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

Кафедра «Теория вероятностей и математическая статистика»

Курсовая работа по математической статистике

на тему «**Проверка гипотезы о нормальном распределении логарифмической доходности по критерию Харке-Бера**»

Вид исследуемых данных: «**Котировки акций компаний Акрон, РусГидро, Магнит, Роснефть, Сбербанк**»

Выполнил:

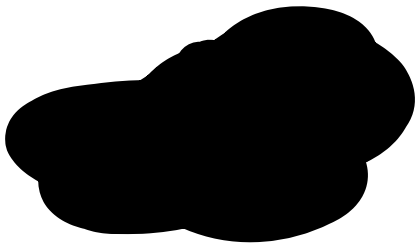
студент группы ПМ18-2,

Поздняков А.Р.

Научный руководитель:

доцент, к.ф.-м.н

Пяткина А.В.



Москва 2020

Введение

В этой работе производится проверка гипотезы о нормальном распределении логарифмической доходности курсов акций ведущих отечественных компаний за каждый год с 2010 по 2019 по критерию Харке-Бера. Цель данной курсовой работы – оценить, на сколько гипотеза о нормальном распределении логарифмических доходностей соотносится с реальностью. В наши дни нормальный закон распределения доходности активов актуален, так как является основой множества экономических моделей.

Данная работа состоит из двух частей, теоретической и практической. В ходе работы будет произведен предварительный анализ данных, дана теоретическая справка по проверке гипотез, проверена гипотеза для модельных и реальных данных.

В работе я буду использовать язык Python для проведения наглядных вычислений, нужных для исследования.

Предварительный анализ

Для рассмотрения были взяты котировки акций за период с 01.01.2010 по 31.12.2014 двенадцати компаний, входящих в состав МосБрижи. Отслеживаем динамику цен у основных отечественных нефтедобывающих компаний:

1. AKRN (Акрон)
2. HYDR (Русгидро)
3. MGNT(Магнит)
4. ROSN (Роснефть)
5. SBER (Сбербанк)

Исследуемый период с начала 2010 года по конец 2019. Данные по котировкам и ценам получены с сайта www.finam.ru

В предварительном анализе исследуем следующие показатели. Использованы поля: Adj CLOSE - скорректированная цена закрытия, единицы измерения – RUB.

Таб. 1. Таблица кол-ва торговых дней.

<YEAR>	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
<TICKER>										
AKRN	248	248	255	250	250	250	252	252	254	252
HYDR	248	248	255	250	250	250	252	252	254	252
MGNT	248	248	255	250	250	250	252	252	254	252
ROSN	248	248	255	250	250	250	252	252	254	252
SBER	248	248	255	250	250	250	252	252	254	252

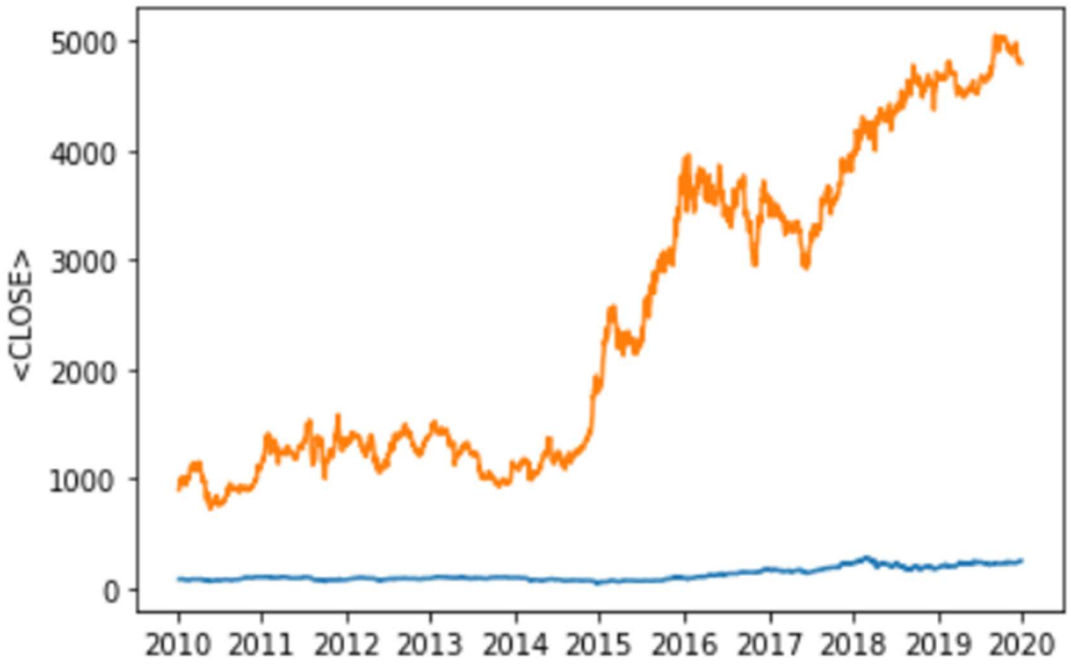
В дальнейшем анализе будут использованы акции компаний имеющих более 200 торговых дней в каждом году исследуемого периода. На основе полученных данных видно, что нам подходят для исследования тикеры всех компаний.

Таб. 2. Таблицы максимальных (вверх и вниз) дневных относительных скачков цен в процентах (по годам и акциям).

<YEAR>	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
<TICKER>										
AKRN	7.361963	9.565619	5.692438	9.138004	8.251748	6.922218	8.168713	5.335731	6.302202	3.285421
HYDR	7.633098	7.990809	7.672209	8.762366	10.286195	11.865028	8.167234	7.733620	6.137184	4.364326
MGNT	11.735805	9.443326	8.193417	7.142857	9.145431	9.726524	7.859889	9.714889	10.095764	4.742547
ROSN	7.736047	6.562068	6.113575	4.045664	4.818139	6.384558	5.586429	4.015657	5.935930	4.080641
SBER	10.284022	8.724544	7.229885	4.326316	10.843819	11.544992	5.750560	5.010737	16.503332	3.793581

Минимальный скачок – 3,28% AKRN, максимальный скачок 16,5% SBER. Построим графики цен для акций с максимальными и минимальными скачками.

Рис. 1. График изменения цен акций SBER и AKRN за всё время



Теоретическая справка

Математическая статистика — наука о математических методах анализа данных, полученных при проведении массовых наблюдений (измерений, опытов). В зависимости от математической природы конкретных результатов наблюдений статистика математическая делится на статистику чисел, многомерный статистический анализ, анализ функций (процессов) и временных рядов, статистику объектов нечисловой природы. Существенная часть статистики математической основана на вероятностных моделях. Выделяют общие задачи описания данных, оценивания и проверки гипотез.

Нормальный закон распределения

Нормальный закон распределения - наиболее часто встречающийся на практике закон распределения. Эту особенность можно объяснить тем, что этот закон проявляется во всех случаях, когда случайная величина является результатом действия большого числа различных факторов. все остальные законы распределения приближаются к нормальному.

Частный случай нормального закона – стандартный нормальный закон $N(0,1)$ с параметрами $\mu = 0$ и $\sigma^2 = 1$, где μ и σ^2 - параметры распределения.

Если $X \sim N(\mu, \sigma^2)$, то функция распределения X:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

Эту функцию называют функцией Лапласа, а случайную величину, обладающую данной функцией распределения – нормальной случайной величиной.

Распределение «хи-квадрат»

Распределение χ^2 (хи-квадрат) с n степенями свободы — это распределение суммы квадратов n независимых стандартных нормальных случайных величин. Пусть $X_1,...,X_n \sim iidN(0,1)$, тогда $\chi^2 = \sum_{i=1}^{n} X_i^2$ распределена по закону χ^2 с $k = n$ степенями свободы. С увеличением числа степеней свободы распределение медленно приближается к нормальному.

Статистическая гипотеза

Статистическая гипотеза – это некоторое предположение о виде известного распределения, о параметрах известных распределений, об отношениях между случайными величинами и т.д. Статистическая гипотеза называется параметрической, если она основана на предположении, что генеральное распределение известно с точностью до конечного числа параметров.

Параметрическая гипотеза называется простой, если она имеет вид $\theta = \theta_0$, где θ_0 - некоторое фиксированное значение параметра θ. Гипотеза вида $\theta \in \Theta$, где Θ - какое либо множество, содержащее по меньшей мере два различных элемента, называется сложной.

Пусть H_0 и H_1 - две взаимоисключающие статистические гипотезы. Где H_0 - основная гипотеза, а H_1 - альтернативная.

Статистическим критерием с критической областью K называется правило, в соответствии с которым H_0 отвергается, если выборка $(x_1,...,x_n) \in K$, и принимается, если $(x_1,...,x_n) \notin K$.

Как правило критическая область задается при помощи неравенства: $K = \{(x_1,...,x_n) \in R^n : t > c\}$ или $K = \{(x_1,...,x_n) \in R^n : t < c\}$ или $K = \{(x_1,...,x_n) \in R^n : t < c_1\} \cup \{(x_1,...,x_n) \in R^n : t > c_2\}$, где $c, c_1, c_2 (c_2 > c_1) = const$, $t = t(x_1,...,x_n)$ - статистика критерия.

Применение статистического критерия может привести к ошибкам двух различных типов. В одном отвергается верная гипотеза H_0, в другом отвергается верная гипотеза. H_1.

Вероятность ошибки первого рода называется уровнем значимости критерия и обозначается α. Вероятность ошибки второго рода обозначается β. Величина $1-\beta$ называется мощностью критерия.

Критерий Харке-Бера

Критерий Харке-Бера используется для проверки гипотезы о том, что исследуемая выборка x_s является выборкой нормально распределенной случайной величины с неизвестным математическим ожиданием и дисперсией. Как правило, этот критерий

применяется перед тем, как использовать методы параметрической статистики, требующие нормальности исследуемых случайных величин.

Для проверки случайной величины на нормальность используется тот факт, что у нормального распределения коэффициент асимметрии и эксцесс равны нолю - отклонение этих величин от нулевого значения может служить мерой отклонения распределения от нормального. На основе выборки строится статистика Харке-Бера

$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$, где: n - размер выборки, $S = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\left(\sigma^2\right)^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3\right)^{3/2}}$

; $K = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{\left(\sigma^2\right)^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right)^2}$.

В нашем случае мы выдвигаем гипотезу H_0, что выборка распределена нормально. Если гипотеза H_0 верна, то статистика критерия имеет асимптотическое распределение χ^2 с 2 степенями свободы. Чтобы высчитать нужную нам критическую точку воспользуемся аппроксимацией Голдштейна получения квантиля распределения χ^2 приближенно.

Согласно данной аппроксимации, чтобы получить квантиль χ^2 с n степенями свободы и уровнем значения α:

$$\chi^2_{\alpha,n} = n \cdot \left[\sum_{i=0}^{6} n^{-\frac{i}{2}} \cdot d \cdot \left(a_i + \frac{b_i}{n} + \frac{c_i}{n^2}\right)\right]^3$$

, где коэффициенты a, b и c берутся из таблицы

a	b	c
1.0000886	-0.2237368	-0.01513904
0.4713941	0.02607083	-0.008986007
0.0001348028	0.01128186	0.02277679
-0.008553069	-0.01153761	-0.01323293
0.00312558	0.005169654	-0.006950356
-0.0008426812	0.00253001	0.001060438

А d определяется по формуле $\mathrm{d} = -2.0637 \cdot \left(\ln\frac{1}{\alpha} - 0.16\right)^{0.4274} + 1.5774$, при α между 0.001 и 0.5, включая левую границу.

Тогда наше критическое значение для нулевой гипотезы будет равно 5.991. Если $JB > 5.991$, то гипотеза H_0 о нормальном распределении отклоняется, т.е. распределение не является нормальным, и принимается гипотеза H_1. Критической областью гипотезы H_0 является $(5.991; +\infty)$. Эту область значений можно сформулировать, как область соответствующих P-значений (0.05, 0), что мы и будем использовать в наших вычислениях далее.

Волатильность

Волатильность (Изменчивость, англ. Volatility) — это статистический показатель, описывающий тенденцию рыночной цены или дохода во времени. Учитывается в управлении финансовыми рисками, где представляет собой меру риска использования финансового инструмента за промежуток времени. Волатильность цены акции в широком

смысле «представляет собой меру неопределённости её доходности». Чаще всего вычисляется среднегодовая мера.

Среднегодовая волатильность σ пропорциональна стандартному отклонению стоимости финансового инструмента деленной на квадратный корень из временного периода:

$\sigma = \frac{\sigma_{SD}}{\sqrt{P}}$, где P — временной период в годах.

Волатильность σ_T за интервал времени T (в годах) рассчитывается на основе среднегодовой волатильности по следующей формуле:

$$\sigma_T = \sigma\sqrt{T} .$$

Бета-коэффициент

Бета-коэффициент — это показатель, рассчитываемый для ценной бумаги или портфеля. Мера рыночного риска, отражает изменчивость доходности ценной бумаги (портфеля) по отношению к доходности портфеля (рынка, в случае портфеля) в среднем (среднерыночного портфеля).

Рассчитывается коэффициента Бета для актива в составе портфеля ценных бумаг по следующей формуле:

$\beta_a = \frac{Cov(r_a, r_p)}{Var(r_p)}$, где r_a — доходность актива, а r_p — доходность портфеля ценных бумаг.

В случае расчёта относительно рынка коэффициента Бета для актива (или портфеля) находится иначе:

$\beta_a = \frac{Cov(r_a, r_m)}{Var(r_m)}$, где r_a — доходность актива (или портфеля), а r_m — доходность рынка.

P-значение

Статистическая гипотеза проверяется путем сравнения наблюдаемого значения критерия с критическим значением, связанным с данным уровнем значимости, что позволяет отклонить или принять основную гипотезу. Однако, если уровень значимости будет другим, то придется вновь вычислять соответствующее критическое значение. Вводимое ниже понятие, ставшее популярным в связи с широким распространением статистических программ, позволяет решить вопрос о принятии или отклонении основной гипотезы одновременно для всех уровней значимости без вычисления критических значений.

Р-значением статистического критерия для фиксированной реализации $\vec{x}$ случайной выборки $\vec{X} = (X_1,...,X_n)$ называется такое число $PV(\vec{x})$, что $PV(\vec{x}) \geq \alpha$ для любого уровня значимости α, при котором гипотеза H_0 принимается, и $PV(\vec{x}) \leq \alpha$, для любого уровня значимости α, при котором гипотеза H_0 отвергается.

Предположим, что Р-значение $PV(\vec{x})$ уже каким-либо способом найдено. Тогда решение о принятии (отклонении) H_0 для заданного α осуществляется на основе следующего простого правила: если $PV(\vec{x}) < \alpha$, гипотеза H_0 отвергается, а если $PV(\vec{x}) > \alpha$ гипотеза H_0 принимается.

Критерий Колмогорова

Для любого $x \in R^n$ **ЧИСЛО** компонент вектора $\vec{x} = (x_1,...,x_n)$, которые меньше *x,* обозначим $m(x,\vec{x})$. Для случайного вектора $\vec{X} = (X_1,...,X_n)$ обозначение $m(x,\vec{X})$ имеет тот же смысл, но при этом $m(x,\vec{X})$ является дискретной случайной величиной с возможными значениями 0,1,...,n. Пусть *x* - реализация случайной выборки *X* объема *n* из некоторого распределения с функцией *F(x).* Эмпирическую функцию распределения, соответствующую выборке *x,* можно записать в виде

$$\hat{F} = \hat{F}(x, \vec{x}) = \frac{m(x, \vec{x})}{n}.$$

Оценка функции *F(x)* по случайной выборке *X* записывается аналогично:

$$\hat{F} = \hat{F}(x, \vec{X}) = \frac{m(x, \vec{X})}{n}.$$

Заметим, $\hat{F}(x, \vec{x})$ - числовая функция, тогда как $\hat{F}(x, \vec{X})$ в каждой точке *x* принимает случайное значение, т.е. является случайным процессом.

Определим расстояние между функциями $\hat{F}(x)$ и *F(x)* формулой

$$d = \sup_x \left| \hat{F}(x) - F(x) \right|.$$

Для функции $\hat{F} = \hat{F}(x, \vec{x})$ расстояние $d = d(\vec{x})$ - это просто число, тогда как для $\hat{F} = \hat{F}(x, \vec{X})$ расстояние $d = d(\vec{X})$ является случайной величиной, принимающей значения на отрезке [0,1].

Согласно доказанной А.Н. Колмогоровым теореме в случае непрерывной функции *F(x)* при любом неотрицательном $u \geq 0$ существует предел

$\lim_{n \to \infty} P(\sqrt{n} d(\vec{X}) < u) = K(u)$, где

$$K(u) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2u^2}.$$

Вследствие этой теоремы критерий согласия с критической областью $\sqrt{n} d(\vec{x}) > u_a$,

где u_a - корень уравнения *K(u) = 1-a,* имеет при $n \to \infty$ уровень значимости, стремящийся к α. Другими словами, α - асимптотический уровень значимости. Именно этот критерий и называется *критерием Колмогорова.*

Поскольку при *n* < 20 фактический уровень значимости заметно отличается от номинального значения *a,* критерий Колмогорова применяется при *n* > 20.

На практике, при вычислении максимального абсолютного отклонения гипотетической функции *F(x)* от эмпирической функции $\hat{F}(x)$ применяется следующая формула:

$$d(\vec{x}) = \max_{1 \le i \le n} \left\{ \left| \frac{i}{n} - F\left(x_{(i)}\right) \right|, \left| \frac{i-1}{n} - F\left(x_{(i)}\right) \right| \right\},$$

где $x_{(i)}$ - i-й член вариационного ряда

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n-1)} \le x_{(n)}.$$

Проверка гипотезы для модельных данных

Пусть гипотеза H_0, что модельные данные распределены нормально, согласно критерию согласия Колмогорова, распределение Р-значений критерия Харке-Бера равномерно.

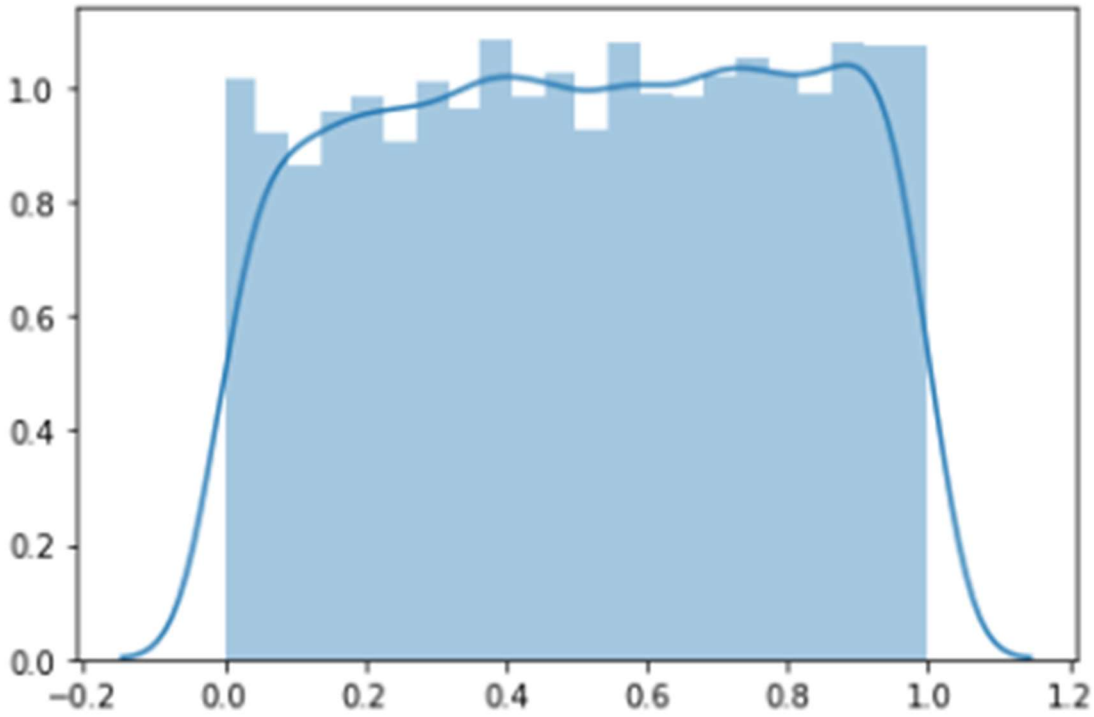
Вычислим таблицу 999 квантилей распределения статистики критерия Харке-Бера для проверки гипотезы на модельных данных методом Монте-Карло, предполагая, что нулевая гипотеза верна. Проверка гипотезы необходима, чтобы удостовериться, что программа, вычисляющая значение критерия и р-значения, работает правильно.

Таб. 3. Таблица квантилей.

quantile	values
0.001	0.001369
0.002	0.003378
0.003	0.004881
0.004	0.007078
0.005	0.009053
...	...
0.994	11.883632
0.995	12.260221
0.996	12.750242
0.997	13.591372
0.998	15.516066

Построим диаграмму Р-значений этих статистик для графического анализа. На рис.2 видно, что Р-значения распределены равномерно, в связи с чем нулевая гипотеза принимается.

Рис. 2. Гистограмма Р-значений.



Проверка гипотезы для реальных данных

В следующем разделе моей работы будет рассмотрена логарифмическая доходность рассматриваемых акций. Все периоды будут исследоваться на уровне 5%-ой значимости. Рассмотрим поведение логарифмической доходности в исследуемый период с 2010 до 2019 гг.

Чтобы проверить гипотезу о нормальном распределении логарифмической доходности по критерию Харке-Бера рассчитаем Р-значения. Для наглядности проведу расчёт Р-значений по годам и по месяцам. И закрашу красным в полученной таблице ячейки с Р-значением больше 0.05.

2010	0.000001	0.000000	0.000000	0.000000	0.000000
2011	0.000004	0.000000	0.000000	0.046470	0.046470
2012	0.000013	0.004577	0.000000	0.077169	0.077169
2013	0.000000	0.021017	0.000000	0.011544	0.011544
2014	0.000000	0.000000	0.000000	0.000001	0.000001
2015	0.028117	0.000000	0.000000	0.060799	0.060799
2016	0.000000	0.000000	0.000000	0.031656	0.031656
2017	0.000000	0.000000	0.000000	0.297176	0.297176
2018	0.000000	0.000000	0.000000	0.000000	0.000000
2019	0.000000	0.000000	0.000000	0.000000	0.000000

1	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.004032	0.004032
3	0.000000	0.000000	0.000000	0.145031	0.145031
4	0.000000	0.000000	0.000000	0.004970	0.004970
5	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.112747	0.008327	0.002359	0.002359
7	0.000000	0.000000	0.082995	0.064504	0.064504
8	0.000000	0.000000	0.025080	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.001216	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000003	0.000003
12	0.000000	0.002648	0.000000	0.001945	0.001945

Красным отмечены те ячейки, в которых P-значение превысило 0.05, т.е. гипотеза сработала. Как видно, случаев, в которых гипотеза принялась подавляющее меньшинство.

Заключение

Данная курсовая работа была написана с целью проверки гипотезы о нормальном распределении логарифмической доходности по критерию Харке-Бера. В ней проверены программы на модельных данных и на реальных данных. Применялись открытые библиотеки языка Python для удобства и ускорения вычислений. В ходе анализа можно сделать вывод о том, что модельные данные подчиняются критерию, в то время как

реальные данные опровергают гипотезу о нормальности. Только модельные данные распределены нормально.

Согласно общему рейтингу критериев нормальности на Википедии, критерий Харке-Бера считается одним из лучших по рангу. (в таблице указан как “Критерий асимметрии и эксцесса”).

Название критерия	**Характеристика альтернативного распределения**					**Ранг**
	асимметричное		**симметричное**		**близкое к нормальному**	
Критерий Шапиро-Уилка	1	1	3	2	2	1
Критерий асимметрии и эксцесса	7	8	10	6	4	2
Критерий Дарбина	11	7	7	15	1	3
Критерий Д'Агостино	12	9	4	5	12	4
Критерий эксцесса	14	5	2	4	18	5

Надо отметить, что среди исследуемых акций не наблюдается тенденции к сохранению определенного Р-значения. Можно сделать вывод, что закономерности в распределении логдоходностей обнаружено не было, в связи с этим сложно прогнозировать цены и их динамику.

Список использованной литературы

1. М.С.Красс, Б.П.Чупрынов. Математика в экономике-М.: Финансы и статистика,2007.
2. А. И. Кобзарь, Прикладная математическая статистика, 2006
3. www.finam.ru4
4. https://ru.wikipedia.org/wiki/Квантили_распределения_хи-квадрат
5. https://ru.wikipedia.org/wiki/Критерии_нормальности

Приложение

В связяи с особенностью пространства Jupyter Notebook для Python, копирую код из ячеек.

В нём своя ‘экосистема’, поэтому лучше открывать с помощью Jupyter Notebook, чем любым другим IDE для Python.

```
def suq(argdate) -> date:
    year = argdate // 10000
    month = (argdate % 10000) // 100
    day = argdate % 100


    return date(year, month, day)
def jarque_bera_mine(x):       #на вход функция получает x - набор наблюдений случайно величины, на выходе
    x = np.asarray(x)
    n = x.size
    if n == 0:
        raise ValueError('At least one observation is required.')
    mu = x.mean()
    diffx = x - mu
    skewness = (1 / n * np.sum(diffx**3)) / (1 / n * np.sum(diffx**2))**(3 / 2.)
    kurtosis = (1 / n * np.sum(diffx**4)) / (1 / n * np.sum(diffx**2))**2
    jb_value = n / 6 * (skewness**2 + (kurtosis - 3)**2 / 4)
    p = 1 - scipy.stats.chi2.cdf(jb_value, 2)
    return jb_value, p      #на выходе даёт jb-value - JB - статистика критерия, p - P-значение
df = pd.concat ((pd.read_csv('AKRN_100101_191231r.csv'), pd.read_csv('MGNT_100101_191231r.csv'),
                 pd.read_csv('HYDR_100101_191231r.csv'),
```

```
            pd.read_csv('ROSN_100101_191231r.csv'),
pd.read_csv('SBER_100101_191231r.csv')),
            ignore_index = True, sort = False)
df=df.drop(['<PER>', '<TIME>'], axis=1)
df['<LOGYIELD>'] = np.log(df['<CLOSE>'].divide(df["<CLOSE>"].shift(1)))
df['<RVOL>'] = df['<CLOSE>']*df['<VOL>']
df['<DATE>'] = df['<DATE>'].apply(lambda x: suq(x))
df = df.fillna(0)
df['<YEAR>'] = 0
df['<MONTH>'] = 0
df['<DAY>'] = 0
i = 0
for date in df['<DATE>']:
    df['<YEAR>'][i] = date.year
    df['<MONTH>'][i] = date.month
    df['<DAY>'][i] = date.day
    i += 1
df #если выдаёт ошибку, она не критична, таблицу выводит в любом случае
#Таблица кол-ва торговых дней
df1 = df.groupby(['<TICKER>', '<YEAR>'])[['<DATE>']].count()
df1.reset_index(inplace = True)
df1 = df1.pivot(index='<TICKER>', columns='<YEAR>', values='<DATE>')
```

```
df1

#Таблицы максимальных (вверх и вниз) дневных относительных скачков цен в процентах (по годам и акциям).

df2 = df

df2['<JUMP>'] = abs((df2['<CLOSE>'] - df2['<OPEN>']) / df2['<OPEN>'])

df2 = df2.groupby(['<TICKER>', '<YEAR>'])[['<JUMP>']].max()

df2.reset_index(inplace = True)

df2 = df2.pivot(index='<TICKER>', columns='<YEAR>', values='<JUMP>')


df2 = df2.apply(lambda x: x*100)

df2

#График изменения цен акций SBER и AKRN за всё время

alldates = df["<DATE>"].unique()

df3 = df.loc[df['<TICKER>'] == 'SBER']

df4 = df.loc[df['<TICKER>'] == 'AKRN']

df3 = df3.reset_index()

df4 = df4.reset_index()

sns.lineplot(x = alldates, y = df3['<CLOSE>'])

sns.lineplot(x = alldates, y = df4['<CLOSE>'])

#Таблица квантилей

nsmpl = np.random.normal(0, 1, 999)

qnt = []
```

```
PVs = []
for i in range(10000):
    nsmpl = np.random.normal(0, 1, 999)
    qnt.append(jarque_bera_mine(nsmpl)[0])
    PVs.append(jarque_bera_mine(nsmpl)[1])
w=np.array([i.round(3) for i in np.arange(0.001, 0.999, 0.001)])

q = pd.DataFrame(index=w, columns=['values'], data=np.quantile(np.array(qnt), w))
q.index.name = 'quantile'
print(np.std(qnt)/math.sqrt(len(qnt)))
q
#Гистограмма P-значений
sns.distplot(PVs)
scipy.stats.kstest(PVs,"uniform")[1]
def check_alpha(s):
    alpha = 0.05
    is_greater = s > alpha
    return ['background-color: red' if v else '' for v in is_greater]
#AKRN,MGNT,HYDR,ROSN,SBER
df3 = df.loc[df['<TICKER>'] == 'AKRN']
df3 = df3.reset_index()
df6 = df.loc[df['<TICKER>'] == 'MGNT']
```

```
df6 = df6.reset_index()

df7 = df.loc[df['<TICKER>'] == 'HYDR']

df7 = df7.reset_index()

df8 = df.loc[df['<TICKER>'] == 'ROSN']

df8 = df8.reset_index()

df9 = df.loc[df['<TICKER>'] == 'SBER']

df9 = df9.reset_index()

dflog=pd.DataFrame(index=df3["<DATE>"], data={"<AKRN>":df3['<LOGYIELD>'].values,

                                              "<MGNT>":df6['<LOGYIELD>'].values,

                                              "<HYDR>":df7['<LOGYIELD>'].values,

                                              "<ROSN>":df8['<LOGYIELD>'].values,

                                              "<SBER>":df8['<LOGYIELD>'].values,

"<YEAR>":df3['<YEAR>'].values,"<MONTH>":df3['<MONTH>'].values,"<DAY>":df3['<D
AY>'].values})

Dflog

dflogyear = dflog.groupby(['<YEAR>']).agg({

'<AKRN>': lambda x: jarque_bera_mine(x)[1],

'<MGNT>': lambda x: jarque_bera_mine(x)[1],

'<HYDR>': lambda x: jarque_bera_mine(x)[1],

'<ROSN>': lambda x: jarque_bera_mine(x)[1],

'<SBER>': lambda x: jarque_bera_mine(x)[1]})

dflogyear.style.apply(check_alpha)
```

```
dfmonth = dflog.groupby(['<MONTH>']).agg({
'<AKRN>': lambda x: jarque_bera_mine(x)[1],
'<MGNT>': lambda x: jarque_bera_mine(x)[1],
'<HYDR>': lambda x: jarque_bera_mine(x)[1],
'<ROSN>': lambda x: jarque_bera_mine(x)[1],
'<SBER>': lambda x: jarque_bera_mine(x)[1]})
dfmonth.style.apply(check_alpha)
dfday = dflog.groupby(['<DAY>']).agg({
'<AKRN>': lambda x: jarque_bera_mine(x)[1],
'<MGNT>': lambda x: jarque_bera_mine(x)[1],
'<HYDR>': lambda x: jarque_bera_mine(x)[1],
'<ROSN>': lambda x: jarque_bera_mine(x)[1],
'<SBER>': lambda x: jarque_bera_mine(x)[1]})
dfday.style.apply(check_alpha)
```