависимости от поставленной задачи нужно выбрать класс Vine-копул. Например, если имеем дело с финансовыми данными с некоторым «ведущим» активом, используем C-Vine. А если моделируем временные ряды, где зависимость «затухает» с увеличением лага, используем D-Vine.

**Оценка парных копул.** Для каждого ребра в выбранной структуре выбирается тип копулы (нормальная, Стюдента, Клейтона и др.) на основе критерия AIC (или другое правило). Далее, параметры выбранной копулы оцениваются методом максимального правдоподобия по правилу (40). Для условных копул (на уровне деревьев  $i \geq 2$ ) используются производные ***h*-функции** в соответствии с формулой (70)

$$h(u|v; \theta) = \frac{\partial C(u, v; \theta)}{\partial v}, \quad (70)$$

с помощью *h*-функции можно рассчитать условные величины, например  $u_{i|k} = h(u_i|u_k; \theta_{ik})$ .

**Оптимизация глобального правдоподобия.** После первоначальной оценки параметры Vine-копулы уточняются, максимизируя полное правдоподобие по формуле (71)

$$\mathcal{L}(\Theta) = \sum_{t=1}^N \sum_{k=1}^{d-1} \sum_{e \in E_k} \ln c_e(u_{i|D_e}(t), u_{j|D_e}(t); \theta_e), \quad (71)$$

где  $E_k$  – ребра дерева  $T_k$ ;

$D_e$  – множество условий для ребра  $e$ .

Vine-копулы являются мощным инструментом для моделирования сложных многомерных зависимостей и преодоления ограничений классических многомерных копул. Главная идея Vine-копул состоит в разложении общей структуры копула на последовательность парных копул, что обеспечивает высокую гибкость при описании различных видов зависимостей между переменными, в том числе и условных. В контексте исследования это означает, что можно создать многомерную модель, описывающую зависимость демографического роста от социально-экономических факторов: она будет состоять из системы парных и условных копул, учитывающих неоднородность данных и эффекты, связанные с экстремальными значениями.

### **3.2 Vine-копула регрессия: предпосылки и ограничения**

Копульные регрессионные модели в последние годы все чаще рассматриваются как удобный инструмент для анализа зависимых данных и выборок со сложной внутренней структурой. В литературе представлен ряд работ, в которых регрессионная конструкция опирается на гауссовскую копулу: зависимость между наблюдениями задается многомерным нормальным распределением, тогда как маргинальные распределения допускаются в достаточно общем виде. В одном из исследований предложен вариант гауссовской копула-регрессии, ориентированный на анализ зависимых откликов, в том числе временных рядов, продольных и пространственных данных, а также наблюдений с дискретными откликами [52]. Отдельно обсуждаются численные подходы к приближению функции правдоподобия; в частности, рассматриваются модификации алгоритма Geweke-Hajivassiliou-Keane. Указанные методические решения послужили основой для реализации в пакете `gcm` для R, который затем был подробно представлен в последующей публикации тех же авторов [53].

Авторы также анализируют использование копул для моделирования полуконкурирующих рисков, то есть ситуаций, в которых наступление одного события (например, отказ трансплантата) может оказаться цензурированным другим событием (смертью пациента) [63]. В работе рассматриваются нормальная копула, а также копулы Клейтона, Франка и Гумбеля; при этом зависимость между событиями специфицируется с учетом бинарных ковариат, которые воздействуют как на маргинальные распределения, так и на параметр копулы. Указанная постановка, как правило, дает возможность более корректно оценивать риски при наличии неоднородности наблюдений и цензурирования.

Особый интерес в контексте задач с дискретными исходами представляет непараметрическая оценка копул. Yang, Frees и Zhang предлагают подход, ориентированный на восстановление латентной структуры зависимости без априорных допущений о конкретном функциональном виде копулы. Авторы акцентируют проблему неидентифицируемости копулы при дискретных распределениях и, в качестве возможного решения, рассматривают использование расширенных областей поддержки в сочетании с непараметрическими методами сглаживания. Полученные оценки применимы не только для количественного описания зависимости, но также могут использоваться как диагностический инструмент при выборе параметрической копулы в прикладных постановках [65].

Методология копула-регрессии в последние годы развивается, в частности, в направлении квантильного анализа. Kraus и Czado предложили полунепараметрическую модель, основанную на D-vine-копуле, предназначенную для прогнозирования условных квантилей. В отличие от классической линейной квантильной регрессии, D-vine-модель формируется как последовательность парных копул, благодаря чему становится возможным описывать нелинейные и асимметричные зависимости между переменными. В рамках этой конструкции реализуется автоматизированный отбор ковариат; кроме того, в значительной степени снимается проблема пересечения

квантилей, типичная для стандартных подходов. Авторы показывают, что предложенный метод демонстрирует высокую точность при большой размерности и на реальных финансовых данных, включая задачу оценки показателя Value-at-Risk [50].

В работе Parsa и Klugman обсуждается прикладной аспект копула-регрессии, в рамках которого копулы предлагаются в качестве альтернативы обобщенным линейным моделям (GLM) для актуарных задач. Авторами отмечается, что применение копул позволяет адекватно описывать зависимости между различными маргинальными распределениями, в частности, и такими, которые не относятся к экспоненциальным семействам. Может быть смоделирована более сложная структура зависимостей, нежели при использовании стандартных обобщенных линейных моделей (GLM) [57]. Поэтому такой подход является эффективным при исследовании страховых убытков, так как в страховании встречается асимметрия распределений, тяжелые хвосты и неоднородность факторов риска.

Современные работы по копула-регрессиям демонстрируют большой потенциал и разнообразие применений: от задач актуарных вычислений и анализа выживаемости до финансового управления рисками. При этом развитие численных методов расширяет возможности описания сложных структур зависимостей. Это особенно важно в случаях, когда данные являются дискретными, цензурированными или содержат нелинейные зависимости. При этом копулы сохраняют свое главное преимущество – возможность отдельно моделировать маргинальные распределения и структуру зависимости, что помогает обойти ограничения классических параметрических моделей.

Далее рассматривается непосредственно подход к непараметрической регрессии, основанной на Vine-копуле: описывается методика, изучаются ограничения, а также указываются основные преимущества и недостатки.

Зададим набор переменных регрессии. Пусть  $(Y, X_1, X_2, \dots, X_k)$  – случайный вектор, для которого предстоит обучить  $k + 1$ -мерную Vine-

копулу. Причем  $Y$  – это эндогенная переменная, которую хотим описать через  $k$  случайных величин  $X_1, \dots, X_k$ . Пусть для каждой случайной величины из вектора определено  $n$  значений, а также определена матрица наблюдений для случайного вектора по формуле (72)

$$\begin{pmatrix} y_1 & x_{11} & \dots & x_{k1} \\ y_2 & x_{12} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ y_n & x_{1n} & \dots & x_{kn} \end{pmatrix}. \quad (72)$$

Процедура построения копульной модели регрессии состоит из нескольких шагов:

**Шаг 1.** По сформированному массиву данных для каждого компонента случайного вектора необходимо оценить плотность и функцию распределения. Оценка функций распределения нужна для того, чтобы обучить копула-функцию, определенную на единичном кубе. Наиболее тривиальным способом оценки функции распределения является построение ECDF (эмпирическая функция распределения) оценки представлена формулой (73)

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i \leq x}, \quad (73)$$

где  $n$  – объем выборки;

$\mathbb{I}_{X_i \leq x}$  – индикаторная функция, равная 1, если  $X_i \leq x$ , и 0 в противном случае.

Оценки такого рода имеют ступенчатый характер, а график эмпирической функции распределения (ECDF) растет линейно на всем протяжении, что часто не отражает реальную ситуацию. В этом случае более подходящими будут ядерные оценки плотности и функции распределения. Следует иметь в виду, что тип ядерной функции  $K(u)$  и ширина окна  $h$  могут выступать в качестве гиперпараметров копула-регрессии. Поэтому их

необходимо тщательно подбирать, ориентируясь на минимизацию функционала ошибки регрессии.

**Шаг 2.** Ключевой этап построения корректной регрессии – обучение копула-функции. Как уже отмечалось в предыдущих главах, подобрать подходящую копула-функцию – задача непростая и весьма важная. Необходимо определить критерий для выбора, например, можно использовать критерий Акаике.

В классе архимедовых и эллиптических копул есть  $k$ -мерные копулы, но, как уже говорилось, они устанавливают одинаковую структуру и тип связи для всех компонентов случайного вектора. В дальнейшем рассматриваются Vine-копулы, поскольку они позволяют описывать систему попарных зависимостей между случайными величинами в многомерной постановке. Кроме того, нужно корректно выбрать типа Vine-копулы в зависимости от задачи. В задачах регрессии естественным исходным вариантом часто выступает C-Vine, так как в такой конструкции выделяется центральная переменная, связь которой с остальными представляет основной интерес. Для частных случаев, например, для авторегрессионных моделей, когда связь с каждым новым лагом ослабевает, имеет место D-Vine копула. В более общей постановке используется R-Vine структура, поскольку она дает более гибкое описание дерева зависимостей на разных уровнях.

**Шаг 3.** Вернемся к определению Vine-копулы и к цепной формуле плотности распределения. После построения иерархического дерева для Vine-копулы, на основе их формул можно оценить совместную плотность распределения. Так как все расчеты делаются численно, выделение аналитической формы совместной плотности является сложной задачей. Поэтому имеем дело с конкретными наблюдениями совместной плотности при заданных значениях маргинальных функций распределений компонент вектора  $\hat{f}_i(y_i, x_{1i}, \dots, x_{ki})$ .

Возвращаясь к регрессии, имеем дело с совместной плотностью распределения. Для решения задачи регрессии нужна не просто совместная

функция плотности, а условное распределения при фиксированных экзогенных переменных в соответствии с формулой (74)

$$f_{Y|X_1, \dots, X_k}(y | x_1, \dots, x_k) = \frac{f_{Y, X_1, \dots, X_k}(y, x_1, \dots, x_k)}{f_{X_1, \dots, X_k}(x_1, \dots, x_k)}. \quad (74)$$

Вспомним свойство плотности распределения – совокупная площадь под кривой для плотности равна 1, как представлено в формуле (75)

$$\begin{aligned} \int_{\mathbb{R}} f_{Y|X_1, \dots, X_k}(y | x_1, \dots, x_k) dy &= \int_{\mathbb{R}} \frac{f_{Y, X_1, \dots, X_k}(y, x_1, \dots, x_k)}{\underbrace{f_{X_1, \dots, X_k}(x_1, \dots, x_k)}_{\text{не зависит от } y}} dy = \\ &= \frac{1}{f_{X_1, \dots, X_k}(x_1, \dots, x_k)} \int_{\mathbb{R}} f_{Y, X_1, \dots, X_k}(y, x_1, \dots, x_k) dy = 1. \end{aligned} \quad (75)$$

Из формулы (75) вытекает, что при фиксированных значениях  $x_1, \dots, x_k$  знаменатель из условной плотности ничто иное как нормирующая константа.

Так, для того чтобы построить условное распределение  $Y$  при фиксированных  $X_i$  достаточно обладать знанием о совместном распределении, а значение совместной плотности для  $X_i$  можно считать, как нормировочную константу, которая сделает площадь под кривой для условной плотности равной 1.

В терминах точечных оценок вышесказанное примет вид:

– Строим с некоторым шагом  $\frac{1}{m-1}$  для диапазона  $[0; 1]$  функции распределения  $F(y)$  точечные оценки совместной плотности при фиксированных  $x_1, \dots, x_k$ , что соответствует формуле (76)

$$\left( f(y_{(1)}, x_1 = a_1, \dots, x_k = a_k), \dots, f(y_{(m)}, x_1 = a_1, \dots, x_k = a_k) \right), \quad (76)$$

где  $m$  – это число точечных оценок с шагом  $\frac{1}{m-1}$ .

– Используя метод трапеций или прямоугольников (или другие численные методы), считаем интеграл под кривой  $S_f$ . Далее каждый элемент множества

(76) умножаем на нормировочную константу  $\frac{1}{S_f}$ , тем самым переходим к значения условной плотности.

Теперь получив информацию об условном распределении, можно построить оценку регрессии, как условное математическое ожидание – формула (77)

$$E(y|x_1, \dots, x_k) = \int_{\mathbb{R}} yf(y|x_1, \dots, x_n)dy. \quad (77)$$

Численно условное математическое ожидание так же можно посчитать с помощью метода трапеций.

Кроме математического ожидания можно в качестве оценки регрессии выбрать моду распределения, то есть самое вероятное значения по формуле (78)

$$Mode(y|x_1, \dots, x_k) = \underset{y}{\operatorname{argmax}}(f(y|x_1, \dots, x_n)). \quad (78)$$

Численно это точечная оценка  $y_{(i)}$ , при которой значение плотности  $f(y_{(i)}, x_1 = a_1, \dots, x_k = a_k)$  максимально из всех элементов множества.

Кроме того, на основании множества (76) можно легко получить любую квантильную оценку, отранжировав ряд и выделив соответствующую часть множества, обеспечивающий сумму ряда равную квантильному уровню. Это очень полезное свойство, которое может считать вероятность экстремального события при определенных вариантах экзогенных переменных.

Предложенный подход к регрессионному анализу, основанный на восстановлении условного распределения с помощью копула-функций и последующем интегрировании по нему для получения регрессионной характеристики (математического ожидания, моды или квантили), обладает рядом фундаментальных преимуществ по сравнению с классическими параметрическими моделями.

**Устранение проблемы мультиколлинеарности.** В классической линейной регрессии высокая корреляция между предикторами приводит к мультиколлинеарности, из-за которой матрица  $X^T X$  становится плохо обусловленной или даже необратимой, что делает оценки нестабильными. В нашем подходе эта проблема не возникает принципиально, так как регрессия строится не через инверсию, а через оценку совместной плотности. Следовательно, предикторы могут быть как коррелированными, так и содержать перекрывающую информацию – *копула-регрессия учитывает как индивидуальные эффекты факторов, так и их совместную зависимость* через структуру копулы.

**Отсутствие предпосылок нормальности, гомоскедастичности и линейности.** Одна из ключевых проблем классических регрессионных моделей, оцениваемых методом наименьших квадратов (МНК), заключается в необходимости выполнения следующих предпосылок:

- нормальность ошибок:  $\varepsilon \sim N(0, \sigma^2)$ ;
- гомоскедастичность:  $Var(\varepsilon_i) = \sigma^2$ ;
- линейная зависимость:  $E(Y|X) = \beta_0 + \sum \beta_i X_i$ .

В копула-регрессии *все эти условия избыточны*:

- структура зависимости задается не через форму уравнения, а через копулу  $C(u_Y, u_1, \dots, u_k)$ , которая инвариантна к монотонным преобразованиям переменных (свойство копул из теоремы Скляра);
- маргинальные распределения моделируются отдельно с использованием ядерных оценок, что позволяет корректно восстанавливать как симметричные, так и асимметричные, многомодальные и хвостатые распределения;
- гетероскедастичность автоматически моделируется через форму условной плотности: если дисперсия  $Y$  зависит от значений  $X$ , это отражается в изменении формы  $f_{Y|X}(y|x)$  – без необходимости делать отдельную спецификацию для дисперсионной функции.

**Гибкость и адаптивность модели.** Модель допускает использование любых регрессионных функционалов.

Математическое ожидание в соответствии с формулой (79)

$$E(Y|X = x) = \int_{\mathbb{R}} y f_{Y|X}(y|x) dy. \quad (79)$$

Мода в соответствии с формулой (80)

$$Mode(Y|X = x) = \operatorname{argmax}_y f_{Y|X}(y|x). \quad (80)$$

Квантиль  $\tau$  в соответствии с формулой (81)

$$Q_{\tau}(Y|X = x) = \inf \left\{ y: \int_{-\infty}^y f_{Y|X}(s|x) ds \geq \tau \right\}. \quad (81)$$

Кроме того, можно рассчитывать значения дисперсии точечных распределений для каждого объекта данных, тем самым считая индивидуальные доверительные интервалы для каждой прогнозной точки в соответствии с формулой (82)

$$\sigma_{Y|X}^2 = \int_{\mathbb{R}} (y - E(x))^2 f_{Y|X}(y|x) dy. \quad (82)$$

**Робастность к выбросам и тяжелым хвостам.** В отличие от классической линейной регрессии, в настоящем исследовании применяется подход, ориентированный на восстановление полной формы условного распределения отклика при заданных значениях факторов. За счет этого регрессия рассматривается не как единственная точечная характеристика, а как класс функционалов, определяемых соответствующим условным распределением.

В частности, если в качестве регрессионной характеристики рассматривать условные квантили, медиану либо моду, то оценки, как правило, оказываются менее чувствительными к асимметрии распределения и единичным выбросам. Следует отметить, что указанная устойчивость обусловлена выбором этого функционала и не связана с какими-то особенностями копула-модели в отдельности. В свою очередь, копульный подход дает возможность гибко и достаточно правдоподобно задавать вид условной плотности, на основе которой впоследствии рассчитываются различные регрессионные характеристики.

Даже при вычислении условного математического ожидания такой метод не требует предположений о нормальности ошибок или симметрии распределения, так как интегрирование идет по восстановленной условной плотности. Это позволяет учитывать тяжелые хвосты и асимметрию распределений, что делает результаты более устойчивыми к нарушениям стандартных параметрических предположений. Так, модель сохраняет свою устойчивость к выбросам, ассиметричным распределениям и аномальным наблюдениям, то есть можно говорить о ее робастности без использования классических робастных методов.

Несмотря на очевидные достоинства предлагаемого подхода, как и любая статистическая методика, он не лишен определенных ограничений. Наиболее существенные из них связаны не столько с концептуальной слабостью модели, сколько с техническими и прикладными аспектами ее реализации. Ниже рассмотрим ключевые из них.

**Требование непрерывности распределений.** Одно из базовых условий для корректного применения копула-функций – непрерывность маргинальных распределений. Теорема Склера гарантирует единственность копулы только в случае, если все маргинальные функции распределения непрерывны. Это означает, что при наличии дискретных переменных (например, категориальных признаков, количеств, рейтингов и тому подобное) возникает проблема неидентифицируемости: одна и та же совокупность маргинальных и

копулы может соответствовать разным совместным распределениям. В качестве решений указанной проблемы в литературе предложен ряд способов адаптации копульного подхода к дискретным переменным. К ним относятся:

- метод сглаживания: добавление небольшого случайного шума (jittering) к дискретным значениям, что позволяет работать с приближенно непрерывной выборкой;

- латентные переменные: переход к моделям со скрытыми непрерывными переменными, распределенными условно на наблюдаемые дискретные.

Таким образом, хотя требование непрерывности остается формальным ограничением, на практике оно может быть преодолено с помощью корректных процедур преобразования данных.

**Непостоянство распределений во времени.** Еще одной серьезной проблемой, особенно актуальной в задачах анализа временных рядов, является временная изменчивость распределений. И маргинальные распределения, и структура зависимости между переменными могут меняться со временем – например, в условиях кризиса активы становятся сильнее коррелированными, а формы распределений приобретают тяжелые хвосты. Маргинальные распределения переменных, включая зависимую переменную часто нестационарны из-за тренда (например, рост продаж) и сезонных колебаний. Это приводит к тому, что:

- положение распределения  $f_Y(y)$  смещается во времени (например, вправо при росте уровне эндогенной переменной);

- форма распределения также может меняться (например, увеличение дисперсии при росте объема);

- оценки копул, построенные на исторических данных, перестают быть релевантными для будущих периодов.

Если не учитывать эту динамику, прогноз будет занижен (или завышен, в зависимости их изменения ряда) или смещен по сезонному паттерну. Однако существуют подходы, помогающие устранить в какой-то степени эти подходы.

1) Декомпозиция ряда перед копульным моделированием. Перед построением модели рекомендуется удалить тренд и сезонность из маргинальных случайных переменных ряда, описанного формулой (83)

$$Y_t = T_t + S_t + R_t \text{ (или } T_t \times S_t \times R_t), \quad (83)$$

где  $T_t$  – трендовая компонента;

$S_t$  – сезонная компонента;

$R_t$  – остаток.

Строится модель копулы для компоненты  $R_t$ . После прогнозирования условного распределения для  $R_{t+1}$ , прогноз возвращается в исходный масштаб через обратную реконструкцию, соответствующую формуле (84)

$$\hat{Y}_{t+1} = \hat{T}_{t+1} + \hat{S}_{t+1} + \hat{R}_{t+1} \text{ (или } \hat{T}_{t+1} \times \hat{S}_{t+1} \times \hat{R}_{t+1}). \quad (84)$$

2) Динамические копулы. Если наряду со сдвигом маргинальных распределений наблюдаются изменения в структуре зависимости, целесообразно обращаться к динамическим копулам, в которых параметры (например, параметр корреляции нормальной копулы) задаются как функции времени в соответствии с формулой (85)

$$C_t(u_Y, u_1, \dots, u_k) = C(u_Y, u_1, \dots, u_k, \theta(t)). \quad (85)$$

Функция  $\theta(t)$  моделируется отдельно, например через ARIMA-подобную модель. В этом случае модель подстраивается под изменения временного ряда как по зависимости, так и по форме распределения. Подход в целом допустим, однако он вычислительно затратен и потому труднореализуем на практике.

Таким образом, предложенный метод можно рассматривать как устойчивую альтернативу традиционным регрессионным моделям. В отличие от параметрических методов, он ослабляет их типичные ограничения и

позволяет более гибко интерпретировать взаимосвязи между переменными, в том числе оценивать различные регрессионные характеристики. Наблюдаемые проблемы, в частности требования к непрерывности и чувствительность к временным сдвигам распределений, в некоторых случаях перекрываются сглаживанием, декомпозицией рядов и адаптивным переобучением, что поддерживает как теоретическую состоятельность, так и практическую пригодность метода для прогнозирования и исследования сложных взаимосвязей.

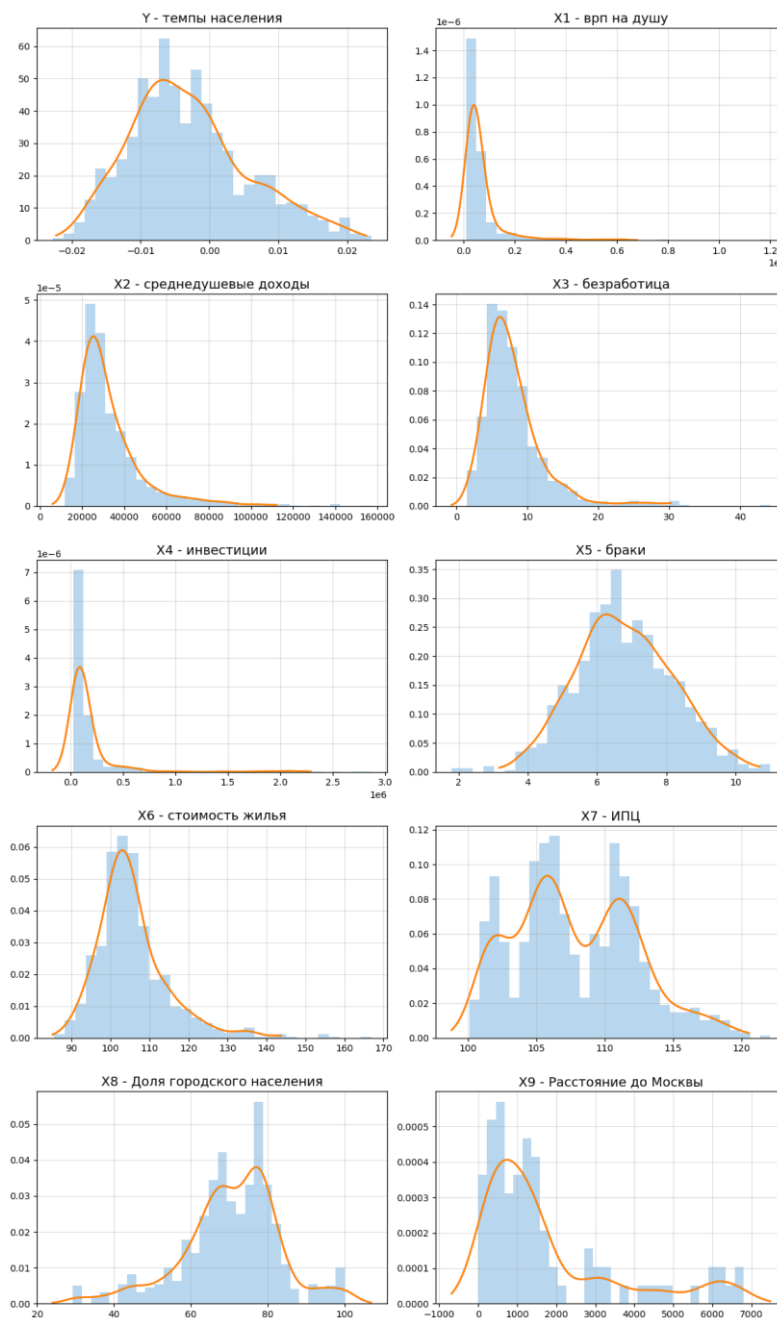
### **3.3 Множественная Vine-копула регрессия для оценки факторов демографического развития**

В настоящем параграфе представлена практическая реализация многомерной регрессионной модели на основе Vine-копул. Построенная с использованием демографических и социально-экономических показателей и дополненная контрольными факторами (временной тренд и фиктивные переменные по федеральным округам), она сводит выявленные нелинейные и условные связи в единую согласованную структуру, что позволяет:

- 1) моделировать условное распределение темпов прироста населения для произвольных сочетаний значений социально-экономических факторов, включая предельные сценарии;
- 2) выявить взаимодействия между факторами, то есть установить, как влияние одной переменной изменяется при разных значениях других (например, как эффект инвестиций связан с конъюнктурой рынка труда);
- 3) оценить сравнительную значимость факторов в многомерной постановке, отделяя основные каналы воздействия от вторичных синергетических эффектов;
- 4) обеспечить более высокое качество прогноза по метрикам MAE, RMSE и  $R^2$  относительно стандартных линейных и параметрических регрессионных моделей, используемых в эконометрическом анализе,

сохранив интерпретируемость и аналитическое описание условных распределений.

Построение модели множественной Vine-копулы выполняется по систематической схеме, теоретически обоснованной ранее и адаптированной к демографическим данным. Процедура включает три шага; на первом этапе непараметрически оцениваются функции плотности и распределения для каждой переменной, включая эндогенную и экзогенные.



Источник: составлено автором.

Рисунок 12 – Оценка ядерных функций плотности для переменных модели

Анализ ядерных оценок плотности на рисунке 12 показывает существенную неоднородность распределений (как демографической переменной  $Y$ , так и социально-экономических факторов  $X_i$ ). Темпы прироста населения характеризуются узким, близким к симметричному вокруг нуля распределением, что согласуется с наблюдаемой стагнацией демографической динамики в регионах Российской Федерации. В то же время показатели ВРП на душу населения, доходы, инвестиции в основной капитал, стоимость жилья характеризуются правосторонней асимметрией, что указывает на наличие небольшого числа регионов, которые заметно смещены на правый край распределения относительно медианной величины. Для части переменных отмечается мультимодальность (например, ИПЦ и доля городского населения), что связано с существованием устойчивых групп регионов и различиями в условиях развития. С учетом указанных свойств распределений, непараметрический подход является обоснованным выбором, поскольку нет оснований для предположения линейности и нормальности, и в таком случае Vine-копулы является релевантным инструментом моделирования сложных, в том числе асимметричных, зависимостей.

После оценки маргинальных распределений и перехода к ранговому представлению  $u_{i,j} \in [0, 1]$  выполняется второй этап – реконструкция парных зависимостей через построение Vine-копулы. Здесь фиксируется топология многомерной зависимости: какие пары переменных образуют основу модели, а какие связи следует рассматривать как условные.

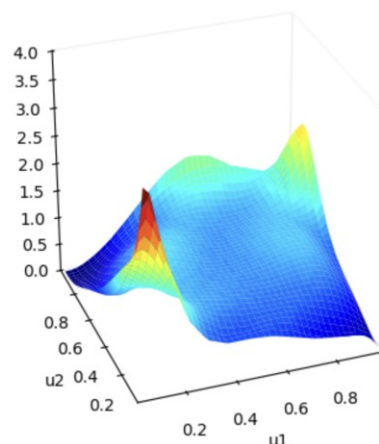
Максимальное остовное дерево  $T_1$ , построенное на модуле рангового коэффициента Кендалла  $|\tau|$ , задает структурой первого дерева. Оно позволяет отобрать наиболее выраженные и устойчивые парные зависимости, которые составят основу многомерной структуры в дальнейшем. Таким образом, в дереве  $T_1$  фиксируются наиболее существенные прямые (безусловные) связи между переменными.

После определения структуры дерева  $T_1$  алгоритм последовательно строит деревья  $T_2, T_3, \dots$ , где каждая последующая структура отражает

условные зависимости, то есть взаимосвязи между переменными при фиксированных значениях остальных. На этих уровнях оцениваются бивариатные копулы вида  $C_{i,j|D}(u_i, u_j|u_D)$ . Таким образом формируется полная R-vine-структура, позволяющая моделировать сложные многомерные зависимости, включая асимметрию, неоднородность и хвостовые эффекты.

В процессе построения R-vine-копультной модели использовались различные семейства бивариатных копул класса эллиптических и архимедовых, включая гауссовскую и t-копулу (для симметричных зависимостей), копулы Клейтона, Гумбеля, Франка и Джо (для асимметричных зависимостей), а также более гибкие семейства BB7, BB8 и Tawn, позволяющие моделировать выраженные хвостовые эффекты. В случаях, когда параметрические копулы не обеспечивали адекватного описания зависимости, применялась непараметрическая t-LL-копула. Для некоторых пар переменных была выявлена статистическая независимость, что отражено использованием копулы независимости. Структуру и параметры R-Vine копулы, обученной на исходных данных, а также матрицу зависимостей можно увидеть в приложениях Б и В.

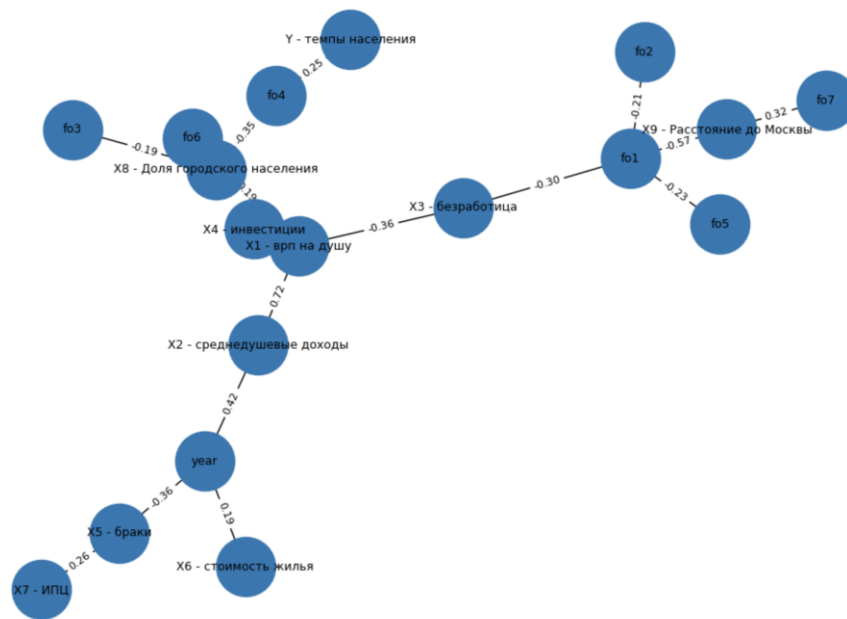
В качестве примера на рисунке 13 проиллюстрируем зависимость, полученную через t-LL-копулу для переменных инвестиций в основной капитал ( $X_4$ ) и ИПЦ ( $X_7$ ), графики для дерева  $T_1$  приведены в приложении Г.



Источник: составлено автором.

Рисунок 13 – t-LL-копула зависимость в структуре дерева  $T_1$  между переменными: инвестициями в основной капитал ( $X_4$ ) и ИПЦ ( $X_7$ )

Построенную Vine-копулу следует трактовать не как простой граф зависимостей, а как иерархически организованную систему связей, где для каждой пары переменных оценивается собственная бивариатная копула, что повышает адаптивность модели. Такая схема позволяет одновременно учитывать как выраженные прямые зависимости (представленные в  $T_1$ ), так и более слабые условные связи, которые проявляются лишь при совместном рассмотрении нескольких факторов; это обстоятельство существенно для анализа пространственной демографической динамики. Для большей наглядности структура дерева  $T_1$  будет графически представлена на рисунке 14.



Источник: составлено автором.

Рисунок 14 – Структура дерева  $T_1$  для рассматриваемой модели (на ребрах указаны значения корреляции  $\tau$  Кендалла)

Структура первого дерева  $T_1$  фиксирует примечательную закономерность: темпы прироста населения непосредственно соотносятся лишь с индикатором Северо-Кавказского федерального округа ( $f_04$ ). Такой вывод в целом ожидаем и согласуется с тем, что для данного региона характерны устойчиво более высокие демографические значения. Вместе с тем трактовка этого результата как «отсутствия влияния» прочих факторов

представляется методологически неверной. В рамках R-vine первое дерево включает только наиболее выраженные безусловные связи; при этом экономические и социальные детерминанты, как правило, проявляются в отношении  $Y$  через условные зависимости на более высоких уровнях модели (деревья  $T_2, T_3, \dots$ ).

Построенное дерево показывает, что экономические переменные группируются в единый кластер: ВРП на душу населения ( $X_1$ ), показатели доходов ( $X_2$ ), безработицы ( $X_3$ ), инвестиций ( $X_4$ ). Для подобных индикаторов характерна тесная взаимосвязанность и совместная динамика; как правило, между ними наблюдаются высокие корреляции. По этой причине в максимальном остовном дереве указанные переменные располагаются рядом друг с другом и формируют компактный экономический блок. Такая структура отражает их внутреннюю согласованность в общей системе социально-экономических связей.

Еще один значимый кластер составляют пространственные факторы (расстояние до Москвы  $X_9$  и производные от него фиктивные переменные федеральных округов). Эти показатели выделяются в самостоятельную группу, что в целом согласуется с наблюдаемой структурой данных: территориальная организация регионов сказывается как на демографических процессах, так и на экономических характеристиках, однако соответствующий эффект проявляется в отдельных комбинациях признаков и не сводится к устойчивой безусловной корреляции.

После оценки параметров Vine-копулы можно перейти к следующему этапу – восстановлению условного распределения эндогенной переменной  $Y$  при фиксированных значениях факторов.

Прогнозные значения  $Y$  определяются на основе выражений (74)-(77), в которых используются совместная плотность, условная плотность и условное математическое ожидание. При этом условная плотность получается из совместного распределения с учетом фиксированных значений экзогенных переменных. Важная особенность такого подхода состоит в том, что он не

требует задания линейной формы связи и не опирается на предпосылку нормальности ошибок. Вместо этого используется структура зависимости, ранее восстановленная с помощью Vine-копулы.

На практике условное математическое ожидание в модели Vine-копул рассчитывается численно, поскольку в явном аналитическом виде оно обычно не выражается. Поэтому дальнейшую процедуру целесообразно представить поэтапно:

- На первом этапе в ранговом пространстве задается регулярная сетка значений эндогенной переменной  $u_{grid} = \{u^{(1)}, u^{(2)}, \dots, u^{(M)}\}$  в ранговом пространстве (например,  $M = 1000$  точек). Через квантильную функцию ядра маргинального распределения  $Y$  сетка преобразуется в исходное пространство:  $y_{grid} = F_Y^{-1}(u_{grid})$ .

- Для каждой точки сетки строится полная матрица размерности  $M \times p$ , где  $p$  – количество переменных. На каждой строке матрицы первый элемент – это текущая  $u^{(j)}$  из сетки  $u$ , а остальные девять элементов – это фиксированные ранги экзогенных переменных  $(\hat{u}_1, \hat{u}_2, \dots)$  для конкретного наблюдения из выборки. Так, матрица описывает: «как плотность копулы меняется, когда варьируется эндогенная переменная, а остальные факторы фиксированы».

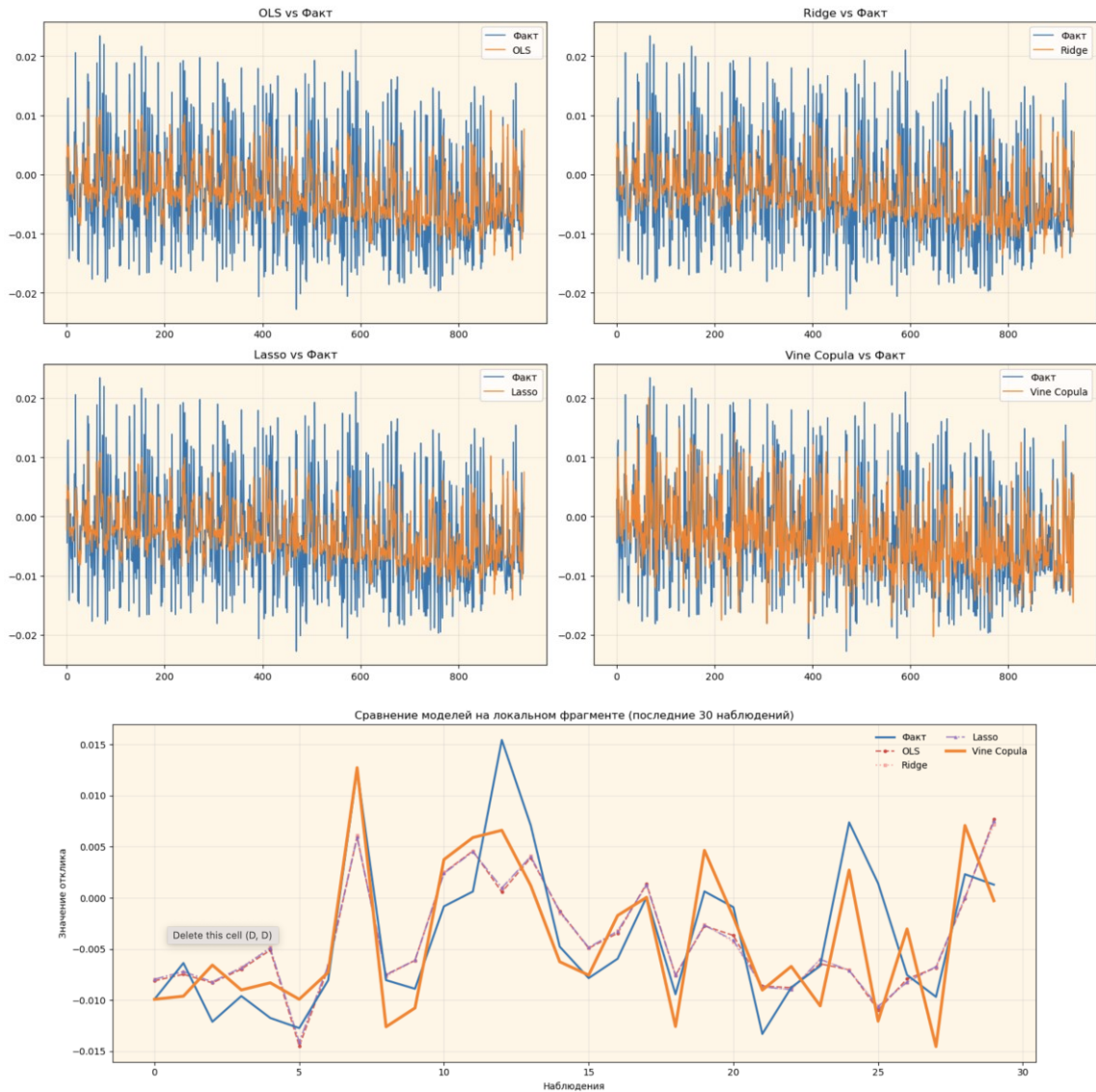
- Для каждой строки матрицы вычисляется значение плотности копулы  $c(u_y^{(j)}, \hat{u}_1, \hat{u}_2 \dots)$  в точке. Результат – вектор длины  $M$  значений плотности вдоль сетки.

- Используется правило трапеций, которая аппроксимирует интегралы суммами, рассчитывается условное математическое ожидание (прогноз регрессии).

- Описанная процедура повторяется для каждого наблюдения в тестовой выборке, результатом чего является вектор прогнозов  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ .

Для оценки предсказательной способности Vine-копул сравним их с линейной регрессией, Lasso и Ridge. Указанные базовые модели предполагают

линейную зависимость и, как правило, не отражают нелинейности, асимметрию и хвостовую зависимость, тогда как Vine-копулы способны их учитывать. Сопоставление метрик точности между Vine-копулами и этими подходами позволяет оценить практический эффект нелинейного моделирования в рассматриваемой задаче.



Источник: составлено автором.

Рисунок 15 – Факт и прогноз темпов прироста населения в различных моделях

На рисунке 15 представлены траектории фактических темпов прироста населения и соответствующие прогнозные значения, полученные по четырем моделям за период 2010-2023 гг.: OLS, Ridge, Lasso и Vine-копуле. Наблюдается, что все линейные подходы (OLS, Ridge, Lasso) формируют относительно сглаженный прогноз, который в целом занижает амплитуду

колебаний по сравнению с фактической динамикой. В частности, данные модели слабо воспроизводят локальные подъемы и снижения, что, по-видимому, связано с заданной линейной спецификацией зависимостей и отсутствием учета возможных нелинейных взаимодействий между факторами.

Vine-копула дает более вариативный прогноз и лучше повторяет форму фактического ряда, что особенно заметно в участках с резкими изменениями. Однако, несмотря на более адекватную динамику, Vine-копула все еще не может полностью воспроизвести ряд, что связано с высокой шумностью показателя и региональными особенностями, которые не описаны исходными факторами.

Таблица 5 – Показатели точности для различных моделей регрессии для темпов прироста населения

model	MAE	RMSE	R <sup>2</sup>
Copula	0,00473	0,00646	0,585040
Ridge	0,00547	0,00718	0,451932
Lasso	0,00548	0,00718	0,452734
OLS	0,00549	0,00717	0,454120

Источник: составлено автором.

Сопоставление метрик в таблице 5 показывает, что Vine-копула обеспечивает заметно более высокую точность по сравнению с линейными моделями. Средняя абсолютная ошибка снижается примерно на 14%, а среднеквадратичная ошибка – почти на 10%.

Для задач моделирования темпов прироста населения, где на результат влияют трудноформализуемые факторы – культурные нормы, ментальные установки, этническая структура, особенности миграционного поведения, институциональная специфика регионов – такое снижение ошибки является значимым. Модель не может полностью устранить системный шум, связанный с демографическими особенностями регионов, однако она демонстрирует способность улавливать форму зависимости и структурные закономерности,

что делает ее пригодной для сценарного анализа и оценки направления влияния ключевых факторов.

Дополнительным аргументом в пользу корректности Vine-копула-подхода служит структура дерева  $T_1$ . В нем именно ВРП на душу населения ( $X_1$ ) выступает центральным узлом, от которого расходятся основные экономические показатели – инвестиции, доходы, безработица, стоимость жилья, браки, инфляция. Это означает, что модель «видит» ядро связей в экономическом блоке и именно через  $X_1$  агрегирует воздействие большинства макрофакторов. В такой конфигурации естественно ожидать, что при выборе ВРП на душу в качестве эндогенной переменной качество описания и прогнозирования будет существенно выше, чем для темпов прироста населения: экономические взаимосвязи для модели более «прозрачны».

Для эмпирической проверки данного вывода далее рассматриваются результаты моделирования, в которых ВРП на душу населения выбирается в качестве эндогенной переменной. Показатели точности для различных спецификаций регрессии ВРП на душу населения приведены в таблице 6. Ниже сопоставляются метрики качества (MAE, RMSE, R2). Такое сравнение позволяет оценить, в какой мере Vine-копула действительно более адекватно отражает экономические зависимости; при этом в дальнейшем анализе и сценарном моделировании основное внимание, как и прежде, уделяется темпам прироста населения как целевому показателю.

Таблица 6 – Показатели точности для различных моделей регрессии ВРП на душу населения

model	MAE	RMSE	R <sup>2</sup>
Copula	101 751	298 048	0,932123
Lasso	157 022	392 240	0,882441
Ridge	161 748	405 064	0,874629
OLS	160 059	391 376	0,882959

Источник: составлено автором.

Проведенный в параграфе анализ показал, что непараметрический подход на основе Vine-копул, как правило, обеспечивает более высокое качество моделирования демографической динамики по сравнению с классическими линейными методами. Использование ядерных оценок маргинальных распределений вместе с полной R-vine-структурой позволило описать нелинейные, асимметричные и условные связи между социально-экономическими показателями, которые обычно не фиксируются линейными моделями. Это подтверждается сравнением точности: Vine-копула демонстрирует меньшие MAE и RMSE, а реконструированный временной ряд лучше совпадает с наблюдениями, в частности на интервалах с резкими изменениями темпов прироста населения.

Таким образом, Vine-копула-регрессия является адекватным инструментом пространственного анализа демографической динамики. Она повышает точность восстановления наблюдаемых значений и задает методическую базу для последующей оценки влияния факторов, построения зависимостей и сценарного прогнозирования, которая рассматривается в следующем параграфе.

### **3.4 Структура влияния социально-экономических факторов на демографический рост: региональные эффекты и сценарное моделирование**

Построенная многомерная модель Vine-копулы дает возможность не только оценить предсказательные характеристики, но и подробно анализировать, каким образом факторы соотносятся с эндогенной переменной. В рамках классических линейных моделей подобная постановка затруднена, поскольку предполагается неизменный аддитивный вклад каждой переменной; при этом нелинейность, асимметрия и взаимозависимости между факторами фактически остаются вне анализа. Копулы, напротив, позволяют моделировать полное совместное распределение переменных и,

следовательно, вычислять условные математические ожидания, а также оценивать вклад каждого фактора как в безусловной, так и в условной интерпретации.

В параграфе представлен количественный анализ влияния социально-экономических переменных на демографическую динамику на базе расчетов частных эффектов и чувствительности  $E(Y|X_k)$  к изменениям факторов. Такой подход дает возможность выяснить, какие показатели статистически значимо связаны с демографическими процессами в среднем по стране, а какие оказываются значимыми только при фиксированных значениях других переменных. Отдельно выделены участки нелинейных зависимостей, когда изменение фактора сопровождается резким усилением или снижением темпов изменения численности населения.

Поскольку Vine-копула моделирует условные зависимости на всех уровнях деревьев  $T_1, T_2, \dots$ , факторы могут воздействовать на темпы прироста населения как напрямую, так и через систему промежуточных экономических связей. Например, влияние инфляции может усиливаться или ослабляться в зависимости от уровня доходов, а эффект стоимости жилья – зависеть от структуры расселения или макрорегиона. Такой многомерный характер связей позволяет более реалистично оценивать факторы демографических рисков и роста.

В рамках параграфа последовательно рассматриваются: оценка вкладов отдельных факторов в качество объяснения модели, анализ кривых влияния основных социально-экономических переменных, а также сценарные расчеты в разрезе федеральных округов. Полученные результаты используются далее для формулирования прикладных выводов и рекомендаций, релевантных задачам демографического блока Стратегии-2030.

Оценка вкладов факторов в рамках Vine-копула-регрессии основана на сравнении точности модели при последовательном исключении отдельных переменных. В качестве базового варианта рассматривается полная модель, включающая все социально-экономические факторы  $X_1, \dots, X_9$ , временной

тренд и дамми-переменные федеральных округов; для нее вычисляется контрольное значение средней квадратической ошибки прогноза  $RMSE_{full}$ . Далее для каждого фактора  $X_j$  строится модифицированная Vine-копула-модель с той же архитектурой, но без данного признака, и оценивается соответствующая ошибка  $RMSE_{-j}$ . Частный вклад  $j$ -го фактора в улучшение качества прогноза определяется как приращение ошибки при его исключении:  $\Delta RMSE_j = RMSE_{-j} - RMSE_{full}$ . Для сопоставимости по шкале вводится также относительная доля вклада фактора в общей сумме приращений по формуле (86)

$$s_j = \frac{\Delta RMSE_j}{\sum_k \Delta RMSE_k} \times 100\%, \quad (86)$$

где суммирование ведется по всем факторам с положительным вкладом. Такой подход интерпретируется как аналог «важности признаков» в нелинейной модели: чем сильнее ухудшается точность при удалении переменной, тем более существенным является ее вклад в объяснение вариации темпов прироста населения.

Таблица 7 – Вклад макроэкономических факторов в оценке темпов прироста населения в Vine-копула регрессии

Фактор	$RMSE_{full}$	$RMSE_{-j}$	$\Delta RMSE_j$	$s_j$
X4 - инвестиции	0,006172	0,006492	0,000320	22,4
X2 - среднедушевые доходы	0,006172	0,006490	0,000319	22,3
X8 - доля городского населения	0,006172	0,006423	0,000252	17,6
X1 - ВРП на душу	0,006172	0,006367	0,000196	13,7
X3 - безработица	0,006172	0,006337	0,000166	11,6
X5 - браки	0,006172	0,006252	0,000081	5,7
X9 - расстояние до Москвы	0,006172	0,006222	0,000050	3,5
X6 - стоимость жилья	0,006172	0,006216	0,000044	3,1
X7 - ИПЦ	0,006172	0,006169	-0,000002	-

Источник: составлено автором.

Анализ таблицы 7 показывает, что в рамках Vine-копула-регрессии ключевыми каналами воздействия на темпы прироста населения являются инвестиции, текущие доходы, степень урбанизации, уровень экономического развития и рынок труда, тогда как брачность, пространственное положение и стоимость жилья играют вспомогательную роль, а инфляция не обладает самостоятельной объясняющей силой. Важным преимуществом такого подхода является то, что вклад оценивается не в линейной модели, а в полной нелинейной структуре зависимостей: каждая  $\Delta RMSE_j$  отражает потерю информации с учетом всех условных и хвостовых эффектов, уже встроенных в Vine-копула-модель.

На следующем этапе анализируются кривые частной зависимости, позволяющие уточнить характер связи между темпами прироста населения и каждым макроэкономическим фактором. В отличие от показателей вкладов, которые дают агрегированную оценку важности переменной через снижение качества модели при ее исключении, эти кривые описывают локальные свойства условного математического ожидания. В результате можно проследить, как изменяются ожидаемые темпы прироста населения при вариации выбранного показателя при фиксированных значениях прочих факторов.

В демографическом анализе такие кривые полезны тем, что позволяют выявлять нелинейные зоны влияния, включая пороги урбанизации или дохода, после которых отклик возрастает; оценивать асимметрию, когда ухудшение фактора действует сильнее, чем сопоставимое улучшение; и учитывать модификацию эффектов региональными особенностями. В Vine-копула-модели эти связи выводятся из многомерного распределения и интерпретируются как структурная зависимость с меньшей примесью случайных колебаний, тогда как в стандартной регрессии обычно задается постоянный линейный коэффициент.

Кривые влияния дают возможность рассматривать факторы по федеральным округам и сопоставлять воздействие экономических

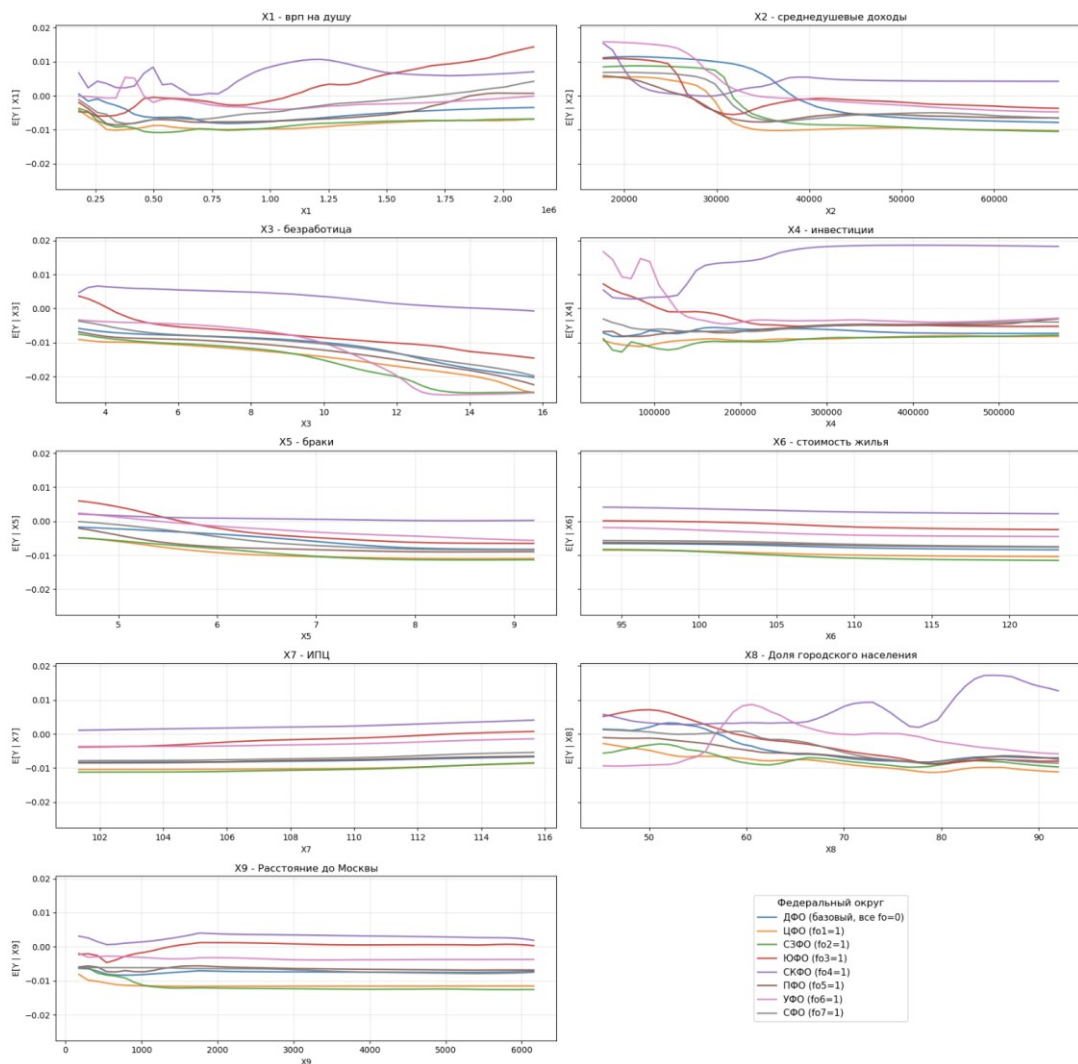
показателей на демографическую динамику в регионах различного типа, включая развитые, аграрные, северные, удаленные и миграционно подвижные. Это позволяет выявлять не только общие зависимости, но и территориально специфические эффекты, необходимые для обоснования адресных мер демографической политики.

Кривые эффектов рассматриваются как оценки условного математического ожидания  $E[Y|X_j = x]$ , получаемые при варьировании одного фактора при фиксированных значениях прочих переменных. В каждом федеральном округе в качестве базы используются медианные по регионам значения показателей за 2023 год, которые задают типичный профиль округа и далее принимаются постоянными.

Затем выбранный фактор  $X_j$  последовательно пробегает характерный для данных диапазон значений, а для каждого значения  $x$  с помощью обученной Vine-копула-модели вычисляется условное ожидание  $E[Y|X_j = x]$  при фиксированных остальных переменных на базовом уровне. Набор точек «уровень фактора – ожидаемые темпы прироста населения» формирует кривую влияния для этого показателя и округа. Сопоставление таких кривых между федеральными округами на единых графиках позволяет анализировать, как один и тот же экономический фактор по-разному воздействует на демографическую динамику в разных частях страны. Кривые влияния факторов на темпы прироста населения по федеральным округам изображены на рисунке 16.

С учетом вкладов макроэкономических факторов в оценке темпов прироста населения, приведенных в таблице 7, в интерпретации кривых влияния имеет смысл сосредоточиться на пяти ключевых показателях: инвестициях, среднедушевых доходах, доле городского населения, ВРП на душу и безработице. Для инвестиций ( $X_4$ ) практически во всех федеральных округах (за исключением южного и уральского федеральных округов) просматривается в целом положительная связь с темпами прироста населения:

при низком инвестиционном уровне условное ожидание темпов прироста отрицательно или близко к нулю, по мере наращивания инвестиций кривая либо выходит в зону слабого положительного прироста, либо заметно растет (особенно в СКФО). Это соответствует интуиции: по мере роста капитальных вложений в регионы улучшается инфраструктура, создаются рабочие места и повышается привлекательность территории для закрепления и притока населения.



Источник: составлено автором.

Рисунок 16 – Кривые влияния факторов на темпы прироста населения (по ФО)

Кривые по среднедушевым доходам ( $X_2$ ) имеют более сложную, «ломаную» форму. При низких доходах наблюдается небольшой рост ожидаемых темпов прироста населения, что отражает эффект догоняющего развития: минимальное повышение доходов в бедных регионах сначала

позволяет «донастроить» базовое потребление и поддержать рождаемость. В диапазоне средних доходов почти во всех округах кривая резко уходит вниз, то есть при достижении определенного уровня обеспеченности темпы прироста населения снижаются. На высоких значениях дохода наклон вновь выравнивается, переходя в пологое плато. Такое поведение хорошо согласуется с тем, что наиболее богатые регионы (Москва, Санкт-Петербург, некоторые административные центры) характеризуются низкой рождаемостью и интенсивной миграцией трудоспособного населения, что тянет демографические показатели вниз.

Доля городского населения ( $X_8$ ) и ВРП на душу ( $X_1$ ) отражают два разных аспекта структурного развития. Для ВРП на душу ожидаемые темпы прироста населения в большинстве округов растут лишь до определенного порога, после чего кривая становится почти горизонтальной: дополнительный рост экономического выпуска сам по себе уже не приводит к ускорению демографической динамики, если не подкреплён улучшением качества жизни и инфраструктуры. Доля городского населения, наоборот, демонстрирует неоднородную картину: в ряде округов рост урбанизации сопровождается легким снижением ожидаемых темпов, что соответствует типичной «урбанизационной ловушке» (меньше детей, выше миграционная подвижность). При этом, для Северо-Кавказского и частично Южного федеральных округов кривая заметно изгибается: при средних показателях уровня урбанизации темпы прироста оказываются максимальными, тогда как при слишком низкой или слишком высокой доле горожан они снижаются. Это можно интерпретировать как стремление найти баланс между традиционными сельскими сообществами и привлекательностью городской жизни.

Кривые безработицы ( $X_3$ ) довольно ожидаемо являются монотонно убывающими: в каждом из округов повышение уровня безработицы сопровождается снижением ожидаемых темпов прироста населения. Причем в Уральском и Северо-Западном ФО наиболее крутые отрицательные склоны,

это значит, что неустойчивость рынка труда быстрее трансформируется в отток населения и снижение рождаемости.

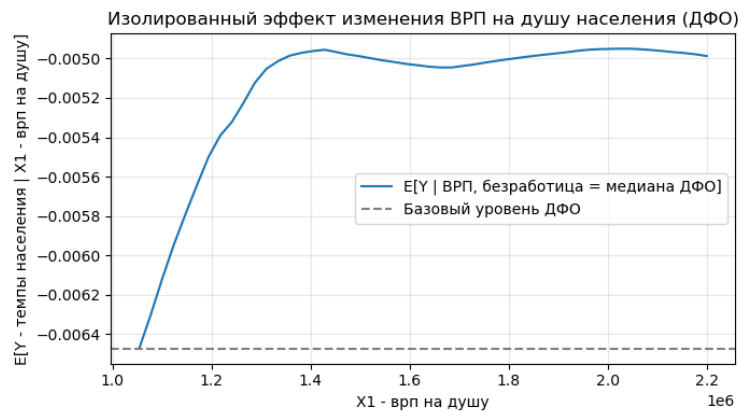
Наконец, визуально «парадоксальные» зависимости, такие как снижение ожидаемых темпов при росте показателя брачности ( $X_5$ ), не следует трактовать как буквальный причинный эффект. Этот фактор имеет небольшой вклад в общей точности модели, а форма кривой, скорее всего, отражает структурные различия между регионами (возрастная структура, миграция молодежи в крупные города, особенности регистрации браков), чем прямое влияние числа браков на темпы прироста населения.

С учетом вышеизложенных выводов в завершение эмпирического анализа выделим отдельный сценарный блок, фокусирующийся на совместном влиянии ключевых экономических факторов, в нашем примере уровня ВРП на душу населения и безработицы. В качестве иллюстративного кейса выбирается Дальневосточный федеральный округ, который за рассматриваемый период продемонстрировал наиболее выраженное сокращение численности населения. Именно для ДФО вопрос, компенсирует ли экономический рост отток населения или, наоборот, высокий уровень безработицы обнуляет эффект роста ВРП, носит принципиальный характер для демографической политики.

Задача сценарного моделирования в этом параграфе формулируется как оценка не только отдельных, но и совместных эффектов двух факторов. На основе построенной Vine-копула-регрессии будут последовательно измеряться:

- «изолированный» эффект изменения ВРП на душу населения при фиксированном уровне безработицы (кривая  $E[Y | X_1, X_3 = const]$ );
- «изолированный» эффект изменения безработицы при фиксированном ВРП (кривая  $E[Y | X_3, X_1 = const]$ );
- совместный эффект одновременного роста ВРП и снижения безработицы относительно базового сценария (кривая  $E[Y | X_3, X_1]$ ).

Сравнение совместного эффекта с суммой двух изолированных позволит оценить, проявляется ли в демографическом отклике синергия (когда одновременное улучшение экономических условий дает больший прирост темпов прироста населения, чем простая сумма отдельных эффектов) или, напротив, наблюдается эффект насыщения/компенсации. Такой формат сценариев не привязан к конкретному периоду и задает универсальную рамку для интерпретации экономико-демографических решений в ДФО и сопоставимых регионах. Кривая изолированного эффекта ВРП на темпы прироста населения в Дальневосточном федеральном округе представлена на рисунке 17.



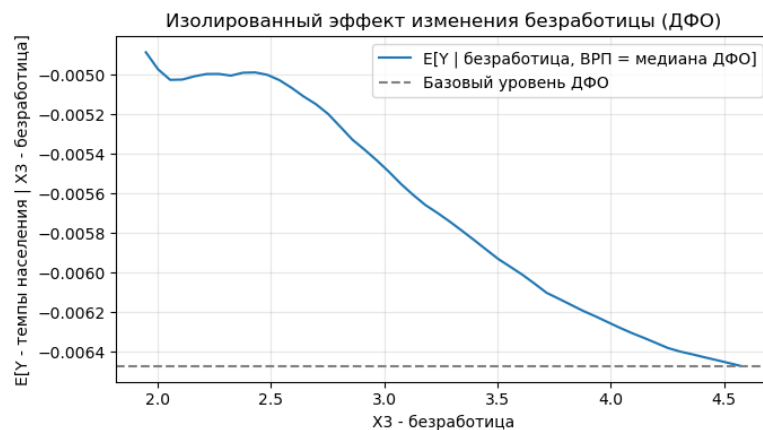
Источник: составлено автором.

Рисунок 17 – Кривая изолированного эффекта ВРП на темпы прироста населения в ДФО

Анализ рисунка 17 показывает, что рост ВРП на душу населения в ДФО оказывает умеренно положительное влияние на темпы прироста населения, однако эффект выражен нелинейно. При повышении ВРП от низких значений до уровня порядка 1,3-1,4 млн руб. фиксируется наиболее выраженное улучшение демографического индикатора по сравнению с базовым уровнем. Последующий рост ВРП сопровождается преимущественно постепенным ослаблением эффекта: показатели прироста населения продолжают улучшаться, однако предельный вклад экономического фактора становится заметно ниже. Конфигурация кривой позволяет предположить наличие диапазона убывающей демографической отдачи от экономического развития; это согласуется с тем, что после достижения минимально достаточной

экономической устойчивости на динамику населения в большей степени воздействуют иные детерминанты, в том числе инфраструктурные условия, социальная среда и миграционные процессы.

Кривая изолированного эффекта, приведенная на рисунке 18, указывает на устойчивую отрицательную зависимость между уровнем безработицы и темпами прироста населения в ДФО. Даже умеренное увеличение безработицы относительно медианного уровня ассоциируется с ощутимым ухудшением демографической динамики; при дальнейшем росте показателя ожидаемые темпы прироста населения снижаются почти линейно.



Источник: составлено автором.

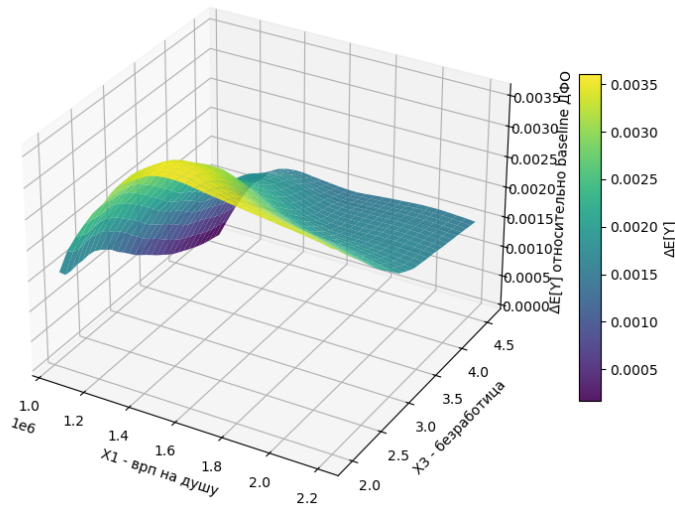
Рисунок 18 – Кривая изолированного эффекта безработицы на темпы прироста населения в ДФО

Наиболее заметное снижение фиксируется при переходе от уровня порядка 2,5-3% к значениям, превышающим 3,5%, что указывает на существенную чувствительность демографических реакций к параметрам регионального рынка труда. В целом конфигурация кривой согласуется с выводом о том, что занятость выступает одним из определяющих факторов демографической устойчивости: ухудшение конъюнктуры на рынке труда, как правило, в короткие сроки проявляется в снижении рождаемости и усилении миграционного оттока, прежде всего в экономически уязвимых регионах ДФО.

Представленная трехмерная поверхность на рисунке 19 отображает, как ожидаемые темпы прироста населения в ДФО изменяются при совместном

варьировании двух ключевых факторов – ВРП на душу населения и уровня безработицы – относительно их базовых (медианных) значений. График позволяет увидеть форму отклика модели в двумерном пространстве факторов и определить области, где изменение обоих показателей приводит к усилению или ослаблению демографической динамики.

Совместный эффект роста ВРП и снижения безработицы (ДФО)

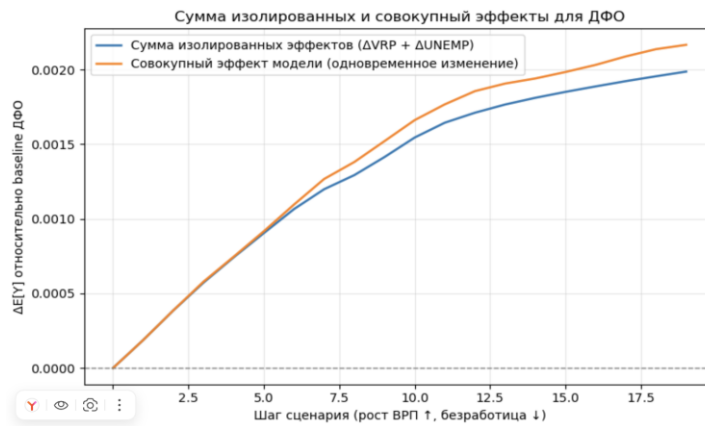


Источник: составлено автором.

Рисунок 19 – Кривая совместных эффектов ВРП и безработицы на темпы прироста населения в ДФО

Сравнение суммы изолированных эффектов (роста ВРП и снижения безработицы по отдельности) с совокупным эффектом их одновременного изменения на рисунке 20 показывает наличие умеренной положительной синергии в ДФО. На первых шагах оба подхода дают почти одинаковый результат, однако по мере усиления изменений разница постепенно возрастает.

В точке максимального расхождения совокупный эффект превышает сумму изолированных воздействий примерно на 0,2 процентного пункта в терминах темпов прироста населения относительно базового уровня. Это говорит о том, что улучшение положения на рынке труда усиливает демографический отклик при одновременном росте экономической базы региона.



Источник: составлено автором.

Рисунок 20 – Сравнение суммы изолированных и совместных эффектов

Модель указывает на то, что в ДФО комплексные меры – одновременное повышение уровня экономической активности (ВРП на душу населения) и снижение безработицы – дают больший демографический эффект, чем простой суммарный результат улучшений по отдельности. Это подтверждает наличие взаимного «поддерживающего» влияния факторов и важность многокомпонентной политики для регионов с устойчивым оттоком населения.

Полученные результаты демонстрируют, что темпы прироста населения определяются совокупным действием макроэкономических факторов, причем влияние большинства из них носит нелинейный, пороговый и регионально неоднородный характер. Наиболее значимые эффекты оказывают инвестиции, среднедушевые доходы, уровень урбанизации, экономическое развитие региона и ситуация на рынке труда. Именно улучшение этих показателей дает заметное увеличение ожидаемых темпов прироста населения, что подтверждается как вкладом факторов в модели, так и формой их кривых влияния.

С учетом выявленных закономерностей демографическая политика должна быть ориентирована не на универсальные меры, а на адресное улучшение ключевых социально-экономических параметров в зависимости от их текущего положения. Рост инвестиционной активности и формирование устойчивых региональных центров экономического роста в субъектах с низким уровнем ВРП могут частично компенсировать накопленные

демографические потери. В подобных условиях увеличение производительности труда и расширение занятости, как правило, заметнее отражаются на ожидаемой динамике численности населения. Вместе с тем следует принимать во внимание установленную нелинейность связи между доходами и рождаемостью: увеличение рождаемости фиксируется главным образом в интервале низких и средних доходов, тогда как в наиболее обеспеченных регионах (Москва, Санкт-Петербург, крупные административные центры) дальнейший рост заработной платы не приводит к сопоставимому увеличению рождаемости. Отсюда следует, что меры семейной политики в высокодоходных регионах целесообразно выстраивать не вокруг прямых финансовых стимулов, а через повышение доступности жилья, снижение транспортной нагрузки, улучшение качества городской среды и развитие механизмов, облегчающих совмещение занятости с воспитанием детей.

Кривые указывают на наличие статистически значимой связи между снижением безработицы и увеличением темпов прироста населения. Это дает основания для мер по целевой поддержке занятости молодежи и групп с низкой мобильностью, а также для стимулирования возвращения людей трудоспособного возраста из миграционных потоков в депрессивные территории. Одновременно оправдано развитие рабочих мест вне крупнейших городов, поскольку высокая доля городского населения, особенно в регионах с мегаполисами, ассоциируется с более медленным ростом рождаемости.

Сценарные расчеты показывают, что вклады факторов не являются аддитивными: при совместном снижении безработицы и улучшении экономической конъюнктуры прирост демографических показателей превосходит сумму эффектов, наблюдаемых при их раздельном анализе. Следовательно, демографическая политика без согласования с экономической будет существенно менее эффективной.

Таким образом, результаты моделирования свидетельствуют о целесообразности комплексной политики, ориентированной на повышение

инвестиционной привлекательности регионов, развитие инфраструктуры и мер занятости, а также на адресное увеличение доходов в бедных территориях и улучшение условий жизни в наиболее урбанизированных регионах. Реализацию таких мер следует выстраивать с учетом выраженной региональной неоднородности и нелинейных зависимостей, выявленных копульной моделью, поскольку это повышает ожидаемую результативность и поддерживает долгосрочный демографический рост.

### **Выводы по главе 3**

В третьей главе выполнено прикладное исследование демографической динамики регионов России в контексте целей и приоритетов Стратегии пространственного развития Российской Федерации до 2030 года с прогнозом до 2026 года. Проведенный анализ позволил получить целостное представление о факторах, формирующих пространственную демографическую динамику регионов России, а также количественно оценить характер и интенсивность их влияния. Использование непараметрической Vine-копула регрессии показало, что демографические процессы существенно нелинейны, асимметричны и зависят от взаимодействия социально-экономических показателей, что невозможно выявить в рамках классических линейных моделей.

Построенная многомерная Vine-копула модель позволила количественно оценить вклад каждого фактора и выявила, что ключевыми драйверами демографической динамики выступают инвестиции на душу населения, среднедушевые доходы, доля городского населения, ВРП на душу и безработица. При этом значимость факторов сильно меняется по макрорегионам: отдача от капитальных вложений и экономического роста наиболее выражена в дальневосточных, северокавказских и сибирских регионах; чувствительность к изменениям на рынке труда выше на территориях с исторически высоким уровнем оттока населения.

Построение кривых условных математических ожиданий позволило выявить зоны нелинейности и пороговые эффекты, что особенно важно для выработки регионально-ориентированной политики. В частности, рост доходов и ВРП оказывает максимальное воздействие на регионы с низкой исходной экономической базой, тогда как для крупнейших агломераций высокие уровни доходов уже не сопровождаются ускорением демографического роста из-за высокой стоимости жизни, насыщенности рынка труда и ограничений городской среды.

Сценарные расчеты подтвердили, что воздействие факторов не является аддитивным: при совместном улучшении условий рынок труда и экономический рост усиливают эффект друг друга. Для Дальневосточного федерального округа, выбранного как наиболее проблемная территория, совокупный эффект одновременного роста ВРП и снижения безработицы оказался выше суммы изолированных эффектов. Это означает наличие синергии: улучшение экономической среды повышает отдачу от мер, направленных на стабилизацию рынка труда, и наоборот.

Полученные результаты позволяют перейти от чисто описательного анализа к формированию ориентиров для государственной демографической политики. Оценки Vine-копула-модели показывают, что наибольший потенциал улучшения демографической динамики связан с комплексным воздействием на инвестиционный климат и занятость, прежде всего в регионах с устойчивой миграционной убылью. Для территорий со средним уровнем социально-экономического развития и сохранным потенциалом роста результаты кривых влияния указывают на важность сочетания повышения доходов населения и развития городской инфраструктуры. В то же время для крупнейших агломераций выявляется ослабление маржинального эффекта традиционных стимулов (рост доходов, дополнительная урбанизация), что косвенно говорит в пользу акцента на снижении ценовой нагрузки и смягчении негативных эффектов чрезмерной концентрации. В целом результаты главы подтверждают, что демографическая политика должна быть

дифференцирована по типам регионов, учитывать нелинейный характер влияния факторов и опираться не на изолированные меры, а на их синергетические комбинации.

## Заключение

В диссертации последовательно рассмотрены теоретические, методологические и прикладные аспекты применения копула-функций для анализа зависимостей и построения непараметрических моделей, а также продемонстрирована возможность их использования для исследования пространственной демографической динамики регионов Российской Федерации. Полученные результаты позволяют сформулировать итоговые выводы, отражающие как теоретическую значимость работы, так и ее прикладную ценность.

В первой главе диссертации сформирована экономическая постановка задачи и обоснован выбор копула-аппарата как методологической основы исследования. Показано, что пространственное развитие регионов Российской Федерации характеризуется устойчивой неоднородностью и поляризацией, а демографическая динамика выступает одним из ключевых факторов и одновременно ограничений социально-экономического развития, что зафиксировано в логике государственной стратегии пространственного развития. В этих условиях описание взаимосвязей между демографическими результатами и социально-экономическими условиями требует инструментария, способного корректно работать с нелинейностью, асимметрией и различием режимов в крайних областях распределений.

В главе изложены ключевые элементы теории копул: теорема Склера, ключевые свойства (включая инвариантность к монотонным преобразованиям), границы меры зависимости и интерпретация хвостовой зависимости. Также обсуждаются методы оценивания и сравнения спецификаций копул; в качестве основной процедуры обоснован метод максимального правдоподобия, который обеспечивает асимптотическую состоятельность и нормальность оценок параметров совместного распределения.

Во второй главе проведен эмпирический анализ парных связей между демографическим ростом и ключевыми социально-экономическими факторами по данным регионов Российской Федерации за 2010-2023 годы. Предварительная диагностика выявила заметные отклонения распределений ряда переменных от нормальности, включая асимметрию и концентрацию наблюдений в хвостах; наиболее выражено это для индикаторов экономического потенциала и инвестиционной активности. Дополнительно выявлена статистически значимая пространственная автокорреляция для большинства показателей, что подтверждает наличие территориальной кластеризации. Эти результаты служат эмпирическим обоснованием применения копульного инструментария как подхода, не требующего предпосылок линейности и нормальности и позволяющего анализировать структуру зависимости на разных участках распределения. Далее в главе введены и интерпретированы ранговые меры зависимости (коэффициенты Спирмена и Кендалла), а также коэффициенты верхней и нижней хвостовой зависимости как инструменты количественного описания согласованности экстремальных режимов. Для наглядного понимания того, как выбор семейства копулы влияет на форму условного ожидания, рассмотрены кривые копула-зависимостей на синтетических примерах. В завершающей части второй главы выполнена оценка верхней и нижней хвостовой зависимости между темпами прироста населения и социально-экономическими факторами при квантилях 0,90 и 0,95. Полученные оценки показали, что связь переменных в крайних областях распределений не является симметричной: для части факторов хвостовая зависимость проявляется преимущественно в нижнем хвосте (совпадение неблагоприятных режимов), для других – в верхнем хвосте (совпадение благоприятных режимов), а в ряде случаев зависимость заметно ослабевает при переходе от  $q = 0,90$  к  $q = 0,95$ . Тем самым выявлено, что одни и те же факторы могут «работать» по-разному в умеренно крайних и в наиболее экстремальных состояниях демографической динамики.

В третьей главе осуществлен переход от парной диагностики к многомерному анализу факторов демографической динамики на основе Vine-копул. Вначале введен многомерный аппарат Vine-построений как способ декомпозиции сложной структуры зависимости на систему парных и условных копул, что обеспечивает гибкость спецификации и позволяет учитывать неоднородные зависимости между отдельными переменными. Далее сформулированы предпосылки и ограничения Vine-копула регрессии как непараметрического подхода, в котором регрессионная зависимость трактуется через условное распределение и может анализироваться не только по условному математическому ожиданию, но и по иным функционалам распределения. На этой основе построена и оценена множественная Vine-копула регрессионная модель для темпов прироста населения с учетом набора социально-экономических факторов и региональных особенностей. Сопоставление с линейными моделями показало, что учет нелинейностей, асимметрий и условных зависимостей обеспечивает более высокое качество описания данных. В заключительной части главы выполнен анализ структуры влияния факторов, выявлена их неодинаковая роль в различных режимах и проведено сценарное моделирование, позволяющее оценивать демографический отклик при совместном изменении ключевых социально-экономических параметров. Тем самым получен инструмент не только для объяснения наблюдаемой динамики, но и для прикладной постановки сценариев в логике пространственной политики.

С научной точки зрения исследование расширяет возможности непараметрического моделирования в экономике, предлагая универсальный и интерпретируемый инструмент анализа многомерных зависимостей. Показано, что копула-регрессия способна заменить или дополнить методы машинного обучения, сохраняя при этом аналитическое выражение модели и интерпретируемость результатов, что важно для задач экономической теории и прикладной политики.

Практические результаты подтверждают, что разработанные модели применимы к анализу широкого круга задач, выходящих далеко за пределы демографического анализа: от моделирования рисков и поведения потребителей до анализа спроса, макроэкономических зависимостей и поведенческих данных. Применение копульной регрессии к региональным данным показало, что настоящий подход позволяет выявлять латентные структуры зависимостей. Такие итоги могут использоваться регулирующими органами, научными центрами и аналитическими подразделениями для составления политических стратегий, оценки воздействия экономических мер, а также для разработки сценариев регионального развития.

Подводя итоги, стоит отметить, что результаты исследования формируют задел для новых теоретических и прикладных исследований непараметрических моделей на основе копул. Дальнейшее развитие Vine-копульных моделей, улучшение методов оценивания, сочетание с методами машинного обучения и применение в новых областях, все это интересные направления для будущей работы. Исследование, проведенное в диссертации, показало высокую эффективность копула-подхода при анализе сложных нелинейных многомерных зависимостей и позволяет рассматривать его как методологически значимый инструмент для решения актуальных экономико-математических задач.

## Список литературы

1. Гранберг, А.Г. Основы региональной экономики : учебник для вузов / А.Г. Гранберг. – 4-е издание – Москва : ГУ ВШЭ, 2004. – 492 с. – ISBN 5-7598-0286-0.
2. Стратегия пространственного развития Российской Федерации на период до 2030 года с прогнозом до 2036 года [Распоряжение Правительства Российской Федерации от 28 декабря 2024 г. № 4146-р]. – Текст : электронный. – URL: <http://government.ru/docs/all/157308/> (дата обращения: 10.01.2025).
3. Федеральная служба государственной статистики: официальный сайт. – Москва. – Обновляется в течение суток. – URL: <https://rosstat.gov.ru/> (дата обращения: 30.03.2025). – Текст: электронный.
4. Естественное движение населения Российской Федерации за 2023 год: оперативные данные. – Электронные данные – Москва : Росстат, 2024. – Текст : электронный. – URL: <https://rosstat.gov.ru/folder/12781> (дата обращения: 29.11.2025).
5. РА Эксперт. Миграции населения в 2024 году: потоки ослабевают // Аналитический отчет. – Москва : 2024. – Текст : электронный. – URL: [https://raexpert.ru/researches/regions/migration\\_regions\\_2024/](https://raexpert.ru/researches/regions/migration_regions_2024/) (дата обращения: 29.11.2025).
6. Аганбегян, А.Г. Как восстановить сохранность народа России / А.Г. Аганбегян // Народонаселение. – 2021. – № 2. Том 24. – С. 4-18. – ISSN 1561-7785.
7. Бачаев, У.А. О многомерном математическом моделировании процессов в экономических системах в период экстремальных событий (пандемии) / У.А. Бачаев // Инновации и инвестиции. – 2024. – № 5. – С. 458-461. – ISSN 2307-180X.
8. Бачаев, У.А. Непараметрический подход к регрессионному моделированию на базе копула-функций на примере парных моделей

/ У.А. Бачаев // Проблемы экономики и юридической практики. – 2025. – № 3. Том 21. – С. 136-145. – ISSN 2541-8025.

9. Бачаев, У.А. Многомерная Vine-копула регрессия: теоретическое обоснование и предпосылки / У.А. Бачаев // Проблемы экономики и юридической практики. – 2026 – № 1. Том 22. – С. 163-169. – ISSN 2541-8025.

10. Бачаев, У.А. Моделирование многомерной структуры активов в задаче портфельной оптимизации на основе копул / У.А. Бачаев, Н.В. Гринева // Информационно-измерительные и управляющие системы. – 2023. – № 6. Том 21. – С. 16-28. – ISSN 2070-0814.

11. Бачаев, У.А. Прогнозирование пространственных эффектов и факторов развития региона с использованием методов машинного обучения / С.С. Михайлова, Н.В. Гринева, Ю.А. Кораблев, У.А. Бачаев // Computational Nanotechnology / ВЫЧИСЛИТЕЛЬНЫЕ НАНОТЕХНОЛОГИИ. – 2025. – № 3. Том 12. – С. 23-30. – ISSN 2313-223X.

12. Бачаев, У.А. Моделирование оптимального портфеля ценных бумаг с помощью многомерных копула-функций / У.А. Бачаев // Современная математика и концепции инновационного математического образования. – 2021. – № 1. Том 8. – С. 386-393. – ISSN 2412-9895.

13. Бачаев, У.А. Сценарное моделирование пространственно-демографической динамики регионов на основе Vine-копула регрессии / У.А. Бачаев // Экономика Строительства. – 2026. – № 3. – С. 564-567. – ISSN 0131-7768.

14. Гранберг, А.Г. Регионы и города России: сценарии-2020 / А.Г. Гранберг // Регион: экономика и социология. – 2000. – № 1. – С. 28-43. – ISSN 0868-5169.

15. Демидова, О.А. Методы пространственной эконометрики и оценка эффективности государственных программ / О.А. Демидова // Прикладная эконометрика. – 2021. – № 4(64). – С. 107–134. – ISSN 1993-7601.

16. Демидова, О.А. Модели экономического роста с неоднородными пространственными эффектами (на примере российских регионов) / О.А.

Демидова, Д.С. Иванов // Экономический журнал Высшей школы экономики. – 2016. – № 1. Том 20. – С. 52–75. – ISSN 1813-8691.

17. Зубаревич, Н.В. Региональное развитие и региональная политика в России / Н.В. Зубаревич // ЭКО. – 2014. – № 4. – С. 6-27. – ISSN 0131-7652.

18. Зубаревич, Н.В. Стратегия пространственного развития: приоритеты и проблемы реализации / Н.В. Зубаревич // Вопросы экономики. – 2019. – № 12. – С. 25-45. – ISSN 0042-8736.

19. Фантаццини, Д. Эконометрический анализ финансовых данных в задачах управления риском / Д. Фантаццини // Прикладная эконометрика. – 2008. – № 2. – С. 91-137. – ISSN 1993-7601.

20. Фантаццини, Д. Моделирование многомерных распределений с использованием копула-функций. I / Д. Фантаццини // Прикладная эконометрика. – 2011. – № 22. – С. 98-134. – ISSN 1993-7601.

21. Фантаццини, Д. Моделирование многомерных распределений с использованием копула-функций. II / Д. Фантаццини // Прикладная эконометрика. – 2011. – № 23. – С. 98-132. – ISSN 1993-7601.

22. Фантаццини, Д. Моделирование многомерных распределений с использованием копула-функций. III / Д. Фантаццини // Прикладная эконометрика. – 2011. – № 25. – С. 100-130. – ISSN 1993-7601.

23. Joe, H. Multivariate Models and Dependence Concepts / H. Joe. – London : Chapman & Hall, 1997. – 424 p. – ISBN 0412073315.

24. Joe, H. Dependence Modeling with Copulas / H. Joe. – Boca Raton : Chapman & Hall/CRC, 2014. – 480 p. – ISBN 1466583223.

25. Kurowicka, D. Dependence Modeling: Vine Copula Handbook / D. Kurowicka, H. Joe [et al.]. – Singapore : World Scientific, 2011. – 368 p. – ISBN 978-981-4299-87-9.

26. Nelsen, R.B. An Introduction to Copulas / R.B. Nelsen. – New York : Springer, 2006. – 277 p. – ISBN 978-0-387-28659-4.

27. Salvadori, G. *Extremes in Nature: An Approach Using Copulas* / G. Salvadori, C. De Michele, N.T. Kottegoda, R. Rosso. – Dordrecht : Springer, 2007. – 292 p. – ISBN 978-1-4020-4414-4.

28. Fang, K.-T. *Symmetric Multivariate and Related Distributions* / K.-T. Fang, S. Kotz, K.W. Ng. – London : Chapman & Hall, 1990. – 230 p. – ISBN 0412314304.

29. European Parliament. *Regional inequalities in the EU* // European Parliamentary Research Service. – Brussels, 2019. – Текст : электронный. – URL: [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_BRI%282019%29637951](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI%282019%29637951) (дата обращения: 01.04.2025).

30. International Monetary Fund. *Regional Disparities, Growth, and Inclusiveness in Emerging Market and Developing Economies* // IMF Working Paper. – WP/21/038. – Washington, D.C., 2021. – Текст : электронный. – URL: <https://www.imf.org/en/publications/wp/issues/2021/02/13/regional-disparities-growth-and-inclusiveness-50076> (дата обращения: 02.04.2025).

31. RegPol2 Project Team. *Territorial Policy Analysis: Central and Eastern Europe* / European Commission, DG for Regional and Urban Policy. – 2023. – Текст : электронный. — URL: <https://regpol2.eu/> (дата обращения: 05.04.2025).

32. Aas, K. *Pair-copula constructions of multiple dependence* / K. Aas, C. Czado, A. Frigessi, H. Bakken // *Insurance: Mathematics and Economics*. – 2009. – № 2. Том 44. – P. 182-198. – ISSN 0167-6687.

33. Baker, J. *Spatial weighting improves accuracy in small-area demographic forecasts of urban census tract populations* / J. Baker, A. Alcántara, H. Riojas-Rodríguez, J. Sargent // *Journal of Population Research*. – 2014. – № 1. Том 31. – P. 1-22. – ISSN 1443-2447.

34. Bathelt, H. *The nature, causes, and consequences of inter-regional inequality* / H. Bathelt, M. Buchholz, M. Storper // *Journal of Economic Geography*. – 2024. – № 3. Том 24. – P. 353-374. – ISSN 1468-2702.

35. Bedford, T. Vines – a new graphical model for dependent random variables / T. Bedford, R.M. Cooke // *Annals of Statistics*. – 2002. – № 4. Том 30. – P. 1031-1068. – ISSN 0090-5364.
36. Beneicke, J. Empowering Citizens' Cognition and Decision Making in Smart Sustainable Cities / J. Beneicke // *IEEE Consumer Electronics Magazine*. – 2020. – № 1. Том 9. – P. 102-108. – ISSN 2162-2248.
37. Brechmann, E.C. Modeling dependence with C- and D-vine copulas: The R package CDVine / E.C. Brechmann, U. Schepsmeier // *Journal of Statistical Software*. – 2013. – № 3. Том 52. – P. 1-27. – ISSN 1548-7660.
38. Chen, X., Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. / X. Chen, Y. Fan // *Journal of Econometrics*, – 2006. – № 1-2. Том 135. – P. 125-154. – ISSN 0304-4076.
39. Czado, C. Vine Copula Based Modeling / C. Czado, T. Nagler // *Annual Review of Statistics and Its Application*. – 2022. Том 9. – P. 453-477. – ISSN 2326-8298.
40. Deheuvels, P. Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes / P. Deheuvels // *Publications de l'Institut de Statistique de l'Université de Paris*. – 1978. Том 23. – P. 1-36. – ISBN 978-90-481-8401-9.
41. Di Lascio, M.L. Copula-based clustering methods / M.L. Di Lascio, F. Durante, R. Pappadà // *Copulas and Dependence Models with Applications*. – 2019. – P. 49-67. – ISBN 978-3-319-64220-8.
42. Dißmann, J. Selecting and estimating regular vine copulae and application to financial returns / J. Dißmann, E.C. Brechmann, C. Czado, D. Kurowicka // *Computational Statistics & Data Analysis*. – 2013. Том 59. – P. 52-69. – ISSN 0167-9473.
43. Duranton, G. Micro-foundations of urban agglomeration economies / G. Duranton, D. Puga // *Handbook of Regional and Urban Economics*. – 2004. Том 4. – P. 2063-2117. – ISSN 1574-0080

44. Embrechts, P. Correlation and dependence in risk management: properties and pitfalls / P. Embrechts, A. McNeil, D. Straumann // Risk Management: Value at Risk and Beyond. – Cambridge University Press, 2002. – ISBN: 09-521-78180-9
45. Frees, E.W. Understanding relationships using copulas / E.W. Frees, E.A. Valdez // North American Actuarial Journal. – 1998. – № 1. Том 2. – P. 1-25. – ISSN 1092-0277.
46. García-Gómez, C. Copula-based analysis of multivariate dependence patterns between dimensions of poverty in Europe / C. García-Gómez, A. Pérez, M. Prieto-Alaiz // Review of Income and Wealth. – 2020. – № 1. Том 67. – P. 107-131. – ISSN 0034-6586.
47. Genest, C. Statistical inference procedures for bivariate Archimedean copulas / C. Genest, L.-P. Rivest // Journal of the American Statistical Association. – 1993. – № 423. Том 88. – P. 1034-1043. – ISSN 0162-1459.
48. Gudendorf, G. Extreme-value copulas / G. Gudendorf, J. Segers // Copula Theory and Its Applications. – 2010. – P. 127-145. – ISBN 978-3-642-12464-8.
49. Iammarino, S. Regional inequality in Europe: examining the participatory governance dimension / S. Iammarino, A. Rodríguez-Pose, M. Storper // Journal of Economic Geography. – 2018. – № 2. Том 19. – P. 273-293. – ISSN 1468-2702.
50. Kraus, D. D-vine copula based quantile regression / D. Kraus, C. Czado // Computational Statistics & Data Analysis. – 2017. Том 107. – P. 1-15. – ISSN 0167-9473.
51. Li, D.X. On Default Correlation: A Copula Function Approach / D.X. Li // Journal of Fixed Income. – 2000. – № 4. Том 9. – P. 43-54. – ISSN 1059-8596.
52. Masarotto, G. Gaussian copula marginal regression / G. Masarotto, C. Varin // Electronic Journal of Statistics. – 2012. Том 6. – P. 1517-1549. – ISSN 1935-7524.

53. Masarotto, G. Gaussian Copula Regression in R / G. Masarotto, C. Varin // *Journal of Statistical Software*. – 2017. – № 8. Том 77. – P. 1-26. – ISSN 1548-7660.
54. McNeil, A. Multivariate Archimedean copulas, d-monotone functions and L1-norm symmetric distributions / A. McNeil, J. Nešlehova // *Annals of Statistics*. – 2009. – № 5. Том 37. – P. 3059-3097. – ISSN 0090-5364.
55. Messaoud, M. Measuring risk of portfolio: GARCH-copula model / M. Messaoud // *Journal of Economic Integration*. – 2015. – № 1. Том 30. – P. 172-205. – ISSN 1225-651X.
56. Nagler, T. Evading the curse of dimensionality in nonparametric density estimation / T. Nagler, C. Czado // *Journal of Multivariate Analysis*. – 2016. Том 151. – P. 69-89. – ISSN 0047-259X.
57. Parsa, R.A. Copula Regression / R.A. Parsa, S.A. Klugman // *Variance*. – 2011. – № 1. Том 5. – P. 45-61. – ISSN 1554-6509.
58. Patton, A.J. Copula methods for forecasting multivariate time series / A.J. Patton // *Handbook of Economic Forecasting*. – 2007. – № 2 – P. 899-960. – ISSN 1574-0706.
59. Sklar, A. Fonctions de répartition à n dimensions et leurs marges / A. Sklar // *Annales de l'ISUP*. – 1959. – VIII (3) – P. 229-231. – ISSN отсутствует.
60. Zoraghein, H. A spatial population downscaling model for integrated human–environment analysis in the United States / H. Zoraghein, B.C. O'Neill // *Demographic Research*. – 2020. Том 43. – P. 1563-1606. – ISSN 1435-9871.
61. Tamura, K. Effects of the distant population density on spatial patterns of demographic dynamics / K. Tamura, N. Masuda // *Royal Society Open Science*. – 2017. – № 8. Том 4. – P. 1-19. – ISSN 2054-5703.
62. Weeks, J.R. The Role of Spatial Analysis in Demographic Research / J.R. Weeks // *Demographic Research*. – 2001. – № 4. Том 5. – P. 133-156. – ISSN 1435-9871.

63. Wei, Y. Bivariate copula regression models for semi-competing risks / Y. Wei, M. Wojtyś, L. Sorrell, P. Rowe // *Statistical Methods in Medical Research*. – 2023. – № 10. Tom 32. – P. 1902-1918. – ISSN 0962-2802.

64. Wilson, T. Methods for Small Area Population Forecasts: State-of-the-Art and Research Needs / T. Wilson // *Population Research and Policy Review*. – 2021. Tom 40. – P. 787-815. – ISSN 0167-5923.

65. Yang, L. Nonparametric estimation of copula regression models with discrete outcomes / L. Yang, E.W. Frees, Z. Zhang // *Journal of the American Statistical Association*. – 2020. – № 530. Tom 115. – P. 707-720. – ISSN 0162-1459.

**Приложение А**  
(информационное)

**Исходный набор данных по Дальневосточному федеральному округу  
для 2023 года**

Таблица А.1 – Исходный набор данных по Дальневосточному федеральному округу для 2023 года переменных Y, X1– X4

Регион	Y - Темпы прироста населения, в процентах	X1 - ВРП на душу населения, руб	X2 - Среднедушевые располагаемые доходы, руб	X3 - безработица, процент изменения	X4 - Инвестиции на душу населения, руб
Республика Бурятия	-0,31	517 755,80	38 412,00	9,47	172 066,00
Забайкальский край	-0,81	726 780,20	42 758,00	8,77	308 889,00
Республика Саха (Якутия)	0,19	2 230 677,60	69 031,00	8,04	745 842,00
Камчатский край	-0,62	1 523 849,90	87 054,00	3,51	470 642,00
Приморский край	-0,97	1 027 409,80	54 494,00	4,58	222 726,00
Хабаровский край	-0,57	985 369,80	58 092,00	3,28	344 798,00
Амурская область	-0,89	1 054 056,20	54 055,00	4,09	989 914,00
Магаданская область	-0,93	3 017 481,00	111 157,00	4,89	673 113,00
Сахалинская область	-0,91	3 538 862,20	87 902,00	4,13	666 229,00
Еврейская автономная область	-1,27	697 096,50	43 479,00	8,09	211 639,00
Чукотский автономный округ	0,13	3 895 053,50	156 988,00	1,95	1 814 385,00

Источник: [3].

Таблица А.2 – Исходный набор данных по Дальневосточному федеральному округу для 2023 года переменных X5– X9

Регион	X5 - браки, количество на 1000 чел,	X6 - Индекс стоимости жилья	X7 - Индекс потребительских цен, в процентах	X8 - Доля городского населения, в процентах	X9 - Расстояние до Москвы, км
Республика Бурятия	6,00	106,12	112,14	59,10	4 430,00
Забайкальский край	7,60	107,27	112,74	69,50	4 750,00
Республика Саха (Якутия)	5,60	107,34	112,49	67,20	4 900,00
Камчатский край	9,00	99,89	114,29	78,00	6 800,00
Приморский край	8,20	101,93	111,13	78,40	6 430,00
Хабаровский край	7,50	100,41	109,65	83,50	6 160,00
Амурская область	7,80	106,17	112,93	68,40	5 630,00
Магаданская область	7,60	118,01	111,80	96,50	5 920,00
Сахалинская область	8,30	102,32	113,13	82,60	6 660,00
Еврейская автономная область	7,60	134,41	112,98	70,80	6 030,00
Чукотский автономный округ	6,10	107,92	106,27	69,09	6 210,00

Источник: [3].

**Приложение Б**  
(информационное)

**Структура и параметры R-vine-копультной модели**

Таблица Б.1 – Структура и параметры R-vine-копультной модели

Дерево	Ребро	Семейство копулы	Параметры копулы	Коэффициент Кендалла
1	2	3	4	5
1	1	gaussian	0,49	0,32
1	2	tawn	0,3; 0,63; 1,87	-0,16
1	3	tawn	1; 0,87; 3,75	0,65
1	4	bb8	2,03; 0,96	-0,32
1	5	tll	grid_matrix	0,17
1	6	tll	grid_matrix	0,24
1	7	tll	grid_matrix	0,36
1	8	tll	grid_matrix	-0,32
1	9	tll	grid_matrix	0,37
1	10	tawn	0,67; 0,36; 3,94	0,27
1	11	tawn	0,82; 1; 3,61	0,62
1	12	tll	grid_matrix	-0,34
1	13	tawn	1; 0,31; 2,6	-0,24
1	14	tawn	1; 0,6; 7	-0,54
1	15	tawn	1; 0,58; 7	-0,53
1	16	tll	grid_matrix	-0,32
1	17	tawn	0,3; 0,94; 6,52	0,28
2	1	tll	grid_matrix	-0,05
2	2	tawn	0,74; 0,86; 7	0,59
2	3	bb8	1,67; 0,89	-0,18
2	4	tawn	0,3; 1; 3,11	-0,25
2	5	student	0,19; 4,99	0,12
2	6	tll	grid_matrix	0,04
2	7	bb7	1,16; 0,33	-0,2
2	8	tawn	0,72; 0,9; 2,48	0,44
2	9	student	-0,06; 9,84	-0,04
2	10	bb8	2; 0,76	-0,18
2	11	tawn	0,76; 0,3; 1,5	-0,14
2	12	gaussian	-0,38	-0,25
2	13	tll	grid_matrix	-0,16
2	14	tawn	0,31; 1; 1,81	-0,2
2	15	tawn	0,38; 1; 2,15	-0,26
2	16	tawn	0,59; 0,96; 7	0,53
3	1	indep	-	0

Продолжение таблицы Б.1

1	2	3	4	5
3	2	gaussian	-0,12;	-0,08
3	3	bb8	1,37; 1	0,17
3	4	clayton	0,36	0,15
3	5	clayton	0,23	0,1
3	6	clayton	0,23	-0,1
3	7	tll	grid_matrix	0
3	8	tawn	0,3; 0,99; 1,5	0,15
3	9	gumbel	1,19	-0,16
3	10	gaussian	0,1	0,06
3	11	bb8	1,47; 0,98	0,18
3	12	gaussian	-0,5	-0,33
3	13	tll	grid_matrix	0,05
3	14	tawn	1; 0,36; 7	-0,34
3	15	tawn	1; 0,36; 7	-0,33
4	1	tll	grid_matrix	0,02
4	2	clayton	0,16	-0,07
4	3	bb8	1,38; 0,79	0,08
4	4	clayton	0,61	0,23
4	5	gumbel	1,05	-0,05
4	6	student	-0,22; 7,98	-0,14
4	7	indep	-	0
4	8	gaussian	0,12	0,07
4	9	student	0,02; 3,13	0,01
4	10	tawn	0,5; 1; 6,07	-0,45
4	11	frank	-0,65	-0,07
4	12	tll	grid_matrix	0,14
4	13	joe	1,18	0,09
4	14	tll	grid_matrix	0,2
5	1	tll	grid_matrix	-0,08
5	2	student	-0,04; 8,53	-0,03
5	3	frank	-0,43	-0,05
5	4	bb7	1,46; 0,02	-0,21
5	5	gaussian	-0,06	-0,04
5	6	joe	1,19	0,1
5	7	tll	grid_matrix	0,01
5	8	tll	grid_matrix	0,03
5	9	joe	1,14	-0,07
5	10	tll	grid_matrix	0,01
5	11	tll	grid_matrix	0,04
5	12	tawn	0,3; 0,93; 1,5	-0,15
5	13	indep	-	0

Продолжение таблицы Б.1

1	2	3	4	5
6	1	frank	-1,17	-0,13
6	2	tll	grid_matrix	-0,06
6	3	indep	-	0
6	4	tawn	1; 0,3; 2,44	-0,23
6	5	tawn	0,3; 0,46; 1,85	-0,14
6	6	indep	-	0
6	7	tawn	1,; 0,31; 1,51	-0,16
6	8	gumbel	1,1	-0,09
6	9	bb8	2,02; 0,7	0,15
6	10	tll	grid_matrix	-0,06
6	11	bb8	1,16; 1	-0,08
6	12	tawn	0,83; 0,37; 1,66	0,19
7	1	gaussian	-0,06	-0,04
7	2	tll	grid_matrix	-0,12
7	3	gaussian	-0,0	-0,04
7	4	frank	0,8;	0,09
7	5	clayton	0,07	-0,03
7	6	indep	-	0
7	7	tll	grid_matrix	0,08
7	8	tll	grid_matrix	-0,08
7	9	clayton	0,23	-0,1
7	10	joe	1,44	-0,2
7	11	gaussian	0,1	0,07
8	1	student	0,11; 6,71	0,07
8	2	indep	-	0
8	3	frank	-2,56	-0,27
8	4	joe	1,17	-0,09
8	5	joe	1,05	-0,03
8	6	indep	-	0
8	7	joe	1,28	0,14
8	8	gaussian	0,2	0,17
8	9	frank	0,34;	0,04
8	10	tawn	0,31; 1; 2,25	0,23
9	1	bb8	1,72; 0,78	0,14
9	2	joe	1,1	0,06
9	3	gaussian	-0,07	-0,05
9	4	tawn	0,3; 1; 1,5	-0,15
9	5	indep	-	0
9	6	clayton	0,14	-0,06
9	7	tll	grid_matrix	-0,15
9	8	gaussian	0,2	0,13
9	9	clayton	0,16	-0,07

Продолжение таблицы Б.1

1	2	3	4	5
10	1	bb8	1,99; 0,76	-0,18
10	2	clayton	0,06	-0,03
10	3	clayton	0,05	-0,03
10	4	bb8	1,44; 0,74	-0,08
10	5	indep	-	0
10	6	student	-0,08; 6,07	-0,05
10	7	clayton	0,19	0,09
10	8	clayton	0,05	0,02
11	1	joe	1,12	-0,06
11	2	tll	grid_matrix	0,08
11	3	bb8	1,13; 0,94	0,05
11	4	indep	-	0
11	5	gumbel	1,06	-0,05
11	6	indep	-	0
11	7	tll	grid_matrix	0
12	1	joe	1,06	-0,04
12	2	tll	grid_matrix	0,01
12	3	indep	-	0
12	4	joe	1,11	-0,06
12	5	bb8	1,12; 1,	-0,06
12	6	student	-0,01; 14,14	-0,01
13	1	tll	grid_matrix	0
13	2	joe	1,09	0,05
13	3	bb7	1,05; 0,21	-0,12
13	4	indep	-	0
13	5	indep	-	0
14	1	indep	-	0
14	2	tll	grid_matrix	-0,04
14	3	bb8	1,51; 0,79	-0,1
14	4	student	0,01; 11,11	0,01
15	1	joe	1,	0,05
15	2	tll	grid_matrix	-0,1
15	3	gaussian	0,09;	0,05
16	1	tll	grid_matrix	-0,15
16	2	indep	-	0
17	1	clayton	0,15	-0,07

Источник: составлено автором.

**Приложение В**  
(информационное)

**R-vine матрица структуры копульной модели**

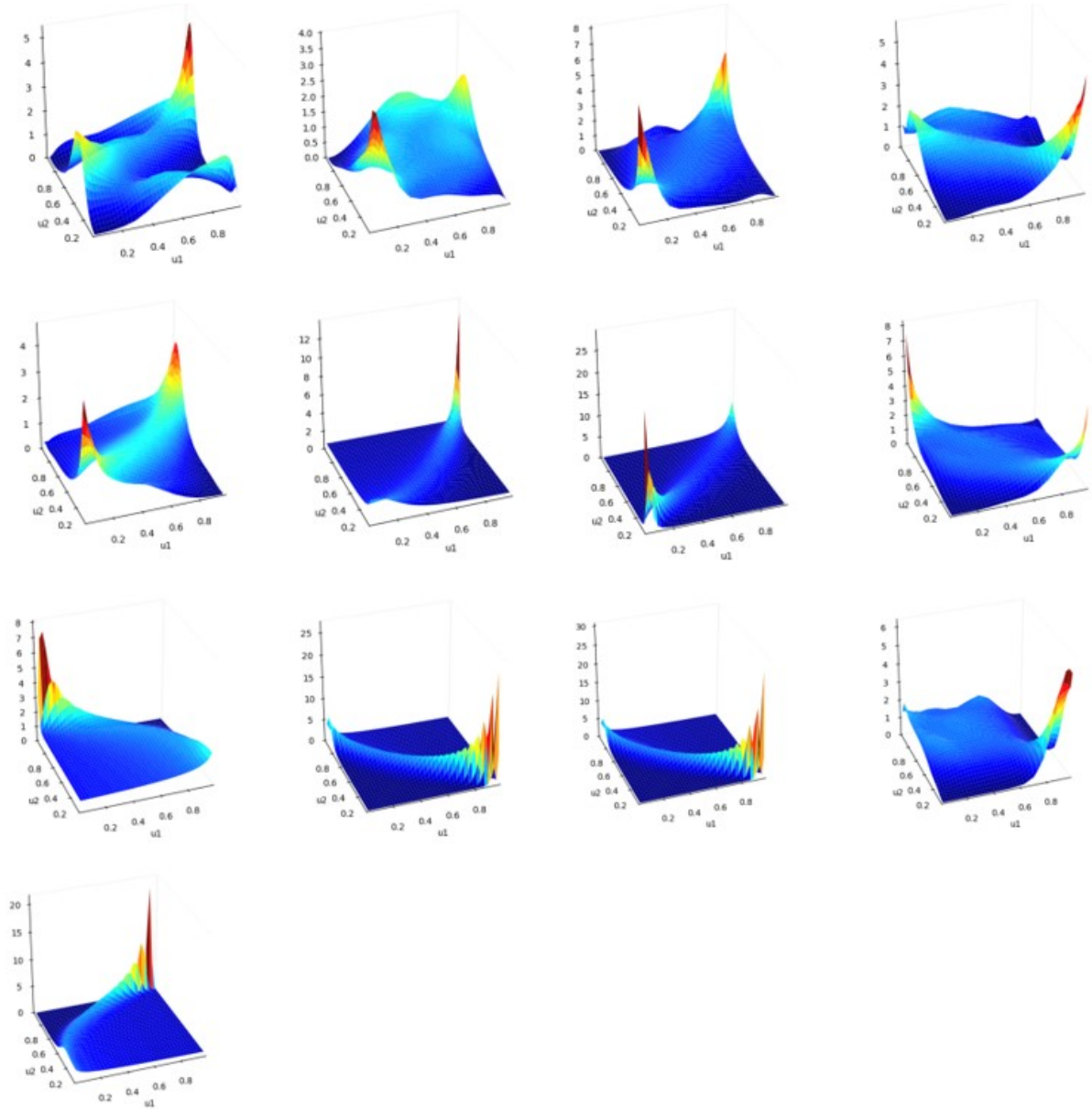
Таблица В.1 – R-vine матрица структуры копульной модели

i \ j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	15	9	2	9	11	6	2	11	3	2	2	4	12	12	12	10	18	18
2	9	15	9	2	6	11	4	3	2	3	4	12	16	10	10	18	10	0
3	14	2	15	4	8	3	3	2	17	4	12	16	10	18	18	12	0	0
4	2	4	4	3	3	2	12	17	4	12	16	10	18	16	16	0	0	0
5	4	3	3	12	2	17	17	4	12	16	10	18	13	13	0	0	0	0
6	3	12	12	17	17	4	11	12	16	10	18	13	4	0	0	0	0	0
7	12	17	17	11	4	12	6	16	10	18	13	2	0	0	0	0	0	0
8	17	5	11	6	12	16	16	10	18	13	3	0	0	0	0	0	0	0
9	5	11	6	16	16	10	10	18	13	17	0	0	0	0	0	0	0	0
10	11	6	16	10	10	9	18	13	11	0	0	0	0	0	0	0	0	0
11	6	16	10	8	9	18	13	6	0	0	0	0	0	0	0	0	0	0
12	16	10	8	18	18	13	9	0	0	0	0	0	0	0	0	0	0	0
13	10	8	18	13	13	8	0	0	0	0	0	0	0	0	0	0	0	0
14	8	18	13	7	7	0	0	0	0	0	0	0	0	0	0	0	0	0
15	18	13	7	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	13	7	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	7	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Источник: составлено автором.

Приложение Г  
(информационное)

R-vine матрица графиков парных копул (дерево  $T_1$ )



Источник: составлено автором.

Рисунок Г.1 – R-vine матрица графиков парных копул (дерево  $T_1$ )