

DOI: 10.15593/2499-9873/2020.3.10

УДК 330.46

Р.В. Гарафутдинов, С.А. Ахуньянова

Пермский государственный национальный
исследовательский университет, Пермь, Россия

АДАПТИРОВАННЫЙ МЕТОД КЛЕТЧНОГО ПОКРЫТИЯ ДЛЯ ОЦЕНИВАНИЯ ФРАКТАЛЬНОЙ РАЗМЕРНОСТИ ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДОВ

Настоящая статья продолжает исследования в рамках научного направления по экономике на кафедре информационных систем и математических методов в экономике экономического факультета ПГНИУ.

Моделирование и прогнозирование финансовых временных рядов является довольно перспективным направлением исследований, так как позволяет участникам финансовых процессов снижать риски и принимать эффективные решения. Например, исследовать процессы, происходящие на финансовых рынках, можно с помощью фрактального анализа. Изучен и детально проработан один из методов фрактального анализа финансовых временных рядов – метод клеточного покрытия для оценивания фрактальной размерности. Данный метод часто используется в исследованиях, проводимых отечественными авторами, однако авторы не углубляются в особенности и проблемы применения клеточного метода при анализе временных рядов, а значит, ответы на интересующие вопросы еще не даны. Основная проблема состоит в том, что, как правило, анализируемым объектом в задачах применения клеточного метода к временным рядам является компьютерное изображение графика ряда. Предложена методика адаптации клеточного метода для оценки фрактальной размерности временных рядов, не требующая формирования компьютерного изображения графика. Рассмотрены возникающие при такой адаптации трудности, среди которых: 1) высокая чувствительность получаемой оценки размерности к входным параметрам метода (соотношение сторон плоскости с графиком, покрываемой клетками; используемый диапазон длин стороны клетки; количество разбиений плоскости на клетки); 2) неочевидность выбора оптимальных значений этих параметров. Проанализированы подходы к выбору этих параметров, предложенные другими авторами, определены наиболее подходящие из них для адаптированного метода клеточного покрытия. Разработаны оригинальные способы определения соотношения сторон плоскости с графиком. Написана компьютерная программа, реализующая разработанную методику, произведена ее апробация на сгенерированных данных.

В ходе исследования получены следующие результаты. Подтвержден факт чувствительности адаптированного метода клеточного покрытия к входным параметрам, что говорит о высокой важности правильного выбора этих параметров. Выявлена несостоятельность предложенных способов автоматического определения соотношения сторон плоскости применительно к искусственным временным рядам. Наиболее точные (в статистическом смысле) оценки фрактальной размерности, найденные с помощью адаптированного клеточного метода, получены при фиксированном соотношении сторон 1:1. При сравнении адаптированного клеточного метода и R/S-анализа наиболее точные оценки фрактальной размерности получены вторым методом.

Сформулированы возможные направления дальнейших исследований: 1) сравнение точности различных методов оценивания фрактальной размерности на рядах разной длины; 2) сравнение методов фрактального анализа и p -адического анализа при моделировании и прогнозировании финансовых временных рядов; 3) определение условий применимости различных методов; 4) апробация разработанных способов определения соотношения сторон плоскости с графиком на реальных экономических данных.

Ключевые слова: метод клеточного покрытия, фрактальная размерность, показатель Херста, финансовые временные ряды, R/S-анализ, фрактальный анализ, экономическая физика, фрактальное броуновское движение, процессы с длинной памятью, вычислительная геометрия, программная реализация, Python, R.

R.V. Garafutdinov, S.A. Akhunyanova

Perm State University, Perm, Russian Federation

ADAPTED BOX-COUNTING METHOD FOR ASSESSMENT OF THE FRACTAL DIMENSION OF FINANCIAL TIME SERIES

This paper continues research within the framework of the scientific direction in econophysics at the Department of Information Systems and Mathematical Methods in Economics of the faculty of Economics of PSU.

Modeling and prediction of financial time series is quite a perspective area of research, because it allows participants of financial processes to reduce risks and make effective decisions. For example, we could research financial processes with the help of fractal analysis. In the article there is studied and worked out in detail one of the methods of fractal analysis of financial time series – the box-counting method for assessment of the fractal dimension. This method is often used in studies conducted by domestic authors, but the authors do not delve into the characteristics and problems of using the box-counting method for analysis of time series, that means that the answers to the interested questions have not yet been given. The main problem is that, as a rule, the analyzed object in the tasks of applying the box-counting method to time series is a computer image of the plot of series. In the article there is proposed the procedure of adaptation of the box-counting method for assessment of the fractal dimension of time series, the procedure does not require the formation of a computer image of the plot. In the article there is considered following difficulties developed from this adaptation: 1) high sensitivity of the resulting estimation of the dimension to the input parameters of the method (the ratio of the sides of the covered by cells plane with the plot; the used range of lengths of the side of the cell; the number of partitions of the plane into cells); 2) the non-obviousness of choosing the optimal values of these parameters. In the article there are analyzed approaches to the selection of these parameters that were proposed by other authors, and there are determined the most suitable approaches for the adapted box-counting method. Also there are developed unique methods for determining the ratio of the sides of the plane with the plot. In the paper there is written the computer program that implements the developed method, and this program is tested on the generated data.

The study obtained the following results. The fact of sensitivity of the adapted box-counting method to input parameters is confirmed, that indicates the high importance of the correct choice of these parameters. According to the study, there is found out inability of the proposed methods of automatic determination the ratio of the sides of the plane in relation to artificial time series. There are obtained the most precise (in a statistical sense) estimates of fractal dimension, those found by means of the adapted box-counting method, with the fixed ratio of the sides 1:1. According to comparing the adapted box-counting method and R/S analysis, there are obtained the most precise estimates by the second method (R/S analysis).

Finally in the paper there are formulated the possible directions for further research: 1) comparison of the accuracy of various methods for assessment of the fractal dimension on series of different lengths; 2) comparison of the methods of fractal analysis and p -adic analysis for modeling and prediction of financial time series; 3) determination of the conditions of applicability of various methods; 4) approbation of the developed methods for determining of the ratio of the sides of the plane with the plot on real economic data.

Keywords: box counting method, fractal dimension, Hurst exponent, financial time series, R/S analysis, fractal analysis, econophysics, fractal Brownian motion, processes with long memory, computational geometry, software implementation, Python, R.

Введение

Одним из наиболее перспективных подходов к изучению поведения финансовых рынков является фрактальный подход. Он основан на гипотезе фрактального рынка Б. Мандельброта [1]. Эта гипотеза гласит, что «процесс ценообразования на рынках глобально детерминирован,

зависим от “начальных условий” [2], прошлые цены влияют на будущие, а графики ценовых рядов финансовых инструментов обладают фрактальными свойствами: самоподобием, масштабированием по степенному закону и фрактальной размерностью» [2, 3]. По мнению ученых – приверженцев данной теории, исследование финансовых рядов методами фрактального анализа позволяет с достаточно высокой точностью судить об их будущем поведении: предсказывать критические точки (кризисы) и строить точечные и/или интервальные прогнозы.

Изначально фрактальный подход, включающий в себя методы физики, использовался при исследовании физических объектов и пространственных структур. Затем его начали применять при исследовании экономических процессов, где в качестве пространственных структур стали рассматривать графики временных рядов [4–7]. Сегодня фрактальный подход является одним из методов особого направления в экономической науке – эконофизике, возникшей на рубеже XX–XXI вв. [8]. Строго говоря, эконофизика – это наука, которая применяет методы физики для изучения экономических систем. В настоящий момент продолжают исследования по эконофизике на кафедре информационных систем и математических методов в экономике экономического факультета Пермского государственного научного исследовательского университета (ПГНИУ). На кафедре ПГНИУ с помощью методов эконофизики изучаются модели финансовых пузырей, модель взаимодействия агентов на рынке [9]. Также на кафедре в рамках эконофизики изучаются фрактальный подход, представленный в работах [3, 10, 11], и p -адический метод, представленный в работах [12, 13]. Далее дадим характеристику фрактальному анализу.

Сущность фрактального анализа состоит «в обработке масштабированием ... фрактальной структуры и описании распределения какой-либо структурной характеристики – или меры – при этом масштабировании» [14]. Для этого определяется фрактальная размерность структуры D – численный показатель, характеризующий то, как объект заполняет пространство, его сложность. Применительно к временным рядам говорят о фрактальной размерности их графиков как пространственных структур. В отличие от других структур, для временных рядов существует собственная мера фрактальности – показатель Херста H , характеризующий свойство ряда, названное персистентностью (способностью ряда сохранять свою тенденцию) [2, 3]. Персистентность тесно

связана с наличием в ряду так называемой длинной памяти – зависимости между далеко отстоящими друг от друга во времени значениями исследуемой переменной. Более подробно связь самоподобия (другими словами, фрактальности) временного ряда, наличия у него длинной памяти и персистентности показана в работах Б. Мандельброта [15]. Если исследуемый временной ряд является реализацией гауссовского случайного процесса, то показатель H связан с фрактальной размерностью D простым линейным соотношением $H = 2 - D$ [16]. Ввиду этого фрактальный анализ временных рядов заключается в определении их фрактальных свойств как с помощью оценки размерности, так и с помощью оценки показателя Херста.

Оценки фрактальных показателей (фрактальной размерности и показателя Херста) исследуемого временного ряда или его фрагмента (в этом случае речь идет о локальной размерности – понятии, введенном Н.В. Старченко [4]) позволяют судить о поведении этого ряда в будущем. И, конечно, при вычислении этих оценок важно получать значения, близкие к истинным. В настоящее время существует большое количество методов и способов оценивания фрактальных показателей, например метод R/S-анализа, метод Фурье, метод максимумов модулей коэффициентов вейвлет-преобразования, методы минимального и клеточного покрытия и другие [17]. На фоне большого количества известных методов работ, посвященных их подробному описанию и сравнению друг с другом, немного.

В данной статье предпринята попытка детальной проработки одного из самых популярных и в то же время простых, интуитивно понятных методов оценки фрактальной размерности – метода клеточного покрытия. Новизна работы заключается в том, что русскоязычные авторы используют этот метод, но подробно его не рассматривают, хотя его применение имеет определенные сложности. Например, в работе [16] сказано о том, что клеточный метод «реализуется при графическом выводе фрактала на экран РС как массива пикселей» (под РС автором подразумевается персональный компьютер). На наш взгляд, анализ временного ряда сложно свести к анализу пикселей его графика, так как пиксели – всего лишь техническая особенность визуализации ряда на экране компьютера. Характеристики массива пикселей зависят от характеристик экрана, в частности от его разрешения, в то время как временной ряд существует сам по себе и никаким образом не должен быть связан с аппаратной конфигурацией

компьютера исследователя. Более оправданным выглядит подход к определению размерности, основанный на аналитических вычислениях. Таким образом, для достижения цели исследования дано описание метода клеточного покрытия; предложен механизм его адаптации к временным рядам; разработана программная реализация метода на языке Python; проведена апробация разработанной программы на сгенерированных временных рядах и выполнено сравнение с реализациями клеточного метода и R/S-анализа из библиотек языка R.

1. Теоретический обзор

Рассмотрим основные результаты и выводы по методу клеточного покрытия тех ученых, которые занимались проблемой применения этого метода для фрактального анализа временных рядов.

Среди отечественных авторов популярность клеточного метода как средства исследования не пространственных, а временных структур невелика. К числу наиболее значимых работ можно отнести диссертацию Н.В. Старченко, в которой им разработана модификация клеточного метода, получившая название метода минимального покрытия. Автором утверждается, что минимальное покрытие точнее клеточного и что использование минимальных покрытий вместо клеточных позволяет существенно уменьшить количество данных, необходимое для определения фрактальной размерности [4]. Предложенный им метод пользуется популярностью у других исследователей [18–20], ссылающихся на выводы Старченко о преимуществах его подхода, изложенные в диссертации и сопутствующих работах. Другим исследователем – Е.К. Кривоносовой – в работе [16] применяются методы как клеточного, так и минимального покрытия для анализа и прогнозирования котировок акций ПАО «Лукойл» и выручки предприятий Приволжского федерального округа. Автором сделан вывод о том, что определенная обоими методами фрактальная размерность может быть использована в качестве индикатора стабильности работы предприятий.

Среди англоязычных авторов клеточному методу как подходу к фрактальному анализу временных рядов уделяется больше внимания. Так, в работе [7] выполнен обзор различных алгоритмов оценки фрактальной размерности как меры гладкости (шероховатости) временных рядов и пространственных данных, в том числе рассмотрены различные вариации клеточного метода (box-counting). Авторами также раз-

работана библиотека `fractaldim` для языка программирования R, содержащая функции, реализующие рассмотренные в указанной работе методы. Подробный анализ подходов к оценке фрактальной размерности временных рядов и возникающих при этом сложностей выполнен I. Pilgrim и R. Taylor. В их публикации [21] озвучены проблемы, характерные для клеточного метода, которые, как правило, не затрагиваются в отечественных источниках. (Следует отметить, что в процессе программной реализации метода мы столкнулись с теми же трудностями, что и Pilgrim и Taylor в своей работе.) Авторами сделан вывод, что клеточный и другие методы фрактального анализа временных рядов должны применяться с осторожностью к данным, «ограниченным по длине или спектральному содержанию», так как подобные ограничения могут препятствовать выявлению их фрактальных свойств. В работе [22] N.C. Kenkel подробно рассмотрены требования, предъявляемые к величине конечной выборки данных, по которой оценивается размерность клеточным методом. Автором сделан вывод, что для надежной оценки показателя требуются очень большие размеры выборки: обычно на порядки больше, чем используемые в большинстве эмпирических исследований, однако отмечено, что, если в качестве данных выступает временной ряд, объем выборки определяется просто как длина этого ряда. Также автором предложена вероятностная модель для вычисления минимально допустимого размера клетки, в которой общее число покрывающих фигуру клеток является функцией от размера клетки и величины клеточной размерности геометрического объекта. Согласно этой модели, минимальный размер клетки выбирается таким образом, чтобы отклонение размерности от величины ее оценки составляло не более $\epsilon = 0,001$. E. Rosenberg в работе [23] предложенную Kenkel модель обобщает на случай произвольного значения ϵ .

Проведенный анализ литературы дает основания полагать, что в русскоязычных источниках клеточный метод освещен далеко не так подробно, как в зарубежных, при этом авторами он сводится к анализу растрового изображения графика временного ряда. В настоящей работе произведена попытка вспомнить этот пробел, логичным образом адаптировать метод к временным рядам, устранив необходимость формирования компьютерного изображения графика ряда, и изложить возникшие при этом проблемы. Но, прежде чем приступить к описанию разработанной методики, следует кратко изложить теоретические основы метода клеточного покрытия как способа определения фрактальной размерности.

Термин «размерность» означает число измерений некоторого геометрического объекта. Топологической размерностью называют минимальное количество координат, необходимое для описания объекта [3]. Эта величина может принимать только целые значения, поэтому подходит лишь для описания «идеальных», гладких объектов (прямая, куб, сфера), однако в природе такие объекты почти не встречаются. Ф. Хаусдорф обратил внимание, что если геометрический объект покрывать шарами радиусом δ , то минимальное число шаров $N(\delta)$ с уменьшением их радиуса возрастает по степенному закону $N(\delta) \sim \left(\frac{1}{\delta}\right)^{D_H}$. Показатель D_H получил название размерности Хаусдорфа. Его формула имеет вид [4]

$$D_H = \lim_{\delta \rightarrow 0} \frac{\ln N(\delta)}{-\ln \delta}. \quad (1)$$

Размерность Хаусдорфа обобщила понятие топологической размерности. Показатель D_H может принимать как целые, так и дробные значения, что оказалось весьма полезным свойством для описания объектов реального мира. D_H показывает, насколько плотно и равномерно объект заполняет пространство, например при малой изрезанности кривой значение ее D_H приближается к 1, при большой – к 2 (рис. 1).

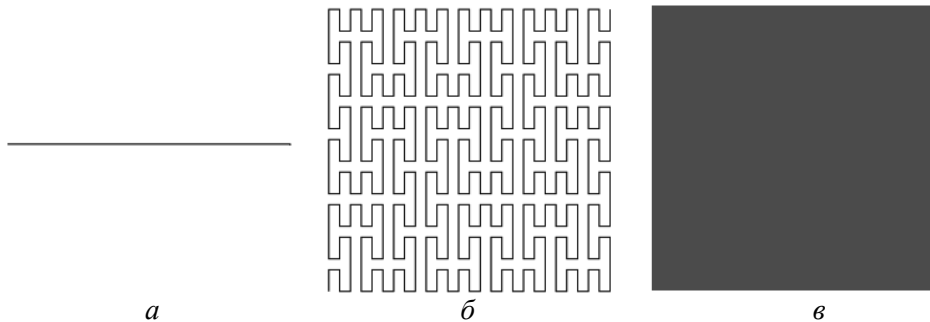


Рис. 1. Геометрические фигуры различной размерности: a – прямая, $D_H = 1$; b – кривая Пеано $D_H \in (1; 2)$; c – полностью заполненное пространство $D_H = 2$

Позднее Б. Мандельбротом был введен термин «фрактальная размерность» для обозначения частного случая размерности Хаусдорфа, в котором она принимает дробные значения, превышающие значе-

ние топологической размерности. Объекты, обладающие такой размерностью, Мандельброт назвал фракталами [1].

Метод клеточного покрытия весьма прост и основан на идее Хаусдорфа о степенной зависимости между количеством фигур элементарной формы, полностью покрывающих геометрический объект, и размером фигуры. Чем меньше фигура, тем больше таких фигур необходимо для полного покрытия объекта. Как следует из названия метода, в качестве фигуры в двумерном пространстве используется клетка, т.е. прямоугольник. В трехмерном пространстве фигурой является «коробка» (box, отчего англоязычное название метода – box-counting method). Далее будет рассмотрена разновидность метода клеточного покрытия, используемая для анализа объектов двумерного пространства.

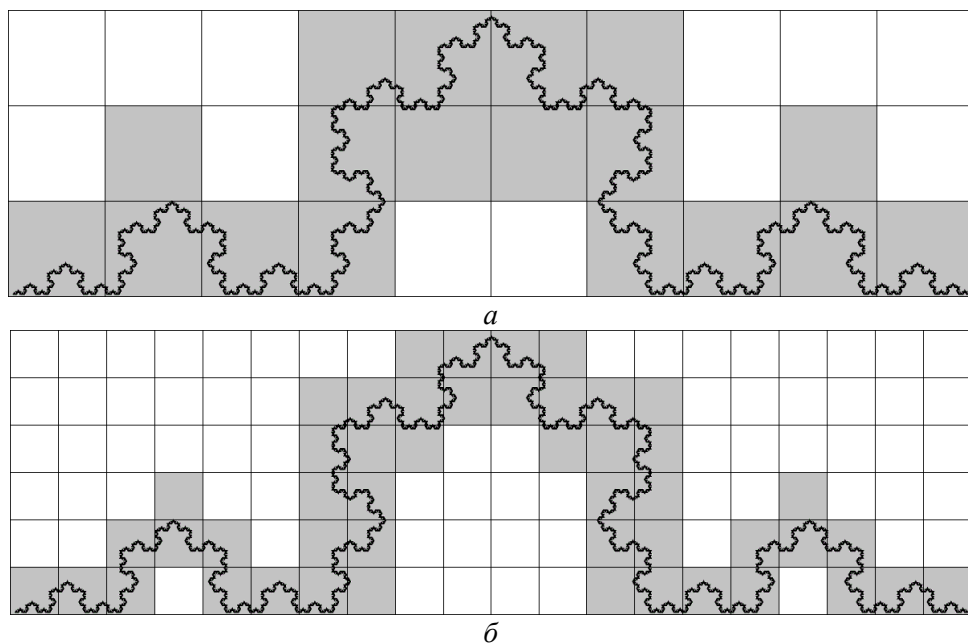


Рис. 2. Графики фрактального объекта (кривой Коха), где серым цветом выделены клетки, содержащие хотя бы одну точку анализируемого объекта. Переходя от a к b , можно заметить, что при уменьшении размера клетки со стороны δ происходит увеличение числа клеток и точности (аккуратности) покрытия объекта

Алгоритм клеточного метода состоит в следующем. Плоскость, содержащая анализируемое изображение объекта, покрывается сеткой, которая делит ее на квадратные клетки стороной δ . Подсчитывается ко-

личество клеток $N(\delta)$, содержащих в себе хотя бы одну точку изображения. Затем поступательно уменьшается δ и подсчитываются соответствующие значения $N(\delta)$ (рис. 2). После в двойном логарифмическом масштабе строится график функции $N(\delta)$, который аппроксимируется прямой с помощью метода наименьших квадратов. Тогда величина клеточной размерности D , являющаяся тангенсом угла наклона данной прямой относительно оси Ox , вычисляется как взятый с противоположным знаком коэффициент при независимой переменной уравнения линейной регрессии $\ln N(\delta) = -D \ln \delta + b$, где b – свободный член.

В работе [21] предлагается модификация данного подхода, названная авторами вариационным методом клеточного покрытия (variational box-counting method), суть которой состоит в том, что клетки каждого столбца располагаются не строго по сетке, а таким образом, чтобы минимизировать их количество, необходимое для покрытия фигуры (рис. 3).

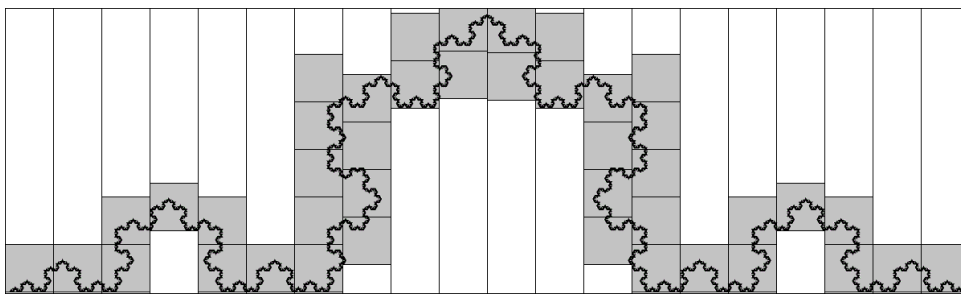


Рис. 3. График фрактального объекта (кривой Коха), где серым цветом выделены клетки, содержащие хотя бы одну точку анализируемого объекта. Количество клеток, по сравнению с рис. 2 б, меньше

В результате модификации количество клеток, соответствующих определенной величине δ , уменьшается. Однако, по утверждению авторов, внесение данных изменений в метод не оказывает значимого влияния на вычисленное значение размерности.

2. Адаптация метода клеточного покрытия для его применения к временным рядам

Как было отмечено во введении, обычно в публикациях предлагается использовать клеточный метод для анализа графика временного ряда как изображения. Нами предпринята попытка отойти от замены

анализа временного ряда анализом компьютерного изображения его графика и применить аналитические расчеты. Опишем ниже механизм адаптации к временным рядам метода клеточного покрытия.

Пусть дан дискретный временной ряд V , состоящий из элементов v_t , где $t = \overline{1, T}$. Очевидно, что для реализации клеточного метода от понятия «график ряда» избавиться не удастся, однако под графиком следует понимать не растровое изображение, а совокупность точек ряда, соединенных между собой. Поскольку временной ряд является численным дискретным описанием непрерывного динамического процесса, значение показателя которого зафиксировано в моменты времени t , видится логичным принять в качестве формы траектории движения показателя наиболее простую из возможных – прямую (линейный тренд), т.е. представить график ряда как некоторую кусочно-линейную функцию, непрерывную и определенную на интервале $[1; T]$. Не менее очевидно, что в подсчете должны участвовать клетки, покрывающие не только точки, но и отрезки между ними.

Прежде всего, необходимо ввести абстрактную плоскость (холст), вмещающую в себя график ряда. Пусть W – ширина холста, H – высота (безразмерные величины). Левый нижний угол холста лежит в точке с координатами $(0; 0)$, правый верхний угол имеет координаты $(W; H)$. Плоскость следует покрыть сеткой, т.е. разбить на квадратные клетки со стороной δ . C_W – ряд, содержащий границы клеток по горизонтали, состоит из элементов $\left[0, \delta, 2\delta, \dots, \left(\left\lfloor \frac{W}{\delta} \right\rfloor + 1 \right) \delta \right]$. C_H – аналогичный ряд границ клеток по вертикали, состоящий из элементов $\left[0, \delta, 2\delta, \dots, \left(\left\lfloor \frac{H}{\delta} \right\rfloor + 1 \right) \delta \right]$. Правые границы рядов вычисляются таким образом, потому что длина и высота холста могут не быть кратными стороне клетки, в то же время клетки, покрывающие график, несомненно, должны быть учтены даже в том случае, если они выходят за границы холста.

Следующим шагом (или, скорее, комплексом шагов) является отображение на холсте графика временного ряда. Каждое значение ряда v_t является точкой с координатами $(t; v_t)$. Возникает необходимость соотнести координатную сетку данных с введенной ранее абстрактной коор-

динатной сеткой холста, разделенного на клетки. Для этого следует «подогнать» первую ко второй, т.е. отмасштабировать ее координаты таким образом, чтобы они соответствовали координатам холста. Известно, что «плоскость данных» ограничена диапазонами $[1; T]$ по горизонтали и $[v_{\min}; v_{\max}]$ по вертикали, где v_{\min} , v_{\max} – соответственно минимальное и максимальное значения ряда, а холст с клетками ограничен диапазонами $[0; W]$ и $[0; H]$. Чтобы привести «шкалу данных» к «шкале холста», имеет смысл воспользоваться следующими формулами:

$$x_{\text{scaled}_t} = t \frac{W}{T - 1}, \quad (2)$$

$$y_{\text{scaled}_t} = H \frac{v_{\max} - v_t}{v_{\max} - v_{\min}}, \quad (3)$$

где x_{scaled_t} , y_{scaled_t} – приведенные к единицам измерения холста координаты точки, соответствующей значению ряда v_t .

Теперь известны границы всех клеток (ячеек сетки), на которые разделен холст, и координаты всех точек ряда, по которым можно определить, входит ли та или иная точка в любую из клеток. Однако, как было сказано выше, недостаточно выявить клетки, покрывающие точки, необходимо также найти и учесть те клетки, внутри которых находятся отрезки (целиком или фрагментом), соединяющие соседние точки. С одной стороны, координатами концов этих отрезков являются координаты самих точек, поэтому клетки, покрывающие точки, не имеет смысла учитывать отдельно: они войдут в набор клеток, пересекемых отрезками. С другой стороны, необходимо выявить клетки, содержащие отрезки, соединяющие соседние точки. Следует рассматривать не каждую очередную точку в отдельности, а пары соседних точек. Пусть A – точка с координатами $(x_{\text{scaled}_t}; y_{\text{scaled}_t})$, соответствующая значению ряда v_t ; B – точка с координатами $(x_{\text{scaled}_{t+1}}; y_{\text{scaled}_{t+1}})$, соответствующая значению ряда v_{t+1} . Точки A и B , помещенные на сетку холста, попадают в клетки, которые являются углами прямоугольника P . Среди всех клеток, составляющих этот прямоугольник, следует выделить те, которые покрывают отрезок AB (рис. 4).

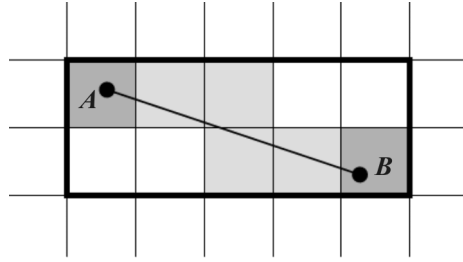


Рис. 4. Прямоугольник P и составляющие его клетки.
Серым цветом выделены клетки, покрывающие отрезок AB

При этом нужно учесть, что точка A может находиться как в левой верхней клетке, так и в левой нижней, а B – соответственно в правой нижней или верхней, что зависит от того, какая из точек на холсте расположена выше (какое значение из v_t, v_{t+1} больше). Таким образом, нужно перебрать все клетки прямоугольника и для каждой определить, содержит ли она фрагмент отрезка AB либо весь отрезок целиком.

Определить границы клетки, содержащей точку с координатами $(p; q)$, можно по следующим формулам. Пусть C_W (ряд, содержащий границы клеток по горизонтали) состоит из элементов c_{Wi} , где $i = \overline{1, I_W}$, I_W – общее число элементов ряда C_W . C_H (ряд, содержащий границы клеток по вертикали) состоит из элементов c_{Hj} , где $j = \overline{1, J_H}$, J_H – общее число элементов ряда C_H . Тогда

$$i_p = \left\lfloor \frac{p}{c_{Wi_W}} (I_W - 1) \right\rfloor, \quad (4)$$

где i_p – номер элемента ряда C_W , ближайшего, не меньшего p ; c_{Wi_W} – последний элемент ряда C_W ;

$$j_p = \left\lfloor \frac{q}{c_{HJ_H}} (J_H - 1) \right\rfloor, \quad (5)$$

где j_p – номер элемента ряда C_H , ближайшего, не меньшего q ; c_{HJ_H} – последний элемент ряда C_H . Номера элементов ряда C_W , являющихся горизонтальными границами клетки, покрывающей точку, можно рассчитать по формулам

$$i_{p_0} = \begin{cases} i_p - 1, i_p = I_W; \\ i_p, i_p \neq I_W; \end{cases} \quad (6)$$

$$i_{p_1} = \begin{cases} i_p, i_p = I_W; \\ i_p + 1, i_p \neq I_W; \end{cases} \quad (7)$$

где i_{p_0} , i_{p_1} – номера левой и правой границ соответственно. Аналогичным образом можно вычислить номера элементов ряда C_H , являющихся вертикальными границами клетки:

$$j_{p_0} = \begin{cases} j_p - 1, j_p = J_H; \\ j_p, j_p \neq J_H; \end{cases} \quad (8)$$

$$j_{p_1} = \begin{cases} j_p, j_p = J_H; \\ j_p + 1, j_p \neq J_H; \end{cases} \quad (9)$$

где j_{p_0} , j_{p_1} – номера нижней и верхней границ соответственно. Подставив в формулы (4)–(9) координаты точек A и B , можно определить номера границ клеток, покрывающих эти точки: i_{a_0} , i_{a_1} , j_{a_0} , j_{a_1} – номера левой и правой, нижней и верхней границ клетки, покрывающей точку A , и соответственно i_{b_0} , i_{b_1} , j_{b_0} , j_{b_1} – номера границ клетки, покрывающей точку B . Тогда границы прямоугольника P можно вычислить следующим образом: $i_{p_0} = i_{a_0}$, $i_{p_1} = i_{b_1}$, $j_{p_0} = \min(j_{a_0}, j_{b_0})$, $j_{p_1} = \max(j_{a_1}, j_{b_1})$, где i_{p_0} , i_{p_1} , j_{p_0} , j_{p_1} – номера левой, правой, нижней и верхней границ соответственно. Теперь, ограничив ряд C_W диапазоном $[i_{p_0}; i_{p_1}]$, а ряд C_H диапазоном $[j_{p_0}; j_{p_1}]$, можно перебрать все клетки прямоугольника P , координатами углов которых будут являться те или иные элементы этих рядов. Количество клеток прямоугольника определяется как произведение числа элементов первого ряда, уменьшенного на единицу, и числа элементов второго ряда, также уменьшенного на единицу.

Далее, для определения того, является ли клетка закрашенной, т.е. содержащей внутри себя фрагмент отрезка AB , следует выявить факт пересечения отрезка AB и хотя бы одной из сторон этой клетки. Стоит отметить, что, несмотря на возможность ситуации, когда отрезок AB находится целиком внутри клетки и не касается ни одной из ее сторон, та-

кой подход к выявлению закрашенных клеток в конечном счете сработает всегда: даже если на одной из итераций, при определенной паре точек A, B , обе из них попадут внутрь клетки, в результате чего она не будет учтена как закрашенная, на следующей (или предыдущей) итерации данная клетка все равно окажется учтена, потому что хотя бы одна ее грань будет пересечена отрезком, соединяющим предыдущую точку и A либо B и следующую точку. В случае же если клетка единственная и покрывает весь холст целиком, ее левая и верхняя границы совпадают с первым значением ряда V и его максимальным значением v_{\max} , а следовательно, как минимум два из возможных отрезков AB пересекутся с одной из сторон этой клетки и она окажется учтена.

Чтобы определить, пересекаются ли отрезки, можно воспользоваться методом, основанным на косом произведении векторов [24]. Пусть CD – отрезок, соответствующий стороне клетки, его концы имеют координаты $C(x_C; y_C)$, $D(x_D; y_D)$. Известно, что концы отрезка AB имеют координаты $A(x_{\text{scaled } t}; y_{\text{scaled } t})$, $B(x_{\text{scaled } t+1}; y_{\text{scaled } t+1})$. Можно ввести следующие обозначения максимумов и минимумов координат точек: $x_{\max AB} = \max(x_{\text{scaled } t}, x_{\text{scaled } t+1})$, $x_{\min AB} = \min(x_{\text{scaled } t}, x_{\text{scaled } t+1})$, $x_{\max CD} = \max(x_C, x_D)$, $x_{\min CD} = \min(x_C, x_D)$, $y_{\max AB} = \max(y_{\text{scaled } t}, y_{\text{scaled } t+1})$, $y_{\min AB} = \min(y_{\text{scaled } t}, y_{\text{scaled } t+1})$, $y_{\max CD} = \max(y_C, y_D)$, $y_{\min CD} = \min(y_C, y_D)$. Тогда условием пересечения отрезков AB и CD будет являться выполнение следующих неравенств (в левой части неравенств 2 и 3 приведены косые произведения векторов):

- 1) $x_{\max AB} \geq x_{\min CD}$, $x_{\max CD} \geq x_{\min AB}$, $y_{\max AB} \geq y_{\min CD}$, $y_{\max CD} \geq y_{\min AB}$;
- 2) $[\overline{AC}, \overline{AB}] \cdot [\overline{AD}, \overline{AB}] \leq 0$;
- 3) $[\overline{CA}, \overline{CD}] \cdot [\overline{CB}, \overline{CD}] \leq 0$.

В случае, если отрезок AB пересекается хотя бы с одной из сторон клетки, эта клетка считается закрашенной и учитывается в подсчете величины $N(\delta)$. Разумеется, одна и та же клетка может быть учтена лишь один раз.

Результатом перебора всех значений ряда V и выполнения описанных шагов является пара значений: $N(\delta)$ – подсчитанное количество закрашенных клеток и δ – размер стороны клетки. Дальнейшая после-

довательность действий полностью соответствует алгоритму метода клеточного покрытия, применяемому к анализу пространственных объектов: многократно уменьшается величина δ , вычисляются соответствующие значения $N(\delta)$, затем логарифмируются полученные ряды и определяется величина фрактальной размерности D из уравнения линейной регрессии $\ln N(\delta) = -D \ln \delta + b$.

В том случае, если размах ряда V равен нулю, т.е. каждый элемент ряда $v_i = \text{const}$, а график представляет собой отрезок прямой, параллельной оси Ox , для вычисления отмасштабированной координаты любой точки ряда $y_{\text{scaled } t}$, согласно формуле (3), необходимо разделить на ноль, что приведет к невозможности реализации приведенной методики. Это частный случай входных данных, который нами предлагается учитывать особо и для такого ряда принять D равной величине теоретической размерности линии – единице.

Несмотря на то, что изложенная методика вычисления клеточной фрактальной размерности временного ряда вполне соответствует сути метода и может быть использована, при реализации ее на практике возникает ряд проблем, требующих разрешения.

Помимо самого ряда V , входными данными алгоритма являются диапазон изменения размера клетки δ и ширина и высота холста W, H . Вопросы определения этих параметров в работах отечественных авторов обычно опускаются. В то же время данные величины влияют на получаемую оценку размерности D , поэтому их произвольный выбор может привести к искажению этой оценки.

Как отмечает E. Rosenberg в работе [23], проблема определения диапазона значений δ широко обсуждается в работах иностранных авторов. Очевидно, что необходимо выбрать такой диапазон, при котором становится возможным исследовать тенденцию масштабирования. Согласно работам [22, 23], не следует использовать величину δ , при

которой $N(\delta) > \frac{N}{10}$, где N – общее количество точек изображения. Поскольку величины δ и $N(\delta)$ имеют обратную зависимость, минимальный размер клетки δ_{\min} достигается при выполнении равенства $N(\delta) = \frac{N}{10}$. Другими словами, при реализации метода можно умень-

шать δ до тех пор, пока не будет выполняться упомянутое равенство. Однако в случае временного ряда остается неясным, какой величине соответствует показатель N . Выше было введено понятие холста абстрактного размера $W \times H$, на который «помещался» график ряда V . Визуализация метода клеточного покрытия на компьютере помогает придать величинам W и H вполне конкретный смысл: их можно рассматривать как ширину и высоту окна с графиком и клетками в пикселях, в этом случае $N = WH$. Но точку, из множества которых состоит холст, нельзя приравнять к пикселю изображения на экране монитора, она не является шагом дискретизации ряда V ни по горизонтали, ни по вертикали. Во-первых, это следует из того факта, что количество значений ряда T может превышать величину W , а значит, «расстояние» между соседними наблюдениями v_t и v_{t+1} будет меньше единицы. Во-вторых, вопрос о дискретизации размаха значений ряда V , который будет рассмотрен ниже, и вовсе является трудноразрешимым. Исходя из этого, на наш взгляд, предложенный способ определения δ_{\min} применим для анализа пространственных объектов, но не временных рядов.

Авторы работы [7] предлагают определять δ путем деления длины ряда V на степени двойки (в нашем случае делимым выступает ширина холста W ; обозначим количество фрагментов холста, на которые он делится по ширине, как $\gamma = \frac{W}{\delta}$). Тогда $\gamma = 2^k$, где $k = \overline{0, K}$, K таково, что $2^k \leq T$. При этом на диапазон значений γ накладываются ограничения: отбрасываются те γ , для которых $N(\delta) > \frac{T}{5}$, а также $\gamma = 1$ и $\gamma = 2$, что улучшает статистическую и вычислительную эффективность алгоритма. Данный подход к определению размеров клеток и предлагается использовать в настоящей статье. Таким образом, размер клетки для анализа графиков временных рядов вычисляется как $\delta = \frac{W}{2^k}$, где $k = \overline{2, k_{\max}}$, k_{\max} таково, что $N\left(\frac{T}{2^{k_{\max}+1}}\right) > \frac{T}{5}$. Выполнение указанного неравенства и будет являться условием прекращения итераций.

Помимо сложностей с определением диапазона размеров клетки, существует еще один вопрос, остро встающий при адаптации метода

клеточного покрытия для его применения к временным рядам: как выбрать значения W и H – ширину и высоту абстрактного холста, на который помещается график ряда? Поскольку размер клетки δ зависит от W , правильнее говорить не об абсолютных значениях W и H (сами по себе они могут быть произвольными), а об их соотношении, т.е. соотношении сторон холста. Строго говоря, величина W может принимать любое значение (логичным выглядит «привязать» его к ширине окна, которое отображает график временного ряда, измеряемой в пикселях). В свою очередь, величина H зависит от величины W и поэтому должна вычисляться как произведение W и некоторого коэффициента $r = \frac{H}{W}$, причем $r \leq 1$ (чтобы холст не оказался в высоту больше, чем в ширину). От выбора значения этого коэффициента зависит, чему будет равна величина D – оценка фрактальной размерности ряда. Проблема выбора значения коэффициента r рассмотрена в работе [21]. В ней также приведена иллюстрация, демонстрирующая, что при различных соотношениях сторон холста величина $N(\delta)$ различается даже при неизменном размере клетки δ (рис. 5).

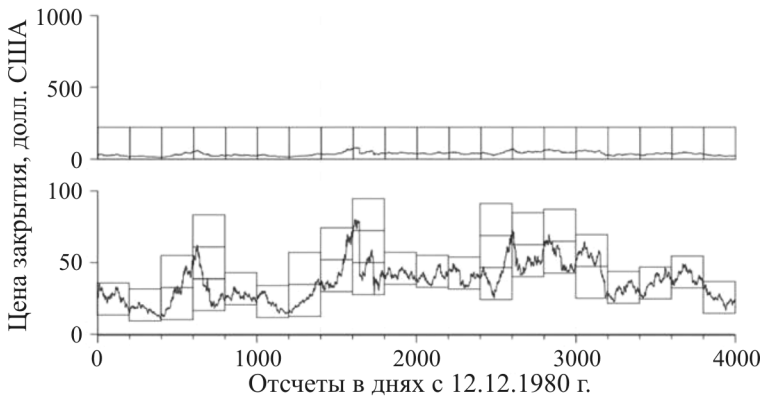


Рис. 5. Графики биржевой цены закрытия, выраженной в долларах США. Количество клеток, покрывающих анализируемый временной ряд, при разных соотношениях сторон холста отличается [21]

Вопрос, каково истинное соотношение W и H , т.е. каково истинное значение r , ответа не имеет. Очевидно, что ширина графика W связана с длиной ряда T , высота H – с размахом $(v_{\max} - v_{\min})$. Однако единицы измерения графика ряда по горизонтали и по вертикали не со-

поставимы друг с другом: мерой ширины графика ряда служит время $t = 1, 2, \dots, T$, единицей измерения W являются, например, торговые дни на бирже; мерой высоты графика ряда служат биржевые показатели (цены, доходности и другие), единицей измерения H являются, например, рубли, если речь идет о ценах, или безразмерная величина, если речь идет о доходностях.

Величину r можно зафиксировать. Например, если $r = 1$, то холст будет квадратным, т.е. иметь соотношение сторон 1:1. Согласие об использовании определенного фиксированного значения r обеспечит сопоставимость результатов при оценке размерности различных временных рядов, однако «абсолютные значения размерности, полученные в результате такого анализа, будут оставаться по существу произвольными» [21]. Ввиду этого Pilgrim, Taylor рекомендуют для анализа временных рядов использовать методы оценки размерности, специально адаптированные для них, в частности метод Херста (R/S-анализ). В то же время, как показал анализ литературы, классический клеточный метод продолжает использоваться для оценки размерности временных рядов, а следовательно, отказаться от его применения (во всяком случае, от сравнения с другими методами), на наш взгляд, неправильно.

Мы предлагаем три способа определения значения коэффициента r :

1. Первый способ состоит в использовании $r = \text{const}$, но не единицы, а, например, 0,7, так как обычно графики временных рядов растянуты по ширине сильнее, чем по высоте.

2. Второй способ заключается в том, чтобы попытаться соотнести «расстояние» между соседними значениями ряда v_t, v_{t+1} по горизонтали с «расстоянием» между ними же по вертикали Δv . Шаг по горизонтали фиксированный и равен единице. В качестве шага по вертикали может быть взята минимальная абсолютная величина изменения между соседними значениями ряда V , а именно: $|\Delta|_{\min} = \min\{|v_t - v_{t-1}|\}$,

$t = \overline{2, T}$, при этом количество шагов вычисляется как $\frac{v_{\max} - v_{\min}}{|\Delta|_{\min}}$. Коэф-

фициент r рассчитывается как $r = \frac{v_{\max} - v_{\min}}{|\Delta|_{\min} T}$. Однако при $|\Delta|_{\min} \rightarrow 0$

коэффициент $r \gg 1$, что приводит к вытянутости графика по вертикали, а при $|\Delta|_{\min} = 0$ коэффициент r вычислить не представляется воз-

возможным. На наш взгляд, для более приближенной к реальности ситуации в качестве шага по вертикали необходимо брать не минимальное, а среднее изменение исследуемой переменной по модулю, т.е.

$$\frac{\sum_{t=2}^T |v_t - v_{t-1}|}{T-1}. \text{ В этом случае } r = \frac{(v_{\max} - v_{\min})(T-1)}{\sum_{t=2}^T |v_t - v_{t-1}| T}.$$

3. Третий способ состоит в том, что коэффициент r определяется как отношение абсолютного минимального значения ряда к размаху:

$$r = \frac{|v_{\min}|}{v_{\max} - v_{\min}}, \text{ т.е. предполагается, что высота холста достигает его}$$

ширины в том случае, если рассматриваемый показатель на протяжении рассматриваемого временного интервала увеличивается вдвое относительно своего минимума.

Наконец, стоит отметить, что величина r может принимать значения как меньше, так и больше единицы. Ввиду этого, чтобы высота холста не превысила его ширину, предлагается вычислять высоту следующим образом:

$$H = \begin{cases} Wr, & r \leq 1; \\ \frac{W}{r}, & r > 1. \end{cases}$$

Далее перейдем к апробации описанной методики нахождения оценки фрактальной размерности с помощью адаптированного к временным рядам метода клеточного покрытия.

3. Апробация разработанной методики

Было решено для программной реализации методики написать функцию на языке программирования Python 3.7.5. Выбор этого высокоуровневого языка обусловлен его простотой, популярностью и существованием большого количества подключаемых библиотек для научных и инженерных расчетов.

На рис. 6 приведены скриншоты окна разработанной программы, содержащего холст с графиком некоторого временного ряда длиной 300 при различных размерах клетки.

На рис. 7 представлен график линейной регрессии для $\ln N(\delta)$, полученный в результате анализа того же ряда.

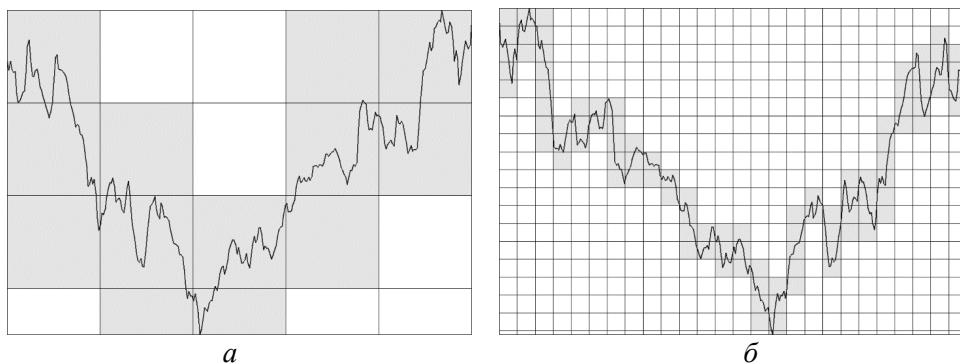


Рис. 6. Визуальная реализация метода клеточного покрытия, где количество фрагментов холста по ширине: $a - \gamma = 5$; $b - \gamma = 26$

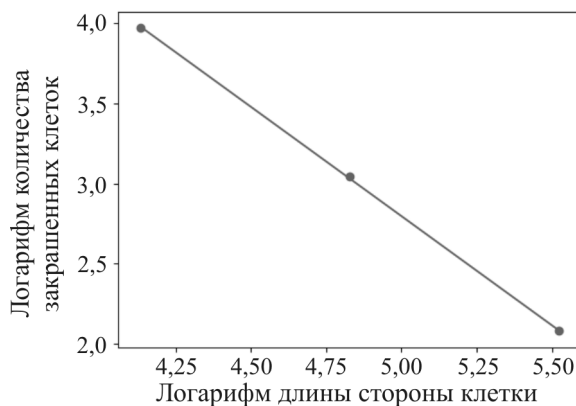


Рис. 7. График линейной регрессии логарифма количества закрасненных клеток от логарифма длины стороны клетки

Несмотря на то, что для написания программы был выбран язык Python, при ее тестировании нами было решено не ограничиваться одним этим инструментом и воспользоваться также возможностями языка R, что связано с наличием в нем некоторых библиотек с подходящими для наших целей готовыми алгоритмами.

Разумным видится в качестве данных взять сгенерированные временные ряды, которые характеризуются различными состояниями и для которых заранее известна величина фрактальной размерности. В зависимости от характера развития случайного процесса временные ряды могут находиться в одном из трех состояний, каждое из которых определяется величиной фрактальной размерности D : 1) тренд ($1 < D < 1,5$); 2) случайное блуждание ($D = 1,5$); 3) флэт ($1,5 < D < 2$) [3]. Также из-

вестно, что размерность является мерой изрезанности или гладкости графика. С нашей точки зрения, имеет смысл проанализировать ряды со следующими характеристиками:

1) ряд, порожденный нестационарным процессом с тенденцией, низкой волатильностью и гладким графиком (его размерность должна стремиться к единице, т.е. $D \rightarrow 1$);

2) ряд, порожденный нестационарным дробно-интегрированным процессом с длинной памятью ($D \rightarrow 1,25$);

3) ряд, порожденный нестационарным процессом «случайное блуждание» ($D \rightarrow 1,5$);

4) ряд, порожденный стационарным дробно-интегрированным процессом ($D \rightarrow 1,75$);

5) ряд, порожденный стационарным процессом с очень высокой волатильностью и изрезанным графиком ($D \rightarrow 2$).

Опишем методику апробации. Чтобы сделать вывод о состоятельности реализованного алгоритма, следует с его помощью оценить размерность сгенерированных с определенными параметрами рядов и сравнить оценку с величиной априорной размерности. Ряды 1 и 5, графики которых представляют собой с некоторыми оговорками прямую и заштрихованную плоскость, можно сформировать с помощью простых формул в единственных вариантах, в то время как ряды 2–4 генерируются случайным образом и при каждой генерации отличаются, поэтому имеет смысл сформировать множество вариантов каждого ряда, оценить размерность каждого из них и вычислить ее математическое ожидание, которое и сравнивать с априорной величиной D . Анализ каждого ряда следует проводить с применением различных способов определения коэффициента r , чтобы посмотреть, как данный параметр влияет на результат. Также было решено сравнить результаты нашего алгоритма адаптированного метода клеточного покрытия с результатами других методов вычисления фрактальных показателей, а именно R/S-анализа (данный метод позволяет вычислить показатель Херста H , и размерность определяется по формуле $D = 2 - H$) и box-counting метода, описанного в работе [7]. Для этого были использованы функции `compute_Hc()` из библиотеки `hurst` (Python) и `fd.estim.boxcount()` из библиотеки `fractaldim` (R). Длина генерируемого ряда – 4096, количество генераций – 5000.

Ряд 1 (обозначим его V_1), представляющий собой гладкую прямую, сформируем с помощью следующей модели: $v_t = v_{t-1} + 0,01$, $v_1 = 0$.

$$\text{Ряд } V_5 \text{ – заштрихованная область: } v_t = \begin{cases} 10 & \text{для } t = 2i - 1; \\ -10 & \text{для } t = 2i, \end{cases} \quad i = \overline{1, 2048}.$$

Ряды V_2, V_3, V_4 нами решено генерировать в виде фрактального броуновского движения (ФБД), которое, с одной стороны, может быть смоделировано с заданной величиной размерности, а с другой стороны, способно правдоподобно описывать реальные процессы на финансовых рынках (например, динамику цен акций) [25]. Для симуляции таких рядов воспользуемся функцией `fbm()` из библиотеки `somebm` (язык R). Входным параметром данной функции является величина H , которая связана с D через уже упомянутое соотношение $H = 2 - D$. Ряд V_2 сгенерирован с $H = 0,75$ ($D = 1,25$); V_4 : $H = 0,25$ ($D = 1,75$); V_3 : $H = 0,5$ ($D = 1,5$). В последнем случае ряд порожден винеровским процессом, являющимся моделью случайного блуждания с непрерывным временем.

На рис. 8 приведены графики тестовых рядов всех видов (D_T означает теоретическую, ожидаемую величину фрактальной размерности D).

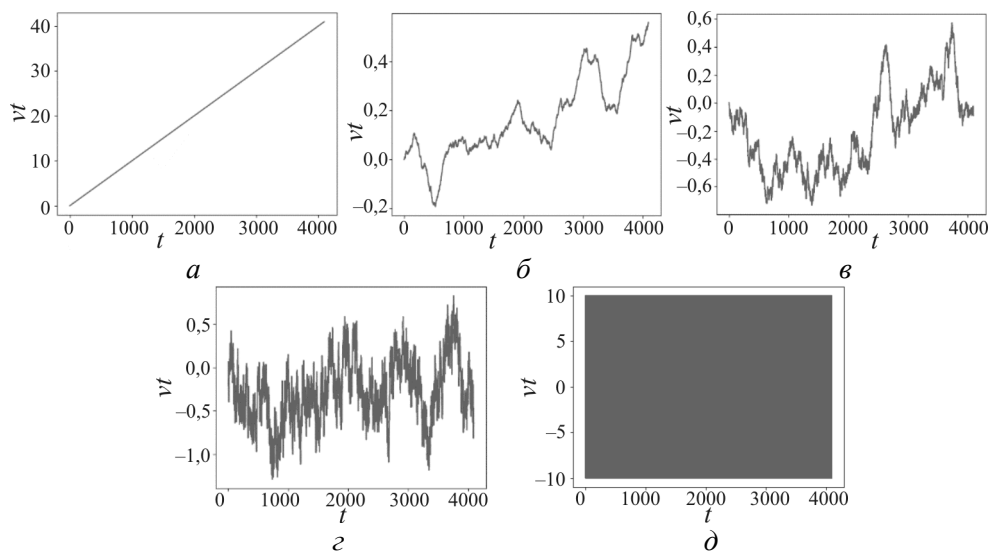


Рис. 8. Графики сгенерированных временных рядов с заданной величиной фрактальной размерности: a – ряд V_1 ($D_T = 1$); $б$ – один из вариантов ряда V_2 ($D_T = 1,25$); $в$ – один из вариантов ряда V_3 ($D_T = 1,5$); $з$ – один из вариантов ряда V_4 ($D_T = 1,75$); $д$ – ряд V_5 ($D_T = 2$)

Также для рядов 2–4 был проведен ADF-тест (расширенный тест Дики – Фуллера на наличие единичного корня), т.е. была выполнена проверка стационарности. Для этого использовалась функция `adf.test()` из библиотеки `tseries (R)` со следующими параметрами: количество лагов – 15, альтернативная гипотеза – ряд стационарен. Тест проводился для каждого из 5000 вариантов рядов, после чего было вычислено математическое ожидание p -значения статистики с 99%-ным доверительным интервалом. Результаты:

$$V_2: p = 0,754 \pm 0,009;$$

$$V_3: p = 0,501 \pm 0,010;$$

$$V_4: p = 0,043 \pm 0,003.$$

В первых двух случаях для отклонения гипотезы о нестационарности оснований недостаточно, поэтому ряд можно считать нестационарным; в последнем случае ряд стационарен.

В следующем подразделе представлены полученные результаты и их анализ.

4. Результаты и их обсуждение

В таблице ниже приведены полученные по итогам проведенных вычислений значения фрактальной размерности.

Оценка фрактальной размерности тестовых рядов различными методами

Ряд	Эмпирически оцененная величина размерности D_E					
	Наша реализация ($r = 0,7$)	Наша реализация ($r = 1$)	Наша реализация (r вычислен способом 2)	Наша реализация (r вычислен способом 3)	R/S-анализ (готовая функция)	Box-counting метод (готовая функция)
V_1	1,024	1,028	1,028	—**	0,920	1,036
V_2	$1,187 \pm 0,002^*$	$1,211 \pm 0,002^*$	$1,108 \pm 0,001^*$	—**	$1,243 \pm 0,001^*$	$1,211 \pm 0,002^*$
V_3	$1,382 \pm 0,002^*$	$1,414 \pm 0,002^*$	$1,096 \pm 0,000^*$	—**	$1,450 \pm 0,001^*$	$1,415 \pm 0,002^*$
V_4	$1,590 \pm 0,002^*$	$1,623 \pm 0,002^*$	$1,062 \pm 0,000^*$	$1,546 \pm 0,003^*$	$1,658 \pm 0,001^*$	$1,624 \pm 0,002^*$
V_5	1,982	2,000	—**	2,000	1,911	2,000

Примечание: *Для рядов, сгенерированных случайным образом по 5000 раз, приведено математическое ожидание размерности с 99%-ным доверительным интервалом.

**При некоторых вариантах рядов вычисление D_E оказалось невозможным.

Можно констатировать, что факт чувствительности метода клеточного покрытия к входным параметрам в ходе экспериментов подтвердился. Коэффициент r , величиной которого определяется соот-

ношение сторон холста, достаточно сильно влияет на полученную оценку размерности: чем его значение выше (т.е. холст более вытянут по высоте), тем больше оказывается D_E . Очевидно, это связано с увеличением количества клеток на плоскости и, как следствие, ростом числа закрашенных клеток $N(\delta)$ при неизменном размере клетки δ , что, в свою очередь, приводит к увеличению угла наклона прямой регрессии в логарифмических координатах. Это обусловлено спецификой методики: клетки квадратные и величина δ зависит только от ширины холста W , но не от его высоты.

Также можно утверждать, что на сгенерированных данных предложенные нами способы автоматического расчета r (способы определения r под номерами 2 и 3) оказались неприменимы. Оба из них базируются на попытке соотнести между собой несопоставимые величины: длину ряда и размах его значений. Способ 2, основанный на сопоставлении длины ряда и числа вхождений в размах среднего абсолютного прироста показателя, в результате дал очень низкий коэффициент высоты холста, так как при большой выборке (4096 в нашем случае) величина Δv оказывается чрезвычайно мала, а низкое значение r приводит к $D \rightarrow 1$ вне зависимости от характера анализируемого ряда. В некоторых же случаях, как можно заметить на примере ряда V_5 , вычисление D с таким подходом к определению коэффициента высоты холста и вовсе невозможно: в каком-то из вариантов ряда значение r получилось настолько малым, что H оказалась меньше единицы, а так как в нашей программной реализации безразмерная величина H соответствует высоте холста в пикселях, т.е. целому числу, то высота составила 0, что и привело к невозможности продолжения работы алгоритма. Что касается способа 3, согласно которому r определяется по абсолютному минимальному значению ряда, отнесенному к размаху, он в нашем случае также оказался несостоятельным по той же причине, что и предыдущий способ: v_{\min} в каких-то из вариантов рядов принимало значения, близкие к нулю, что приводило к $H = 0$ и невозможности вычисления D_E . Кроме того, при данном подходе r зависит не только от размаха значений ряда, но от минимального значения: если к каждому члену ряда прибавить некоторую константу, это не изменит форму графика и вероятностные характеристики ряда, но повлияет на оценку D , что является некорректным. Использование рассмотренных

подходов, несостоятельных в случае анализа сгенерированных рядов, может иметь смысл при исследовании рядов, отражающих реальные экономические процессы (динамику курсов акций, валютных пар и так далее), так как при разработке этих подходов предполагалось, что характер изменения значений анализируемых переменных (процессов и явлений) подчиняется природным закономерностям. Например, сложно себе представить, чтобы цена акции или биржевой индекс многократно увеличивались относительно своего минимума на рассматриваемом временном интервале при условии, что процесс на этом интервале существенно не меняет законов своего развития, т.е. исследователем производится локальный анализ процесса (рассмотрен в работе [4]). Вопросы применимости методов фрактального анализа к рядам различной длины, отражающим динамику процесса на разных временных интервалах, требуют отдельного исследования, а пока можно сделать вывод о неприменимости предложенных способов определения соотношения сторон холста при анализе сгенерированных рядов и необходимости их тестирования на реальных экономических рядах.

Из двух вариантов r меньшую разницу между D_E и D_T в четырех случаях из пяти показало значение $r=1$, уступив лишь при анализе ряда V_1 , где графиком является прямая линия. Поскольку ряды 2–4, несомненно, являются более приближенными к реальным процессам, можно заключить, что квадратная форма холста позволяет получать более точные оценки размерности.

Наиболее точные результаты были получены с помощью R/S-анализа – метода, специально разработанного для оценки фрактальных характеристик хаотических рядов с длинной памятью. Значения D_E , вычисленные на основе этого метода, оказались наиболее близки к D_T , хотя и не достигли их с высокой точностью: абсолютная разница между математическим ожиданием D_E и D_T составила от 0,007 до 0,092. В случае ряда V_1 величина $D_E = 0,92$ оказалась некорректной (плоская фигура не может иметь размерность меньше единицы, как и значение показателя Херста H не должно выходить за границы диапазона $[0;1]$), но это можно объяснить весьма специфическим характером ряда: на практике такие временные ряды, где графиком является идеальная прямая, вряд ли существуют. Следует заметить, что для генерации

фрактального броуновского движения использовалась функция в R, а для оценки H была применена функция из библиотеки Python. Эта же библиотека имеет функцию для генерации ФБД, и возможно, что оценки H сгенерированных с ее помощью рядов оказались бы ближе к величинам, указанным при генерации, однако нами сознательно было решено для большей объективности исследования использовать программные библиотеки разных авторов. Другой вывод, который вытекает из того, что данные для исследования сформированы случайным образом, но с явным указанием вида распределения и параметра H , состоит в следующем. Даже в «идеальных» условиях (т.е. в условиях искусственного характера исследуемых временных рядов) оценивание фрактального показателя с высокой точностью является неординарной задачей. При анализе рядов, отражающих реальные процессы (т.е. в «естественных» условиях), сложность этой задачи, несомненно, повышается в связи с неизвестностью истинного значения H этих рядов. Также стоит отметить, что методика R/S-анализа требует для достаточно надежного определения H нескольких тысяч наблюдений [4]. В нашем случае использовались ряды из 4096 значений. При исследовании более коротких рядов точность R/S-анализа относительно точности клеточного метода может снизиться, однако этот момент требует провести сравнение нескольких методов в разных условиях.

Наконец, реализация метода клеточного покрытия из библиотеки `fractaldim (R)` показала значения D_E , отличающиеся от вычисленных нашим алгоритмом с $r = 1$ на уровне погрешности. Особенностью реализации метода клеточного покрытия в R является то, что по умолчанию для анализа используют только первые $2^{k_{\max}} + 1$ значений входного ряда, остальные отбрасывают. В эксперименте мы использовали ряды длиной $2^{12} = 4096$, а в действительности сгенерировано было по 4097 значений и ряды именно такой длины подавались на вход функции в R. Если бы количество значений в ряду было меньше хотя бы на единицу, тогда величина D_E оценивалась бы по 2049 наблюдениям, т.е. практически по половине всего ряда. По нашему мнению, отбрасывание такого объема данных является некорректным, может привести к искаженным оценкам фрактальных показателей, поэтому в предложенной нами методике используются все значения ряда вне зависимости от того, является длина ряда степенью двойки или нет.

Таким образом, разработанный алгоритм метода клеточного покрытия, адаптированный к временным рядам, корректно воплощает в себе идею метода оценивания фрактальной размерности; его программная реализация на языке Python может быть использована при анализе финансовых временных рядов. Проведенные эксперименты показали, что получаемые с помощью нашей реализации оценки D не уступают в точности оценкам, полученным с помощью другой существующей в R реализации метода. Однако наша программа уступает программе в R по вычислительной эффективности, что выражено в большей длительности ее выполнения, поэтому требуется ее оптимизация.

Заключение

В настоящее время моделирование и прогнозирование финансовых временных рядов является одним из наиболее актуальных и необходимых направлений исследований, так как позволяет участникам, которых непосредственно затрагивают финансовые процессы, определять благоприятные исходы и выработать перспективные решения. В качестве инструментов, помогающих выработать данные решения, можно использовать методы фрактального анализа временных рядов. Один из таких методов – метод клеточного покрытия – нами был изучен и детально проработан. В результате была разработана методика, адаптирующая метод клеточного покрытия для оценки размерности временных рядов без формирования компьютерного изображения графика ряда, рассмотрены возникающие при такой адаптации трудности, предложены способы их разрешения. Также была разработана программа для использования данной методики на практике и произведена ее апробация на сгенерированных данных. По итогам проведенной работы можно сделать следующие основные выводы:

1. Метод клеточного покрытия для фрактального анализа финансовых временных рядов достаточно часто используется в исследованиях, проводимых отечественными авторами. Однако последними не уделяется должного внимания особенностям и проблемам применения метода к временным рядам.

2. В то же время эти проблемы являются достаточно важными. Во-первых, метод обладает высокой чувствительностью к регулирующим параметрам, а именно: соотношению сторон плоскости с графиком, покрываемой клетками; используемому диапазону значений дли-

ны сторон клетки; количеству разбиений на клетки плоскости. Во-вторых, выбор наиболее подходящих значений этих параметров неочевиден. Наибольшие трудности вызывает определение корректного соотношения сторон, потому что оно требует сопоставления несоизмеримых величин. Результаты проведенного эксперимента свидетельствуют, что наиболее точными являются оценки размерности при соотношении сторон 1:1, т.е. квадратной форме плоскости. При этом предложенные нами способы автоматического определения соотношения сторон плоскости применительно к сгенерированным рядам показали себя несостоятельными.

3. R/S-анализ лишен вышеупомянутых проблем, так как разработан непосредственно для анализа временных рядов. В нашем эксперименте он позволил получить более точные оценки размерности. Однако у этого метода есть ограничение, которое заключается в необходимости исследования выборок большого объема. Таким образом, утверждения о его более высокой эффективности во всех случаях преждевременны, необходимо выполнить сравнение точности различных методов в тех или иных условиях, а именно – при анализе временных рядов различной длины.

Итак, в ходе изучения метода клеточного покрытия для оценивания фрактальной размерности пространственных структур была разработана методика, адаптирующая этот метод к временным рядам. По результатам проведенного исследования видно, что разработанный метод на сгенерированных рядах уступает по точности оценивания фрактальных характеристик широко известному методу в области фрактального анализа данных – R/S-анализу, однако не уступает реализации метода клеточного покрытия из библиотеки `fractaldim` языка R. Вполне закономерными дальнейшими шагами в области исследований фрактального анализа временных рядов могут являться следующие: выполнить сравнение точности различных методов оценивания фрактальной размерности на сгенерированных и реальных временных рядах различной длины; осуществить сравнение методов фрактального анализа и p -адического анализа при моделировании и прогнозировании финансовых временных рядов; определить условия применимости различных методов; провести апробацию разработанных способов определения соотношения сторон плоскости с графиком на реальных экономических данных.

Благодарность

Статья посвящается светлой памяти доктора физико-математических наук, профессора Петра Михайловича Симонова [(1 января 1957 – 18 июля 2020) – примеч. отв. секр.], настоящего Ученого, Наставника и Друга, чьи ценные замечания и советы помогли в ее написании.

Список литературы

1. Мандельброт Б. Фрактальная геометрия природы / Ин-т комп. ис-след. – М., 2002. – 656 с.
2. Петерс Э. Фрактальный анализ финансовых рынков: применение теории хаоса в инвестициях и экономике. – М.: Интернет-трейдинг, 2004. – 304 с.
3. Симонов П.М., Гарафутдинов Р.В. Моделирование и прогнозирование динамики курсов финансовых инструментов с применением эконометрических моделей и фрактального анализа // Вестник Пермского университета. Серия «Экономика» = Perm University Herald. ECONOMY. – 2019. – Т. 14, № 2. – С. 268–288. DOI: 10.17072/1994-9960-2019-2-268-288
4. Старченко Н.В. Индекс фрактальности и локальный анализ хаотических временных рядов: дис. ... канд. физ.-мат. наук: 05.13.18, 01.01.03. – М., 2005. – 122 с.
5. Антонова И.В., Чикина Н.А. Применение методов фрактального анализа к исследованию временных рядов // Вестник Национального технического университета «Харьковский политехнический институт». Информатика и моделирование. – 2015. – № 32 (1141). – С. 4–10.
6. Балагула Ю.М. Фрактальные характеристики длинной памяти в ценах на электроэнергию // Финансы и бизнес. – 2017. – № 1. – С. 104–113.
7. Gneiting T., Ševčíková H., Percival D.B. Estimators of fractal dimension: assessing the roughness of time series and spatial data // Statistical Science. – 2012. – Vol. 27, no. 2. – P. 247–277. DOI: 10.1214/11-STS370
8. Водолазский А.А. Начала экономофизики и количественная определенность первых экономических законов. – Новочеркасск: НОК, 2013. – 227 с.
9. Андрианов Д.Л., Симонов П.М. Обзор методов экономико-математического моделирования, основанных на принципах экономофизики. Часть I // Прикладная математика и вопросы управления. – 2020. – № 1. – С. 161–181. DOI: 10.15593/2499-9873/2020.1.10
10. Garafutdinov R., Gurova E. The formation of investment portfolios based on forecasted income with the use of fractal models // Mathematical Modeling. – 2019. – Vol. 3, iss. 3. – P. 86–88.
11. Гарафутдинов Р.В. К задаче классификации участков финансовых временных рядов различной фрактальной размерности // Татищевские чте-

ния: актуальные проблемы науки и практики: материалы XVII Междунар. науч.-практ. конф., г. Тольятти, 24–25 апреля 2020 г.: в 3 т. Т. 1. Гуманитарные и социальные науки, образование. Актуальные проблемы социально-экономического развития / Волж. ун-т им. В.Н. Татищева. – Тольятти: Изд-во ВУиТ, 2020. – С. 226–230.

12. Симонов П.М., Филимонова С.А. *P*-адическое моделирование динамики индекса РТС в зависимости от таймфреймов // Вестник Пермского университета. Серия «Экономика» = Perm University Herald. ECONOMY. – 2016. – № 4(31). – С. 74–85. DOI: 10.17072/1994-9960-2016-4-74-85

13. Симонов П.М., Ахуньянова С.А. Сравнительный анализ методик AR-GARCH и *p*-адического прогнозирования волатильности финансового рынка // Вестник Пермского университета. Серия «Экономика» = Perm University Herald. ECONOMY. – 2019. – Т. 14, № 1. – С. 69–92. DOI: 10.17072/1994-9960-2019-1-69-92

14. Кривоносова Е.К., Первадчук В.П. Использование фрактального подхода для анализа стабильности многоуровневых структур // Вестник Пермского национального исследовательского политехнического университета. Машиностроение, материаловедение. – 2013. – Т. 15, № 1. – С. 63–69.

15. Балагула Ю.М., Абакумова Ю.А. Длинная память на рынке нефти: спектральный подход. – СПб.: Изд-во ЕУСПб., 2011. – 40 с. (Препринт Ес-01/11.)

16. Кривоносова Е.К. Разработка методов прогнозирования и анализа кредитных и инвестиционных рисков с применением фрактальных и мультифрактальных характеристик: дис. ... канд. экон. наук: 08.00.13. – Пермь, 2015. – 167 с.

17. Гарафутдинов Р.В. Обзор методов оценивания фрактальных характеристик финансовых временных рядов // Аналитические и численные методы моделирования естественно-научных и социальных проблем: материалы XIII Междунар. науч.-техн. конф., г. Пенза, 4–6 декабря 2018 г. / Пенз. гос. ун-т. – Пенза: Изд-во ПГУ, 2018. – С. 97–103.

18. Аптуков В.Н., Митин В.Ю. Фрактальный анализ метеорологических рядов с помощью метода минимального покрытия // Географический вестник. – 2019. – № 2 (49). – С. 67–79. DOI: 10.17072/2079-7877-2019-2-67-79

19. Белолицев И.И., Фархиева С.А. Предсказание финансовых временных рядов на основе индекса фрактальности [Электронный ресурс] // Мир Науки. – 2014. – Вып. 3. – URL: <https://mir-nauki.com/PDF/01EMN314.pdf> (дата обращения: 10.05.2020).

20. Сизов А.А. Модели, способы и программные средства поддержки принятия решений на основе прогнозирования временных рядов с переменной структурой: дис. ... канд. техн. наук: 05.13.17. – Смоленск, 2014. – 139 с.

21. Pilgrim I., Taylor R. Fractal analysis of time-series data sets: methods and challenges. Fractal analysis. – UK: InTech, 2018. – Chapter 2. – 26 p. DOI:10.5772/intechopen.81958
22. Kenkel N.C. Sample size requirements for fractal dimension estimation // Community Ecology. – 2013. – Vol. 14. – P. 144–152. DOI: 10.1556/ComEc.14.2013.2.4
23. Rosenberg E. Minimal box size for fractal dimension estimation // Community Ecology. – 2016. – Vol. 17. – P. 24–27. DOI: 10.1556/168.2016.17.1.4
24. Андреева Е.В., Егоров Ю.Е. Вычислительная геометрия на плоскости // Информатика. – 2002. – № 40. – С. 28–31.
25. Чичаев И.А., Попов В.Ю. Об одном подходе к вычислению индекса Херста финансовых временных рядов и их аппроксимации фрактальным броуновским движением [Электронный ресурс] // Современные проблемы науки и образования. – 2013. – № 2. – URL: <http://www.science-education.ru/ru/article/view?id=8698> (дата обращения: 10.05.2020).

References

1. Mandelbrot B. The Fractal Geometry of Nature. Henry Holt and Company, 1983, 468 p.
2. Peters E. Fractal Market Analysis: Applying Chaos Theory to Investment and Economics. Wiley, 1994, 336 p.
3. Simonov P.M., Garafutdinov R.V. Modelirovanie i prognozirovanie dinamiki kursov finansovykh instrumentov s primeneniem ekonometricheskikh modelei i fraktal'nogo analiza [Modeling and forecasting of financial instruments dynamics using econometrics models and fractal analysis]. *Perm University Herald. ECONOMY*, 2019, vol. 14, no. 2, pp. 268–288. DOI: 10.17072/1994-9960-2019-2-268-288 (in Russian).
4. Starchenko N.V. Indeks fraktal'nosti i lokal'nyi analiz khaoticheskikh vremennykh riadov [Fractality index and local analysis of chaotic time series]. Ph. D. thesis. Moscow, 2005, 122 p. (in Russian).
5. Antonova I.V., Chikina N.A. Primenenie metodov fraktal'nogo analiza k issledovaniyu vremennykh riadov [Application of fractal analysis methods to the study of time series]. *Vestnik Natsional'nogo tekhnicheskogo universiteta «Khar'kovskii politekhnicheskii institut»*. *Informatika i modelirovanie*, 2015, no. 32 (1141), pp. 4–10 (in Russian).
6. Balagula Iu.M. Fraktal'nye kharakteristiki dlinnoi pamiaty v tsenakh na elektroenergii [Fractal characteristics of long memory in electricity prices]. *Finansy i biznes*, 2017, no. 1, pp. 104–113 (in Russian).
7. Gneiting T., Ševčíková H., Percival D.B. Estimators of Fractal Dimension: Assessing the Roughness of Time Series and Spatial Data. *Statistical Science*, 2012, vol. 27, no. 2, pp. 247–277. DOI: 10.1214/11-STS370

8. Vodolazskii A.A. Nachala ekonofiziki i kolichestvennaia opredelennost' pervykh ekonomicheskikh zakonov [The beginnings of econophysics and the quantitative certainty of the first economic laws]. Novocherkassk, NOK, 2013, 227 p. (In Russian).

9. Andrianov D.L., Simonov P.M. Obzor metodov ekonomiko-matematicheskogo modelirovaniia, osnovannykh na printsipakh ekonofiziki. Chast' I [A review of the methods of economic and mathematical modeling based on the principles of econophysics. Part I]. *Applied Mathematics and Control Sciences*, 2020, no. 1, pp. 161–181. DOI: 10.15593/2499-9873/2020.1.10 (in Russian).

10. Garafutdinov R., Gurova E. The formation of investment portfolios based on forecasted income with the use of fractal models. *Mathematical Modeling*, 2019, vol. 3, iss. 3, pp. 86–88.

11. Garafutdinov R.V. K zadache klassifikatsii uchastkov finansovykh vremennykh riadov razlichnoi fraktal'noi razmernosti [To the task of classification of sections of financial time series of various fractal dimensions]. *Proceedings of the 17th International Scientific and Practical Conference «Tatishchevskie chteniia: aktual'nye problemy nauki i praktiki» in 3 volumes. Vol. 1. Gumanitarnye i sotsial'nye nauki, obrazovanie. Aktual'nye problemy sotsial'no-ekonomicheskogo razvitiia, 24–25 April 2020*, Tol'iatti, Volzhskii universitet imeni V.N. Tatishcheva, 2020, pp. 226–230 (in Russian).

12. Simonov P.M., Filimonova S.A. P -adicheskoe modelirovanie dinamiki indeksa RTS v zavisimosti ot taimfreimov [P -adic modeling of the RTS index dynamics depending on the timeframes]. *Perm University Herald. ECONOMY*, 2016, no. 4(31), pp. 74–85. DOI: 10.17072/1994-9960-2016-4-74-85 (in Russian).

13. Simonov P.M., Akhunyanova S.A. Sravnitel'nyi analiz metodik $AR-GARCH$ i p -adicheskogo prognozirovaniia volatil'nosti finansovogo rynka [Comparative analysis of $AR-GARCH$ and p -adic methods of the prediction of the financial market volatility]. *Perm University Herald. ECONOMY*, 2019, vol. 14, no. 1, pp. 69–92. DOI: 10.17072/1994-9960-2019-1-69-92 (in Russian).

14. Krivonosova E.K., Pervadchuk V.P. Ispol'zovanie fraktal'nogo podkhoda dlia analiza stabil'nosti mnogourovnevnykh struktur [Using the fractal approach for analysis of the stability of multilevel structures]. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Mashinostroenie, materialovedenie*, 2013, no. 1, pp. 63–69 (in Russian).

15. Balagula Iu.M., Abakumova Iu.A. Dlinnaia pamiat' na rynke nefti: spektral'nyi podkhod. Preprint Es-01/11 [Long memory on the oil market: spectral approach. Preprint Es-01/11]. Saint Petersburg, EUSP, 2011, 40 p. (in Russian).

16. Krivonosova E.K. Razrabotka metodov prognozirovaniia i analiza kreditnykh i investitsionnykh riskov s primeneniem fraktal'nykh i mul'tifraktal'nykh kharakteristik [Development of methods for prediction and analysis the

credit and the investment risks using fractal and multifractal characteristics]. Ph.D. thesis, Perm, 2015, 167 p. (in Russian).

17. Garafutdinov R.V. Obzor metodov otsenivaniia fraktal'nykh kharakteristik finansovykh vremennykh riadov [Overview of methods for assessing the fractal characteristics of financial time series]. *Proceedings of the 13th International Scientific and Practical Conference «Analiticheskie i chislennye metody modelirovaniia estestvenno-nauchnykh i sotsial'nykh problem», 4–6 December 2018*, Penza, Penzenskii gosudarstvennyi universitet, 2018, pp. 97–103 (in Russian).

18. Aptukov V.N., Mitin V.Yu. Fraktal'nyi analiz meteorologicheskikh riadov s pomoshch'iu metoda minimal'nogo pokrytiia [Fractal analysis of meteorological series based on the minimal covering method]. *Geographical bulletin*, 2019, no. 2(49), pp. 67–79. DOI: 10.17072/2079-7877-2019-2-67-79 (in Russian).

19. Beloliptsev I.I., Farkhieva S.A. Predskazanie finansovykh vremennykh riadov na osnove indeksa fraktal'nosti [Prediction of financial time series based on fractality index]. *Mir Nauki*, 2014, iss. 3, available at: <https://mir-nauki.com/PDF/01EMN314.pdf> (accessed 10 May 2020) (in Russian).

20. Sizov A.A. Modeli, sposoby i programmnye sredstva podderzhki priniatiia reshenii na osnove prognozirovaniia vremennykh riadov s peremennoi strukturoi [Models, methods and software tools for decision support based on prediction of time series with variable structure]. Ph.D. thesis, Smolensk, 2014, 139 p. (in Russian).

21. Pilgrim I., Taylor R. Fractal Analysis of Time-Series Data Sets: Methods and Challenges. *Fractal Analysis*. InTech, UK, 2018, Chapter 2, 26 p. DOI:10.5772/intechopen.81958

22. Kenkel N.C. Sample size requirements for fractal dimension estimation. *Community Ecology*, 2013, vol. 14, pp. 144–152. DOI: 10.1556/ComEc.14.2013.2.4

23. Rosenberg E. Minimal box size for fractal dimension estimation. *Community Ecology*, 2016, vol. 17, pp. 24–27. DOI: 10.1556/168.2016.17.1.4

24. Andreeva E.V., Egorov Iu.E. Vychislitel'naia geometriia na ploskosti [Computational geometry on a plane]. *Informatika*, 2002, no. 40, pp. 28–31 (in Russian).

25. Chichaev I.A., Popov V.Iu. Ob odnom podkhode k vychisleniiu indeksa Khersta finansovykh vremennykh riadov i ikh approksimatsii fraktal'nym brounovskim dvizheniem [About one approach to calculating the Hurst index of financial time series and their approximation by the fractal Brownian motion]. *Sovremennye problemy nauki i obrazovaniia*, 2013, no. 2, available at: <http://www.science-education.ru/ru/article/view?id=8698> (accessed 10 May 2020) (in Russian).

Статья получена: 03.06.2020

Статья принята: 01.09.2020

Сведения об авторах

Гарафутдинов Роберт Викторович (Пермь, Россия) – аспирант кафедры «Информационные системы и математические методы в экономике», Пермский государственный национальный исследовательский университет (614990, Пермь, ул. Букирева, 15; e-mail: rvgarafutdinov@gmail.com).

Ахуньянова Софья Алексеевна (Пермь, Россия) – аспирант кафедры «Информационные системы и математические методы в экономике», Пермский государственный национальный исследовательский университет (614990, Пермь, ул. Букирева, 15; e-mail: sofi_perm@mail.ru).

About the authors

Robert V. Garafutdinov (Perm, Russian Federation) – Postgraduate Student, Department of Information Systems and Mathematical Methods in Economics, Perm State University (614990, Perm, Bukireva st., 15, e-mail: rvgarafutdinov@gmail.com).

Sofya A. Akhunyanova (Perm, Russian Federation) – Postgraduate Student, Department of Information Systems and Mathematical Methods in Economics, Perm State University (614990, Perm, Bukireva st., 15, e-mail: sofi_perm@mail.ru).

Библиографическое описание статьи согласно ГОСТ Р 7.0.100–2018:

Гарафутдинов, Р.В. Адаптированный метод клеточного покрытия для оценивания фрактальной размерности финансовых временных рядов / Р. В. Гарафутдинов, С. А. Ахуньянова. – DOI 10.15593/2499-9873/2020.3.10. – Текст : непосредственный // Прикладная математика и вопросы управления = Applied Mathematics and Control Sciences. – 2020. – № 3. – С. 185–218.

Цитирование статьи в изданиях РИНЦ:

Гарафутдинов Р.В., Ахуньянова С.А. Адаптированный метод клеточного покрытия для оценивания фрактальной размерности финансовых временных рядов // Прикладная математика и вопросы управления. – 2020. – № 3. – С. 185–218. DOI: 10.15593/2499-9873/2020.3.10

Цитирование статьи в references и международных изданиях:

Cite this article as:

Garafutdinov R.V., Akhunyanova S.A. Adapted box-counting method for assessment of the fractal dimension of financial time series. *Applied Mathematics and Control Sciences*, 2020, no. 3, pp. 185–218. DOI: 10.15593/2499-9873/2020.3.10 (in Russian)