

# ОБНАРЖЕНИЕ КРАТНОЙ РАЗЛАДКИ В ARCH-ПРОЦЕССАХ.

Н. М. Феропонтова

Национальный исследовательский Томский государственный университет

## Введение

В настоящее время различные процессы и явления, происходящие в экономике, являются объектами анализа. И в большинстве своём эти процессы нестационарны и иногда их поведение достаточно сильно изменяется во времени.

Большое число известных процессов описываются параметрическими моделями. И в случае отсутствия изменений в последовательностях наблюдаемых данных методы построения адекватных моделей хорошо изучены.

Построение модели в интервале, содержащем изменение его свойств, приводит к ошибкам модели и как следствие к ошибкам при анализе и построении прогнозов. Поэтому очень важно выделять интервалы, на которых процесс ведет себя неизменно, называемые интервалами стационарности.

Целью данной работы является рассмотрение, реализация и выбор необходимых параметров для процедур обнаружения разладок в ARCH-процессах, которые описывают такие финансовые показатели как цены акций, биржевые индексы и другие. Рассматриваемые процедуры – это процедура BASTA, обнаружения кратной разладки, и процедура CUSUM для обнаружения одиночной разладки.

Таким образом поставленная задача – это обнаружение одиночной и кратной разладки в ARCH-процессах. Предполагается, что в неизвестный момент разладки  $\tau$  параметры модели изменяются скачкообразно, при том что влияние шумов, распределение которых неизвестно, остаётся неизменным. В случае кратной разладки, этих изменений несколько и моменты их наступления в поставленной задаче также неизвестны.

## 1. Описание модели.

Модель ARCH(p), описывающая процесс  $h_t$ , относится к числу нелинейных стохастических условно-гауссовских моделей в дискретном времени.

Согласно этой модели, величины  $h_t$  удовлетворяют уравнению:

$$h_t = \sigma_t \varepsilon_t, \quad (1)$$

где «волатильность»  $\sigma_t$  представляет собой условную дисперсию, которая является линейной функцией квадратов прошлых возмущений, т.е.

$$\sigma_t^2 = \alpha_0^0 + \sum_{i=1}^p \alpha_i^0 h_{t-i}^2,$$

где  $\varepsilon = \{\varepsilon_t\}_{t \geq 0}$  - последовательность независимых нормально распределенных случайных величин с математическим ожиданием  $M\{\varepsilon_t\} = 0$  и дисперсией  $D\{\varepsilon_t^2\} = 1$ , а  $\alpha_0^0, \alpha_i^0, i = 1 \dots p$  – параметры. Для обеспечения положительности условной дисперсии параметры

должны удовлетворять следующим условиям  $\alpha_0^0 > 0$ ,  $\alpha_i^0 \geq 0$ ,  $i = 1 \dots p$ , ( $p$  - порядок модели).

## 2. Метод трансформации и бинарной сегментации BASTA для обнаружения кратной разладки.

*Постановка задачи.* Рассматривается последовательность  $(h_t)_{0 \leq t \leq T}$  наблюдений на промежутке  $[0, T]$ , удовлетворяющих ARCH(p)-модели, которая имеет  $N$  моментов изменения параметров («разладок»)  $\tau_i, i = 0, \dots, N$ , таких, что  $0 < \tau_1 < \dots < \tau_N < T - 1$  ( $\tau_0 = 0, \tau_{N+1} = T - 1$ ). В каждый из моментов разладки  $\tau_i, i = 0, \dots, N$  вектор параметров  $\{\alpha_j(\tau_i)\}_i$  модели изменяется скачком, т.е.  $\alpha_j(\tau_i) \neq \alpha_j(\tau_{i-1})$ . Предполагается, что количество разладок  $N$ , а также моменты их наступления  $\tau_i$  заранее неизвестны. Задача состоит в том, чтобы оценить количество моментов разладки и их месторасположения.

При описании процедуры BASTA (Binary segmentation for transformed auto-regressive conditional heteroscedasticity) будем следовать работе Фрайзлевича и Субба Рао, опубликованной в 2013 году, [1]. Алгоритм включает в себя 2 этапа:

- 1) трансформация данных
- 2) бинарная сегментация.

На первом этапе осуществляется преобразование данных с помощью некоторой функции  $g(\cdot)$  в последовательность  $U_t$ :

$$U_t = g(h_t, h_{t-1}, \dots, h_{t-T}).$$

Виды преобразующих функций  $g(\cdot)$  делятся на 2 класса, в одном из которых используются остатки, а в другом – усреднение. Назовем алгоритм, применяющий остатки для построения преобразующей функции, BASTA-res, а алгоритм, использующий усреднение, BASTA-avg. Виды конкретных функций можно найти в [1]. Применение разных преобразований ведет к одной цели – уменьшению автокорреляции данных и снижению неоднородности дисперсии.

На втором этапе алгоритма BASTA выполняется бинарная сегментация преобразованной на первом этапе последовательности  $U_t$  с целью обнаружения точек «разладки». Метод бинарной сегментации, представляющий собой итерационную процедуру, реализует рекурсивный поиск кратной «разладки». Для этого вводится дополнительный параметр – пороговая константа, которая определяет момент остановки алгоритма.

### Этап 1. Трансформация данных.

На первом этапе проводится выбор трансформирующей функции, на основе которой формируется последовательность  $U_t = g(h_t, h_{t-1}, \dots, h_{t-T})$ . Наиболее часто используется преобразование

$$U_t = \ln \left( \varepsilon + \frac{h_t^2}{\hat{\alpha}_0 + \sum_{i=1}^p \hat{\alpha}_i h_{t-1}^2 + \varepsilon h_t^2} \right),$$

где  $\varepsilon = 0.001$ ,  $\hat{\alpha}_i$  – оценки параметров.

## Этап 2. Бинарная сегментация

Второй этап представляет собой итерационную процедуру, исходные данные для которой – преобразованная на первом этапе последовательность  $U_t$ , объем данных  $T$ , а также дополнительные параметры, задаваемые по умолчанию, и зависящие от выбора трансформирующей функции.

До начала работы алгоритма также необходимо определить значение пороговой константы  $\tilde{b}_T$ .

Алгоритм бинарной сегментации:

1. Задаются индексы  $(j, l) = (1, 1)$  и начальные значения  $s_{j,l} = 0$ ,  $u_{j,l} = T - 1$ .
2. Положив  $n = u_{j,l} - s_{j,l} + 1$  вычисляются следующие значения

$$\tilde{U}_{s_{j,l}, u_{j,l}}^b = \frac{\sqrt{u_{j,l} - b}}{\sqrt{n(b - s_{j,l} + 1)}} \sum_{t=s_{j,l}}^b U_t - \frac{\sqrt{b - s_{j,l} + 1}}{\sqrt{n(u_{j,l} - b)}} \sum_{t=b+1}^{u_{j,l}} U_t, \quad (2)$$

для всех  $b \in (s_{j,l}, u_{j,l})$ . И также на каждом интервале  $(s_{j,l}, u_{j,l})$  вычисляется момент «разладки» по формуле

$$b_{j,l} = \arg \max_b \left| \tilde{U}_{s_{j,l}, u_{j,l}}^b \right|.$$

3. Для заданной пороговой константы  $\tilde{b}_T$  проверяется неравенство  $\left| \tilde{U}_{s_{j,l}, u_{j,l}}^b \right| < \tilde{b}_T$ , если оно выполняется, то закончить алгоритм на интервале  $[s_{j,l}, u_{j,l}]$ . Иначе фиксируется факт наличия разладки и момент  $b_{j,l}$  добавляется к множеству найденных точек разладки (англ. change-points) и продолжается алгоритм:
  - а) запомнить значения индексов  $(j_0, l_0) = (j, l)$ , далее изменить следующие значения:  $(s_{j+1, 2l-1}, u_{j+1, 2l-1}) = (s_{j,l}, b_{j,l})$ ,  $j = j + 1$ ,  $l = 2l - 1$  и перейти на шаг 2.
  - б) вернуть  $(j, l) = (j_0, l_0)$ , сохраненные в а) и изменив значения  $(s_{j+1, 2l}, u_{j+1, 2l}) = (b_{j,l} + 1, u_{j,l})$ ,  $j = j + 1$ ,  $l = 2l$  перейти на шаг 2.

Этапы 3а) и 3б) описывают бинарную рекурсию влево и вправо от каждой обнаруженной «разладки», именно поэтому алгоритм носит название бинарной сегментации.

Полученные в результате работы алгоритма оценки  $N$  точек «разладки» и их месторасположений  $b_{j,l}$  располагаются в возрастающем порядке, т.е. мы получаем  $\hat{\tau}_1 < \dots < \hat{\tau}_N$ . Пороговая константа  $\tilde{b}_T$  определяет минимальные величины интервалов  $[s_{j,l}, u_{j,l}]$ .

## 3. Метод CUSUM обнаружения одиночной разладки.

Для решения задачи обнаружения одиночной разладки в ARCH-процессах рассмотрим последовательность  $T$  величин, удовлетворяющих (1). Предположим, что параметры процесса изменяются скачком в некоторый неизвестный момент времени  $\tau^*$ ,  $0 < \tau^* < T$ . Векторы параметров модели до и после разладки отличаются по крайней мере по одной координате. Согласно алгоритму CUSUM оценка  $\hat{\tau}$  момента «разладки»  $\tau^*$  определяется из уравнения:

$$\hat{t} = \min \left\{ t : |H_t| = \max_{1 \leq j \leq T} H_j \right\}, \quad (3)$$

где

$$H_t = \frac{t(T-t)}{T^2} \left( \frac{1}{t} \sum_{j=1}^t h_j^2 - \frac{1}{T-t} \sum_{j=t+1}^T h_j^2 \right). \quad (4)$$

Этот алгоритм дает приемлемое качество обнаружения моментов разладки в случае, если разладка произошла в средней части отрезка наблюдений, т.е. достаточно отдаленно от начала и конца выборки.

#### 4. Численное моделирование.

На практике были рассмотрены 3 модели:

1. ARCH(1)-модель без разладки с параметрами  $\alpha_0 = 0.091, \alpha_1 = 0.48$
2. ARCH(1)-модель с одиночной разладкой в момент  $\tau = 941$ .  
Параметры до и после разладки:  
 $\alpha_0^{(0)} = 0.091, \alpha_1^{(0)} = 0.48$   
 $\alpha_0^{(1)} = 0.055, \alpha_1^{(1)} = 0.33$
3. ARCH(1)-модель с двумя разладками в моменты  $\tau_1 = 350$  и  $\tau_2 = 735$ .  
Параметры изменяются в моменты  $\tau_1$  и  $\tau_2$ :  
 $\alpha_0^{(0)} = 0.091, \alpha_1^{(0)} = 0.48$   
 $\alpha_0^{(1)} = 0.055, \alpha_1^{(1)} = 0.21$   
 $\alpha_0^{(2)} = 0.078, \alpha_1^{(2)} = 0.69$

Для реализации алгоритмов были построены оценки параметров этих моделей по методу максимального правдоподобия.

*Результаты работы алгоритмов.*

На этапе моделирования и исследования ARCH(1)-процессов с разладками и без разладок были получены следующие результаты ( $\hat{t}^{(n)}$  – момент ложной тревоги):

| $T = 1500$ | разладки<br>нет       | $\tau = 941$      | $\tau_1 = 350,$<br>$\tau_2 = 735$ |
|------------|-----------------------|-------------------|-----------------------------------|
| BASTA      | $\hat{t}^{(n)} = 839$ | $\hat{t}_B = 910$ | $\hat{t}_B = 869$                 |
| CUSUM      | $\hat{t}^{(n)} = 839$ | $\hat{t}_C = 837$ | $\hat{t}_C = 869$                 |

Выводы:

- 1) Оценки моментов разладок по методу BASTA и CUSUM достаточно близки.
- 2) Обе процедуры работоспособны.
- 3) Процедура BASTA сложнее в реализации, однако не требует наличия априорной информации о числе разладок в ARCH(1)-процессе.
- 4) CUSUM более прост для реализации, однако рассчитан на обнаружения одиночной разладки.

## 5. Обнаружение моментов разладки в последовательности реальных данных – доходностей цен акций компании ОАО «Газпром» за период с 11.01.09 по 30.12.13.

Методом BASTA обнаружены 2 момента разладки, соответствующие датам 21 августа 2009 года и 20 декабря 2011 года (156 и 737 значения).

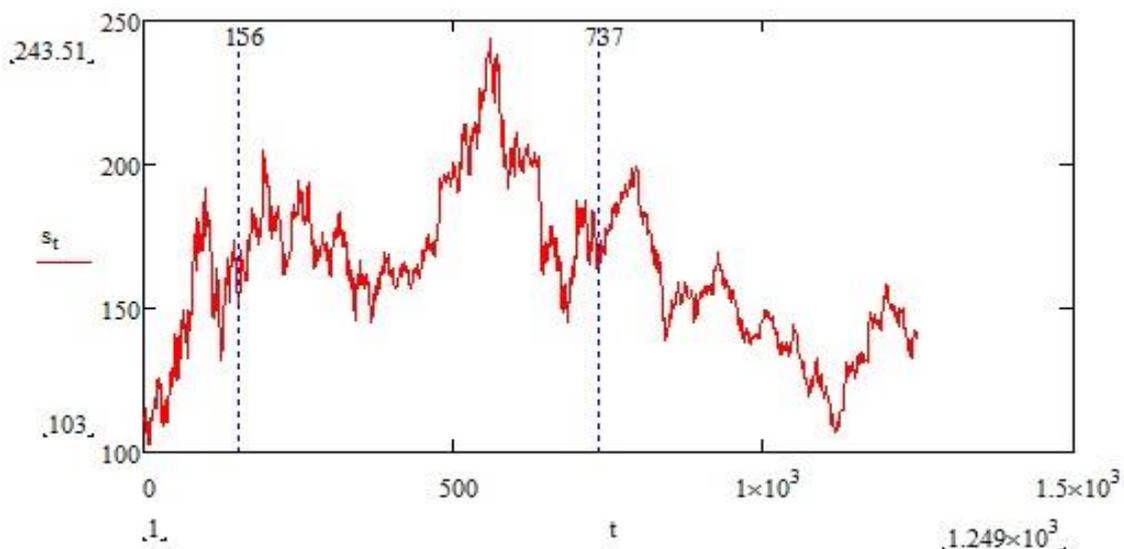


Рис. 1 График цен акций ОАО "Газпром" за период с 11.01.09 по 30.12.13, а также моментов разладок

Данные изменения можно заметить и на рисунке 2 с изображенными на нем доходностями и найденными моментами разладки.

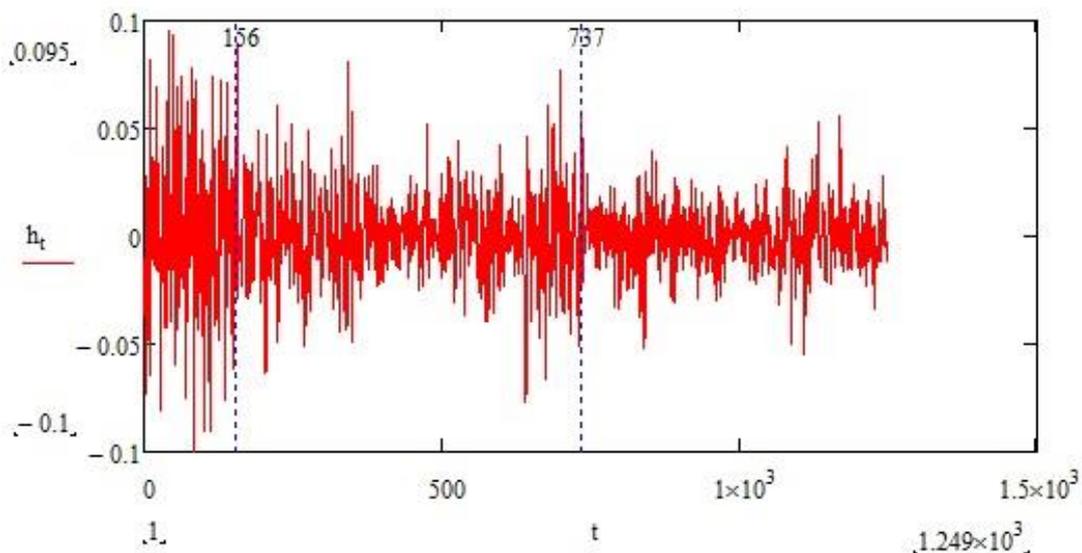


Рис.2. График цен акций ОАО "Газпром" за период с 11.01.09 по 30.12.13 а также моментов разладок.

Алгоритм CUSUM обнаружил также 2 момента разладки (156 и 737 значения), соответствующих датам 22 августа 2009 года и 20 декабря 2011 года. Но CUSUM-алгоритм – это изначально метод обнаружения одиночной разладки и, если мы разделим массив

исходных данных на два равных, и для каждого отдельно проведем CUSUM-поиск разладки, то мы получим 2 момента разладки, отличных от найденного ранее. Необходимо отметить, что эти оценки моментов разладок очень близки к оценкам, обнаруженным методом BASTA, а именно - это значения 157 и 737.

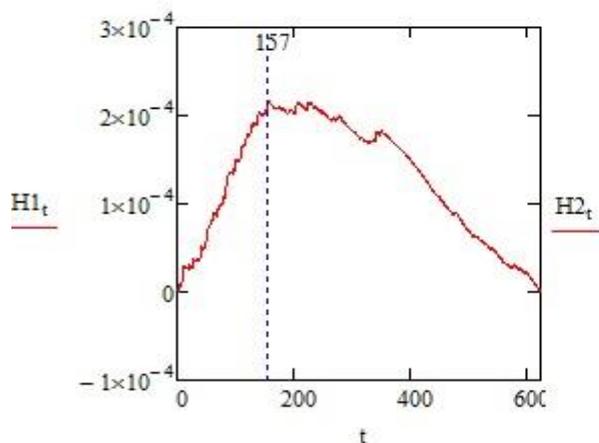


рис.3а.

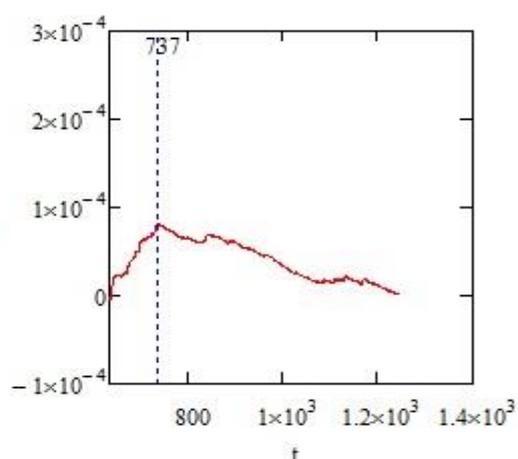


рис.3б.

На рисунках 3а и 3б изображены две последовательности вычисленных статистик  $H_t$  на интервалах  $[0, T/2]$  и  $[T/2, T]$  соответственно.

Заметим, что разбиение на части исходного интервала не противоречит условиям алгоритма CUSUM

Выводы:

1. Оцененные моменты разладок совпали с периодами изменения в налоговом законодательстве для нефтегазового сектора экономики.
2. Один из этих периодов относится к началу августа 2009 года, когда были введены некоторые налоговые послабления во время кризиса 2008.
3. Вторая разладка связана с периодом 2011 года, когда эти льготы были отменены.

### Список литературы:

1. Fryzlewicz P., S. Subba Rao / "BaSTA: consistent multiscale multiple change-point detection for ARCH processes"// Journal of the Royal Statistical Society, 2013. – 35с.
2. Kokoszka P., R. Leipus / "Change-point estimation in ARCH models"// Bernoulli 6(3), 2000, p. 513-539.
3. Basseville M., Nikiforov I.V. Detection of abrupt changes theory and application: New Jersey: Prentice Hall in Information and System Sciences, 1993. – 469с.