

РОССИЙСКАЯ АКАДЕМИЯ НАУК

**ФИЗИЧЕСКИЙ  
ИНСТИТУТ**



*имени  
П. Н. Лебедева*

**Ф И А Н**

ПРЕПРИНТ

**19**

Д.С. ЧЕРНАВСКИЙ, А.П. НИКИТИН,  
О.Д. ЧЕРНАВСКАЯ, О.И. КРИВОШЕЕВ

**ГЕНЕРАЦИЯ И АНАЛИЗ  
КВАЗИХАОТИЧЕСКИХ  
ВРЕМЕННЫХ РЯДОВ**

**Часть 1. Модель  
перемешивающего слоя**

МОСКВА 2005

В работе предложена динамическая модель, включающая шумовую компоненту, которая позволяет генерировать квазихаотические временные ряды, имитирующие явление, называемое «перемешивающий слой», т.е. сценарий «хаотическое поведение – бистабильный режим (перескоки между двумя существенно разными состояниями) – выбор одного стабильного состояния». Такой сценарий характерен для многих процессов в экономике, медицине, и т.д. Предложен также метод анализа сгенерированных рядов, основанный на исследовании некоторых статистических характеристик. Показано, что анализ (с предварительным созданием обучающего множества) позволяет определить «момент истины», т.е. тот момент времени, в который возможно, с заданной вероятностью, предсказать, какое именно стационарное состояние выберет данная система. Исследованы некоторые дополнительные характеристики временных рядов в зависимости от параметров модели и характера шума. Показано, что эти характеристики оказываются существенно различными в случае гауссова шума и шума Парето.

In the present paper, the dynamical model involved the “noise” component is suggested, which enabled us to generate quasi-chaotic time sets imitating the effect of the “mixing layer”, that is, the scenario “chaotic behavior — bistable regime (jumps between two considerably different states) — choice of one specific stable state”. This scenario is typical in economics, medicine, and so on. The method of analysis of these sets based on study of some statistical characteristics is suggested to determine (with using the preliminary created the “learning set”) the “moment of truth”, that is the moment when it possible, with the given probability, to predict, which specific stable state will be chosen. Some additional characteristics of the sets have been analyzed as functions of the model parameters and of the noise type. It is shown that these characteristics demonstrate considerably different behavior in the cases of the Gaussian noise and the Pareto one.

## 1. Введение

В последнее время, в самых разных областях науки и человеческой деятельности вообще, возникает задача анализа квазихаотических временных рядов. Такие проблемы характерны, например, для биржевых котировок акций, сейсмологии, медицины и т.д. При этом, характер временных рядов может быть самым разным (от более-менее регулярного, как, например, в кардиограммах, до полностью хаотического). Однако, реальные ряды, как правило, квазислучайны, т.е. имеют как динамическую, так и хаотическую составляющие. Последнее неизбежно, так как любая реальная система подвержена случайным воздействиям, или находится в «шумовом поле». Однако в зависимости от динамики системы шум может оказывать пренебрежимое, или, напротив, решающее воздействие.

При анализе временных рядов могут преследоваться различные цели:

1. Выяснить динамику процесса, которая, как правило, неизвестна.
2. Дать кратковременный прогноз поведения ряда на основе каких-либо наблюдаемых признаков. При этом интерес представляют сравнительно небольшие флуктуации.
3. Дать прогноз редких, но значительных изменений. Крупные (катастрофические) изменения возможны в том случае, если система мультистабильна, то есть имеет два (или более) устойчивых состояния.

В данной работе мы сосредоточим внимание именно на последней задаче. Обсудим ее более подробно.

Для очень широкого диапазона приложений, характерна следующая картина временного ряда (или его отрезков) [Чернавский, 2004]: хаотический кусок сменяется режимом резких перескоков между двумя существенно различными состояниями, затем система выбирает одно из этих состояний, в котором живет достаточно долго; затем процесс повторяется. Такая картина характерна, например, при конкуренции двух крупных фирм, причем в конце концов побеждает одна из них. Это означает, что *a priori* возможны два стабильных состояния, т.е. система бистабильна. Предсказать *заранее*, какая фирма окажется победителем, задача, важная в чисто практическом отношении и на начальном (хаотическом) этапе невыполнимая. Такое предсказание можно сделать в определенные моменты (или интервалы) времени — так называемые «моменты истины», и оно неизбежно будет иметь вероятностный характер. Термин «момент истины» был предложен Дж. Соросом в книге [Сорос, 1998]; там же обсуждаются примеры подобных ситуаций. Стадия развития, в которой имеет место хаотический режим, получила название «джокер» [Белайчук, Малинецкий, 1997; Малинецкий, Потапов, 2000], или «перемешивающий слой» [Колупаев, Чернавский, 1997, Чернавский и др., 2003; Курдюмов и др., 2005]. «Момент истины» наступает в конце перемешивающего слоя.

Ниже предлагается простейшая математическая модель, в рамках которой удается описать именно такой сценарий и дать «рецепт» анализа с целью определения моментов истины.

## 2. Формулировка модели

Динамическая система в шумовом поле описывается уравнением типа Ланжевена:

$$\frac{\partial x}{\partial t} = f(x) + G(t), \quad (1)$$

где  $x$  – наблюдаемая величина,  $f(x)$  – функция, описывающая динамику (в общем случае, нелинейную).  $f(x)$  можно представить как  $dU/dx$ , где  $U(x)$  – потенциал системы.  $G(t)$  — случайная величина с заданным распределением (т.е. шум), в общем случае зависящая от времени. Подчеркнем два основных фактора, необходимых для того, чтобы описать интересующий нас сценарий (хаос → перемешивающий слой → выбор одного из возможных вариантов):

а) потенциал должен быть бистабилен, т.е. допускаются два существенно различных стационарных состояния;

б) амплитуда шума медленно (по сравнению с частотой флуктуаций) падает со временем.

Отметим, что подобные модели достаточно широко исследуются с теоретических позиций и применяются для описания социально-экономических процессов, например, динамики курсов акций и цен [Bouchaud, Cont, 1998; Friedrich et al, 2000; Mariani et al, 2000], доходности финансовых инструментов [Sornette, 2001], формирования общественного мнения [Tessone, Toral, 2005] и др. Однако использовавшаяся гипотеза о постоянстве (отсутствии зависимости от времени) шума позволяла только изучать характеристики генерируемого временного ряда, но не определять моменты истины (при такой постановке их просто нет).

Мы рассматриваем конкретное уравнение:

$$\frac{dx}{dt} = (x - x^3) + g(t) \cdot \xi(t) \quad (2)$$

Здесь, величина  $\xi(t) = \xi_t(t) \cdot \xi_A$  — случайный дельта-коррелированный шум с нормированной амплитудой  $\xi_A$ . Плотность вероятности реализации конкретного значения  $\xi_A$  описывается некоторым заданным распределением. Там, где это не указано особо, подразумевается стандартное гауссово распределение с нулевым средним и дисперсией  $\sigma^2$ . Шумовая компонента  $\xi_t(t)$  отлична от нуля только в моменты времени  $t = \tau \cdot i$ , где  $i$  — целое число. Таким образом, случайные импульсы с частотой  $1/\tau$  вызывают мгновенное изменение координаты  $x$ .

Величина  $g(t)$  — монотонно убывающая функция. Ниже рассматривается

вариант  $g(t) = g_0 e^{-\nu t}$ , где  $\nu$  — параметр модели.

Таким образом, формально имеется 4 параметра, определяющих поведение модели:  $\tau$ ,  $g_0$ ,  $\nu$ ,  $\sigma$ . Отметим, что влияние параметра  $\sigma$  аналогично влиянию  $g_0$ , т.е. изменение стандартного отклонения  $\sigma$  приводит к тому же изменению произведения  $g(t)\xi(t)$ , что и соответствующее изменение  $g_0$ . (Это утверждение легко доказывается, так как известно свойство нормального распределения: если случайная величина  $\xi$  имеет нормальное распределение  $N(a, \sigma^2)$ , то случайная величина  $k\xi+b$  также имеет нормальное распределение —  $N(a+b, k^2 \sigma^2)$ ). Поэтому ниже положим  $\sigma=1$  и будем рассматривать поведение модели при различных значениях трех параметров:  $\tau$ ,  $g_0$  и  $\nu$ . Динамическая часть модели параметров не имеет, поскольку величины  $x$  и  $t$  выбраны, не ограничивая общности, в безразмерных единицах, поэтому характерный масштаб изменения как  $x$ , так и  $t$  — порядка единицы.

Динамическая система имеет три стационарных состояния: два устойчивых ( $x = +1, -1$ ) и одно неустойчивое ( $x = 0$ ). В отсутствие шума она стремится в одно из устойчивых состояний, выбор которого определяется начальными условиями. Включение шума, в зависимости от конкретных значений параметров, может либо слабо повлиять на поведение системы, либо привести (в начале) к полностью хаотическому поведению и (в конце) к случайному выбору одного состояния, никак не связанному с начальными условиями. Таким образом, нетривиальное поведение системы проявляется при определенных комбинациях параметров  $\tau$ ,  $g_0$  и  $\nu$ .

Обсудим роль параметра  $\tau$ . При  $\tau \ll 1$  система испытывает практически непрерывное случайное воздействие («дрожание»); в противоположном пределе  $\tau \sim 1/2$  (т.е. порядка характерного динамического времени), напротив, поведение определяется главным образом динамикой, по модулю редких случайных «толчков». Вообще говоря, в реальных системах, как правило, присутствует оба типа воздействий, причем они имеют разную природу. По-видимому, при попытке описания каких-либо реальных явлений необходимо вводить два (вероятнее всего, не связанных между собой) шума, имеющих разную временную и амплитудную шкалу. Здесь мы решаем задачу имитации и анализа перемешивающего слоя, для чего одного шума достаточно. Однако зависимость характеристик системы от  $\tau$  будет прослеживаться в достаточно широком диапазоне  $0.01 < \tau < 0.7$ .

### 3. Генерация и анализ временных рядов.

В свете поставленных целей при разумных значениях параметров  $\tau$ ,  $\nu$  и  $g_0$  была проведена генерация временных рядов  $X(t_i) = x(t_i = t_0 + \tau i)$ , где  $i = 1 \dots N$ . Общая продолжительность «наблюдения» поведения системы, таким образом, составляет  $T = N\tau$  и была выбрана равной  $T = 10000$ .

Временные ряды подобного рода, порождаемые уравнением Ланжевена с шумовой составляющей  $\xi(t)$ , изучались достаточно активно (см., например, [Siegert et al, 1998; Gradisek et al, 2003]). Однако подчеркнем, что в перечисленных работах использовался шум с нормальным гауссовым распределением и постоянной амплитудой  $\xi_A$ .

Между тем, в настоящее время не вызывает сомнения тот факт, что случайные изменения во многих экономических процессах на относительно малом масштабе времени имеют существенно негауссов характер. В качестве примера можно привести работу [Bouchaud, Cont, 1998], где проводился анализ изменения курсов акций. В этой связи в настоящей работе нами были рассмотрены, помимо гауссова, и другие типы шумов, в частности, экспоненциальный шум и так называемый Парето-шум.

Получение экспоненциального шума основано на [Dagpunar, 1988], согласно которому шум  $\xi(t) = a \ln(r)$ , где  $r$  — случайное равномерно распределенное число  $r \in [0,1)$ ,  $a$  — случайно выбираемый знак  $\xi(t)$  ( $a = \pm 1$ ). Для  $\ln(r)$  справедливо следующее: среднее значение и стандартное отклонения равны 1.

Парето-шум, подчиняющийся распределению с плотностью вероятности  $\rho(x) \sim \left[ \frac{1}{x+1} \right]^2$ , был имитирован с помощью свободно распространяемой программы *RandGen* [Lohninger]. Эта программа позволяет получить псевдослучайную последовательность, распределение элементов которой имеет плотность вероятности, задаваемую пользователем. Пользователь определяет конечный диапазон по  $x \in [a,b]$ , число интервалов  $k$  разбиения  $[a,b]$  и значения  $\rho(x)$  в точках  $x_i = a + (b-a)/k$ . Распределение Парето было имитировано путем вычисления функции  $\rho(x)$  на четырех диапазонах:  $[0,1)$ ;  $[1,5)$ ;  $[5,25)$ ;  $[25, 225)$  с шагом, соответственно, 0.005, 0.02, 0.1, 1.0. Таким образом, распределение было обрезано по  $x = 225$ , т.е.  $\rho(x) = 0$  при  $x > 225$ . Последнее значение  $\rho(225) = 0.0000195787$ . Вероятность попадания случайной величины в перечисленные 4 диапазона была оценена следующим образом:  $\approx 50.10\%$  для  $[0,1)$ ;  $\approx 33.50\%$  для  $[1,5)$ ;  $\approx 12.93\%$  для  $[5,25)$ ;  $\approx 3.47\%$  для  $[25, 225)$ . Симметричность распределения достигалась путем случайной генерации знака  $x$ .

В качестве генератора псевдослучайных чисел использовался алгоритм Дж.Марсальи, обеспечивающий получение равномерно распределенных случайных величин в диапазоне  $[0,1)$  с длиной неповторяющейся последовательности  $\sim 3 \cdot 10^{47}$ . В качестве реализации алгоритма использовался свободно распространяемый модуль *MRNG.pas* [EFD].

Для получения нормально распределенных последовательностей использовался алгоритм Марсальи-Брея, реализованный в функции *RandG* из

модуля *Math* среды разработки *Delphi5.0*. Данная функция обеспечивает нормальное распределение с заданными средним и дисперсией (в нашем случае, с нулевым средним и единичной дисперсией). Также использовался модуль *AMRandom.pas* [ESB], позволяющий получить нормально и экспоненциально распределенные псевдослучайные последовательности.

На рис. 1. приведены типичные временные ряды, описывающие поведение модели.

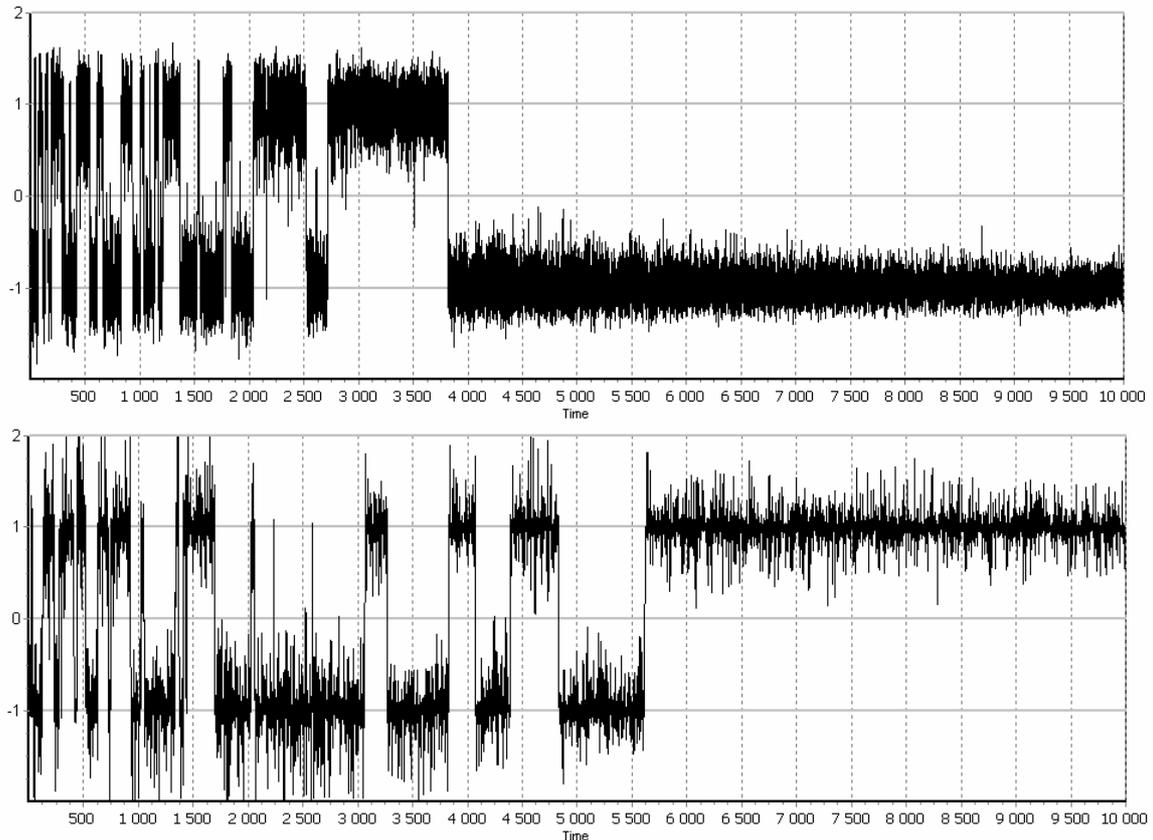


Рис. 1. Реализация при гауссовом шуме (1а,верху) и Парето-шуме (1б,внизу).

Из рис. 1 видно, что весь «жизненный цикл»  $X(t)$  можно разделить на три стадии: «стохастический резонанс» (на рис.1а приблизительно до  $t = 500$ ), «бистабильный режим» (до  $t = 3800$ ), «униmodalный режим».

Визуальные признаки каждой стадии ясны из графиков. На стадии стохастического резонанса временной ряд заполняет собой все доступное пространство и области  $x = \pm 1$  не выделены.

В бистабильном режиме можно выделить «состояния», в которых величина  $x$  достаточно долгое время проводит либо в области  $+1$ , либо  $-1$ . Иными словами, переходы (перескоки) в которых изменение величины  $x$  на величину порядка  $\Delta x = 2$  достаточно редкие, т.е. времена между ними много больше времени  $\tau$ . При этом текущее состояние системы можно отнести к одному из стационарных состояний. Для этого был построен следующий

алгоритм определения текущего состояния системы. (Справедливости ради отметим, что данный алгоритм непригоден для стадии стохастического резонанса.)

1. Определено три состояния системы  $S(t_i)$ : «+1», «-1», «0».
2. Состояние системы оценивается дискретно, через интервалы  $\Delta t_y = n_y \tau = 10\tau$ .
3. Рассчитывается среднее значение  $X(t_i)$  на предшествующем интервале

$$\Delta t_y Y_i = \frac{1}{n_y} \sum_{j=i-n_y+1}^i X(t_j).$$

4. Определена функция оценки текущего состояния в момент времени  $t_i$  (на

$$i\text{-м интервале)} Z(t_i) = \begin{cases} +1, Y_i > L^+ \\ -1, Y_i < L^- \\ 0, Y_i \in [L^-, L^+] \end{cases}. \text{ Экспериментально подтверждена}$$

применимость очевидных значений  $L^+ = 0.5$  и  $L^- = -0.5$ .

5. Оценивалось состояние системы на предыдущих 10 интервалах, т.е.  $S(t_i - n_y \tau j)$ ,  $j = 1 \dots 10$ .

6. Если состояние системы  $S(t_i - n_y \tau j)$  не меняло знак по сравнению с  $S(t_i)$ , то все  $S(t_i - n_y \tau j)$  переопределяются на  $S(t_i)$ .

Таким образом производится «сшивка» близко расположенных интервалов одинаковой «полярности» ( $S = +1$  или  $S = -1$ ), разделенных «серыми» интервалами (т.е. только такими, что  $S = 0$ ).

7. Определялось финальное состояние системы  $S(T)$  на конце периода наблюдения.

8. Временной ряд просматривался с конца для нахождения момента последнего перескока  $t_s$ , при котором  $S(t_s) = -S(T)$ .

Данный алгоритм позволяет надежно определять момент последнего перескока, исключая случайные (короче чем на  $\Delta t_y$ ) переходы в противоположное состояние и не учитывая кратковременные (не дольше  $10\Delta t_y$ ) выходы системы в «нулевое» состояние.

Существует несколько способов анализа полученных рядов: вейвлет-анализ, фрактальный анализ (индекс амплитудной вариации [Dubovikov et al, 2004]), «распознавание образа» при помощи обучающего множества. Ниже мы рассмотрим именно последний метод и продемонстрируем, как с его помощью определяются «моменты истины».

Обратимся предварительно к работе [Чернавский и др., 2004] и сделаем несколько вводных замечаний. Главная задача теории распознавания — дать прогноз развитию событий, а ее фундаментом служит утверждение: «Если наблюдаемый объект «похож» на уже известный прототип, то его поведение тоже будет сходно с поведением этого прототипа». В общем виде задача

распознавания состоит в идентификации предъявляемого объекта с одним из известных классов, т.е. в его классификации. Вообще говоря, допускается ответ, означающий отказ от классификации конкретного объекта («отказ от диагноза»). Если правило идентификации неизвестно, то возникает задача построения решающего правила отнесения объектов к одному из указанных классов. Для этого необходимо обучающее множество (материал обучения) — набор известных прецедентов, объектов и/или явлений, каждый из которых характеризуется набором значений количественных или качественных признаков. Любое обучающее множество формируется с определенной целью, т.е. целесообразно. Экзаменуемое множество — это набор объектов, которые сравниваются с известными прецедентами, т.е. объектами обучающего множества. Предполагается, что любому объекту, подлежащему распознаванию, найдется некоторый аналог в обучающем множестве. Реально это допущение оправдано не всегда, т.е. обучающее множество может оказаться недостаточным (нерепрезентативным) и потребуются его пополнение. И, наконец, обучение распознаванию — процесс, в результате которого интеллектуальная система становится способной отнести данный объект из экзаменуемого множества к одному из классов обучающего множества.

Итак, в интересующей нас задаче:

- объект — это конкретная реализация временного ряда;
- признаки — статистические и прочие характеристики этого ряда;
- обучающее множество — набор реализаций, полученных при тех или иных значениях параметров модели;
- экзаменуемый объект — наблюдаемая реализация, для которой необходимо сделать прогноз;
- решающее правило — алгоритм, позволяющий соотнести предъявленный экзаменуемый объект к одному из классов обучающего множества.

#### **4. Поиск моментов истины**

Будем считать, что моментом истины является момент времени  $t_{ист}$ , начиная с которого с заданной вероятностью можно утверждать, что система останется в текущем состоянии (дальнейших переходов/перескоков не будет). Введем обозначения  $t_{ист}(30)$ ,  $t_{ист}(20)$ ,  $t_{ист}(10)$  для «моментов истины», соответствующих вероятности последующего перескока в 30, 20 и 10%.

Общая постановка задачи заключается в определении моментов истины, исходя из характеристик самого временного ряда  $X(t)$  при априорно неизвестных значениях параметров и неизвестном характере зависимости  $g(t)$ .

Были проведены вычислительные эксперименты, позволившие исследовать поведение квазихаотических временных рядов в диапазоне

$\tau \in [0.01, 0.7]$ .

Интересующие нас диапазоны допустимых значений параметров  $g_0$  и  $\nu$  связаны с динамическим поведением модели. Можно сформулировать качественное правило: частота и амплитуда шума должны быть таковы, чтобы среднее значение модуля амплитуды  $|g(t) \cdot \xi_A|$  было сопоставимо с характерными приращениями  $x$  при релаксации системы в одно из стационарных состояний, причем усреднение производится по интервалу  $\Delta t : \tau \ll \Delta t \ll 1/\nu$ .

Таким образом,  $\nu$  должно быть сравнительно мало, так чтобы система находилась в бистабильном режиме достаточное время для осуществления хотя бы нескольких четко идентифицируемых перескоков из одного состояния в другое.

Для каждого  $\tau$  и каждого типа шума подбиралось  $g_0$ , позволяющее «наблюдать» все состояния системы (стохастический резонанс, бистабильный и унимодальный режимы) и при этом обеспечивающее «почти наверняка» отсутствие дальнейших перескоков после окончания времени наблюдения за системой  $T$ . Выбор  $g_0$  связан с  $\nu$  и фактически предопределяет продолжительность наблюдения этапа стохастического резонанса до перехода системы в бистабильный режим. В этом смысле выбор конкретного значения  $g_0$  определяет начальный момент  $t_0$ .

Для первого этапа анализа выбраны следующие диапазоны значений параметров:  $\nu \in [0.00008; 0.0002]$ ,  $g_0 \in [0.04; 0.4]$  для гауссова шума,  $g_0 \in [0.03; 0.2]$  для экспоненциального шума и  $g_0 \in [0.002; 0.009]$  для Парето-шума.

Общая продолжительность наблюдения  $T$  выбиралась из соотношения  $T \geq 1/\nu$ , т.е. амплитуда шума должна была уменьшиться по крайней мере в  $e$  раз. При этом можно «гарантировать», что дальнейших перескоков (на разумном интервале времени, например,  $100 T$ ) уже не произойдет.

В общем случае для нахождения «момента истины» с заданной вероятностью полезно использовать несколько независимых признаков. Однако в данной модели наиболее информативным признаком оказывается величина  $D_4$ . Алгоритм ее определения состоит в следующем.

Строится буферный массив, каждый элемент которого представляет собой модуль разности значений  $X$  в моменты времени, соответствующие  $i$ -му и  $(i-1)$ -му шумовому импульсу:  $D_{mix}[i] = |X(\tau \cdot i) - X(\tau \cdot (i-1))|$ . Массив  $D_{mix}[i]$  разбивается на интервалы по  $q$  элементов. Далее рассчитывается статистика  $D_4$  — средние значения на  $j$ -х интервалах для последовательности  $D_{mix}[i]$ :

$$D_4[j] = \sum_{i=(j-1)q+1}^{jq} D_{mix}[i].$$

Характер экспериментальных кривых для  $D_4$  позволяет рассматривать этот признак как перспективный для последующего анализа. Для получения устойчивых данных необходимо провести «обучение», заключающееся в генерации массива реализаций временного ряда и усреднении статистики по этому массиву (см. рис. 2).

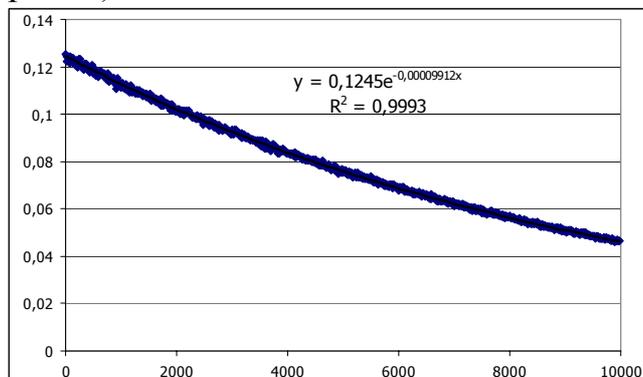


Рис. 2. Статистика  $D_4$  (усреднение по 100  $\tau$  и 100 реализациям).

На графике построена кривая экспоненциальной регрессии  $ER(t)=k_0 \exp[-k_1 t]$ . Значение коэффициента детерминированности ( $R^2 > 0.999$ ) свидетельствует об адекватности такой модели для аппроксимации соответствующей статистики. Обратим внимание, что коэффициент  $k_1$  является очень хорошей аппроксимацией коэффициента  $\nu$  (по крайней мере, при выбранном  $\tau$ ).

На большом числе реализаций проверено соответствие наблюдаемой вероятности исхода (финальное состояние «+1» или «-1») по распределению Лапласа (например, попадание в  $2\sigma_L = 2\sqrt{mpq}$ , где  $m$  — число экспериментов,  $p = 0.5$ ,  $q = 0.5$  — равная вероятность исходов «+1» и «-1» соответственно). Например, для параметров  $g_0=0.25$ ;  $\tau = 0.1$ ,  $\nu = 0.0002$  и гауссова шума было последовательно проведено 10000 реализаций временного ряда  $X(t)$ . В результате в 4983 случаях было достигнуто унимодальное состояние «-1» и в 5017 случаях — состояние «+1». Так как  $\sigma_L = 50$ , то полученное в численном эксперименте соотношение заведомо попадает в пределы  $\pm \sigma_L$  от ожидаемого.

Исследуем вопрос о моменте истины  $t_{ucm}$ , для чего построим зависимости времени последнего перескока системы из одного состояния в противоположное («-1» → «+1» или «+1» → «-1») при различных комбинациях параметров. Можно построить зависимость  $P(t)$ , описывающую зависимость вероятности последующего перескока от времени. Такая зависимость должна строиться для каждого комплекта параметров модели (см. рис. 3). Из  $P(t)$  можно получить интересные моменты истины, в частности,  $t_{ucm}(30)$ ,  $t_{ucm}(20)$  и  $t_{ucm}(10)$  (см. рис. 4).

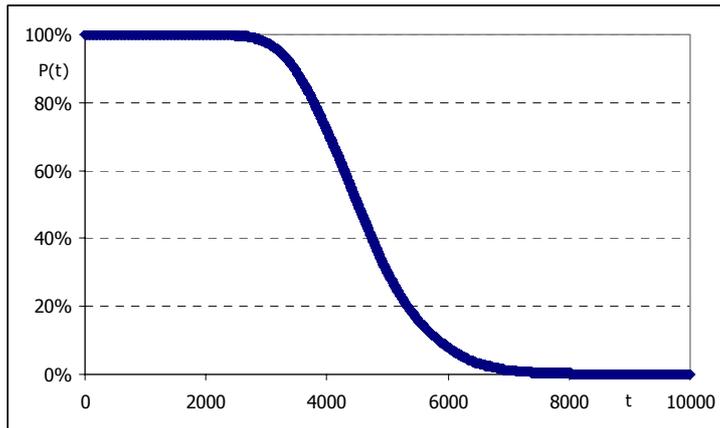


Рис. 3. Зависимость вероятности последующего перескока от времени для стандартного набора параметров (гауссов шум,  $g_0 = 0.15$ ,  $\tau = 0.1$ ;  $\nu = 10^{-4}$ )

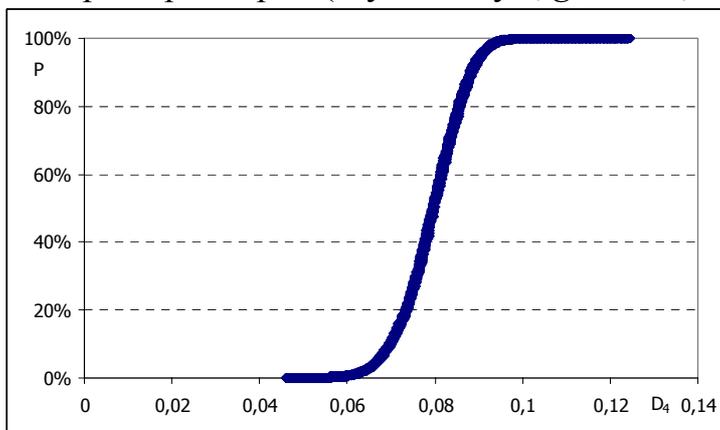


Рис. 4. Зависимость вероятности последующего перескока от статистики  $D_4$  для стандартного набора параметров (гауссов шум,  $g_0 = 0.15$ ,  $\tau = 0.1$ ;  $\nu = 10^{-4}$ )

Эта зависимость может быть аппроксимирована формулой

$$P(D_4) = 1 - \exp\left(-\frac{\nu_0}{\nu} \left[\frac{D_4}{D_4^{(0)}}\right]^A\right), \quad (3)$$

где  $D_4^{(0)}$  — характерная величина  $D_4$ , полученная при фиксированном  $\nu_0 = 10^{-4}$ , а  $A$  — параметр, не зависящий от  $\nu$ . С помощью средств регрессионного анализа можно для модели с гауссовым шумом получить (например, при  $\tau = 0.1$ ) следующие оценки коэффициентов:  $D_4^{(0)} = 0.082$  и  $A = 13.6$ .

Для определения момента истины генерировалось достаточно обширное обучающее множество (в описываемых вычислительных экспериментах использовались различные варианты с генерацией от 1000 до 10000 реализаций). Как было представлено выше, рассчитывался массив  $D_{mix}[i]$  и на его основе дополнительная статистика  $D_4(t)$  (по интервалам из  $q = 10 \div 20$  точек,

т.е. длительностью  $10 \div 20 \tau$ ), которая далее усреднялась по обучающему множеству. На том же обучающем множестве строилась зависимость наблюдаемых «последних перескоков» от времени  $P(t)$ . На этой зависимости выделяются характерные точки «моменты истины», например,  $t_{ucm2}(10)$ ,  $t_{ucm2}(20)$ ,  $t_{ucm2}(30)$ . Для каждого момента истины находится соответствующее значение статистики  $D_4$ , обозначаемое как  $D_4(10)$ ,  $D_4(20)$  и т.д. После обработки обучающих множеств для различных  $\tau$  строятся зависимости  $D_4(10, \tau)$ ,  $D_4(20, \tau)$ ,  $D_4(30, \tau)$  и т.д. (см. рис. 5). Для проверки устойчивости обнаруженных особенностей варьировались: количество реализаций, интервалы усреднения  $D_{mix}$  и  $D_4$ , продолжительность наблюдения  $T$ , Рунге-куттовский шаг разностной схемы.

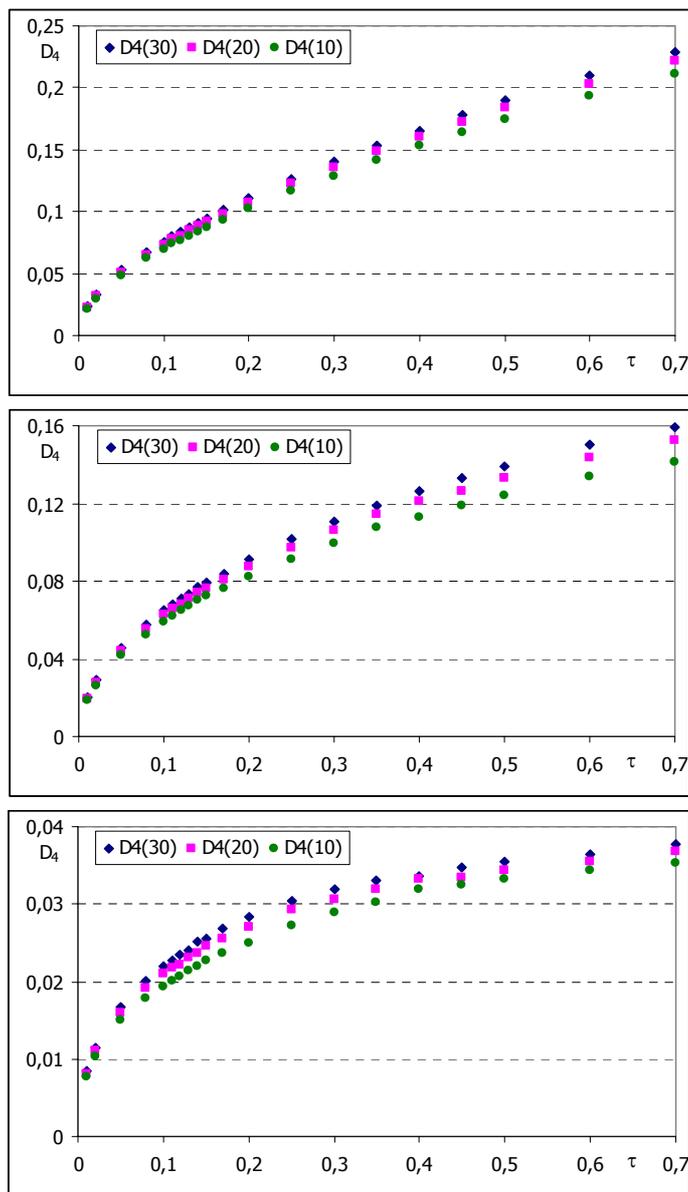


Рис. 5. Зависимости моментов истины от  $\tau$  для гауссова (ба, вверху), экспоненциального (бб, в середине) и Парето шума (бс, внизу)

На кривых можно выделить интервалы линейности. Так,  $D_4(30, \tau)$  для гауссова шума в диапазоне  $\tau = 0.05 \div 0.2$  может быть аппроксимирована линейной зависимостью (см. рис. 6):  $D_4(30, \tau) = 0.387\tau + 0.036$ . Это наблюдение позволяет при постоянном  $\nu$  использовать единое решающее правило в достаточно широком диапазоне по  $\tau$ . Сформулируем это правило: «если экспериментально наблюдаемая величина  $D_4$  падает ниже  $0.387\tau + 0.036$  при известном  $\tau$ , то вероятность перескока из текущего состояния системы не превышает 30%».

Аналогичный вид имеют зависимости характерных значений  $D_4$  и для других моментов истины.

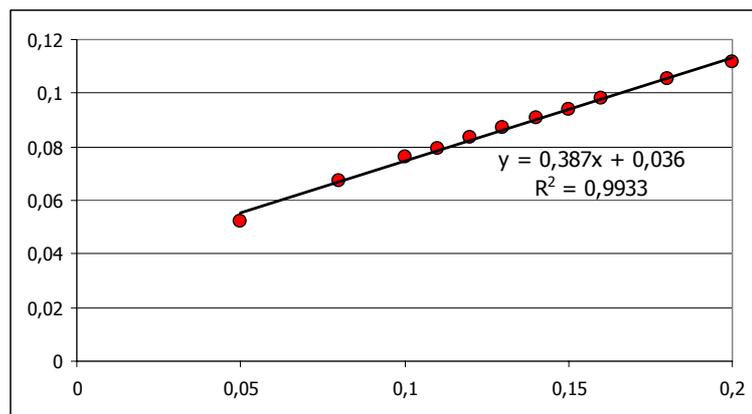


Рис. 6. Участок линейности для зависимости  $D_4(30)$  от  $\tau$ .

Вернемся к кривым, отображенным на рис. 5, и кратко представим еще некоторые результаты, полученные в диапазоне  $\tau \in [0.01; 0.7]$ .

**Гауссов шум.** Видно, что в исследованном диапазоне монотонный рост зависимостей  $D_4(30, \tau)$  сохраняется, постепенно замедляясь. На начальном этапе ( $\tau = 0.01 \div 0.05$ ) коэффициент наклона составляет  $a = 0.77$ , на участке  $(0.1 \div 0.15)$  он падает до  $a = 0.37$ , на участке  $0.45 \div 0.7$  он составляет  $a = 0.21$ . Таким образом, на заключительном этапе он составляет  $\eta = 0.56$  от коэффициента  $a$  на втором участке.

**Экспоненциальный шум.** Характерное значение  $\eta = 0.39$ , т.е. рост несколько медленнее, чем для Гауссова шума.

**Парето-шум.** Обратим внимание, что при сохранившемся характере кривых, рост после  $\tau = 0.35$  существенно замедляется. Характерное значение составляет  $\eta = 0.15$ , что почти в 4 раза меньше, чем для гауссова шума (значения коэффициента наклона равны соответственно 0.23, 0.074 и 0.011).

Также была проанализирована вариабельность зависимостей  $D_4(10, \tau)$ ,  $D_4(20, \tau)$  и  $D_4(30, \tau)$  при изменении параметра  $\nu$ . Помимо стандартного значения  $\nu_0 = 0.0001$  были взяты два дополнительных значения  $\nu_1 = 0.00012$ ,

$v_2 = 0.00008$ . Выяснено, что на всех зависимостях  $D_4(\tau)$  для трех типов шумов (гауссов, экспоненциальный, Парето) увеличение  $v$  на 20% ( $v_0 \rightarrow v_1$ ) приводит к увеличению  $D_4(\tau)$  на  $1.5 \pm 1.0\%$  в исследованном диапазоне  $\tau$ . Аналогично снижение  $v$  на 20% ( $v_0 \rightarrow v_2$ ) приводит к изменению  $D_4(\tau)$  на  $1.5 \pm 1.0\%$ . Так как характерный масштаб изменения  $D_4(30, \tau) \rightarrow D_4(10, \tau)$  составляет порядка 10% (см. рис. 35–37), то можно оценить погрешность в определении момента истины при оценке  $v$  с точностью  $\pm 20\%$ . Так вместо  $D_4(30, \tau)$  будут определяться соответственно,  $D_4(33, \tau)$  или  $D_4(27, \tau)$ .

Определение моментов истины для конкретной реализации может быть осуществлено с помощью линейной регрессии статистики  $D_4$  и сопоставления с результатами численных экспериментов на обучающей выборке (см. табл. 1).

Таблица 1

Экспериментальное определение момента истины (для  $g_0 = 0.15$  и  $\tau = 0.1$ )

Диапазон временного ряда и метод аппроксимации	$t_{ucm}(30)$	$t_{ucm}(20)$	$t_{ucm}(10)$
из «обучения»	4950	5282	5755
экспоненциальная регрессия по предыстории $t = 0-60000$	4930	5263	5738
линейная регрессия по предыстории $t = 20000-60000$	4985	5291	5710
линейная регрессия по предыстории $t = 40000-60000$	5023	5323	5734
линейная регрессия по предыстории $t = 45000-55000$	5023	5322	5731
линейная регрессия по предыстории $t = 30000-45000$	4785	5050	5413

Применение линейной регрессии (см. рис. 7) вместо экспоненциальной представляется более оправданным из-за характерного разброса статистики  $D_4$ . Отметим, что при длине фрагмента ряда короче  $10000\tau$  коэффициент детерминированности  $R^2$  падает до  $0.12 \div 0.15$ , поэтому дальнейшее уменьшение анализируемого фрагмента неоправданно без дополнительных исследований устойчивости.

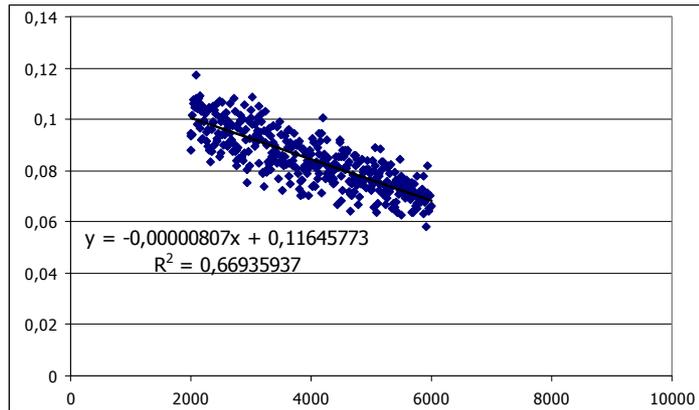


Рис. 7. Линейная регрессия  $D_4$ .

Обратим внимание, что определение момента истины  $t_{ист}$  с удовлетворительной точностью достигается и по фрагменту ряда, не включающему собственно  $t_{ист}$ , т.е. возможна экстраполяция, т.е. «предсказание»  $t_{ист}$  как вперед, так и назад (за пределы периода наблюдения).

Таким образом, определение момента истины возможно при условии предварительного обучения (по крайней мере, на 100 реализациях ряда для каждого набора параметров) и достаточно длительного (порядка  $10000\tau$ ) периода наблюдения за поведением системы.

### 5. Исследование дополнительных характеристик рядов при различных распределениях шума.

Предметом дополнительного исследования являются зависимости относительной ширины «коридора моментов истины» от  $\tau$ . Более конкретно, исследуются зависимости

$$Z_1 = \frac{2(D_4(30) - D_4(10))}{D_4(30) + D_4(10)}, \quad Z_2 = \frac{2(D_4(70) - D_4(30))}{D_4(70) + D_4(30)}, \quad Z_3 = \frac{2(D_4(70) - D_4(10))}{D_4(70) + D_4(10)}$$

от  $\tau$ . Величины  $Z_1$ ,  $Z_2$  и  $Z_3$  отражают относительную эффективную ширину коридора, в котором происходит (статистически) переход системы от бистабильного к унимодальному режиму. Причем  $Z_2$  отражает поведение линейной части  $P(t)$ ,  $Z_1$  — ее «правого хвоста», а  $Z_3$  — обобщенный характер спада  $P(t)$ . Оценим поведение характеристик  $Z_1$ ,  $Z_2$  и  $Z_3$  при разных распределениях шума (см. рис. 8).

Для гауссова шума характеристики  $Z_1$ ,  $Z_2$  и  $Z_3$  практически постоянны в исследованном диапазоне  $\tau$ . Для экспоненциального шума обнаруживается слабый линейный рост эффективной ширины коридора от  $\tau$ .

Наиболее интересно поведение  $Z_1$ ,  $Z_2$  и  $Z_3$  для Парето-шума. Для  $Z_1$  обнаруживается максимум в области  $\tau=0.13 \div 0.15$  с последующим падением (ниже чем соотв. «асимптотические» значения для гауссова шума). Для  $Z_2$  аналогичный максимум достигается чуть раньше при  $\tau=0.1$ , далее  $Z_2$  падает,

достигая минимума при  $\tau=0.25\div 0.3$ , после чего вновь растет.

Из приведенных рисунков видно, что рассматриваемые характеристики ведут себя различно при разных распределениях шума. Это позволяет надеяться на то, что с их помощью окажется возможным определить характер шума в реальных рядах.

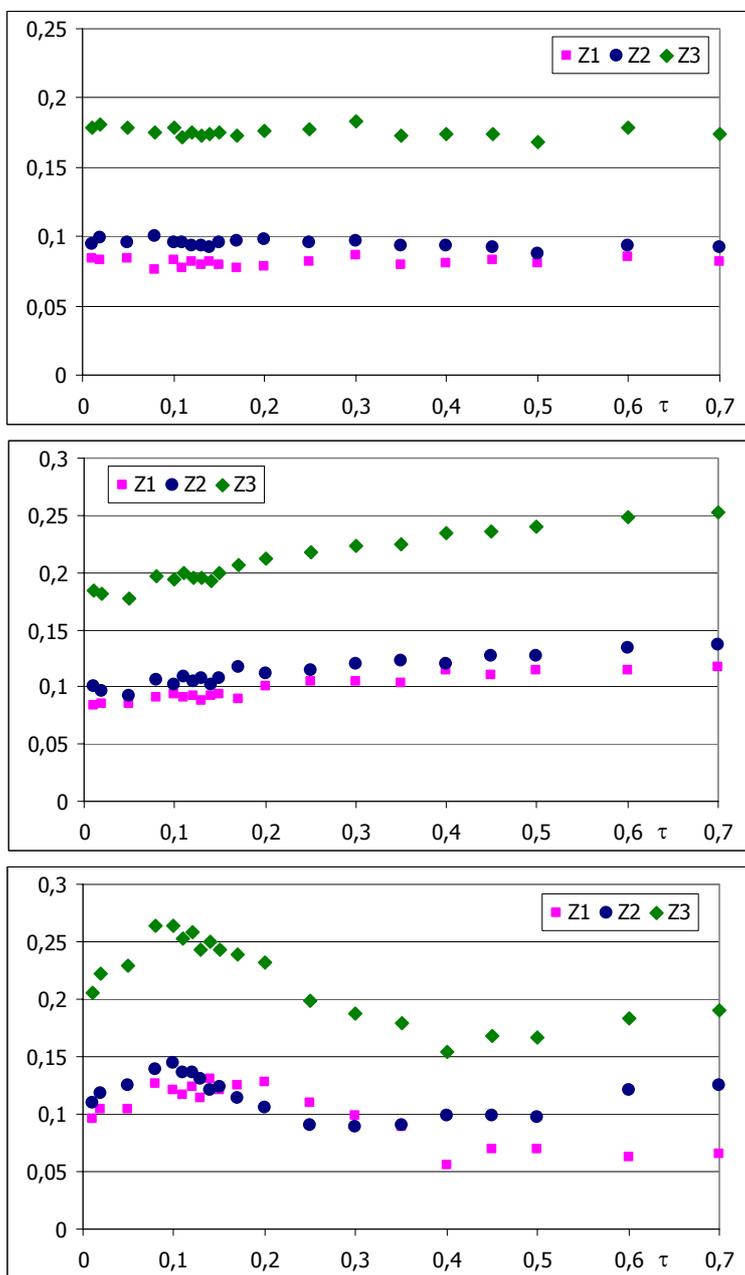


Рис. 8. Зависимости относительной ширины «коридора истины» от  $\tau$  для гауссова (8а, вверху), экспоненциального (8б, в середине) и Парето-шума (8с, внизу).

## 6. Заключение

Мы рассмотрели класс временных рядов, порождаемых простейшей динамической моделью с аддитивным шумом. В этом классе продемонстрирована возможность определения момента истины с заданной точностью на основе анализа одной (но достаточно длинной) реализации временного ряда. Предложен алгоритм определения момента истины. Он позволяет как определить «момент истины» в данном временном ряде, так и прогнозировать его наступление. В этом алгоритме используются лишь два признака: модернизированная дисперсия ( $D_4$ ) и ее производная по времени.

Однако, как сама величина  $D_4(P)$ , так и соответствующие им значения  $t_{\text{ист}}(P)$  определяются со сравнительно невысокой точностью. Поиск других (дополнительных) признаков, уточняющих определение  $t_{\text{ист}}(P)$ , в рамках рассматриваемой модели не имеет смысла. Дело в том, что в модели (2) фактически отсутствует долговременная память. Точнее, временной ряд, рассмотренный на конечном интервале, содержит информацию (т.е. помнит) о том, в каком состоянии находилась система, какова амплитуда шума в данный момент времени, и как она изменялась со временем. Однако, корреляция между близкими по времени флуктуациями отсутствует.

Кроме того, исследована зависимость решающего признака  $D_4(P)$  от характера шума и параметра  $\tau$ . Интерес представляет сопоставление шумов, распределенных по нормальному закону и по Парето, поскольку они существенно отличаются. Предложены и исследованы дополнительные признаки (характеристики)  $Z_1, Z_2, Z_3$ , которые позволяют определить характер шума. Показано, что в случае гауссова шума, они практически не зависят от  $\tau$ ; при этом различие между величинами  $Z_1$  и  $Z_2$  не велико. Те же величины в случае шума типа Парето существенно зависят от  $\tau$ , и при любых  $\tau$  разница между  $Z_1$  и  $Z_2$  заметно больше, чем в случае гауссова шума.

В целом, исследованные результаты характерны для простейшей модели. В моделях с более сложной динамикой, свойства, вообще говоря, будут иными. При аналогичном исследовании моделей другого класса, необходимо знать свойства простейшей модели. При сравнении, можно решить задачу определения момента истины в общем случае.

Авторы выражают благодарность проф. В.П. Карп за обсуждение и ценные замечания.

Работа выполнена при поддержке грантов РГНФ №04-03-00069а и НШ-2071.2003.4.

### Список литературы.

*Bouchaud J.-P., Cont R. A Langevin approach to stock market fluctuations and crashes // Eur. Phys. J. B. 1998, v.6, p. 543-550*

Dagpunar J. *Principles of random variate generation*. Clarendon Press, Oxford, 1988

Dubovikov M.M., Starchenko N.V., Dubovikov M.S. *Dimension of the minimal cover and fractal analysis of time series* // *Physica A*. 2004, v.339, No. 3-4, p. 591–608.

EFD Systems © MRNG.PAS [www.home.mindspring.com](http://www.home.mindspring.com) (freeware)

ESB Consultancy © AMRandom.pas <http://www.esbconsult.com> (freeware)

Friedrich R., Peinke J., Renner Ch. *How to quantify deterministic and random influences on the statistics of the foreign exchange market* // *Phys. Rev. Lett.* 2000, v. 84, No. 22, p.5224–5227

Gradišek J., Friedrich R., Govekar E., Grabec I. *Examples of analysis of stochastic processes based on time series data* // *Meccanica*. 2003, v. 38, p. 33–42.

Lohninger H. RandGen © [www.lohninger.com](http://www.lohninger.com) (freeware)

Mariani S., Rotundo G., Tirozzi B. *Hedging strategy with Langevin evolution* // *Int. J. Theoretical and Applied Finance*. 2000, v. 3, No. 3, p.553-556.

Siegert S., Friedrich R., Peinke J. *Analysis of data sets of stochastic systems* // *Physics Letters A*. 1998, v. 243, p. 275–280.

Sornette D. *Fokker-Planck equation of distributions of financial returns and power laws* // *Physica A*. 2001, v.290, p.211–217.

Tessone C.J., Toral R. *System size stochastic resonance in a model for opinion formation*. // *Physica A*. 2005, v. 351, p. 106-116.

Белайчук Л.В., Малинецкий Г.Г., Проделки джокеров на одномерных отображениях. Препринт ИПМ №24, 1997.

Колупаев А.Г., Чернавский Д.С. *Перемешивающий слой* // *Краткие сообщения по физике*. 1997, №1-2, с.12–18.

Курдюмов С.П., Малинецкий Г.Г., Чернавский Д.С., *Режимы с обострением, джокеры, перемешивающий слой. Новый взгляд на нелинейную динамику*. Препринт ИПМ, 2005. Представлено в ДАН.

Малинецкий Г.Г., Потапов А.Б. *Джокеры, русла и третья парадигма / Сб. Синергетическая парадигма: многообразие поисков и подходов*. Под ред. В.И. Аршинова, В.Г. Буданова, В.Э. Войцеховича. 2000. с. 138-156.

Сорос Дж. *Алхимия финансов*. – М.: ИНФРА–М, 1998.

Чернавский Д.С. *Синергетика и информация (динамическая теория информации)*. Изд. второе. М.: Эдиториал УРСС, 2004. 288 с.

Чернавский Д.С., Карп В.П., Родитат И.В., Никитин А.П., Чернавская Н.М. *Распознавание. Аутодиагностика. Мышление*. – М.: Радиотехника, 2004. – 272 с.

Чернавский Д.С., Колупаев А.Г., Веселова Т.В., Веселовский В.А. *Исследование перемешивающего слоя методом точечных отображений* // *Биофизика*. 2003, т.48, №2, с. 361–367.