

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЕ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет экономических наук

Образовательная программа "Статистическое моделирование и актуарные
расчеты"

Новиков Лев Ильич

*«Метод обнаружения структурных сдвигов в GARCH-моделях с помощью
бутстрэпа»*

Выпускная квалификационная работа – МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ
По направлению подготовки 38.04.01 Экономика

Рецензент
PhD, доцент, Департамент статистики
и анализа данных
Веретенникова Мария Александровна

Руководитель
Доцент, к.э.н.
Фурманов Кирилл
Константинович

Консультант
старший преподаватель,
научный сотрудник
Международной лаборатории
стохастического анализа и
его приложений
Борzych Дмитрий
Александрович

Москва 2020

СОДЕРЖАНИЕ

Введение.....	2
Глава 1: Классический CUSUM-тест.....	4
1.1 Свойства финансовых временных рядов.....	4
1.2 Критерий об отсутствии структурных сдвигов.....	6
1.3 Критерий, выявляющий наличие одного структурного сдвига.....	8
1.4 ICSS-метод.....	10
Глава 2. Метод выявления структурных сдвигов на основе бутстрэпа.....	14
2.1 Модифицирование критерия с помощью независимых инноваций: критерий проверки на отсутствие структурных сдвигов.....	14
2.2 Усовершенствованный критерий проверки нулевой гипотезы о наличии одного структурного сдвига.....	16
2.3 Критерий, с помощью бутстрэпа проверяющий нулевую гипотезу об отсутствии структурных сдвигов (VS-0) против LRD-модели.....	18
2.4 Критерий, с помощью бутстрэпа проверяющий нулевую гипотезу о наличии одного структурного сдвига (VS-1) против LRD-модели.....	19
2.5 Реализация алгоритма и его исследование на симулированных данных.....	21
2.6 Структурные сдвиги в реальных данных.....	23
Заключение.....	26
Список литературы.....	27
Приложение.....	29

ВВЕДЕНИЕ

В процессе оценивания параметров любой эконометрической модели необходимо учитывать стандартные требования к моделям, например, такие, как $\varepsilon \sim N(0, \sigma^2)$ или независимость ошибок регрессии между собой. Для состоятельности оценок параметров требуется сравнительно большое количество наблюдений, и если обеспечить независимость ошибок в большинстве случаев можно, то большое количество наблюдений при работе с временными рядами доступно не всегда. Обычно при построении прогноза доступны только наблюдения до текущего момента времени (Т) включительно, что значительно ухудшает оценки коэффициентов модели.

Но даже если нехватки в наблюдениях нет, со временем параметры модели, описывающей данные, могут случайным образом меняться. Такая ситуация характерна для моделирования стоимости ценных бумаг, поскольку состояние финансового рынка в значительной степени зависит от новостей: так, при смене новостного фона, параметры модели могут измениться скачкообразно (например, вызывая резкий рост волатильности). Подобные моменты времени, в которые наша модель перестает объяснять имеющиеся закономерности, называются разладками случайного процесса или структурными сдвигами. Существуют различные методы выявления структурных сдвигов, однако большинство из них в своей основе имеет критерий основанный на асимптотических квантилях, что вызывает искажение в случае маленьких выборок.

В реальной жизни, достаточно часто при оценивании параметра целью является как можно быстрее узнать об изменениях в закономерностях на финансовом рынке при моделировании цены актива, что уменьшает количество доступных наблюдений. В связи с этим упомянутые выше методы дают неточные результаты – определяют время сдвига неточно или не выявляют его вовсе. Поэтому актуальной задачей в анализе временных рядов остается выявление структурных сдвигов при малом числе доступных наблюдений.

Объектом нашего исследования являются методы обнаружения структурных сдвигов в GARCH-моделях, а **предметом** – метод обнаружения структурных сдвигов с помощью бутстрэпа.

Цель данной работы – исследование метода обнаружения структурных сдвигов в GARCH-моделях с помощью бутстрэпа и его проверка на симулированных и реальных данных. Чтобы достичь указанной цели, мы поставили перед собой следующие **задачи**:

- изучить теоретическое обоснование работы ICSS-метода для обнаружения структурных сдвигов;
- протестировать на симулированных данных CUSUM-тест, лежащий в основе ICSS-метода;
- изучить теоретическое обоснование работы метода на основе бутстрэпа для обнаружения структурных сдвигов;
- протестировать метод на основе бутстрэпа на симулированных данных;

– сравнить результаты тестирования CUSUM-теста и метода на основе бутстрэпа на симулированных данных;

– сделать выводы о преимуществах и недостатках обоих методов, а также проверить их работу на реальных данных.

Основная использованная нами **литература** включает в себя следующие работы: классическая статья Inclan C. и Tiao G. C. "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance"¹, которая описывает алгоритм выявления произвольного числа моментов разладки; две последовательно вышедшие статьи^{2 3} Kokoszka P. и Leipus R., показывающие состоятельность CUSUM-оценки момента разладки и устанавливающие скорость сходимости этой оценки. Помимо упомянутых выше рассматривается статья "Tests for volatility shifts in GARCH against long-range dependence"⁴ Lee T., Kim M. и Baek C. и статья "Bootstrap-based Test for Volatility Shifts in GARCH Against Long-range Dependence"⁵ Wang Y., Park C. и Lee T., посвященная методу, основанному на применении бутстрэпа.

Научная новизна нашей работы заключается в сравнении работы классического CUSUM-теста и нового метода, использующего бутстрэп для улучшения CUSUM-теста, на примере альтернативной гипотезы об изменении параметра w в GARCH-модели. Значимость работы состоит в том, что мы показали нецелесообразность применения данного метода для улучшения оценок параметров в случае маленьких выборок.

Работа состоит из введения, двух глав, заключения, списка литературы и приложений.

¹ Inclan C., Tiao G. C. Use of cumulative sums of squares for retrospective detection of changes of variance // Journal of the American Statistical Association. – 1994. – Т. 89. – №. 427. – С. 913-923.

² Kokoszka P., Leipus R. Testing for parameter changes in ARCH models //Lithuanian Mathematical Journal. – 1999. – Т. 39. – №. 2. – С. 182-195.

³ Kokoszka P., Leipus R. Change-point estimation in ARCH models //Bernoulli. – 2000. – Т. 6. – №. 3. – С. 513-539.

⁴ Lee T., Kim M., Baek C. Tests for Volatility Shifts in Garch Against Long-Range Dependence //Journal of Time Series Analysis. – 2015. – Т. 36. – №. 2. – С. 127-153.

⁵ Wang Y., Park C., Lee T. Bootstrap-based test for volatility shifts in GARCH against long-range dependence // Communications for Statistical Applications and Methods. – 2015. – Т. 22. – №. 5. – С. 495-506.

ГЛАВА 1: КЛАССИЧЕСКИЙ CUSUM-ТЕСТ

1.1 Свойства финансовых временных рядов

В сфере финансов волатильность – это мера изменения доходности финансового инструмента с течением времени. Измерение волатильности является одним из наиболее важных шагов в финансовом моделировании. Модель волатильности может использоваться для прогнозирования абсолютной величины доходности, стоимости активов и распределения весов активов в портфеле.

Большинству финансовых временных рядов свойственны следующие факты⁶:

- Цены финансовых активов в большинстве случаев не стационарны, в отличие от их доходностей;
- Автокорреляция для ряда доходностей как правило низкая или совсем отсутствует;
- Для волатильности ряда доходностей чаще всего наблюдается кластеризация.

Для большинства финансовых рядов естественной мерой волатильности является логарифмическая доходность, определяемая как $r_t = \log P_t - \log P_{t-1}$, где P_t – цена базового актива или биржевого индекса в момент времени t .

Для объяснения большинства имеющихся закономерностей финансовых временных рядов Engle R. F. ещё в 1982 году определил процесс авторегрессионной условной гетероскедастичности (ARCH) и упомянул несколько привлекательных характеристик ARCH моделей⁷. Однако, более общим процессом является обобщенная модель авторегрессионной условной гетероскедастичности (GARCH), которая является популярным инструментом для анализа волатильности финансовых временных рядов с момента появления у Bollerslev T. в статье 1986 г.⁸

Модель GARCH (p, q) задаётся следующими уравнениями:

$$r_t = \varepsilon_t, \quad \varepsilon_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = w + \sum_{i=1}^q \alpha_i r_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

где (w, α, β) – параметры процесса, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ – инновации, последовательность независимых, одинаково распределённых стандартизованных случайных величин, (то есть таких, что $\mathbb{E}\varepsilon_t = 0$, $\mathbb{E}\varepsilon_t^2 = 1$), а σ_t^2 – условная дисперсия процесса Y_t . Для обеспечения положительности условной дисперсии предполагается, что $w > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ $\forall i = 1, \dots, q, j = 1, \dots, p$. Говорят, что волатильность высоко устойчива, если сумма оцененных коэффициентов при квадратах лагов доходностей и условных дисперсиях в модели GARCH близка к 1. Случай, когда $\alpha + \beta = 1$, был назван IGARCH моделью и

⁶ Engle R. F., Patton A. J. What good is a volatility model? //Quantitative Finance. – 2001. – Т. 1. – С. 237–245.

⁷ Engle R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation //Econometrica: Journal of the Econometric Society. – 1982. – С. 987-1008.

⁸ Bollerslev T. Generalized autoregressive conditional heteroskedasticity //Journal of econometrics. – 1986. – Т. 31. – №. 3. – С. 307-327.

является одним из способов описать высокую устойчивость волатильности. Однако, возникают сомнения, что IGARCH является лучшим выбором при моделировании финансовых временных рядов: если в данных присутствует структурный сдвиг или если спецификация модели определена неверно, то будут получаться неопределенные результаты⁹.

Другим способом справиться с высокой устойчивостью волатильности – это использовать моделирование долгосрочной зависимости (LRD) на основе преобразованных данных, например, квадратов доходностей. Задолго до того, как были разработаны подходящие модели, было отмечено, что между наблюдениями замечаются корреляции, которые затухают до нуля с медленной скоростью. Ding Z., Granger C. W. J. и Engle R. F. исследовали свойство долговременной памяти ряда доходностей фондового рынка и установили, что степенное преобразование модулей доходностей имеет довольно высокие автокорреляции для длинных лагов¹⁰. Временной ряд LRD $\{x_t\}_{t \in \mathbb{Z}}$ определяется как модель стационарного временного ряда второго порядка с медленно затухающей автоковариационной функцией,

$\gamma(h) = \text{Cov}(x_t, x_{t+h}) \sim Ch^{2d-1}$ при $h \rightarrow \infty$, где $C > 0$ – константа, а $d \in (0, \frac{1}{2})$ – LRD параметр.

Однако, LRD-модели обладают нестационарными свойствами, из-за чего ведутся споры о возможности их использования. Поскольку LRD имеет медленно затухающую положительную автоковариационную функцию, это означает, что сдвиг волатильности может приводить к "поддельной" LRD-модели¹¹. Таким образом, до того, как оценивать волатильность временного ряда, следует принять решение о том, какой вид имеет процесс, порождающий наблюдаемые данные (DGP): наличие изменения волатильности (volatility shift, VS) или модель с длинной памятью (LRD).

Все предложенные тесты для различения LRD и VS делятся на два класса согласно нулевой гипотезе:

1. Нулевой гипотезой выступает LRD модель^{12 13};
2. Нулевой гипотезой считается модель, имеющая структурный сдвиг¹⁴.

⁹ Mikosch T., Stărică C. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects // Review of Economics and Statistics. – 2004. – Т. 86. – №. 1. – С. 378-390.

¹⁰ Ding Z., Granger C. W. J., Engle R. F. A long memory property of stock market returns and a new model // Journal of empirical finance. – 1993. – Т. 1. – №. 1. – С. 83-106.

¹¹ Mikosch T., Stărică C. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects // Review of Economics and Statistics. – 2004. – Т. 86. – №. 1. – С. 378-390.

¹² Qu Z. A test against spurious long memory // Journal of Business & Economic Statistics. – 2011. – Т. 29. – №. 3. – С. 423-438.

¹³ Kuswanto H. A new simple test against spurious long memory using temporal aggregation // Journal of Statistical Computation and Simulation. – 2011. – Т. 81. – №. 10. – С. 1297-1311.

¹⁴ Zhang A., Gabrys R., Kokoszka P. Discriminating between long memory and volatility shifts // Austrian Journal of Statistics. – 2007. – Т. 36. – №. 4. – С. 253-275.

В нашей работе мы будем придерживаться последнего подхода. Значит, критерий должен проверять H_0 : VS-R модель против альтернативы H_1 : LRD-модель. Здесь и далее мы будем кратко обозначать модель с R структурными сдвигами как VS-R.

Как уже было упомянуто ранее, r_t между собой не коррелируют, в отличие от их квадратов r_t^2 (поскольку $\mathbb{E}\epsilon_t = 0$, но $\mathbb{E}\epsilon_t^2 = 1$). В связи с этим в качестве критерия рассматривается CUSUM-тест для кумулятивной суммы *квадратов* рекурсивных остатков.

1.2 Критерий об отсутствии структурных сдвигов

Рассмотрим критерий об отсутствии структурных сдвигов: учитывая всё вышесказанное, мы будем проверять гипотезу H_0 (наблюдения $\{r_t\}_{t \in \mathbb{Z}}$ соответствуют VS-0 модели) против альтернативы H_1 (наблюдения $\{r_t^2\}_{t \in \mathbb{Z}}$ соответствуют LRD-модели).

В статье Basrak B., Davis R. A., Mikosch T.¹⁵ доказывается, что модели GARCH(p, q) являются SRD (краткосрочно зависимыми), и что любое степенное преобразование доходностей $\{|r_t|^d\}_{t \in \mathbb{Z}}$, $d > 0$ так же будет SRD. Вследствие этого, вышеупомянутый критерий эффективно различает SRD и LRD модели.

В 1999 г. Kokoszka P. и Leipus R. показали, что при выполнении нулевой гипотезы, существовании моментов восьмого порядка, и $n \rightarrow \infty$, процесс сходится по распределению:

$$KL(k) = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^k r_i^2 - \frac{k}{n} \sum_{i=1}^n r_i^2 \right) \xrightarrow{d} \sigma B^0 \left(\frac{k}{n} \right)$$

(сходимость процесса в пространстве Скорохода $D[0,1]$), где σ – истинное стандартное отклонение модели, а $B^0(t)$, $t \in [0; 1]$ – процесс Броуновского моста¹⁶.

Более того, было доказано¹⁷, что это утверждение также верно для оценки σ^2 :
 $\hat{s}_{n,q}^2 = \sum_{j=-q}^q \omega_j(q) \hat{\gamma}_j$, где $\hat{\gamma}_j = \frac{1}{n} \sum_{i=1}^{n-|j|} (X_i - \bar{X})(X_{i+|j|} - \bar{X})$ – выборочные ковариации,
 $\omega_j(q) = \left(1 - \frac{|j|}{q+1}\right) \cdot I\{|j| \leq q\}$ – веса Бартлетта (коэффициенты для треугольного ядра), а

¹⁵ Basrak B., Davis R. A., Mikosch T. Regular variation of GARCH processes // Stochastic processes and their applications. – 2002. – Т. 99. – №. 1. – С. 95-115.

¹⁶ Kokoszka P., Leipus R. Testing for parameter changes in ARCH models. С. 185.

¹⁷ Berkes I. et al. On discriminating between long-range dependence and changes in mean // The annals of statistics. – 2006. – Т. 34. – №. 3. – С. 1140-1165.

$q \in n$ – константа, зависящая от n . Единственное условие, которое накладывается на q – сходимость $\frac{q}{n}$ при T , стремящемся к бесконечности: $\lim_{n \rightarrow \infty} \left(\frac{q}{n}\right) = 0$.

Итак, $\frac{KL(k)}{\hat{s}_{n,q}} \rightarrow B^0(k)$. Таким образом, у нас есть возможность свести нашу задачу к исследованию статистики, поведение которой нам хорошо известно – статистике супремума модуля броуновского моста. Нулевая гипотеза об отсутствии структурного сдвига отвергается на уровне значимости $1-p$ при $\frac{|KL(k^*)|}{\hat{s}_{n,q}} \geq q_p$, где q_p – квантиль уровня p для супремума модуля броуновского моста, а k^* – момент структурного сдвига для альтернативной гипотезы.

Поскольку очевидно, что таких k^* может быть много, логично положить в формуле выше $k^* = \min\{k: |KL(k)| = \max |KL(k)|\}$.

В дальнейшем будем обозначать за $T_n(k)$ нормированный процесс $\frac{|KL(k)|}{\hat{s}_{n,q}}$, а за $T_n = \max \left(\frac{|KL(k)|}{\hat{s}_{n,q}} \right)$ – статистику данного CUSUM-теста.

Распределение статистики супремума модуля броуновского моста было найдено Resnick S. I. в 1992 году¹⁸ и задается функцией

$$Pr\left(\sup_{0 \leq k \leq 1} |B^0(k)| \leq v\right) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 v^2}, \quad \text{где } v > 0$$

Асимптотические квантили для T_n – квантили для $\sup |B^0(k)|$ равны

$$q_{0.9} = 1.224, \quad q_{0.95} = 1.358, \quad q_{0.99} = 1.628.$$

Пример применения критерия: симуляция GARCH(1, 1)-процесса с одним структурным сдвигом

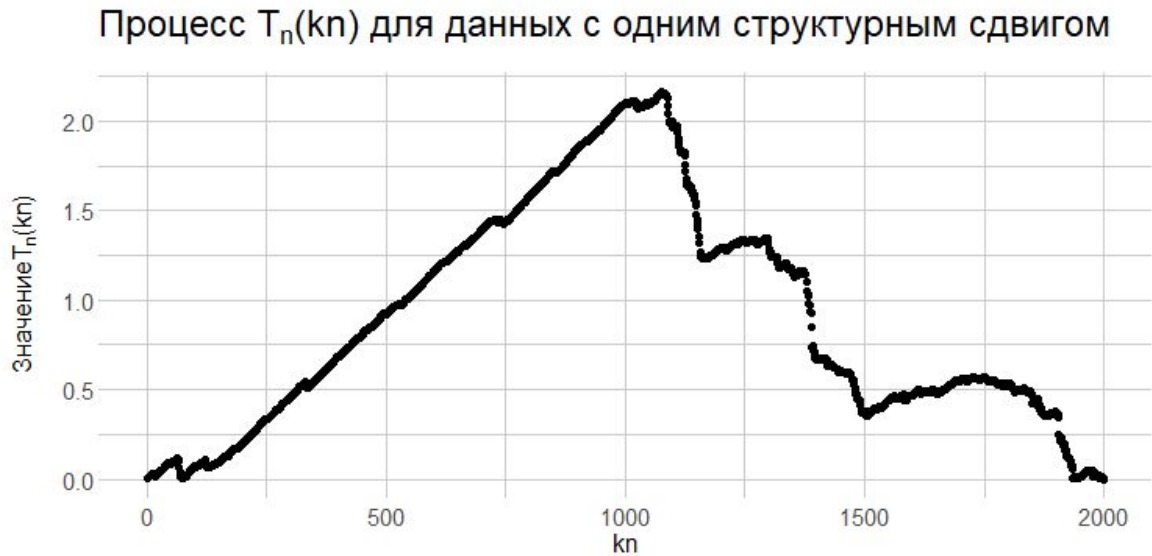
В качестве примера рассмотрим результат работы указанного критерия на 2000 симулированных наблюдений, соответствующим GARCH(1, 1) со структурным сдвигом на 1001 наблюдении.

Данные и параметры для обеих GARCH моделей указаны в таблице ниже:

Параметры			
	w	β	α
$t \leq 1000$	0,1	0,7	0,2
$t \geq 1001$	0,3	0,7	0,2

¹⁸ Resnick S. I. Adventures in stochastic processes. Birkhäuser, Boston. – 1992. – С. 533.

Процесс $T_n(k)$ для этого временного ряда изображен на данном графике:



Максимум указанного процесса достигается на $\hat{k} = 1076$ -ом наблюдении, что в данном случае означает наличие там структурного сдвига: максимальное значение процесса на указанном промежутке $[0, 2000]$ составляет $T_n(1076) = 2.165$, что больше $1.628 = q_{0.99}$, и гипотеза об отсутствии структурного сдвига отвергается на 1%-ном уровне значимости.

Как мы можем видеть, точность работы алгоритма ещё предстоит оценить: истинный момент структурного сдвига $k^* = 1001$. При этом, в данном примере нам было заранее известно, что сдвиг в модели в точности один, а в реальной жизни сдвигов может быть произвольное количество.

1.3 Критерий, выявляющий наличие одного структурного сдвига

Рассмотрим критерий, выявляющий наличие одного структурного сдвига.

Проверяется нулевая гипотеза H_0 (наблюдения $\{r_t\}_{t \in Z}$ соответствуют VS-1 модели) против альтернативы H_1 (наблюдения $\{r_t^2\}_{t \in Z}$ соответствуют LRD-модели).

При нулевой гипотезе получаем, что

$$\begin{cases} r_{1,t} = \sigma_{1,t}\epsilon_t, & \sigma_{1,t}^2 = w_1^\circ + \sum_{i=1}^q \alpha_{1,i}^\circ r_{t-i}^2 + \sum_{j=1}^p \beta_{1,j}^\circ \sigma_{1,t-j}^2 \\ r_{2,t} = \sigma_{2,t}\epsilon_t, & \sigma_{2,t}^2 = w_2^\circ + \sum_{i=1}^q \alpha_{2,i}^\circ r_{t-i}^2 + \sum_{j=1}^p \beta_{2,j}^\circ \sigma_{2,t-j}^2 \end{cases}$$

где инновации $\{\epsilon_t\}_{t \in \mathbb{Z}}$ – стандартизованные, независимые, одинаково распределенные случайные величины, k^* – неизвестный момент структурного сдвига и

$$r_t = \begin{cases} r_{1,t}, & 1 \leq t < k^* \\ r_{2,t}, & k^* \leq t \leq n \end{cases}$$

Согласно Kokoszka P. и Leipus R., k^* может быть оценено¹⁹ как

$$\hat{k} = \arg \max_{1 \leq k \leq n} \left| \frac{(n-k)k}{n^2} \left(\frac{1}{k} \sum_{i=1}^k r_i^2 - \frac{1}{n-k} \sum_{i=1}^n r_i^2 \right) \right|$$

Они же показали, что $|\hat{k} - k^*| = O(1)$, в силу чего данная оценка является состоятельной. Вследствие этого свойства может быть предъявлена следующая статистика^{20 21} для проверки гипотезы о наличии единственного структурного сдвига:

$M_n = \max(T_{n_1}, T_{n_2})$, где статистики

$$T_{n_1} = \frac{1}{\hat{s}_{N,q} \sqrt{\hat{k}}} \max_{1 \leq k \leq \hat{k}} \left| \sum_{i=1}^k r_i^2 - \frac{k}{\hat{k}} \sum_{i=1}^{\hat{k}} r_i^2 \right|,$$

$$T_{n_2} = \frac{1}{\hat{s}_{N,q} \sqrt{n - \hat{k}}} \max_{\hat{k} \leq k \leq n} \left| \sum_{i=\hat{k}+1}^k r_i^2 - \frac{k - \hat{k}}{n - \hat{k}} \sum_{i=\hat{k}+1}^n r_i^2 \right|$$

аналогичны статистике T_n , а \hat{k} – оценка момента структурного сдвига согласно формуле выше. Идея заключается в том, что при верной нулевой гипотезе H_0 , T_{n_1} и T_{n_2} имеют одинаковое асимптотическое распределение, как если бы наблюдения $1 - \hat{k}$ и наблюдения $\hat{k} + 1 - n$ не содержали структурных сдвигов. Это происходит из-за того, что разница между \hat{k} и k^* асимптотически пренебрежимо мала. При верной же альтернативе, T_{n_1} и T_{n_2} стремятся к бесконечности.

¹⁹ Kokoszka P., Leipus R. Change-point estimation in ARCH models. – С. 527.

²⁰ Zhang A., Gabrys R., Kokoszka P. Discriminating between long memory and volatility shifts. – С. 257.

²¹ Lee T., Kim M., Baek C. Tests for Volatility Shifts in Garch Against Long-Range Dependence // Journal of Time Series Analysis. – 2015. – Т. 36. – №. 2. – С. 132.

Итого, при нулевой гипотезе, статистика $M_n = \max(T_{n_1}, T_{n_2})$ сходится по распределению $M_n \xrightarrow{d} \max(\sup_{0 \leq k \leq 1} |B_1^0(k)|, \sup_{0 \leq k \leq 1} |B_2^0(k)|)$, где B_1^0, B_2^0 – независимые Броуновские мосты²².

Асимптотические квантили для M_n – квантили для $\max(\sup |B_1^0(k)|, \sup |B_2^0(k)|)$ равны $q_{0.9} = 1.353$, $q_{0.95} = 1.478$, $q_{0.99} = 1.730$.

Так как два вышеупомянутых критерия не могут выявить все сдвиги в том случае, если истинное количество сдвигов больше одного, рассмотрим итеративную процедуру Iterated Cumulative Sums of Squares (ICSS), которая обобщает данный тест на случай неизвестного числа структурных сдвигов (впервые идея упомянута в 1994 г. авторами Inclan, Tiao²³, а затем строго описана Bai J.²⁴)

1.4 ICSS-метод

Краткое описание ICSS-процедуры²⁵:

1. Проводим тест на отсутствие структурных сдвигов на заданном уровне значимости. Если нулевая гипотеза не отвергается, то алгоритм останавливается, иначе переходим к шагу 2.
2. Проводим тест на наличие одного структурного сдвига на том же уровне значимости. Если нулевая гипотеза не отвергается, алгоритм останавливается (и возвращает единственный структурный сдвиг в момент \hat{k}). Иначе запоминаем этот момент структурного сдвига и переходим к шагу 3.
3. После нахождения момента \hat{k} структурного сдвига для выборки Y_1, \dots, Y_n , мы рассматриваем левую подвыборку $Y_1, \dots, Y_{\hat{k}-1}$ и проводим тест на наличие одного структурного сдвига в ней, после чего повторяем эти действия с новым найденным \hat{k} , и так до тех пор, пока не окажется, что левая подвыборка не содержит сдвигов. Последний самый левый найденный структурный сдвиг мы запоминаем как \hat{k}_{left} .
4. Повторяем аналогичную процедуру для правых подвыборок, стартуя с выборки $Y_{\hat{k}}, \dots, Y_n$. Полученный в конце самый правый сдвиг запоминаем как \hat{k}_{right} .
5. Итеративно применяем шаги 1-4 для новой выборки $Y_{[\hat{k}_{left}; \hat{k}_{right} - 1]}$, состоящей из самого левого и самого правого структурных сдвигов, найденных на предыдущей

²² Berkes I. et al. On discriminating between long-range dependence and changes in mean. – С. 1146.

²³ Inclan C., Tiao G. C. Use of cumulative sums of squares for retrospective detection of changes of variance // Journal of the American Statistical Association. – 1994. – Т. 89. – №. 427. – С. 913-923.

²⁴ Bai J. Estimating multiple breaks one at a time // Econometric theory. – 1997. – Т. 13. – №. 3. – С. 324.

²⁵ Борзых Д. А., Хасыков М. А. Процедура уточнения ICSS алгоритма обнаружения структурных сдвигов в GARCH-моделях // Прикладная эконометрика. 2018. Т. 51. С. 126-139.

итерации алгоритма, до тех пор пока размер оставшейся выборки не станет слишком мал (используется ограничение $n > 2$).

6. Пусть в итоге всего найдено m возможных моментов структурных сдвигов. Они упорядочиваются по возрастанию и записываются в вектор $K = [k_1, \dots, k_m]$, также туда добавляются значения $k_0 = 1$ и $k_{m+1} = n + 1$. После этого производится проверка: ко всем выборкам $Y[k_{i-1}; k_{i+1} - 1]$, $i = 1, \dots, m$ последовательно применяется тест об отсутствии структурных сдвигов. Если в соответствующем случае гипотеза об отсутствии структурного сдвига отвергается, то точка i сохраняется, иначе — удаляется. Шаг 6 повторяется до тех пор, пока число обнаруженных моментов не станет постоянным, а соответствующие структурные сдвиги на различных итерациях не будут близки друг к другу (в программном коде используется условие — разность между положениями соответствующих структурных сдвигов должна быть не больше четырех наблюдений).

Для предотвращения заикливания процесса перепроверки найденных структурных сдвигов вводится ограничение сверху на максимальное число итераций, в качестве верхней границы используется число 10: на практике алгоритм редко требует больше трёх итераций.

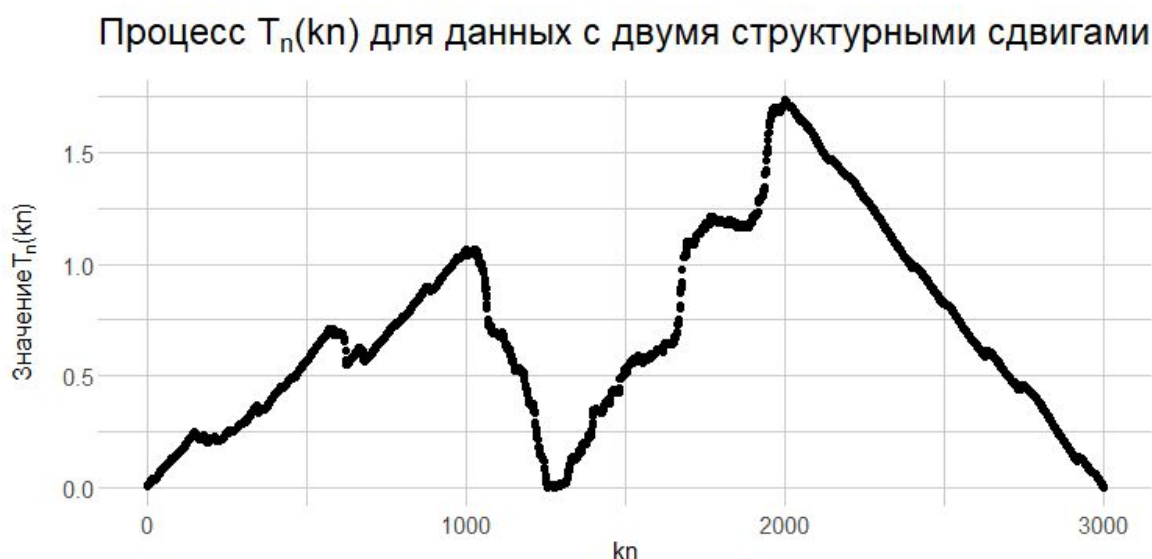
Также для того, чтобы алгоритм работал корректно, вводится дополнительное разумное условие – не проверять выборки единичного размера, так как в противном случае мы получим оценку дисперсии $\hat{s}_{n,q}^2 = 0$.

Пример: симуляция GARCH(1, 1)-процесса с двумя структурными сдвигами

Опять же, будем проверять работоспособность данного алгоритма на симулированных данных, теперь симулируем GARCH(1, 1)-процесс, состоящий из 3000 наблюдений, со структурными сдвигами в моменты времени $k_1^* = 1001$ и $k_2^* = 2001$ и заданными параметрами.

Параметры			
	w	β	α
$t \leq 1000$	0,1	0,7	0,2
$1001 \leq t \leq 2000$	0,3	0,7	0,2
$2001 \leq t \leq 3000$	0,1	0,6	0,2

Процесс $T_n(k)$ для данного временного ряда изображен на графике ниже:



Процедура ICSS выявляет два потенциальных момента структурных сдвигов – наблюдения $\hat{k}_1 = 1030$ и $\hat{k}_2 = 2001$ (два локальных пика на графике), что в данном случае означает наличие там структурных сдвигов (максимальное значение процесса на промежутке составляет $T_n(2001) = 1.738$, что больше $1.628 = q_{0.99}$, и гипотеза об отсутствии структурных сдвигов отвергается на 1%-ном уровне значимости).

Подводя итоги, у нас есть два критерия на основе CUSUM-статистики: для проверки нулевой гипотезы об отсутствии структурных сдвигов и проверки нулевой гипотезы о наличии ровно одного сдвига. Также можно использовать ICSS алгоритм для выявления произвольного количества структурных сдвигов в данных. Однако, несмотря на то, что у нас имеется алгоритм для проверки гипотезы, возникает сразу несколько проблем:

1. В ходе построения статистики $T_n(k)$ требуется оценить дисперсию $\hat{s}_{n,q}^2$. В силу коррелированности наблюдений, оценка дисперсии методом моментов невозможна. При непараметрическом же оценивании дисперсии с помощью треугольного ядра, в том случае, если наблюдения сильно коррелированы (что наблюдается в случае с GARCH(p, q)), оценка будет сильно зависеть от величины носителя ядра. В моей курсовой работе предыдущего года "Методы обнаружения структурных сдвигов в GARCH-моделях"²⁶ при исследовании было зафиксировано $q(n) = \sqrt{n}$ (другие возможные варианты, такие как $q = \ln(n)$ или $\log(n)$ дают гораздо меньшую величину q , из-за чего носитель ядра получается слишком маленьким, и точность оценок ощутимо падает). Хотелось бы получить метод, который не требует настройки гиперпараметров алгоритма.
2. При работе с данными, объём выборки всегда конечен и фиксирован, в связи с чем не соблюдается базовое требование: так как $n \not\rightarrow \infty$, мы не можем утверждать, что

²⁶ Новиков Л.И. Методы обнаружения структурных сдвигов в GARCH-моделях: Курсовая работа, ФЭН НИУ ВШЭ, 2019 (на правах рукописи). С. 2.

$KL(k) = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^k r_i^2 - \frac{k}{n} \sum_{i=1}^n r_i^2 \right) \xrightarrow{d} \sigma B^0 \left(\frac{k}{n} \right)$. И если для больших выборок, содержащих порядка 1000 наблюдений, проблема не так серьезна, то в случае же маленьких выборок результаты работы данного критерия будут сильно искажаться.

В связи с этим, необходимо улучшить наш алгоритм, для того чтобы повысить его точность и получить возможность выявлять структурные сдвиги даже при малом размере выборок.

ГЛАВА 2. МЕТОД ВЫЯВЛЕНИЯ СТРУКТУРНЫХ СДВИГОВ НА ОСНОВЕ БУТСТРЭПА

В 1979 году Efron В. предложил бутстрэп как улучшение метода складного ножа (jackknife)²⁷. Это универсальный метод, который и сейчас активно используется при непараметрическом подходе. В результате применения получается оценить квантили любой статистики, независимо от её сложности.

Классический бутстрэп, если данные $\{x_t\}_{t=1}^n$ являются независимыми наблюдениями, состоит в том, чтобы выбрать m наблюдений $\{x_1^*, \dots, x_m^*\}$, из исходной выборки с возвращением (сгенерировать бутстрэп-выборку), а затем вычислить оценку или интересующую нас статистику t^* на основе выбранных наблюдений. После этого данная процедура повторяется B раз, и в результате получается вектор $\{t_1^*, \dots, t_B^*\}$. Тогда распределение полученного в ходе бутстрэпа вектора аппроксимирует распределение оценки или тестовой статистики при верной нулевой гипотезе. Если мы выполним бутстрэп для тестовой статистики, то правостороннее p -значение для бутстрэпа может быть рассчитано как $\hat{p}^* = \frac{1}{B} \sum_{j=1}^B I(t_j^* \geq t)$, где t – тестовая статистика, оцененная по исходной выборке, а $I(x)$ – индикаторная функция.

В случае же временных рядов, когда наблюдения автокоррелированы, они не являются независимыми, и использование методов генерации повторной выборки может в итоге дать результаты хуже, чем асимптотический подход. В этом случае существует несколько возможных вариантов, как бороться с коррелированностью наблюдений: можно использовать свободный бутстрэп (wild bootstrap)²⁸, или блочный бутстрэп (block bootstrap), а также его вариации, такие как бутстрэп сдвигающимся окном (moving block bootstrap) и стационарный бутстрэп (stationary bootstrap).

Однако, в силу того, что $\{\epsilon_t\}_{t \in \mathbb{Z}}$ являются независимыми одинаково распределенными случайными величинами, возможен и другой путь – рассмотреть классический бутстрэп для инноваций ϵ_t временного ряда r_t .

2.1 Модифицирование критерия с помощью независимых инноваций: критерий проверки на отсутствие структурных сдвигов

В работе 2014 г. Lee Т., Kim М. и Baek С. показали, что при применении статистики, разобранной в § 1.2 к ряду стандартизованных ошибок (инноваций), асимптотическое

²⁷ Efron В. Bootstrap methods: another look at the jackknife // The Annals of Statistics. – 1979. – Т. 7. – №. 1. – С. 1-26.

²⁸ Ahlgren N., Catani P. Wild bootstrap tests for autocorrelation in vector autoregressive models. – 2012. – С. 6.

распределение уже упомянутых выше статистик T_n и M_n при выполнении нескольких условий регулярности остаётся таким же, и сходится к статистике супремума модуля Броуновского моста²⁹.

Условия регулярности для асимптотических свойств статистики критерия об отсутствии структурных сдвигов следующие:

1. Параметры GARCH-модели $\theta = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)$ принадлежат пространству параметров $\Theta \subset [c_1, c_2]^{p+q+1}$ для некоторых $0 < c_1 < c_2 < \infty$ (показано, что точки θ , не являющиеся внутренними точками указанного множества Θ , также удовлетворяют условию)³⁰;
2. Процесс $\{r_t\}_{t=1}^n$ – строго стационарен и $\sup_{\theta \in \Theta} \sum_{j=1}^p \beta_j < 1$.

Тогда, если выполняются условия 1 и 2, то при нулевой гипотезе и $n \rightarrow \infty$ имеется сходимость по распределению $\tilde{T}_n \xrightarrow{d} \sup |B^0(k)|$, где $B^0(k)$ – процесс Броуновского моста, а \tilde{T}_n – статистика, построенная с помощью CUSUM-теста для оценок инноваций ряда доходностей $\{r_t\} \sim \text{GARCH}(p, q)$:

$$\tilde{T}_n = \frac{1}{\hat{\tau}\sqrt{n}} \max_{1 \leq k \leq n} \left| \sum_{i=1}^k \tilde{\epsilon}_i^2 - \frac{k}{n} \sum_{i=1}^n \tilde{\epsilon}_i^2 \right|, \text{ где } \hat{\tau}^2 = \frac{1}{n} \sum_{i=1}^n \tilde{\epsilon}_i^4 - \left(\frac{1}{n} \sum_{i=1}^n \tilde{\epsilon}_i^2 \right)^2$$

– оценка дисперсии квадратов инноваций методом моментов.

Поскольку инновации не наблюдаемы, для применения метода необходимо получить их оценки: $\tilde{\epsilon}_t = \frac{r_t}{\hat{\sigma}_t(\hat{\theta})}$, $t = 1, \dots, n$, где условная дисперсия вычисляется рекурсивно по

$$\tilde{\sigma}_t^2(\hat{\theta}) = \hat{w} + \sum_{i=1}^q \hat{\alpha}_i r_{t-i}^2 + \sum_{j=1}^p \hat{\beta}_j \tilde{\sigma}_{t-j}^2(\hat{\theta})$$

формуле

$\hat{\theta} = (\hat{w}, \hat{\alpha}_1, \dots, \hat{\alpha}_q, \hat{\beta}_1, \dots, \hat{\beta}_p)$ – оцененный вектор параметров GARCH-модели, являющийся решением оптимизационной задачи: $\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \tilde{l}_i(\theta)$, $\tilde{l}_i(\theta) = \frac{r_i^2}{\hat{\sigma}_i^2(\theta)} + \log(\hat{\sigma}_i^2(\theta))$ – оценка метода псевдо-максимального правдоподобия.

²⁹ См. док-во теорем 1 и 3: Lee T., Kim M., Baek C. Tests for Volatility Shifts in Garch Against Long-Range Dependence. – С. 140-152.

³⁰ Francq C., Zakoian J. M. Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero // Stochastic Processes and their Applications. – 2007. – Т. 117. – №. 9. – С. 1268.

2.2 Усовершенствованный критерий проверки нулевой гипотезы о наличии одного структурного сдвига

Проведем аналогичное улучшение для критерия, выявляющего единственный структурный сдвиг – проверяется нулевая гипотеза H_0 (наблюдения $\{r_t\}_{t \in \mathbb{Z}}$ соответствуют VS-1 модели) против альтернативы H_1 (наблюдения $\{r_t\}_{t \in \mathbb{Z}}$ соответствуют LRD-модели).

При нулевой гипотезе со сдвигом в неизвестный момент времени k^* ряд доходностей задаётся двумя GARCH(p, q) моделями с одинаковыми стандартизованными ошибками (инновациями), то есть:

$$\begin{cases} r_{1,t} = \sigma_{1,t} \epsilon_t, & \sigma_{1,t}^2 = w_1^\circ + \sum_{i=1}^q \alpha_{1,i}^\circ r_{t-i}^2 + \sum_{j=1}^p \beta_{1,j}^\circ \sigma_{1,t-j}^2 \\ r_{2,t} = \sigma_{2,t} \epsilon_t, & \sigma_{2,t}^2 = w_2^\circ + \sum_{i=1}^q \alpha_{2,i}^\circ r_{t-i}^2 + \sum_{j=1}^p \beta_{2,j}^\circ \sigma_{2,t-j}^2 \end{cases}$$

где инновации $\{\epsilon_t\}_{t \in \mathbb{Z}}$ – стандартизованные, независимые, одинаково распределенные случайные величины

То есть VS-1 модель определяется как:

$$r_t = \begin{cases} r_{1,t}, & 1 \leq t < k^* \\ r_{2,t}, & k^* \leq t \leq n \end{cases}$$

Для того, чтобы построить статистику критерия, оценкой k^* является

$\hat{k} = \arg \max_{1 \leq k \leq n} \left| \frac{(n-k)k}{n^2} \left(\frac{1}{k} \sum_{i=1}^k r_i^2 - \frac{1}{n-k} \sum_{i=k+1}^n r_i^2 \right) \right|$, предложенная Kokoszka P. и Leipus R.³¹.

Статистика критерия для последовательности инноваций (см. ниже) задаётся как

$$\begin{aligned} \widetilde{M}_n &= \max(\widetilde{T}_{n_1}, \widetilde{T}_{n_2}), \quad \text{где} \quad \widetilde{T}_{n_1} = \frac{1}{\hat{\tau}_1 \sqrt{\hat{k}}} \max_{1 \leq k \leq \hat{k}} \left| \sum_{i=1}^k \tilde{\epsilon}_i^2 - \frac{k}{\hat{k}} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^2 \right| \\ \widetilde{T}_{n_2} &= \frac{1}{\hat{\tau}_2 \sqrt{n-\hat{k}}} \max_{\hat{k} \leq k \leq n} \left| \sum_{i=\hat{k}+1}^k \tilde{\epsilon}_i^2 - \frac{k-\hat{k}}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^2 \right|, \quad \text{и} \\ \hat{\tau}_1^2 &= \frac{1}{k} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^4 - \left(\frac{1}{k} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^2 \right)^2, \quad \hat{\tau}_2^2 = \frac{1}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^4 - \left(\frac{1}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^2 \right)^2 \end{aligned}$$

³¹ Kokoszka P., Leipus R. Change-point estimation in ARCH models. – С. 514.

– оценки дисперсий методом моментов.

Условные дисперсии вычисляются рекурсивно, по отдельности для каждой части, с оцененными векторами параметров GARCH-модели $\hat{\theta}_1$ и $\hat{\theta}_2$. Инновации вычисляются как

$$\tilde{\epsilon}_t = \frac{r_t}{\hat{\sigma}_{1,t}(\hat{\theta}_1)}, t = 1, \dots, \hat{k} - 1 \quad \text{и} \quad \tilde{\epsilon}_t = \frac{r_t}{\hat{\sigma}_{2,t}(\hat{\theta}_2)}, t = \hat{k}, \dots, n$$

В целом процесс аналогичен процессу построения статистики \tilde{T}_n , рассмотренной в предыдущем параграфе.

Для данного критерия добавляется еще одно условие регулярности по сравнению с предыдущим:

1. Параметры $\theta_1 = (\omega_1, \alpha_{1,1}, \dots, \alpha_{1,q}, \beta_{1,1}, \dots, \beta_{1,p})$, $\theta_2 = (\omega_2, \alpha_{2,1}, \dots, \alpha_{2,q}, \beta_{2,1}, \dots, \beta_{2,p})$ обеих GARCH-моделей принадлежат пространству параметров $\Theta \subset [c_1, c_2]^{p+q+1}$ для некоторых $0 < c_1 < c_2 < \infty$: показано, что точки θ , не являющиеся внутренними точками указанного множества Θ , также удовлетворяют условию;
2. Процессы $\{r_{m,t}\}_{t=1}^n$ – строго стационарны и $\sup \sum_{j=1}^p \beta_{m,j} < 1$ (для $m=1, 2$);
3. Величина $\Delta = \frac{\omega_1}{1 - \sum_{i=1}^q \alpha_{1,i} - \sum_{j=1}^p \beta_{1,j}} - \frac{\omega_2}{1 - \sum_{i=1}^q \alpha_{2,i} - \sum_{j=1}^p \beta_{2,j}} \neq 0$.

Тогда, если выполняются условия 1, 2 и 3, то при нулевой гипотезе и $n \rightarrow \infty$, статистика $\tilde{M}_n = \max(\tilde{T}_{n_1}, \tilde{T}_{n_2})$ сходится по распределению к $\max(\sup |B_1^0(k)|, \sup |B_2^0(k)|)$, где B_1^0, B_2^0 – независимые Броуновские мосты³².

Отметим, что добавленное третье условие регулярности здесь устраняет случай, когда изменение параметров не приводит к появлению скачка волатильности.

Итак, подытожим всё то, что мы получили на текущий момент: мы можем проверять гипотезы о наличии и отсутствии структурных сдвигов в рядах логарифмических доходностей с помощью статистик \tilde{T}_n и \tilde{M}_n , поскольку данные статистики имеют известные асимптотические распределения, сводящиеся к статистике супремума модуля Броуновского моста. Более того, данные статистики являются функциями от ряда стандартизованных ошибок (инноваций) $\{\tilde{\epsilon}_t^2\}_{t \in \mathbb{Z}}$, которые одинаково распределены и независимы.

Следовательно, мы можем применить бутстрэп к стандартизованным ошибкам (инновациям), для того чтобы хотя бы отчасти справиться со второй проблемой – недостатком числа наблюдений. Искажение размера имеет место, поскольку мы знаем

³² Berkes I. et al. On discriminating between long-range dependence and changes in mean. – С. 1146.

лишь асимптотическое распределение для $n \rightarrow \infty$, а в реальной ситуации размер выборки стремиться к бесконечности не может.

2.3 Критерий, с помощью бутстрэпа проверяющий нулевую гипотезу об отсутствии структурных сдвигов (VS-0) против LRD-модели

Описание построения критерия может быть также найдено у Yu Wang³³:

1. Для данной выборки $\{r_t\}_{t \in \mathbb{Z}}$ оцениваются коэффициенты GARCH(p, q) согласно оптимизационной задаче $\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \tilde{l}_i(\theta)$, где

$$\tilde{l}_i(\theta) = \frac{r_i^2}{\hat{\sigma}_i^2(\theta)} + \log(\hat{\sigma}_i^2(\theta)) \quad \text{— оценка псевдо-максимального правдоподобия;}$$

2. Рекурсивно оцениваются условные дисперсии $\hat{\sigma}_t^2(\hat{\theta}) = \hat{w} + \sum_{i=1}^q \hat{\alpha}_i r_{t-i}^2 + \sum_{j=1}^p \hat{\beta}_j \hat{\sigma}_{t-j}^2(\hat{\theta})$, после чего вычисляются оценки ряда инноваций $\tilde{\epsilon}_t = \frac{r_t}{\hat{\sigma}_t(\hat{\theta})}, t = 1, \dots, n$;
3. Вычисляется статистика критерия

$$\tilde{T}_n = \frac{1}{\hat{\tau}\sqrt{n}} \max_{1 \leq k \leq n} \left| \sum_{i=1}^k \tilde{\epsilon}_i^2 - \frac{k}{n} \sum_{i=1}^n \tilde{\epsilon}_i^2 \right|, \quad \text{где} \quad \hat{\tau}^2 = \frac{1}{n} \sum_{i=1}^n \tilde{\epsilon}_i^4 - \left(\frac{1}{n} \sum_{i=1}^n \tilde{\epsilon}_i^2 \right)^2;$$

4. Используя бутстрэп, выбираем n наблюдений из исходной выборки $\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_n$ с возвращением (генерируем бутстрэп-выборку инноваций $\tilde{\epsilon}_1^*, \dots, \tilde{\epsilon}_n^*$)

- 4.1. На основе полученной выборки инноваций рекурсивно находим бутстрэп-выборку $r_t^* = \sigma_t^* \tilde{\epsilon}_t^*$, $\sigma_t^{*2}(\hat{\theta}) = \hat{w} + \sum_{i=1}^q \hat{\alpha}_i r_{t-i}^{*2} + \sum_{j=1}^p \hat{\beta}_j \sigma_{t-j}^{*2}(\hat{\theta})$, с фиксированными начальными значениями для $r_0^{*2}, \dots, r_{1-q}^{*2}$ и $\sigma_0^{*2}, \dots, \sigma_{1-p}^{*2}$. В алгоритме в качестве начального приближения использована выборочная дисперсия $\{r_t\}_{t \in \mathbb{Z}}$, однако показано³⁴, что асимптотически начальными приближениями можно пренебречь;

- 4.2. Для полученной бутстрэп-выборки вычисляется статистика критерия $\tilde{T}_n^* = \frac{1}{\hat{\tau}^* \sqrt{n}} \max_{1 \leq k \leq n} \left| \sum_{i=1}^k \tilde{\epsilon}_i^{**2} - \frac{k}{n} \sum_{i=1}^n \tilde{\epsilon}_i^{**2} \right|$, где $\hat{\tau}^{*2}$ — оценки дисперсии $\tilde{\epsilon}_i^{**2}$ методом моментов, а $\{\tilde{\epsilon}_t^{**}\}$ определяется с помощью найденных на предыдущем шаге $\{r_t^*\}$ согласно формуле $\tilde{\epsilon}_t^{**} = \frac{r_t^*}{\hat{\sigma}_t^{**}(\hat{\theta})}, t = 1, \dots, n$. Условные дисперсии находятся рекурсивно

³³ Wang Y., Park C., Lee T. Bootstrap-based test for volatility shifts in GARCH against long-range dependence // Communications for Statistical Applications and Methods. — 2015. — Т. 22. — №. 5. — С. 495-506.

³⁴ См. Лемма 12: Francq C., Zakoian J. M. Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. — С. 1279.

$$\tilde{\sigma}_t^{**2}(\hat{\theta}) = \hat{w}^* + \sum_{i=1}^q \hat{\alpha}_i^* r_{t-i}^{*2} + \sum_{j=1}^p \hat{\beta}_j^* \tilde{\sigma}_{t-j}^{**2}(\hat{\theta}) \quad \text{с фиксированными начальными значениями для } r_0^{*2}, \dots, r_{1-q}^{*2} \text{ и } \tilde{\sigma}_0^{**2}, \dots, \tilde{\sigma}_{1-p}^{**2};$$

5. Шаги 4, 4.1, 4.2 повторяются B раз (в нашем алгоритме для уменьшения трудоемкости используется $B=100$). В итоге получается выборка из бутстрэп-статистик $\tilde{T}_{n_1}^*, \dots, \tilde{T}_{n_B}^*$;
6. Для данного критерия оценивается р-значение: $\hat{p}^* = \frac{1}{B} \sum_{j=1}^B I(\tilde{T}_{n_j}^* \geq \tilde{T}_n)$, исходя из которого делается вывод о том, отвергается или не отвергается нулевая гипотеза об отсутствии структурных сдвигов на данном уровне значимости.

2.4 Критерий, с помощью бутстрэпа проверяющий нулевую гипотезу о наличии одного структурного сдвига (VS-1) против LRD-модели

Описание построения критерия:

1. Для данной выборки получаем оценку момента структурного сдвига $\hat{k} = \arg \max_{1 \leq k \leq n} \left| \frac{(n-k)k}{n^2} \left(\frac{1}{k} \sum_{i=1}^k r_i^2 - \frac{1}{n-k} \sum_{i=k+1}^n r_i^2 \right) \right|$, после чего для $\{\mathbf{r}_t\}_{t \in \mathbb{Z}}$ оцениваются коэффициенты GARCH(p, q) согласно оптимизационной задаче $\hat{\theta}_1 = \arg \min_{\theta \in \Theta} \sum_{i=1}^{\hat{k}-1} \tilde{l}_i(\theta_1)$, $\hat{\theta}_2 = \arg \min_{\theta \in \Theta} \sum_{i=\hat{k}}^n \tilde{l}_i(\theta_2)$,

$$\tilde{l}_i(\theta_m) = \frac{r_i^2}{\tilde{\sigma}_i^2(\theta_m)} + \log(\tilde{\sigma}_i^2(\theta_m)), \quad \tilde{\sigma}_i^2(\theta_m) = \tilde{\sigma}_{1,i}^2(\theta_m)I(i < \hat{k}) + \tilde{\sigma}_{2,i}^2(\theta_m)I(i \geq \hat{k}), \quad m = 1, 2;$$

2. Для обеих выборок (левее и правее момента сдвига \hat{k}) аналогично предыдущему случаю рекурсивно оцениваются условные дисперсии $\tilde{\sigma}_{m,t}^2(\hat{\theta}) = \hat{w}_m + \sum_{i=1}^q \hat{\alpha}_{m,i} r_{m,t-i}^2 + \sum_{j=1}^p \hat{\beta}_{m,j} \tilde{\sigma}_{m,t-j}^2(\hat{\theta})$, $m=1, 2$, после чего вычисляются оценки ряда инноваций $\tilde{\epsilon}_t = \frac{r_t}{\tilde{\sigma}_{m,t}(\hat{\theta})}$, $t = 1, \dots, n$;

3. Вычисляется статистика критерия $\tilde{M}_n = \max(\tilde{T}_{n_1}, \tilde{T}_{n_2})$, где

$$\tilde{T}_{n_1} = \frac{1}{\hat{\tau}_1 \sqrt{\hat{k}}} \max_{1 \leq k \leq \hat{k}} \left| \sum_{i=1}^k \tilde{\epsilon}_i^2 - \frac{k}{\hat{k}} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^2 \right|,$$

$$\tilde{T}_{n_2} = \frac{1}{\hat{\tau}_2 \sqrt{n-\hat{k}}} \max_{\hat{k} \leq k \leq n} \left| \sum_{i=\hat{k}+1}^k \tilde{\epsilon}_i^2 - \frac{k-\hat{k}}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^2 \right|, \quad \text{и}$$

$$\hat{\tau}_1^2 = \frac{1}{\hat{k}} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^4 - \left(\frac{1}{\hat{k}} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^2 \right)^2, \quad \hat{\tau}_2^2 = \frac{1}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^4 - \left(\frac{1}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^2 \right)^2;$$

4. Используя бутстрэп, выбираем n наблюдений из исходной выборки $\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_n$ с возвращением (генерируем бутстрэп-выборку инноваций $\tilde{\epsilon}_1^*, \dots, \tilde{\epsilon}_n^*$)

- 4.1. На основе полученной выборки инноваций рекурсивно находим бутстрэп-выборку согласно формуле $r_t^* = \sigma_t^* \tilde{\epsilon}_t^*$, $\sigma_t^{*2}(\hat{\theta}_1) = \hat{w}_1 + \sum_{i=1}^q \hat{\alpha}_{1,i} r_{t-i}^{*2} + \sum_{j=1}^p \hat{\beta}_{1,j} \sigma_{1,t-j}^{*2}(\hat{\theta}_1)$, $t = 1, \dots, \hat{k} - 1$ и

$\sigma_t^{*2}(\hat{\theta}_2) = \hat{w}_2 + \sum_{i=1}^q \hat{\alpha}_{2,i} r_{t-i}^{*2} + \sum_{j=1}^p \hat{\beta}_{2,j} \sigma_{2,t-j}^{*2}(\hat{\theta}_2)$, параграф $t = \hat{k}, \dots, n$ с фиксированными начальными значениями для $r_0^{*2}, \dots, r_{1-q}^{*2}$ и $\sigma_0^{*2}, \dots, \sigma_{1-p}^{*2}$ для обеих моделей. Так же, как и в алгоритме предыдущего параграфа, для инициализации используется выборочная дисперсия;

4.2. Для полученной бутстрэп-выборки вычисляется статистика критерия

$$\widetilde{M}_n^* = \max(\widetilde{T}_{n_1}^*, \widetilde{T}_{n_2}^*), \quad \text{где} \quad \widetilde{T}_{n_1}^* = \frac{1}{\hat{\tau}_1^* \sqrt{\hat{k}}} \max_{1 \leq k \leq \hat{k}-1} \left| \sum_{i=1}^k \tilde{\epsilon}_i^{**2} - \frac{k}{\hat{k}} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^{**2} \right|,$$

$$\widetilde{T}_{n_2}^* = \frac{1}{\hat{\tau}_2^* \sqrt{n-\hat{k}}} \max_{\hat{k} < k \leq n} \left| \sum_{i=\hat{k}+1}^k \tilde{\epsilon}_i^{**2} - \frac{k-\hat{k}}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^{**2} \right|, \quad \text{и}$$

$$\hat{\tau}_1^2 = \frac{1}{\hat{k}} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^{**4} - \left(\frac{1}{\hat{k}} \sum_{i=1}^{\hat{k}} \tilde{\epsilon}_i^{**2} \right)^2$$

$$\hat{\tau}_2^2 = \frac{1}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^{**4} - \left(\frac{1}{n-\hat{k}} \sum_{i=\hat{k}+1}^n \tilde{\epsilon}_i^{**2} \right)^2, \quad \text{где } \hat{\tau}_m^2 - \text{оценка дисперсии } \tilde{\epsilon}_i^{**2}$$

методом моментов, а $\tilde{\epsilon}_t^{**} = \frac{r_t^*}{\tilde{\sigma}_t^{**}(\hat{\theta})}$, для $t = 1, 2, \dots, n$. Условные дисперсии находятся рекурсивно, согласно уравнениям

$$\tilde{\sigma}_t^{**2}(\hat{\theta}_1) = \hat{w}_1^* + \sum_{i=1}^q \hat{\alpha}_{1,i}^* r_{t-i}^{*2} + \sum_{j=1}^p \hat{\beta}_{1,j}^* \tilde{\sigma}_{1,t-j}^{**2}(\hat{\theta}_1) \quad \text{и}$$

$$\tilde{\sigma}_t^{**2}(\hat{\theta}_2) = \hat{w}_2^* + \sum_{i=1}^q \hat{\alpha}_{2,i}^* r_{t-i}^{*2} + \sum_{j=1}^p \hat{\beta}_{2,j}^* \tilde{\sigma}_{2,t-j}^{**2}(\hat{\theta}_2) \quad \text{с фиксированными}$$

начальными значениями для $r_0^{*2}, \dots, r_{1-q}^{*2}$ и $\tilde{\sigma}_{m,0}^{**2}, \dots, \tilde{\sigma}_{m,1-p}^{**2}$ для обеих моделей;

5. Шаги 4, 4.1, 4.2 повторяются B раз (в нашем алгоритме для уменьшения трудоемкости используется $B=100$). В итоге получается выборка из бутстрэп-статистик $\widetilde{M}_{n_1}^*, \dots, \widetilde{M}_{n_B}^*$;

6. Для данного критерия оценивается p -значение $\hat{p}^* = \frac{1}{B} \sum_{j=1}^B I(\widetilde{M}_{n_j}^* \geq \widetilde{M}_n)$, исходя из которого делается вывод о том, отвергается или не отвергается нулевая гипотеза на данном уровне значимости.

2.5 Реализация алгоритма и его исследование на симулированных данных

На языке R был написан программный код (см. Приложение), реализующий тесты, упомянутые в двух предыдущих параграфах. В ходе проверки работающей программы, возникла проблема в работе алгоритма, связанная с решением оптимизационной задачи $\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \tilde{l}_i(\theta)$: метод, встроенный в пакет fGarch, для оценки ошибок коэффициентов GARCH-модели должен вычислить матрицу (гессиан), и иногда (довольно редко), эта матрица оказывается сингулярной, в результате чего обратная матрица не определена, и ошибки коэффициентов не вычисляются. Поскольку ошибки коэффициентов GARCH в данном исследовании не нужны, такие случаи мы будем обрабатывать как обычно, оставляя вычисленные оценки параметров.

Мы хотим сравнить результаты работы классического алгоритма ICSS (см. § 1.4) и критерия на основе бутстрэпа (см. § 2.2). Уже было показано³⁵, что для данного метода ошибки в спецификации порядка модели не приводят к существенному изменению эмпирической ошибки первого рода по сравнению с номинальным уровнем значимости, то есть, к примеру, если оценивается модель GARCH(1, 1) вместо истинной модели GARCH(2, 1). В связи с этим мы ограничимся рассмотрением модели GARCH(1, 1).

Чтобы получаемые результаты были приближены к реальным данным, мы будем рассматривать такие параметры модели, которые соответствуют реальной ситуации на финансовом рынке. Для этой цели мы собрали данные о ценах закрытия для акций 13 компаний, торгуемых на американской бирже³⁶, за последние 10 лет, построили ряды логарифмических доходностей (см. § 1.1) и для каждого ряда оценили GARCH(1, 1). Итоговые результаты представлены в таблице №1 (рассмотрены компании Apple, Boeing, Chevron, Disney, IBM, Intel Corp., Johnson&Johnson, JPMorgan, Coca-Cola, McDonalds, Microsoft, WalMart, Exxon Mobil).

Параметры оцененных GARCH(1, 1) моделей для логарифмических доходностей акций крупных американских компаний:

Таблица №1													
TICKER	AAPL	BA	CVX	DIS	IBM	INTC	JNJ	JPM	KO	MCD	MSFT	WMT	XOM
α	0.11	0.10	0.10	0.14	0.11	0.07	0.10	0.11	0.07	0.18	0.13	0.32	0.13
β	0.82	0.86	0.87	0.78	0.78	0.88	0.87	0.84	0.88	0.64	0.79	0.54	0.84

³⁵ Lee T., Kim M., Baek C. Tests for Volatility Shifts in Garch Against Long-Range Dependence //Journal of Time Series Analysis. – 2015. – Т. 36. – №. 2. – С. 127-153.

³⁶ Информационный сервис Финам: экспорт котировок

URL: <https://www.finam.ru/profile/akcii-usa-bats/apple-inc/export/?market=25> (дата обращения 15.05.20)

Как следствие, мы будем рассматривать модели GARCH(1, 1) с параметрами $\alpha=0.1$ и $\beta=0.8$. Параметр w для всех оцениваемых рядов очень мал (порядка 10^{-5}) и поэтому в таблице не указан, в нашем исследовании мы положили $w=0.1$. Мы будем сравнивать эмпирические ошибки первого рода и мощности критериев на уровнях значимости 0.05 и 0.01.

Все результаты основаны на 1000 симуляций для выборок размеров 200, 500, и 2000, рассматривается уровень значимости равный 5% и 1%. Предполагается, что стандартизованные ошибки (инновации) имеют стандартное нормальное распределение, $\epsilon_t \sim N(0, 1)$, гиперпараметр нашего алгоритма на основе бутстрэпа устанавливается равным $B=100$.

В таблицах №2-3 представлены эмпирические ошибки первого рода, а в таблицах №4-5 - мощности для проверки нулевой гипотезы $H_0 (\{r_t\}_{t \in Z} \sim VS-0)$ против альтернативы $H_1 (\{r_t^2\}_{t \in Z} \sim \text{LRD-модели})$. В качестве альтернативы для симуляций рассматривается модель со структурным сдвигом на наблюдении $k^*=501$ с изменением параметра w на 0,2, то есть первая половина наблюдений ведёт себя согласно модели GARCH(1, 1) с параметрами (0.1, 0.1, 0.8), а вторая - согласно модели GARCH(1, 1) с параметрами (0.3, 0.1, 0.8).

Таблица №2			
Ошибка первого рода при уровне значимости 0.01			
	n=200	n=500	n=2000
KL	0	0.01	0.005
Bootstrap	0	0.005	0.01

Таблица №3			
Ошибка первого рода при уровне значимости 0.05			
	n=200	n=500	n=2000
KL	0.02	0.07	0.05
Bootstrap	0.03	0.07	0.045

Таблица №4			
Мощность теста при уровне значимости 0.01			
(w_1, α_1, β_1) и (w_2, α_2, β_2)	n=200	n=500	n=2000
KL (0.1, 0.1, 0.8) – (0.3, 0.1, 0.8)	0.11	0.755	0.995
Bootstrap (0.1, 0.1, 0.8) – (0.3, 0.1, 0.8)	0.075	0.195	0.945

Таблица №5			
Мощность теста при уровне значимости 0.05			
(w_1, α_1, β_1) и (w_2, α_2, β_2)	n=200	n=500	n=2000
KL (0.1, 0.1, 0.8) – (0.3, 0.1, 0.8)	0.66	0.95	0.995
Bootstrap (0.1, 0.1, 0.8) - (0.3, 0.1, 0.8)	0.195	0.45	0.995

Как мы видим, при большом размере выборки ошибка первого рода для обоих тестов становится близкой к номинальному уровню значимости 0.01 и 0.05, а эмпирическая мощность стремится к 1, что подтверждается результатами других исследований³⁷. Таким образом, асимптотические свойства у статистики, построенной с использованием бутстрэп-метода, сохраняются. Однако, при малом числе наблюдений новый критерий не даёт большой мощности; более того, мощность критерия оказалась существенно ниже, чем у давно изученного CUSUM-теста, в связи с чем можно сделать вывод о том, что данный метод хорошо себя проявляет лишь на больших выборках.

2.6 Структурные сдвиги в реальных данных

Для выявления структурных сдвигов в реальных данных используем ICSS-алгоритм, способный выявить произвольное их число. Поскольку мы выяснили, что на малых выборках результаты обоих алгоритмов испытывают сильные искажения, возьмём в качестве реальных данных длинные временные ряды - цены закрытия для крупных американских компаний, уже рассмотренных ранее в § 2.5. Наблюдаемый диапазон времени - с 30.04.2010 по 30.04.2020, всего в каждой выборке порядка 2500 наблюдений.

Для всех временных рядов была произведена предобработка на случай наличия в данных сплита или обратного сплита цены акции (в случае, если бы он не был учтен, это был бы выброс во временном ряде логарифмических доходностей). Затем строится ряд

³⁷ Борзых Д. А., Языков А. А. KS-метод обнаружения структурного сдвига в GARCH(1,1) моделях // Прикладная эконометрика. 2019. Т. 54. С. 90-104. doi

логарифмических доходностей $r_t = \log P_t - \log P_{t-1}$, где P_t – цена акции в момент времени t , после чего итеративно выявляются структурные сдвиги.

Результаты для рассмотренных компаний получились следующие:

Таблица №6													
TICKER	AAPL	BA	CVX	DIS	IBM	INTC	JNJ	JPM	KO	MCD	MSFT	WMT	XOM
\hat{k}	-	2533	2539	-	2191	2012	2010	2533	2219	2533	-	1413	-

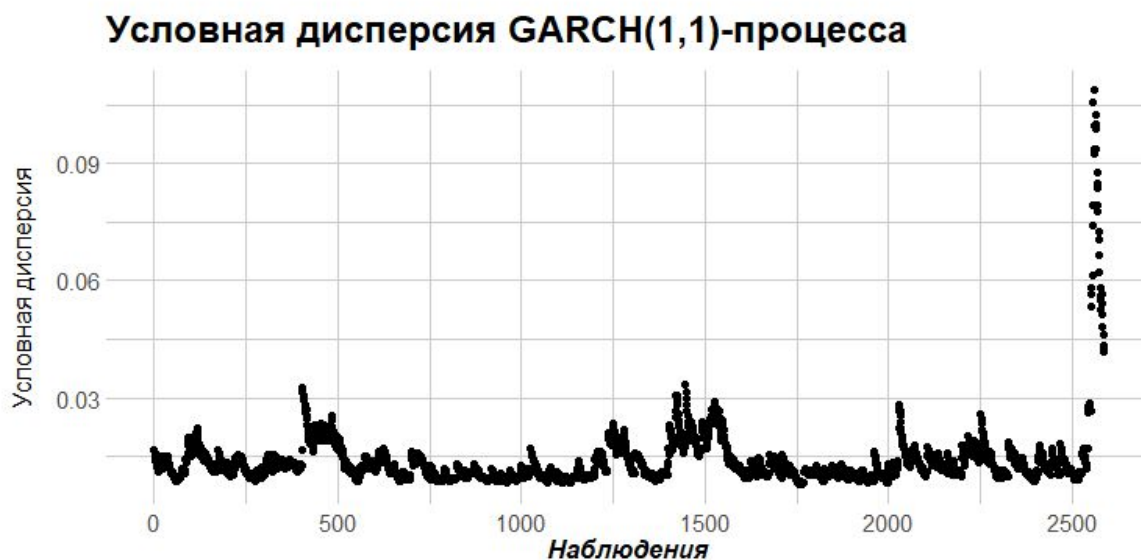
Как мы можем видеть, 4 компании из разных секторов экономики - Boeing, Chevron, JPMorgan и McDonald's согласно тесту имеют структурный сдвиг приблизительно в одно и то же время. Это соответствует 20.02.2020 (для McDonald's 21.02.2020). В это время на финансовом рынке США начался крах из-за развала сделки ОПЕК, а также опасений по поводу коронавирусной инфекции, которые спустя несколько недель получили подтверждение³⁸. В итоге волатильность резко возросла, что и замечает наш критерий.

Данный всплеск волатильности настолько сильный, что замечен даже на маленьком объеме выборки: если взять данные для других компаний за 2019-2020 гг., структурный сдвиг также выявляется для Amazon, Pepsi и Facebook (оценка момента времени структурного сдвига - 29.01.20, 25.02.20 и 20.02.20, соответственно). Полученная информация может быть применена в практических исследованиях, например, при моделировании волатильности при помощи GARCH модели, мы можем улучшить оценки, оценив GARCH модель для каждой из полученных подвыборок. Покажем пример этого для логарифмических доходностей цены акции компании Chevron:

³⁸ Обвал фондового рынка 2020 года. URL:

[\(https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D0%B2%D0%B0%D0%BB_%D1%84%D0%BE%D0%BD%D0%B4%D0%BE%D0%B2%D0%BE%D0%B3%D0%BE_%D1%80%D1%8B%D0%BD%D0%BA%D0%B0_\(2020\)\)](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D0%B2%D0%B0%D0%BB_%D1%84%D0%BE%D0%BD%D0%B4%D0%BE%D0%B2%D0%BE%D0%B3%D0%BE_%D1%80%D1%8B%D0%BD%D0%BA%D0%B0_(2020)) (дата обращения 15.05.20)

Оцениваем GARCH(1, 1)-модель и получаем условные дисперсии для всего временного ряда:



Оцениваем две GARCH(1, 1) модели отдельно:



ЗАКЛЮЧЕНИЕ

Моделирование волатильности является критически важным при работе с временными рядами, однако в реальной ситуации параметры модели, порождающей данные, могут внезапно меняться, в результате чего оценки параметров становятся смещенными, а информация, получаемая из модели - неверной. В связи с этим важно уметь выявлять моменты разладки случайного процесса - структурные сдвиги в модели для наблюдаемой в данных волатильности.

В ходе данной работы мы ознакомились с основными теоретическими результатами и имеющимися методами, позволяющими выявлять структурные сдвиги. Фундаментом для них является ICSS метод, который позволяет применять тест итеративно, выявляя произвольное количество сдвигов. Критически важным для работы алгоритма свойством является состоятельность оценки момента структурного сдвига \hat{k} : при $n \rightarrow \infty$, $|\hat{k} - k^*| = O(1)$, что обосновывает возможность применения ICSS-алгоритма.

Был показан результат работы ICSS-алгоритма на симулированных данных с двумя структурными сдвигами, а также выполнено тестирование CUSUM-теста на симулированных данных с параметрами, приблизительно соответствующим реальным данным. В то время, как для больших размеров выборок наблюдаются хорошие результаты: мощность теста близка к 1, эмпирическая ошибка первого рода стремится к номинальной, в случае же маленькой выборки результаты теста сильно искажаются, мощность падает.

В данной работе была использована классическая бутстрэп-схема для улучшения работы алгоритма: исследовано поведение модифицированной статистики \tilde{T}_n , вычисленной для ряда инноваций и показано, что она обладает теми же асимптотическими свойствами, что и статистика CUSUM-теста, сходясь при $n \rightarrow \infty$ к статистике супремума модуля Броуновского моста. Данный алгоритм был запрограммирован на языке R (см. Приложение), после чего были проведены симуляции на выборках размера 200, 500, и 2000 наблюдений. Ожидалось, что тест на основе бутстрэпа покажет повышение мощности по сравнению с классическим CUSUM-тестом, однако после проведения исследований на симуляциях мы получили, что оба теста показывают примерно одинаковые эмпирические ошибки первого рода на уровнях значимости $\alpha = 0.05$ и $\alpha = 0.01$, их мощности стремятся к 1 при больших размерах выборки, но на малых выборках мощность теста оказывается невысокой: меньше, чем мощность CUSUM-теста. Таким образом, предпочтительнее использовать CUSUM-тест.

Мы протестировали новый метод на реальных данных - логарифмических доходностях акций крупных компаний, и обнаружили, что алгоритм выявляет структурные сдвиги в данных, в том числе точную дату начала последнего кризиса - 20.02.2020. Более того, данный вывод в этом случае можно сделать и для меньшей выборки из 250 наблюдений: структурный сдвиг замечен для доходностей акций Facebook, Amazon, и Pepsi.

СПИСОК ЛИТЕРАТУРЫ

1. Ahlgren N., Catani P. Wild bootstrap tests for autocorrelation in vector autoregressive models. – 2012. С. 1-31.
2. Bai J. Estimating multiple breaks one at a time //Econometric theory. – 1997. – Т. 13. – №. 3. – С. 315-352.
3. Basrak B., Davis R. A., Mikosch T. Regular variation of GARCH processes // Stochastic processes and their applications. – 2002. – Т. 99. – №. 1. – С. 95-115.
4. Berkes I. et al. On discriminating between long-range dependence and changes in mean // The annals of statistics. – 2006. – Т. 34. – №. 3. – С. 1140-1165.
5. Bollerslev T. Generalized autoregressive conditional heteroskedasticity // Journal of econometrics. – 1986. – Т. 31. – №. 3. – С. 307-327.
6. Ding Z., Granger C. W. J., Engle R. F. A long memory property of stock market returns and a new model // Journal of empirical finance. – 1993. – Т. 1. – №. 1. – С. 83-106.
7. Efron B. Bootstrap methods: another look at the jackknife // The Annals of Statistics. – 1979. – Т. 7. – №. 1. – С. 1-26.
8. Engle R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation // Econometrica: Journal of the Econometric Society. – 1982. – С. 987-1008.
9. Engle R. F., Patton A. J. What good is a volatility model? // Quantitative Finance. – 2001. – Т. 1. – С. 237-245.
10. Francq C., Zakoian J. M. Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero //Stochastic Processes and their Applications. – 2007. – Т. 117. – №. 9. – С. 1265-1284.
11. Inclan C., Tiao G. C. Use of cumulative sums of squares for retrospective detection of changes of variance //Journal of the American Statistical Association. – 1994. – Т. 89. – №. 427. – С. 913-923.
12. Kokoszka P., Leipus R. Testing for parameter changes in ARCH models // Lithuanian Mathematical Journal. – 1999. – Т. 39. – №. 2. – С. 182-195.
13. Kokoszka P., Leipus R. Change-point estimation in ARCH models //Bernoulli. – 2000. – Т. 6. – №. 3. – С. 513-539.
14. Kuswanto H. A new simple test against spurious long memory using temporal aggregation // Journal of Statistical Computation and Simulation. – 2011. – Т. 81. – №. 10. – С. 1297-1311.
15. Lee T., Kim M., Baek C. Tests for Volatility Shifts in Garch Against Long-Range Dependence //Journal of Time Series Analysis. – 2015. – Т. 36. – №. 2. – С. 127-153.
16. Mikosch T., Stărică C. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects //Review of Economics and Statistics. – 2004. – Т. 86. – №. 1. – С. 378-390.
17. Qu Z. A test against spurious long memory //Journal of Business & Economic Statistics. – 2011. – Т. 29. – №. 3. – С. 423-438.
18. Resnick S. I. Adventures in stochastic processes. Birkhäuser, Boston. – 1992. – 626 с.

19. Wang Y., Park C., Lee T. Bootstrap-based test for volatility shifts in GARCH against long-range dependence // Communications for Statistical Applications and Methods. – 2015. – Т. 22. – №. 5. – С. 495-506.
20. Zhang A., Gabrys R., Kokoszka P. Discriminating between long memory and volatility shifts // Austrian Journal of Statistics. – 2007. – Т. 36. – №. 4. – С. 253–275.
21. Борzych Д. А., Хасыков М. А. Процедура уточнения ICSS алгоритма обнаружения структурных сдвигов в GARCH-моделях // Прикладная эконометрика. 2018. Т. 51. С. 126-139.
22. Борzych Д. А., Хасыков М. А., Языков А. А. Численное сравнение V-MLR-и CUSUM-методов обнаружения структурных сдвигов для кусочно-заданных GARCH-моделей // Труды Московского физико-технического института. – 2017. – Т. 9. – №. 3 (35).
23. Борzych Д. А., Языков А. А. KS-метод обнаружения структурного сдвига в GARCH(1, 1) моделях // Прикладная эконометрика. – 2019. – Т. 54. – С. 90-104.
24. Новиков Л.И. Методы обнаружения структурных сдвигов в GARCH-моделях: Курсовая работа, ФЭН НИУ ВШЭ, 2019 (на правах рукописи).
25. Информационный сервис Финам: экспорт котировок. URL: <https://www.finam.ru/profile/akcii-usa-bats/apple-inc/export/?market=25> (дата обращения 15.05.20)
26. Обвал фондового рынка 2020 года. URL: [https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D0%B2%D0%B0%D0%BB_%D1%84%D0%BE%D0%BD%D0%B4%D0%BE%D0%B2%D0%BE%D0%B3%D0%BE_%D1%80%D1%8B%D0%BD%D0%BA%D0%B0_\(2020\)](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D0%B2%D0%B0%D0%BB_%D1%84%D0%BE%D0%BD%D0%B4%D0%BE%D0%B2%D0%BE%D0%B3%D0%BE_%D1%80%D1%8B%D0%BD%D0%BA%D0%B0_(2020)) (дата обращения 15.05.20)

ПРИЛОЖЕНИЕ

Использованный код на языке R:

```
# Libraries:
{library("readxl")
 library("fGarch")}

# Functions:
{# Converting and plotting
Preprocess_data <- function(X){
  X <- as.matrix(X)
  storage.mode(X) <- "numeric"
  X <- as.vector(X)
  return(X)
} # preprocess observations
Preprocess_dates <- function(dats){
  dats <- as.matrix(dats)
  storage.mode(dats) <- "character"
  dats <- as.vector(dats)
  return(dats)
} # preprocess dates
# import data and show plot for splits detection
take_data <- function(path, show_plot = TRUE){
  data <- read_xlsx(path, col_names = FALSE, range = cell_cols(5)) # <CLOSE> for data downloaded from finam.ru
  data <- Preprocess_data(data) # convert to numeric vector
  data <- data[2:length(data)] # delete first row with <ticker>
  if (show_plot == TRUE)
    plot(data) # checking for splits
  return(data)
}
dates <- function(path){
  dats <- read_xlsx(path, col_names = FALSE, range = cell_cols(3)) # <DATE> for data downloaded from finam.ru
  dats <- Preprocess_dates(dats) # convert to character vector
  dats <- dats[2:length(dats)] # delete first row with <ticket>
  for (i in 1:length(dats)){ # converting date to type yyyy-mm-dd
    dats[i] <- paste(substr(dats[i], 1, 4), substr(dats[i], 5, 6), substr(dats[i], 7, 8), sep="-")
  }
  return(dats)}
```

```

# detect split (if necessary)
detect_split <- function(data){
  mod1 <- data[2:length(data)]
  mod2 <- data[1:(length(data)-1)]
  ratio <- mod1/mod2
  k1 <- which(ratio == min(ratio)) # candidate for usual split
  k2 <- which(ratio == max(ratio)) # candidate for reversed split
  print(paste0('Возможный момент для сплита акций: Стоимость акции ', data[k2], " до и ", data[k2+1], "
после."))
  print(paste0('Возможный момент для обратного сплита акций: Стоимость акции ', data[k1], " до и ",
data[k1+1], " после."))
  return(c(k1, k2)) # two candidates for split moment
}

# Estimate parameters of GARCH(1,1) and get sigmas
get_params <- function(Y){
  suppressWarnings(est_model <- garchFit(formula = ~ garch(1, 1), data = Y,
include.mean = FALSE, trace = FALSE))
  suppressWarnings(params <- est_model@fit$par[1:3]) # We are getting omega, alpha1 and beta1
  return(params) # comment: warnings are suppressed, because garchFit sometimes have problems with std errors
}
get_sigmas <- function(Y){
  suppressWarnings(est_model <- garchFit(formula = ~ garch(1, 1), data = Y,
include.mean = FALSE, trace = FALSE))
  suppressWarnings(sigmas <- est_model@sigma.t) # We are getting sigmas series
  return(sigmas) # comment: warnings are suppressed, because garchFit sometimes have problems with std errors
}

# Builds bootstrapped observations using bootstrapped innovation series
Obs_boot <- function(inn, innov_boot, coefs) {
  sigmas_boot <- c(var(inn)*(length(inn)-1)/length(inn)) # Bootstrapped sigmas - start with sample variance
  obs_boot <- c(sqrt(var(inn)*(length(inn)-1)/length(inn))*innov_boot[1])

  for (i in 2:length(inn)) {
    new_sigma_boot <- coefs[1] + coefs[2]*(obs_boot[i-1])^2 +
      coefs[3]*sigmas_boot[i-1]
    sigmas_boot <- c(sigmas_boot, new_sigma_boot)
    obs_boot <- c(obs_boot, sqrt(new_sigma_boot)*innov_boot[i])
  }
  return(obs_boot)
}

```

```

}

# Calculates sum of squares for statistic for each observation from the sample
T_k <- function(Y, uncorrelated=TRUE) {
  X <- Y*Y
  N <- length(Y) # number of observations in our sample
  Xmean <- mean(X) # mean value of Y^2 series

  KL <- numeric(0) # for Kokoszka-Leipus statistics
  for (k in 1:N){
    KL <- c(KL, abs(sum(X[1:k]) - k*Xmean)/sqrt(N)) # calculating KL statistic
  }

  # estimation of variance (two options)
  # new version: for innovation series, which are uncorrelated observations
  if (uncorrelated == TRUE){
    Z <- X*X
    s <- sqrt(mean(Z)-mean(X)*mean(X))
  }

  # old version: for series, which are correlated - but Bartlett weights cause size distortion
  if (uncorrelated == FALSE){
    q <- floor(sqrt(N)) # using square root function (logarithm finds more breaks, but also gives more mistakes)
    w <- c((1:q)/(q+1), ((q+1):1)/(q+1)) # Bartlett weights (triangular kernel with window [-q; q])
    C <- c()
    for (j in 0:q){
      Cj <- 0
      for (i in 1:(N-j)){
        Cj <- Cj + (X[i]-Xmean)*(X[i+j]-Xmean) # sample covariances in [-q; q] window
      }
      C <- c(C, Cj/N)
    }
    C <- c(rev(C[-1]), C) # here we use symmetry of covariance to simplify the code
    s <- sqrt(sum(w*C)) # triangular kernel
  }
  return(KL/s)
}

# Calculates k_hat - estimated moment of structural break
k_hat <- function(Y, uncorrelated=TRUE) {
  X <- Y*Y
  N <- length(Y) # number of observations in our sample

```



```

k_hats <- numeric(0)
for (k in 1:(N-1)){
  k_hats <- c(k_hats, abs((N-k)*sum(X[1:k]) - k*sum(X[(k+1):N])))
}
k <- min(which(k_hats == max(k_hats))) # maximum
return(k)
}

# Returns suspicious time moment and value of KL statistic
BB_stat <- function(Y, uncorrelated=TRUE) {
  X <- Y*Y
  N <- length(Y) # number of observations in our sample
  Xmean <- mean(X) # mean value of Y^2 series

  KL <- numeric(0)
  for (k in 1:N){
    KL <- c(KL, abs(sum(X[1:k]) - k*Xmean)/sqrt(N)) # calculating KL statistic
  }
  tau <- min(which(KL == max(KL))) # supremum for statistics

  # estimation of variance
  # for innovation series, which are uncorrelated obs.
  if (uncorrelated == TRUE){
    Z <- X*X
    s <- sqrt(mean(Z)-mean(X)*mean(X))
  }
  # for series, which are correlated
  if (uncorrelated == FALSE){
    q <- floor(sqrt(N)) # using square root function (logarithm finds more breaks, but also gives more mistakes)
    w <- c((1:q)/(q+1), ((q+1):1)/(q+1)) # Bartlett weights (triangular kernel with window [-q; q])
    C <- c()
    for (j in 0:q){
      Cj <- 0
      for (i in 1:(N-j)){
        Cj <- Cj + (X[i]-Xmean)*(X[i+j]-Xmean) # sample covariances in [-q; q] window
      }
      C <- c(C, Cj/N)
    }
    C <- c(rev(C[-1]), C) # here we use symmetry of covariance to simplify the code
    s <- sqrt(sum(w*C)) # triangular kernel
  }
}

```

```

}

return(c(tau, KL[tau]/s))
}

# Tests:
# For 0 structural breaks with bootstrap (returns p-value)
# Null hypothesis: 0 breaks, alternative: 1 or more breaks
test_for_no_breaks <- function(Y, B=100){
  #Step 1: take data and estimate coefficients for GARCH(1,1)
  params0 <- get_params(Y)
  # We are getting omega, alpha1, beta1

  #Step 2: build innovation series
  innov <- Y/get_sigmas(Y)

  #Step 3: calculate residual-based CUSUM test statistic
  T_start <- BB_stat(innov)[2]

  #Step 4: use bootstrap
  T_boot <- numeric(0)
  for (i in 1:B){
    #Step 4.1: build bootstrapped observations recursively
    innov_boot <- sample(innov, replace = TRUE) # bootstrapping innovation series
    x_new <- Obs_boot(innov, innov_boot, params0)

    #Step 4.2: calculate residual-based bootstrapped T_n statistic
    innov_boot <- x_new/get_sigmas(x_new)
    T_boot <- c(T_boot, BB_stat(innov_boot)[2])
  }

  #Step 5: calculate p-value for the test
  p_val <- mean(T_boot>T_start)
  return(c(p_val, T_start, T_boot)) # returns p-value and bootstrapped statistic
}

# For presence of 1 structural break with bootstrap (returns p-value)
# Null hypothesis: exactly 1 structural break, alternative: 2 or more breaks
test_for_one_str_break <- function(Y, B=100){
  #Step 1: get the estimated change point and estimate coefficients for two GARCH(1,1) models
  K <- k_hat(Y)

```

```

# get estimated change point
Y1 <- Y[1:K]
Y2 <- Y[(K+1):length(Y)]
params1 <- get_params(Y1)
params2 <- get_params(Y2)
# get omega, alpha1, beta1 for both models

#Step 2: build innovation series for both models
innov1 <- Y1/get_sigmas(Y1)
innov2 <- Y2/get_sigmas(Y2)

#Step 3: calculate residual-based CUSUM test statistic
T_st_1 <- BB_stat(innov1)[2]
T_st_2 <- BB_stat(innov2)[2]
M_start <- max(T_st_1, T_st_2)

#Step 4: use bootstrap
M_boot <- numeric(0)
for (i in 1:B){
  #Step 4.1: build bootstrapped observations recursively
  innov_boot_1 <- sample(innov1, replace = TRUE)
  innov_boot_2 <- sample(innov2, replace = TRUE)
  # bootstrapping innovation series
  x_new_1 <- Obs_boot(innov1, innov_boot_1, params1)
  x_new_2 <- Obs_boot(innov2, innov_boot_2, params2)

  #Step 4.2: calculate bootstrapped M_n statistic for x_new
  innov_boot_1 <- x_new_1/get_sigmas(x_new_1)
  innov_boot_2 <- x_new_2/get_sigmas(x_new_2)
  T_boot_1 <- BB_stat(innov_boot_1)[2]
  T_boot_2 <- BB_stat(innov_boot_2)[2]
  M_boot <- c(M_boot, max(T_boot_1, T_boot_2))
}

#Step 5: calculate p-value
p_val <- mean(M_boot>M_start)
return(c(p_val, M_start, K, M_boot))
}

# Iterated algorithm, which uses two previous tests

```

```

ICSS <- function(Y, B=100, p = 0.05) {
  T <- length(Y) # number of observations in our sample
  t1 <- 1
  t2 <- T # starting endpoints
  tfirst <- 1
  tlast <- T
  tau <- c(1, T+1) # for all possible points of structural break

  while (tfirst < tlast - 1) {
    if (test_for_no_breaks(Y[t1:t2], B)[1] > p){
      tau <- sort(unique(tau))
      return(tau) # stop the procedure, if there are no breaks at all
    }
    if (test_for_one_str_break(Y[t1:t2], B)[1] > p){
      tcus <- k_hat(Y[t1:t2]) + t1-1 # taking into account that k_hat returns relative position
      tau <- c(tau, tcus)
      tau <- sort(unique(tau))
      return(tau) # stop the procedure, if there is only 1 break on this interval
    }
    #if (BB_test(Y[t1:t2], p)[1] == 'No break'){
    #print('t2:')
    #print(t2)
    # tau <- sort(unique(tau))
    # return(tau) # stop the procedure, if there are no more breaks
    #}
    else {
      tcus <- k_hat(Y[t1:t2]) + t1-1 # taking into account that k_hat returns relative position
      tau <- c(tau, tcus)
      t2 <- tcus-1 # looking at left interval

      repeat {
        if (test_for_no_breaks(Y[t1:t2], B)[1] > p){
          tau <- c(tau, tcus)
          break # go further, if there are no breaks at all on this interval
        }
        if (test_for_one_str_break(Y[t1:t2], B)[1] > p){
          tcus <- k_hat(Y[t1:t2]) + t1-1 # taking into account that k_hat returns relative position
          tau <- c(tau, tcus)
          break # go further, if there are no more breaks to the left
        }
      }
    }
  }
}

```

```

else {
  tcus <- k_hat(Y[t1:t2]) + t1-1 # remember for the case if there are no breaks to the left
  t2 <- k_hat(Y[t1:t2])-1 + t1-1
}
}

tfirst <- tcus
t1 <- tfirst
t2 <- T # new endpoints for next part of an algorithm

repeat {
  if (test_for_no_breaks(Y[t1:t2], B)[1] > p){
    tau <- c(tau, tcus)
    break # go further, if there are no breaks at all on this interval
  }
  if (test_for_one_str_break(Y[t1:t2], B)[1] > p){
    tcus <- k_hat(Y[t1:t2]) + t1-1 # taking into account that k_hat returns relative position
    tau <- c(tau, tcus)
    break # go further, if there are no more breaks to the right
  }
  else {
    tcus <- k_hat(Y[t1:t2]) + t1-1 # remember for the case if there are no breaks to the right
    t1 <- k_hat(Y[t1:t2]) + t1-1
  }
}

tlast <- t1
t1 <- tfirst
t2 <- tlast-1 # new endpoints for next iteration of an algorithm
}
}
tau <- sort(unique(tau))
return(tau) # stop the procedure, if the remaining interval is too small
}

# Algorithm, refining the moments of structural breaks given by ICSS procedure
ICSS_refinement <- function(Y, B = 100, p = 0.05) {
  tau <- ICSS(Y, B, p) # getting series of possible structural breaks
  if ((length(tau) <= 2)) { # in case if there are no structural breaks at all
    return(NULL)
  }
}

```

```

#print(tau) # uncomment if we want to see all steps
tau_ref <- tau[2:(length(tau)-1)] # series of possible structural breaks without 0 and T
iteration <- 1
while (iteration <= 5) { # setting limit to number of iterations; though, it coincides quickly
  tau <- c(1, tau_ref, length(Y)+1)
  tau_ref <- c()
  K <- length(tau) # remember total number of structural breaks

  for (n in 2:(K-1)){
    # check if the moment is really a structural break
    # for interval between two adjacent potential breaks
    tprev <- tau[n-1]
    tnext <- tau[n+1]-1
    #print(c(tprev, tnext)) # uncomment if we want to see all steps
    #print(iteration) # uncomment if we want to see all steps
    if (test_for_no_breaks(Y[tprev:tnext], B)[1] < p){
      tau_ref <- c(tau_ref, tau[n-1]-1 + k_hat(Y[tprev:tnext]))
    }
  }
  #print(tau) # uncomment if we want to see all steps
  #print(tau_ref) # uncomment if we want to see all steps
  iteration <- iteration + 1

  if (is.null(tau_ref)) # end refinement process if there are no breaks after all
    break
  if (length(tau_ref) == length(tau[2:(K-1)])){
    if (max(abs(tau_ref-tau[2:(K-1)]))<5)
      break # end refinement process if there is no change in tau or if change is very small
  }
}
return(sort(unique(tau_ref)))
}

# Functions for GARCH simulations
GARCH_0breaks <- function(N = 1000, w = 0.1, b = 0.7, g = 0.2){
  ksi <- rnorm(N-1, mean = 0, sd = 1)
  sigma <- c(0)
  Y <- c(rnorm(1, mean = 0, sd = 1))

  for (i in 2:N){

```

```

    sigma[i] <- sqrt(w + b*sigma[i-1]^2 + g*Y[i-1]^2)
    Y[i] <- ksi[i-1]*sigma[i]
  }
  return(Y) # generates GARCH(1,1) with given parameters and length N
}

GARCH_1break <- function(N = 2000, w = 0.1, b = 0.7, g = 0.2, dw = 0.2, db = 0, dg = 0) {

  ksi <- rnorm(N-1, mean = 0, sd = 1)
  sigma <- c(0)
  Y <- c(rnorm(1, mean = 0, sd = 1))

  for (i in 2:(N/2)){
    sigma[i] <- sqrt(w + b*sigma[i-1]^2 + g*Y[i-1]^2)
    Y[i] <- ksi[i-1]*sigma[i]
  }

  w <- w + dw # external shock w_2 = w_1 + dw (0.2 by default)
  b <- b + db # external shock b_2 = b_1 + db (0 by default)
  g <- g + dg # external shock g_2 = g_1 + dg (0 by default)

  for (i in (N/2+1):N){
    sigma[i] <- sqrt(w + b*sigma[i-1]^2 + g*Y[i-1]^2)
    Y[i] <- ksi[i-1]*sigma[i]
  }
  return(Y) # generates two GARCH(1,1) with given parameters and length N in total
}

GARCH_2breaks <- function(w = 0.1, b = 0.7, g = 0.2, dw1 = 0.2, db1 = 0, dg1 = 0, dw2 = 0.1, db2 = -0.1, dg2 = 0)
{

  ksi <- rnorm(2999, mean = 0, sd = 1)
  sigma <- c(0)
  Y <- c(rnorm(1, mean = 0, sd = 1))

  for (i in 2:1000){
    sigma[i] <- sqrt(w + b*sigma[i-1]^2 + g*Y[i-1]^2)
    Y[i] <- ksi[i-1]*sigma[i]
  }

  w <- w + dw1 # external shock w_2 = w_1 + dw1 (0.2 by default)
  b <- b + db1 # external shock b_2 = b_1 + db1 (0 by default)

```

```

g <- g + dg1 # external shock  $g_2 = g_1 + dg1$  (0 by default)

for (i in 1001:2000){
  sigma[i] <- sqrt(w + b*sigma[i-1]^2 + g*Y[i-1]^2)
  Y[i] <- ksi[i-1]*sigma[i]
}

w <- w + dw2 # external shock  $w_3 = w_2 + dw2$  (0.1 by default)
b <- b + db2 # external shock  $b_3 = b_2 + db2$  (-0.1 by default)
g <- g + dg2 # external shock  $g_3 = g_2 + dg2$  (0 by default)

for (i in 2001:3000){
  sigma[i] <- sqrt(w + b*sigma[i-1]^2 + g*Y[i-1]^2)
  Y[i] <- ksi[i-1]*sigma[i]
}
return(Y)
}
}

# Significance level
p <- 0.01 # the default one for this work

# Importing data:
{
  # Data from January 2010 to April 2020 (~2500 obs.)
  # technology
  {
    msft <- take_data("MSFT.xlsx", TRUE)
    msft_dates <- dates("MSFT.xlsx")
    msft_log_yields <- diff(log(msft))

    aapl <- take_data("AAPL.xlsx", TRUE)
    aapl_dates <- dates("AAPL.xlsx")
    moments <- detect_split(aapl)
    aapl <- c(aapl[1:moments[1]], 7*aapl[(moments[1]+1):length(aapl)])
    plot(aapl) # checking for result
    aapl_log_yields <- diff(log(aapl))

    ibm <- take_data("IBM.xlsx", TRUE)
    ibm_dates <- dates("IBM.xlsx")

```



```

ibm_log_yields <- diff(log(ibm))

intc <- take_data("INTC.xlsx", TRUE)
intc_dates <- dates("INTC.xlsx")
intc_log_yields <- diff(log(intc))
}

# entertainment
{
dis <- take_data("DIS.xlsx", TRUE)
dis_dates <- dates("DIS.xlsx")
dis_log_yields <- diff(log(dis))
}

# financial
{
jpm <- take_data("JPM.xlsx", TRUE)
jpm_dates <- dates("JPM.xlsx")
jpm_log_yields <- diff(log(jpm))
}

# healthcare
{
jnj <- take_data("JNJ.xlsx", TRUE)
jnj_dates <- dates("JNJ.xlsx")
jnj_log_yields <- diff(log(jnj)) # get logarithmic yields
}

# retail
{
wmt <- take_data("WMT.xlsx", TRUE)
wmt_dates <- dates("WMT.xlsx")
wmt_log_yields <- diff(log(wmt))
}

# energy
{
xom <- take_data("XOM.xlsx", TRUE)
xom_dates <- dates("XOM.xlsx")
xom_log_yields <- diff(log(xom))
}

```

```

cvx <- take_data("CVX.xlsx", TRUE)
cvx_dates <- dates("CVX.xlsx")
cvx_log_yields <- diff(log(cvx))
}

# aerospace & defence
{
ba <- take_data("BA.xlsx", TRUE)
ba_dates <- dates("BA.xlsx")
ba_log_yields <- diff(log(ba))
}

# consumers - food and beverages
{
ko <- take_data("KO.xlsx", TRUE)
ko_dates <- dates("KO.xlsx")
moments <- detect_split(ko)
ko <- c(ko[1:moments[1]], 2*ko[(moments[1]+1):length(ko)])
plot(ko) # checking for result
ko_log_yields <- diff(log(ko))

mcd <- take_data("MCD.xlsx", TRUE)
mcd_dates <- dates("MCD.xlsx")
mcd_log_yields <- diff(log(mcd))
}

#####

# Data from April 2019 to April 2020 (~ 250 obs.)
# communication services and entertainment
{
googl <- take_data("GOOGL.xlsx", TRUE)
googl_dates <- dates("GOOGL.xlsx")
googl_log_yields <- diff(log(googl))

fb <- take_data("FB.xlsx", TRUE)
fb_dates <- dates("FB.xlsx")
fb_log_yields <- diff(log(fb))

```

```

nflx <- take_data("NFLX.xlsx", TRUE)
nflx_dates <- dates("NFLX.xlsx")
nflx_log_yields <- diff(log(nflx))
}

#retail
{
amzn <- take_data("AMZN.xlsx", TRUE)
amzn_dates <- dates("AMZN.xlsx")
amzn_log_yields <- diff(log(amzn))
}

# consumers - food and beverages
{
pep <- take_data("PEP.xlsx", TRUE)
pep_dates <- dates("PEP.xlsx")
pep_log_yields <- diff(log(pep))

sbux <- take_data("SBUX.xlsx", TRUE)
sbux_dates <- dates("SBUX.xlsx")
sbux_log_yields <- diff(log(sbux))
}

#####

# Extra:
# Zoom (startup)
zm <- take_data("ZM.xlsx", TRUE)
zm_dates <- dates("ZM.xlsx")
zm_log_yields <- diff(log(zm))

#####
}

# Results:

#estimate common GARCH(1,1) parameters
mean(get_params(aapl_log_yields)[2], get_params(ba_log_yields)[2],
get_params(cvx_log_yields)[2], get_params(dis_log_yields)[2],
get_params(ibm_log_yields)[2], get_params(intc_log_yields)[2],

```

```

get_params(jnj_log_yields)[2], get_params(jpm_log_yields)[2],
get_params(ko_log_yields)[2], get_params(mcd_log_yields)[2],
get_params(msft_log_yields)[2], get_params(wmt_log_yields)[2],
get_params(xom_log_yields)[2])

mean(get_params(aapl_log_yields)[3], get_params(ba_log_yields)[3],
      get_params(cvx_log_yields)[3], get_params(dis_log_yields)[3],
      get_params(ibm_log_yields)[3], get_params(intc_log_yields)[3],
      get_params(jnj_log_yields)[3], get_params(jpm_log_yields)[3],
      get_params(ko_log_yields)[3], get_params(mcd_log_yields)[3],
      get_params(msft_log_yields)[3], get_params(wmt_log_yields)[3],
      get_params(xom_log_yields)[3])

# healthcare
print("Johnson&Johnson: ")
plot(KL(jnj_log_yields))
print(ICSS_refinement(jnj_log_yields, p)) # one break
print(jnj_dates[2010]) # signals that volatility shift is 2018/01/19
plot(get_sigmas(jnj_log_yields), xlab = "Observations", ylab = "IBM: Conditional variances before detecting breaks")
renewed <- c(get_sigmas(jnj_log_yields[1:2009]), get_sigmas(jnj_log_yields[2010:length(jnj_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "IBM: Conditional variances after detecting breaks")
# least affected by coronacrisis, as healthcare organisation

print("Disney: ")
print(ICSS_refinement(dis_log_yields, p)) # no breaks
print("Walmart: ")
print(ICSS_refinement(wmt_log_yields, p)) # one break
print(wmt_dates[1413]) # signals that volatility shift was long ago, 2015/08/14

# technology sector
print("Microsoft: ")
print(ICSS_refinement(msft_log_yields, p)) # no breaks
print("Apple: ")
print(ICSS_refinement(aapl_log_yields, p)) # no breaks
print("IBM: ")
print(ICSS_refinement(ibm_log_yields, p)) # one break
print(ibm_dates[2191]) # signals that volatility shift is 2018/10/08
print("Intel: ")
print(ICSS_refinement(intc_log_yields, p)) # one break

```

```

print(intc_dates[2012]) # signals that volatility shift is 2018/01/24
plot(get_sigmas(ibm_log_yields), xlab = "Observations", ylab = "IBM: Conditional variances before detecting breaks")
renewed <- c(get_sigmas(ibm_log_yields[1:2190]), get_sigmas(ibm_log_yields[2191:length(ibm_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "IBM: Conditional variances after detecting breaks")
plot(get_sigmas(intc_log_yields), xlab = "Observations", ylab = "Intel: Conditional variances before detecting breaks")
renewed <- c(get_sigmas(intc_log_yields[1:2011]), get_sigmas(intc_log_yields[2012:length(intc_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "Intel: Conditional variances after detecting breaks")

# financial sector
print("JPMorgan: ")
print(ICSS_refinement(jpm_log_yields, p)) # one break
print(jpm_dates[2533]) # signals that volatility shift is 2020/02/20
plot(get_sigmas(jpm_log_yields), xlab = "Observations", ylab = "JPMorgan: Conditional variances before detecting breaks")
renewed <- c(get_sigmas(jpm_log_yields[1:2532]), get_sigmas(jpm_log_yields[2533:length(jpm_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "JPMorgan: Conditional variances after detecting breaks")

# energy sector
print("ExxonMobil: ")
print(ICSS_refinement(xom_log_yields, p)) # no breaks
print("Chevron: ")
print(ICSS_refinement(cvx_log_yields, p)) # one break
print(cvx_dates[2539]) # signals that volatility shift is 2020/02/20
plot(get_sigmas(cvx_log_yields), xlab = "Observations", ylab = "Chevron: Conditional variances before detecting breaks")
renewed <- c(get_sigmas(cvx_log_yields[1:2538]), get_sigmas(cvx_log_yields[2539:length(cvx_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "Chevron: Conditional variances after detecting breaks")

# aerospace & defence
print("Boeing: ")
print(ICSS_refinement(ba_log_yields, p)) # one or three breaks
# 2533th observation is definitely a structural break
print(ba_dates[2533]) # signals that volatility shift is 2020/02/20
plot(get_sigmas(ba_log_yields), xlab = "Observations", ylab = "Boeing: Conditional variances before detecting breaks")
renewed <- c(get_sigmas(ba_log_yields[1:2532]), get_sigmas(ba_log_yields[2533:length(ba_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "Boeing: Conditional variances after detecting breaks")

```

```

# consumers: food and beverages
print("Coca-cola: ")
print(ICSS_refinement(ko_log_yields, p)) # one break
print(ko_dates[2219]) # signals that volatility shift is 2018/11/16
plot(get_sigmas(ko_log_yields), xlab = "Observations", ylab = "Coca-cola: Conditional variances before detecting
breaks")
renewed <- c(get_sigmas(ko_log_yields[1:2218]), get_sigmas(ko_log_yields[2219:length(ko_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "Coca-cola: Conditional variances after detecting breaks")
print("McDonalds: ")
print(ICSS_refinement(mcd_log_yields, p)) # one break
print(mcd_dates[2533]) # signals that volatility shift is 2020/02/21
plot(get_sigmas(mcd_log_yields), xlab = "Observations", ylab = "McDonalds: Conditional variances before detecting
breaks")
renewed <- c(get_sigmas(mcd_log_yields[1:2532]), get_sigmas(mcd_log_yields[2533:length(mcd_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "McDonalds: Conditional variances after detecting breaks")

#For only last year
print("Google: ")
print(ICSS_refinement(googl_log_yields, p)) # no breaks
print("Netflix: ")
print(ICSS_refinement(nflx_log_yields, p)) # no breaks
print("Starbucks: ")
print(ICSS_refinement(sbox_log_yields, p)) # no breaks
print("Zoom: ")
print(ICSS_refinement(zm_log_yields, p)) # no breaks

print("Amazon: ")
print(ICSS_refinement(amzn_log_yields, p)) # one break
print(amzn_dates[193]) # signals that volatility shift is 2020/01/29
plot(get_sigmas(amzn_log_yields), xlab = "Observations", ylab = "Amazon: Conditional variances before detecting
breaks")
renewed <- c(get_sigmas(amzn_log_yields[1:192]), get_sigmas(amzn_log_yields[193:length(amzn_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "Amazon: Conditional variances after detecting breaks")

print("Facebook: ")
print(ICSS_refinement(fb_log_yields, p)) # two breaks
print(fb_dates[37]) # signals about volatility shift at 2019/06/14
print(fb_dates[208]) # signals that volatility shift is 2020/02/20
plot(get_sigmas(fb_log_yields), xlab = "Observations", ylab = "Facebook: Conditional variances before detecting
breaks")

```

```

renewed <- c(get_sigmas(fb_log_yields[1:36]), get_sigmas(fb_log_yields[37:207]),
get_sigmas(fb_log_yields[208:length(fb_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "Facebook: Conditional variances after detecting breaks")
print(get_params(fb_log_yields[1:36])) # looks strange, but data is fine

print("Pepsi: ")
print(ICSS_refinement(pep_log_yields, p)) # three breaks most likely
print(pep_dates[59]) # signals about volatility shift at 2019/07/17
print(pep_dates[115]) # signals about volatility shift at 2019/10/04
print(pep_dates[211]) # signals that volatility shift is 2020/02/25
plot(get_sigmas(pep_log_yields), xlab = "Observations", ylab = "Pepsi: Conditional variances before detecting
breaks")
renewed <- c(get_sigmas(pep_log_yields[1:58]), get_sigmas(pep_log_yields[59:114]),
get_sigmas(pep_log_yields[115:210]), get_sigmas(pep_log_yields[211:length(pep_log_yields)]))
plot(renewed, xlab = "Observations", ylab = "Pepsi: Conditional variances after detecting breaks")
print(get_params(pep_log_yields[1:58]))
print(get_params(pep_log_yields[115:210]))

# Conclusion:

# plenty of observations show presence of structural break at February 20th
# (financial (JPM), energy (CVX), aerospace (BA), consumer-food (McD))
# on 0.01 significance level.

# Moreover, this remains even if sample size is small (~ 260 obs. for this one year):
# Amazon and Facebook shares show the presence of this structural break

```