

Сравнительный анализ различных методов моделирования риска и доходности портфолио для российских акций.

И.О.Панин¹

¹Московский физико-технический институт (государственный университет)

(предварительная версия)

Вступление.

В данной работе рассматриваются несколько различных моделей составления портфолио. В ходе работы проводится моделирование дневных риска и доходности, а также, проводится оценка точности различных моделей. В первой части приводится описание использованных данных. Во второй рассматриваются способы вычисления и очистки ковариационных матриц. В третьей части рассматривается прогнозирование риска и доходности для различных моделей, также, приводится оценка точности на основе получаемых реальных распределений риска-доходности, на основе чего выбирается наиболее точная модель. В четвертой части для наиболее точной модели вычисляются дневные значения VaR ряда портфолио. В пятой части проводится анализ отклонений от модели.

1. Данные.

Для анализа были выбраны акции компаний, удовлетворяющие двум критериям: суточный объем торгов превышает 10 млн. рублей на конец ноября 2018 года и акции торгуются на ММВБ с начала 2010 года или ранее. Были загружены данные по с периодом 1 день из архива на сайте ФИНАМ. Далее, из анализа исключена компания "Интер РАО ЕЭС" поскольку визуальный осмотр выявил несоответствие загруженных данных и графика, представленного на сайте ММВБ. Сомнительным также является включение префов транснефти по- причине высокой цены акции(стоимость одной акции 170 тыс. рублей). Итого имеем 39 инструментов, данные по которым приведены в следующей таблице.

Инструмент	Наблюдений	Инструмент	Наблюдений	Инструмент	Наблюдений
AFKS	2226	MGNT	2226	RTKM	2217
AFLT	2226	MRKP	2222	RTKMP	2217
AKRN	2226	MSNG	2226	SBER	2226
CHMF	2226	MTLR	2226	SBERP	2226
ENRU	2226	MTSS	2217	SIBN	2226
FEES	2226	MVID	2226	SNGS	2226
GAZP	2226	NVTK	2226	SNGSP	2226
GMKN	2226	OGKB	2226	TATN	2226
HYDR	2226	PIKK	2226	TATNP	2226
IRKT	2225	PLZL	2226	TRMK	2226

LKOH	2226	RASP	2226	TRNFP	2226
LSRG	2226	ROSN	2226	UPRO	1960
MAGN	2226	RSTI	2226	VTBR	2226

Видно, что по некоторым инструментам меньше данных, чем по остальным. Возможно несколько решений данной проблемы. Вариант 1, вычислять корреляционную матрицу, используя в каждом случае наибольшее возможное число общих данных. В таком случае, правда, полученная матрица может оказаться не положительно определенной, тогда необходимо использовать метод поиска ближайшей положительно определенной матрицы см. [Higham 2002](#). Или можно оставить в рассмотрении только общие для всех инструментов наблюдения. Мы выбираем второй способ, так как задача заключается в прогнозировании суточных значений риска и доходности. Код находится в файле “Select common parts.R”. В итоге получаем 38 временных рядов по 2203 значения.

1.1 Дивидендная доходность.

Не секрет, что на Российском рынке доход от курсовой разницы как правило, сильно превышает дивидендную доходность. Тем не менее, включим в рассмотрение дивиденды. В файле “Dividend yield.xlsx” приводятся данные по ежегодным дивидендным выплатам для всех компаний. При расчете ожидаемой доходности добавим дивидендный доход в виде равномерного положительного смещения. Безусловно, это далеко от истины, однако упрощает вычисления, поскольку компании выбирают совершенно различные даты дивидендных отсечек.

В файле “Percentage equity returns.csv” приводится годовая доходность в процентах с учетом дивидендов (разница между ценой в последний и первый рабочие дни года + выплаченные в течение года дивиденды). Эти данные и используются далее в вычислениях.

2. Вычисление корреляционных и ковариационных матриц.

2.1 Обозначения

$\{p_t\}$ – временной ряд, зависимость цены актива от времени

Гипотеза 1: изменение цены намного меньше цены $\Delta p_t = p_{t+1} - p_t \ll p_t$

Следствие: можно заменить процентное изменение цены на разность логарифмов

$$\{\chi_t\} = \{\ln p_t - \ln p_{t-1}\} \approx \frac{p_t - p_{t-1}}{p_{t-1}}$$

Гипотеза 2: ряд процентных изменений цены строго стационарен (стационарность первого порядка проверяется, стационарность более высоких порядков, как правило, не соблюдается)

Следствие: можно рассматривать временной ряд $\{\chi_t\}$ как выборку некоторой случайной величины, $x = \{\chi_1, \chi_2, \dots, \chi_t\}$, введем в рассмотрение оценки мат. ожидания и дисперсии:

$$E(x) = \frac{1}{t} \sum_{i=1}^t \chi_i$$

$$D(x) = \frac{1}{t-1} \sum_{i=1}^t (\chi_i - E(x))^2$$

$$sd(x) = \sqrt{D(x)}$$

Пусть даны N временных рядов на некотором временном промежутке (1).

$$\{p_t^i\}, \quad i = \overline{1, N}, \quad t = \overline{1, Q}$$

Перейдем к рассмотрению рядов процентных изменений цен $\{\chi_t^i\}$

x^i – случайная величина, соотв. ряду $\{\chi_t^i\}$

$$Cov(x^i, x^j) = E\{(x^i - E(x^i)) * (x^j - E(x^j))\}$$

$$Cor(x^i, x^j) = \frac{Cov(x^i, x^j)}{sd(x^i) * sd(x^j)}$$

2.3 Eigenvalues clipping

Один из методов очистки корреляционных матриц от шума, описан например здесь [Risk](#) или в оригинальной работе (Bouchaud & Potters 2011).

3. Различные модели построения портфолио.

Разделим общий период из 2203 наблюдений на 2: 2000 и 203 измерения соответственно. Будем определять точность предсказаний по первому периоду, состоящему из 2000 измерений.

Замечание: Ожидаемая доходность по каждому из инструментов складывается из курсовой разницы и дивидендной доходности.

3.1 Модели случайных блужданий.

Код в файле: “Improved portfolio optimization.r”

В этой модели предполагается, что активы описываются формулой

$$p_t = a + p_{t-1} + \varepsilon_t, \quad \varepsilon \sim N(0, \sigma)$$

$$(1 - B)p = a + 1 * \varepsilon$$

Введем обозначения:

$w = (w_1, w_2, \dots, w_N)$ – весовые коэффициенты

$E = (E_1, E_2, \dots, E_N)$ – ожидаемые доходности по каждому из активов

Нетрудно показать, что ожидаемая доходность портфолио нормально распределенная случайная величина со следующими значениями мат. ожидания и дисперсии:

$$A = E * w, \quad D(A) = w * Cov * w$$

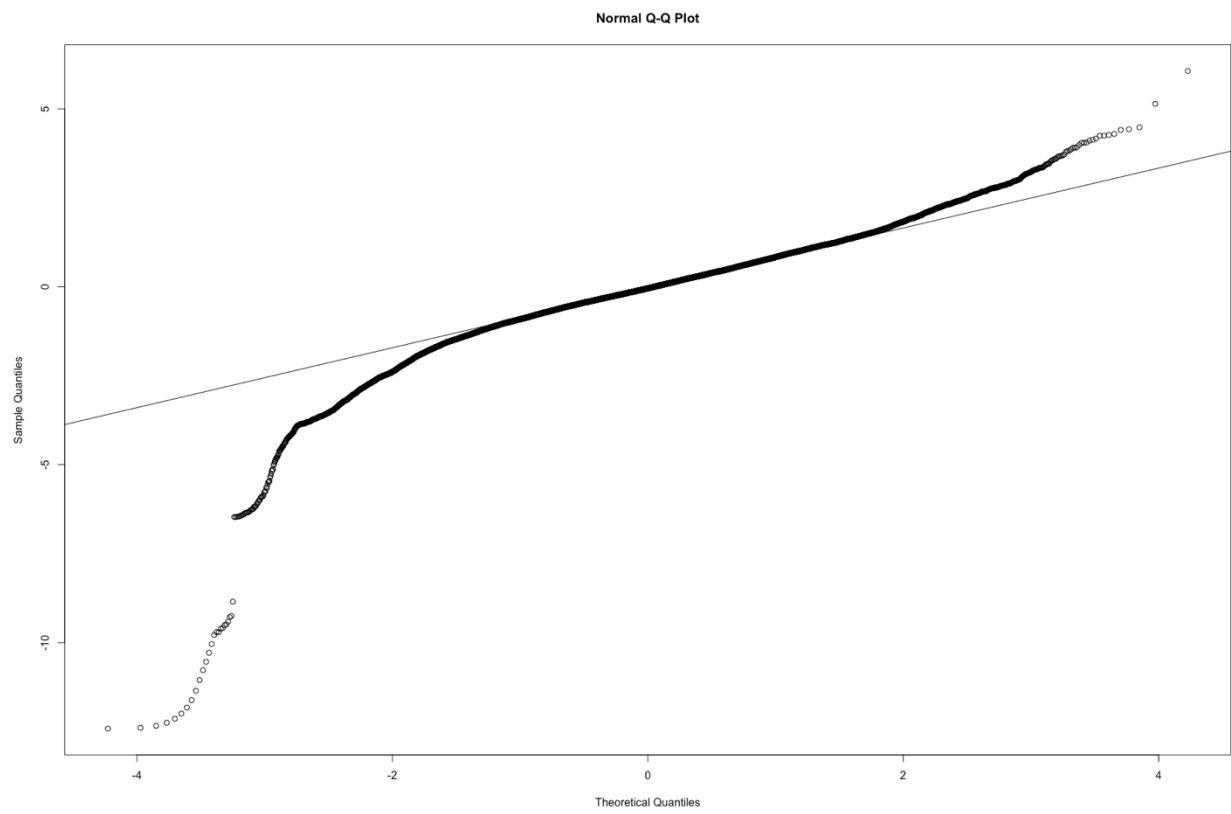
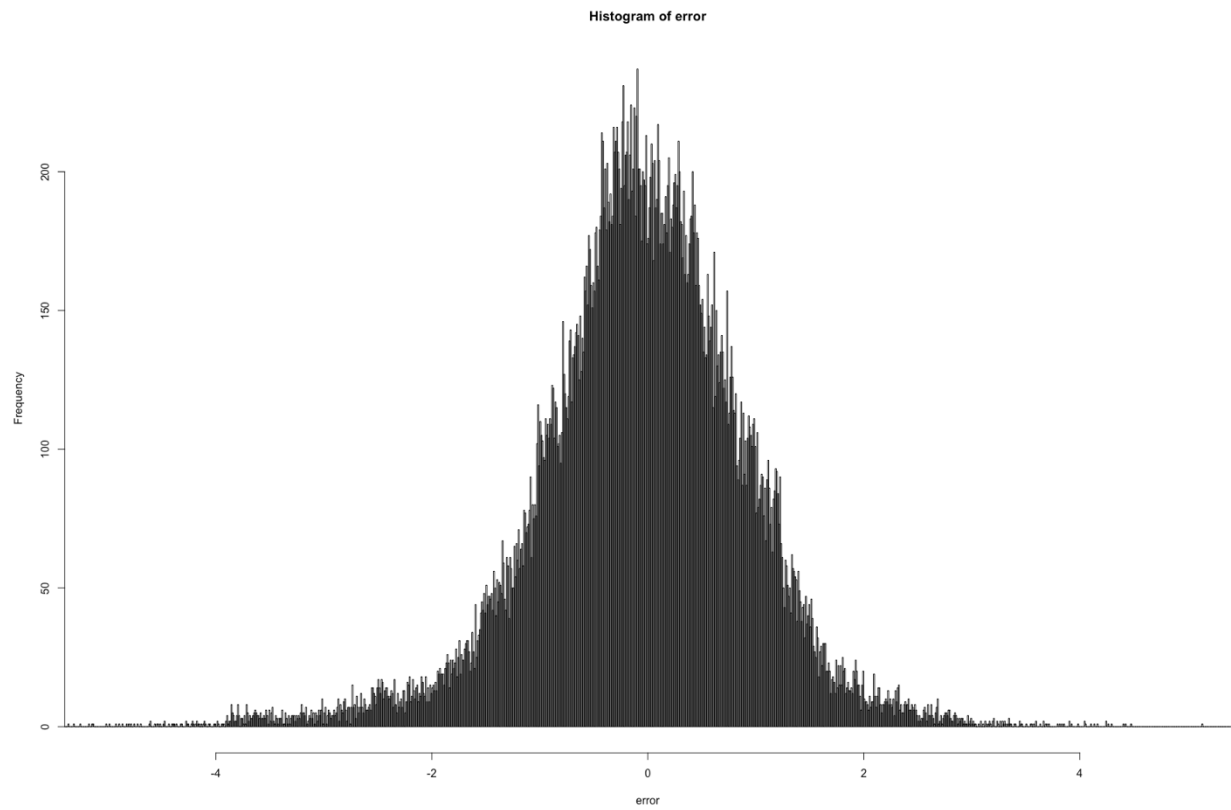
Для проверки гипотезы выберем набор портфолио и построим распределение величины

$$\frac{E_{real} - A}{sd(A)} - \text{если модель справедлива, должно получиться } N(0,1).$$

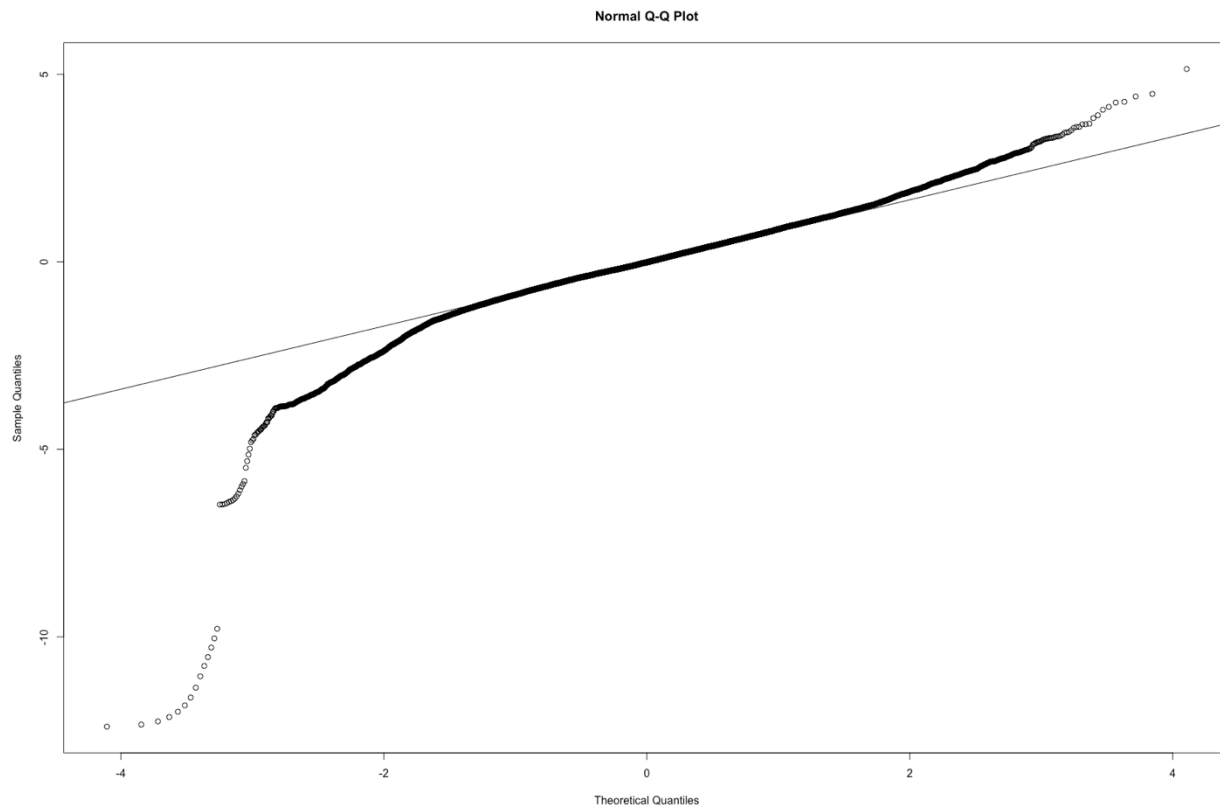
3.1.1 По периоду в 200 рабочих дней (чуть меньше одного года).

Наибольший интерес представляют т.н. эффективные портфолио. Для оценки выберем эффективные портфолио с ожидаемой доходностью 0.01 – 0.25 % в сутки (примерно 2 - 50 % годовых). Итого имеем $1800 * 25 = 45000$ точек.

На следующих графиках представлено распределение величины $\frac{E_{real} - A}{sd(A)}$ и квантильный график. Очевидны значительные отклонения от нормального распределения. Отметим, что большие отрицательные ошибки случаются намного чаще, чем положительные.



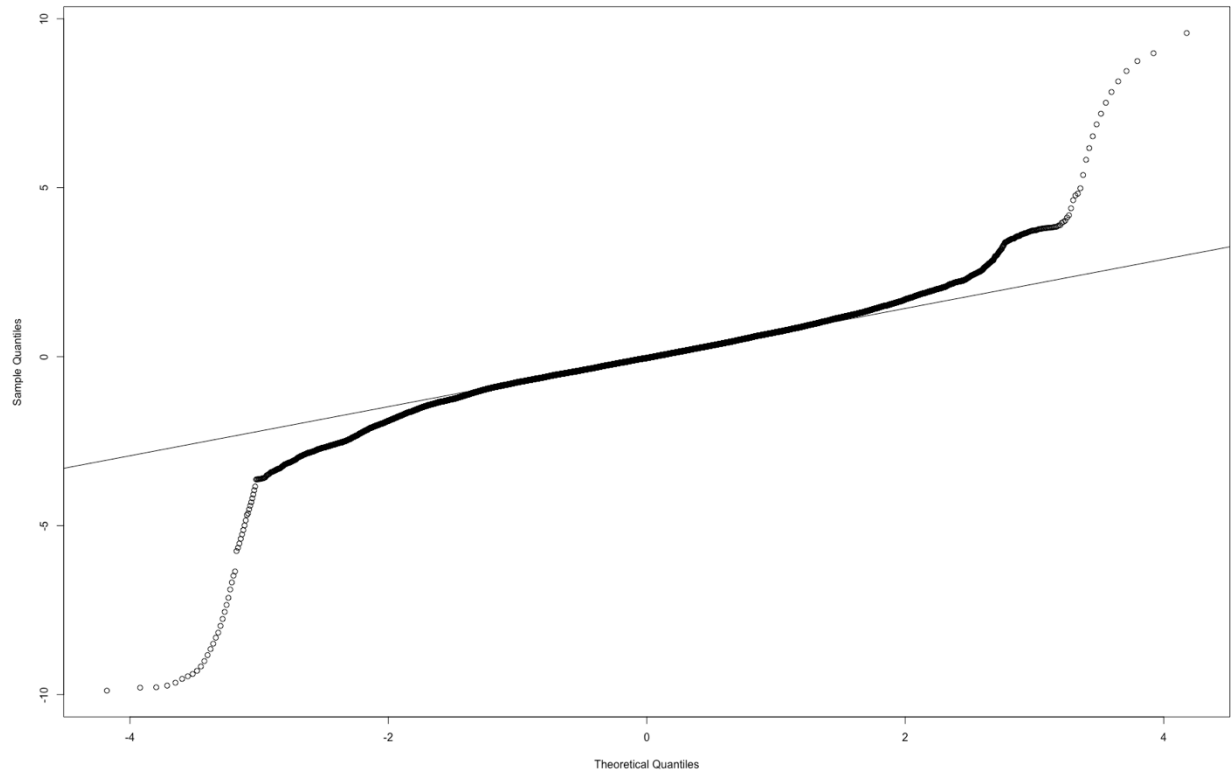
Рассмотрим возможные причины несоответствия. Для начала, сузим диапазон на значения ожидаемой доходности 5-30%. Чтобы исключить возможно слишком оптимистичные прогнозы намного выше рыночной доходности. Квантильный график показывает, что это не улучшает ситуацию.



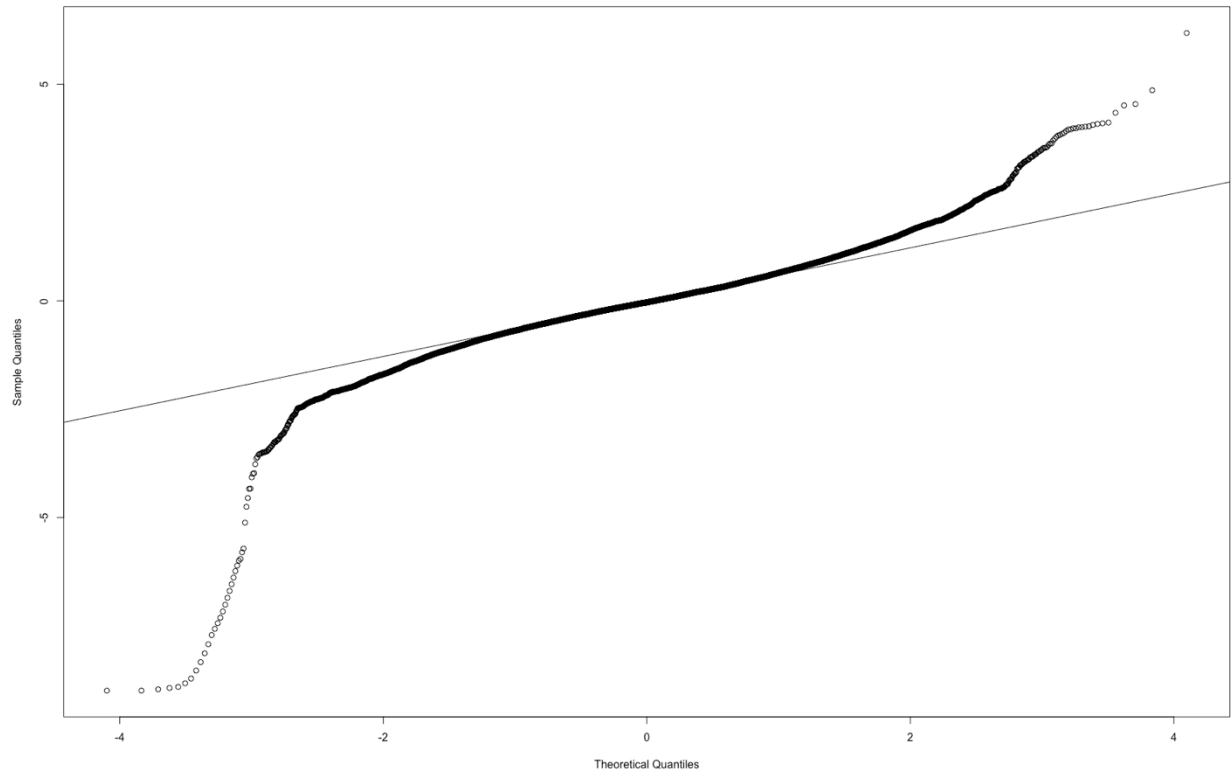
3.1.2 По периоду в 500 и 1000 рабочих дней.

Попробуем увеличить период обсчета корреляционной матрицы. Ниже приводятся соответствующие графики. Таким способом не удастся достичь значительного улучшения прогнозирования. Можно предположить, что распределение ошибки является симметричным Парето-стабильным распределением. По форме своей оно мало зависит от периода расчета корреляционных матриц. Попробуем другой способ.

Normal Q-Q Plot

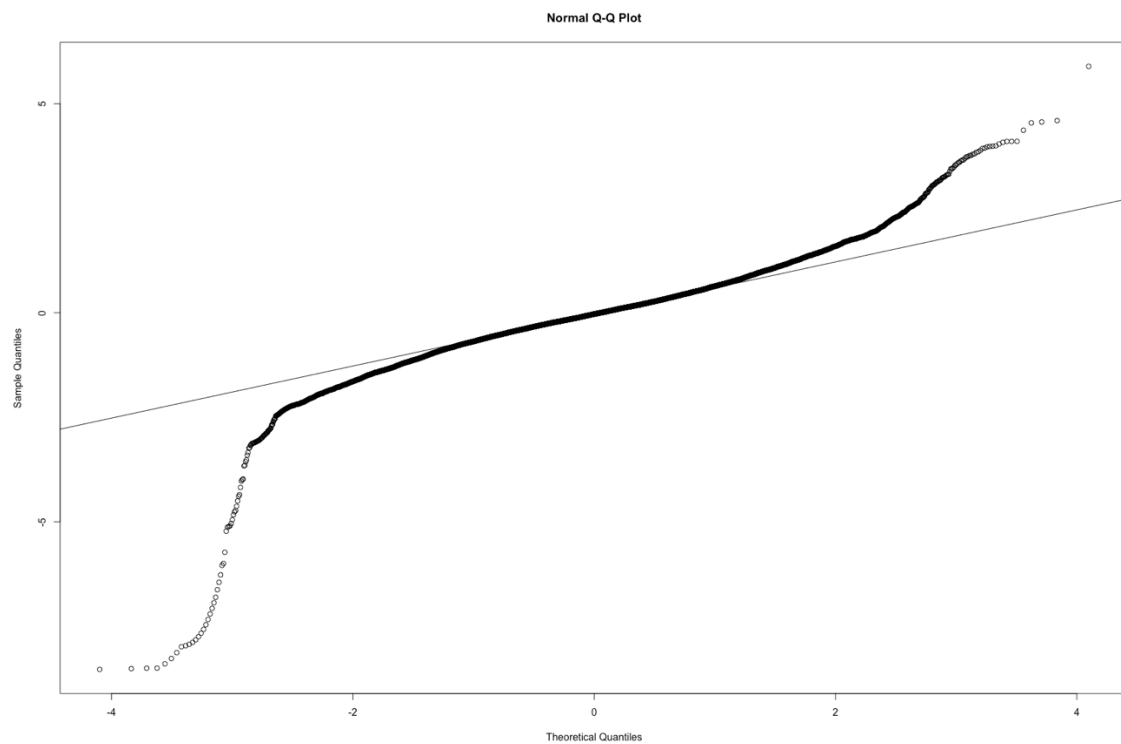
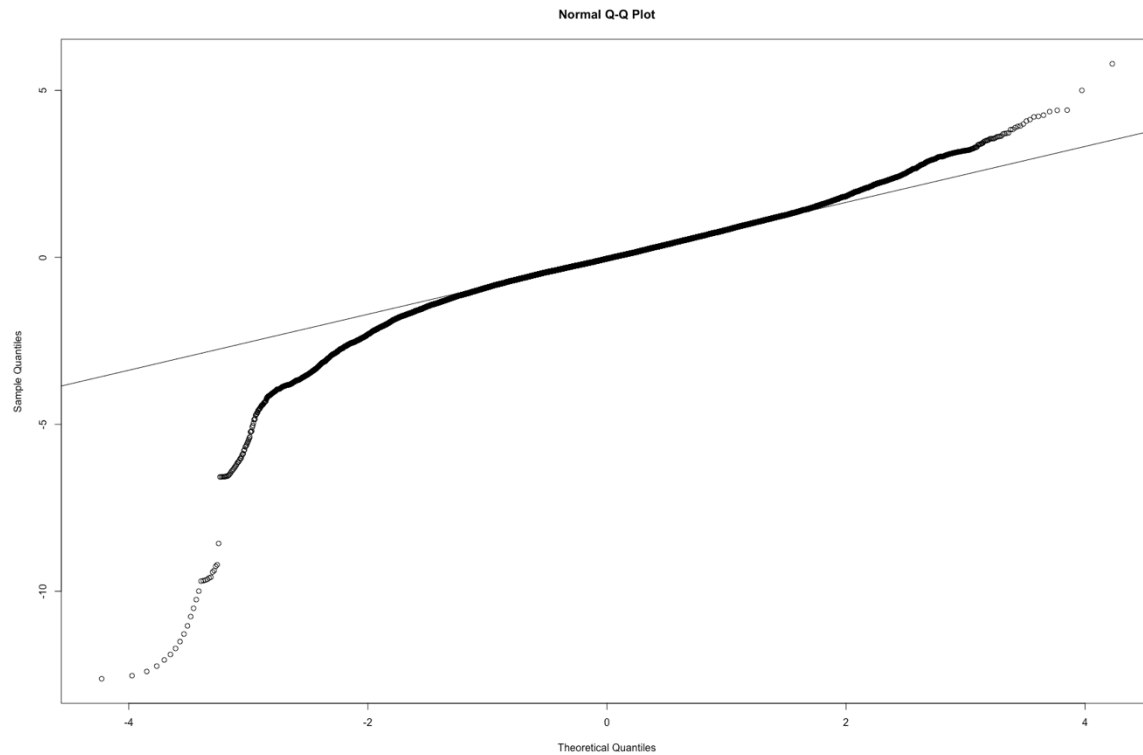


Normal Q-Q Plot



3.2.1 С очисткой корреляционных матриц (для периодов расчета 200, 500 и 1000 дней)

Применим теперь к вычисленным корреляционным матрицам технику очистки, называемую eigenvalues clipping [Risk](#). Ниже приводятся графики для 200 и 500 дней. Видно, что использование этой техники не приводит к радикальным улучшениям.



3.3 Парето-стабильные распределения.

Квантильные графики показывают, что имеет смысл подобрать распределение доходности в классе стабильных распределений. [Wiki](#)

Будем использовать 'S0' параметризацию ([Info](#)). Таким образом:

alpha, beta, gamma, delta

value of the index parameter alpha in the interval= (0, 2]; skewness parameter beta, in the range [-1, 1]; scale parameter gamma; and location (or 'shift') parameter delta.

Проведем подбор в классе симметричных распределений $\beta = 0$ методом наименьших квадратов. Диапазон варьирования $\alpha \in [1,2], \gamma \in [0.2,1]$ с шагом 0.1 установлен эмпирически.

Код в файле “Stable Paretian optimization.r” Графики в папке “Stability tests”

Q	α	γ
200	1.7	0.6
500	1.7	0.5
1000	1.5	0.4

В зависимости от ожидаемой доходности.

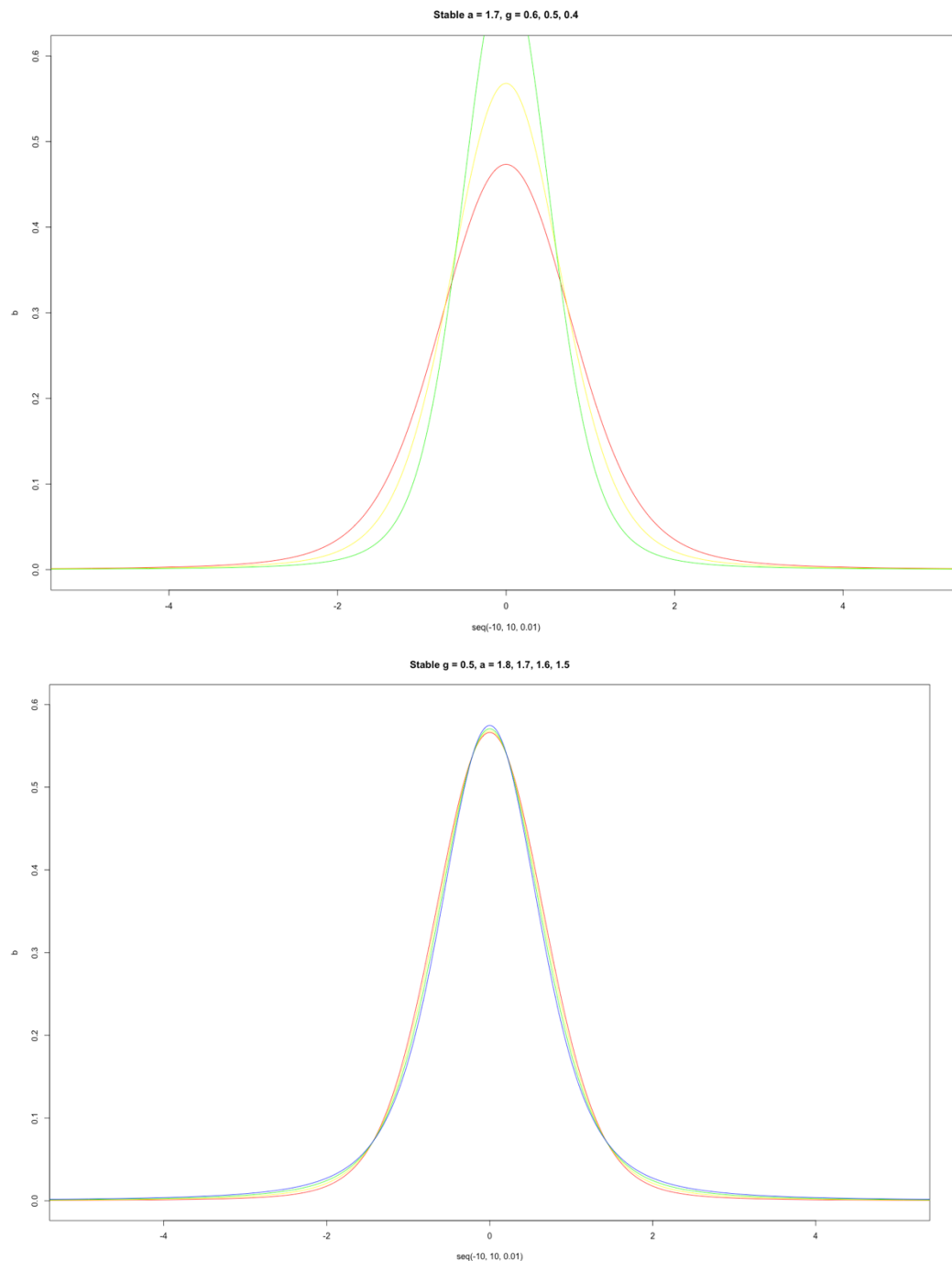
E	α	γ
0.01-0.05 (200)	1.7	0.6
0.05-0.15 (200)	1.7	0.6
0.15-0.25 (200)	1.7	0.6
0.01-0.05 (500)	1.6	0.5
0.05-0.15 (500)	1.7	0.5
0.15-0.25 (500)	1.7	0.5
0.01-0.05 (1000)	1.6	0.4
0.05-0.15 (1000)	1.5	0.4
0.15-0.25 (1000)	1.5	0.5

В зависимости от временного промежутка.

E	α	γ
1-600 (200)	1.5	0.6
601-1200 (200)	1.7	0.6
1201-1800 (200)	1.8	0.6
1-500 (500)	1.8	0.5
501-1000 (500)	1.8	0.6
1001-1500(500)	1.8	0.4
1-333 (1000)	1.7	0.6
334-666 (1000)	1.5	0.4
667-1000 (1000)	1.6	0.4

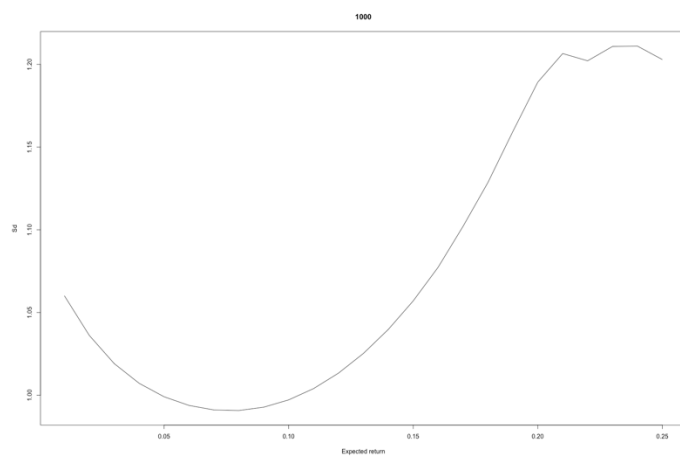
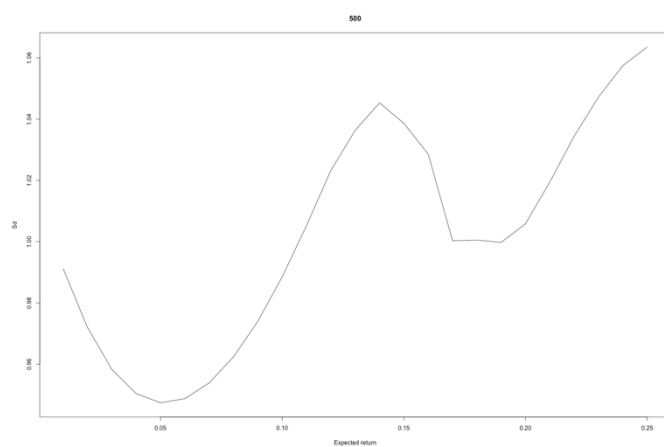
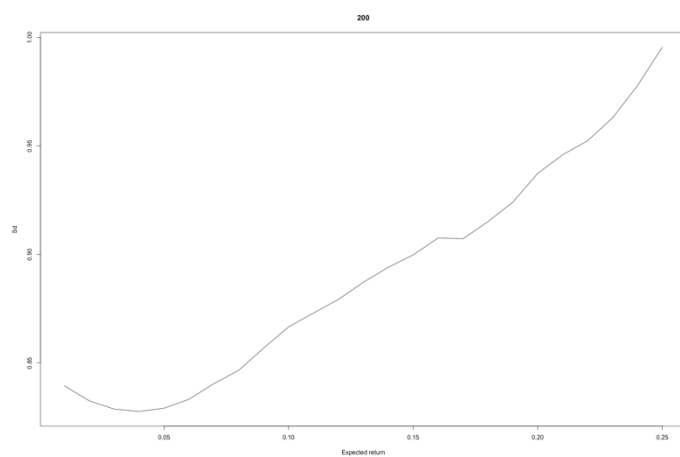
В целом стабильные распределения обеспечивают хорошее соответствие наблюдаемой картине. Параметры оптимальных распределений также довольно близки и не сильно меняются для разных значений времени и ожидаемой доходности. Ниже приводится зависимость формы функций распределения от α и γ . Видно, что чем меньше γ – тем

быстрее спадает плотность вероятности, что означает более высокую точность прогнозирования, тогда как в полученном диапазоне значений α зависимость слабее. Таким образом, модели с более длинным периодом вычисления корреляционной матрицы дают более точные предсказания.



Итак, мы установили, что величина $\frac{E_{real} - A}{sd(A)}$ хорошо приближается стабильными распределениями. Посмотрим теперь, как меняется величина $sd(A)$ в зависимости от A .

Ниже приводятся зависимости усредненного по времени $sd(A)$ от A для трех периодов вычисления корреляционных матриц (200, 500, 1000 дней).



4. Построение модели.

Предположим, что в некоторый момент времени t мы хотим вычислить доходность портфолио, характеризуемого вектором весовых коэффициентов $w = w(w_1, w_2, \dots, w_N)$.

1. Определим ожидаемую суточную доходность для каждого из активов, как разницу между рыночной стоимостью на момент времени t и рыночной стоимостью на момент $t - 1000$ плюс накопленный дивидендный доход за 1000 дней. Ожидаемая доходность представляется вектором $E = (E_1, E_2, \dots, E_N)$
2. Определим ковариационную матрицу по процентным изменениям цен активов за указанный период 1000 дней. $Cov(x_i, x_j)$
3. Определим ожидаемую доходность и уровень риска.

$$A = E * w$$

$$Var = w * Cov * w$$

$$Sd = \sqrt{Var}$$

4. Величина $\frac{E_{real}-A}{sd(A)}$ распределена стабильно со значениями параметров

$$\alpha = 1.5, \quad \beta = 0, \quad \gamma = 0.4, \quad \delta = 0$$

5. $P(E_{real} - A < x) = P(\frac{E_{real}-A}{sd(A)} < \frac{x}{sd(A)})$

$$pdf_{E_{real}-A}(x) = \frac{1}{sd(A)} pdf(\frac{x}{sd(A)})$$

$$pdf_{E_{real}}(x + A) = pdf_{E_{real}-A}(x) = \frac{1}{sd(A)} pdf(\frac{x}{sd(A)})$$

Таким образом, можно получить распределение E_{real} и определить значения Var для различных уровней риска

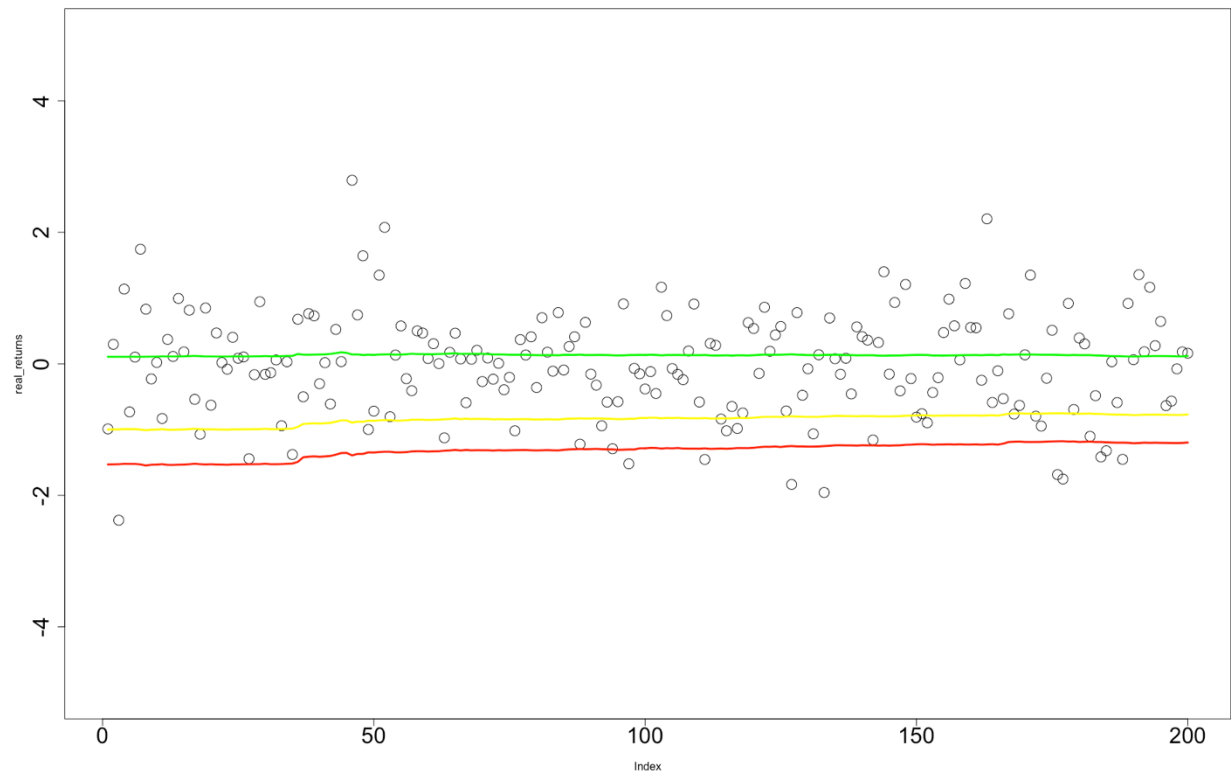
4.1 Пример.

Для расчета Var выберем 4 портфолио, состоящих соответственно из бумаг 1-9, 10-19, 20-29, 30-38 в равных пропорциях. Вычислим Var для 5% и 10% уровней. В качестве периода предсказаний выберем последние 200 рабочих дней согласно данным.

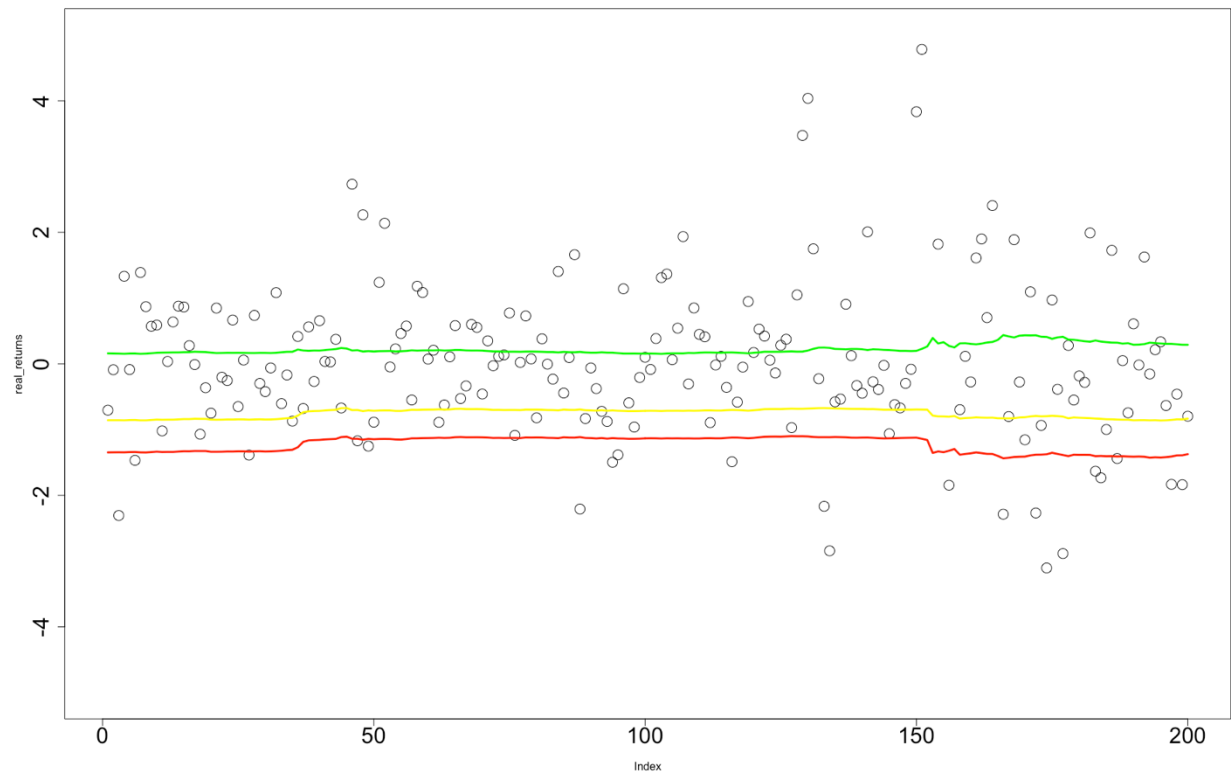
Portfolio	< 0.1	< 0.05
1	31	11
2	39	20
3	29	10
4	29	10

Можно заметить отклонения от модели в области больших отрицательных значений. Как следует из данных, полученных методом Монте-Карло, вероятность получить более 30 значений, превышающих 0.1 около 1%, а вероятность получить 20 и более значений, превышающих 0.05- 0.25%, эти отклонения не объясняются простой случайностью. Ниже приведены графики. Зеленая линия показывает ожидаемую доходность, желтая соответствует $Var(0.1)$, красная $Var(0.05)$.

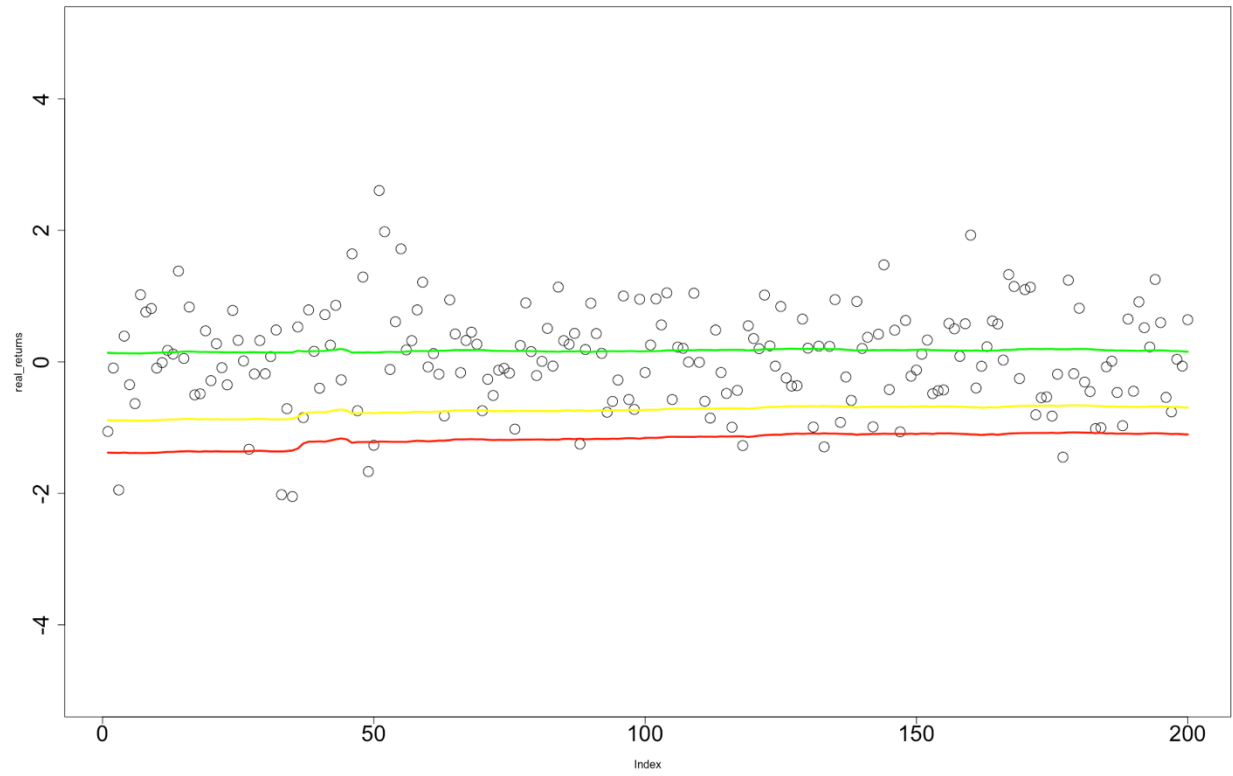
Portfolio1



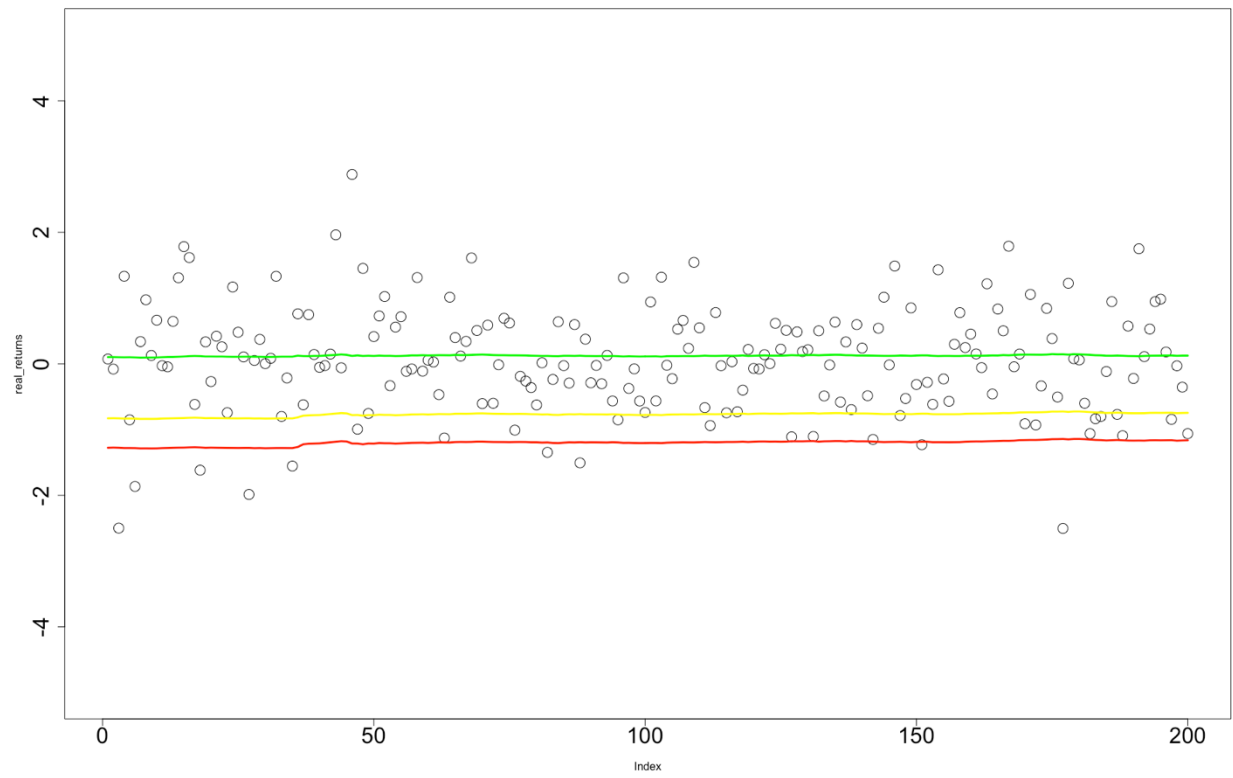
Portfolio2



Portfolio3



Portfolio4



5. Рассмотрим отклонения от модели.

В следующих таблицах собраны результаты, которые не находят объяснения в модели.

Portfolio	N	Return
1	45	-6.75%
2	45	-7%
2	153	-6.8%
2	155	-5.5%
2	152	13.44%
2	157	7.69%
2	165	5.7%
3	45	-8.63%
4	45	-5.94%

Во всех 4-х случаях фигурирует день 45- 9 апреля 2018 года. Что странно, этот день не соответствует никаким дивидендным отсечкам. А по портфолио 2 это еще диапазон со 153 по 165 дни, что соответствует 07.09 – 25.09.

Тут объяснение найти проще- к тому времени портфолио состоял уже на 35-40% из акций компании Иркут. Которая показывала какую-то невероятную динамику.

Литература

1. *Marcovitz H.* Portfolio selection: efficient diversification of investments, V.16 Yale university press, 1968
2. *Haff L.* Empirical Bayes estimation of the multivariate normal covariance matrix, Annals of Statistics, V.8, pages 586-597, 1980
3. *Bouchaud J-P and M Potters*, Theory of financial risks and derivative pricing: from statistical physics to risk management, Cambridge university press, 2008
4. *Bouchaud J-P and M Potters*, Financial applications of random matrix theory: a short review, The Oxford Handbook of Random Matrix Theory, 2011
5. *Conlon T., Ruskin J., Crane M.*, Cross-Correlation Dynamics in Financial Time-Series, Statistical Mechanics and its Applications, Feb. 2010
6. *Fama F.*, The Behavior of Stock-Market Prices, Journal of Business V.38, pages 34-105, N.1, Jan. 1965
7. *Mandelbrot B.*, The Variation of Certain Speculative Prices, Journal of Business V.36, pages 394-419, N.4, Oct. 1963
8. *Diethelm Wuertz, Martin Maechler*, Stable Distribution Functions, Rmetrics Association, Sep. 2016