**Федеральное государственное бюджетное образовательное учреждение высшего образования**
**«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА и ГОСУДАРСТВЕННОЙ СЛУЖБЫ при Президенте Российской Федерации»**

ИНСТИТУТ ЭКОНОМИКИ, МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ
НАПРАВЛЕНИЕ 38.03.01

НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА ПО ТЕМЕ:

# «ИСПОЛЬЗОВАНИЕ БАЙЕСОВСКИХ МЕТОДОВ ДЛЯ АНАЛИЗА ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДОВ»

студент-бакалавр:
Носков Никита Александрович /______________/__________/
(подпись) (дата)

научный руководитель:
научно-исследовательской работы
д.т.н. Шилин Кирилл Юрьевич /______________/__________/
(подпись) (дата)

**МОСКВА**
**2021 г.**

# Оглавление

## Введение

В современных реалиях, одним из наиболее популярных методов сохранения и приумножения капитала является инвестирование в рынок ценных бумаг.

Применительно к России, по данным ЦБ РФ, количество инвесторов на внутреннем рынке выросло более чем в 2 раза и достигло 9,9 млн. чел., это составляет 12% от всего экономически активного населения страны.

Снижение ставок по депозитам до минимальных уровней, происходящее на фоне общемировых трендов по смягчению денежно-кредитной политики, и упрощение открытия брокерских и депозитарных счетов, связанное с глубокой цифровизацией банковского сектора, привлекает все больше и больше новых инвесторов на фондовый рынок. Некоторые из них передают свои деньги в доверительное управление, а некоторые пытаются создать собственный инвестиционный портфель.

При формировании инвестиционного портфеля основной задачей является определение ожидаемых доходностей ценных бумаг и рисков, связанных с изменением цены бумаг. В нынешних реалиях, люди обладают огромным количеством различных данных, которые можно использовать для моделирования доходностей и, исходя из, смоделированных данных подбирать веса активов в портфеле. В данной работе основное внимание уделено моделированию ожидаемых доходностей с помощью байесовских методов и подбору весов для бумаг в портфеле.

Основными задачами являются:

- Изучение основ современной портфельной теории;
- Изучение основ байесовского анализа;
- Моделирование доходностей ценных бумаг;
- Формирование инвестиционного портфеля, исходя из полученных данных.

# 1. Современная портфельная теория

## 1.1 Портфельная теория Марковица

В основе современной портфельной теории лежит подход, предложенный Гарри Марковицем в его статье в журнале “The Journal of Finance” в 1952 году.

Процесс формирования инвестиционного портфеля состоит из двух этапов. Вначале, инвестор изучает доступную ему информацию, касающуюся выбранных ценных бумаг, и формирует убеждения относительно будущих изменений цены бумаги, и, соответственно, относительно ее доходности и связью с доходностями остальных ценных бумаг. На втором же этапе он выбирает обоснованные убеждения относительно доходностей исследуемых ценных бумаг и, на основании полученных доходностей, формирует свой инвестиционный портфель.

Критерием оптимальности портфеля, как правило, считается использование всех возможностей его диверсификации, то есть снижения риска портфеля.

Пусть у нас есть $N > 1$ ценных бумаг, $r_i$ – доходность ценной бумаги $i$ (случайная величина), $R_i$ – ожидаемая доходность ценной бумаги $i$, $X_i$ – вес бумаги в нашем портфеле. Короткие продажи запрещены, следовательно $X_i \geq 0, \forall\, i \in (1; N)$. Тогда взвешенная средняя доходностей ценных бумаг в портфеле или, иначе, доходность портфеля с заданными весами будет иметь вид[1]:

$$R_p = \sum_{i=1}^{N} X_i * R_i, \text{ где } R_i = E[r_i]$$

В качестве меры риска принимается стандартное отклонение доходности ценной бумаги[2]:

$$\sigma_i = \sqrt{Var(r_i)}$$

Для того, чтобы учесть взаимосвязь между изменениями доходностей различных бумаг в портфеле используются ковариации доходностей, из которых формируется ковариационная матрица $V$.[3]

---

[1] Harry Markowitz, The Journal of Finance, Vol. 7, No. 1. (Mar., 1952), pp. 77-91.

[2] Там же, 1952 г.

[3] Там же, 1952 г.

$$\sigma_{ij} = cov(r_i, r_j) = E[(r_i - E[r_i])(r_j - E[r_j])]$$

$$V = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{pmatrix}$$

Имея заданные значения ожидаемых доходностей $R_i$ и коэффициенты ковариации бумаг $\sigma_{ij}$, инвестор, путем изменения весов бумаг в портфеле $X_i$, выбирает такой портфель, который будет иметь максимальную доходность $R_p$, при минимально возможном уровне риска $\sigma_p = X'VX$, где $X$ - вектор-столбец весов активов в портфеле. Тогда множество комбинаций возможных доходностей и риска портфеля ценных бумаг можно представить на графике:

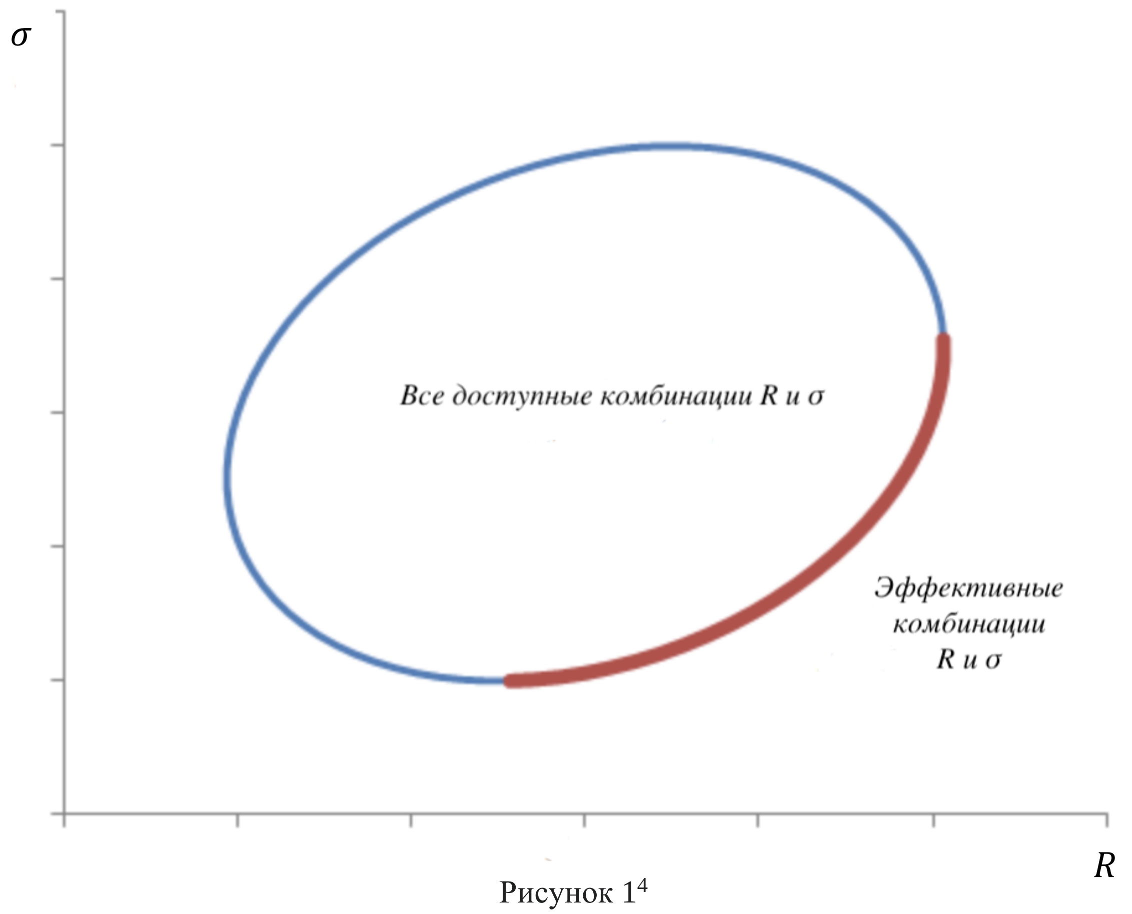



Рисунок 1[4]

Как видно из рисунка, эффективные комбинации доходности и риска портфеля, выделенные красным цветом, позволяют выбрать между более высокой доходностью, при

[4] Там же, 1952 г.

минимально возможном уровне риска, либо между более низким риском, при максимально возможном уровне доходности.

Формирование портфеля — это всегда выбор между наилучшей для инвестора комбинацией величин риск-доходность.

Теория Марковица предполагает, что инвестор выбирает между двумя вещами: Он либо максимизирует доходность, при заданном максимально возможном для него уровне риска, либо минимизирует риск, при заданном минимально допустимом уровне доходности портфеля.

### 1.1.1 Формирование портфеля с заданным уровнем риска

При составления оптимального портфеля с заданным уровнем риска инвестор выбирает максимально допустимый уровень риска, и максимизирует доходность портфеля.

Пусть[5]:

$X = (X_1, X_2, \dots, X_n)'$ - вектор-столбец весов активов;

$R = (R_1, R_2, \dots, R_n)'$ - вектор-столбец доходностей активов;

$E = (E_1, E_2, \dots, E_n)'$ - единичный вектор-столбец;

$\sigma_p$ – заданный уровень риска портфеля;

$V$ – ковариационная матрица размера $n \times n$.

В таком случае решается следующая оптимизационная задача[6]:

$$\begin{cases} \max\limits_{X_i}(X' * R) \\ X' * V * X = \sigma_p \\ X' * E = 1 \\ X_i \geq 0 \end{cases}$$

В результате решения получаем набор весов активов для оптимального портфеля с заданным уровнем риска и максимальной возможной доходностью.

### 1.1.2 Формирование портфеля с заданным уровнем доходности

При составления оптимального портфеля с заданным уровнем доходности инвестор выбирает минимальный возможный уровень доходности, и минимизирует риск портфеля.

Пусть[7]:

[5] Там же, 1952 г.
[6] Там же, 1952 г.
[7] Там же, 1952 г.

$X = (X_1, X_2, \dots, X_n)'$ - вектор-столбец весов активов;

$R = (R_1, R_2, \dots, R_n)'$ - вектор-столбец доходностей активов;

$E = (E_1, E_2, \dots, E_n)'$ - единичный вектор-столбец;

$R_p$ – заданный уровень доходности;

$V$ – ковариационная матрица размера $n \times n$.

В таком случае решается следующая оптимизационная задача[8]:

$$\begin{cases} \min\limits_{X_i}(X' * V * X) \\ X' * R = R_p \\ X' * E = 1 \\ X_i \geq 0 \end{cases}$$

В результате решения получаем набор весов активов для оптимального портфеля с заданным уровнем доходности и минимально возможным риском.

### 1.1.3 Коэффициент Шарпа

Коэффициент, предложенный Уильямом Шарпом, определяющий эффективность инвестиционного портфеля, как отношение риск-премии к риску данного портфеля.

Он определяется по формуле[9]:

$$SR = \frac{R - R_f}{\sigma}$$

Где $R$ – доходность портфеля, $R_f$ – безрисковая доходность, $\sigma$ – риск портфеля, выраженный в виде стандартного отклонения. В качестве доходности, как правило, берется доходность по среднесрочным облигациям государственного займа.

### Выводы

Современная портфельная теория не ограничивается только работой Марковица. В частности, развитием методов, предложенных им, является работа Фишера Блэка и Роберта Литтермана[10]. Методы, предложенные в ранее указанной работе, могут быть использованы для дополнения моделей и продолжения исследования.

---

[8] Там же, 1952 г.

[9] William F. Sahrpe, The Journal of Portfolio Management Fall 1994, 21 (1) 49-58

[10] Black F. and Litterman R.: Global Portfolio Optimization, Financial Analysts Journal, September 1992, pp. 28–43

Из портфельной теории Марковица следует, что основную сложность при формировании оптимального портфеля играют определение ожидаемых доходностей ценных бумаг, и их рисков. Если построить матрицу ковариаций доходностей ценных бумаг представляется возможным, то формирование ожиданий относительно доходностей, как взятия среднего из исторических данных, представляется слишком упрощенной аппроксимацией. Методам решения данной проблемы и будет посвящена следующая глава.

# 2. Байесовские методы

## 2.1 Основные концепции байесовской статистики

Методы байесовской статистики используют и имеют в своей основе теорему Байеса для расчета и вероятностей и их обновления, при получении новых данных.

Теорема Байеса выглядит следующим образом[11]:

$$P(X|Y) = \frac{P(Y|X) * P(X)}{P(Y)}$$

В данном случае, $X$ – это предполагаемая гипотеза, $Y$ – это имеющиеся в нашем распоряжении данные. То есть теорема Байеса показывает каким образом можно вычислить вероятность выполнении нашей гипотезы, при условии наличия определенных данных. Если рассуждать терминами теории вероятностей, то элементы данной теоремы имеют следующие обозначения

- $P(X)$ – априорная вероятность;
- $P(Y|X)$ – правдоподобие;
- $P(X|Y)$ – апостериорная вероятность;
- $P(Y)$ – предельное правдоподобие.

Рассмотрим каждый из элементов подробнее.

Априорная вероятность определяется распределением, которое наиболее полно соответствует нашей информации относительно параметра $X$. Если же наших знаний недостаточно для того, чтобы выбрать априорную вероятность, то может быть выбрано распределение, не содержащее сколько-нибудь значимую информацию.

Правдоподобие является комбинацией априорной вероятности и модели правдоподобия и определяет, как будут представлены наши данные в дальнейшем анализе.

Апостериорная вероятность является результатом байесовского моделирования, который показывает всю нашу информацию о выдвинутой гипотезе, при учете всех имеющихся у нас данных. По сути, апостериорная вероятность является обновленной, с учетом наших данных, априорной вероятностью.

[11] https://ru.wikipedia.org/wiki/Теорема_Байеса

Предельное правдоподобие, определяет вероятность наших данных, при условии, что априорная вероятность исключена из рассмотрения.

Поскольку при байесовском анализе основное внимание уделяется изучению распределений случайных величин, и интерес представляют относительные показатели параметров, а не абсолютные, то можно не учитывать предельное правдоподобие, и тогда формула теоремы Байеса примет следующий вид[12]:

$$P(X|Y) \propto P(Y|X) * P(X)$$

## 2.2 Выбор априорной вероятности

Выбранное априорное распределение оказывает прямое влияние на результаты байесовского анализа. Это можно видеть на примере с задачей о подбрасывании монетки, приведенной Освальдо Мартином в его книге[13].

Предположим, что у нас есть данные о 150 бросках монеты и частоте выпадения орла и решки, и мы хотим определить, является ли монета симметричной, или она смещена в сторону выпадения или решки. И считается, что монета с отклонением равным 1 всегда падает орлом вверх, с отклонением 0 всегда падает решкой вверх, а при отклонении равном 0,5 в половине случаев выпадает орел, а в половине решка, что равносильно предположению о том, что монета является симметричной. Опустим этап, связанный с определением правдоподобия, так как в текущий момент он не представляет для нас интереса.

В качестве априорной вероятности и правдоподобия были выбраны бета-распределение и биномиальное распределение, соответственно. Проведя несколько математических преобразований, получаем, что формула для расчета апостериорной вероятности приняла вид[14]:

$$P(\theta|y) \propto Beta(\alpha_{prior} + y, \beta_{prior} + N - y)$$

Теперь зададим три варианта априорных распределений вероятностей:

- Равномерное, при котором все возможные значения отклонения равновероятны (синий цвет);
- Нормальное, со средним значением в точке 0,5 (оранжевый цвет);

[12] Мартин О. Байесовский анализ на Python/ пер. с анг. А.В. Снастина. – М.:ДМК Пресс, 2020. – 340 с.: ил
[13] Там же, 2020.
[14] Там же, 2020.

- Асимметричное, при котором больший вес придается результатам с отклонением в сторону выпадения решки (зеленый цвет).

И, вычислим и отобразим на графиках этапы расчета апостериорного распределения (рисунок 3).

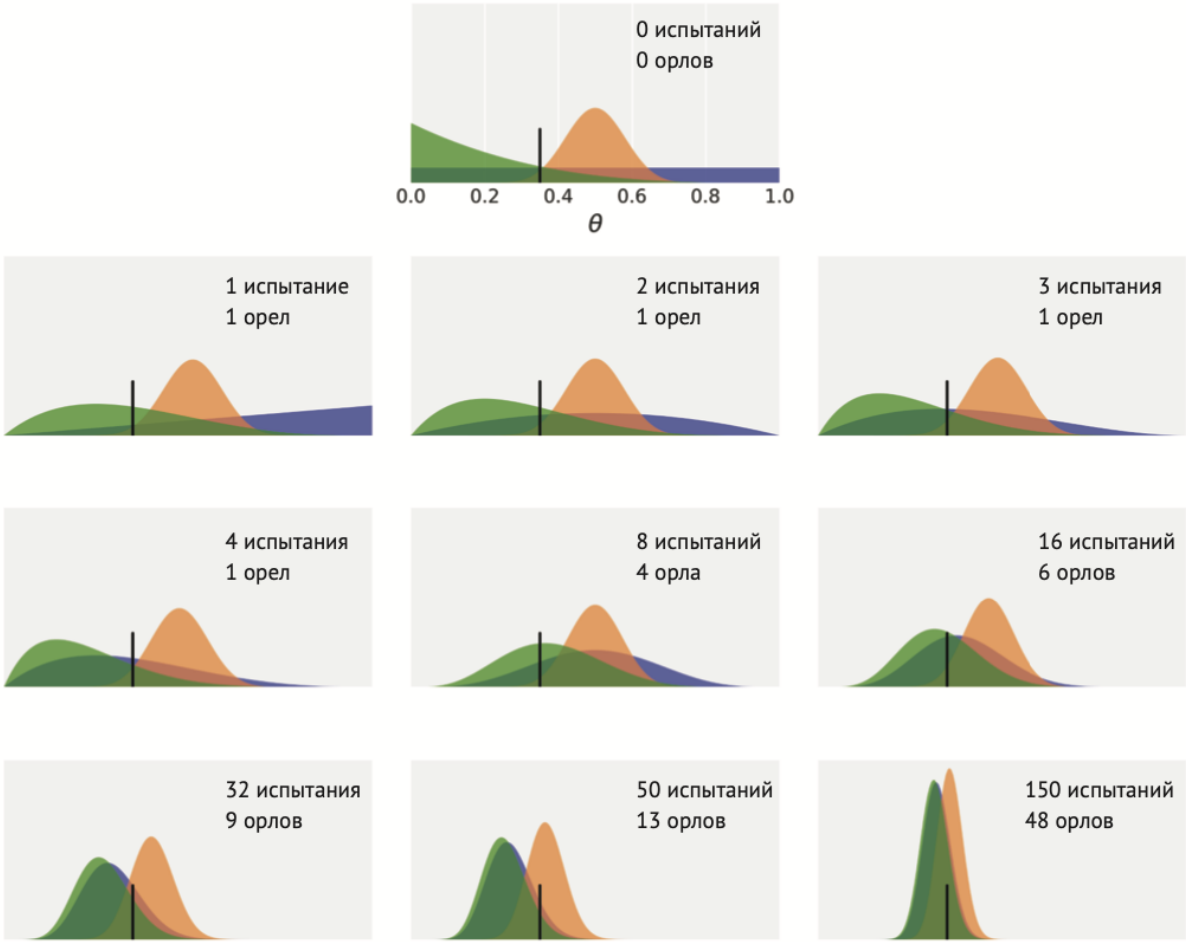



Рисунок 3[15]

Исходя из рисунка 3 можно видеть, что скорость сходимости различных апостериорных распределений к одной форме зависит от выбранных априорных распределений. Также важно заметить, что, при наличии достаточно большого объема данных, байесовские модели с различными априорными распределениями будут сходится к одному и тому же результату.

При решении задач, как правило известны минимальные данные об исследуемом параметре, например границы или знак принимаемых значений и так далее, в следствие чего, при наличии данной информации, более эффективным является применение

[15] Там же, 2020.

информативных или слабо информативных априорных распределений, нежели чем применение неинформативных априорных распределений (оказывающих наименьшее возможное воздействие на анализ).

### 2.3 Вероятностная модель простой линейной регрессии

Зачастую, нам необходимо спрогнозировать зависимость одного фактора от другого, или от совокупности других факторов. Одна из самых простых моделей, которая может помочь в этом – это модель линейной регрессии. Она позволяет выявить линейную зависимость между прогнозируемой переменной $y$ и прогнозирующей переменной $x$.

Существование линейной зависимости между переменными может быть выражено в следующей форме:

$$y_t = \alpha + \beta x_t$$

Данное уравнение говорит о том, что на переменную $y_t$ влияет регрессор $x_t$, учтенный с коэффициентом $\beta$, а также свободный член $\alpha$.

Если $x$ представлен в виде вектора, то такая регрессия называется множественной линейной регрессией, и она имеет вид:

$$y_t = \alpha + BX$$

Где:

- $B$ — вектор коэффициентов $\beta_i, i \in [0; n]$;
- $X$ — вектор регрессоров (зависимых переменных) $x_i, i \in [0; n]$.

Значения коэффициентов $\alpha, \beta$, для линейных регрессионных моделей, находятся с помощью метода наименьших квадратов, который минимизирует среднеквадратичную ошибку модели и дает наилучшие значения коэффициентов.

Вероятностный подход позволяет получить не только наилучшие оценки параметров модели, но также и неопределенность, присущую этим оценкам. При использовании данного подхода мы делаем предположение не о том, что переменная $y$ линейно зависит от $x$, а о том, что она имеет вероятностное распределение со средним значением, которое линейно зависит от переменной $x$.

Для случая моделирования $y$ с помощью нормального распределения, модель вероятностной линейно регрессии будет иметь следующий вид[16]:

[16] Там же, 2020.

$$y \sim \mathcal{N}(\mu = \alpha + x\beta, \sigma)$$

В данным случае, массив данных $y$ определяется нормальным распределением, со средним значением $\mu = \alpha + x\beta$ и стандартным отклонением равным $\sigma$.

Так как нам неизвестны значение параметров $\alpha, \beta$ и $\sigma$, то, для данной постановки задачи, их также необходимо задать априорными распределениями. Зачастую используют нижеследующие распределения[17]:

$$\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha)$$
$$\beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta)$$
$$\sigma \sim |\mathcal{N}(0, \sigma_\sigma)|$$

Для наглядного представления модели вероятностной линейной регрессии удобно использовать диаграммы Крушке[18], пример такой диаграммы для описанной выше модели представлен на рис. 4.

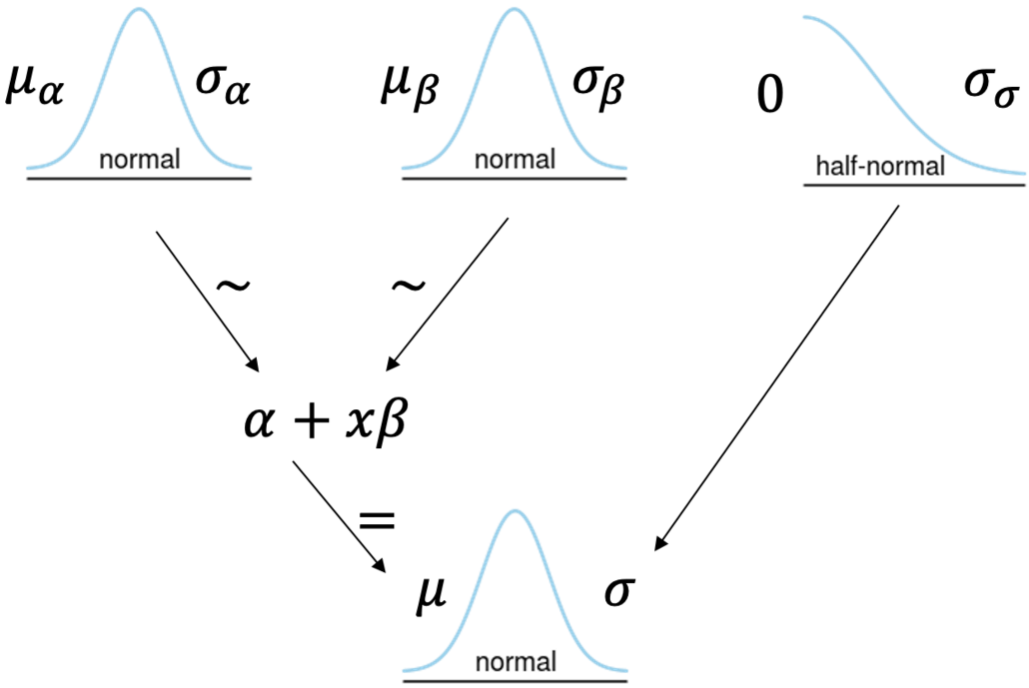



Рисунок 4 - Диаграмма Крушке для вышеописанной модели[19]

[17] Там же, 2020.
[18] http://www.sumsar.net/blog/2013/10/diy-kruschke-style-diagrams/
[19] Мартин О. Байесовский анализ на Python/ пер. с анг. А.В. Снастина. – М.:ДМК Пресс, 2020. – 340 с.: ил

### 2.4 Вероятностная модель робастной линейной регрессии

Использование нормального распределения для аппроксимации наших данных является обоснованным для многих задач, но, если данные имеют промахи, или не так сильно центрированы относительно своего среднего значения, то имеет смысл использовать распределения, которые менее центрированы, по сравнению с гауссовым, и плотности которых имеют более «тяжелые» хвосты.

Примером такого распределения является распределение Стьюдента[20].

Плотность распределения Стьюдента в сравнении с плотностью нормального распределения изображена на рис. 5, где параметр $\nu$ — это число коэффициентов свободы, так же называемое параметром нормальности распределения. Известно, что при $\nu \geq 30$ распределение Стьюдента практически совпадает с нормальным распределением[21].

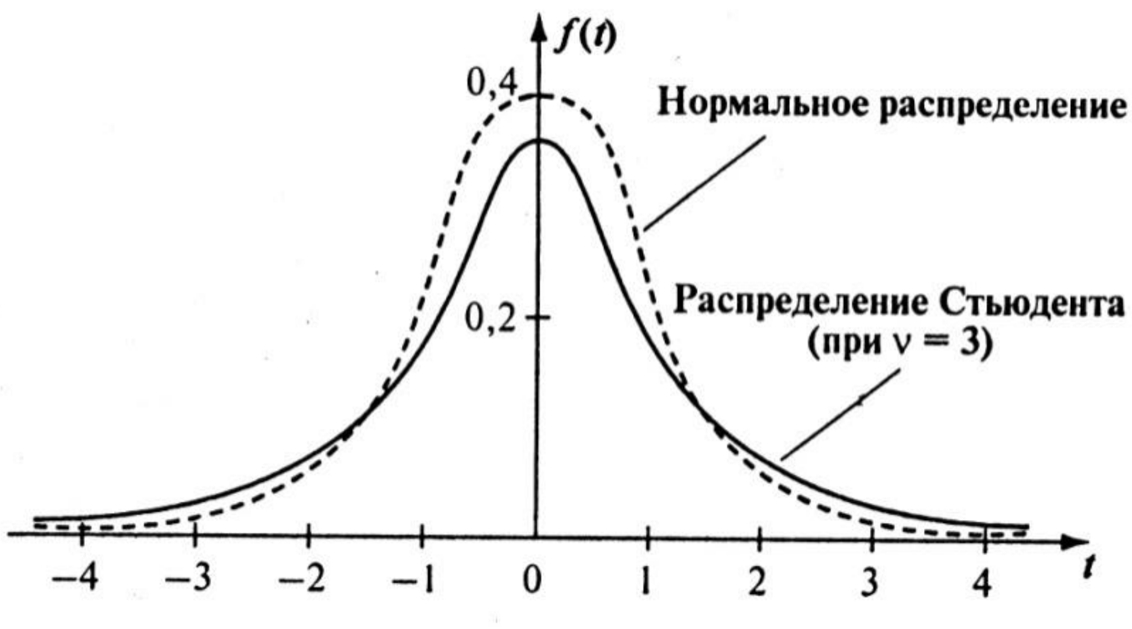



Рисунок 5[22]

Для иллюстрирования эффективности использования распределения Стьюдента, по сравнению с гауссовым распределением, Освальдо Мартин приводит пример с использованием третий группы данных из квартета Энскомба.

Данный набор данных показан на рис. 6

[20] https://ru.wikipedia.org/wiki/Распределение_Стьюдента
[21] Там же
[22] https://userdocs.ru/matematika/3696/index.html?page=5

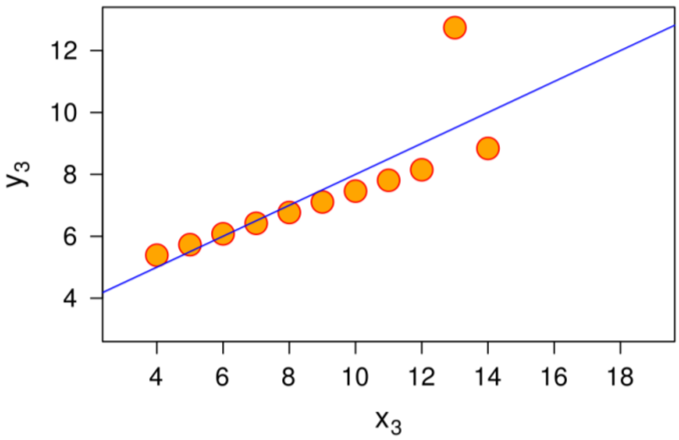



Рисунок 6[23]

На Рис. 7 показаны результаты моделирования с помощью простой линейной регрессии и с помощью вероятностной робастной регрессии. Можно заметить, что моделировании данной группы данных с помощью простой линейной регрессии находит компромисс, и смещает линию регрессии для учета всех точек, в то время как линия робастная байесовская модель просто отбрасывает данную точку, являющуюся выбросом или сильным отклонением от общего массива данных. «Благодаря своему более медленно убывающему (более широкому) хвосту t-распределение Стьюдента способно снижать значимость точек, которые расположены слишком далеко от основного массива данных»[24].

### 2.5 Сравнение вероятностных моделей

Любая модель является всего лишь аппроксимацией нашей реальности, и никогда полностью ей не соответствует. Важнейшей частью анализа, наряду с составлением самих моделей, является выбор наилучшей модели из имеющихся.

Для сравнения моделей, зачастую, используют информационные критерии.

«Информационный критерий — применяемая в эконометрике (статистике) мера относительного качества эконометрических (статистических) моделей, учитывающая степень «подгонки» модели под данные с корректировкой (штрафом) на используемое количество оцениваемых параметров.»[25]

---

[23] https://ru.wikipedia.org/wiki/Квартет_Энскомба
[24] Мартин О. Байесовский анализ на Python/ пер. с анг. А.В. Снастина. – М.:ДМК Пресс, 2020. – 340 с.: ил.
[25] (https://ru.wikipedia.org/wiki/Информационный_критерий

Как правило, информационные критерии используются исключительно для сравнения моделей между собой, без анализа какой-либо содержательной составляющей самих критериев.

Первым предложенным информационным критерием является информационный критерий Акаике[26], определяемы следующей формулой[27]:

$$AIC = -2\sum_{i=1}^{n} logp\left(y_i \middle| \hat{\theta}_{mle}\right) + 2pAIC$$

Где:

- $pAIC$- количество параметров;
- $\hat{\theta}_{mle}$- максимальная оценка правдоподобия значения $\theta$;
- $\sum_{i=1}^{n} logp\left(y_i \middle| \hat{\theta}_{mle}\right)$ – логарифмическая функция правдоподобия.

Первое слагаемое критерия учитывает степень соответствия выбранной модели данным, а второе штрафует за большое количество переменных, так как лучшей моделью, согласно информационным критериям, является та, у которой значение информационного критерия ниже.

В своей книге Освальдо Мартин указывает, что информационный критерий Акаике хорошо работает при использовании небайесовских методов, в то время как его применение в байесовском анализе не всегда целесообразно, так как он никак не учитывает апостериорное распределение, а значит не рассматривает информацию о неопределенности наших оценок.

Наиболее предпочтительным критерием, для использования в байесовским анализе, является информационный критерий Ватанабэ-Акаике. Он вычисляется по следующей формуле[28]:

$$WAIC = -2\sum_{i}^{n} log\left(\frac{1}{S}\sum_{S=1}^{s} p(y_i|\theta^S)\right) + 2\sum_{i}^{n}\left(\sum_{S=1}^{s} logp(y_i|\theta^S)\right)$$

Первое слагаемое – это логарифм поточечно определяемой прогнозируемой плотности, он вычисляет среднее значение правдоподобия по $S$ выборкам апостериорного распределения. Фактически, это логарифмическая функция правдоподобия, но вычисленная по апостериорному распределению. Второе слагаемое вычисляет дисперсию

[26] Там же
[27] Мартин О. Байесовский анализ на Python/ пер. с анг. А.В. Снастина. – М.:ДМК Пресс, 2020. – 340 с.: ил.

[28] Там же, 2020.

логарифмической функции правдоподобия по *S* выборкам апостериорного распределения. Интуитивно можно предположить, что чем больше параметров в модели, тем более широким будет разброс апостериорного распределения, следовательно более высокое значение дисперсии и штраф, за использование большого числа параметров.

Еще одним инструментом, используемым для сравнения моделей при байесовском анализе, являются коэффициенты Байеса.

Исходя из формулы (байеса) можно составить зависимость статистического вывода от конкретной модели, она будет иметь вид:

$$P(X|Y, M_i) = \frac{P(Y|X, M_i) * P(X, M_i)}{P(Y|\ M_i)}$$

Если ранее, рассматривая саму формулу Байеса, мы говорили, что слагаемое $P(Y|\ M_i)$, являющееся предельным правдоподобием, можно опустить, без снижения качества расчетов, то сейчас, для сравнения моделей, это является важной числовой характеристикой. Коэффициентом Байеса называют отношение предельных правдоподобий двух моделей[29]:

$$BF = \frac{P(Y|\ M_i)}{P(Y|\ M_j)}$$

Если $BF > 1$, то модель $M_i$ объясняет данные лучше, чем модель $M_j$.

Мы можем использовать предельные правдоподобия, если модели имеют одинаковые априорные вероятности, в противном случае необходимо вычислять апостериорные шансы по нижеследующей формуле:

$$\frac{P(M_i|Y)}{P(M_i|Y)} = \frac{P(Y|\ M_i)}{P(Y|\ M_i)} \frac{P(M_i)}{P(M_i)}$$

Где:

- $\frac{P(M_i|Y)}{P(M_i|Y)}$ – апостериорные шансы;
- $\frac{P(Y|\ M_i)}{P(Y|\ M_i)}$ – коэффициенты Байеса;
- $\frac{P(M_i)}{P(M_i)}$ – априорные шансы.

**Выводы**

В данной главе были рассмотрены основные концепции байесовского анализа, который рассматривает получение не точечных оценок и прогнозов, интересующих нас

[29] Там же, 2020.

параметров, а их вероятностных распределений. Также были рассмотрены вероятностные регрессии, которые будут использоваться для построения прогнозов в практической части этой работы, и способы сравнения моделей и выбора наилучшей с помощью информационных критериев и коэффициентов Байеса.

# 3. Анализ доходностей акций российского нефтегазового сектора

В рамках данной главы будет рассматриваться построение и анализ моделей, прогнозирующих доходности акций компаний из российского нефтегазового сектора, входящих в индекс Нефти и Газа Московской Биржи. В него входят следующие бумаги[30]:

- Башнефть;
- Газпром;
- Лукойл;
- Новатэк;
- РуссНефть;
- Роснефть;
- Сургутнефтегаз;
- Сургутнефтегаз-привилегированные;
- Татнефть;
- Татнефть- привилегированные;
- Транснефть- привилегированные.

## 3.1 Выгрузка и подготовка данных

Для анализа были выбраны недельные доходности, так как они обеспечивают достаточное количество точек для подгонки модели, и при этом модели не так часто дают рекомендации по ребалансировке портфеля, как это было бы при использовании высокочастотных данных. Весь анализ проводится с помощью средств языка программирования Python.

Выгрузка данных выполнялась с помощью библиотеки investpy[31], код для выгрузки представлен ниже.

[30] https://www.moex.com/ru/index/MOEXOG/constituents
[31] https://investpy.readthedocs.io/

```
import investpy

cols = ('RNFT', 'BANE_P', 'TRNF_P', 'SNGS_P', 'SNGS', 'ROSN', 'TATN_P',
'TATN', 'GAZP',
        'LKOH', 'NVTK')

for i in cols:

    df = investpy.get_stock_historical_data(stock = i,
                                            country = 'Russia',
                                            from_date = '01/01/2017',
                                            to_date = '01/04/2021')
    df.to_csv(f'{i}.csv')
```

С помощью переменной cols задаются биржевые тикеры бумаг. Данные выгружаются с 01/01/2017 по 01/04/2021 и сохраняются в виде файлов .csv.

После этого из каждого массива данных о бумаге берутся цены на момент закрытия торгов и объединяются в новую таблицу, которая будет использоваться для анализа.

```
    close_prices = pd.concat([msci['Close_msci'],brent['Close_brent'],
banep['Close_banep'],gazp['Close_gazp'],lkoh['Close_lkoh'],
nvtk['Close_nvtk'],rnft['Close_rnft'], rosn['Close_rosn'],
sngsp['Close_sngsp'], sngs['Close_sngs'], tatnp['Close_tatnp'],
tatn['Close_tatn'],trnfp['Close_trnfp'], usdrub['Close_usdrub']],axis = 1,
join = 'inner')
```

Вид полученной таблицы представлен в Приложении 1.

Затем, с помощью библиотеки для расчета значений технических индикаторов ta-lib[ссылка] производится расчет значений индикатора Скользящее среднее с окном в 14 периодов, сдвинутых на один период вперед, так как для прогнозирования доходности в период *t* мы будем использовать значение параметра в период *t-1*. Аналогичная процедура сдвига проводится и для цен бумаг.

```
for i in cols:
    close_prices[f'MA_{i}_t-1'] = ta.trend.sma_indicator(close =
close_prices.loc[:, f'Close_{i}'], window = 14, fillna = True).shift(1)
    close_prices[f'MA_{i}'] = ta.trend.sma_indicator(close =
close_prices.loc[:, f'Close_{i}'], window = 14, fillna = True)
    close_prices[f'Close_{i}_t-1'] = close_prices.loc[:,
f'Close_{i}'].shift(1)
```

Как видно в Приложении 1, в качестве индекса используется дата, но, поскольку мы хотим рассматривать недельные доходности, то необходимо перейти от дат к дням недели и считать процентное изменение цены, что эквивалентно доходности, день недели к дню недели. Этот переход выполняется с помощью нижеследующего фрагмента кода:

```
close_prices.index = close_prices.index.weekday
week_name = {0: 'Mon', 1: 'Tue', 2: 'Wed', 3: 'Thu', 4: 'Fri', 5: 'Sat', 6: 
'Sun'}
close_prices = close_prices.rename(index = week_name)
close_prices.dropna(inplace = True)
```

После смены дат на дни недели мы можем перейти к составлению таблицы с недельными доходностями:

```
returns = pd.DataFrame()
for i in ['Mon', 'Tue', 'Wed', 'Thu', 'Fri']:
    returns = returns.append(close_prices[close_prices.index == 
i].pct_change())
returns.dropna(inplace = True)
```

### 3.2 Спецификация моделей

Теперь определим три вида моделей, но вначале определим входные данные:

$y$ - недельная доходность акций компании;

$X$ – вектор, состоящий из курса доллара в рублях за предыдущий период, цены нефти марки Brent за предыдущий период, доходности акции компании за предыдущий период, значения 14-дневного скользящего среднего за предыдущий период и значения индекса MSCI Russia за предыдущий период.

Данные параметры были выбраны для использования в качестве регрессоров, так как они достаточно сильно коррелируют с доходностью бумаг, что можно видеть на примере корреляционной матрицы для бумаг компании «Новатэк» на Рис. 7. Также можно предположить, что логично наличие данных взаимосвязей, ведь цены акций компаний нефтегазового сектора могут зависеть от цен на нефть, так как это основной выходной продукт их деятельности, они могут зависеть от индекса MSCI Russia, так как он отражает ситуацию на фондовом рынке, еще они могут зависеть от курса доллара, так как он оказывает влияние на нефтегазовые доходы и от среднего значения доходности за последние несколько периодов.

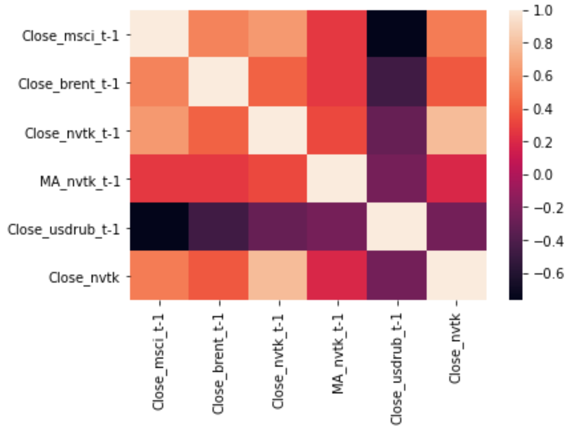



Рисунок 7

Входные данные будут идентичными для двух моделей.

Первая модель использует в качестве предполагаемого распределения наших данных распределение Стьюдента и имеет следующий вид:

$$\alpha \sim \mathcal{N}(0.1, 0.01)$$

$$\beta \sim \mathcal{N}(0.1, 0.01)$$

$$\sigma \sim |\mathcal{N}(0.1, 0.01)|$$

$$\gamma \sim Gamma(0.1, 0.01)$$

$$y \sim Student(\mu = \alpha + \beta X, \sigma, \gamma)$$

Диаграмма Крушке для данной модели изображена на Рис. 8.

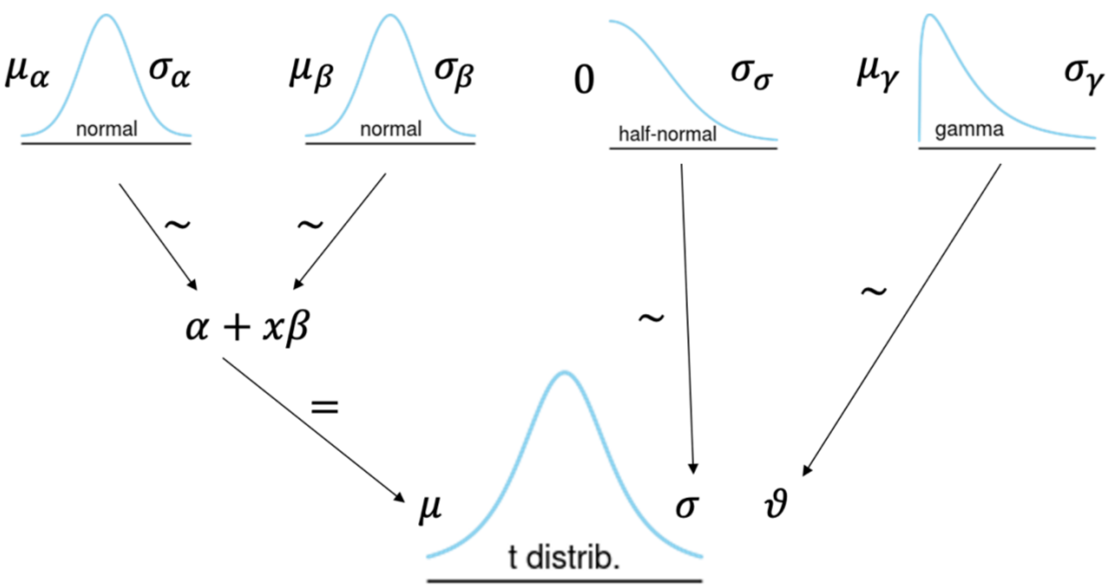



Рисунок 8

Для задания вероятностных моделей существует библиотека pymc3[32], которую автор и будет использовать для определения моделей и прогнозирования доходностей. Для создания вышеописанной модели была написана функция, которая в качестве входной переменной использует название бумаги, таблицу с историческими доходностями данной бумаги и всех влияющих на нее переменных и, задаваемое для априорных распределений, математическое ожидание и стандартное отклонение, а на выходе дает спецификацию модели для указанной компании. Данная функция приведена ниже:

```
def returns_oil_gas_student(company, data = returns, mu_input = 0.1, sd_input = 0.01):

    y = data.loc[:, f'Close_{company}'].to_numpy()
    x = data.loc[:,['Close_usdrub_t-1' ,'Close_brent_t-1', f'Close_{company}_t-1',
                        f'MA_{company}_t-1', 'Close_msci_t-1']].to_numpy()

    with pm.Model() as model_company:

        α = pm.Normal('α', mu = mu_input, sd = sd_input)
        β = pm.Normal('β', mu = mu_input, sd = sd_input, shape = 5)
        σ = pm.HalfNormal('σ', sd = 0.1)
        γ = pm.Gamma('γ', 2, 0.1)
        x_shared = pm.Data('x_shared', x)
        μ = α + pm.math.dot(x_shared, β)
        y_pred = pm.StudentT('y_pred', mu = μ, sd = σ, nu = γ, observed = y)

    return model_company
```

Вторая модель использует в качестве предполагаемого распределения наших данных распределение Гаусса или, иначе, нормальное распределение и имеет следующий вид:

$$\alpha \sim \mathcal{N}(0.1, 0.01)$$

$$\beta \sim \mathcal{N}(0.1, 0.01)$$

$$\sigma \sim |\mathcal{N}(0.1, 0.01)|$$

$$y \sim \mathcal{N}(\mu = \alpha + \beta X, \sigma)$$

Диаграмма Крушке для данной модели изображена на Рис. 9.

---

[32] http://docs.pymc.io/

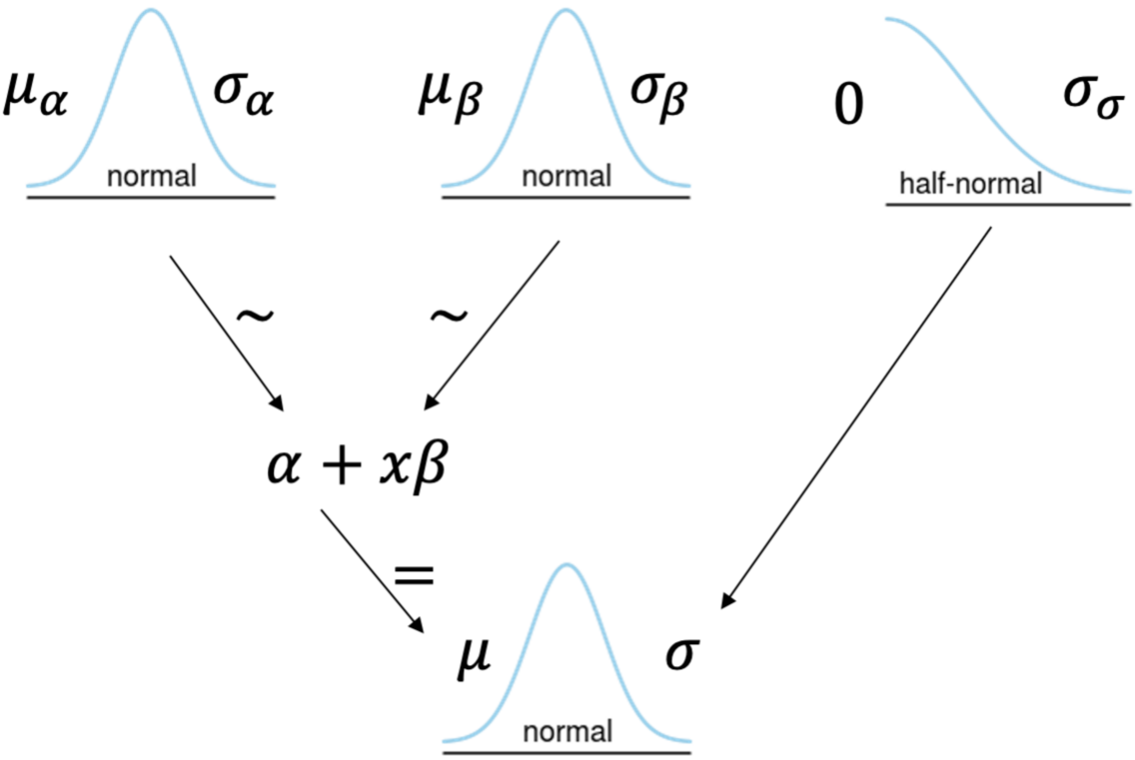



Рисунок 9

Для создания вышеописанной модели была написана функция, аналогичная функции для предыдущей модели. Данная функция приведена ниже:

```
def returns_oil_gas_gaussian(company, data = returns, mu_input = 0.1, sd_input = 0.01):

    y = data.loc[:, f'Close_{company}'].to_numpy()
    x = data.loc[:,['Close_usdrub_t-1' ,'Close_brent_t-1', f'Close_{company}_t-1',
                    f'MA_{company}_t-1', 'Close_msci_t-1']].to_numpy()

    with pm.Model() as model_company:

        α = pm.Normal('α', mu = mu_input, sd = sd_input)
        β = pm.Normal('β', mu = mu_input, sd = sd_input, shape = 5)
        σ = pm.HalfNormal('σ', 0.1)
        x_shared = pm.Data('x_shared', x)
        μ = α + pm.math.dot(x_shared, β)
        y_pred = pm.Normal('y_pred', mu = μ, sd = σ, observed = y)

    return model_company
```

Третья модель практически аналогична первой, за исключением того, что стандартное отклонение в данной модели определяется с использованием индекса волатильности российского рынка (*RVI*). Данный индекс рассчитывается исходя из стоимости опционов на российский индекс РТС, и отражает степень неопределённости или «страха» на фондовом рынке. По мнению автора, использование значения данного индекса вполне обоснованно, и может помочь в составлении более качественных прогнозов. Модель будет определяться следующим образом:

$$\alpha \sim \mathcal{N}(0.1, 0.01)$$
$$\beta \sim \mathcal{N}(0.1, 0.01)$$
$$\delta \sim |\mathcal{N}(0.1, 0.01)|$$
$$\gamma \sim Gamma(0.1, 0.01)$$
$$y \sim Student(\mu = \alpha + \beta X, \sigma = \delta \cdot volatility, \gamma)$$

Диаграмма Крушке для данной модели представлена на Рис. 10.

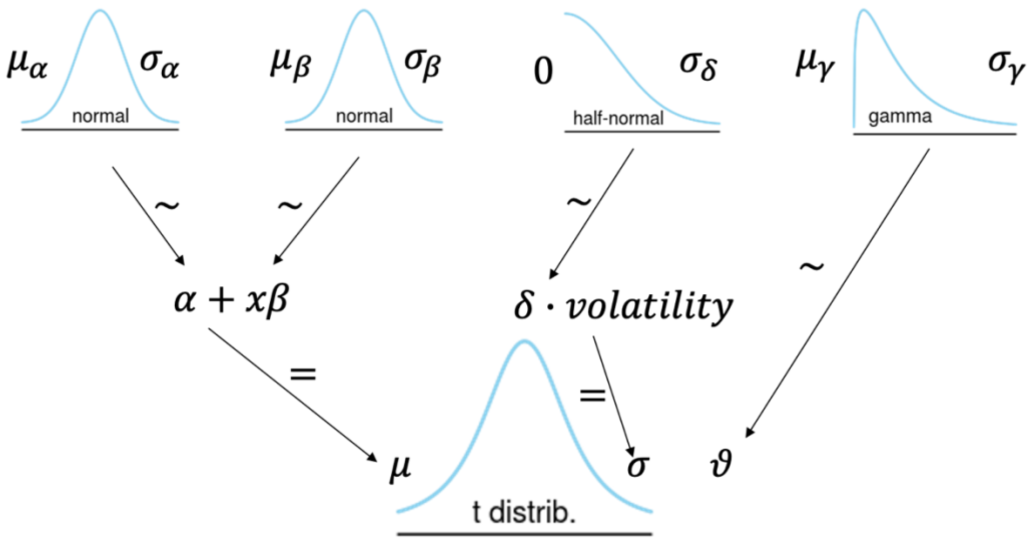



Рисунок 10

Для данной модели, аналогично двум предыдущим, была создана своя функция:

```
def returns_oil_gas_volatility(company, data = returns, mu_input = 0.1, sd_input =
0.01):

    y = data.loc[:, f'Close_{company}'].to_numpy()
    x = data.loc[:,['Close_usdrub_t-1' ,'Close_brent_t-1', f'Close_{company}_t-1',
                          f'MA_{company}_t-1', 'Close_msci_t-1']].to_numpy()
    volatility = data.loc[:, 'Close_rvi_t-1'].to_numpy()

    with pm.Model() as model_company:

        α = pm.Normal('α', mu = mu_input, sd = sd_input)
        β = pm.Normal('β', mu = mu_input, sd = sd_input, shape = 5)
        δ = pm.HalfNormal('δ', sd = 0.1)
        volatility_shared = pm.Data('volatility_shared', volatility)
```

```
        σ = δ * volatility
        γ = pm.Gamma('γ', 2, 0.1)
        x_shared = pm.Data('x_shared', x)
        μ = α + pm.math.dot(x_shared, β)
        y_pred = pm.StudentT('y_pred', mu = μ, sd = σ, nu = γ, observed = y)

    return model_company
```

### 3.3 Построение моделей

Далее автором будут рассмотрены результаты применения моделей только для бумаг компании «Новатэк», результаты трассировки, статистические характеристики и прогнозируемые распределения данных для акций всех остальных компаний, включенных в индекс Нефти и Газа Московской биржи, можно найти в Приложении 2.

При построении и трассировке модели 1, использующей распределение Стьюдента, для акций компании «Новатэк» были получены следующие статистические характеристики параметров Рис. 11, а также графики плотностей распределения параметров Рис. 12.

| | mean | sd | hdi_3% | hdi_97% | mcse_mean | mcse_sd | ess_bulk | ess_tail | r_hat |
|---|---|---|---|---|---|---|---|---|---|
| **α** | 0.001 | 0.001 | -0.001 | 0.003 | 0.000 | 0.000 | 14469.0 | 9371.0 | 1.0 |
| **β[0]** | 0.097 | 0.010 | 0.079 | 0.115 | 0.000 | 0.000 | 16284.0 | 10014.0 | 1.0 |
| **β[1]** | 0.110 | 0.009 | 0.093 | 0.125 | 0.000 | 0.000 | 14927.0 | 9580.0 | 1.0 |
| **β[2]** | 0.179 | 0.010 | 0.161 | 0.197 | 0.000 | 0.000 | 14768.0 | 10069.0 | 1.0 |
| **β[3]** | 0.102 | 0.010 | 0.085 | 0.121 | 0.000 | 0.000 | 14197.0 | 9517.0 | 1.0 |
| **β[4]** | 0.130 | 0.009 | 0.112 | 0.147 | 0.000 | 0.000 | 15256.0 | 9524.0 | 1.0 |
| **σ** | 0.023 | 0.001 | 0.022 | 0.025 | 0.000 | 0.000 | 10418.0 | 8866.0 | 1.0 |
| **γ** | 4.183 | 0.553 | 3.196 | 5.202 | 0.005 | 0.004 | 11388.0 | 8916.0 | 1.0 |

Рисунок 11

На Рис. 11 можно увидеть: средние значения параметров модели (mean), стандартные отклонения (sd), а также границы 94% доверительного интервала для каждого из параметров. Для данной модели, параметр нормальности распределения $\gamma$ приблизительно равен 4, а это означает, что распределение отличается от нормального и имеет тяжелые хвосты, что связано с большим разбросом данных, чем при нормальном распределении.

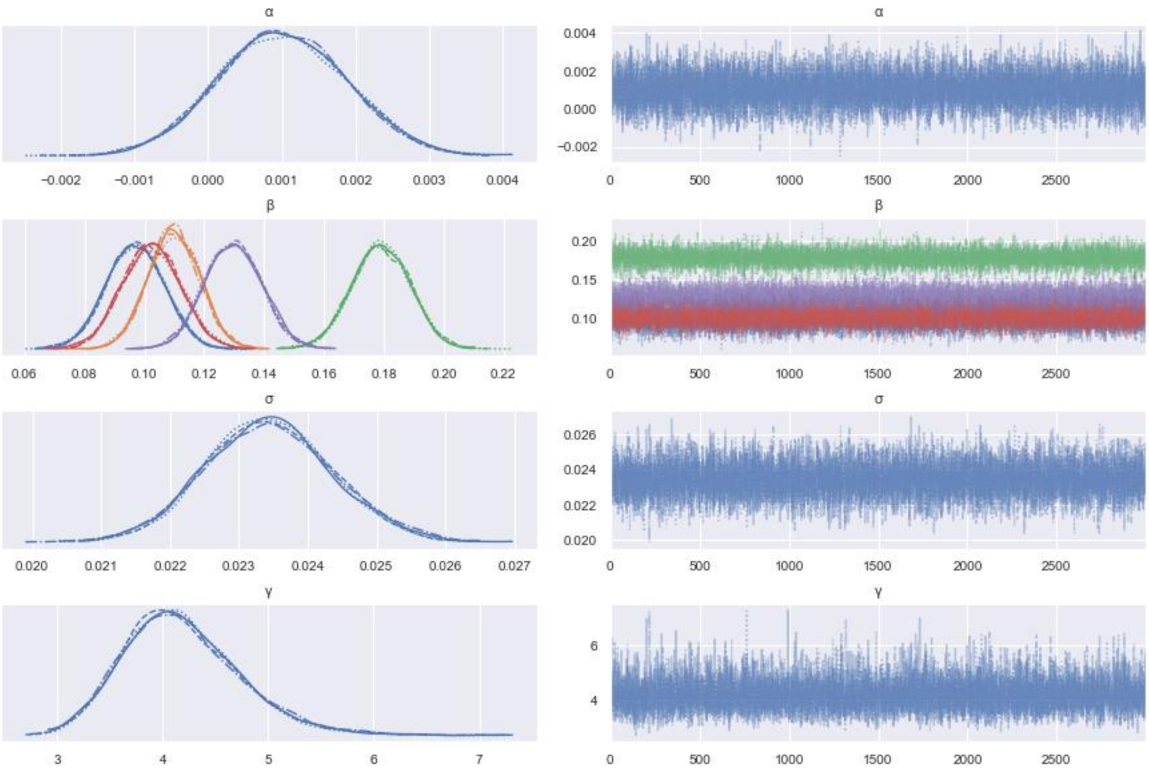



Рисунок 12

После определения и обучения модели были взяты выборки из полученного апостериорного распределения, они изображены на Рис. 13. Было сгенерировано 1500 выборок, которые обеспечили такой широкий разброс для «колокола» распределения. По рисунку видно, что полученное апостериорное распределение достаточно хорошо описывает наши данные.

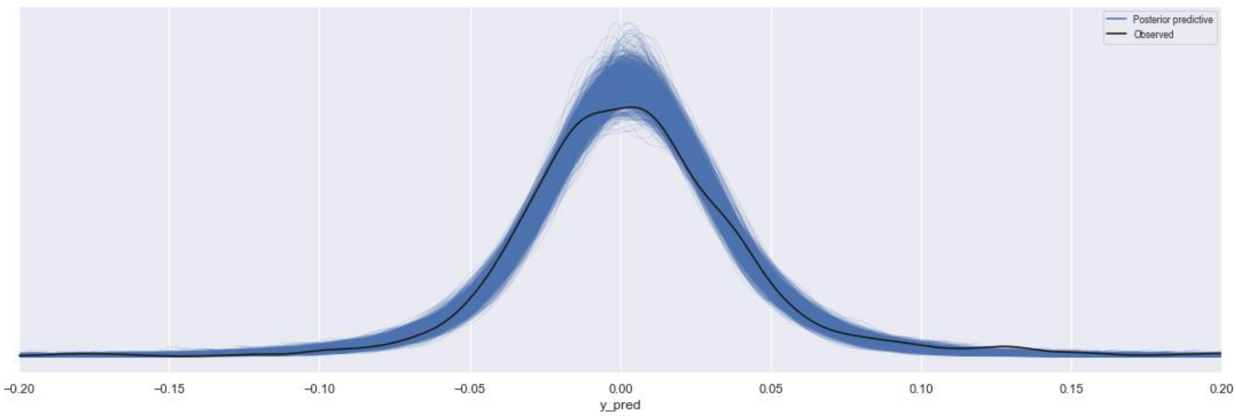



Рисунок 13

При применении модели, использующей нормальное распределение для прогнозирования доходности, получаем следующие результаты. На Рис. 14, 15, 16 изображены статистические характеристики параметров, полученные распределения параметров и выборки из апостериорного распределения, соответственно.

| | mean | sd | hdi_3% | hdi_97% | mcse_mean | mcse_sd | ess_bulk | ess_tail | r_hat |
|---|---|---|---|---|---|---|---|---|---|
| **α** | 0.003 | 0.001 | 0.002 | 0.005 | 0.0 | 0.0 | 20761.0 | 9800.0 | 1.0 |
| **β[0]** | 0.097 | 0.010 | 0.078 | 0.115 | 0.0 | 0.0 | 16103.0 | 9036.0 | 1.0 |
| **β[1]** | 0.116 | 0.008 | 0.101 | 0.133 | 0.0 | 0.0 | 17599.0 | 9824.0 | 1.0 |
| **β[2]** | 0.175 | 0.010 | 0.157 | 0.194 | 0.0 | 0.0 | 15719.0 | 9933.0 | 1.0 |
| **β[3]** | 0.100 | 0.010 | 0.082 | 0.119 | 0.0 | 0.0 | 20232.0 | 10198.0 | 1.0 |
| **β[4]** | 0.132 | 0.009 | 0.114 | 0.150 | 0.0 | 0.0 | 18736.0 | 10228.0 | 1.0 |

Рисунок 14

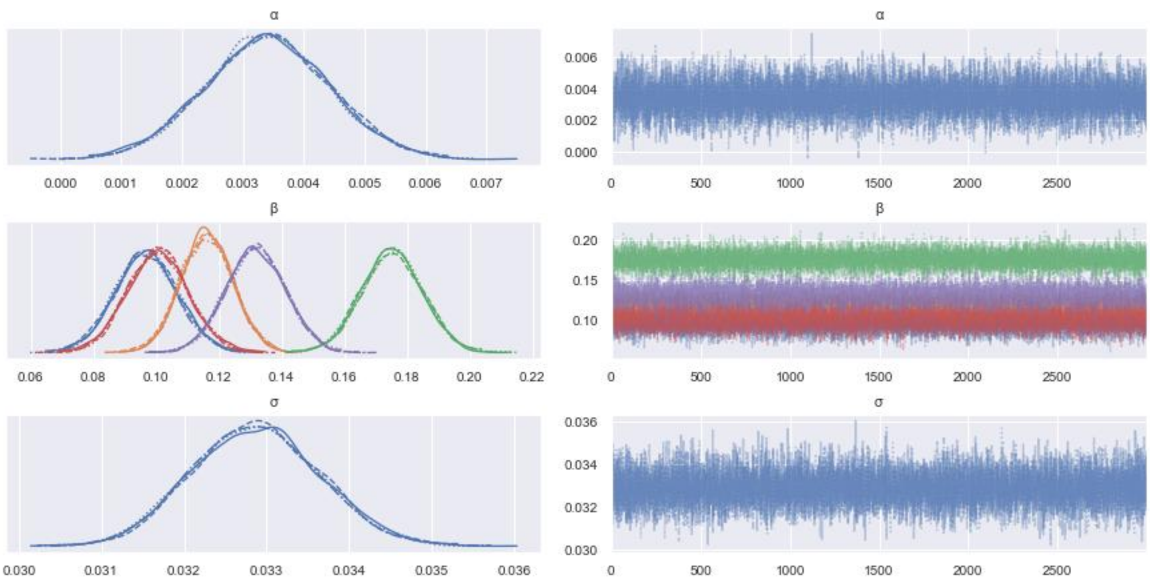



Рисунок 15

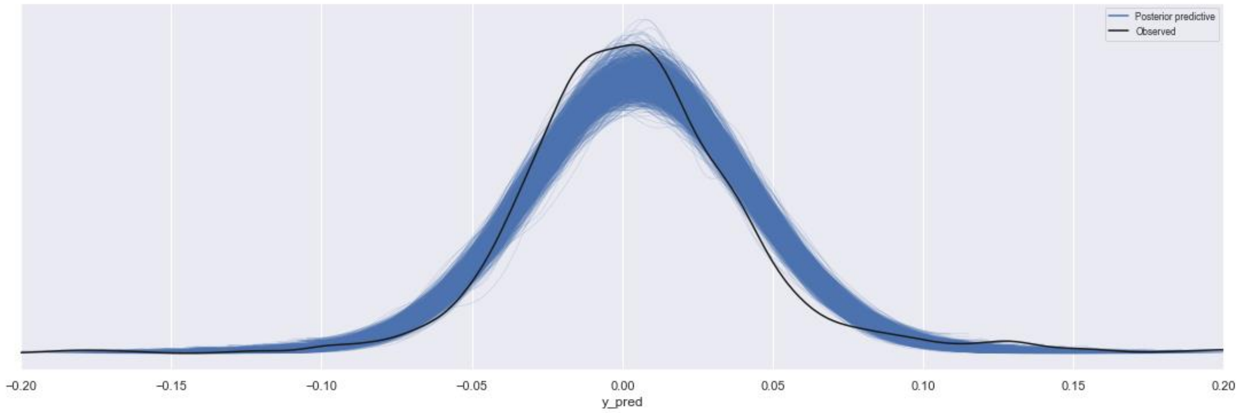



Рисунок 16

Модель, использующая нормальное распределение, незначительно отличается от робастной модели, использующей распределение Стьюдента, однако, это только в случае рассматриваемой ценной бумаги, для некоторых других бумаг, даже исходя из графика апостериорного распределения можно сделать вывод, что нормальное распределение в должной степени не описывает распределение доходности бумаги. Подобные примеры можно найти в Приложении 2.

Статистические характеристики параметров и апостериорное распределение модели, использующей распределение Стьюдента и индекс волатильности, приведены на рисунках 17, 18, 19.

| | mean | sd | hdi_3% | hdi_97% | mcse_mean | mcse_sd | ess_bulk | ess_tail | r_hat |
|---|---|---|---|---|---|---|---|---|---|
| **α** | 0.001 | 0.001 | -0.000 | 0.003 | 0.000 | 0.000 | 13681.0 | 9329.0 | 1.0 |
| **β[0]** | 0.099 | 0.010 | 0.081 | 0.118 | 0.000 | 0.000 | 14987.0 | 8933.0 | 1.0 |
| **β[1]** | 0.107 | 0.009 | 0.091 | 0.124 | 0.000 | 0.000 | 15358.0 | 9751.0 | 1.0 |
| **β[2]** | 0.178 | 0.010 | 0.160 | 0.196 | 0.000 | 0.000 | 14881.0 | 9897.0 | 1.0 |
| **β[3]** | 0.105 | 0.010 | 0.087 | 0.123 | 0.000 | 0.000 | 15491.0 | 9167.0 | 1.0 |
| **β[4]** | 0.124 | 0.009 | 0.106 | 0.142 | 0.000 | 0.000 | 15500.0 | 9351.0 | 1.0 |
| **δ** | 0.001 | 0.000 | 0.001 | 0.001 | 0.000 | 0.000 | 10825.0 | 8379.0 | 1.0 |
| **γ** | 8.698 | 2.167 | 5.340 | 12.670 | 0.023 | 0.017 | 10530.0 | 8111.0 | 1.0 |

Рисунок 17

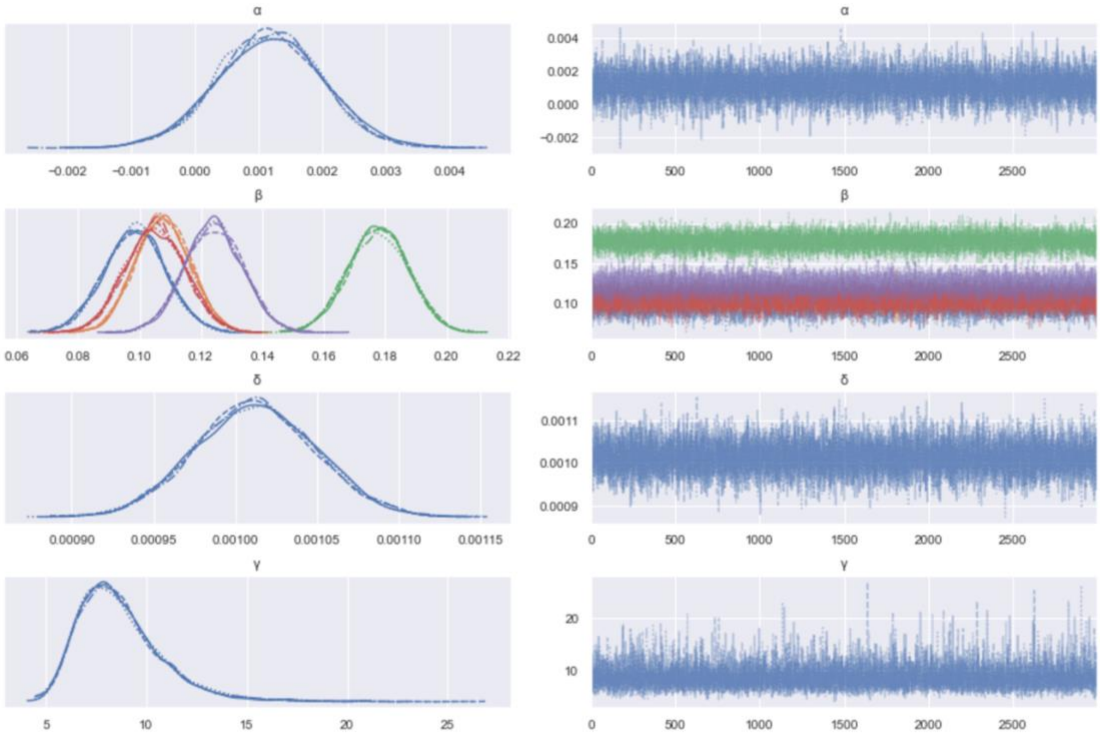



Рисунок 18

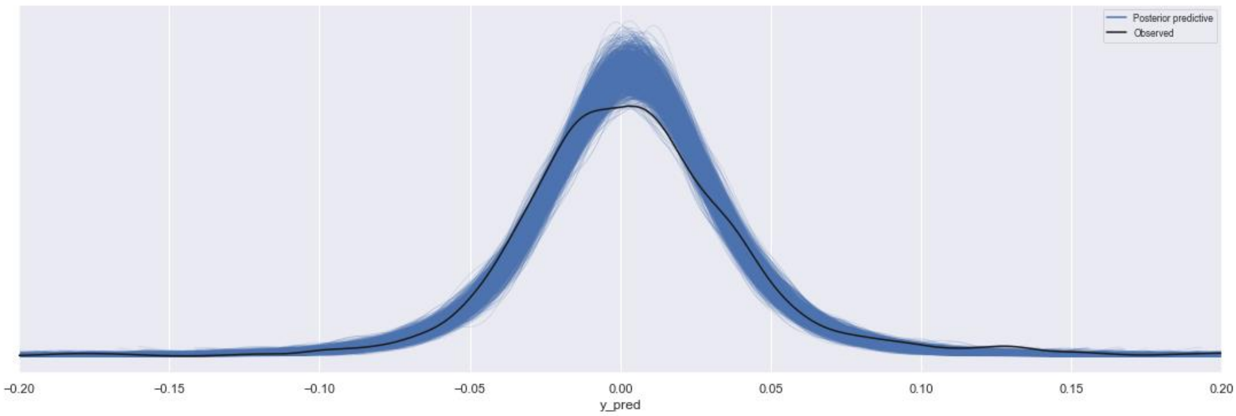



Рисунок 19

Исходя из графиков выборок из апостериорных распределений, более предпочтительными и лучше описывающими данные выглядят модели с распределением Стьюдента. Необходимо сравнить модели с помощью информационных критериев, чтобы сделать обоснованный выбор.

Результаты расчета критерия WAIC приведены на Рис. 20. Значение информационного критерия ниже для модели, использующей распределение Стьюдента и индекс волатильности, но при расчете критерия для данной модели, как и для модели с нормальным распределением, были вычислительные трудности, на что указывает значение True в колонке warning. Поэтому были проведены расчеты с использованием метода парето-сглаженной выборки по значимости для перекрестной проверки. Результаты представлены на Рис. 21.

| | rank | waic | p_waic | d_waic | weight | se | dse | warning | waic_scale |
|---|---|---|---|---|---|---|---|---|---|
| **volatility** | 0 | -4453.699279 | 3.865395 | 0.000000 | 0.928286 | 58.154396 | 0.000000 | True | deviance |
| **student** | 1 | -4343.022437 | 4.102836 | 110.676842 | 0.071714 | 64.251986 | 27.449199 | False | deviance |
| **gauss** | 2 | -4168.315706 | 7.400202 | 285.383573 | 0.000000 | 98.055278 | 74.450602 | True | deviance |

Рисунок 20

| | rank | loo | p_loo | d_loo | weight | se | dse | warning | loo_scale |
|---|---|---|---|---|---|---|---|---|---|
| **volatility** | 0 | -4453.661416 | 3.884326 | 0.000000 | 0.928258 | 58.163093 | 0.00000 | False | deviance |
| **student** | 1 | -4343.017371 | 4.105369 | 110.644045 | 0.071742 | 64.253222 | 27.46697 | False | deviance |
| **gauss** | 2 | -4168.276887 | 7.419612 | 285.384529 | 0.000000 | 98.082533 | 74.48773 | False | deviance |

Рисунок 21

Ошибок в вычислениях не произошло, и по данному показателю видно, что наилучшей является робастная модель, использующая индекс волатильности. Эту модель автор и будет использовать для дальнейшей работы.

### 3.4 Симуляция работы модели

После подгонки модели происходит задание переменных *x_shared* и *volatility_shared,* которые обеспечивают прогнозирование доходностей при условии полученного апостериорного распределения и последних, доступных на момент прогнозирования значений параметров.

Получив прогнозируемую выборку из апостериорного распределения, при условии заданных параметров, можно взять среднее данной выборки. Полученное среднее и будет считаться ожидаемой доходностью ценной бумаги в следующем моменте времени.

Сформировав ожидаемые доходности для каждой из бумаг, можно перейти к расчету весов для портфеля, оптимизированного методом Марковица.

Для того, чтобы проверить насколько эффективен данный метод определения ожидаемых доходностей проведем симуляцию. Для этого из доступного массива данных выбирается часть, в данном случае массив состоит из 1046 наблюдений и были выбраны первые 500. Затем происходит подгонка модели с использованием только этих 500 точек, после чего, с помощью подогнанной модели строятся прогнозы на 5 периодов вперед, то есть на одну неделю. Каждые 10 недель, или 50 периодов, к модели добавляются данные за эти 50 периодов и происходит переподгонка. Так продолжается до тех пор, пока не будет исчерпан весь датасет. Программный код, необходимый для реализации данной процедуры, представлен в Приложении 3.

В результате вышеописанных действий получается таблица прогнозов доходностей, которые могут быть использованы для расчета весов инвестиционного портфеля. Расчет производился с использованием оптимизационного пакета библиотеки *SciPy*. Веса для портфелей определялись по 3 методам:

- Максимизация коэффициента Шарпа;
- Максимизация доходности, при заданном максимальном уровне риска в 0.005;
- Минимизация риска, при заданном минимальном уровне доходности в 0.003.

Короткие продажи разрешены. Рассчитывалась накопленная доходность инвестиций, а не период к периоду. Была применена упрощенная методика расчета ребалансировки активов в портфеле, то есть при каждой ребалансировке считалось, что портфель полностью распродается и заново пересобирается. Данная методика не является достаточно объективной, так как в данном случае могу возникнуть высокие транзакционные издержки.

В результате данной симуляции были получены три временных ряда доходностей портфелей, сформированных различными методами. Они сравнивались с так называемым рыночным портфелем, роль которого выполнял индекс Нефти и Газа Московской биржи. Результаты можно видеть на Рис. 22.

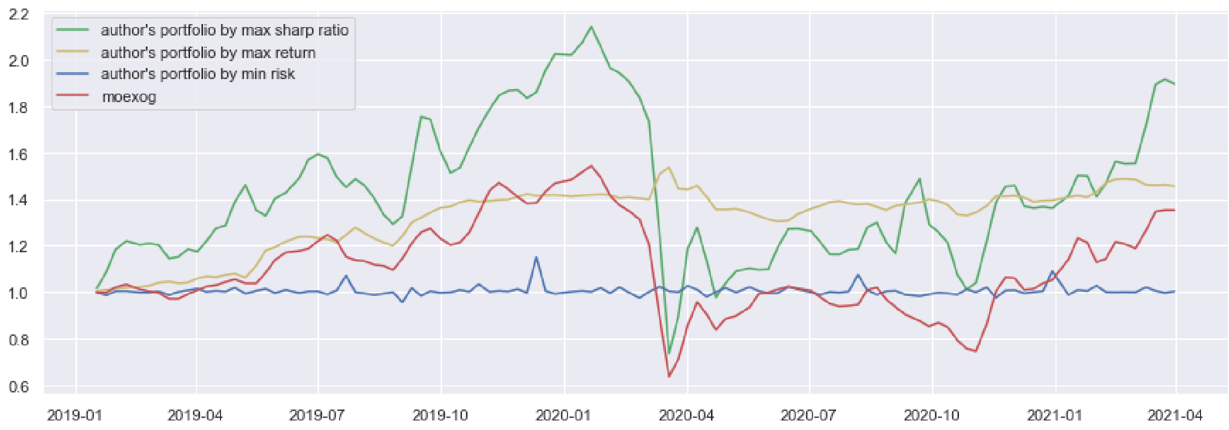



Рисунок 22

Зеленым цветом показана накопленная доходность портфеля, сформированного и ребалансировавшегося с помощью максимизации коэффициента Шарпа. Желтым цветом показана доходность портфеля, сформированного методом максимизации доходности при заданном целевом уровне риска. Синим цветом показана доходность портфеля, сформированного методом минимизации риска, при заданном целевом уровне доходности. Красным цветом показана накопленная доходность индекса MOEXOG за этот же период.

Исходя из полученных результатов можно сделать вывод, что с точки зрения максимизации доходности, наиболее эффективным методом формирования портфеля является максимизация коэффициента Шарпа. Метод минимизации риска при заданном уровне доходности дает относительно низкорискованный, но и низкодоходный портфель. В целом, построение портфеля на основании ожидаемых доходностей, спрогнозированных с помощью байесовского моделирования, показало результат лучше рыночного. Портфель, сформированный методом максимизации коэффициента Шарпа, на протяжении всего периода оказался более высокодоходным, нежели чем отраслевой индекс.

## Заключение

В данной работе были рассмотрены основные аспекты современной портфельной теории, в частности, различные методики формирования инвестиционных портфелей путем подбора весов для составляющих их активов. Также были рассмотрены базовые методы байесовского анализы, определены и построены несколько конкурирующих моделей для моделирования недельных доходностей бумаг нефтегазового сектора московской биржи. В результате симуляционного моделирования на исторических данных, было выявлено, что прогнозы, сделанные с помощью байесовских методов, оказываются достаточно эффективным инструментом для формирования инвестиционного портфеля. Портфель, сформированный с помощью максимизации коэффициента Шарпа, показал более привлекательную динамику, нежели чем отраслевой индекс нефти и газа, в моменты роста рынка он имел более взрывной рост, а в моменты падения оказывался более устойчив, по сравнению с рыночным.

В качестве развития данной работы можно применить более инновационные методы формирования инвестиционного портфеля, например, модель Блэка-Литтермана. Также можно расширить модель, анализируя бумаги, входящие в индекс МосБиржи, и определив новые параметры модели. Еще одним вариантом развития работы является более обоснованное моделирование волатильности доходностей ценных бумаг. Возможно, эффективным методом окажется связка авторегрессионных моделей или моделей для нестационарных временных рядов, байесовских моделей и моделей машинного обучения.

## Список литературы


1. Мартин О. Байесовский анализ на Python/ пер. с анг. А.В. Снастина. – М.:ДМК Пресс, 2020. – 340 с.: ил.
2. Harry Markowitz, The Journal of Finance, Vol. 7, No. 1. (Mar., 1952), pp. 77-91.
3. William F. Sahrpe, The Journal of Portfolio Management Fall 1994, 21 (1) 49-58
4. Fischer Black & Robert Litterman (1992) Global Portfolio Optimization, Financial Analysts Journal, 48:5,28-43
5. Fabozzi, Frank J., et al. Robust portfolio optimization and management. John Wiley & Sons, 2007.
6. Лефевр Э. Воспоминания биржевого спекулянта. – Olympus Business, 2012. – 164 с.
7. Alexander Elder. «Trading for a Living: Psychology, Trading Tactics, Money Management»., 1993.
8. John Murphy. Technical Analysis of the Futures Markets. Prentice Hall Press. March 3, 1986. I
9. Найман Э. Малая энциклопедия трейдера. - Olympus Business, 2008. – 700 с.
10. Берзон Н. И., Аршавский А. Ю., Буянова Е. А. Фондовый рынок. – М.: Вита-Пресс, 1999. – 396 с
11. Abarbanell J. S., Bushee B. J. Fundamental analysis, future earnings, and stock prices // Journal of accounting research. – 1997. – Vol. 35, № 1. – P. 1-24.
12. Achelis S. B. Technical Analysis from A to Z. – New York : McGraw Hill, 2001. – 231 p.
13. Hendershott T., Jones C. M., Menkveld A. J. Does algorithmic trading improve liquidity? // The Journal of Finance. – 2011. – Vol. 66, № 1. – P. 1-33.
14. Трегуб И. В. Моделирование динамики цены биржевых инструментов на российском фондовом рынке методами технического анализа // Лесной вестник/Forestry bulletin. – 2005. – № 3. – С. 156-170.
15. Мицель А. А., Ефремова Е. А. Прогнозирование динамики цен на фондовом рынке // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2006. – Т. 309, № 8. – С. 197-201.

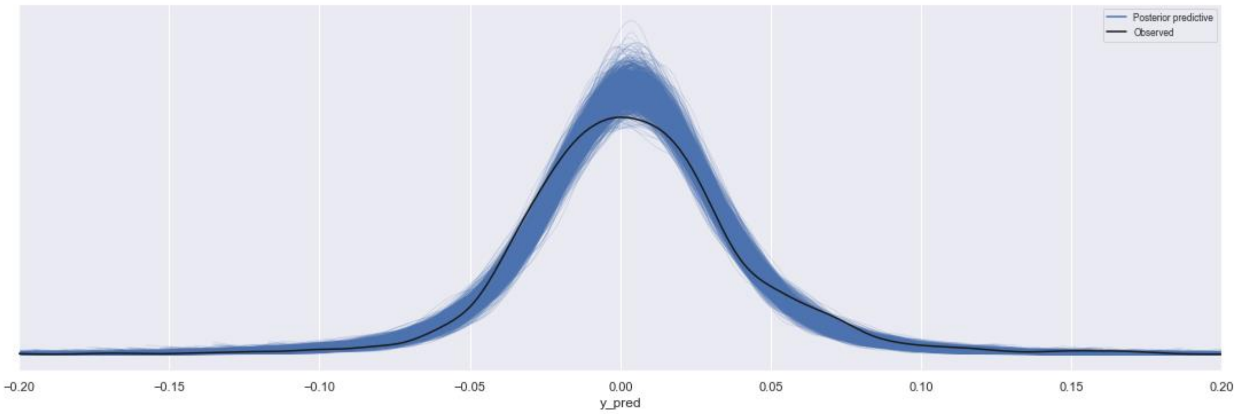
## Приложение 1

| Date | Close_msci | Close_brent | Close_banep | Close_gazp | Close_lkoh | Close_nvtk | Close_rnft | Close_rosn | Close_sngsp | Close_sngs | Close_tatnp | Close_tatn |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2017-01-03 | 622.00 | 57.73 | 1271.0 | 157.70 | 3575.0 | 783.0 | 559.2 | 419.50 | 33.000 | 31.930 | 244.1 | 445.20 |
| 2017-01-04 | 616.25 | 55.59 | 1279.0 | 157.44 | 3479.5 | 775.0 | 559.9 | 406.40 | 32.900 | 31.800 | 244.0 | 446.40 |
| 2017-01-05 | 612.33 | 57.06 | 1274.5 | 154.00 | 3379.0 | 756.4 | 559.9 | 399.00 | 32.250 | 31.595 | 235.8 | 424.25 |
| 2017-01-06 | 610.95 | 56.61 | 1280.5 | 154.10 | 3365.0 | 754.9 | 559.1 | 392.50 | 32.545 | 31.690 | 233.3 | 423.45 |
| 2017-01-09 | 604.17 | 55.65 | 1285.5 | 154.40 | 3384.5 | 755.4 | 559.3 | 394.00 | 32.220 | 31.300 | 235.3 | 425.80 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2021-03-26 | 689.66 | 64.50 | 1296.0 | 227.30 | 6283.0 | 1518.6 | 166.0 | 582.20 | 41.630 | 35.250 | 566.8 | 603.40 |
| 2021-03-29 | 696.50 | 64.60 | 1287.0 | 228.74 | 6257.5 | 1497.6 | 161.0 | 575.30 | 41.710 | 34.885 | 562.8 | 603.00 |
| 2021-03-30 | 693.38 | 64.25 | 1288.0 | 225.93 | 6155.0 | 1484.0 | 155.0 | 570.25 | 41.465 | 34.585 | 560.5 | 596.90 |
| 2021-03-31 | 701.05 | 63.91 | 1299.5 | 227.24 | 6111.5 | 1488.2 | 161.6 | 572.60 | 42.010 | 34.530 | 558.3 | 594.20 |
| 2021-04-01 | 690.02 | 63.60 | 1285.5 | 226.94 | 6112.0 | 1489.0 | 165.8 | 568.90 | 42.445 | 36.070 | 569.2 | 606.00 |

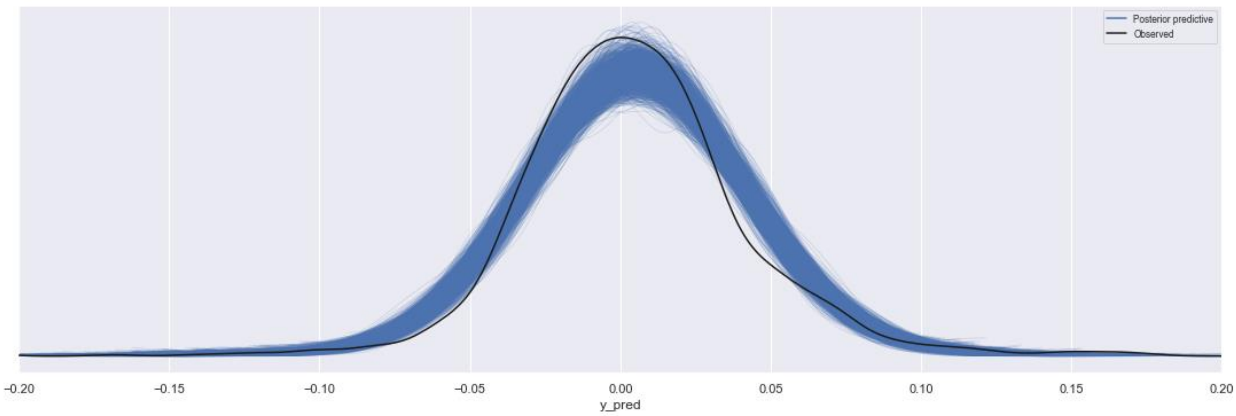
## Приложение 2

Лукойл

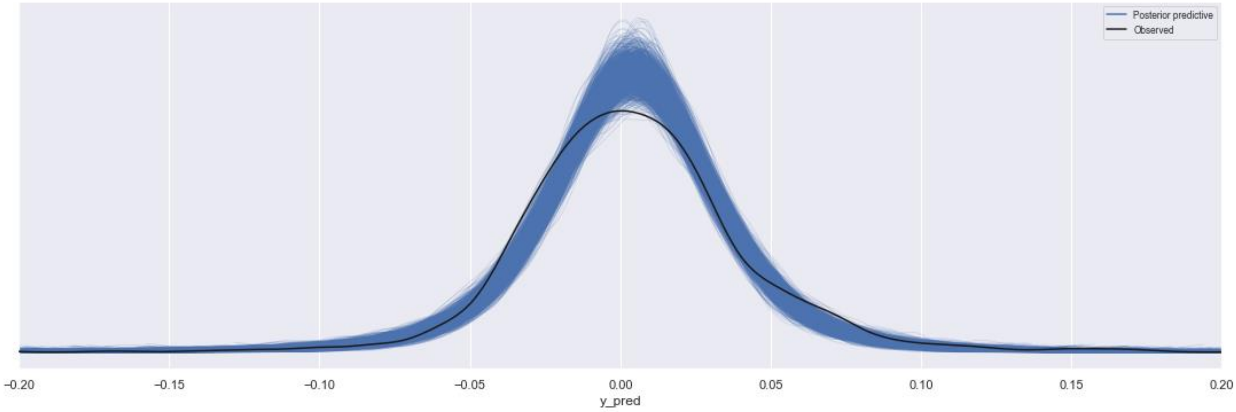
Распределение Стьюдента




Нормальное распределение

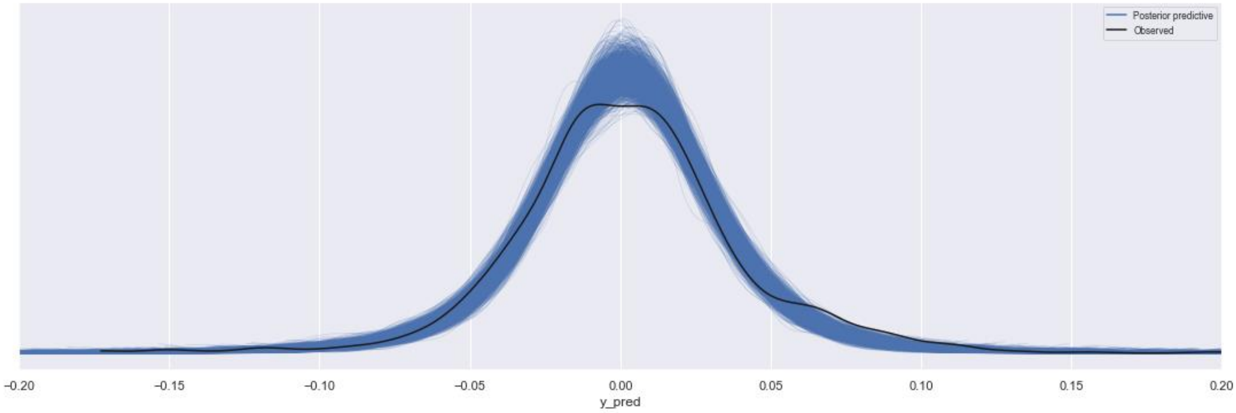


Распределение Стьюдента + волатильность

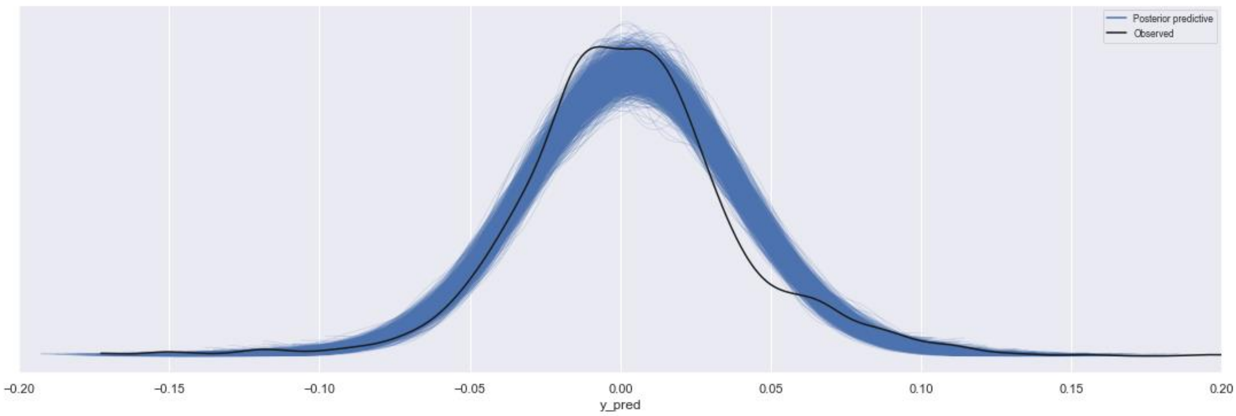



Газпром

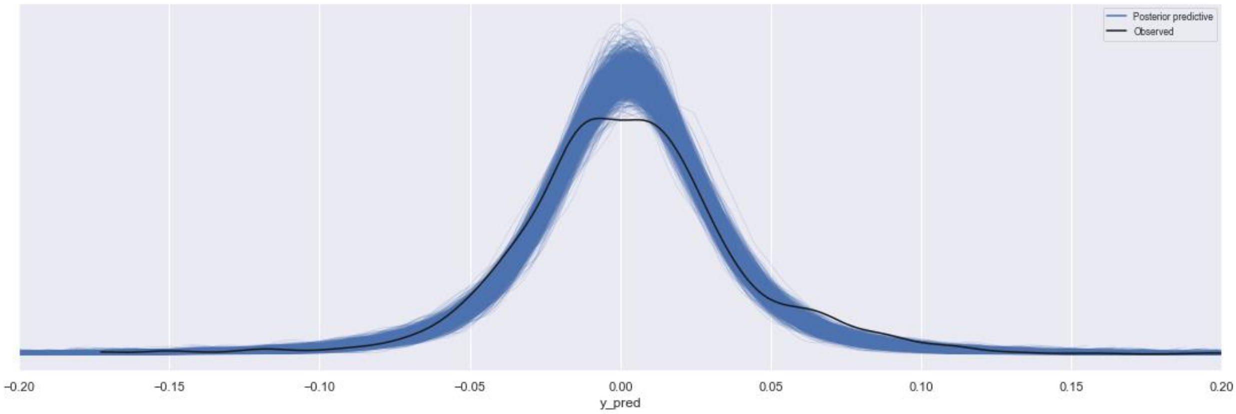
Распределение Стьюдента




Нормальное распределение

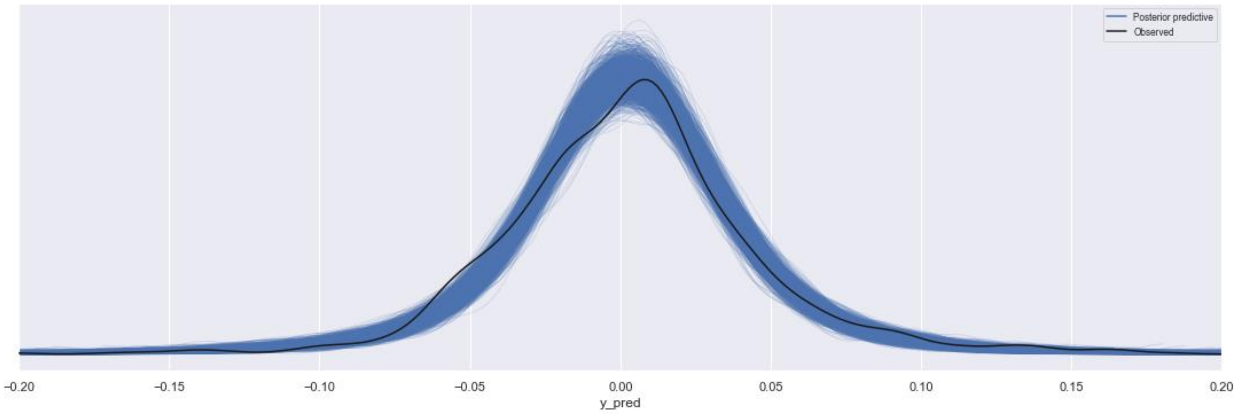



Распределение Стьюдента + волатильность

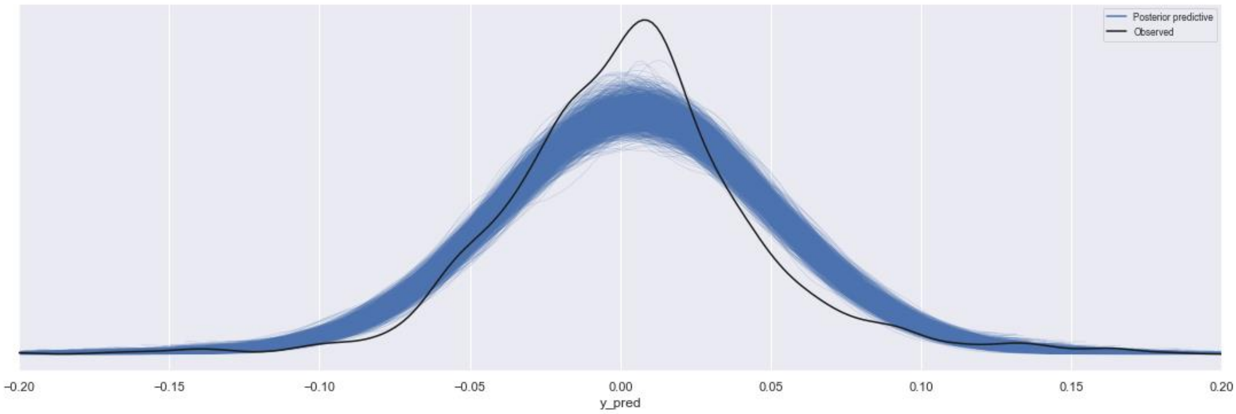



Татнефть

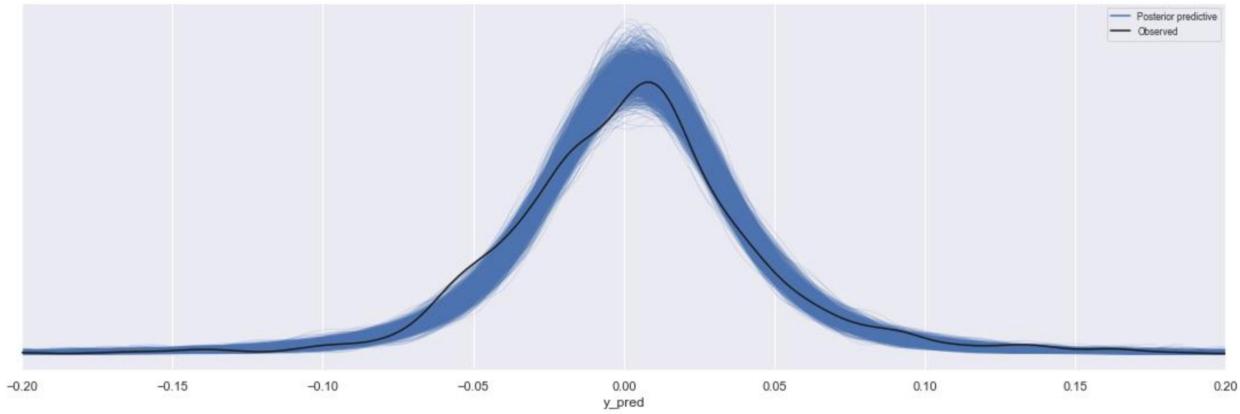
Распределение Стьюдента




Нормальное распределение

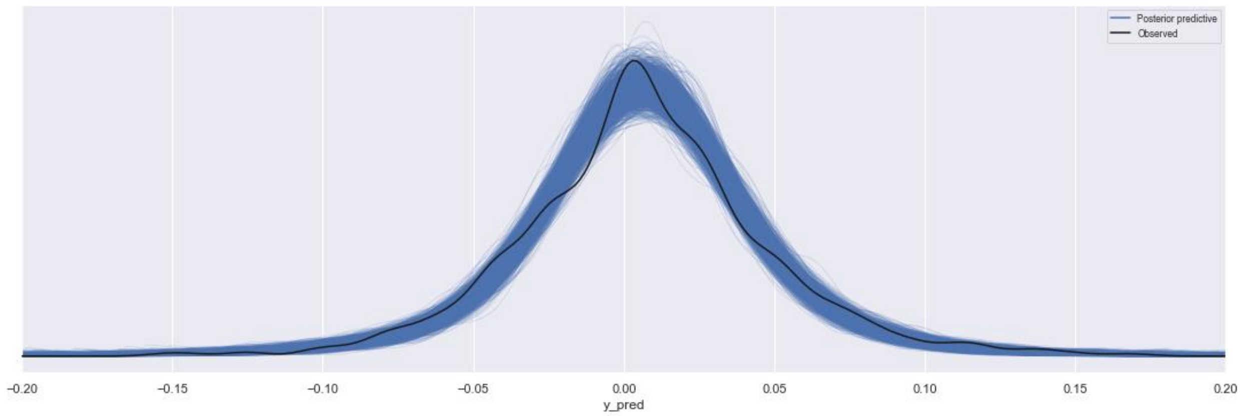



Распределение Стьюдента + волатильность

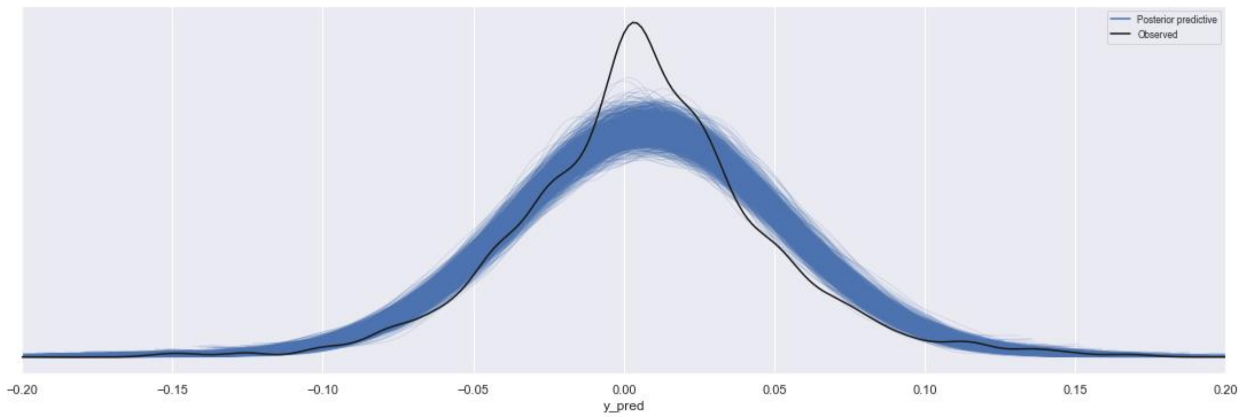



Татнефть-п

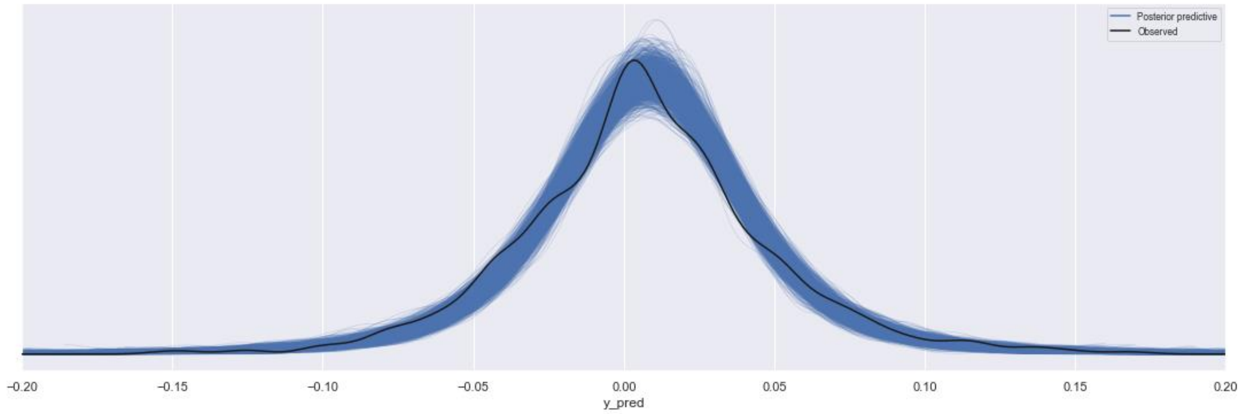
Распределение Стьюдента




Нормальное распределение

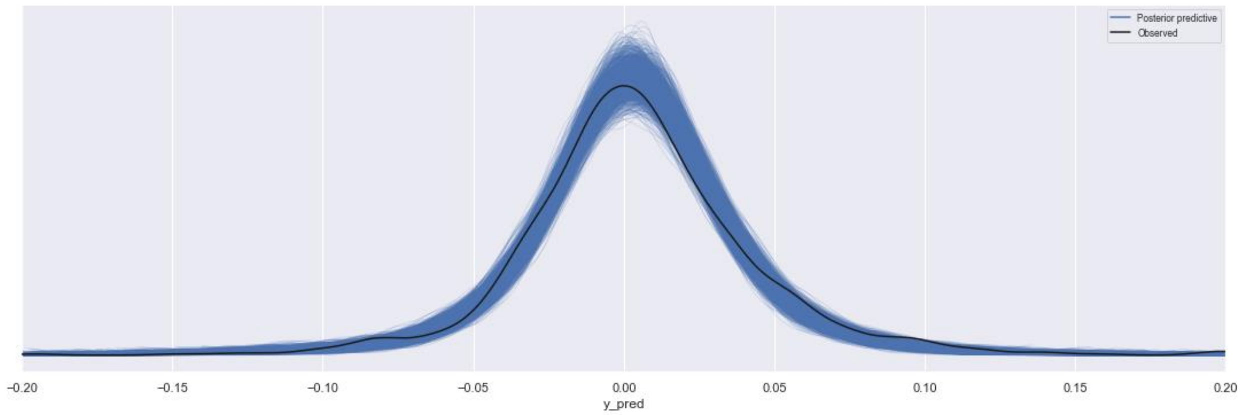



Распределение Стьюдента + волатильность

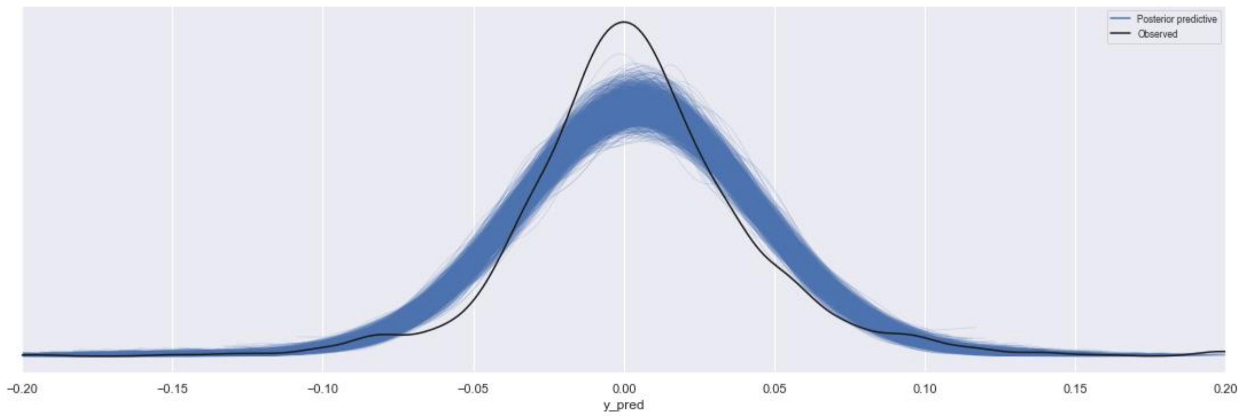



Роснефть

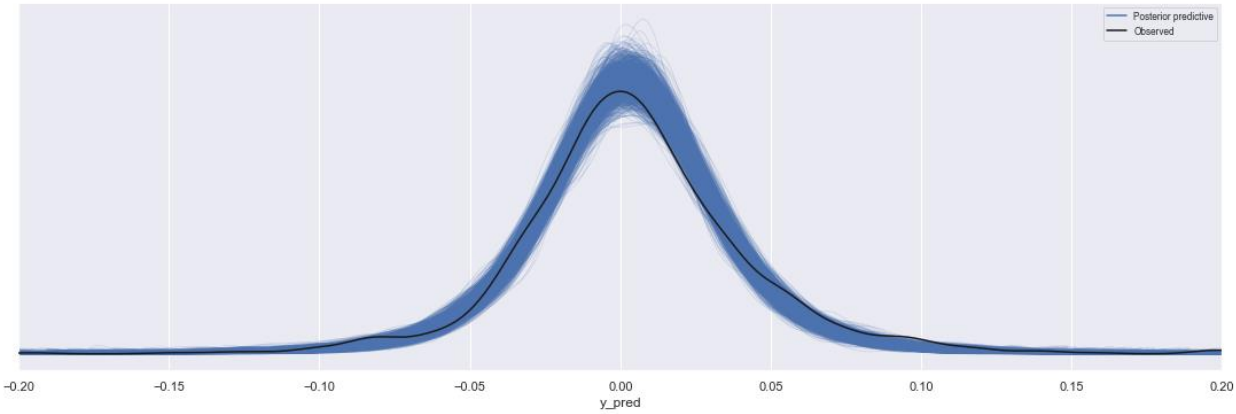
Распределение Стьюдента




Нормальное распределение

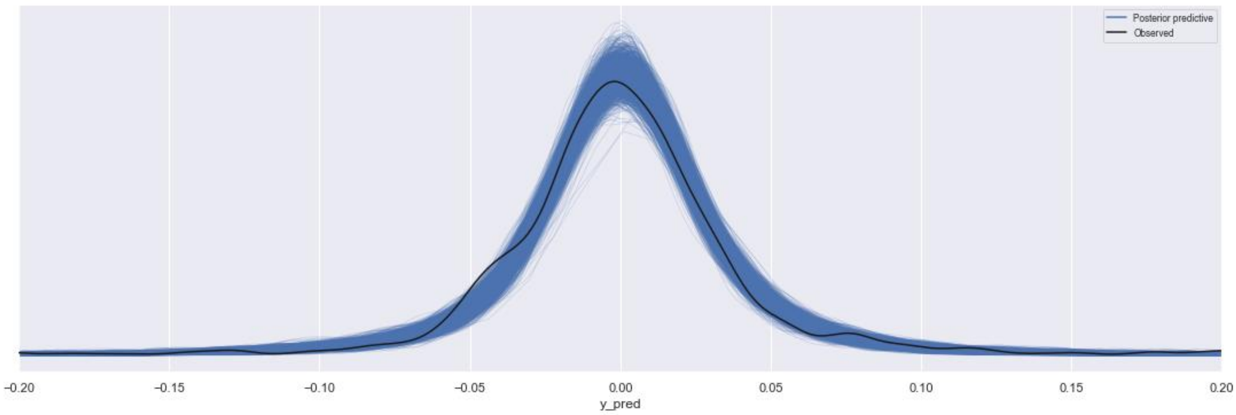



Распределение Стьюдента + волатильность

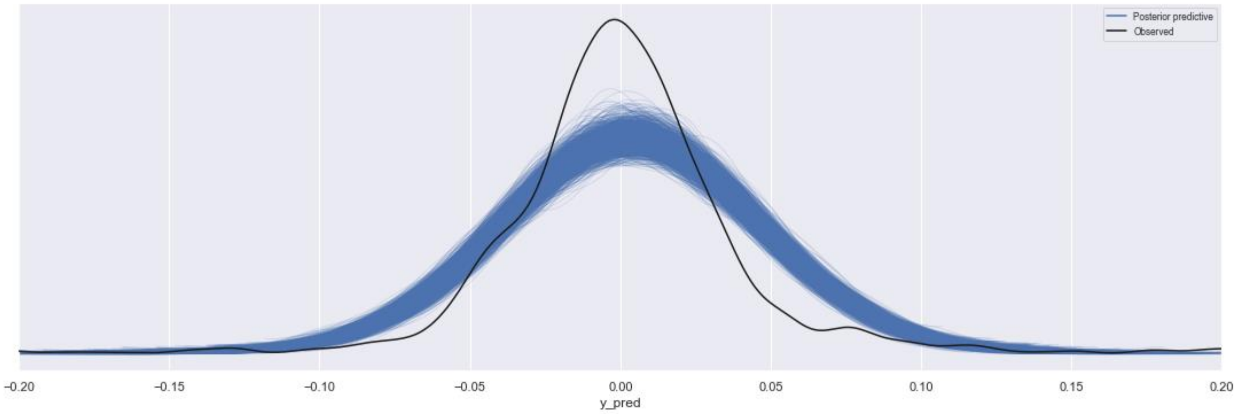



Сургутнефтегаз

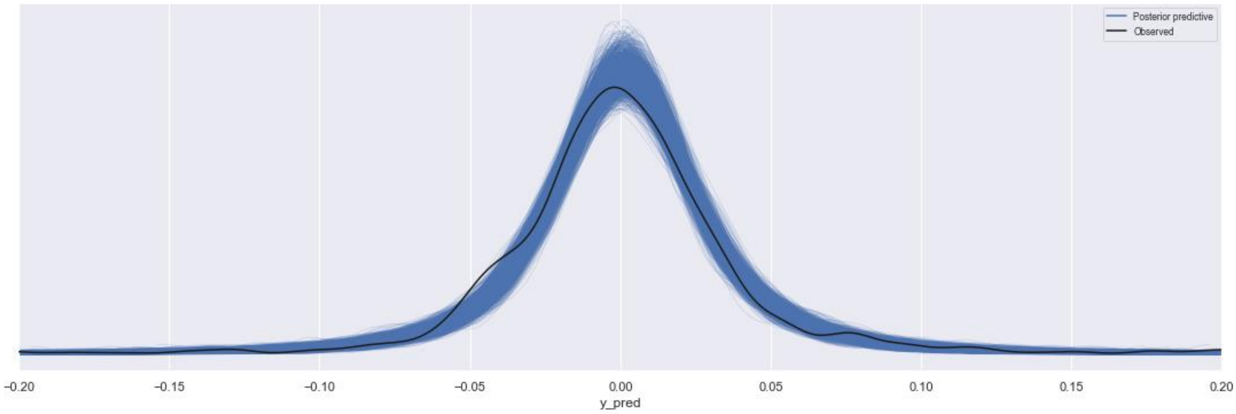
Распределение Стьюдента




Нормальное распределение

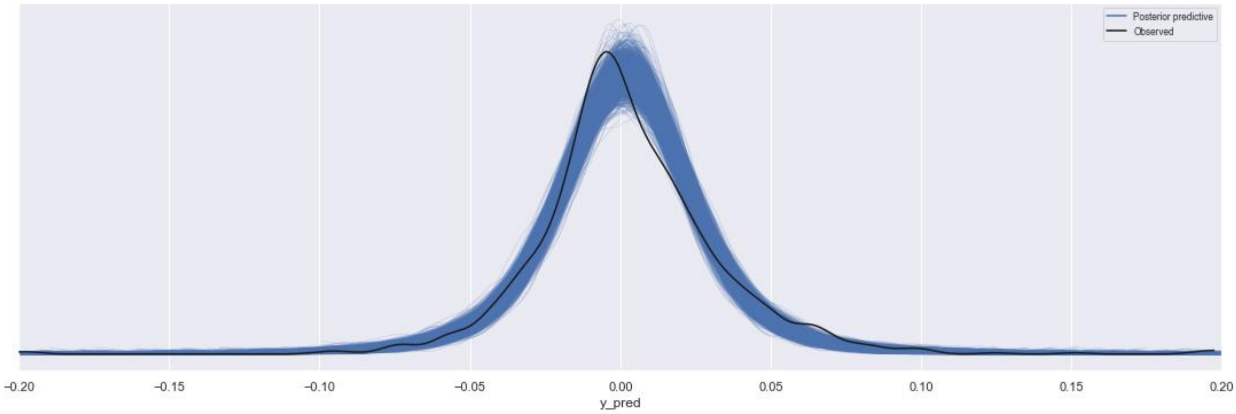



Распределение Стьюдента + волатильность

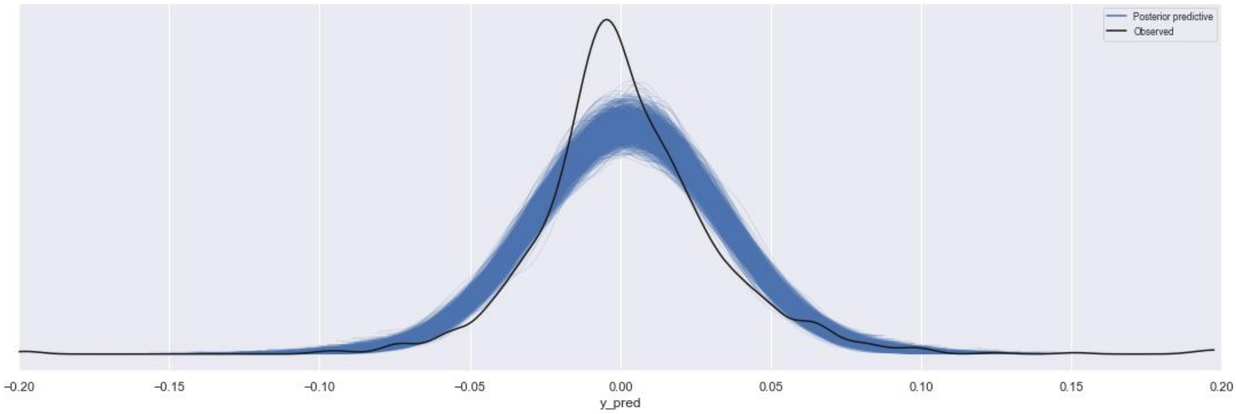



Сургутнефтегаз-п

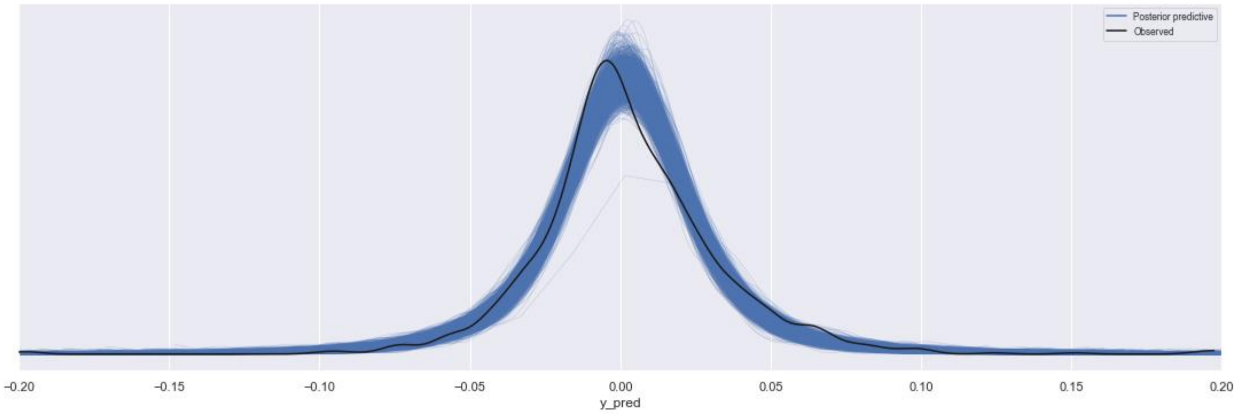
Распределение Стьюдента




Нормальное распределение

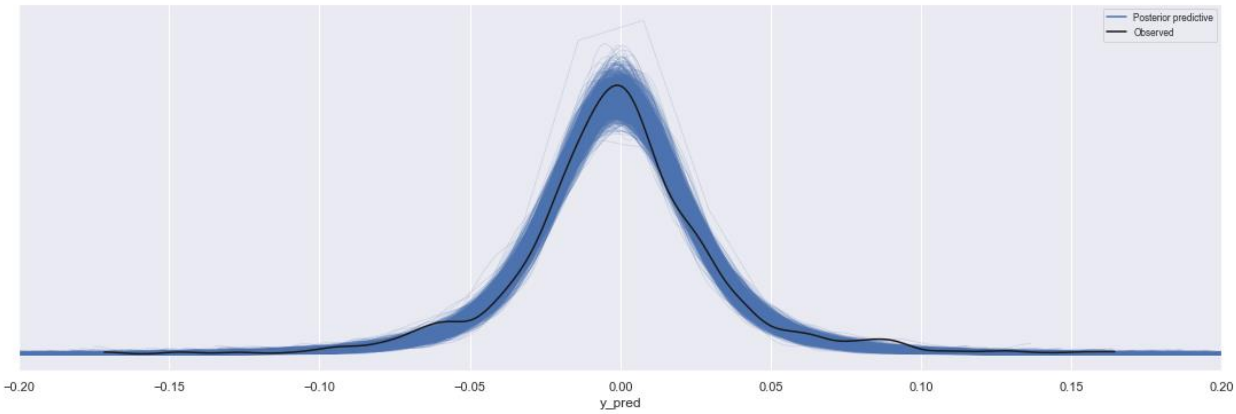



Распределение Стьюдента + волатильность

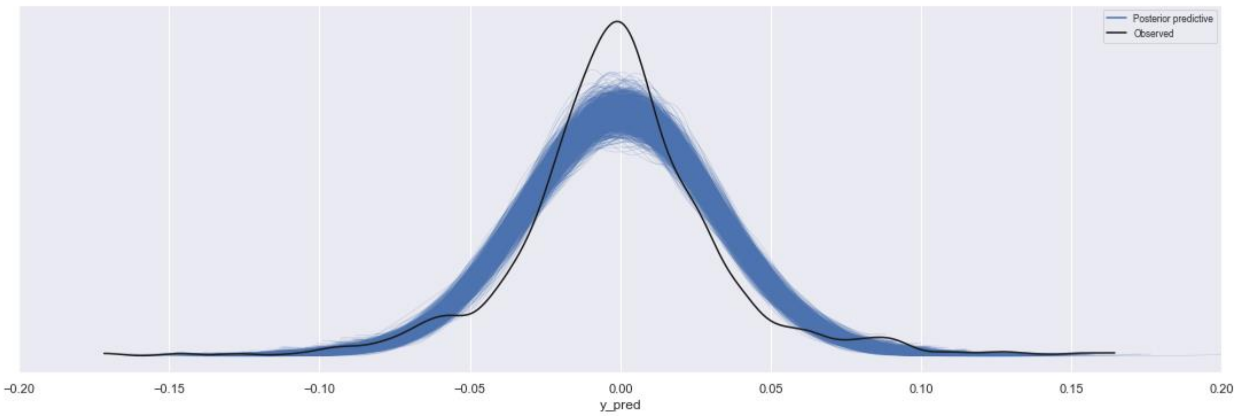



Транснефть-п

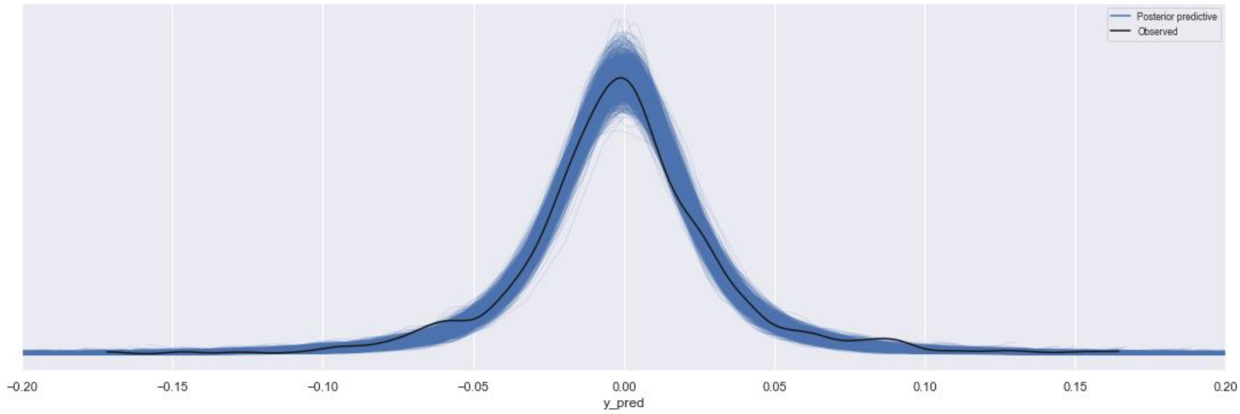
Распределение Стьюдента




Нормальное распределение

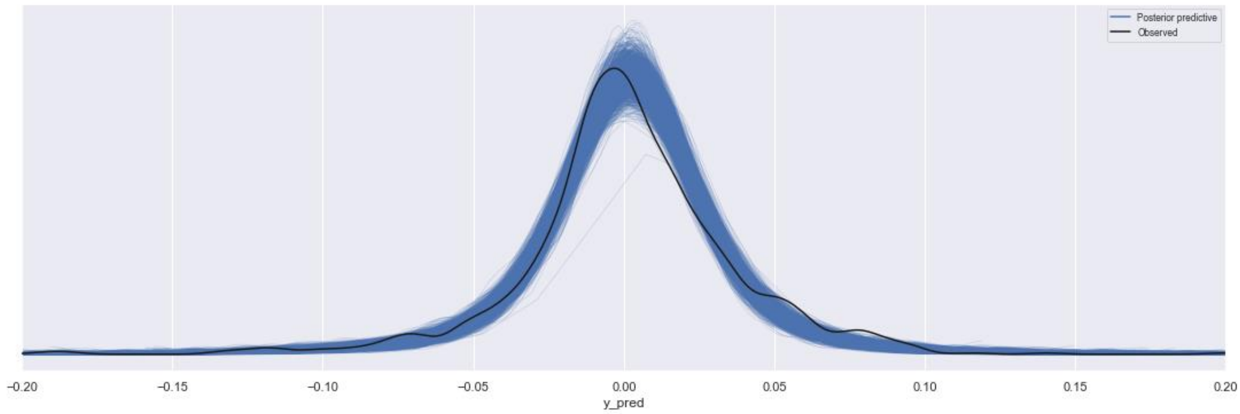



Распределение Стьюдента + волатильность

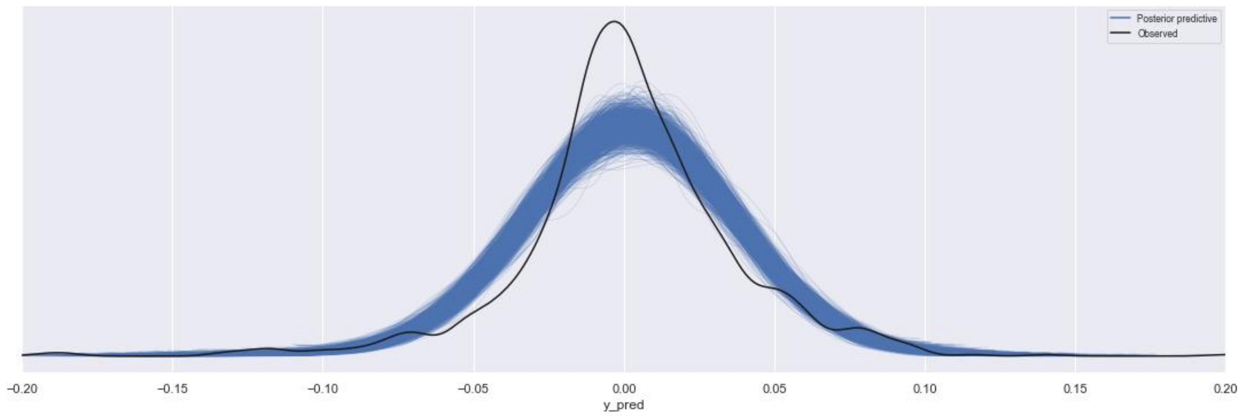



Башнефть

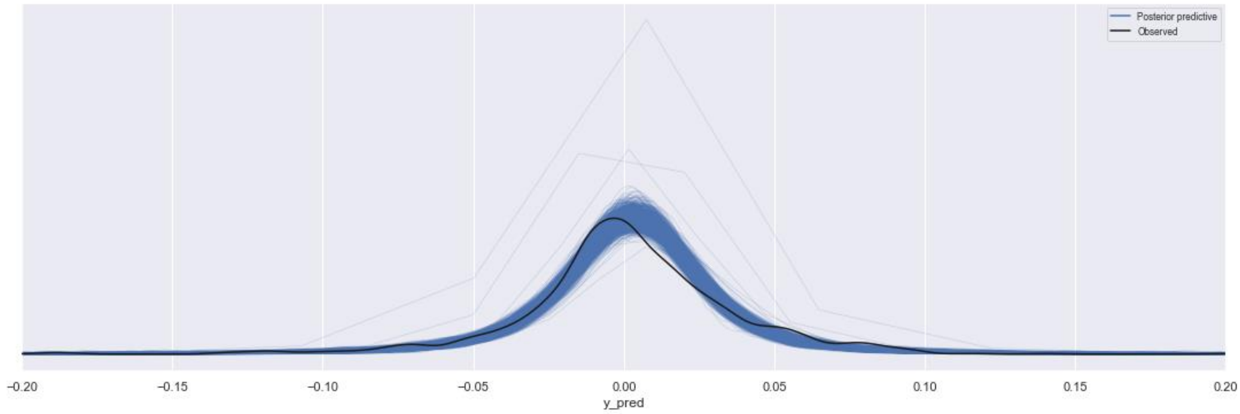
Распределение Стьюдента




Нормальное распределение

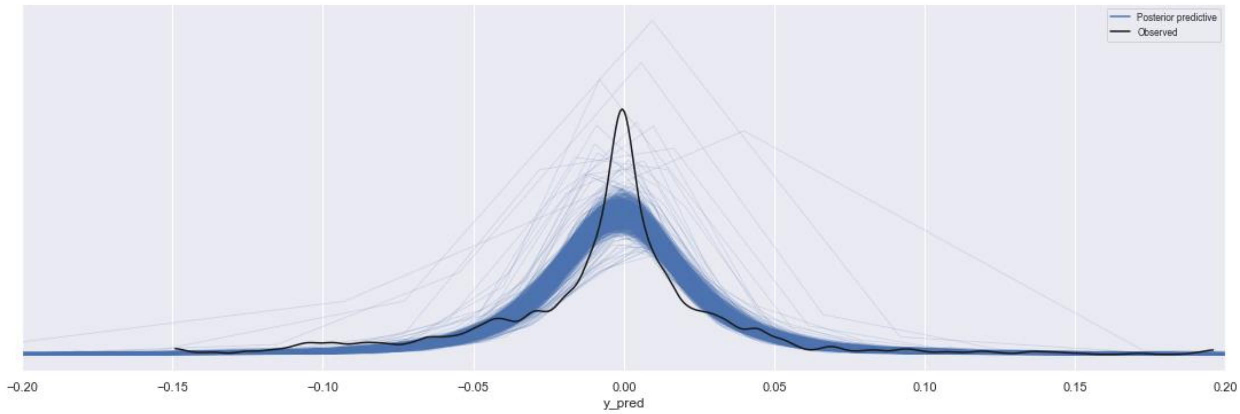



Распределение Стьюдента + волатильность

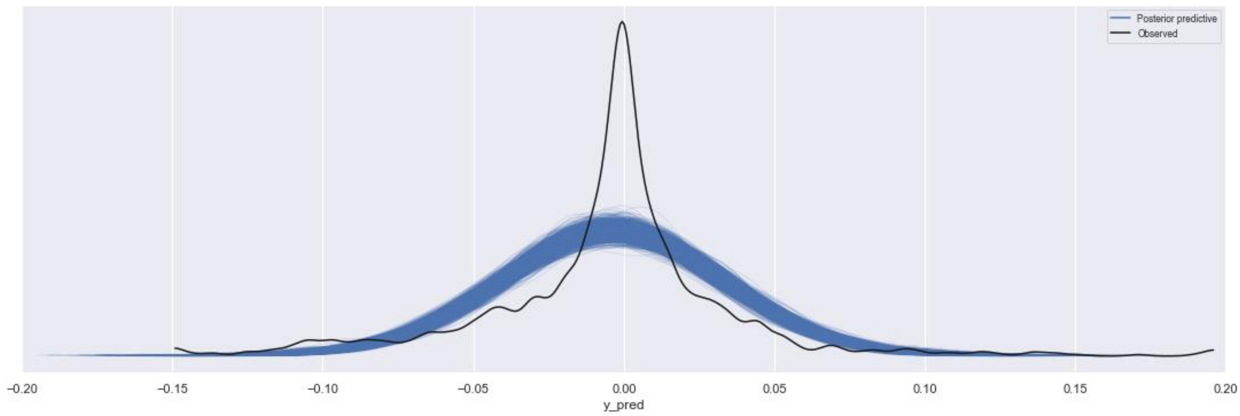



РуссНефть

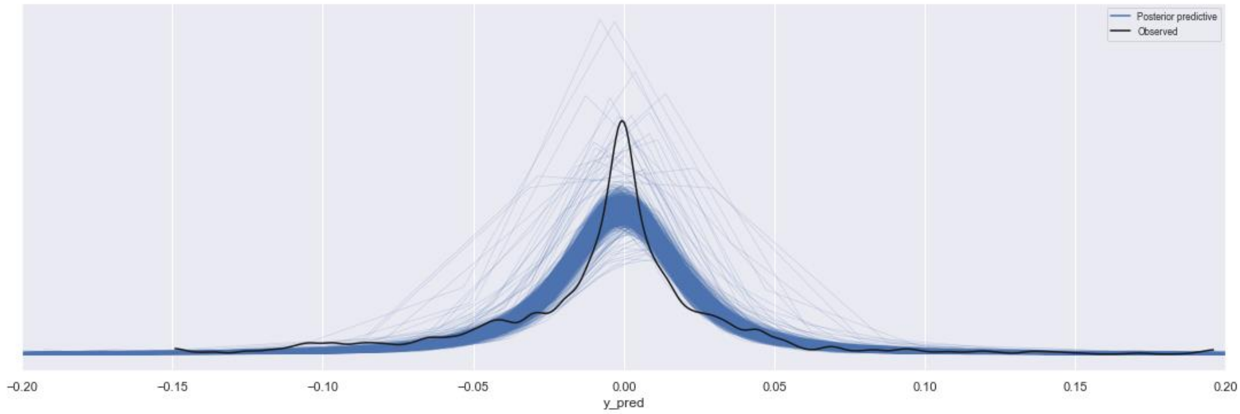
Распределение Стьюдента




Нормальное распределение




Распределение Стьюдента + волатильность




# Приложение 3

```
cols_for_simulation = ('rnft', 'banep', 'trnfp', 'sngsp', 'sngs', 'rosn', 'tatnp', 'tatn',
'gazp', 'lkoh', 'nvtk')

simulated_returns = pd.DataFrame(columns = cols_for_simulation)

for i in range(500, close_prices_for_simulation.shape[0], 50):

    close_prices_for_simulation_i = close_prices_for_simulation.iloc[0 : min(i + 50,
          close_prices_for_simulation.shape[0]), :]
    close_prices_for_simulation_i.index = close_prices_for_simulation_i.index.weekday
    week_name = {0: 'Mon', 1: 'Tue', 2: 'Wed', 3: 'Thu', 4: 'Fri', 5: 'Sat', 6: 'Sun'}
    close_prices_for_simulation_i = close_prices_for_simulation_i.rename(index = week_name)

    yields = pd.DataFrame()

    for j in ['Mon', 'Tue', 'Wed', 'Thu', 'Fri']:

        yields = yields.append(
         close_prices_for_simulation_i[close_prices_for_simulation_i.index == j].apply(lambda x:
       x.pct_change() if x.name !='Close_rvi_t-1' else x))

    dropped_na = yields.isna().sum().max()
    yields.dropna(inplace = True)

    simulated_returns_k = pd.DataFrame(columns = cols_for_simulation)

    for stock in cols_for_simulation:

        model = returns_oil_gas_volatility(stock, data = yields.iloc[0 : i, :])
        container = pd.Series()

        with model:

            trace = pm.sample(1500, tune = 500)

        for k in range(i, min(i + 50, close_prices_for_simulation.shape[0]), 5):

            with model:

                pm.set_data({'x_shared': [yields.loc[:,['Close_usdrub_t-1' ,
                   'Close_brent_t-1', 'Close_nvtk_t-1',
                   'MA_nvtk_t-1', 'Close_msci_t-1']].iloc[k - 1 - dropped_na, :].to_numpy()],
                'volatility_shared': [yields.loc[:, 'Close_rvi_t-1'][k - 1 - dropped_na]]})
                pred = pm.sample_posterior_predictive(trace, samples = 1000)

            pred_mean = pred['y_pred'].mean()
            container = container.append(pd.Series(pred_mean))

        simulated_returns_k[stock] = container

    simulated_returns = simulated_returns.append(simulated_returns_k)
    print('Сделана запись №', simulated_returns.shape[0])

simulated_returns.to_csv('simulated_returns_new.csv')
```