

Федеральное государственное автономное образовательное учреждение высшего образования

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

**Факультет экономических наук**

## **КУРСОВАЯ РАБОТА**

Методы обнаружения структурных сдвигов в моделях временных рядов

Выполнил:

Студент группы МСМ-181

Кафтанов И. А.

Руководитель:

Борzych Д. А.

Москва 2019

# Содержание

Введение	3
1 Метод детектирования одного сдвига	6
2 Итеративные алгоритмы для поиска всех возможных сдвигов	8
3 Апробация методов на реальных данных	10
Литература	16

Задача анализа и прогнозирования временных рядов является одной из самых востребованных и актуальных: практически во всех областях науки и прикладных отраслей существуют процессы, которые упорядочены во времени. Развитие теорий, методов и технологий в области решения этой задачи в первую очередь подкреплено огромным спросом на анализ подобных процессов.

Временной ряд представляет собой последовательность наблюдений, упорядоченных по какому-либо параметру (обычно, реальному времени или номеру события) для которых порядок наблюдения, играет существенную роль. Существует масса примеров процессов, где значения упорядочены во времени и скоррелированы между собой, и в зависимости от области, в которой данный ряд порождается, перед исследователями ставятся различные цели и задачи по его анализу и прогнозированию. С помощью статистического анализа находятся аномальные значения во временных рядах, прогнозируют его поведение или изучают механизм порождения.

Основным средством для этого служат так называемые модели. Понятие модели включает в себя два ключевых компонента – как модель временного ряда и как прогнозная модель. Модель временного ряда описывает правило генерации членов ряда, а прогнозирование даёт оценку будущих значений членов ряда. Во многих случаях модель можно определить с точностью до конечного числа параметров. Задача получения статистических выводов базируется на данных параметрах. При этом важно понимать, что для нахождения наиболее точных оценок параметров модели, как правило, необходимо большое количество наблюдений.

Такое ограничение подразумевает, что для нахождения оптимальных оценок коэффициентов для различного рода моделей, например эконометрических, требуется большой объем данных (количество наблюдений). Ограничение влечёт за собой логичный вывод, что существуют процессы, структура которых, с определённой вероятностью, будет изменяться с течением времени, а при увеличении количества наблюдений это вероятность лишь увеличивается. Конечно, это не относится к временным рядам, которые

являются стационарными и не имеют внешних факторов воздействия. Но в силу специфики предметной области эконометрических моделей (которые будут рассмотрены ниже) мы, как правило, сфокусированы на временных рядах, в которых влияние внешних факторов существенно.

Проверка изменения параметров (коэффициентов) моделей во времени не является новой задачей в статистике, ещё Page [3] в своей работе исследовал возможность определения изменения параметров во времени. В последствии этой задачей заинтересовалось все больше исследователей из различных областей науки, было опубликовано большое количество статей в журналах, например Wichern, Miller and Hsu (1976) [4], Zacks (1983) [6]. Классический пример этой задачи можно найти в работе Hamilton [5]. Исследованием методов нахождения структурных сдвигов занимались Kokoszka, Leipus (1999) [2], Lee S. [7], Inclan, Tiao (1994) [1]. Эти работы посвящены исследованию выбора критериев для оптимального определения сдвигов. Основным источником этой работы является работа Kokoszka, Leipus [2], где описан способ нахождения одного структурного сдвига. Так же используется работа Tiao для применения алгоритма Iterated Cumulative Sums of Squares (ICSS).

Таким образом исследование такого рода временных рядов затруднено наличием структурных сдвигов (разладки случайного процесса) игнорирование которых в большинстве случаев приводит к некорректным результатам. Поэтому задача обнаружения структурных сдвигов является актуальной и востребованной при анализе временных рядов.

В работе рассматривается задача детектирования структурных сдвигов для кусочно-заданной  $GARCH(1, 1)$  модели. Пусть имеется временной ряд  $(Y_t)_{t=1}^T$  для которого существует  $k \geq 0$  структурных изменений (сдвигов), тогда обозначив моменты разладок как  $\tau_1, \tau_2, \dots, \tau_k$  уравнение для  $j$ -того фрагмента будет выглядеть следующим образом:

$$Y_t = \varepsilon_t, \quad \sigma_t^2 = w_j + \delta_j \sigma_{t-1}^2 + \gamma_j \varepsilon_{t-1}^2 \quad \text{где } t \in [\tau_{j-1}; \tau_j - 1] \quad (1)$$

а  $(w_j, \delta_j, \gamma_j)$  тройка неизвестных параметров модели которые принадлежат множеству значений  $\Theta = \{(w, \delta, \gamma) : w > 0, \delta \geq 0, \gamma \geq 0, \delta + \gamma \leq 0\}$

В результате задача заключается в нахождении алгоритмов поиска всех возможных сдвигов во временном ряде по двум параметрам:

1. Точность нахождения моментов сдвига;
2. Временная и пространственная сложность алгоритмов.

В работе рассматривается ICSS процедура с применением KL-CUSUM для детектирования изменений  $garch$  модели. Мы предполагаем, что используя данный подход нам удастся определить моменты времени  $\tau$ , которые отражают изменения в мире и на рынке в частности, с некоторой величиной погрешности.

Для проверки выдвигаемых гипотез методы апробируются на данных о цене акций следующих компаний:

- QIWI:MCX (QIWI) На период с 2013-10-08 по 2019-05-16. Размер выборки: 1400;
- LKOH:MCX (Лукойл) На период с 1997-05-29 по 2019-06-06. Размер выборки: 5495.
- RTKMP:MCX (Ростелеком) На период с 1999-04-19 по 2019-06-13. Размер выборки: 4968.

# 1 Метод детектирования одного сдвига

Метод обнаружения одного структурного сдвига KL-CUSUM, предложенный основан на разнице двух взвешенных сумм по временным промежуткам разной длины. Вводятся статистика KL и  $\hat{v}$  на основании которых принимается решение о наличии структурного сдвига или его отсутствия. Критерий был сформулирован (ссылка на книжку) и опирается на теорему 2.2 [2]:

**Теорема 1.** *Предположим, что  $q \rightarrow \infty$  и  $q/N \rightarrow 0$  Тогда согласно гипотезе  $H_0$*

$$\frac{U_N(T)}{s_{\hat{N},q}} \xrightarrow{D(0,1)} W^0(t) \quad (2)$$

где гипотезы:

- $H_0$ :  $X_1, \dots, X_N$  это выборка полученная из уравнения (1)
- $H_1$ :  $\exists(w_1, \delta_1, \gamma_1), (w_2, \delta_2, \gamma_2)$  такие что  $X_1 \dots X_\tau \in$  выборке  $s(w_1, \delta_1, \gamma_1)$  и  $X_\tau \dots X_N \in$  выборке  $s(w_2, \delta_2, \gamma_2)$

Соответственно необходимо показать, что сходимость по распределению отсутствует.

Пусть существует кусочно-заданный GARCH(1,1)-процесс  $(Y_t)_{t=1}^T$  допускающий наличие одного структурного сдвига в момент времени  $\tau \in [2; T]$  и описывающийся следующей системой уравнений:

$$\begin{cases} Y_t = \epsilon_t, \epsilon_t = \sigma_t \xi, \sigma_t^2 = w_1 + \delta_1 \sigma_{t-1}^2 + \gamma_1 \epsilon_{t-1}^2, & t \in [1; \tau - 1] \\ Y_t = \epsilon_t, \epsilon_t = \sigma_t \xi, \sigma_t^2 = w_2 + \delta_2 \sigma_{t-1}^2 + \gamma_2 \epsilon_{t-1}^2, & t \in [\tau; T] \end{cases} \quad (3)$$

Рассчитывается статистика  $KL$  такая что

$$\begin{aligned} KL(k) &= \sqrt{T} \frac{k(T-k)}{T^2} \left( \frac{1}{k} \sum_{t=1}^k Y_t^2 - \frac{1}{T-k} \sum_{t=k+1}^T Y_t^2 \right) = \\ &= \frac{1}{\sqrt{T}} \left( \sum_{t=1}^k - \frac{k}{T} \sum_{t=1}^T Y_t^2 \right) \quad k \in \{1, \dots, T\} \end{aligned} \quad (4)$$

Далее среди всех возможных значений статистики  $KL$  выбирается  $k$  по следующему правилу:

$$\tau^* = \min\{k : |KL(k)| = \max_{j \in \{1, \dots, T\}} |KL(j)|\} \quad (5)$$

Для проверки критерия наличия структурного сдвига, введём следующую статистику:

$$v_{r,T}^2 = \sum_{|j| \leq r} w_j \hat{c}_j, \quad \text{где} \quad (6)$$

$$w_j = 1 - \frac{|j|}{r+1}, \quad r \in N \quad (7)$$

$$\hat{c}_j = \frac{1}{T} \sum_{i=1}^{T-|j|} (Y_i^2 - \bar{Y}^2)(Y_{i+|j|}^2 - \bar{Y}^2) \quad (8)$$

Из теоремы 1 следует, что про  $T \rightarrow \infty$  и  $\frac{r}{T} \rightarrow 0$  в случае отсутствия структурного сдвига имеет место сходимость по распределению:

$$\frac{|KL(\tau^*)|}{v_{r,T}^2} \xrightarrow{d} \sup_{0 \leq u \leq 1} |B^0(u)| \quad (9)$$

где  $B^0(u)$ ,  $u \in [0; 1]$ , - процесс броуновского моста. При этом эмпирически  $r$  может быть выбрана двумя способами (ссылка на статьи):

$$1. \quad r = \lfloor \ln T \rfloor$$

$$2. \quad r = \lfloor \sqrt{T} \rfloor$$

Тогда критерий наличия структурного сдвига можно определить следующим образом: если  $\frac{|KL(\tau^*)|}{v_{r,T}^2} \geq q_p$ , то гипотеза об отсутствии структурного сдвига в момент времени  $\tau^*$  отвергается на уровне значимости  $1 - p$ , где  $q_{0.95} = 1.358$  и  $q_{0.99} = 1.628$  - значение квантилей 0.95 и 0.99 соответственно для  $\sup_{0 \leq u \leq 1} |B^0(u)|$  (супремум броуновского моста)

## 2 Итеративные алгоритмы для поиска всех возможных сдвигов

Идея процедуры поиска всех возможных структурных сдвигов во временном ряду заключается в последовательном применении алгоритма KL-CUSUM (KL) (описанный в разделе 1) к временным рядам полученных в результате деления исходного ряда по предполагаемым сдвигам  $\tau_j$ .

Соответственно, задача поиска формируется стандартным способом: рассматриваются подмножество исходного временного ряда границами которых являются предполагаемые структурные сдвиги. Алгоритм можно представить в виде последовательных шагов:

1. Находим первый сдвиг  $\tau_0$
2. Движемся влево меняя правую границу постоянно (т.е. рассматриваем новый ряд  $Y [0; KL(Y [0; T])]$ , продолжая операцию до тех пор пока предполагаемые сдвиги существуют слева)

$$\tau_{first} = KL(Y [0; KL(Y [0; \dots KL(Y [0; T])])]) \quad (10)$$

3. Аналогично, находим новую правую границу

$$\tau_{last} = KL(Y [KL(Y [\dots KL(Y [0; T])]); T]) \quad (11)$$

4. Теперь рассматриваем новый ряд  $Y [\tau_{first} + 1; \tau_{last}]$
5. Далее в полученный вектор сдвигов временного ряда добавляются точки начала и конца исходного ряда (т.е.  $\tau_{vec} = (0, \dots, T)$ )
6. После формируются интервалы вида  $[\tau_{j-1} + 1; \tau_{j+1}]$  И применяется алгоритм KL для проверки является ли точка сдвигом или нет.

На языке описания алгоритмов процесс поиска всех возможных структурных сдвигов выглядит следующим образом [ICSS алгоритм 1]:



---

**Algorithm 1:** Алгоритм распространения близости
 

---

**Function** *ICSS* (*Y*)

```

   $t_1 = 1$ 
  // применяем метод для определения первого сдвига
  KL-CUSUM(Y [ $t_1 : T$ ])
  if KL-CUSUM then
    // если сдвиг был найден считаем его правой границей
    // нового ряда, который вложен в Y
     $t_2 = \tau^*(Y [t_1 : T])$ 
    while not KL-CUSUM(Y [ $t_1 : t_2$ ]) do
      |  $t_2 = \tau^*(Y [t_1 : t_2])$ 
    end
     $\tau_{first} = t_2$ 
    while not KL-CUSUM(Y [ $t_1 : t_2$ ]) do
      |  $t_1 = \tau^*(Y [t_1 : t_2])$ 
    end
     $\tau_{last} = t_1 - 1$ 
    if  $\tau_{first} = \tau_{last}$  then
      | return  $\tau_{first}$  // считается что сдвиг только один
    end
    if  $\tau_{first} < \tau_{last}$  then
      |  $\tau_{vec} \leftarrow (\tau_{first}, \tau_{last})$ 
      | ICSS(Y[ $\tau_{first} + 1; \tau_{last}$ ])
    else
      |  $\tau_{vec} \leftarrow (0, T)$ 
      | sort( $\tau_{vec}$ , type="ascending") // Сортируем всевозможные
      |   моменты сдвигов по возрастанию
      | foreach  $\tau_j$  do
      |   | if KL-CUSUM(Y [ $\tau_{j-1} + 1 : \tau_{j+1}$ ]) then
      |   |   |  $ts\_breakpoints \leftarrow \tau_j$ 
      |   |   end
      |   end
      | return  $ts\_breakpoints$  // Массив уже не будет содержать
      |   начальную и конечные точки
    end
  else
    | return no change-points found
  end

```

---

### 3 Апробация методов на реальных данных

Для проверки метод были реализованы алгоритмы на языке программирования Python. Проверка проводилась на выборках указанных в .

Данные представляют собой цены закрытия по инструменту с интервалом один день и преобразованные по правилу:

$$Y_t = \ln \frac{y_t}{y_{t-1}} \quad (12)$$

Все выборки были протестированы на отсутствие единичных корней (стационарности) тестом Дики-Фуллера (его расширенной версией):

- QIWI:MCX - Результат теста = (0.0) - Гипотеза о единичном корне отклоняется;
- LKOH:MCX - Результат теста = (0.0) - Гипотеза о единичном корне отклоняется;
- RTKMP:MCX - Результат теста =  $(2.26e - 20$  или  $2.26 \times 10^{-20})$  что близко к нулю и гипотеза о единичном корне отклоняется.

#### Рассмотрим QIWI:MCX

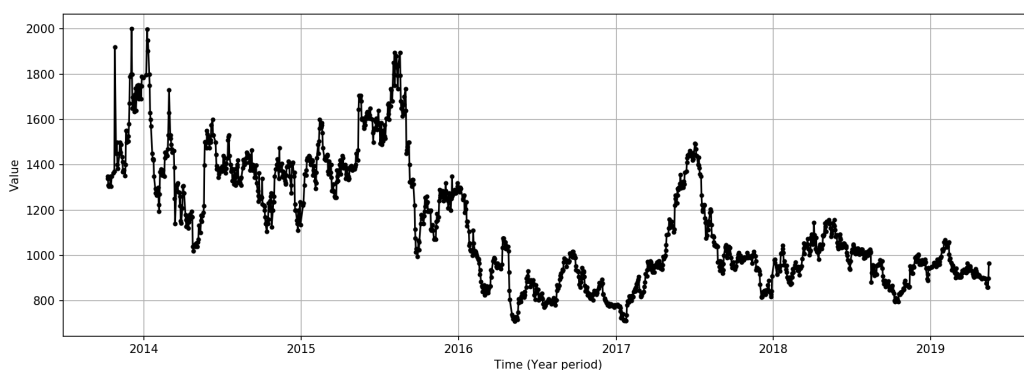


Рис. 1: Цена акций Qiwi

Преобразовав ряд по формуле (12) получаем:

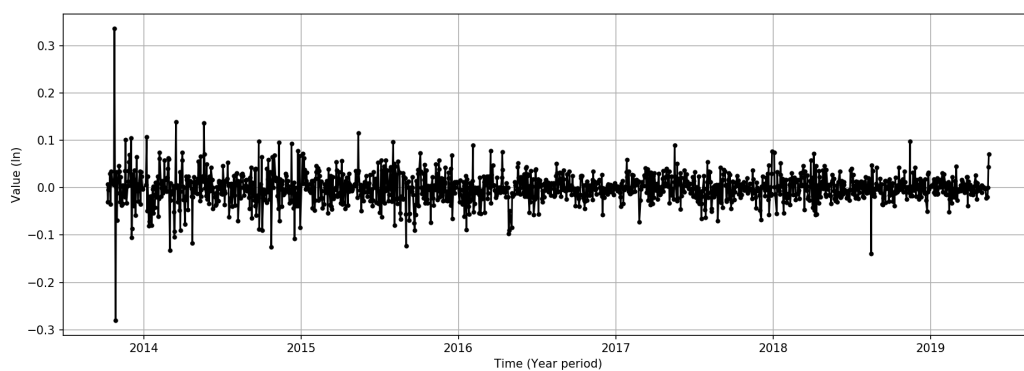


Рис. 2: Натуральный логарифм отношения цены во времени с лагом 1

Теперь применим ICSS процедуру (код реализации доступен в репозитории 1)

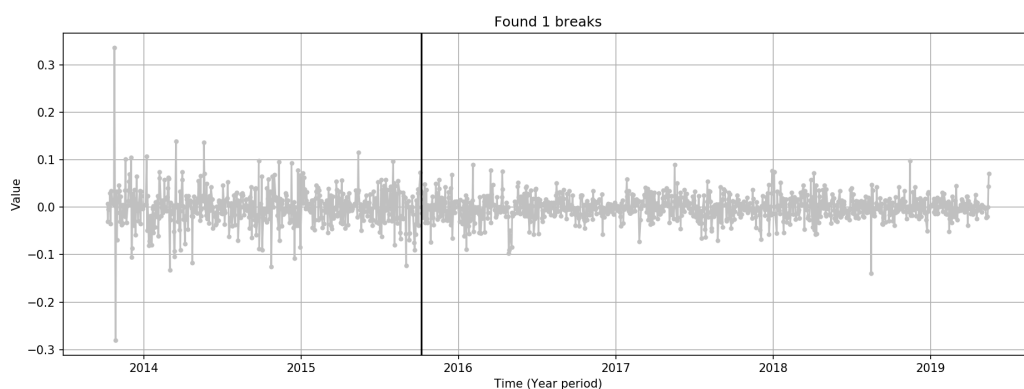


Рис. 3: Структурные сдвиги временного ряда

В результате метод обнаружил 1 возможный сдвига в точке:

1. 2015-10-07 - Qiwi начинает работу на мировом рынке. Это привело к росту заинтересованности компанией со стороны инвесторов.

**Рассмотрим LKOH:MCX**

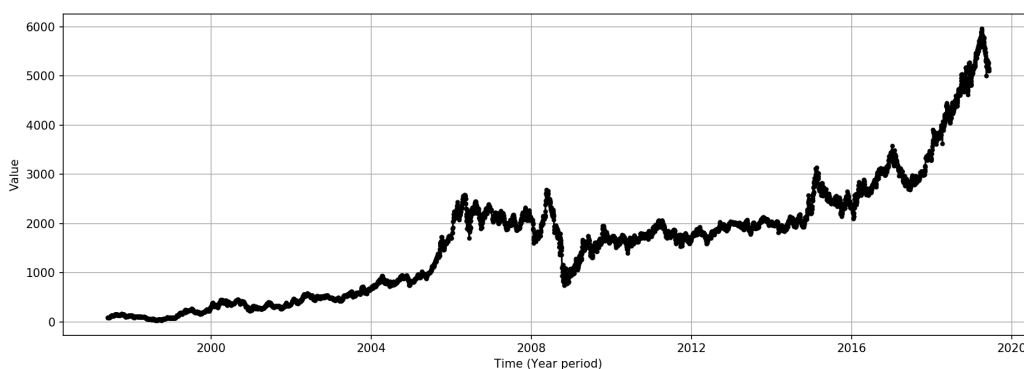


Рис. 4: Цена акции Лукойл

Выполнив аналогичное преобразование (12)

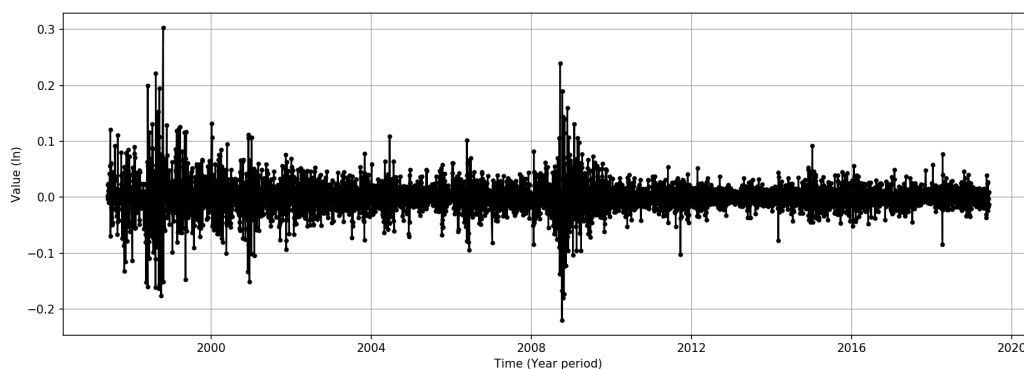


Рис. 5: Натуральный логарифм отношения цены во времени с лагом 1

В результате применения ICSS процедуры было получено 3 возможных сдвига временного ряда:

1. 2001-01-30 - British Petroleum объявляет о намерении продать 7% акций ЛУКОЙЛа.
2. 2008-07-24 - Публикация оценок прибыли за первое полугодие 2008 года. Расширения рынка сбыта (Lukoil Eurasia Petrol A.S. приобретает 100% акций турецкой сбытовой компании Akpet)
3. 2009-11-17 - Выход из тенденции падения добычи нефти. На изменение положительно повлияло несколько факторов: приближение даты публикации годовых отчётов; общее увеличение добычи нефти; покупка новых НПЗ в Голландии и победа в тендере на разработку месторождения Западная Курна-2 в Ираке.

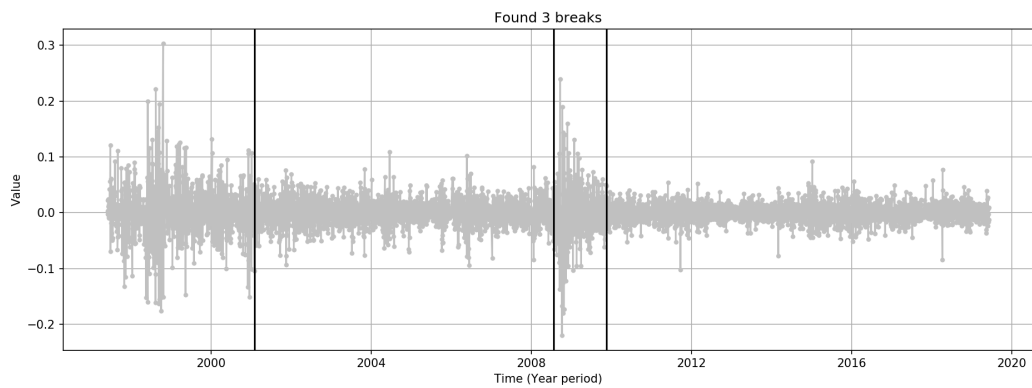


Рис. 6: Структурные сдвиги временного ряда для ЛКОН

Рассмотрим РТКМР:МСХ

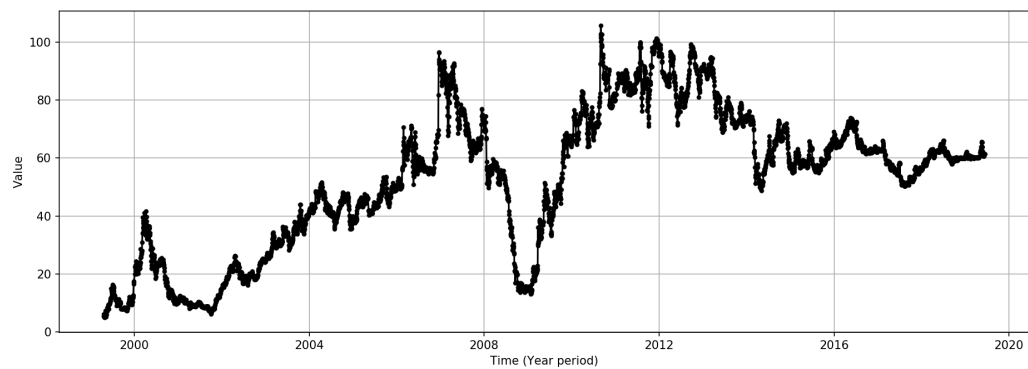


Рис. 7: Цена акций Ростелеком

Преобразовав ряд по формуле (12) получаем:

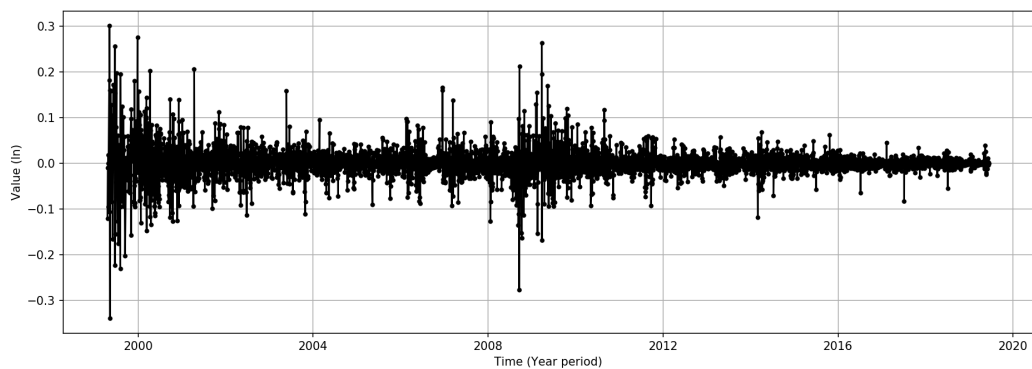


Рис. 8: Натуральный логарифм отношения цены во времени с лагом 1

Теперь применим ICSS процедуру:

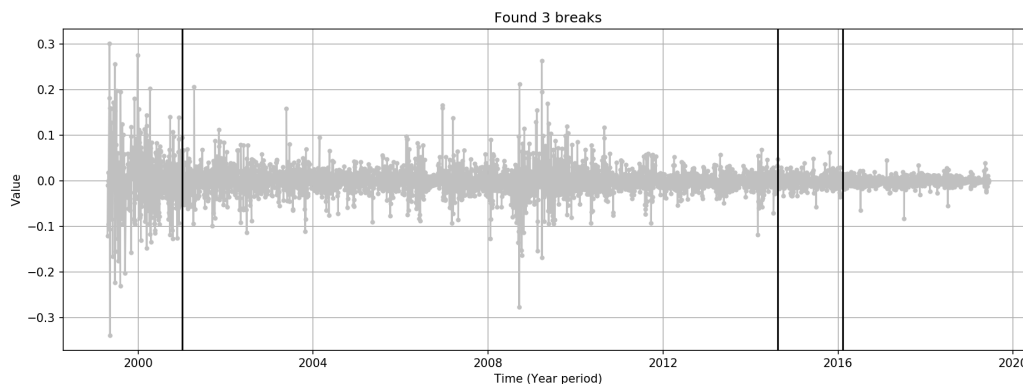


Рис. 9: Структурные сдвиги временного ряда

В результате метод обнаружил 3 возможных сдвига в точках:

1. 2001-01-05 - Обвал индекса Nasdaq. Так же это связано с таким эффектом как "пузырь доткомов". Т.е. повышенный интерес к интернет стартапам на тот момент привёл как серьёзному оттоку средств во время кризиса.
2. 2014-08-14 - Завершение слияния сотовых активов Tele2 и Ростелеком. Таким образом Ростелеком передал 100% акций "РТ-Мобайл"
3. 2016-02-11 - Продолжительный рост интереса инвесторов к Ростелеком и объявление о конкурсе на предоставление кредитной линии.

## Дальнейшее исследование

Важно заметить, что в работе мы использовали ICSS алгоритм, который не гарантирует что все сдвиги были найдены и точность найденных сдвигов может варьироваться на разных наборах данных. Поэтому данное исследование можно продолжить более детальным изучением поведения алгоритма на разнородных данных (возможно синтетических). Так же применение этим исследованиям можно найти не только в экономике, но и в других областях где временной ряд имеет место быть. Например, возможность применения алгоритмов обнаружения сдвигов для задача мониторинга в инженерии, медицине и др.

Так же работу можно продолжить в области исследования/разработке новых алгоритмов для детектирования изменений параметров эконометрических моделей и не только.

## Список литературы

- [1] *Carla Inclin, George C. Tiao* (1994) Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance // Journal of the American Statistical Association, Vol. 89, No. 427., pp. 913-923.
- [2] *Kokoszka P., Leipus R.* (2000) Change-point estimation in ARCH models // Bernoulli. V. 6, N3. P. 513–539.
- [3] *Page E. S.* (1955) A test for change in a parameter occurring at an unknown point // Biometrika 42, 523-527.
- [4] *Wichern, D. W., Miller, R. B. and Hsu, D. A.* (1976) Changes of variance in first-order autoregressive time series models - with an application. Appl. Statist. 25, 248-256.
- [5] *Hamilton, J. D.* (1994) Time Series Analysis. Princeton University Press, New Jersey.
- [6] *Zacks, S.* (1983) Survey of classical and Bayesian approaches to the change-point problem : fixed sample and sequential procedures of testing and estimation. In Recent Advances in Statistics. M. H. Rivzi et al. eds. 245-269, Academic Press, New York.
- [7] *Lee S.* (2000) The Cusum Test for Parameter Change in GARCH(1,1) Models
- [8] *Д.А. Борзых, М.А. Хасыков, А.А. Языков* (2017) Численное сравнение V-MLR и CUSUM-методов обнаружение структурных сдвигов для кусочно-заданных GARCH-моделей // Труды МФТИ. Том 9, № 3
- [9] *Д. А. Борзых, М. А. Хасыков, А. А. Языков* (2017) Новый способ обнаружения структурных сдвигов в GARCH-моделях // Национальный исследовательский университет «Высшая школа экономики» // Вестник ВГУ. серия: Экономика И Управление. № 2



- [10] *Д. А. Борзых, М. А. Хасыков* (2018) Процедура уточнения ICSS алгоритма обнаружения структурных сдвигов в GARCH-моделях // Прикладная эконометрика, т. 51, с. 126–139.

## Приложение

1. Код доступен в репозитории *coursework* <https://github.com/Kaftanov/coursework>
2. Данные были взяты с сайта <https://www.finam.ru>