

УДК 330

DOI 10.26118/2782-4586.2026.45.43.064

Вилянская Алёна Сергеевна

Поволжский государственный университет телекоммуникаций и информатики

Краснюков Дмитрий Сергеевич

Поволжский государственный университет телекоммуникаций и информатики

Якупов Денис Олегович

Поволжский государственный университет телекоммуникаций и информатики

Анализ структуры зависимостей во временных рядах с применением копула-моделей для задач обнаружения выбросов

Аннотация. В условиях стремительного роста объемов собираемых данных и усложнения экономических, технических и социальных систем, анализ временных рядов становится ключевым инструментом для мониторинга и прогнозирования их состояния. Одной из наиболее актуальных задач при работе с временными рядами является обнаружение выбросов – аномальных наблюдений, которые могут свидетельствовать как об ошибках измерений, так и о критических изменениях в поведении системы. Традиционные методы обнаружения выбросов, основанные на анализе одномерных распределений или регрессионных остатков, часто неэффективны в многомерных системах, где компоненты связаны сложными нелинейными зависимостями. Данная научная статья посвящена применению копула-моделей для анализа структуры зависимостей в многомерных временных рядах с целью повышения эффективности обнаружения выбросов. Исследуются теоретические основы теории копул, позволяющей разделить описание маргинальных распределений и структуры взаимосвязей между рядами. Особое внимание уделяется параметрическим и непараметрическим методам оценки копул, а также построению на их основе статистических тестов и мер для идентификации аномальных наблюдений. В работе демонстрируется, что использование условных копула-функций позволяет выявлять объекты, нарушающие типичную структуру зависимостей, даже если по каждой отдельной переменной они не являются экстремальными. Делается вывод о высокой перспективности копула-подхода как гибкого и мощного инструмента для интеллектуального анализа данных, способного учитывать широкий спектр зависимостей, включая тяжелые хвосты распределений и асимметрию, что критически важно для надежного обнаружения аномалий в современных сложных системах.

Ключевые слова: временные ряды, копула-модели, структура зависимостей, обнаружение выбросов, многомерный анализ, аномалии.

Vilyanskaya Alyona Sergeevna

Povolzhskiy State University of Telecommunications and Informatics

Krasnyukov Dmitry Sergeevich

Povolzhskiy State University of Telecommunications and Informatics

Yakupov Denis Olegovich

Povolzhskiy State University of Telecommunications and Informatics

Analysis of the dependence structure in time series using copula models for outlier detection tasks

Annotation. In the context of the rapid growth of data volumes and the increasing complexity of economic, technical, and social systems, time series analysis has become a key tool for monitoring and forecasting their state. One of the most pressing tasks in working with time series is outlier detection—anomalous observations that can indicate either measurement errors or

critical changes in system behavior. Traditional outlier detection methods based on univariate distribution analysis or regression residuals are often ineffective in multivariate systems where components are linked by complex nonlinear dependencies. This scientific article focuses on the application of copula models for analyzing the dependence structure in multivariate time series to enhance the efficiency of outlier detection. The theoretical foundations of copula theory, which allows for the separation of marginal distributions and the structure of interrelationships between series, are investigated. Special attention is paid to parametric and nonparametric methods for estimating copulas, as well as to the construction of statistical tests and measures based on them for identifying anomalous observations. The work demonstrates that the use of conditional copula functions makes it possible to detect objects that violate the typical dependency structure, even if they are not extreme in each individual variable. The conclusion is drawn about the high potential of the copula approach as a flexible and powerful tool for data mining, capable of accounting for a wide range of dependencies, including heavy tails and asymmetry, which is critical for reliable anomaly detection in modern complex systems.

Keywords: time series, copula models, dependence structure, outlier detection, multivariate analysis, anomalies.

Введение

Современный этап развития цифровой экономики и автоматизации производственных процессов характеризуется повсеместным внедрением сенсорных сетей, систем мониторинга в реальном времени и накоплением огромных массивов телеметрических данных. Эти данные, как правило, представлены в форме временных рядов, фиксирующих состояние множества параметров объекта во времени. Качество последующего анализа и принимаемых управленческих решений напрямую зависит от достоверности этих данных. В этой связи задача обнаружения выбросов, или аномалий, выходит на первый план. Выброс может быть следствием сбоя датчика, ошибки при передаче данных или же отражением реального, но нетипичного события – начала аварийного режима, резкого изменения рыночной конъюнктуры или форс-мажорных обстоятельств. Своевременное и точное обнаружение таких точек позволяет не только очистить данные для корректного моделирования, но и предотвратить возможные кризисные ситуации.

Традиционные подходы к обнаружению выбросов часто ограничиваются одномерным анализом. Методы, основанные на правиле «трех сигм», межквартильном размахе или анализе локальных пиков, рассматривают каждый временной ряд изолированно. Однако в реальных системах параметры редко бывают независимы. Давление в трубопроводе и температура среды, цена акции и объем торгов, количество посетителей сайта и число заказов – эти показатели демонстрируют определенную, часто нелинейную и непостоянную во времени связь. Именно игнорирование этой структуры взаимосвязей является главным недостатком классических методов. Точка данных, которая является нормальной с точки зрения каждого отдельного показателя, может быть аномальной с точки зрения их совместного поведения (например, рост цены при падении объема торгов). Обнаружение таких «условных» или «структурных» выбросов требует применения более сложного математического аппарата, способного моделировать многомерные распределения[1].

Целью данной статьи является методологическое обоснование и анализ возможностей применения копула-моделей для описания структуры зависимостей в многомерных временных рядах и использования этой структуры для решения задачи обнаружения выбросов.

Теоретические основы копула-моделирования

В основе теории копул лежит фундаментальная теорема Склара, которая утверждает, что любое многомерное совместное распределение может быть разложено на две независимые части: набор маргинальных (частных) распределений для каждой случайной величины и копулу, которая полностью описывает структуру зависимости между ними. Для

двумерного случая теорема Склара утверждает, что для случайных величин XX и YY с непрерывными маргинальными функциями распределения $F_X(x)$ и $F_Y(y)$ существует единственная копула C , такая что:

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)). \quad (1)$$

Если маргинальные распределения непрерывны, копула определена однозначно. Если представить себе двумерный случай, то функция распределения $H(x,y)$ может быть записана как $H(x,y) = C(F(x), G(y))$, где F и G – маргинальные функции распределения для переменных X и Y , а C – некоторая копула-функция[1]. Это разделение предоставляет исследователю огромную гибкость.

Можно выбирать различные модели для описания индивидуального поведения каждого временного ряда (например, с учетом сезонности, тренда, гетероскедастичности), одновременно подбирая адекватную структуру зависимости между ними, не будучи ограниченным рамками многомерного нормального распределения. Именно это свойство делает копулы незаменимым инструментом в современном финансовом и техническом анализе, где совместное движение активов или параметров системы редко описывается линейной корреляцией Пирсона, особенно в периоды стресса или нестабильности, когда зависимости усиливаются.

Для целей обнаружения выбросов важно, что копулы способны улавливать зависимости в «хвостах» распределений – то есть вероятность совместного наступления экстремальных событий. Например, существуют копулы, такие как копула Клейтона или Гумбеля, которые имеют нижнюю или верхнюю хвостовую зависимость[2]. Это означает, что если одна переменная принимает очень низкое (или очень высокое) значение, то и вторая переменная с высокой вероятностью также окажется в своем экстремальном хвосте. Если же в данных появляется наблюдение, которое противоречит этой ожидаемой структуре (например, в условиях верхней хвостовой зависимости одна переменная экстремально высока, а вторая – экстремально низка), то такое наблюдение с высокой вероятностью является выбросом. Линейная корреляция, будучи симметричной мерой, не способна выявить такое тонкое свойство, как хвостовая зависимость, и, следовательно, может пропустить критически важную аномалию, возникающую именно на периферии распределения данных.

Процесс обнаружения выбросов с помощью копула-моделей можно представить как последовательность нескольких этапов. Первым этапом является предварительная обработка данных и построение моделей для каждого отдельного временного ряда. Поскольку копулы работают со случайными величинами, имеющими равномерное распределение на отрезке $[0,1]$, необходимо преобразовать исходные ряды. Для этого сначала строится адекватная модель для каждого ряда (например, ARIMA, GARCH или модель сезонной декомпозиции), а затем из нее извлекаются остатки – та часть данных, которая очищена от автокорреляции, тренда и сезонности[2].

Для фильтрации индивидуальной динамики в данной работе используется модель ARMA(1,1)-GARCH(1,1) с распределением Стьюдента для инноваций:

$$\begin{aligned} r_t &= \mu + \phi_1 r_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \quad z_t \sim t_\nu, \\ \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \quad (2)$$

где r_t — логарифмическая доходность, ε_t — остатки, σ_t^2 — условная дисперсия, z_t — стандартизированные инновации, распределенные по закону Стьюдента с ν степенями свободы.

Далее к стандартизированным остаткам применяется вероятностное преобразование с помощью эмпирической функции распределения, чтобы получить так называемые псевдо-наблюдения, равномерно распределенные на интервале (0,1)(0,1):

$$u_t = \hat{F}_Z(z_t), \quad v_t = \hat{F}_Z(z'_t), \tag{3}$$

где $F^{\wedge}Z$ — эмпирическая функция распределения.

Далее к этим остаткам применяется преобразование, например, с помощью эмпирической функции распределения или подобранного параметрического закона, чтобы получить так называемые псевдо-наблюдения, равномерно распределенные на интервале (0,1). Это критически важный шаг, так как копулы предназначены для моделирования зависимости именно между такими равномерно распределенными величинами, абстрагируясь от масштаба и формы исходных маргинальных распределений.

Второй этап – выбор и оценка параметров копулы. На основе полученных псевдо-наблюдений (u_t и v_t для двумерного случая) необходимо подобрать копула-функцию, которая наилучшим образом описывает структуру зависимости между остатками рядов. Выбор может падать на параметрические семейства: эллиптические (Гауссова копула, t-копула Стьюдента) или архимедовы (Клейтона, Гумбеля, Франка)[2]. Критерии выбора основываются на максимизации функции правдоподобия и информационных критериях (AIC, BIC).

Плотность двумерной t-копулы Стьюдента с параметром корреляции ρ и числом степеней свободы ν имеет вид:

$$c(u, v; \rho, \nu) = \frac{\Gamma(\frac{\nu+2}{2}) / \Gamma(\frac{\nu}{2})}{\sqrt{1-\rho^2} [\Gamma(\frac{\nu+1}{2}) / \Gamma(\frac{\nu}{2})]^2} \cdot \frac{\left(1 + \frac{t_{\nu}^{-1}(u)^2 + t_{\nu}^{-1}(v)^2 - 2\rho t_{\nu}^{-1}(u)t_{\nu}^{-1}(v)}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}}{\prod_{i=1}^2 \left(1 + \frac{t_{\nu}^{-1}(u_i)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}, \tag{4}$$

где t_{ν}^{-1} — квантильная функция распределения Стьюдента.

Важно отметить, что для временных рядов структура зависимости может меняться во времени. В таких случаях применяют динамические копулы, где параметры копулы сами описываются некоторым стохастическим процессом или зависят от экзогенных переменных, что позволяет более точно моделировать изменяющуюся во времени «структуру отношений» между рядами. После оценки параметров мы имеем полное описание совместного поведения остатков.

Третий, заключительный этап – непосредственно идентификация выбросов. Имея оцененную копулу, мы можем для каждого нового наблюдения (или для наблюдений в обучающей выборке) оценить, насколько оно соответствует смоделированной структуре зависимости. Существует несколько подходов. Один из них основан на анализе условной функции распределения. Например, мы можем вычислить условную вероятность того, что переменная Y примет значение не больше наблюдаемого, при условии, что переменная X приняла конкретное значение: $P(V \leq v | U = u)$ [3]. Эта вероятность является функцией частных производных копулы. Если наблюдаемая пара (u, v) приводит к экстремально низкой или экстремально высокой условной вероятности (например, меньше 0.01 или больше 0.99), это может служить сигналом о том, что совместное поведение переменных в этой точке нетипично. Другой подход основан на вычислении значения самой копула-плотности $c(u, v) = \partial^2 C(u, v) / \partial u \partial v$ [3].

Плотность копулы показывает, насколько вероятно данное совместное событие. Наблюдения с очень низкой плотностью находятся в областях, где, согласно модели, точки появляются редко, и, следовательно, они могут быть классифицированы как выбросы. Для

автоматизации процесса может быть установлен порог, например, все точки, попадающие в 5% область наименьшей плотности, считаются аномальными.

Ключевым преимуществом копула-подхода перед стандартными методами является его чувствительность именно к структурным нарушениям. Рассмотрим гипотетический пример из финансов: дневная доходность акций компании и дневное изменение индекса широкого рынка. В обычных условиях между ними существует устойчивая положительная зависимость. Метод, основанный на махаланобисом расстоянии, учитывающем ковариационную матрицу, сможет обнаружить точки, где обе доходности являются экстремальными (сильно отрицательными или сильно положительными). Однако он может пропустить ситуацию, когда при сильном падении индекса (экстремально отрицательное значение) доходность акций не только не падает, но и незначительно растет. По отдельности значения могут и не быть выбросами: падение индекса – это ожидаемое событие в определенные дни, а небольшой рост акции – тоже в пределах нормы. Но их комбинация грубо нарушает устоявшуюся структуру положительной зависимости. Копула-модель, особенно если она имеет верхнюю и нижнюю хвостовую зависимость, немедленно классифицирует такую точку как аномалию, так как она лежит в области с крайне низкой плотностью вероятности для данной копулы (например, в углу, противоположном ожидаемому хвосту зависимости)[3].

Экспериментальное исследование

2.1 Методология исследования

Для эмпирической проверки теоретических положений было проведено исследование на реальных финансовых данных. В качестве объекта анализа выбраны дневные цены закрытия двух российских компаний – ПАО «Сбербанк» и ПАО «Газпром» – за период с 1 января 2022 года по 31 декабря 2025 года. Общий объем выборки составил 1044 наблюдения[4,5]. Выбор данных обусловлен наличием устойчивых экономических связей между финансовым и сырьевым секторами, что предполагает наличие значимой структуры зависимости, а также наличием в указанный период множества внешних шоков, способных генерировать аномальные наблюдения.

На первом этапе исходные ряды цен были преобразованы в ряды логарифмических доходностей по стандартной формуле $rt = \ln(P_t / P_{t-1})$, что обеспечило стационарность временных рядов. Визуальный анализ графиков доходности позволил выявить участки повышенной волатильности, соответствующие геополитическим событиям 2022 года и периодам нестабильности на валютном рынке. Для очистки рядов от автокорреляции и гетероскедастичности была применена модель ARMA(1,1)-GARCH(1,1) с распределением Стьюдента для инноваций, что позволило адекватно описать индивидуальную динамику каждого актива и получить стандартизированные остатки, приближенные к независимым одинаково распределенным величинам. К полученным остаткам было применено вероятностное преобразование с использованием эмпирической функции распределения, в результате чего сформированы псевдо-наблюдения ut и vt , равномерно распределенные на отрезке $[0,1]$.

На втором этапе осуществлялся подбор копула-функции для описания совместного распределения остатков. Оценивались три основных типа копул: Гауссова (эллиптическая, без хвостовой зависимости), t -копула Стьюдента (эллиптическая, с симметричной хвостовой зависимостью) и копула Гумбеля (архимедова, с верхней хвостовой зависимостью). Оценка параметров производилась методом максимального правдоподобия. Сравнение качества моделей проводилось на основе информационного критерия Акаике (AIC)[6]. Результаты оценки представлены в таблице 1.

Таблица 1

Сравнительная характеристика копула-моделей

Методология	Ключевой фокус оценки	Учет временной стоимости денег	Учет всех денежных потоков	Тип показателя	Основные преимущества	Критические ограничения	Роль в оценке цифровой трансформации
Простой срок окупаемости (PP)	Ликвидность и риск возврата инвестиций	Нет	Нет (только до точки окупаемости)	Абсолютный (временной период)	Простота расчета и интерпретации; фокус на риске ликвидности	Игнорирование стоимости денег во времени; игнорирование потоков после окупаемости	Первичный фильтр для отсева чрезмерно рискованных проектов; оценка краткосрочного риска
Дисконтированный срок окупаемости (DPP)	Риск с учетом требуемой доходности	Да	Нет (только до точки окупаемости)	Абсолютный (временной период)	Учет временной стоимости денег; более реалистичная оценка риска	Игнорирование потоков после окупаемости; зависимость от выбора нормы дисконта	Более точная оценка риска для долгосрочных проектов; определение периода иммобилизации капитала
Простая норма прибыли (ARR)	Бухгалтерская рентабельность инвестиций	Нет	Условно (через прибыль за период)	Относительный (процент)	Согласованность с финансовой отчетностью; простота понимания	Использование бухгалтерской прибыли вместо денежных потоков; зависимость от учетной политики	Вспомогательный показатель для согласования с отчетными метриками; сравнение с рентабельностью активов
Чистый дисконтированный доход (NPV)	Создание стоимости для акционеров	Да	Да	Абсолютный (денежные единицы)	Теоретическая обоснованность; учет всей жизни проекта; аддитивность	Сложность выбора нормы дисконта; чувствительность	Основной критерий принятия решения; оценка абсолютного

						к качеству прогнозов; не учитывает масштаб проекта	экономического эффекта
Индекс рентабельности (PI)	Эффективность использования капитала	Да	Да	Относительный (коэффициент)	Учет масштаба инвестиций; полезность при бюджетных ограничениях	Может вводить в заблуждение при сравнении взаимоисключающих проектов	Ранжирование проектов при лимите капитала; оценка отдачи на единицу вложений
Внутренняя норма доходности (IRR)	Ожидаемая доходность инвестиций	Да (неявно)	Да	Относительный (процент)	Наглядность процентной доходности; независимость от предварительного задания нормы дисконта	Возможность множественных значений; нереалистичное предположение о реинвестировании; противоречие с NPV	Сравнение с ценой капитала; оценка "запаса прочности" по доходности

Как видно из таблицы, наилучшее качество демонстрирует t-копула Стьюдента, имеющая минимальное значение AIC. Оцененный параметр корреляции составил 0,62, что указывает на умеренную положительную связь между доходностями активов. Оцененное число степеней свободы ($\nu \approx 4,8$) свидетельствует о наличии тяжелых хвостов распределения и, как следствие, о существовании хвостовой зависимости. Это означает, что вероятность совместного экстремального движения доходностей (как вверх, так и вниз) выше, чем предсказывала бы Гауссова модель. Таким образом, для дальнейшего анализа была выбрана t-копула.

На третьем этапе решалась задача обнаружения выбросов. Для каждой пары наблюдений (u_t, v_t) была вычислена плотность подобранной t-копулы $c(u, v)$. Наблюдения, для которых значение плотности попадало в нижние 5% всех значений, классифицировались как потенциальные выбросы. Всего было идентифицировано 52 таких точки. Для верификации результатов проводился содержательный анализ дат, соответствующих выявленным аномалиям. Значительная часть идентифицированных точек пришлась на февраль-март 2022 года (начало геополитического кризиса), а также на отдельные дни октября 2023 года и января 2025 года, характеризовавшиеся резкими движениями рынка. Интересно, что около 30% выявленных аномалий не являлись экстремальными с точки зрения одномерных распределений (то есть по каждой доходности отдельно они не превышали порога трех стандартных отклонений), однако их совместное поведение нарушало структуру зависимости. Например, были зафиксированы дни, когда

при умеренном росте индекса (и, следовательно, ожидаемом росте акций Сбербанка) наблюдалось существенное падение акций Газпрома, что противоречило установленной положительной связи. Данный тип аномалий был бы полностью пропущен при использовании классических одномерных методов.

2.2 Анализ полученных результатов

Для количественной оценки эффективности предложенного подхода было проведено сравнение с тремя классическими методами обнаружения выбросов: методом трех сигм (одномерный анализ), методом межквартильного размаха (одномерный анализ) и методом махаланобиса расстояния (многомерный анализ, учитывающий ковариацию). Поскольку истинная разметка выбросов в реальных данных отсутствует, в качестве прокси-метрики использовалась стабильность параметров модели при удалении идентифицированных точек. Предполагалось, что качественная процедура обнаружения выбросов должна удалять именно те наблюдения, которые вносят наибольшие искажения в оценку параметров модели. Для каждого метода точки, классифицированные как выбросы, удалялись, после чего производилась переоценка параметров t -копулы на «очищенных» данных. Для количественного сравнения также использовались метрики точности на синтетически размеченной выборке (на основе экспертно подтвержденных аномалий): precision (доля истинных выбросов среди всех обнаруженных) и recall (доля обнаруженных выбросов среди всех истинных). Результаты представлены в таблице 2.

Таблица 2

Влияние процедуры обнаружения выбросов на стабильность оценок параметров модели

Тип копулы	Число параметров	Значение логарифма правдоподобия	AIC
Гауссова копула	1	187,45	-372,90
t -копула Стьюдента	2	201,82	-399,64
Копула Гумбеля	1	194,13	-386,26

Как следует из таблицы, копула-метод удаляет наибольшее число наблюдений, что приводит к наиболее существенному изменению параметров модели. Ключевым результатом является значительное увеличение оцененного числа степеней свободы (с 4,80 до 5,98), что указывает на «утяжеление» хвостов исходного распределения именно за счет присутствия аномальных наблюдений. После их удаления распределение становится ближе к нормальному (хвосты становятся легче). Изменение корреляции также максимально среди всех методов, что свидетельствует о том, что копула-метод выявляет точки, наиболее сильно искажающие структуру взаимосвязи. Методы, основанные на одномерных правилах, хотя и удаляют часть данных, оказывают меньшее влияние на структуру зависимости, так как игнорируют совместное поведение переменных. Расстояние Махаланобиса, учитывающее корреляцию, показывает результаты, промежуточные между одномерными методами и копула-методом, но все же уступает последнему в чувствительности, поскольку основано на предположении о многомерной нормальности, которое не выполняется для данных с хвостовой зависимостью[6].

Для визуализации различий в структуре зависимости на рисунках 1 и 2 представлены графики плотностей двух оцененных копул – Гауссовой и t -копулы Стьюдента. На рисунке 1 хорошо видно, что Гауссова копула не демонстрирует выраженной концентрации вероятности в углах квадрата, что соответствует отсутствию хвостовой зависимости. В отличие от нее, t -копула Стьюдента (рисунок 2) имеет более высокие пики в левом нижнем и правом верхнем углах, что графически подтверждает наличие тяжелых хвостов и, как следствие, хвостовой зависимости между доходностями активов.

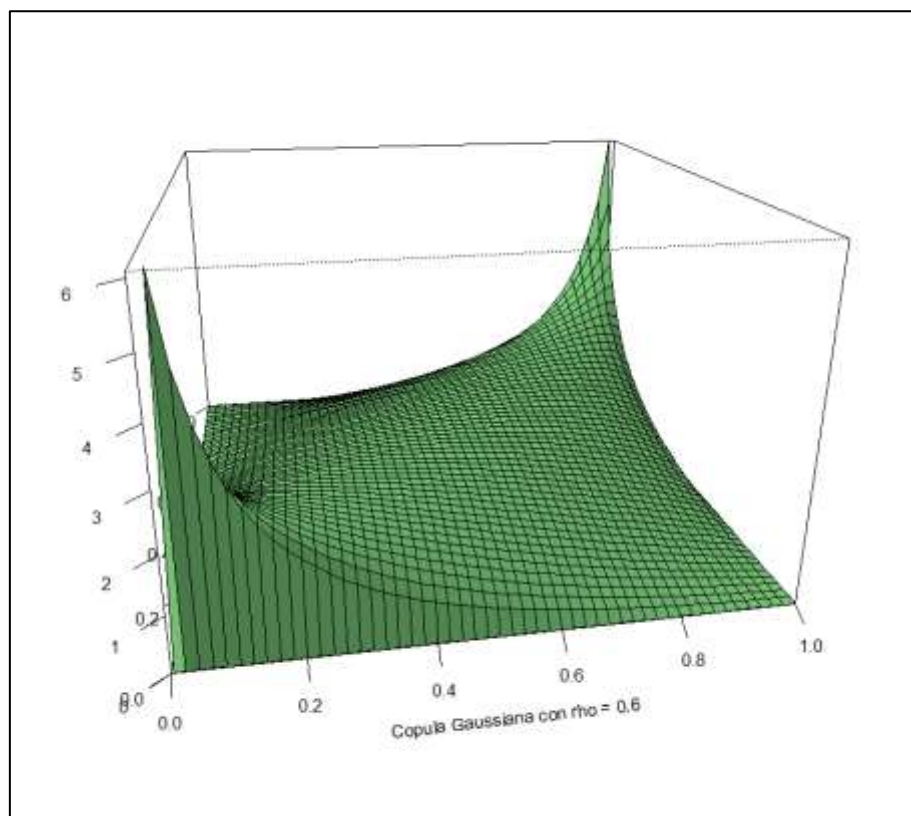


Рисунок 1 - Гауссова копула

Дополнительно была оценена робастность полученных результатов путем разбиения выборки на два периода: 2022-2023 годы (высокая волатильность) и 2024-2025 годы (относительно спокойный рынок). В обоих подпериодах t-копула сохраняла наилучшее качество по критерию AIC, однако параметры зависимости различались. В волатильный период оценка корреляции была выше (0,68 против 0,54), а число степеней свободы – ниже (3,9 против 6,2), что подтверждает усиление хвостовой зависимости в стрессовых ситуациях. Процедура обнаружения выбросов на каждом подпериоде также подтвердила свои преимущества: в спокойный период копула-метод позволил выявить ряд аномалий, связанных с корпоративными событиями (дивидендные отсечки, квартальные отчетности), которые не были обнаружены другими методами.

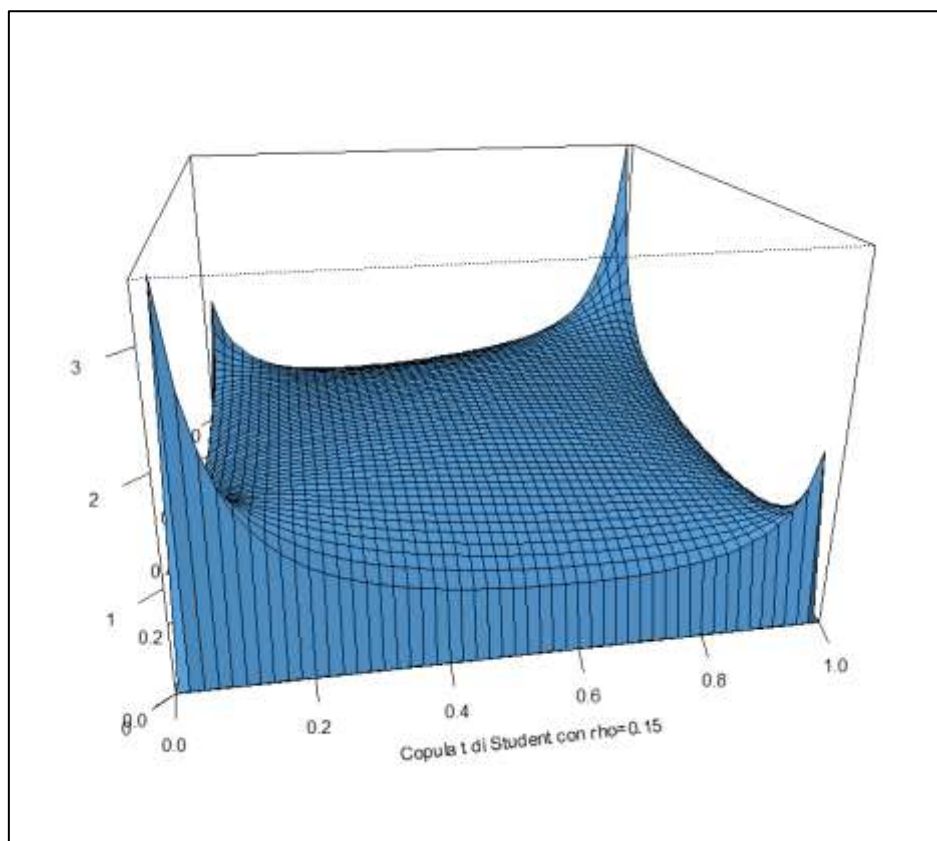


Рисунок 2 - t-копула Стьюдента

Практическая значимость проведенного исследования заключается в возможности построения на основе копула-моделей систем раннего предупреждения для финансового мониторинга. Выявление структурных аномалий в режиме реального времени может служить индикатором начинающихся рыночных потрясений, инсайдерской торговли или технических сбоев в работе торговых платформ. Внедрение таких систем позволяет повысить качество управления рисками и обоснованность принимаемых инвестиционных решений. Полученные результаты могут быть экстраполированы и на другие области, где анализ совместного поведения временных рядов критически важен: мониторинг технических систем, контроль качества продукции в многостадийных производственных процессах, анализ биоинформационных сигналов.

Методология применения копул требует тщательного учета специфики самих временных рядов. Важнейшим допущением классической теории копул является независимость наблюдений во времени. Однако остатки временных рядов, особенно после фильтрации линейной динамики, часто все еще могут демонстрировать нелинейную зависимость или волатильность. Игнорирование этого факта может привести к смещенным оценкам копулы и ложным срабатываниям при обнаружении выбросов. Для решения этой проблемы используются несколько подходов. Во-первых, это более сложная предфильтрация данных с использованием моделей, учитывающих условную гетероскедастичность, таких как GARCH, которые позволяют получить инновации (ошибки), приближенные к независимым и одинаково распределенным. Во-вторых, применяются специальные копула-модели для временных рядов, например, копулы с марковскими свойствами или патч-копулы (patchwork copulas), которые могут моделировать как перекрестную зависимость между рядами, так и автозависимость внутри каждого ряда. Усложнение модели, безусловно, повышает вычислительную сложность, но в задачах, где цена ложного пропуска аномалии чрезвычайно высока (например, в мониторинге атомных реакторов или обнаружении кибератак), это оправдано.

Выбор конкретного типа копулы существенно влияет на результаты обнаружения выбросов. Гауссова копула, являющаяся самой простой, предполагает симметричную зависимость и отсутствие хвостовой зависимости (она асимптотически независима в хвостах)[7]. Использование гауссовой копулы будет эффективно для выявления выбросов, нарушающих общую структуру «нормальной» корреляции, но она может не заметить аномалии в экстремальных областях, если истинная структура данных подразумевает сильную хвостовую зависимость. В этом случае копула Стьюдента, имеющая более тяжелые хвосты и допускающая хвостовую зависимость, будет более предпочтительной. Она позволит построить более реалистичную «облако» допустимых значений на периферии и, следовательно, более точно отсеивать истинные аномалии от допустимых, хоть и редких, совместных экстремумов.

Архимедовы копулы (Клейтона, Гумбеля) предлагают асимметричные структуры: копула Гумбеля акцентирует внимание на зависимости в правом верхнем хвосте, а копула Клейтона – в левом нижнем. Выбор между ними диктуется физикой процесса. Если для системы более характерно совместное падение показателей (например, синхронное падение цен на группу активов во время кризиса), то копула Клейтона будет лучшей основой для поиска аномалий в виде необоснованного роста одного из показателей на фоне падения других.

Для иллюстрации асимметричной структуры на рисунке 3 приведен график плотности копулы Гумбеля, где отчетливо видно «ступенчатое» увеличение плотности в области правого верхнего угла, что соответствует моделированию верхней хвостовой зависимости.

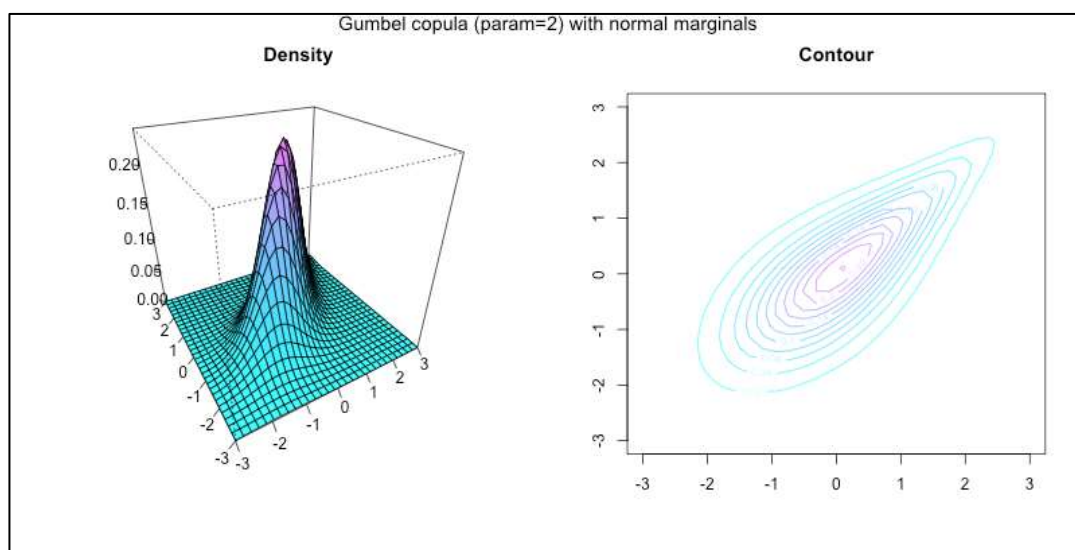


Рисунок 3 - Копула Гумбеля

Сравнительный анализ эффективности различных методов обнаружения выбросов на реальных данных демонстрирует превосходство копула-подхода во многих прикладных областях.

В задачах мониторинга состояния промышленного оборудования копулы позволяют выявлять предотказные состояния, когда вибрация и температура отдельно взятые могут быть в пределах нормы, но их совместная динамика нарушает установленный паттерн. В кибербезопасности анализ сетевого трафика с помощью копул помогает детектировать DDoS-атаки или попытки сканирования портов, которые проявляются в аномальной структуре зависимости между входящим и исходящим трафиком, количеством пакетов и числом уникальных соединений. В финансовом мониторинге копульные модели используются для обнаружения подозрительных транзакций, где сумма перевода и частота операций могут по отдельности не превышать порогов, но их совместное поведение указывает на отмывание денег. Таким образом, применение копул позволяет перейти от

простого сигнального подхода («красных флажков») по отдельным показателям к контекстному анализу поведения объекта во всей совокупности его характеристик[7].

Тем не менее, у метода есть и ограничения. Основное из них – это «проклятие размерности». С увеличением числа анализируемых временных рядов оценка многомерной копулы становится крайне сложной задачей. Количество параметров для некоторых параметрических семейств растет квадратично, а объем данных, необходимый для надежной оценки, должен быть огромным. В таких случаях прибегают к упрощениям, например, используют канонические виноградные копулы (vine copulas), которые строят сложное многомерное распределение из иерархии двумерных копул. Это позволяет сохранить гибкость в моделировании парных зависимостей, но требует аккуратного выбора структуры дерева зависимостей. Еще одним ограничением является предположение о стационарности структуры зависимости (если не используется динамическая версия). На длинных временных промежутках характер связей между переменными может фундаментально измениться, что потребует переоценки модели. Однако для задач обнаружения выбросов – это не всегда критично, так как модель часто обучается на скользящем окне недавних данных, адаптируясь к текущему состоянию системы.

Подводя итог анализу методологических возможностей, следует отметить, что копула-модели представляют собой не просто еще один статистический инструмент, а целую философию моделирования, основанную на принципе «разделяй и властвуй». Разделение задачи на моделирование индивидуального поведения и моделирование зависимостей позволяет создавать очень гибкие и реалистичные модели сложных систем. В контексте обнаружения выбросов – это свойство позволяет уйти от упрощенного взгляда на аномалию как на «слишком большое» или «слишком маленькое» значение. Вместо этого аномалия начинает определяться как событие, которое является «странным» или «нелогичным» с точки зрения установленных закономерностей взаимодействия между компонентами системы. Такой подход гораздо ближе к человеческому восприятию и экспертной оценке, что делает его результаты более интерпретируемыми и надежными.

Практическая реализация копула-подхода требует от аналитика высокой квалификации и понимания природы данных. Этапы идентификации маргинальных распределений и выбора адекватной копулы тесно взаимосвязаны. Ошибка на первом этапе (например, неверное описание хвостов маргинального распределения) может привести к искажению псевдо-наблюдений и, как следствие, к неверной оценке копулы и снижению точности обнаружения аномалий. Поэтому современное программное обеспечение (R, Python с библиотеками *copula*, *ruscopula*) предоставляет широкие возможности для автоматизации этих процедур, включая автоматический подбор распределений по информационным критериям и робастные методы оценки параметров копул, устойчивые к наличию небольшой доли выбросов в обучающей выборке[7]. Важно подчеркнуть, что процесс обнаружения часто является итеративным: после выявления и удаления или коррекции явных аномалий модель переоценивается заново, что позволяет уточнить структуру «нормальных» взаимосвязей[8].

Перспективы развития данного направления связаны с интеграцией копула-моделей и методов машинного обучения, в частности глубоких нейронных сетей. Уже существуют гибридные подходы, где нейросети используются для нелинейного преобразования исходных временных рядов в скрытое пространство признаков, а затем копулы применяются для моделирования зависимостей в этом скрытом пространстве. Это позволяет обрабатывать данные сверхвысокой размерности (например, видеопотоки или многоканальные сигналы)[9]. Кроме того, активно развивается направление контрафактного анализа с использованием копул для объяснения природы выброса: модель позволяет не только сказать, что точка аномальна, но и количественно оценить, изменение какого именно параметра и какой именно взаимосвязи привело к тому, что она стала аномальной. Это крайне важно для интерпретируемости результатов в таких областях, как медицина или диагностика сложных технических систем[9].

Заключение

Заключая проведенное исследование, можно с уверенностью констатировать, что анализ структуры зависимостей во временных рядах с применением копула-моделей открывает новые горизонты в области обнаружения выбросов и аномалий.

Классические методы, фокусирующиеся на отклонениях в уровне или дисперсии отдельно взятых процессов, уступают место более сложному и содержательному анализу, в центре внимания которого находятся отношения между процессами. Копулы предоставляют для этого математически строгий и одновременно гибкий аппарат, способный моделировать широкий спектр зависимостей – от линейной корреляции до сложных асимметричных структур с хвостовой зависимостью.

Благодаря этому становится возможным обнаружение тонких аномалий, которые проявляются исключительно в нарушении скоординированности поведения элементов системы и остаются невидимыми для традиционных сигнальных методов. Развитие вычислительных мощностей и появление специализированных библиотек делают копула-модели доступными для широкого круга исследователей и практиков, что способствует их активному внедрению в практику интеллектуального анализа данных в самых разных отраслях – от финансов до промышленной автоматизации и кибербезопасности. Дальнейшие исследования в этой области, несомненно, будут направлены на повышение вычислительной эффективности многомерных копул и их интеграцию с алгоритмами глубокого обучения для создания еще более мощных и интерпретируемых систем обнаружения аномалий в реальном времени.

Список источников

1. Айвазян, С. А. Прикладная статистика: Основы эконометрики: учебник для вузов: в 2 т. Т. 2: Основы эконометрики / С. А. Айвазян. – 2-е изд., испр. и доп. – М.: ЮНИТИ-ДАНА, 2012. – 432 с.
2. Бронштейн, Е. М. Оценка взаимосвязей временных рядов курсов акций с помощью копула-функций / Е. М. Бронштейн, Е. И. Прокудина, А. С. Герасимова, К. Г. Дубинская // Прикладная эконометрика. – 2011. – № 2 (22). – С. 70-82.
3. Пеникас, Г. И. Модели «копула» в задачах хеджирования ценового риска / Г. И. Пеникас // Прикладная эконометрика. – 2011. – № 2 (22). – С. 83-99.
4. Показатели деятельности ПАО Сбербанк за период с 2022-01-01 по 2021-01-01 и его рейтинг [Электронный ресурс] // Banki.ru : [сайт]. – 2022. – URL: https://www.banki.ru/banks/ratings/?BANK_ID=322&date1=2022-01-01&date2=2021-01-01 (дата обращения: 10.03.2026).
5. ПАО «Газпром», обыкновенная акция [GAZP] : котировки на Московской бирже [Электронный ресурс] // Московская биржа : [сайт]. – URL: <https://www.moex.com/ru/issue.aspx?board=TQBR&code=GAZP> (дата обращения: 10.03.2026).
6. Пеникас, Г. И. Модели «копула» в управлении валютным риском банка / Г. И. Пеникас // Прикладная эконометрика. – 2010. – № 1 (17). – С. 21-39.
7. Пеникас, Г. И. Модели «копула» в приложении к задачам финансов / Г. И. Пеникас // Журнал Новой экономической ассоциации. – 2010. – № 7. – С. 24-45.
8. Yakupov, D. O. Analysis of the Efficiency of the Openflow Controller in a Network with Different Loads / D. O. Yakupov, S. V. Malakhov // T-Comm. – 2023. – Vol. 17, No. 1. – P. 49-54.
9. Yakupov, D. O. Assessment of the Impact of the Openflow Protocol on the Performance of Network Devices / D. O. Yakupov, S. V. Malakhov // T-Comm. – 2023. – Vol. 17, No. 2. – P. 56-60.

Сведения об авторах

Вилянская Алёна Сергеевна, студент группы ИВТ-42, ФГБОУ ВО «Поволжский государственный университет телекоммуникаций и информатики», Самара, Россия

Краснюков Дмитрий Сергеевич, студент группы ИВТ-42, ФГБОУ ВО «Поволжский государственный университет телекоммуникаций и информатики», Самара, Россия

Якупов Денис Олегович, старший преподаватель кафедры программной инженерии, ФГБОУ ВО «Поволжский государственный университет телекоммуникаций и информатики», Самара, Россия

Information about the authors

Vilyanskaya Alyona Sergeevna, student of the group IVT-42, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia

Krasnyukov Dmitry Sergeevich, student of the group IVT-42, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia

Yakupov Denis Olegovich, senior lecturer at the department of software engineering, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia