

УДК: 51.77

## Кластеризация по времени крупных падений фондовых индексов

А. М. Казарян<sup>1,a</sup>, А. Б. Шаповал<sup>1,2,b</sup>

<sup>1</sup>Финансовый университет при правительстве РФ,  
Россия, 125993, г. Москва, Ленинградский пр-т, д. 49

<sup>2</sup>Институт теории прогноза землетрясений и математической геофизики РАН,  
Россия, 117997, г. Москва, ул. Профсоюзная, д. 84/32

E-mail: <sup>a</sup> annakaz90@mail.ru, <sup>b</sup> shapoval@mccme.ru

Получено 25 мая 2012 г.

В статье оценивается повторяемость падений фондовых индексов S&P100, CAC40, DAX, FTSE, AMEX, ATX, NASDAQ, BEL20. Введена количественная мера повторяемости, основанная на ошибках первого и второго рода. Установлено, что за первую четверть времени между падениями происходит в среднем более трех четвертей всех падений. Этот результат распространяется с достаточно крупных падений, которые фиксируются в среднем два раза в год, на меньшие падения, наблюдаемые в среднем один раз в 1.5–2 месяца.

Ключевые слова: распределение времени между событиями, ошибки первого и второго рода

### Time-clustering of stock indices' big fall

A. M. Kazaryan<sup>1</sup>, A. B. Shapoval<sup>1,2</sup>

<sup>1</sup>Financial University under the Government of the Russian Federation, 49 Leningradsky Prospekt, Moscow, 125993, Russia

<sup>2</sup>Institute of Earthquake Prediction Theory and Mathematical Geophysics, 84/32 Profsoyuznaya str., Moscow, 117997, Russia

**Abstract.** — The paper estimates the recurrence rate of stock indices S&P100, CAC40, DAX, FTSE, AMEX, ATX, NASDAQ, BEL20. The introduced quantitative measure of the recurrence rate underlies type I and type II errors. We show that more than three quarters of the indices' falls occur on average during the first quarter of the time between them. This result expands from sufficiently large falls, which are observed on average two times a year, over smaller falls, which occur approximately once 1.5–2 months.

Keywords: inter-event distribution, errors of the first and second kind

Citation: *Computer Research and Modeling*, 2012, vol. 4, no. 3, pp. 631–638 (Russian).

Работа выполнена в рамках бюджетной НИР в Финансовом университете при Правительстве РФ при поддержке грантов РФФИ 11-01-00339-а, 11-01-00887-а, 11-06-00278.

## Введение

Хорошо известны статистические закономерности (stylized facts), наблюдаемые у финансовых временных рядов [Mantegna, Stanley, 2000; Romanovsky, Vidov, 2011]. Видимо, самыми известными являются «тяжелые хвосты» распределений логарифмических доходностей и кластеризация волатильности [Романовский, Романовский, 2007]. Второе свойство означает, что амплитуда колебаний цен меняется достаточно медленно: имеют место интервалы как относительно малых, так и относительно больших колебаний цен. В частности, после значительного падения цены финансового инструмента часто наблюдаются значительные колебания, в том числе последующие падения.

Эффект повторяемости экстремальных событий зафиксирован при анализе последовательности сильных землетрясений [Keilis-Borok, 2003], изменения уровня преступности [Кузнецов и др., 2006], распространения компьютерного спама [Ciamarra, Coniglio, 2008], колебания интернет-трафика [Abe, Suzuki, 2002], эволюции валютных курсов [Siokis, 2012]. Видимо, это следствие общего закона эволюции сложных систем, установленного на модельном уровне [Blanter et al., 1997].

Количественное описание повторяемости экстремальных событий дается в терминах распределения  $\xi$  времени между ними. При «отсутствии памяти» у этих событий распределение  $\xi$  являлось бы показательным. Для различных финансовых показателей, однако, это распределение лучше аппроксимируется степенной функцией (на основной части области определения), чем экспоненциальной [Bogachev, Bunde, 2008; Lee et al., 2006]. Этот факт, впрочем, напрямую соответствует существованию *аномально больших* интервалов между крупными падениями цен. Область малых интервалов между падениями обычно игнорируется при аппроксимациях.

Более сложное описание последствий крупных падений связано с попыткой распространения так называемого закона Омори [Omori, 1894] на финансовые ряды. Закон Омори, адаптированный к финансовым рядам, утверждает, что непосредственно после краха количество падений временного ряда в скользящем временном окне убывает как степенная функция времени. Универсальность этого вывода требует по крайней мере определенной проверки [Mu, Zhou, 2008; Lillo, Mantegna, 2003; Selcuk, 2004; Weber et al., 2007]. Кроме того, из закона Омори не следуют ни ожидаемые направления трендов, ни оценка амплитуды последующих колебаний.

Наличие определенной памяти у последовательности финансовых крахов и существование более тонких закономерностей финансовых временных рядов являются вызовом для специалистов в теории прогноза экстремальных событий. Развитые методы, прогнозирующие время наступления крахов, несомненно, эффективнее прогнозирования наугад [Sornette, Zhou, 2006]. Обоснование эффективности проводится в терминах ошибок первого и второго рода, адаптированных Молчаном [Молчан, 1991] к оценке прогноза экстремальных событий временных рядов.

В статье [Sharoval, 2010], с одной стороны, обсуждается адекватность сравнения эффективности алгоритмов, прогнозирующих финансовые крахи с результатом случайного прогноза, с другой стороны, оценивается повторяемость крупных падений индексов Доу Джонс (DJIA) и Ханг Сенг (HSI). Обе задачи решаются с помощью алгоритма, ожидающего наступление следующего крупного падения в течение фиксированного времени после предыдущего. Эффективность этого алгоритма можно считать мерой повторяемости крахов.

В настоящей работе развивается подход статьи [Sharoval, 2010]. Оценка повторяемости крахов устанавливается для индексов S&P100, CAC40, DAX, FTSE, AMEX, ATX, NASDAQ, BEL20. Показано, что оценка повторяемости устойчива относительно величины исследуемых падений, времени, характеризующего последствие падения и промежутка между данными временного ряда.

## Данные и метод

### Фондовые индексы

В работе исследуются наиболее значимые мировые фондовые индексы: американские NASDAQ (за 1985–2011 годы), AMEX (за 1995–2011 годы), S&P100 (за 1982–2011 годы); европейские CAC40 (за 1990–2011 годы), ATX (за 1992–2011 годы), BEL20 (за 2005–2011 годы), DAX (за 1990–2011 годы), FTSE (за 1984–2011 годы). Ежедневные данные доступны на сайте [finance.yahoo.com](http://finance.yahoo.com).

### Объект исследования

Пусть  $P(t_1), P(t_2), \dots$  — цены закрытия фондового индекса в моменты времени  $t_1, t_2, \dots$ . Интервал времени между двумя последовательными моментами времени равен  $G$  дней:  $t_{i+1} - t_i = G$ . Через  $r(t) = \ln p(t)/p(t - G)$  обозначается логарифмическая доходность индекса на промежутке  $[t - G, t]$ . На самом деле  $r(t) = r_G(t)$  зависит от  $G$ . Однако индекс  $G$  будет опущен, когда это не приводит к неопределенности. Падение индекса на промежутке  $[t - G, t]$  называется крупным, если  $r(t) < r^*$ , где  $r^*$  — некоторый, по смыслу задачи, отрицательный параметр. При вычислениях это падение ставится в соответствие моменту времени  $t$ . Пусть  $S$  — выборка крупных падений. Исследуется вопрос кластеризации по времени крупных падений выборки  $S$  в зависимости от параметров  $r^*$  и  $G$ .

### Количественная интерпретация повторяемости

Повторяемость крупных падений означает, что в течение определенного времени  $T$  после крупного падения произойдет новое крупное падение. Определяется алгоритм  $A$ , который ожидает очередное падение в течение  $T$  дней после предыдущего, параметр  $T$  считается кратным  $G$ . Пусть крупное падение зафиксировано в  $t_0$ . Успехи и неудачи алгоритма  $A$  определяются следующим образом. Если в течение времени ожидания  $(t_0, t_0 + T]$  действительно произошло крупное падение, скажем в  $t_1$ , то количество успехов алгоритма  $A$  увеличивается на единицу, и время ожидания продлевается на  $(t_1, t_1 + T]$ . Если же  $t_1 > t_0 + T$  (крупное падение произошло по окончании времени ожидания), то количество неудач алгоритма  $A$  увеличивается на единицу. Следующее время ожидания крупного события также устанавливается как  $(t_1, t_1 + T]$ .

Пусть  $N_s$  и  $N_f$  — это количество успехов и неудач алгоритма,  $N = N_s + N_f$  — количество крупных падений. Через  $n = N_f/N$  обозначается доля неудач, через  $\tau$  — доля суммарного времени ожидания в общей протяженности временного ряда. Тогда  $\varepsilon = 1 - n - \tau$  характеризует повторяемость последовательности крупных событий. Если распределение времени между крупными событиями показательное, то  $\varepsilon = 0$  (при практическом вычислении окажется близким к нулю). Напротив, если крупные падения следуют непосредственно одно за другим в моменты времени  $t_0, t_0 + G, t_0 + 2G, \dots$ , то  $\varepsilon = 1$ . Очевидно, что величины  $n$  и  $\tau$  являются адаптированными ошибками первого и второго рода, применяемыми в статистике.

При фиксированных  $r^*$  и  $G$  величина  $\varepsilon$  максимизируется по протяженности  $T$  интервала, в течение которого ожидается повторение крупного события. Другими словами, оценивается интервал, на котором последствие крупного падения наиболее заметно. Используемый шаг при сканировании значений  $T$  равен пяти рабочим дням. Оптимальное время ожидания обозначается  $T^*$ .

### Связь с алгоритмами прогноза

Методология  $(n, \tau)$  применяется для оценки эффективности прогноза экстремальных событий [Molchan, Keilis-Borok, 2008]. Произвольный алгоритм, прогнозирующий моменты наступления экстремальных событий, разделяет временную ось на промежутки двух видов: на одних

прогнозируется наступление экстремального события, то есть объявляется тревога, на других — нет. Тогда  $n$  и  $\tau$  — это доля непредсказанных событий и доля продолжительности тревог соответственно. Величина  $\varepsilon$  показывает эффективность алгоритма: чем ближе эта величина к единице, тем эффективнее прогноз (корректнее говорить о близости абсолютного значения  $\varepsilon$  к единице; подробности, здесь несущественные, можно найти в [Molchan, Keilis-Borok, 2008]). Следовательно, величина  $\varepsilon$ , полученная с помощью алгоритма А, который, несомненно, алгоритмом прогноза не является, должна быть превзойдена любым алгоритмом, разработанным для прогноза.

## Результаты наблюдений

### Среднее время между событиями

Сравнение повторяемости падений различных индексов имеет смысл, если падения различных индексов соответствуют друг другу. Значение  $r^*$  устанавливает не лучшее соответствие, поскольку пороги, определяющие значимые падения индексов, различны для разных индексов. Например, 3%-ое падение, несомненно значимое для индекса S&P100, не столь существенно для индекса DAX. Поэтому в работе, не претендуя на наилучшее решение, в качестве нормировки используется среднее время  $T_a$  между падениями. Для каждого временного ряда параметры  $r^*$  и  $T_a$  взаимнооднозначно определяют друг друга. Чем меньше  $T_a$ , тем меньше абсолютное значение  $r^*$  и выборка содержит большее количество падений.

### Повторяемость событий при $G = 1$

Высокая повторяемость событий,  $\varepsilon > 0.5$ , наблюдается для всех исследуемых индексов на значительном диапазоне  $T_a$  при  $G = 1$ . Рис. 1 иллюстрирует это наблюдение для индексов ATX, CAC40, DAX и NASDAQ. Диапазон значений  $T_a = [50, 100]$ , на котором повторяемость индекса NASDAQ, несомненно, превышает половину, соответствует нижней границе падений  $r^* \in [-0.0503, -0.0413]$ . При  $r^*$ , равном  $-0.0503$  ( $-0.0413$ ), выборка содержит 65 (соответственно, 130) падений. Индекс DAX характеризуется повторяемостью  $\varepsilon > 0.6$  на интервале  $T_a \in [60, 100]$ .

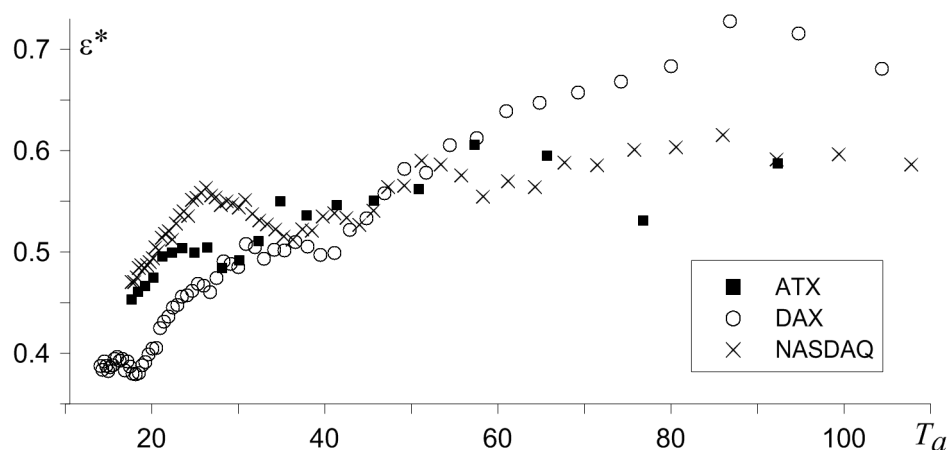


Рис. 1. Повторяемость  $\varepsilon(T_a)$  различных фондовых индексов

Уменьшение  $T_a$  (например, в область  $T_a < 40$ ) добавляет в выборку падения стандартной величины, свойства которых, видимо, отличаются от свойств редких падений. В результате повторяемость  $\varepsilon$  уменьшается. Увеличение  $T_a$ , напротив, оставляет в выборке только редкие падения, которым соответствуют существенные падения индексов. Графики на рис. 1 прорисованы

вправо до тех пор, пока в выборке крахов сохраняется несколько десятков событий и вычисления повторяемости статистически значимы. При  $\varepsilon$ , близких к 0.5, и нескольких десятках падений в выборке стандартное отклонение повторяемости приблизительно равно 0.1. Функция  $\varepsilon(T_a)$  не убывает в исследуемой области  $T_a$ , за исключением индекса DAX, у которого, однако, значение  $\varepsilon$  при больших  $T_a$  больше, чем у других индексов.

### Зависимость повторяемости от $G$

Некоторые крупные падения естественно называть крахами. На качественном уровне крах — это значительное падение цены финансового инструмента за определенный промежуток времени. Выбор величины и длительности падения не однозначен. В статье [Sornette, Zhou, 2006] крахи определяются как трехдневные падения. При анализе связи фондовых индексов с наиболее значимыми макропоказателями рассматривают более протяженные интервалы времени. Можно ожидать, что с увеличением  $G$  повторяемость крахов станет менее заметной. В самом деле, исчезнут «ложные» крахи, представляющие собой падения в серии последовательных колебаний, которые состоят из скачкообразных беспорядочных движений вверх и вниз. В результате крахи, соответствующие большому  $G$ , будут отделены друг от друга. Повторяемость падений для рассматриваемых индексов следует этому сценарию (мы подтверждаем утверждение рисунком, рис. 2, только для индексов DAX и NASDAQ). Некоторое исключение составляет NASDAQ, у которого повторяемость повышается при переходе от  $G = 1$  к  $G = 2$  именно в области больших падений.

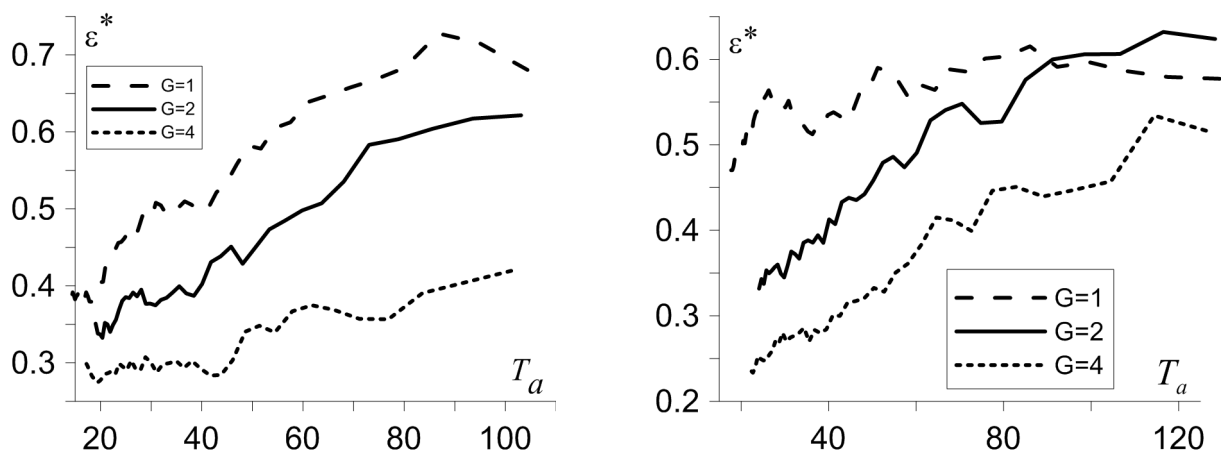


Рис. 2. Повторяемость  $\varepsilon(T_a)$  индексов DAX (слева) и NASDAQ (справа) при различных  $G$

### Продолжительность интервала повторяемости

На рис. 3 представлена зависимость оптимального  $T^*$  от среднего значения  $T_a$  между падениями для индексов DAX и NASDAQ. Очевидно, что уменьшение  $T_a$  ( $T_a < 40$ ) приводит к уменьшению времени повторяемости крупных падений, в частности к уменьшению  $T^*$ . Однако увеличение  $T_a$  приводит к очень медленному росту  $T^*$  в рассматриваемом диапазоне  $T_a$ . Так как повторяемость  $\varepsilon(T_a)$  также слабо чувствительна к  $T^*$ , то с увеличением времени между падениями (то есть при изучении больших и редких падений) доля продолжительности интервала, на котором фиксируется последствие, уменьшается, однако доля повторных падений на этом интервале также уменьшается. Другими словами, последствие проявляется на меньшем интервале времени, причем вероятность последствия понижается.

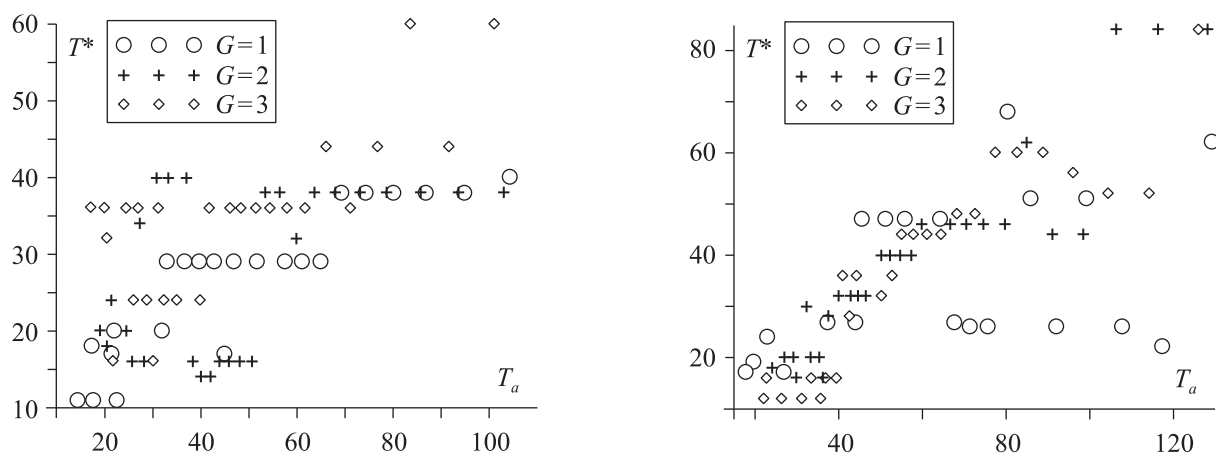


Рис. 3. Зависимость интервала повторяемости  $T^*$  от среднего времени  $T_a$  между падениями

Имеют место также значительные скачки  $T^*$  при малом изменении  $T_a$  (× на рис. 3, слева при  $T \approx 80$ , ● на рис. 3, справа на  $T_a \in [80, 110]$ ). Видимо, оптимальное значение  $\varepsilon$  слабо чувствительно к точному выбору  $T^*$  для редких событий. Чтобы проверить эту гипотезу, повторяемость  $\varepsilon = \varepsilon_T(T_a)$  как функция от  $T_a$  вычисляется для фиксированных  $T$  — протяженности последействия. Для индексов DAX и FTSE вычисления проведены для  $T = 30, 35$  и  $45$ . Вычисленные функции сравниваются с повторяемостью  $\varepsilon(T_a)$ , полученной при оптимальных  $T^*$ , вообще говоря, различных при разных  $T_a$ . Несмотря на существенный разброс значений  $T$ , наибольшее отклонение между функциями  $\varepsilon(T_a)$  составляет 0.055 (рис. 4).

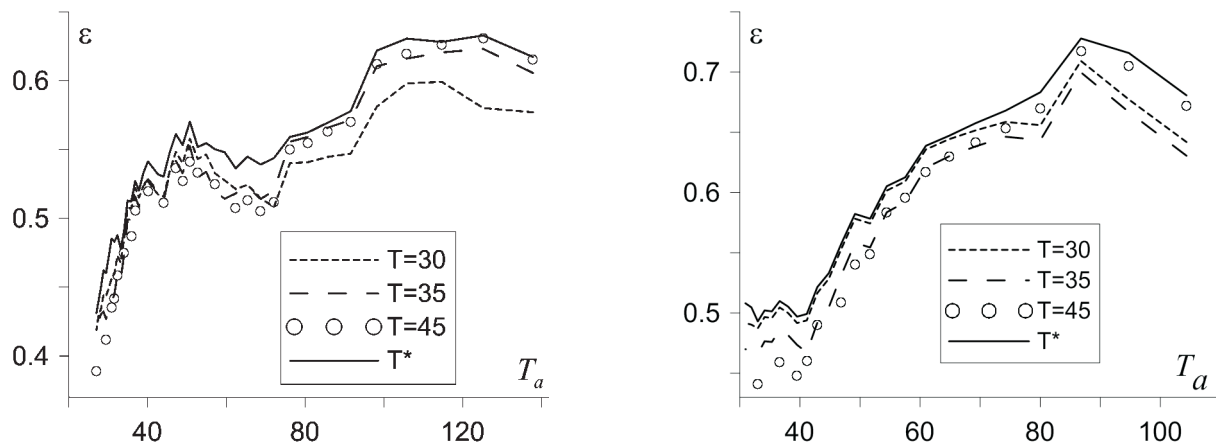


Рис. 4. Повторяемость  $\varepsilon(T_a)$  как функция от  $T_a$  при  $T = 30, 35, 45$  и оптимальном  $T^*$  для FTSE (слева) и DAX (справа)

## Заключение

В статье оценена повторяемость крупных падений наиболее значимых фондовых индексов с помощью доли  $\tau$  длительность интервала, в течение которого вероятность повторного падения высока, и доли  $n$  падений, произошедших за пределами этих интервалов. Сумма  $\varepsilon = n + \tau$  вводится в качестве количественной меры повторяемости падений. Величина  $\varepsilon$  лежит в диапазоне  $[0, 1]$  в зависимости от величины последействия. Установлено, в частности, что за первую четверть



времени между падениями происходит в среднем более трех четвертей всех падений. Соответствующее этой оценке значение  $\varepsilon$ , несколько превосходящее половину, устойчиво по параметрам: величина и продолжительность падений и времени последствия.

По существу, в статье предложена мера отклонения распределения  $\xi$  времени между падениями от показательного. В многочисленных статьях [Lillo, Mantegna, 2003; Selcuk, 2004; Weber et al., 2007] указывается на «тяжелые хвосты» этого распределения для различных финансовых инструментов. Из наших результатов следует, что распределение  $\xi$  для наиболее значимых фондовых индексов существенно отличается от показательного в области малых значений. Указанные отклонения распределения  $\xi$  от показательного свидетельствуют об  $U$ -образном виде функции риска (hazard rate), известной также как функция смертности (или функция отказов). При ее  $U$ -образном виде естественно ожидать наступления события (смерти, отказа, в нашем случае — финансового краха) в начале времени наблюдения или спустя продолжительное время. Это рассуждение дает элементарный прогноз времени наступления событий — объявлять тревогу сразу после краха на фиксированное время  $T_1$  и спустя время  $T_2$  до наступления события, если ко времени  $T_2$  новый крах еще не произошел. Оценка повторяемости  $\varepsilon$  соответствует объявлению тревог первого вида. Поэтому  $\varepsilon \in [0.4, 0.5]$ , устойчивость которого обоснована в нашей статье, — это ориентир для алгоритмов прогноза финансовых крахов, который, насколько мы знаем, до сих пор не превзойден.

## Список литературы

- Кузнецов И. В., Родкин М. В., Серебряков Д. В., Урядов О. Б. Иерархический подход к динамике преступности // Новое в синергетике. Новая реальность, новые проблемы, новое поколение. Сборник статей / Под редакцией Г. Г. Малинецкого. Часть 1. М.: Радиотехника, 2006, 120 с.
- Молчан Г. М. Оптимальные стратегии в прогнозе землетрясений: современные методы интерпретации сейсмологических данных // Вычислительная сейсмология. — 1991. — Т. 24. — С. 3–18.
- Романовский М. Ю., Романовский Ю. М. Введение в эконофизику. Статистические и динамические модели. — М.: ИКИ, 2007. — 150 с.
- Abe S., Suzuki N. Omori's law in the Internet traffic LANL archive // Journal of Anthropological Archaeology. — 2002. — V. 12. — P. 1–40.
- Al-Anaswah N., Wilfling B. Identification of Speculative Bubbles Using State-Space Models with Markov-Switching // Journal of Banking and Finance. — 2011. — V. 35. — P. 1073–1086.
- Blanter E. M., Shnirman M. G., Le Mouél J.-L. and Allegre C. J. Scaling laws in blocks dynamics and dynamic self-organized criticality // Physics of the Earth and Planetary Interiors. — 1997. — V. 99. — P. 295–307.
- Bogachev M. I., Bunde A. Memory effects in the statistics of interoccurrence times between large returns in financial records // Physical Review E. — 2008. — V. 78. — P. 036114–036121.
- Ciamarra M. P., Coniglio A., de Arcangelis L. Correlations and Omori law in spamming. — Europhysics Letters. — 2008. — V. 84. — P. 28004–28007.
- Keilis-Borok V. I. Fundamentals of Earthquake Prediction: Four Paradigms // Nonlinear Dynamics of the Lithosphere and Earthquake Prediction. / V. I. Keilis-Borok, A. A. Soloviev (eds.). Springer-Verlag. — 2003. — P. 1–36.
- Lee J. W., Lee K. E., Rikvold P. A. Waiting-Time distribution for Korean Stock-Market Index KOSPI // Journal of the Korean Physical Society. — 2006. — V. 48. — P. S123–S126.
- Lillo F., Mantegna R. N. Power law relaxation in a complex system: Omori law after a financial market crash // Physical Review E. — 2003. — V. 68. — P. 016119–016123.

- Mantegna R. N., Stanley H. E.* An Introduction to Econophysics: Correlations and Complexity in Finance. — Cambridge University Press. — 2000. — Cambridge, England. — 150 p.
- Molchan G., Keilis-Borok V.* Earthquake prediction: probabilistic aspect // Geophysical J. Int. — 2008. — V. 173. — P. 1012–1017.
- Mu G.-H., Zhou W.-X.* Relaxation dynamics of aftershocks after large volatility shocks in the SSEC index // Physica A. — 2008. — V. 387. — P. 5211–5218.
- Omori F.* On the aftershocks of earthquakes // J. of the College of Science. — 1894. — V. 7. — P. 111–120.
- Romanovsky M. Yu., Vidov P. V.* Analytical representation of stock and stock-indexes returns: Non-Gaussian random walks with various jump laws // Physica A. — 2011. — V. 390. — P. 3794–3805.
- Selcuk F.* Financial earthquakes, aftershocks and scaling in emerging stock markets // Physica A. — 2004. — V. 333. — P. 306–316.
- Shapoval A. B.* Prediction problem for target events based on the inter-event waiting time // Physica A. — 2010. — V. 389. — P. 5145–5154.
- Siokis F. M.* The dynamics of a complex system: The exchange rate crisis in Southeast Asia // Economics Letters. — 2012. — V. 114. — P. 98–101.
- Sornette D., Zhou W.-X.* Predictability of large future changes in major financial indices // International Journal of Forecasting. — 2006. — V. 22. — P. 153–168.
- Weber P., Wang F., Vodenska-Chitkushev I., Havlin S., Stanley H. E.* Relation between volatility correlations in financial markets and Omori processes occurring on all scales // Physical Review E. — 2007. — V. 76. — P. 016109–016116.