



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 004.81

КОГНИТИВНЫЙ ФРАКТАЛЬНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ: R/S АНАЛИЗ И V-СТАТИСТИКА

Романко О.Р.

*Национальный технический университет Украины «Киевский политехнический институт»
г. Киев, Украина*

romanko1993@gmail.com

Фрактальный анализ временных рядов – новая методология, использующая инструментарий фрактальной математики для анализа и прогнозирования финансовых и экономических процессов. Представлены описания и алгоритмы нового подхода к анализу временных рядов, позволяющие привнести некий семантический смысл в исследуемые данные – методология R/S анализа, позволяющая идентифицировать процессы с длинной и короткой памятью, и V-статистика, для нахождения длины неперiodических циклов.

Ключевые слова: когнитивный фрактальный анализ; R/S анализ; V-статистика.

Введение

Современный подход к ведению бизнеса в любой сфере человеческой деятельности состоит, прежде всего, в возможности предугадать или предсказать ближайшее развитие событий. Для этого, в первую очередь, нужно «понять» данные, провести так называемый технический и фундаментальный анализ процесса. Последний подразумевает под собой соединение всех известных знаний о нем в единую взаимосвязанную сеть, соединенную причинно-следственными связями. Большинство стараний исследователя направлено на получение более семантически значимого представления о процессе или явлении.

В данной работе изложены следующие модели и инструменты анализа, позволяющие, по мнению автора, получить достаточно значимые знания относительно исследуемых процессов, являющиеся определенным маячком при дальнейшем их исследовании:

- Метод R/S – анализа, который позволяет исследователю идентифицировать процесс с длинной памятью и классифицировать по его особенностям (персистентный или антиперсистентный).
- V-статистика, позволяющая исследовать среднюю длину неперiodических циклов, возникающих в процессах вследствие сезонности определенных явлений.
- Альтернативный метод оценки памяти – с помощью модели дробного броуновского движение

Данная методология позволяет дать количественную оценку такого явления, как память процесса. Этой оценкой является параметр Хёрста.

Параметр Хёрста $H \in (0,1)$ – показывает характер изменения процесса. При $H \in (0.5,1]$ процесс является персистентным, он сохраняет имеющуюся тенденцию, то есть возрастание в прошлом более вероятно приводит к возрастанию в дальнейшем, и наоборот. Биржевые маклеры описывают такое поведение возникновением «трендов» и «пробоев». При $H \in [0,0.5)$ процесс называется антиперсистентным, любая тенденция стремится смениться противоположной, формируется так называемый «коридор», в котором процесс осуществляет колебания. При $H = 0.5$ процесс является полностью случайным, явной тенденции не выражено. Стоит отметить, что природные процессы имеют параметр Хёрста $H = 0.72 - 0.73$

1.Предпосылки возникновения нового фрактального подхода к анализу временных рядов

Появление нового фрактального подхода, связано с обнаружением некоторых ошибочных представлений о природе и поведении многих финансовых процессов.

Во-первых, это широко распространенное применение основной теоремы статистики – центральной предельной теоремы. Эта теорема гласит о том, что сумма большого числа независимых одинаково распределенных величин

имеет нормальное (гауссово) распределение. Но почти всегда временные показатели финансового процесса не являются независимыми. Поэтому центральная теорема может дать сбой.

Явления «тяжелых» хвостов и высоких пиков являются весьма распространенными, для объяснения «тяжелых» хвостов были разработаны отдельные вероятностные распределения (Коши, Парето, Леви).

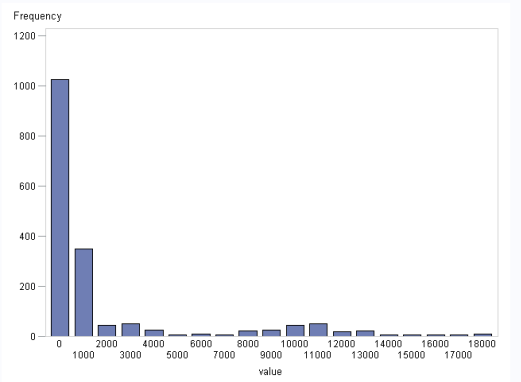


Рисунок 1 -- Пример данных с «тяжелыми» хвостами – 20-тидневный индекс Доу–Джонса

Для проведения качественного исследования, необходимо понять природу процесса, по возможности провести анализ политических, экономических, социальных составляющих, которые повлияли на поведение исследуемого процесса [Schwager,2012].

2.Применение фрактального анализа – анализ финансовых данных предприятия

Для наглядного примера возьмем данные ежедневных финансовых потоков предприятия «А» за 2014 год и проведем фрактальный анализ (данные и программу на языке SAS можно найти на сайте https://sites.google.com/site/sas4ua2015/romanko_alex)

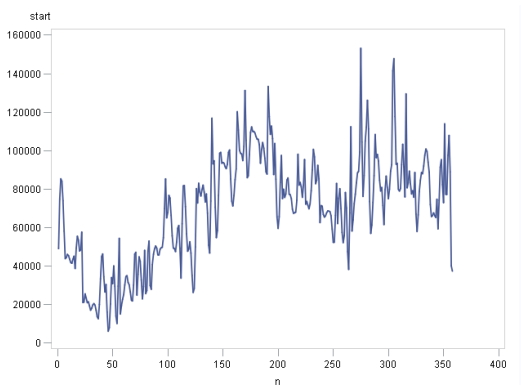


Рисунок 2 -- Данные о входящих денежных потоках (income) предприятия за 358 дней

3.Метод поиска параметра Хёрста. R/S анализ и V-статистика

Как упоминалось выше, некоторые временные

ряды недостаточно хорошо описываются стандартной гауссовой статистикой, поскольку системы изначально не являются независимыми и идентично распределенными (independent and identically-distributed, IID).

Х. Е. Хёрстом была открыта непараметрическая методология, позволяющая изучать системы без предварительных предположений о форме распределения вероятностей. Хёрст включил в свое исследование многие естественные системы и дал нам новую статистическую методологию для различения случайных и неслучайных систем, постоянства трендов и продолжительности циклов если таковые имеются [Mandelbrot and others,2008]. Идея метода заключается в том, что рассчитывается так называемый «нормированный размах», который показывает расстояние S , на которое перемещается система за дискретное время n . Известно, что случайное блуждание (например случайное броуновское движение молекул) за время n проходит расстояние $S = c * \sqrt{n}$ (где c – произвольная константа). Хёрст обобщил это уравнение для систем, не являющихся «IID»: $S = c * n^H$, где $H \in [0,1]$. Следовательно, мы можем выполнить простую линейную регрессию на $\log(n)$

как независимой переменной и $\log(r/s)n$ как зависимой переменной. Наклон уравнения является оценкой показателя Хёрста H .

Алгоритм R/S анализа следующий:

1. Начните с временного ряда длины M . Преобразуйте его во временной ряд длины $N = M - 1$ из логарифмических отношений

$$N_i = \log\left(\frac{M_{i+1}}{M_i}\right), i = 1, 2, 3, \dots, (M - 1). \quad (1)$$

Или используя $AR(1)$ разности - построив регрессию M_{i+1} как зависимой переменной против M_i как независимой переменной.

$$N_i = M_{i+1} - (c + a \cdot M_i), i = 1, 2, 3, \dots, (M - 1). \quad (2)$$

2. Разделите этот период времени на A смежных подпериодов длины n , так что $A * n = N$. Пометьте каждый подпериод I_a , с учетом того, что $a=1, 2, 3, \dots, A$. Каждый элемент в I_a помечен N_k при этом $k = 1, 2, 3, \dots, n$. Для каждого I_a длины n среднее значение определяется как:

$$E_a = (1/n) \sum_{k=1}^n N_{k,a} \quad (3)$$

где E_a - среднее значение N_k содержащегося в подпериоде I_a длины n .

3. Временной ряд накопленных отклонений($X_{k,a}$) от среднего значения для каждого подпериода I_a определяется как:

$$X_{k,a} = \sum_{i=1}^k (N_{i,a} - e_a), k = 1, 2, \dots, n. \quad (4)$$

4. Диапазон определяется как максимальное значение за вычетом минимального значения $X_{k,a}$ в пределах каждого подпериода I_a :

$$R_{k,s} = \max(X_{k,a}) - \min(X_{k,a}), k = 1, 2, 3, \dots, n. \quad (5)$$

5. Выборочное стандартное отклонение, рассчитываемое для каждого подпериода I_a :

$$S_{ia} = \left(\frac{1}{n} \cdot \sum_{k=1}^n (N_{k,a} - e_a^2)^{0.5} \right). \quad (6)$$

6. Каждый диапазон R_{ia} теперь нормализуется путем деления на соответствующий S_{ia} . Поэтому повторно нормированный размах в течении каждого I_a подпериода равен R_{ia} / S_{ia} . В шаге 2 выше мы получили смежные подпериоды длины n . Следовательно, среднее значение R/S для длины n определяется как:

$$(R/S)_n = \frac{1}{A} \cdot \sum_{a=i}^a (R_i / S_{i,a}). \quad (7)$$

7. Длина n увеличивается до следующего более высокого значения, а $(M-I)/n$ является целочисленным значением. Мы используем значения n , включающие начальные и конечные точки временного ряда, и шаги 1-6 повторяются до $n = (M - 1)/2$. Теперь мы можем выполнить простую регрессию методом наименьших квадратов на $\log(n)$ как независимой переменной и $\log(r/s)_n$ как зависимой переменной. Отрезок, отсекаемый на координатной оси, является оценкой $\log(c)$, константой. Наклон уравнения является оценкой показателя Хёрста H .

Ожидаемое значение нормированного размаха с поправкой Энниса и Ллойда [Peters, 1994]:

$$E(R/S_n) = \frac{n - 0.5}{n} \left(n \cdot \frac{\pi}{2} \right)^{-0.5} \sum_{r=1}^{n-1} \sqrt{\frac{n-r}{r}}. \quad (8)$$

Ожидаемое значение соответствует размаху полностью случайного независимого процесса (то есть с параметром Хёрста, равным 0.5).

Variable	DF	Estimate
Intercept	1	0.1947
Hurst_par	1	0.3285

Рисунок 3 -- Параметр Хёрста, определенный с помощью авторегрессии-0.3285

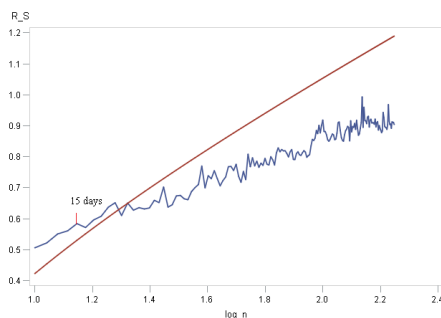


Рисунок 4 -- График результатов R/S – анализа

Горизонтальной оси графика соответствует $\lg(n)$, вертикальной - $\lg(R/S_n)$. На графике видно что реальные значения размаха (синяя ломанная) лежат ниже ожидаемых значений (красная линия). Это значит, наш процесс имеет длинную память (поскольку его размах намного отклоняется от ожидаемого) и антиперсистентен (поскольку синяя ломанная лежит ниже красной линии).

Теперь воспользуемся V-статистикой. Как было сказано ранее, она позволяет находить среднюю длину неперiodических циклов [Peters, 1994]. Простейшим примером естественных неперiodических циклов являются смены периодов года (ведь мы не знаем наверняка, сколько дней будет длиться осень, а сколько – зима, но знаем среднюю продолжительность времен года).

Общая идея V-статистики – когда цикл какого-либо процесса начинает закончился, нормированный размах (R/S_n) перестает расти (поскольку значения опять начинают повторяться, следовательно не вносят изменений в диапазон значений). Если вычертить R/S_n против $\lg(n)$, то при окончании цикла, график начнет стремительно падать вниз.

Конечная формула: $V_n = \frac{R/S_n}{\sqrt{n}}$. Это отношение приведет к горизонтальной линии, если процесс случаен. Если он персистентен, линия на графике будет наклонена вверх. График будет иметь наклон вниз, если процесс антиперсистентен.

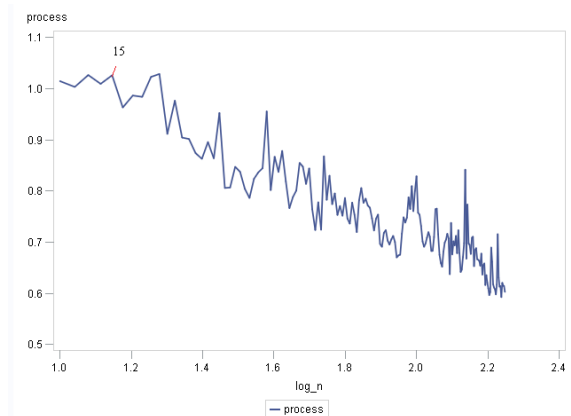


Рисунок 5 График V-статистики процесса

Как мы видим на рис 7, первый «разрыв» на графике (резкое его снижение), соответствует $n=15$. Действительно, если посмотреть на график приростов y_k , можно увидеть похожие друг на друга паттерны, длина которых составляет от 13 до 17 дней.

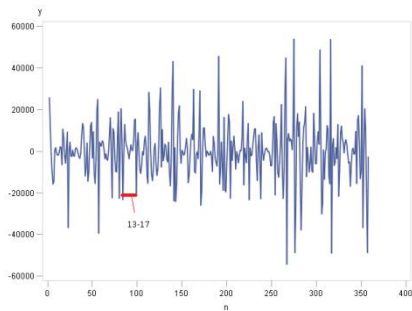


Рисунок 6 -- График приростов y_k

Заключение

В данной работе описан принципиально новый подход к анализу временных рядов, для которых корреляционная функция постепенно уменьшается со временем («длинная память»).

Использование новых инструментов фрактального анализа, таких как R/S – анализ и V-статистика дает более глубокое понимание природы временного ряда и его особенностей, возможность исследовать периодические и непериодические циклы. При развитии и дополнении фрактальных инструментов, они открывают новые возможности анализа и прогнозирования временных рядов различной природы: всевозможных финансовых индексов, финансовых потоков, различных природных явлений.

Библиографический список

[Mandelbrot and others, 2008] Mandelbrot, Benoit B., and Richard L. Hudson. Misbehaviour of Markets. Profile Books, 2008.

[Peters, 1994] Peters, Edgar E. Fractal market analysis: applying chaos theory to investment and economics. Vol.

[Schwager, 2012] Schwager, Jack D. Market wizards: Interviews with top traders. John Wiley & Sons, 2012.

COGNITIVE FRACTAL ANALYSIS OF TIME SERIES: R / S ANALYSIS AND V-STATISTICS

Romanko A.R.

National University of Ukraine «Kiev Politechnic Institute »

romanko1993@gmail.com

Fractal Time Series Analysis - a new methodology using modern vision of the financial market and fractal mathematics tools for analyzing financial and economic processes. There are descriptions and algorithms of a new approach to the analysis of time series - methodology of R/S analysis and V-statistic.

Introduction

The emergence of a new fractal approach is due to the discovery of some of the misconceptions about the nature and behavior of many financial processes.

Firstly, it is the widespread misuse of the fundamental theorem of statistics - the central limit theorem. But many processes are not independent and

identically distributed - so the central theorem can not be used.

Secondly, the use of fractal approach is due to the emergence of a new concept of the behavior of the global market, and financial markets in particular. Classical theory is that markets are "fair", where participants are guided only by the principle of maximizing profits therefore fluctuations in the market are mostly random, aimed at stabilizing processes.

Practical approach

The modern approach to business in any sphere of human activity is, above all, the ability to anticipate or predict the near future developments. The quality and reliability of doing business depends on the most accurate answer to the question "what will happen tomorrow?".

To answer this question correctly, we must do some fundamental and technical research, "understand" data. The goal is to aggregate all knowledge about some process into one logical whole

Following analysis tools, that help us to gain some considerable knowledge on how to use the information available, are outlined:

- The method of R/S analysis allows the researcher to identify a process with a long memory, and classify according to its characteristics (persistent or antipersistent) and find the Hurst exponent.
- V-statistic allows to investigate the average length of non-periodic cycles that occur in the process due to the seasonality of certain phenomena.

The Hurst exponent $H \in (0,1)$ – shows the nature of change in the process. When $H \in (0.5,1]$ the process is persistent, it maintains the existing tendency, i.e. the increase in the past are more likely to lead to an increase in the future, and vice versa. Stock brokers describe this behavior by the emergence of "trends" and "structural breaks."

When $H \in [0,0.5)$ the process is called antipersistent, any tendency tends to be replaced by its counterpart. If $H=0.5$ it is a completely random process, a clear trend is not expressed. It is worth noting that natural processes have $H = 0.72-0.73$

Conclusion

This paper describes a new approach to the analysis of time series, for which the correlation function decreases gradually with time ("long memory").

Using the new fractal analysis, such as R / S - analysis and V-statistics gives a better understanding of the nature of the time series and its features, the ability to investigate non-recurring and periodic cycles.

With the development and addition of fractal tools, they open up new possibilities of analysis and forecasting of time series of different nature: all kinds of financial indices, financial flows of various natural phenomena.