

АНАЛИЗ ВЗАИМОЗАВИСИМОСТИ ВРЕМЕННЫХ РЯДОВ БИТКОИНА И АКТИВНОСТИ СООБЩЕСТВ В СОЦИАЛЬНЫХ СЕТЯХ

Людмила Кириченко, Тамара Радивилова, Виталий Булах, Вадим Чакрян

Аннотация: В работе проведен сравнительный кросскорреляционный и мультифрактальный анализ временных рядов курса биткоина и рядов активностей сообществ в социальных сетях связанных с этой криптовалютой. Была обнаружена значительная корреляция между курсом Биткоина и активностью сообществ временных рядов, когда интерес к биткоину и системе блокчейн возрастал вместе с ростом курса биткоина. Фрактальный анализ временных рядов выявил наличие самоподобных и мультифрактальных свойств для рядов активностей сообществ. Результаты исследований показывают, что ряды, которые имеют сильную корреляционную зависимость, имеют близкие мультифрактальные свойства.

Ключевые слова: самоподобные и мультифрактальные временные ряды, обобщенный показатель Херста, кросскорреляция, биткоин, социальные сети

ACM Classification Keywords: G.3 Probability and statistics - Time series analysis, Stochastic processes, G.1 Numerical analysis, G.1.2 Approximation - Wavelets and fractals.

Введение

Биткоин является основной криптовалютой и производным финансовым инструментом, который пользуется большим спросом в последние годы. Биткоин - это децентрализованная электронная платежная система, созданная в 2008 году, рыночная капитализация которой неуклонно растет. Биткоин использует сложные криптографические методы для подписи транзакций и определения средств управления. Биткоин — это одноранговая система, пользователи которой проводят транзакции непосредственно без необходимости посредников, и в этих транзакциях записывается информация для последующей проверки всеми узлами в открытой распределенной книге называемой Блокчейн. [Nakamoto, 2008].

В настоящее время стало общепризнанным, что финансовые рынки обладают фрактальными свойствами [Peters, 2003]. В последние годы был проведен анализ цены биткоина на наличие фрактальных свойств, который выявил сильно выраженные фрактальные свойства (см, например [Meléndez, 2014]).

Анализ финансовых рынков при инвестиционных решениях полагается на традиционные источники информации, такие как раскрытие информации о компаниях, новости на рынке, отчеты, и так далее. Данные социальных сетей использовались для коммерческих целей для автоматического получения отзывов клиентов о товаре или бренде, поиска новых клиентов, анализа предпочтения и влияния на пользователей социальных сетей при голосовании, выборке, на рынке и т.д. За прошедшие несколько лет социальные сети быстро проникли в различные аспекты нашей жизни, и содержат огромный объем информации; данные социальных сетей представляют альтернативный источник информации для аналитиков и инвесторов финансовых рынков [Arnaboldi, 2015, Do, 2015, Bollen, 2011, Mittal, 2013, Matta, 2016]. Одно из направлений исследований данных в социальных сетях и финансовом секторе – это анализ взаимосвязи между информацией о биткойне и блокчейне, содержащейся в социальных сетях, и цене биткойна [Feng Mai, 2016, Peng Xie, 2017, Greaves, 2015].

Социальные сети и взаимодействия между участниками рынка могут существенно влиять на курс биткойна. В [Garcia, 2014] показано, что растущая популярность биткойна вызывает увеличение объема поиска, что, в свою очередь, приводит к более высокой активности в социальных сетях, связанной с биткойном. Покупка биткойнов пользователями приводит к большему интересу, что приводит к более высоким ценам, которые в конечном итоге будут возвращены в объем – образуется цикл социальной обратной связи.

Объектом данного исследования являются процессы взаимосвязи и взаимовлияния между динамикой финансовых рынков и социальных сетей. Субъектом исследования является корреляционная зависимость и фрактальная структура финансовых рядов и активности социальных групп. Целью работы является анализ взаимозависимости фрактальных характеристик временных рядов курса криптовалюты биткойн и активности сообществ в социальной сети фейсбук, связанных с биткойном.

Самоподобные и мультифрактальные временные ряды

Самоподобие случайных процессов заключается в сохранении вероятностных характеристик при изменении масштаба времени. Стохастический процесс $X(t)$ является самоподобным с параметром H , если процесс $a^{-H} X(at)$ описывается теми же законами конечномерных распределений, что и $X(t)$:

$$\text{Law}\{a^{-H} X(at)\} = \text{Law}\{X(t)\}, \quad \forall a > 0. \quad (1)$$

Параметр H , $0 < H < 1$, называемый показателем Херста, представляет собой степень самоподобия процесса. Наряду с этим свойством, показатель $H > 0.5$ характеризует меру долгосрочной зависимости стохастического процесса. Это означает, что если временной ряд в течение какого-то времени возростал (убывал), то с вероятностью, близкой к показателю Херста, ряд сохранит эту тенденцию в течение аналогичного промежутка времени [Feder, 1988]. Начальные моменты самоподобного случайного процесса можно выразить как

$$M\left[|X(t)|^q\right] = C(q) \cdot t^{qH}, \quad (2)$$

где величина $C(q) = M\left[|X(1)|^q\right]$.

В сравнении с самоподобными процессами (1-2) мультифрактальные процессы проявляют более сложное скейлинговое поведение:

$$\text{Law}\{X(at)\} = \text{Law}\{\mathcal{M}(a) \cdot X(t)\}, \quad a > 0, \quad (3)$$

где $\mathcal{M}(a)$ – случайная функция, независимая от $X(t)$. В случае самоподобия (монофрактальности) процесса $\mathcal{M}(a) = a^H$.

Для моментов мультифрактальных процессов выполняется отношение:

$$M\left[|X(t)|^q\right] = c(q) \cdot t^{qh(q)}, \quad (4)$$

где $c(q)$ – некоторая детерминированная функция; $h(q)$ – обобщенный показатель Херста, являющийся в общем случае нелинейной функций. Значение $h(q)$ при $q = 2$ совпадает со значением степени самоподобия H . Для монофрактальных процессов обобщенный показатель Херста не зависит от параметра q : $h(q) = H$. [Reidi, 2002].

Мультифрактальные объекты являются статистически неоднородными самоподобными объектами. Для мультифрактальных временных рядов статистическая неоднородность объекта выражается в неоднородности распределения данных ряда, т.е. наличии тяжелых хвостов функции плотности распределения вероятностей временного ряда.

Метод мультифрактального детрендрованного флуктационного анализа

Одним из наиболее популярных методов исследования мультифрактальной структуры временных рядов является метод мультифрактального детрендрованного флуктационного анализа (МФ ДФА) [Kantelhardt, 2002, 2008], который основан на обработке нестационарных временных рядов с трендами.

Согласно МФ ДФА для начального временного ряда $x(t)$ формируется кумулятивный временной ряд $y(t) = \sum_{i=1}^t x(i)$, который в дальнейшем делится на N непересекающихся сегментов длиной τ , и для таких сегментов $y(t)$ в дальнейшем рассчитывается флуктуационная функция:

$$F^2(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} (y(t) - Y_m(t))^2, \quad (5)$$

где $Y_m(t)$ - локальный m -полиномиальный тренд на выбранном сегменте. Усредненная по всему временному ряду $y(t)$ функция $F(\tau)$ имеет масштабирование на участке длиной τ :

$$F(\tau) \propto \tau^H \quad (6)$$

При изучении мультифрактальных свойств зависимость флуктуационной функции от параметра τ рассматривается следующим образом:

$$F_q(\tau) = \left\{ \frac{1}{N} \sum_{i=1}^N [F^2(\tau)]^{\frac{q}{2}} \right\}^{\frac{1}{q}} \quad (7)$$

Если исследуемая серия является мультифрактальной и имеет долгосрочную зависимость, флуктуационная функция $F_q(\tau)$ представлена степенным законом:

$$F_q(\tau) \propto \tau^{h(q)}, \quad (8)$$

где $h(q)$ – обобщенный показатель Херста.

Очевидно, что оценивание показателя Херста по экспериментальным данным играет важнейшую роль в изучении процессов, обладающими мультифрактальными свойствами. Однако, прежде чем проводить анализ мультифрактальных свойств, необходимо убедиться, что временной ряд действительно является самоподобным, что соответствует наличию линейного участка флуктуационной функции $F(\tau)$ в (6). При оценивании показателя Херста методом ДФА (5), необходимо сначала провести прикидочное оценивание, используя локальные полиномы увеличивающейся степени и определить наименьшую степень полинома, начиная с которой оценка показателя Херста перестает изменяться. После этого для оценивания самоподобия временного ряда необходимо удалять локальный полиномиальный тренд найденной степени.

Методика расчета медианной кросскорреляции

При анализе данных, в которых присутствуют существенные выбросы, что характерно для мультифрактальных временных рядов, возникает вопрос о степени влияния данных выбросов на стандартные статистические оценки. Появление в выборке нескольких выбросов может сильно ухудшить оценку математического ожидания, дисперсии или автокорреляционной функции, но очень слабо повлиять на значение медианы. Поэтому использование медианы вместо математического ожидания в формуле коэффициента корреляции приводит к тому, что он становится менее восприимчив к сильным выбросам и такая оценка является более не смещенной:

$$r_{median} = \frac{\text{Median}[(x - \text{Median}[x]) (y - \text{Median}[y])]}{\sqrt{\text{Median}[(x - \text{Median}[x])^2] \text{Median}[(y - \text{Median}[y])^2]}} \quad (9)$$

Проиллюстрируем применение данного подхода на примере оценивания корреляционной функции процесса авторегрессии. На рис. 1 вверху слева представлен ряд авторегрессии первого порядка с коэффициентом авторегрессии $\phi = 0.9$, для которого, как известно, теоретическая нормированная автокорреляционная функция $k(\tau) = \phi^\tau$. Выборочная автокорреляционная функция, полученная с помощью среднего значения, представлена под временным рядом, а выборочная корреляционная функция, полученная с помощью медианного оценивания, показана внизу слева. Вверху справа изображен тот же авторегрессионный временной ряд, но с добавлением сильных выбросов в случайные моменты времени, а под ним соответствующие выборочные автокорреляционные функции. Очевидно, что оценка автокорреляционной функции, полученная медианным способом практически не изменилась, в то время как обычная выборочная корреляционная функция для ряда с выбросами не соответствует первоначальному ряду.

Результаты исследований

Для проведения исследований в сети Facebook было выбрано три больших сообщества, интересы которых связаны с криптовалютой: Bitcoin Product/service (около 200 тыс. участников), Bitcoin Finance company (70 тыс. пользователей) и группа Blockchain (62 тыс. пользователей). Для периода с октября 2016 по октябрь 2017 были выбраны данные по лайкам и репостам для этих групп и данные курса криптовалюты Биткоин.

Таким образом, были сформированы семь временных рядов ежедневных данных: *Bitcoin* – курс криптовалюты Биткоин относительно доллар США, *BitPr_likes* и *BitPr_reposts* – соответственно количество лайков и репостов в сообществе Bitcoin Product/service, *BitFin_likes* и *BitFin_reposts* – количество лайков и репостов в сообществе Bitcoin Finance company, *Blchain_likes* и *Blchain_reposts* – количество лайков и репостов в сообществе Blockchain. На рис. 2 показаны временные ряды значений курса криптовалюты *Bitcoin* (вверху) и количества лайков (посередине) и репостов (внизу) сообщества Bitcoin Product/service (*BitPr_likes* и *BitPr_reposts*).

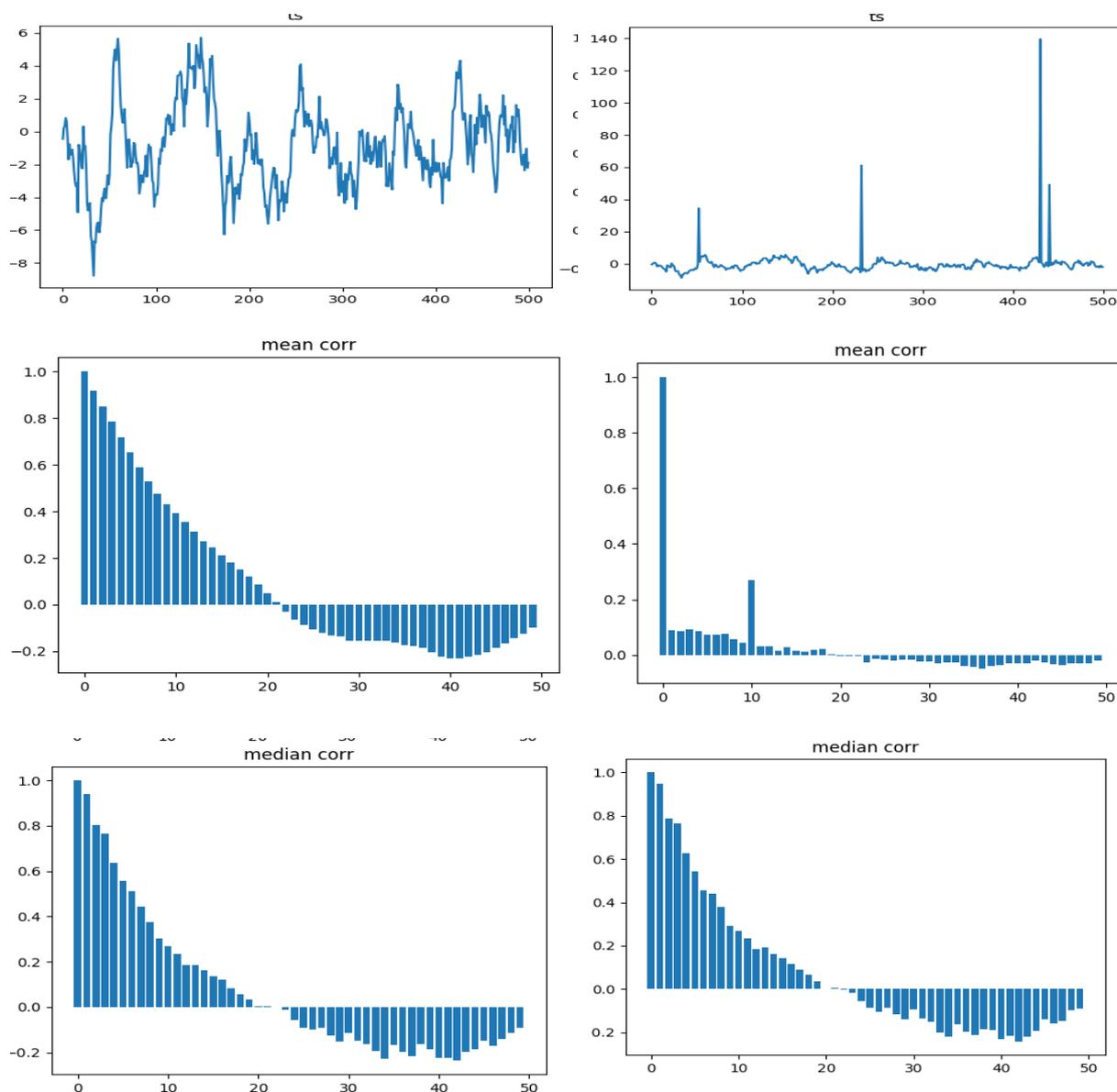


Рисунок 1. Ряд авторегрессии (вверху слева) и этот же ряд с выбросами (вверху справа) и их автокорреляционные функции: в средней части на основе математического ожидания, в нижней - на основе медианы.

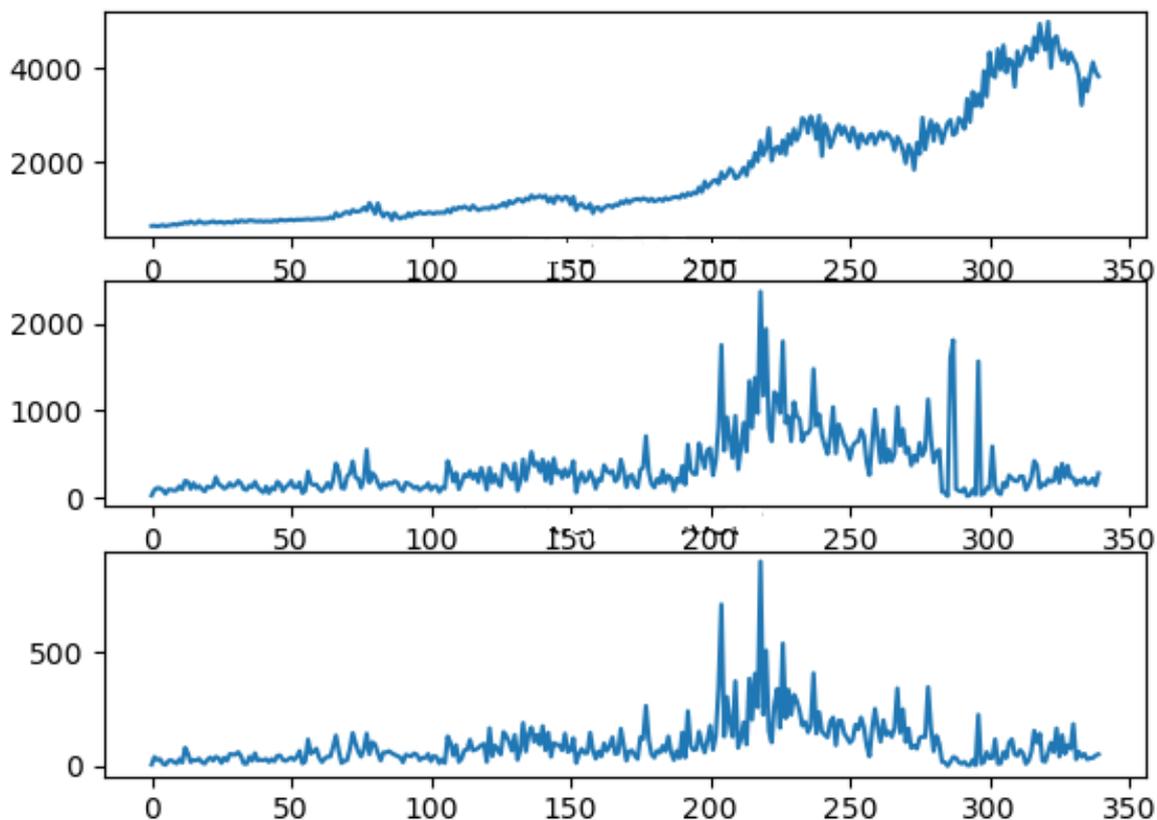


Рисунок 2. Временные ряды *Bitcoin*, *BitPr_likes* и *BitPr_reposts*

В работе проведен кросскорреляционный анализ между рядом курса криптовалюты Биткоин и рядами активностей сообществ. Рассчитаны выборочные коэффициенты кросскорреляции для первых 100 значений временных рядов, когда интерес к биткоину и системе блокчейн в сообществах возрастал вместе с ростом курса биткоина (см. рис. 2) и для последних 100 значений, когда пик интереса постепенно начал спадать. Также были рассчитаны коэффициенты кросскорреляции для близких по структуре рядов лайков и репостов в одних и тех же группах. Поскольку ряды активностей обладают значительными выбросами, при расчете коэффициентов корреляции был использован медианный способ расчета по формуле (9).

Анализ показал значимую корреляцию между курсом биткоина и активностью сообществ. Была обнаружена значительная корреляция между курсом Биткоина и активностью сообществ временных рядов, когда интерес к биткоину и системе блокчейн возрастал вместе с ростом курса биткоина. Для последних значений временных рядов, когда пик интереса постепенно начал спадать, корреляция стремится к нулю.

В табл.1 представлены максимальные коэффициенты корреляции $r(\tau)$ между рядом *Bitcoin* и рядами лайков выбранных сообществ. Между рядами лайков и репостов одного сообщества существует сильная кросс-корреляция в диапазонные значений $0.85 \leq r(0) \leq 0.95$

Таблица 1. Максимальные коэффициенты кросскорреляции

Ряды	$r(\tau)$	
	Начальные значения ряда	Конечные значения ряда
<i>BitPr_likes</i>	0.81	0.18
<i>BitFin_likes</i>	0.58	0.12
<i>Blchain_likes</i>	0.43	-0.08

Проведен мультифрактальный анализ всех рассмотренных временных рядов. Особое внимание уделено проверке наличия свойства самоподобия (2) методом ДФА с использованием формулы (5) и подбором степени полиномиального тренда. Для проверки наличия самоподобия были построены и исследованы флуктуационные функции $F(\tau)$ для каждого временного ряда. Самоподобное поведение соответствует наличию участка функции $F(\tau)$ с аппроксимированной линейной зависимостью (6). На рис. 3 показан график $F(\tau)$ для временного ряда *BitPr_likes*, который был представлен на рис.1 (посредине). Полученная кривая достаточно хорошо аппроксимируется прямой линией с наклоном, определяющим показатель Херста. Исследования показали, что для данного ряда степень полиномиального тренда равна 2.

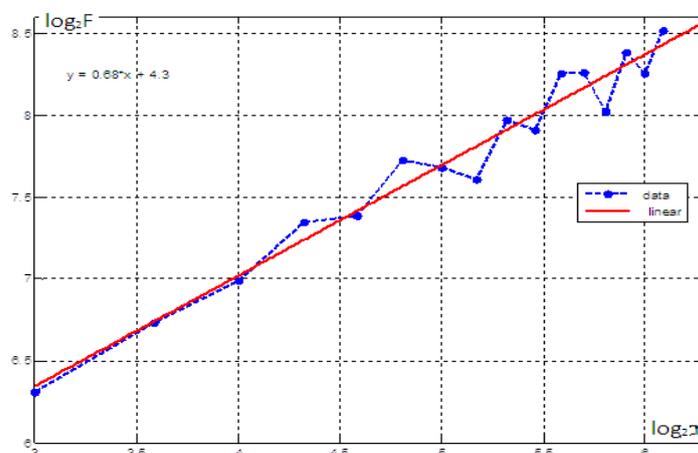
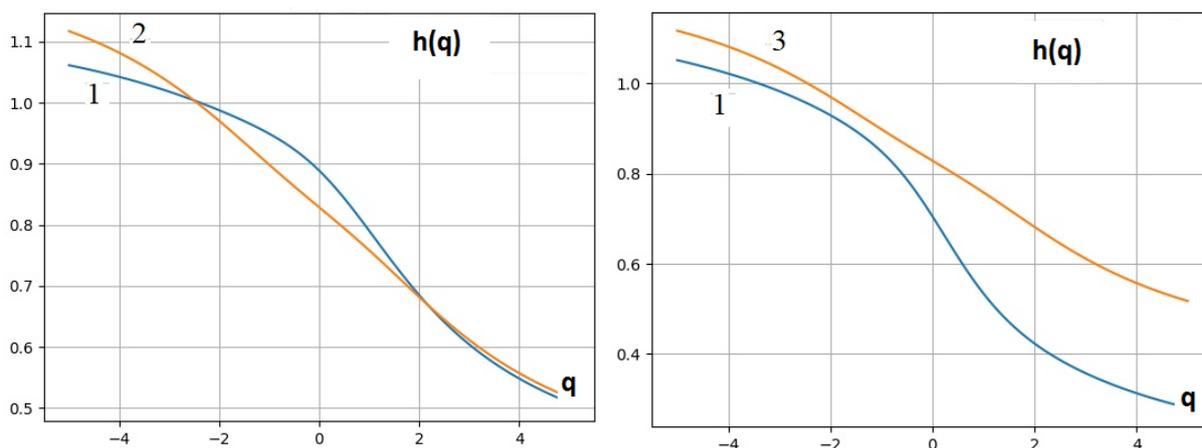


Рисунок 3. Флуктуационная функция $F(\tau)$ для ряда *BitPr_likes*

Исследование флуктуационных функций для временных рядов групп активности показало наличие самоподобных свойств. После этого был проведен мультифрактальный анализ методом МФДФА по формулам (7-8). Исследования показали наличие мультифрактальных свойств для всех рядов групповой активности.

Особое внимание было уделено фрактальному анализу первых значений временных рядов, при возрастающем интересе к биткоину и для последних значений, когда интерес постепенно начал спадать. Рис. 4 представляет функции обобщенного показателя Херста для рядов *Bitcoin* и *BitPr_likes* (в левой части), *Bitcoin* и *Blchain_likes* (в правой части) на интервалах времени соответствующим значениям 100-200 (значимая корреляция) и значениям 250-350 (незначимая корреляция) временных рядов. В верхней части рисунка показаны функции обобщенного показателя Херста для начальных значений рядов, в нижней- для последних значений.

Для начальных значений рядов



Для последних значений рядов

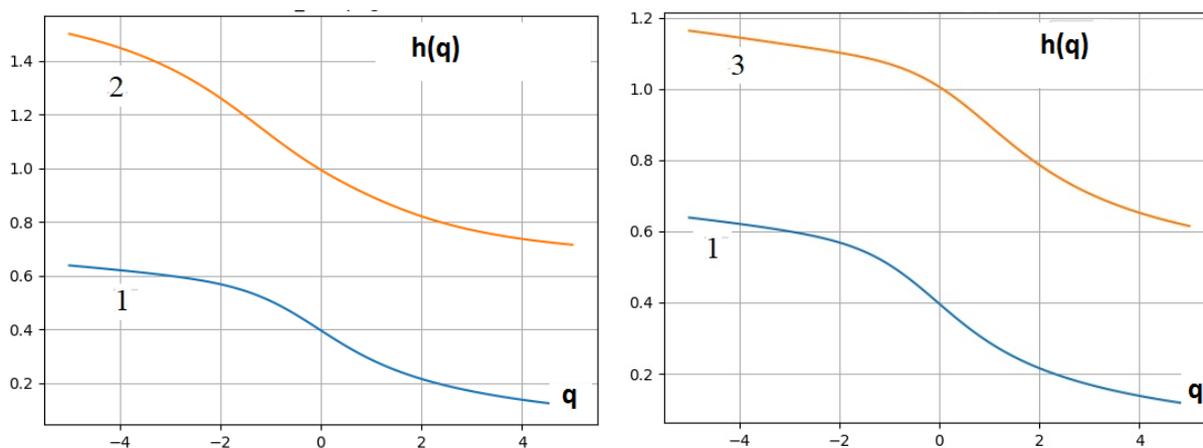


Рисунок 4 Обобщенный показатель Херста $h(q)$ для *Bitcoin* (линия 1) и *BitPr_likes* (линия 2), *Bitcoin* и *Blchain_likes* (линия 3). Вверху значимая корреляция, внизу незначимая корреляция между рядами.

Результаты исследования показали, что ряды с сильной корреляционной зависимостью имеют сходную мультифрактальную структуру. Ряды *Bitcoin* и *BitPr_likes* имеют сильную корреляцию в начальных значениях рядов и близкие значения мультифрактальной характеристики. Слабая корреляция между *Bitcoin* и *Blockchain_likes* проявляется в различии между мультифрактальной структурой этих рядов. При незначимой корреляции между рядами функции обобщенного показателя Херста не имеют сходства.

Поскольку почти всегда ряды лайков и репостов сильно связаны друг с другом, они имеют близкие значения обобщенного показателя Херста. Ряды *BitPr_likes* и *BitPr_reposts* показаны на рис. 2. Их сильная корреляционная зависимость очевидна. На рис. 5 представлены функции обобщенного показателя Херста для рядов *BitPr_likes* и *BitPr_reposts* в начале и конце исследуемого периода времени.

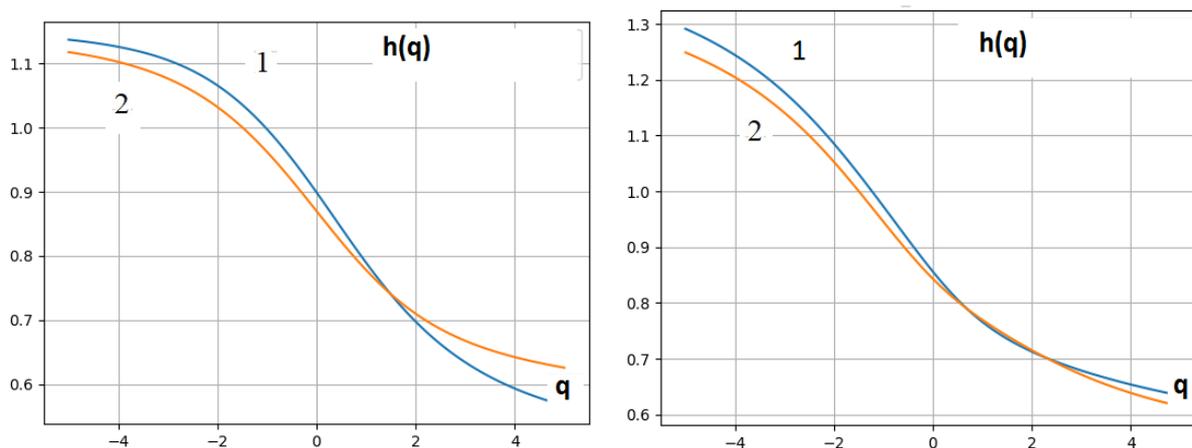


Рисунок 5. Обобщенный показатель Херста $h(q)$ для рядов *BitPr_likes* (линия 1) и *BitPr_reposts* (линия 2), в начале (слева) и конце (справа) исследуемого периода времени.

Аналогичные результаты справедливы и для других исследованных рядов. Чем больше корреляция между фрактальными временными рядами, тем более близки их фрактальные характеристики.

Благодарности

Работа опубликована при поддержке общества ITHEA ISS (www.ithea.org) и ADUIS (www.aduis.com.ua).

Заключение

В работе был проведен корреляционный анализ временных рядов, который выявил значимую корреляцию между курсом Биткоина и активностью связанных с ним социальных сообществ. При расчете коэффициентов корреляции был использован медианный способ расчета, который показал хорошие результаты. Фрактальный анализ временных рядов выявил наличие самоподобных и мультифрактальных свойств как для ряда цен биткоина, так и для рядов активностей сообществ.

Результаты исследований показали, что ряды, имеющие корреляционную зависимость, имеют сходную мультифрактальную структуру: чем больше корреляция между фрактальными рядами, тем более сходны их фрактальные характеристики.

Таким образом, фрактальная структура временных рядов криптовалюты и активности социальных сообществ и, как следствие, динамика их развития взаимосвязаны. Это указывает на необходимость при прогнозировании курсов криптовалют учитывать данные, полученные из анализа социальных сетей.

Литература

- [Arnaboldi, 2015] Valerio Arnaboldi, Andrea Passarella, Marco Conti, Robin I.M. Dunbar, "Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs", Elsevier, 2015, p.116.
- [Bollen, 2011] Bollen, Johan, Huina Mao, and Xiaojun Zeng, "Twitter mood predicts the stock market." Journal of Computational Science, vol. 2.1, p.1-8, 2011.
- [Do, 2015] Do, Rotimi Opeke, James Webb, "CS224W Final Project: Predicting Yelp Ratings From Social Network Data", Department of Computer Science, Stanford University, p.8, December 9, 2015.
- [Feder, 1988] Feder J.: Fractals. Plenum, New York 1988.
- [Feng Mai, 2016] Feng Mai, Qing Bai, Zhe Shan, Xin (Shane) Wang, Roger H. L. Chiang, "From Bitcoin to Big Coin: The Impact of Social Media on Bitcoin Performance", 2016. Electronic copy available at: <http://ssrn.com/abstract=2545957>
- [Garcia, 2014] Garcia D, Tessone CJ, Mavrodiev P, Perony N, "The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy", J. R. Soc. Interface 11, 2014. Electronic copy available at: <http://dx.doi.org/10.1098/rsif.2014.0623>
- [Greaves, 2015] Alex Greaves, Benjamin Au, "Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin", Department of Computer Science, Stanford University, p.8, December 8, 2015.

- [Kantelhardt, 2002] J.W. Kantelhardt, E. Koscielny-Bunde, H.H.A. Rego, S. Havlin, A. Bunde, "Detecting long-range correlations with detrended fluctuation analysis", *Physica A.*, vol. 295, pp. 441-454, 2001.
- [Kantelhardt, 2008] J.W. Kantelhardt, "Fractal and multifractal time series", *Mathematics of complexity and dynamical systems*, pp. 463-487, 2012.
- [Matta, 2016] Martina Matta, "The Predictor Impact of Web Search and Social Media", Theses doctoral research. University of Cagliari, p.95, 2016.
- [Meléndez, 2014] Guillermo Romero Meléndez, "The fractal nature of bitcoin: evidence from wavelet power spectra", *Fundacion Universidad de las Americas Puebla*, 2014, p.39. Electronic copy available at: <http://ssrn.com/abstract=2547281>.
- [Mittal, 2013] A. Mittal, A. Goel, "Stock Prediction Using Twitter Sentiment Analysis" in proceeding of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, p.1-5, 2013.
- [Nakamoto, 2008] Nakamoto, Satoshi, "Bitcoin: A Peer-to-Peer Electronic Cash System", 31 October 2008. Electronic copy available at: <https://bitcoin.org/bitcoin.pdf>
- [Peng Xie, 2017], Hailiang Chen, Yu Jeffrey Hu, "Network Structure and Predictive Power of Social Media for the Bitcoin Market", *Georgia Tech Scheller College of Business Research Paper No. 17-5*, 2017. Electronic copy available at: <https://ssrn.com/abstract=2894089>.
- [Peters, 1994], Peters E.E.: *Fractal Market Analysis: applying chaos theory to investment and economics*. Wiley, 2003.
- [Riedi, 2002] R.H.Riedi, "Multifractal processes", in Doukhan P., Oppenheim G., Taqqu M.S. (Eds.), *Long Range Dependence: Theory and Applications*: Birkhuser, pp. 625–715, 2002.
- [Sho Tsugawa, 2014] Sho Tsugawa, Hiroyuki Ohsaki, "Emergence of Fractals in Social Networks: Analysis of Community Structure and Interaction Locality", in *38th Annual Computer Software and Applications Conference*, 2014.

Информация об авторах



Людмила Кириченко – д.т.н., профессор Харьковского национального университета радиозлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: lyudmyla.kirichenko@nure.ua.

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, фрактальный анализ, вейвлет-анализ, детерминированные хаотические системы.



Тамара Радивилова – к.т.н., доцент Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: tamara.radivilova@gmail.com.

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, телекоммуникационные системы, управление трафиком, информационная безопасность.



Виталий Булах - аспирант Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: bulakhvitalii@gmail.com.

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, интеллектуальный анализ временных рядов.



Вадим Чакрян -, к.т.н., ассистент Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: vadym.chakrian@nure.ua

Основные области научных исследований: менеджмент безопасности, телекоммуникационные сети и системы, риск менеджмент, методы противодействия кибер-атакам

ANALYSIS OF INTERDEPENDENCE OF BITCOIN TIME SERIES AND TIME SERIES OF SOCIAL NETWORKS ACTIVITIES

Lyudmyla Kirichenko, Tamara Radivilova, Vitalii Bulakh, Vadim Chakrian

Abstract. The comparative fractal and crosscorrelation analysis for community groups in the social network and the Bitcoin currency rate is carried out. The time series of daily data of likes and reposts for three social communities and the Bitcoin rate were formed. Time series crosscorrelation analysis revealed a significant correlation between the crypto currency rate and the social communities activity. Fractal analysis indicated the presence of multifractal properties, both for the series of Bitcoin prices, and for ones of community activities. The results have shown that the series having a significant correlation dependence have a similar multifractal structure.

Keywords: self-similar and multifractal time series, generalized Hurst exponent, crosscorrelation, bitcoin, social networks.