**Министерство науки и высшего образования Российской Федерации**

федеральное государственное автономное образовательное учреждение   
высшего образования

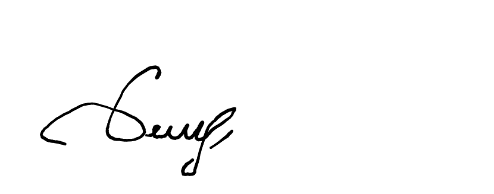
«Национальный исследовательский Томский политехнический университет»

|  |  |
| --- | --- |
| Школа / филиал | ИЯТШ ТПУ |
| Обеспечивающее подразделение | ОЭФ |
| Направление подготовки / специальность | 01.03.02  Прикладная математика и информатика |
| Образовательная программа (направленность (профиль)) | Математические и программные средства исследования операций в экономике;  Математические средства эконофизики |

**КУРСОВАЯ РАБОТА**

**«Моделирование многомерного распределения доходностей СОВОКУПНОСТИ АКЦИЙ предприятий НЕФТЕ-ГАЗОдобывающего СЕКТОРА РОССИИ НА ОСНОВЕ КОПУЛА-ФУНКЦИЙ»**

|  |  |
| --- | --- |
| Выполнил обучающийся | Саматов Денис Сергеевич |
| Группа | 0В01 |



(подпись обучающегося)

**Согласовано:**

Дата проверки    13.12.22

Итоговая оценка\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(традиционная оценка, балл)

         Доцент ОЭФ ТПУ       \_\_\_\_\_\_\_\_\_\_     /Шинкеев М.Л./

Томск – 2022 г.

**Оглавление**

[Задание 3](#_Toc121654442)

[Теоретическая часть 4](#_Toc121654443)

[Практическая часть 5](#_Toc121654444)

[Сбор данных 5](#_Toc121654445)

[Анализ полученных данных 5](#_Toc121654446)

[Заключение 12](#_Toc121654447)

[Приложение A 13](#_Toc121654448)

# **Задание**

**Цель работы:** *Оценить возможность использования копула-функций для описания многомерного распределения доходностей*

**Основные вопросы, подлежащие разработке:**

1. Получить исходные для расчета данные - совокупность цен закрытия акций для предприятий нефтегазодобывающего сектора России, за период с 01.01.2021 по 31.08.2021 с периодичностью 1 день (выбрать 3-5 компаний с наибольшей капитализацией).
2. Подобрать подходящие одномерные распределения для **доходностей (относительных приращений цен акций)** каждой из акций.
3. Определить структуру копула-функции (при использовании иерархической копулы), а также тип(ы) копула-функции для описания многомерного распределения.
4. Оценить параметры копула-функций.
5. На основе полученной копула-функции смоделировать многомерные выборки и проверить гипотезу однородности полученных выборочных распределений и распределения исходного набора данных.

# **Теоретическая часть**

Зависимость между несколькими случайными величинами можно определить через совместную функцию распределения – копулу. Копула – это функция вероятности, а не распределение величин как таковое, графически ее изображают как поверхность, у которой каждая точка равна совместной вероятности двух величин. Иначе говоря – это график плотности совместного распределения. В общем, копулы дают возможность показать любую зависимость нескольких величин друг от друга, расширяя диапазон моделирования.

Копула-функция включает в себя две составляющие: структура зависимости между величинами и маргинальные (частные) распределения для каждой из величин.

Существует несколько способов представления данных в виде копул. В рамках данной работы будут рассмотрены копула-функции, которые довольно часто используются при работе с финансовыми данными: копулы, использующие многомерное распределение Гаусса и многомерное распределение Стьюдента.

Копула Гаусса имеет следующее представление:

|  |  |
| --- | --- |
|  | (1) |

где Ф – функция многомерного распределения Гаусса, Ф-1 – обратная функция маргинального распределения Гаусса, – матрица корреляции, ui = F(x), F – маргинальная функция распределения Гаусса.

Копула Стьюдента выглядит следующим образом:

|  |  |
| --- | --- |
|  | (2) |

где t – функция многомерного распределения Стьюдента, t-1 – обратная функция маргинального распределения Стьюдента, – матрица корреляции, – это число степеней свободы, ui = F(x), F – маргинальная функция распределения Стьюдента.

Существует несколько методов для оценки параметров и построения копулярных моделей: полный параметрический, полупараметрический и непараметрический метод.

Полный параметрический метод включает два этапа: оценка параметров маргинальных распределений и оценка параметров копулы. Полупараметрический предусматривает эти же этапы, но в первом вместо маргинальных рассчитываются эмпирические распределения. В непараметрическом методе на обоих этапах оцениваются эмпирические функции распределения.

В рамках данной работы будет рассмотрен только полный параметрический метод оценки параметров копулярной модели.

# **Практическая часть**

Сбор данных

Используя открытые источники (сайт finam.ru), проведен сбор исходных статистических данных для моделирования многомерного распределения: получены данные цен закрытия акций для предприятий нефтегазодобывающего сектора России: Газпрома, Роснефти, Газпромнефти – , за период с 01.01.2021 по 31.08.2021 с периодичностью 1 день. Было рассмотрено изменение цен акций в течение 170 дней. Графически данных выглядят следующим образом.

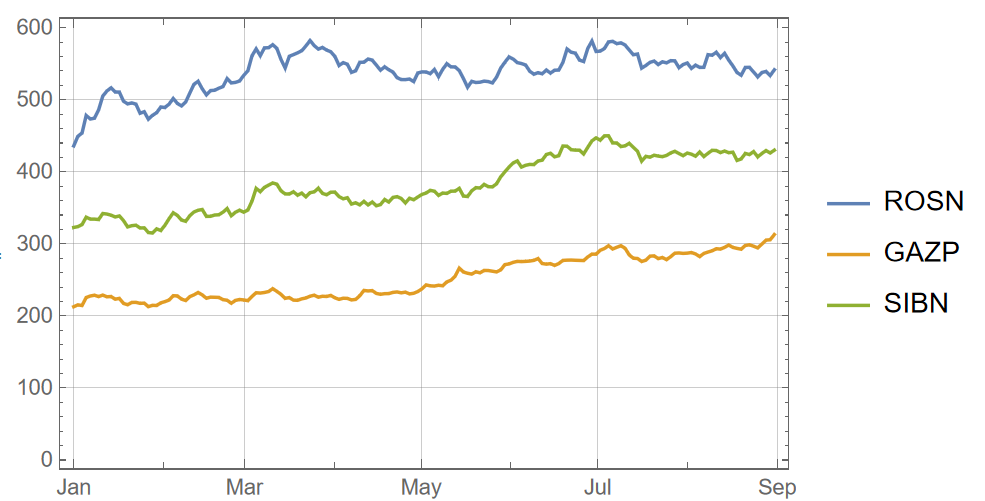


Рисунок 1 – Изменение цен закрытия акция на заданный период времени.

Анализ полученных данных

В первую очередь, имеющиеся исходные данные, представляющие собой временные ряды, необходимо преобразовать в логарифмические доходности. Таким образом, мы получим набор данных, который будем использовать для оценки параметров частных (маргинальных) распределений и параметров копул. Уравнение 3 преобразует ряд дневных цен закрытия активов 𝑝𝑖,*t* в ряд дневных лог-доходностей 𝑟𝑖,*t* для каждого актива 𝑖:

|  |  |
| --- | --- |
|  | (3) |

где i ϵ – количество исследуемых активов, t ϵ – время в днях.

Построим попарное распределение, рис. 2.

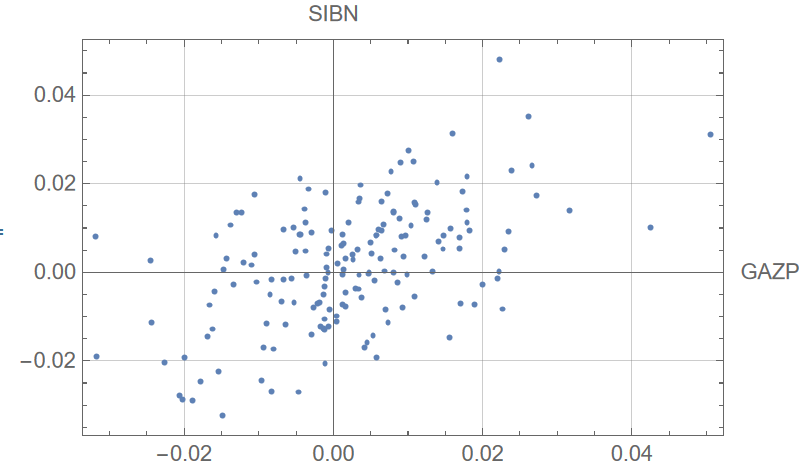
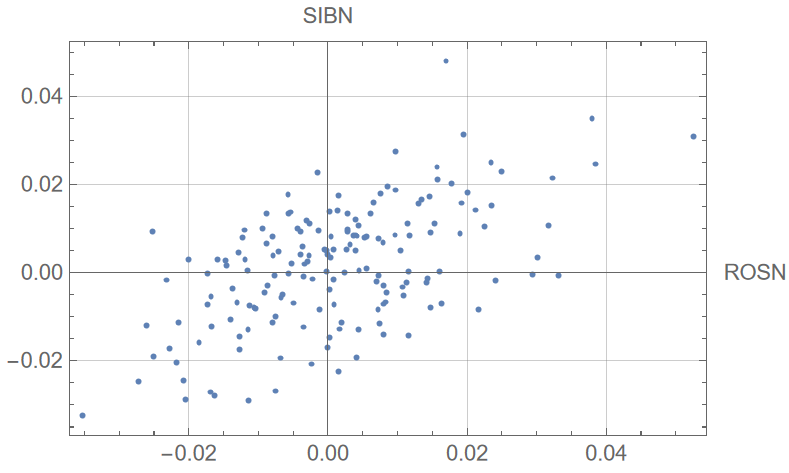
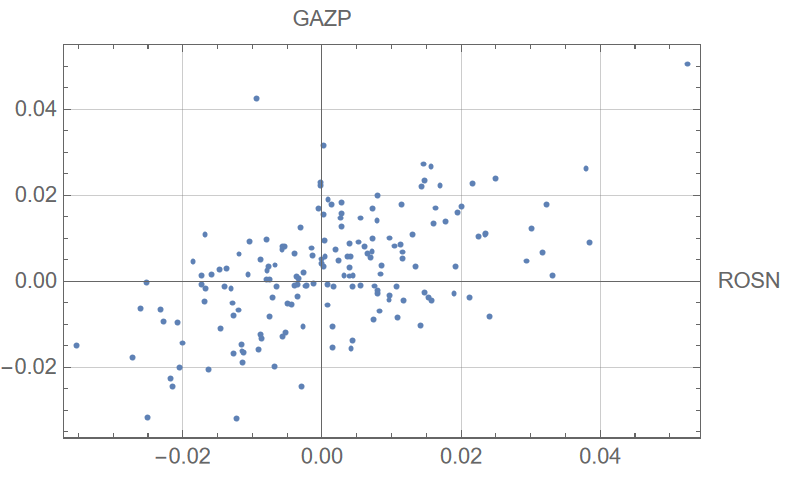


Рисунок 2 – Точечные графики попарных распределений.

На следующем шаге нужно подобрать подходящие маргинальные распределения для логарифмических доходностей акций каждой компании. Для этого следует рассмотреть распределения, которые учитывают асимметрию финансовых временных рядов. В качестве кандидатов были рассмотрены следующие четырех- параметрические распределения:

* Гиперболическое;
* Устойчивое;
* Мейкснера.

Приведем краткое описание указанных распределений. Гиперболическое распределение определяется четырьмя параметрами: ɵ показывает крутизну, 𝜁 – асимметрию, 𝛿 определяет сдвиг и 𝜇 – масштаб. Уравнение 4 описывает плотность гиперболического распределения:

|  |  |
| --- | --- |
|  | (4) |

где К1 – это модифицированная функция Бесселя третьего рода первого порядка, 𝜋 ∈ ℝ, 𝜁 > 0, 𝜇 > 0, 𝛿 ∈ ℝ.

Устойчивое распределение также описывается четырьмя параметрами: 𝛼 определяет эксцесс, 𝜁 — асимметрию, 𝜇 является параметром масштаба, а 𝛿 — сдвига. Характеристическая функция данного распределения имеет следующий вид:

|  |  |
| --- | --- |
|  | (5) |

где 𝛼 ∈ (0; 2], 𝜁 ∈ [−1; 1], > 0, 𝜇 ∈ ℝ, 𝑖 – мнимая единица.

Распределение Мейкснера задаётся следующими четырьмя параметрами: 𝜇 – параметр масштаба, 𝜁 – асимметрии, π – формы, 𝛿 – сдвига. Плотность распределения Мейкснера описывается следующим уравнением:

|  |  |
| --- | --- |
|  | (5) |

где Г – гамма функция, 𝜇 > 0, | 𝜁 | < , π > 0, 𝛿 ∈ ℝ.

Оценим параметры каждого распределения для всех выборок, табл. 1.

Таблица 1 – Оценка параметров маргинальных распределений

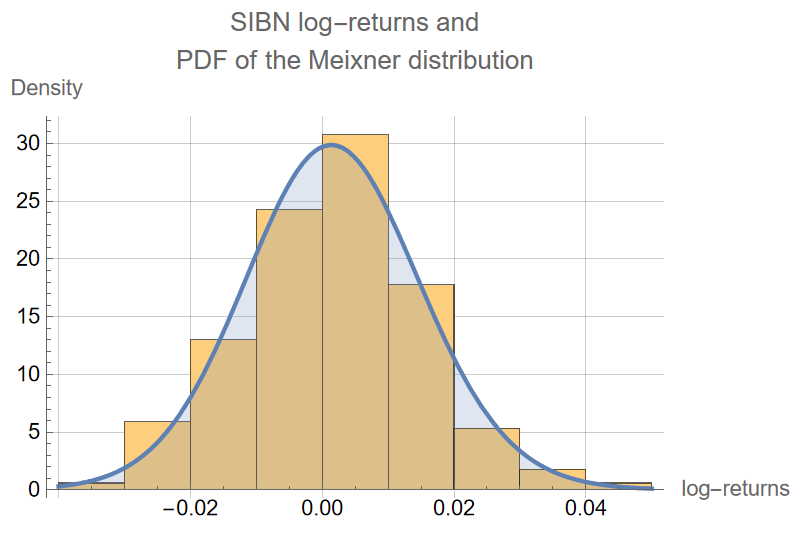
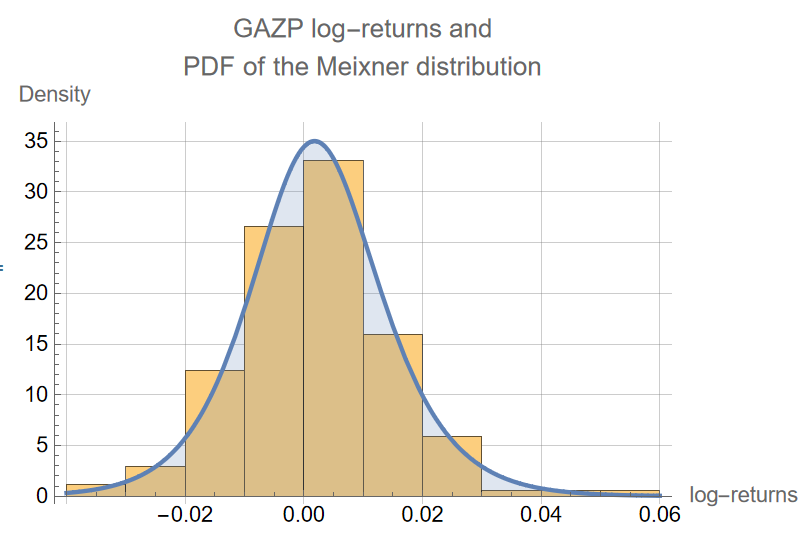
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Распределения и параметры | | GAZP | ROSN | SIBN |
| Гиперболическое распределение | ɵ |  |  |  |
| 𝜁 |  |  |  |
| 𝛿 |  |  |  |
| 𝜇 |  |  |  |
| Устойчивое распределение | 𝛼 |  |  |  |
| 𝜁 |  |  |  |
| 𝜇 |  |  |  |
| 𝛿 |  |  |  |
| Распределение Мейкснера | 𝜇 |  |  |  |
| 𝜁 |  |  |  |
| π |  |  |  |
| 𝛿 |  |  |  |

Для оценки качества полученных параметров будем использовать тесты Колмогорова – Смирнова, а также Крамера – фон Мизеса. В этих тестах будем сравнивать эмпирические наблюдения, т. е. непосредственно реальные логарифмические доходности, со значениями модельных данных полученных с использованием указанных распределений. Результаты тестов приведены в табл. 2.

Таблица 2 – Значения 𝑝-value статистических тестов для распределений

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Тест | Распределение | GAZP | ROSN | SIBN |
| Колмогорова-Смирнова | Гиперболическое |  |  |  |
| Устойчивое |  |  |  |
| Мейкснера |  |  |  |
| Крамера-фон Мизеса | Гиперболическое |  |  |  |
| Устойчивое |  |  |  |
| Мейкснера |  |  |  |

Для выбора одного распределения будем использовать значения 𝑝-value, полученные в результате теста Крамера – фон Мизеса, который является самым мощным статистическим критерием из рассматриваемых в данной работе. По его результатам выберем распределение Мейкснера в качестве маргинального, которое будем использовать в дальнейших расчетах. На рис. 3 приведены графические результаты оценивания параметров распределений. На нем изображены гистограммы лог-доходностей и наложенные на них графики плотности выбранного маргинального распределения Мейкснера.



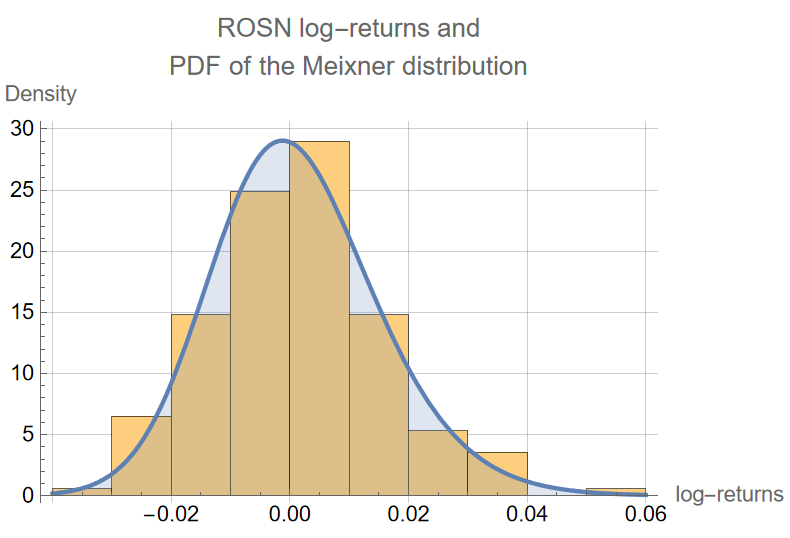


Рисунок 3 – Результаты оценки параметров распределений

Рассмотрим полученные данные на примере двух копулярных моделей, построенных с помощью: Гауссовой копулы, t-копулы Стьюдента.

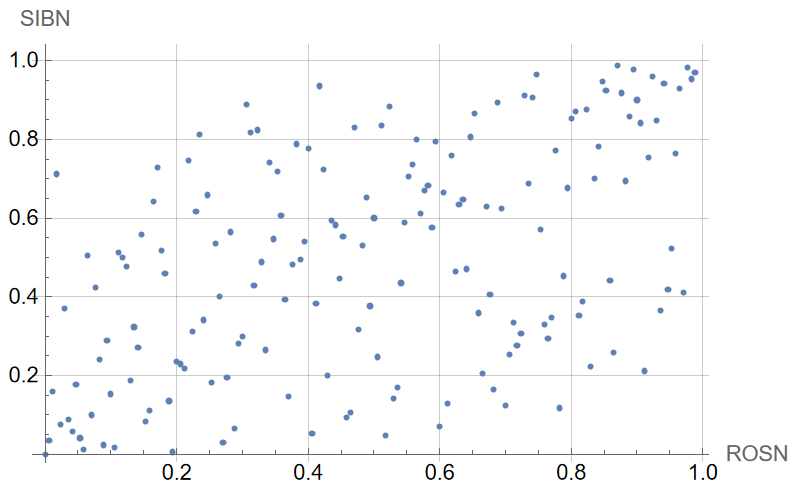
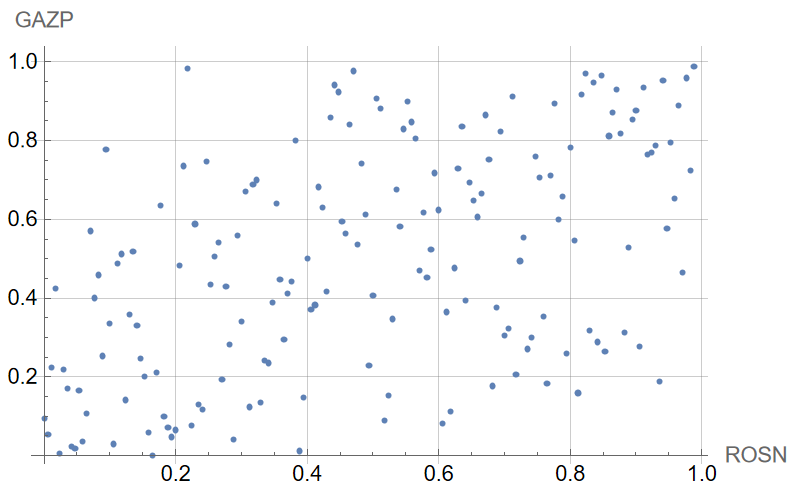
Для оценки параметров копулярных моделей необходимо исходные данные отобразить в пространство области определения копулы, то есть в единичный гиперкуб: ℝn → [0, 1]n .

Параметрами для копулы Гаусса и копулы Стьюдента является корреляционная матрица для псевдо-наблюдений – доходностей, отображённых в единичный гиперкуб. Псевдонаблюдения формируются по следующему правилу:

|  |  |
| --- | --- |
|  | (5) |

где r – логарифмическая доходнось, T – число наблюдений.

На рис. 4 приведено графическое представление псевдо-наблюдений.



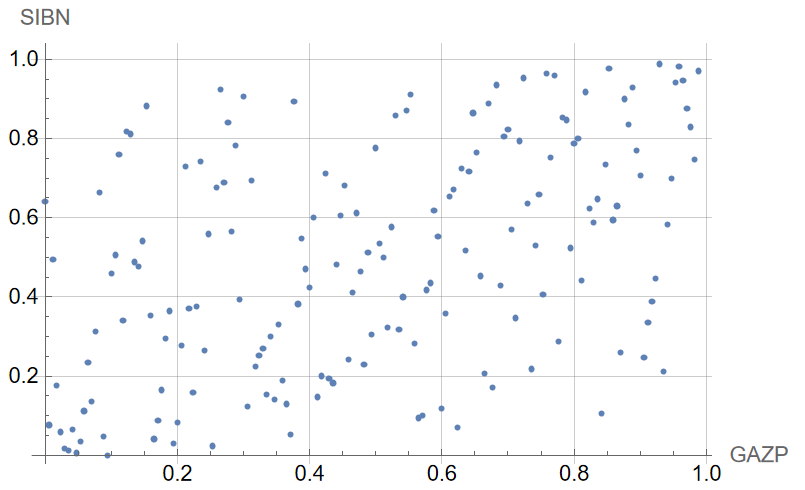


Рисунок 4 – Графики псевдо-наблюдений.

Полученные псевдо-наблюдения далее будем использовать для построения оценок параметров копулярных моделей. Для оценки параметра копул была использована ранговая корреляция Кендалла, поскольку она менее чувствительна к сильным отклонениям, возникающим в финансовых временных рядах.

Коэффициент ранговой корреляции Кендалла вычисляется по следующей формуле:

|  |  |
| --- | --- |
|  | (5) |

где – коэффициент корреляции Пирсона.

Для модели Гауссовой копулы корреляционная матрица Кендалла выглядит следующим образом:

Для копулы Стьюдента корреляционная матрица Кендалла и число степеней свободы v выглядят следующим образом:

Как можно заметить, корреляционные матрицы для обеих копул идентичны. Это происходит потому, что оценка первого параметра многомерной копулы определяется только зависимостью между переменными.

Соответствующие копулярные модели были построены в Wolfram Mathematica с помощью функции CopulaDistribution, аргументами которой является ядро (вид копула-функции с её параметрами) и маргинальные распределения.

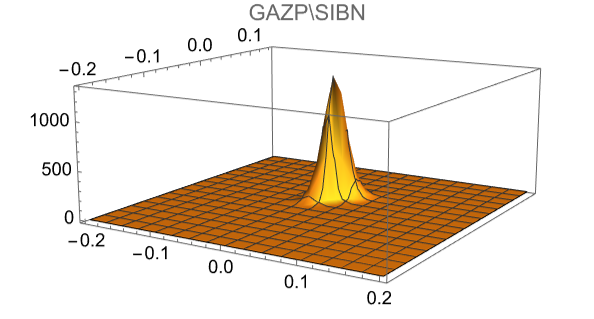
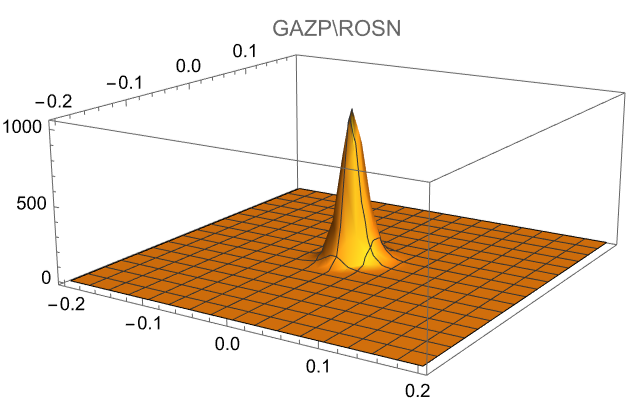
Для оценки адекватности моделей воспользуюсь критериям Крамера-фон Мизеса и Колмогорова-Смирнова. Для этого три раза сгенерирую многомерную выборку случайных значений, распределённых по построенным копулярным моделям. Критерии проверяет гипотезу об однородности двух выборок.

Таблица 3 – Значения 𝑝-value статистических тестов для сгенерированных выборок.

|  |  |  |  |
| --- | --- | --- | --- |
| № генерации | Копул функция | Тест Колмогорова-Смирнова (𝑝 -value) | Тест Крамера-фон Мизеса (𝑝 -value) |
| 1 | Гаусса |  |  |
| Стьюдента |  |  |
| 2 | Гаусса |  |  |
| Стьюдента |  |  |
| 3 | Гаусса |  |  |
| Стьюдента |  |  |

Стоит отметить, что во всех случаях достигнутый уровень значимости позволяет принять гипотезу об однородности выборок, построенных по копулам Гаусса и Стьюдента, и исходных данных. Это говорит о том, что наши логарифмические доходности можно представить с помощью построенных копулярных моделей. Однако нельзя точно сказать, какая из моделей лучшей описывает данные.

Также построим копулярные модели Гаусса и Стьюдента, которые отражают зависимость только двух величин, рис. 5-6.



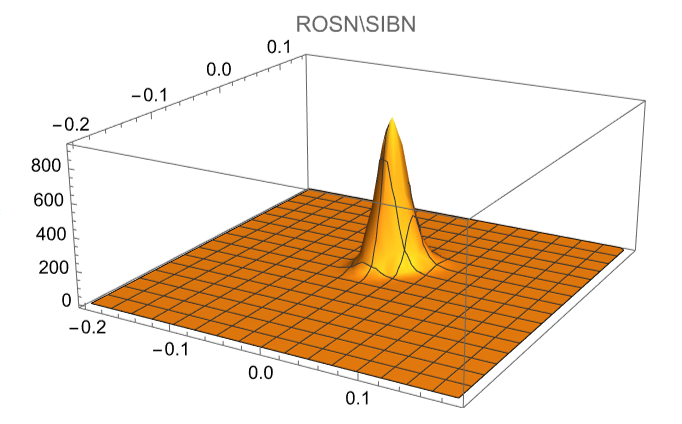
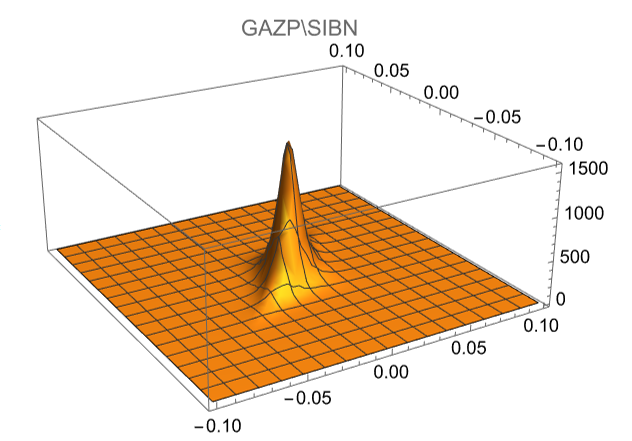
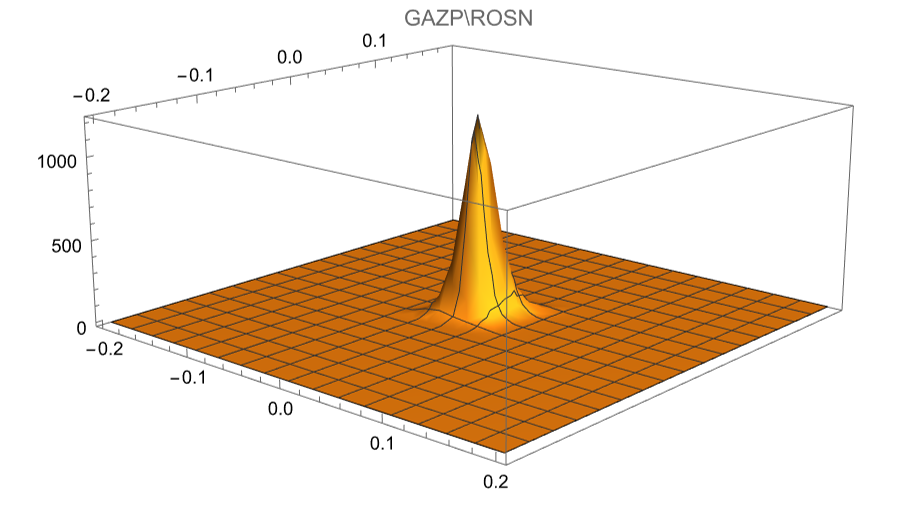


Рисунок 5 – Эмпирические плотности копулярных моделей Гаусса двух величин.



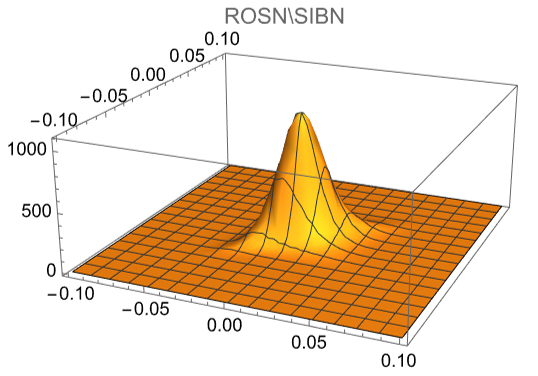


Рисунок 6 – Эмпирические плотности копулярных моделей Стьюденте двух величин.

# **Заключение**

В ходе выполнения работы, cмоделировано многомерное распределение акций для предприятий нефтегазодобывающего сектора России: Газпрома, Роснефти, Газпромнефти – , за определенный период:

* Рассмотрен параметрический метод построения копулярной модели логарифмических доходностей предприятий;
* Были оценены параметры маргинальных распределений для каждой акции;
* Построены две копулярные модели: на основе многомерного распределения Гаусса и на основе многомерного распределения Стьюдента;
* Проверена значимость данных моделей на основе сгенерированной многомерной выборки.

# **Приложение A**

LogData = Flatten[Import["import\_data\_cupola.xlsx"],1];

data=Flatten[Import["import\_hist\_data.xlsx"],1];

rosn=Transpose[data][[1]];

gazp=Transpose[data][[2]];

sibn=Transpose[data][[3]];

ROSN=Transpose[LogData][[1]];

GAZP=Transpose[LogData][[2]];

SIBN=Transpose[LogData][[3]];

DateListPlot[{rosn,gazp,sibn},{{2021,1,1},{2021,8,31}},PlotRange->All, GridLines->Automatic,Frame->True,PlotLegends->{"ROSN","GAZP","SIBN"}]

unionROSNGAZP=Table[{ROSN[[i]],GAZP[[i]]},{i,Length[GAZP]}];

ListPlot[unionROSNGAZP,PlotRange->All, GridLines->Automatic,Frame->True,AxesLabel->{"ROSN","GAZP"}]

unionROSNSIBN=Table[{ROSN[[i]],SIBN[[i]]},{i,Length[GAZP]}];

ListPlot[unionROSNSIBN,PlotRange->All, GridLines->Automatic,Frame->True,AxesLabel->{"ROSN","SIBN"}]

unionGAZPSIBN=Table[{GAZP[[i]],SIBN[[i]]},{i,Length[GAZP]}];

ListPlot[unionGAZPSIBN,PlotRange->All, GridLines->Automatic,Frame->True,AxesLabel->{"GAZP","SIBN"}]

T=Length[GAZP];

psROSN={};

psGAZP={};

psSIBN={};

For[i=1,i<=T,i++,

f = 1/(1+T)\*Sum[If[ROSN[[j]]<ROSN[[i]],1,0],{j,T}];

AppendTo[psROSN,f]

]

For[i=1,i<=T,i++,

f = 1/(1+T)\*Sum[If[GAZP[[j]]<GAZP[[i]],1,0],{j,T}];

AppendTo[psGAZP,f]

]

For[i=1,i<=T,i++,

f = 1/(1+T)\*Sum[If[SIBN[[j]]<SIBN[[i]],1,0],{j,T}];

AppendTo[psSIBN,f]

]

unionpsROSNpsGAZP=Table[{psROSN[[i]],psGAZP[[i]]},{i,Length[GAZP]}];

ListPlot[unionpsROSNpsGAZP,PlotRange->All, GridLines->Automatic,AxesLabel->{"ROSN","GAZP"}]

unionpsROSNpsSIBN=Table[{psROSN[[i]],psSIBN[[i]]},{i,Length[GAZP]}];

ListPlot[unionpsROSNpsSIBN,PlotRange->All, GridLines->Automatic,AxesLabel->{"ROSN","SIBN"}]

unionpsGAZPpsSIBN=Table[{psGAZP[[i]],psSIBN[[i]]},{i,Length[GAZP]}];

ListPlot[unionpsGAZPpsSIBN,PlotRange->All, GridLines->Automatic,AxesLabel->{"GAZP","SIBN"}]

paramsForGAZP1=FindDistributionParameters[GAZP,HyperbolicDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

paramsForROSN1=FindDistributionParameters[ROSN,HyperbolicDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

paramsForSIBN1=FindDistributionParameters[SIBN,HyperbolicDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

CramerVonMisesTest[GAZP,HyperbolicDistribution[θ,ξ,δ,μ]/.paramsForGAZP1]

CramerVonMisesTest[ROSN,HyperbolicDistribution[θ,ξ,δ,μ]/.paramsForROSN1]

CramerVonMisesTest[SIBN,HyperbolicDistribution[θ,ξ,δ,μ]/.paramsForSIBN1]

KolmogorovSmirnovTest[GAZP,HyperbolicDistribution[θ,ξ,δ,μ]/.paramsForGAZP1]

KolmogorovSmirnovTest[ROSN,HyperbolicDistribution[θ,ξ,δ,μ]/.paramsForROSN1]

KolmogorovSmirnovTest[SIBN,HyperbolicDistribution[θ,ξ,δ,μ]/.paramsForSIBN1]

paramsForGAZP2=FindDistributionParameters[GAZP,StableDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

paramsForROSN2=FindDistributionParameters[ROSN,StableDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

paramsForSIBN2=FindDistributionParameters[SIBN,StableDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

CramerVonMisesTest[GAZP,StableDistribution[θ,ξ,δ,μ]/.paramsForGAZP2]

CramerVonMisesTest[ROSN,StableDistribution[θ,ξ,δ,μ]/.paramsForROSN2]

CramerVonMisesTest[SIBN,StableDistribution[θ,ξ,δ,μ]/.paramsForSIBN2]

KolmogorovSmirnovTest[GAZP,StableDistribution[θ,ξ,δ,μ]/.paramsForGAZP2]

KolmogorovSmirnovTest[ROSN,StableDistribution[θ,ξ,δ,μ]/.paramsForROSN2]

KolmogorovSmirnovTest[SIBN,StableDistribution[θ,ξ,δ,μ]/.paramsForSIBN2]

paramsForGAZP3=FindDistributionParameters[GAZP,MeixnerDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

paramsForROSN3=FindDistributionParameters[ROSN,MeixnerDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

paramsForSIBN3=FindDistributionParameters[SIBN,MeixnerDistribution[θ,ξ,δ,μ],ParameterEstimator->"MaximumLikelihood"]

CramerVonMisesTest[GAZP,MeixnerDistribution[θ,ξ,δ,μ]/.paramsForGAZP3]

CramerVonMisesTest[ROSN,MeixnerDistribution[θ,ξ,δ,μ]/.paramsForROSN3]

CramerVonMisesTest[SIBN,MeixnerDistribution[θ,ξ,δ,μ]/.paramsForSIBN3]

KolmogorovSmirnovTest[GAZP,MeixnerDistribution[θ,ξ,δ,μ]/.paramsForGAZP3]

KolmogorovSmirnovTest[ROSN,MeixnerDistribution[θ,ξ,δ,μ]/.paramsForROSN3]

KolmogorovSmirnovTest[SIBN,MeixnerDistribution[θ,ξ,δ,μ]/.paramsForSIBN3]

gazpDistribution=MeixnerDistribution[θ,ξ,δ,μ]/.paramsForGAZP3;

sibnDistribution=MeixnerDistribution[θ,ξ,δ,μ]/.paramsForSIBN3;

rosnDistribution=MeixnerDistribution[θ,ξ,δ,μ]/.paramsForROSN3;

paramsModel = KendallTau[Transpose[{psGAZP, psROSN, psSIBN}]]

GaussCopula =

CopulaDistribution[{"Multinormal", paramsModel}, {gazpDistribution,

rosnDistribution, sibnDistribution}];

StudentCopula =

CopulaDistribution[{"MultivariateT", paramsModel,

3}, {gazpDistribution, rosnDistribution, sibnDistribution}];

sampleGauss=RandomVariate[GaussCopula,500];

sampleStudent=RandomVariate[StudentCopula,500];

KolmogorovSmirnovTest[LogData,sampleGauss]

KolmogorovSmirnovTest[LogData,sampleStudent]

CramerVonMisesTest[LogData,sampleGauss]

CramerVonMisesTest[LogData,sampleStudent]

sampleGauss=RandomVariate[GaussCopula,500];

sampleStudent=RandomVariate[StudentCopula,500];

KolmogorovSmirnovTest[LogData,sampleGauss]

KolmogorovSmirnovTest[LogData,sampleStudent]

CramerVonMisesTest[LogData,sampleGauss]

CramerVonMisesTest[LogData,sampleStudent]

sampleGauss=RandomVariate[GaussCopula,500];

sampleStudent=RandomVariate[StudentCopula,500];

KolmogorovSmirnovTest[LogData,sampleGauss]

KolmogorovSmirnovTest[LogData,sampleStudent]

CramerVonMisesTest[LogData,sampleGauss]

CramerVonMisesTest[LogData,sampleStudent]