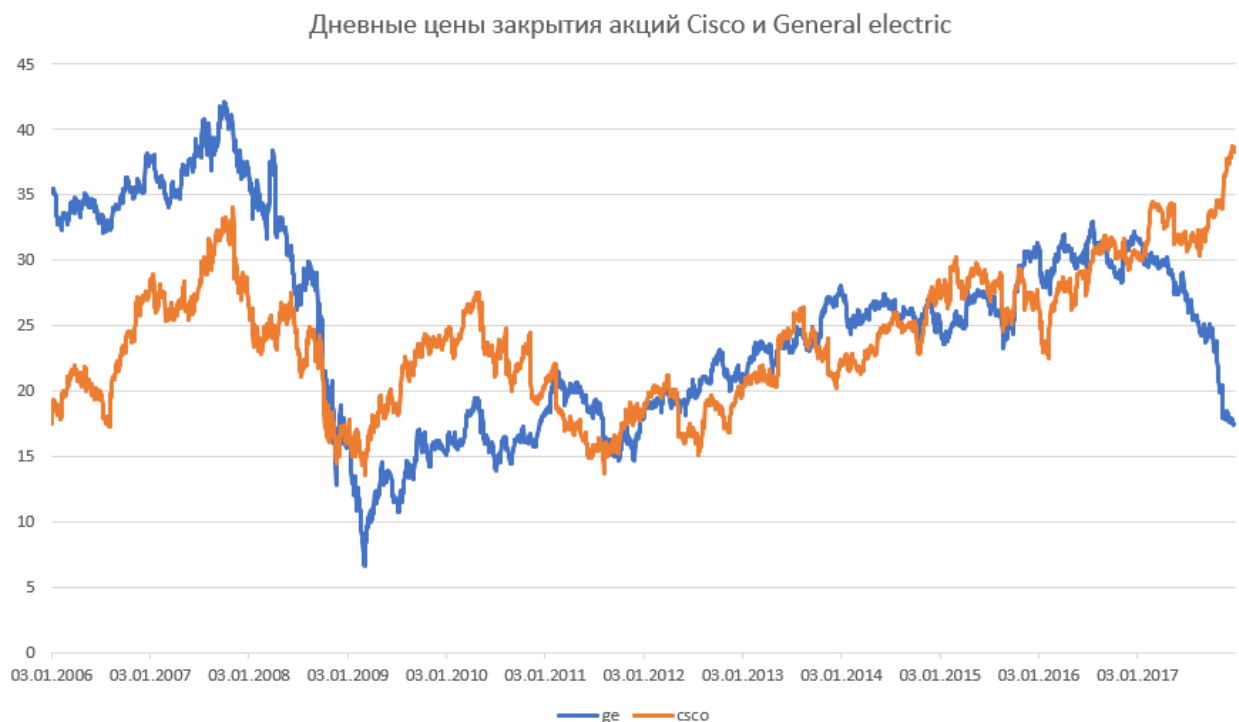


- 1) 5 баллов: взять из открытых данных динамику любых двух биржевых индексов или любую пару финансовых активов (доллар-евро);
- 2) 5 баллов: составить лог-приросты каждого показателя за все время и изобразить точками на двумерной плоскости получившиеся величины;
- 3) 5 баллов: с помощью непараметрических методов провести на плоскости регрессионную кривую, выбрав оптимальную ширину окна;
- 4) 5 баллов: сравнить качество полученной регрессионной кривой с качеством линейной регрессии;
- 5) 5 баллов: разделить смесь распределений одного из показателей;
- 6) 5 баллов: теперь 50% данных (первых по времени) использовать для построение регрессии на двумерной плоскости, на оставшихся 50% необходимо сравнить качество прогноза непараметрической модели регрессии с фактически наблюдаемыми величинами. В качестве прогноза взять построенную кривую регрессии. Т.е. фактически нужно сравнить построенную непараметрическую кривую регрессии с будущими истинными значениями по критерию суммы квадратов отклонений истинных значений от построенной регрессионной кривой.

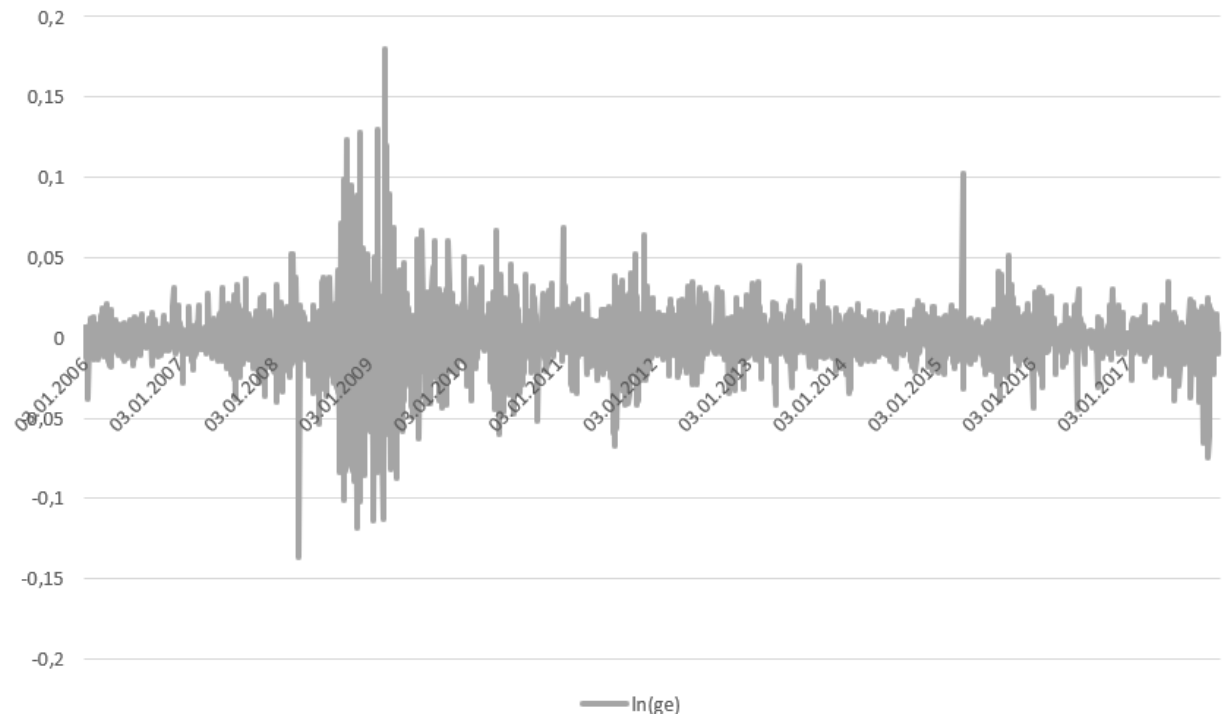
Демонстрация используемых данных.



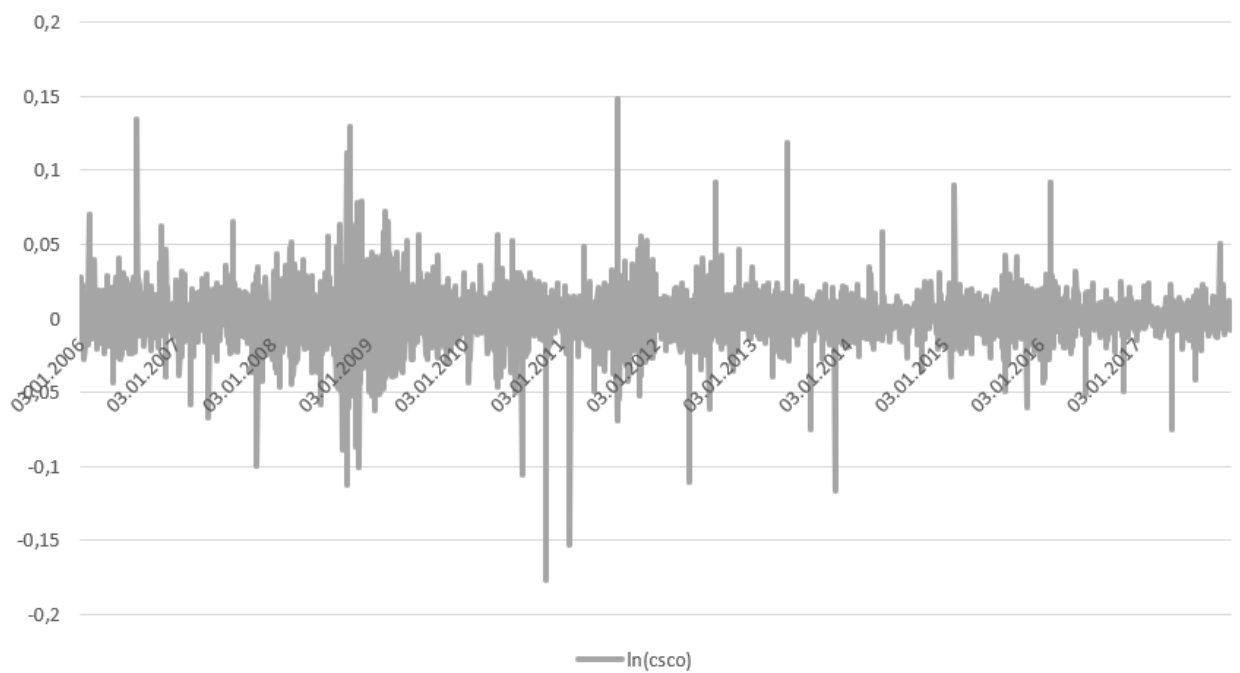
Составление лог приростов используемых показателей и представление данных.

$$y = \ln \frac{x_t}{x_{t-1}}, x - \text{значение индекса}, y - \text{новое значение}$$

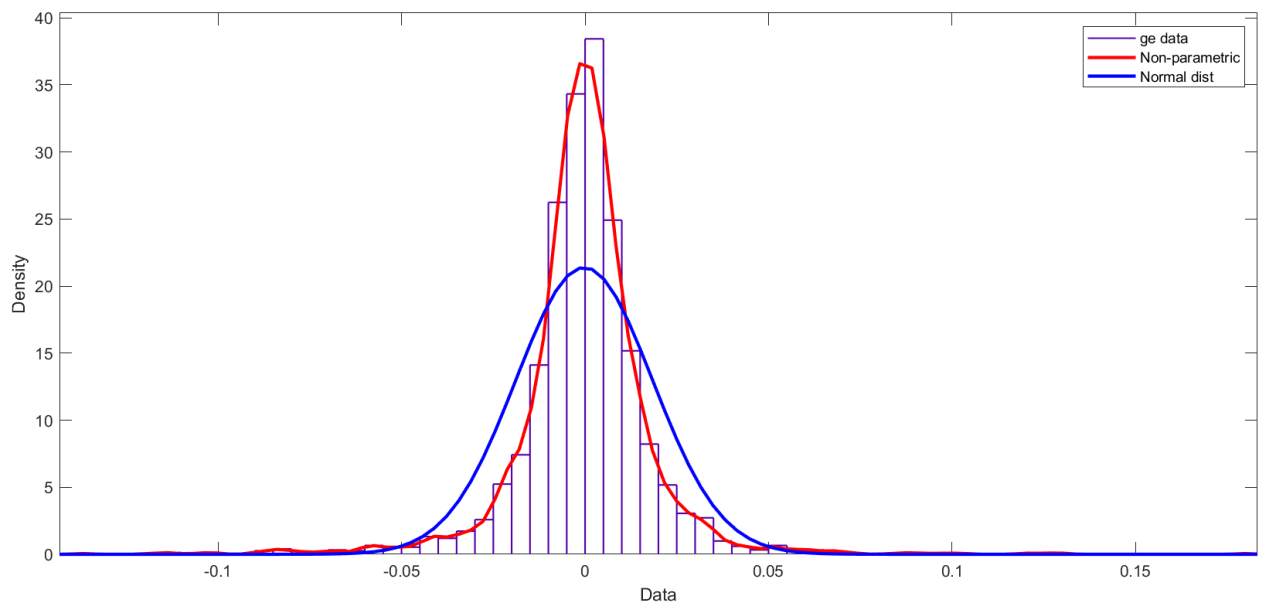
Логорифмические прирост цен закрытия акций General Electric



Логорифмические прирост цен закрытия акций Cisco



Ниже приведён график плотности распределения, непараметрической оценки и нормального распределения.

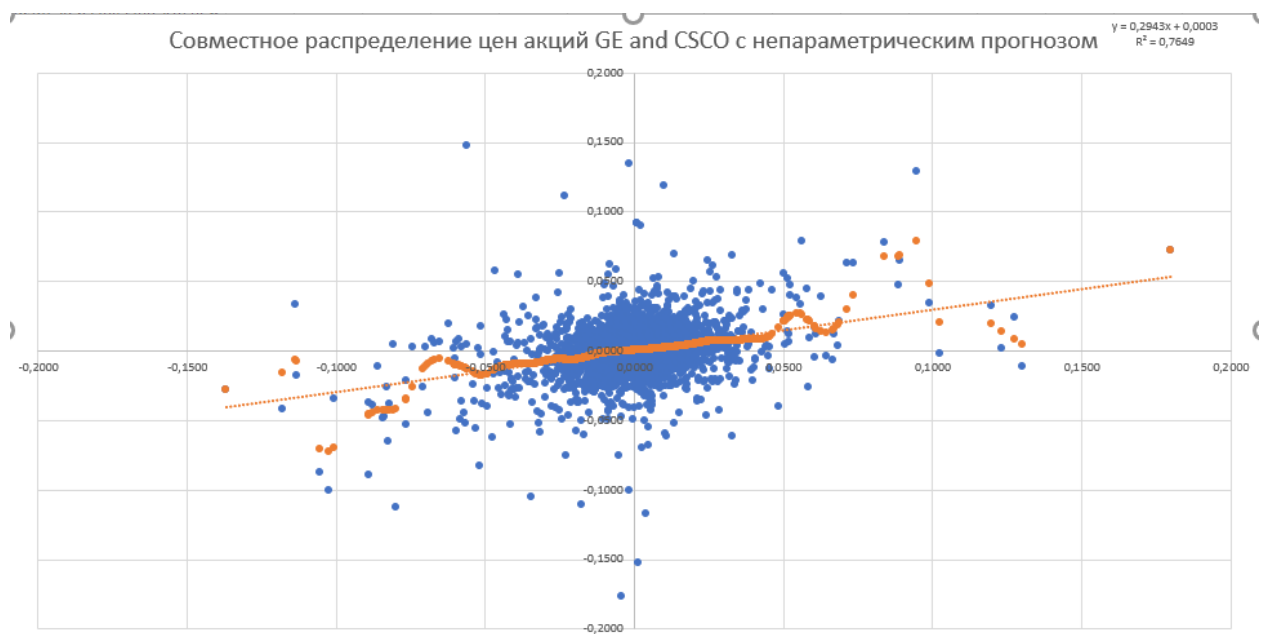


Пункт 3 и 4.

Ширина окна для Гауссовского ядра

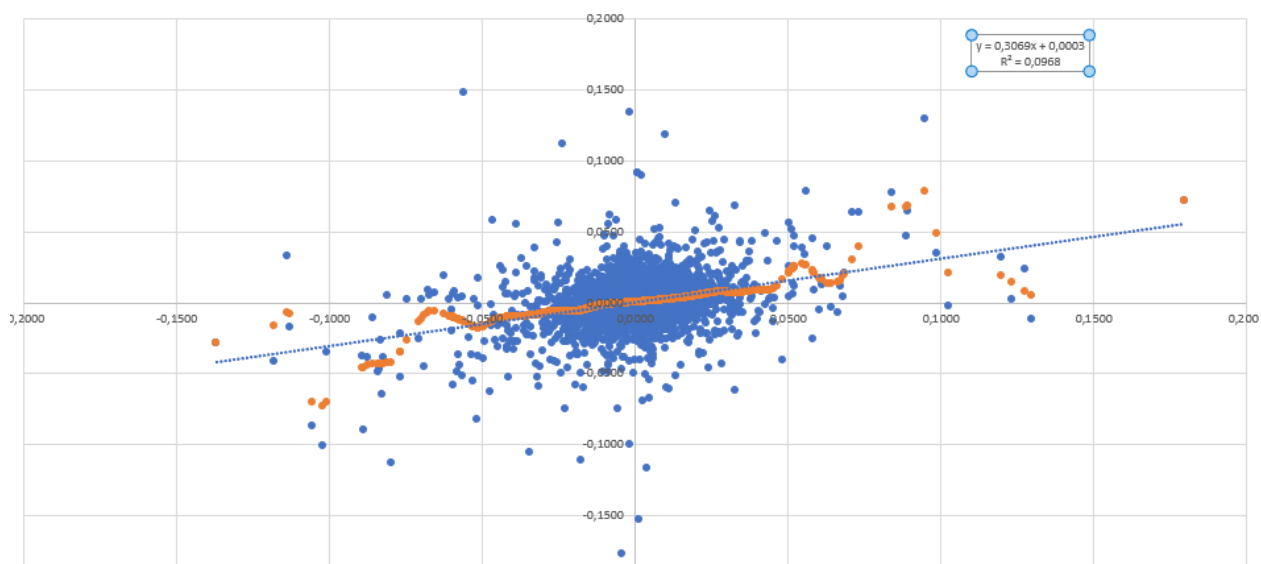
$$h^S = 1.06\hat{\sigma}_x n^{-\frac{1}{5}}.$$

Ниже график: примерное распределение с расчётом ядра.



Модель регрессии без непараметрических методов

Совместное распределение цен акций GE and CSCO с непараметрическим прогнозом

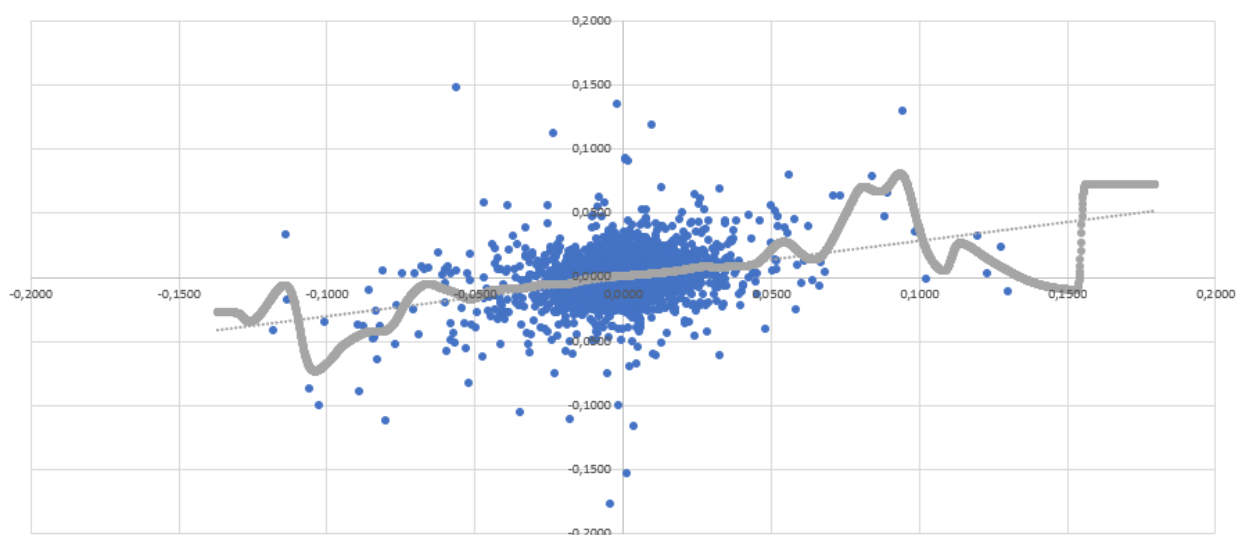


И сглаженный график для пиков.

Совместное распределение цен акций GE and CSCO с заполнением пропущенных значений

$$y = 0,2923x - 0,0012$$

$$R^2 = 0,5878$$

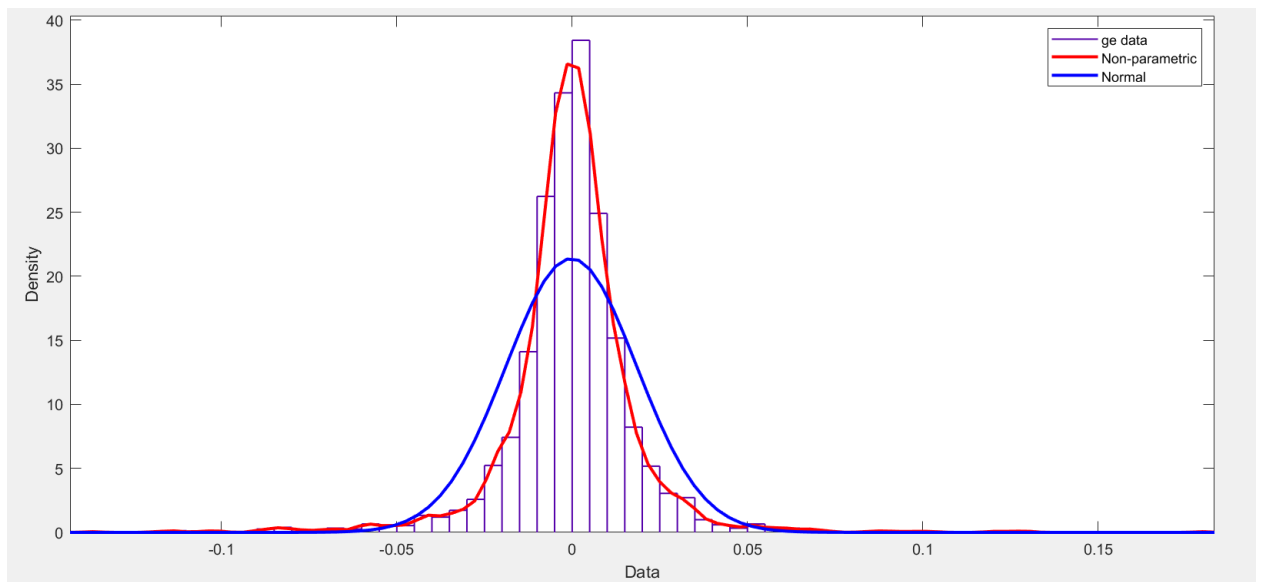


Интересно, точно во 2 случае точность линии тренда уменьшилась.

Пункт 5.

Profit – как в 1 так и во 2 случае нужно правильно выбирать данные для построения линии тренда, и R2 будет достаточно хорошим.

(Работаем с акциями компании General Electric) Затем определим оптимальные параметры по графику non-parametric distribution.



Решил взять  $k=5$  для алгоритма, так как только для 5 стабильное положительно значение функции правдоподобия, и такой же вывод был сделан после созерцания графика. Наши коэффициенты:

```
>> [W,M,V,L] = EM_GM(ge,5,[],[],1)
```

CPU time used for EM\_GM: 3.39s

Number of iterations: 55

Elapsed time is 2.969804 seconds.

W = 0.0497 0.0407 **0.1913** 0.1322 0.5859

M = -0.0245 0.0117 0.0119 -0.0137 0.0001

V(:,1) = 0.0013

V(:,2) = 0.0028

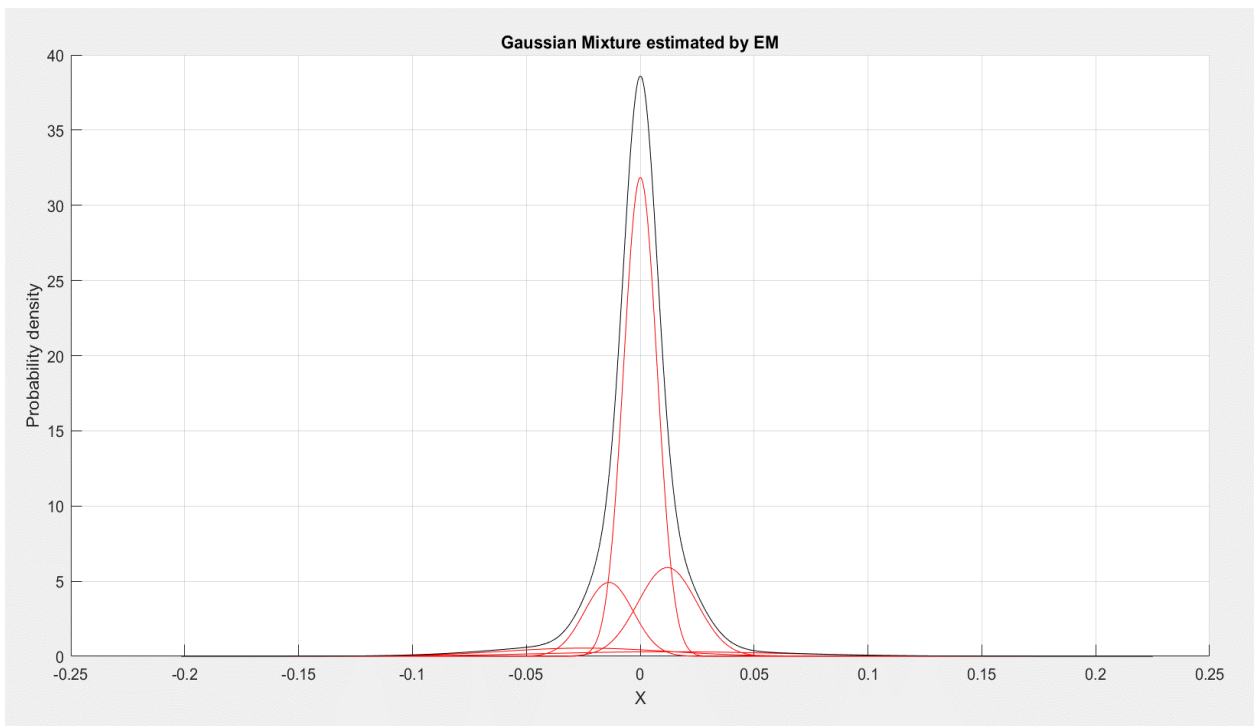
V(:,3) = 1.6701e-04

V(:,4) = 1.1518e-04

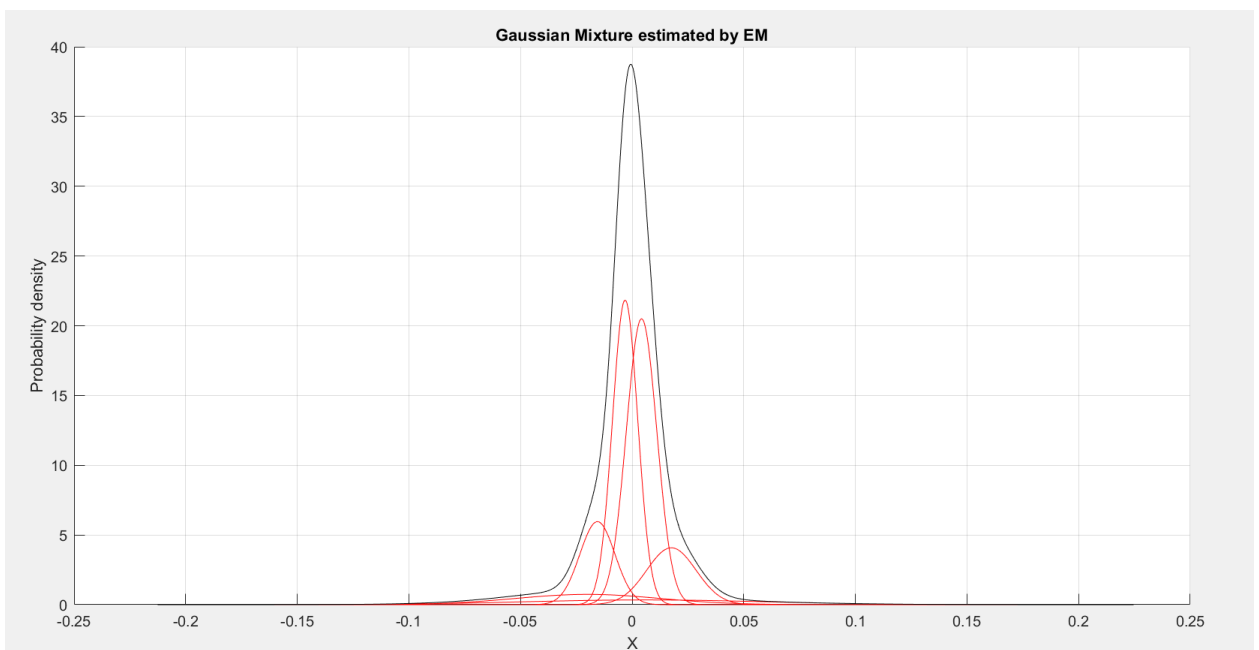
V(:,5) = 5.3851e-05

**L = 3.5469e+03 (достаточно стабильное значение)**

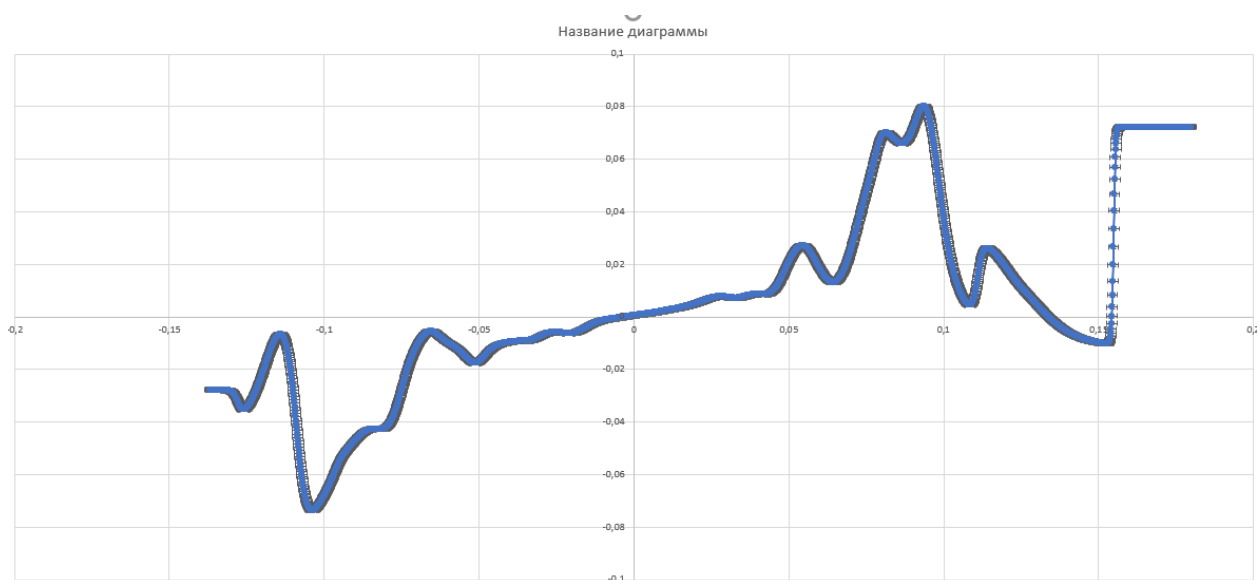
Также наглядно продемонстрируем из чего же строится наше распределение.



По изначальному графику плотности предполагалось что будет около 6-7 режимов / распределений. Но если взять для них разбивку как на графике выше, видно что общий график становится хуже, да и значение функции правдоподобия  $L = -354.8632$ .

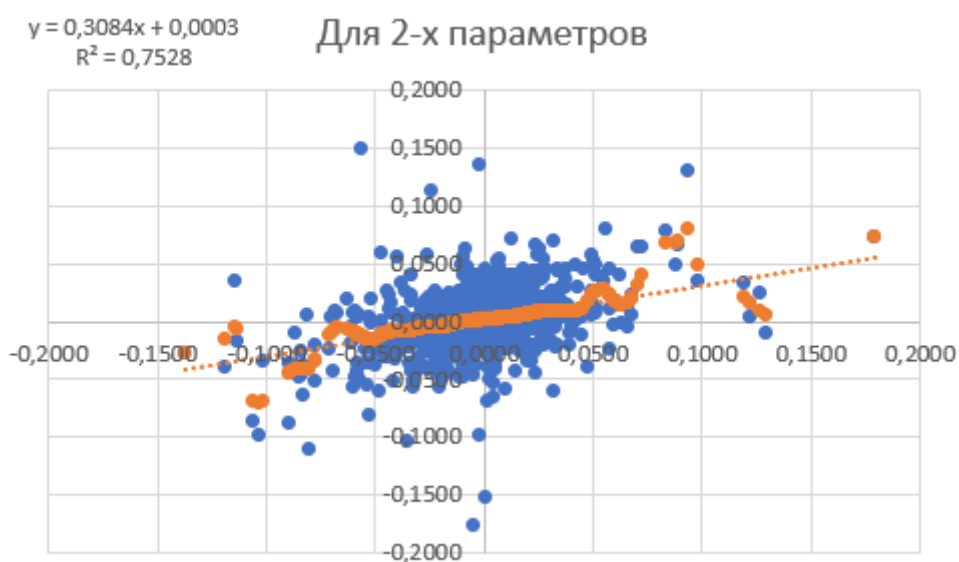


Добавление стандартных отклонений не показало большей дисперсии на пиках.



## Часть 6

Если я верно понял, то на половине данных я построил вот такое приближение применив непараметрический метод.



Воспользуемся уравнением линейного тренда выше для 1 половины данных чтобы восстановить 2ю половину. Сравним MSE для 2 половины непараметрических методов и оцененных при помощи линейного тренда. Увидим, что MSE для непараметрической модели будет меньше.

MSE non	MSE est
0,000193	0,000198

УРА!