



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 36 за 2017 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Орлов Ю.Н., Плешаков Р.В.

Программный комплекс для
моделирования
нестационарных
неэквидистантных
временных рядов

Рекомендуемая форма библиографической ссылки: Орлов Ю.Н., Плешаков Р.В. Программный комплекс для моделирования нестационарных неэквидистантных временных рядов // Препринты ИПМ им. М.В.Келдыша. 2017. № 36. 15 с. doi:[10.20948/prepr-2017-36](https://doi.org/10.20948/prepr-2017-36)
URL: <http://library.keldysh.ru/preprint.asp?id=2017-36>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

Ю.Н. Орлов, Р.В. Плешаков

**Программный комплекс
для моделирования нестационарных
неэквидистантных временных рядов**

Москва — 2017

Орлов Ю.Н., Плешаков Р.В.

Программный комплекс для моделирования нестационарных неэквидистантных временных рядов

Описывается методика построения ансамбля траекторий временного ряда с нестационарным потоком событий и нестационарным же выборочным распределением значений. Описывается программный комплекс, решающий эту задачу. Приводятся примеры его практического применения.

Ключевые слова: нестационарный временной ряд, нестационарный поток событий, моделирование ансамбля траекторий

Orlov Yu.N., Pleshakov R.V.

Numerical simulation of non-stationary non-equidistant time-series modeling

The method of generation of non-stationary ensemble of trajectories for time-series with non-equidistant probability flow is formulated. The corresponding simulation program with practical examples is constructed.

Key words: non-stationary distribution function, non-equidistant time-series

Работа выполнена при поддержке гранта РФФИ, проект № 17-01-00361

Содержание

Введение	3
1. Структура программного комплекса.....	3
2. Метод генерации неэквидистантного временного ряда.....	7
3. Алгоритм моделирования нестационарного потока событий	10
4. Алгоритм вычисления показателя Херста	11
5. Примеры вычислительных экспериментов	12
Заключение.....	14
Литература	15

Введение

В работах [1-4] представлена модель прогнозирования выборочной функции распределения нестационарного временного ряда на некоторый горизонт, определяемый уровнем нестационарности ряда. Уровень нестационарности представляет собой специальную статистику, собираемую по встык-выборкам заданной длины, в виде распределения расстояний между выборочными распределениями в норме C или $L1$. Итогом указанных работ явилось создание программного комплекса, который генерирует ансамбль траекторий временного ряда, распределение которого эволюционирует в соответствии с кинетическим уравнением, сохраняющим нормировку и отвечающим наблюдаемым свойствам ряда: сохранению тренда или изменению его на противоположный. Временем в этих работах считался порядковый номер события, т.е. наблюдение за случайным процессом осуществлялось через постоянные промежутки времени.

На практике часто встречаются ситуации, когда промежутки времени сами представляют собой некоторый случайный процесс. Это специфика систем массового обслуживания, частным случаем реализации которой является динамика биржевых сделок [5-7]. Цена купли-продажи финансового инструмента и промежутки времени между двумя последовательными актами купли-продажи представляют собой два измерения, характеризующие данный временной ряд. Аналогичными свойствами обладают ряды длительностей телефонных или интернет-соединений, последовательности магнитуд землетрясений и другие случайные события, моменты наступления которых, как и их значения, случайны.

Генерация ансамбля траекторий неэквидистантного нестационарного временного ряда является, таким образом, практически важной задачей, решение которой позволит моделировать различные функционалы управления наблюдаемым случайным процессом и проводить их оптимизацию. В данной работе представлен программный комплекс, реализующий эту задачу.

1. Структура программного комплекса

В данном разделе содержатся сведения о программном комплексе для моделирования и расчета статистик нестационарных неэквидистантных временных рядов.

1. Общие сведения

1.1 Наименование программы – «Модуль для моделирования и расчета статистик для нестационарных неэквидистантных временных рядов» MNTS.

1.2 Для функционирования комплекса требуется .NET 4.0.

1.3 Система разработана с использованием языков программирования C#/C++.

2. Функциональное назначение комплекса

2.1 Комплекс предназначен для построения множества нестационарных временных рядов, обладающих свойствами, характерными для заданного ряда либо набора рядов. Реализована возможность расчета индикативных статистик для нестационарных временных рядов со случайным распределением промежутков времени между последовательными событиями.

2.2 Модуль MNTS реализует следующие основные функции:

- генерация заданного количества нестационарных временных рядов с нестационарным потоком событий;
- расчет различных статистик для набора временных рядов;
- расчет средних значений, дисперсий и собственно выборочных распределений для значений функционалов, заданных вдоль траектории таких двумерных временных рядов.

2.3 Функциональных ограничений по описываемым операциям не имеется.

3. Описание логической структуры

3.1 Комплекс состоит из следующих основных компонентов:

- проверка валидности входных данных;
- расчет маски плотности тиков в зависимости от времени;
- расчет плотностей распределения тиковых приростов;
- расчет плотности распределения длительностей подряд идущих тиков величиной наиболее вероятного прироста;
- блок статистических функций;
- блок возможных функций финансовой математики для построения статистик;
- построение стационарного ряда для наиболее вероятного по модулю приращения;
- построение нестационарных рядов на основании масок распределений тиковых приращений;
- совмещение стационарной и нестационарной компонент.

3.2 В основе работы модуля генерации неэквидистантных временных рядов лежит метод, основанный на декомпозиции рассматриваемого временного ряда на стационарную и нестационарную компоненты. На основе анализа большого количества биржевых временных рядов был сделан вывод о том, что ряд из тиков (значения цен отдельных сделок), приращение которых по модулю равно наиболее вероятному для рассматриваемого ряда, (в дальнейшем – базовые тики) является стационарным. А направленное движение создается за счет менее частых тиков большего прироста. Количество базовых тиков между ними является индивидуальной характеристикой временного ряда, для каждого ряда строится плотность распределения их количества, т.е. длительности базовой серии. Для тиков, не являющихся базовыми, строится плотность распределения размеров прироста. Далее на основании полученных распределений строится набор временных рядов. Схема программы

представлена на рис.1. На рис.2 представлен список реализованных функций в данном модуле.

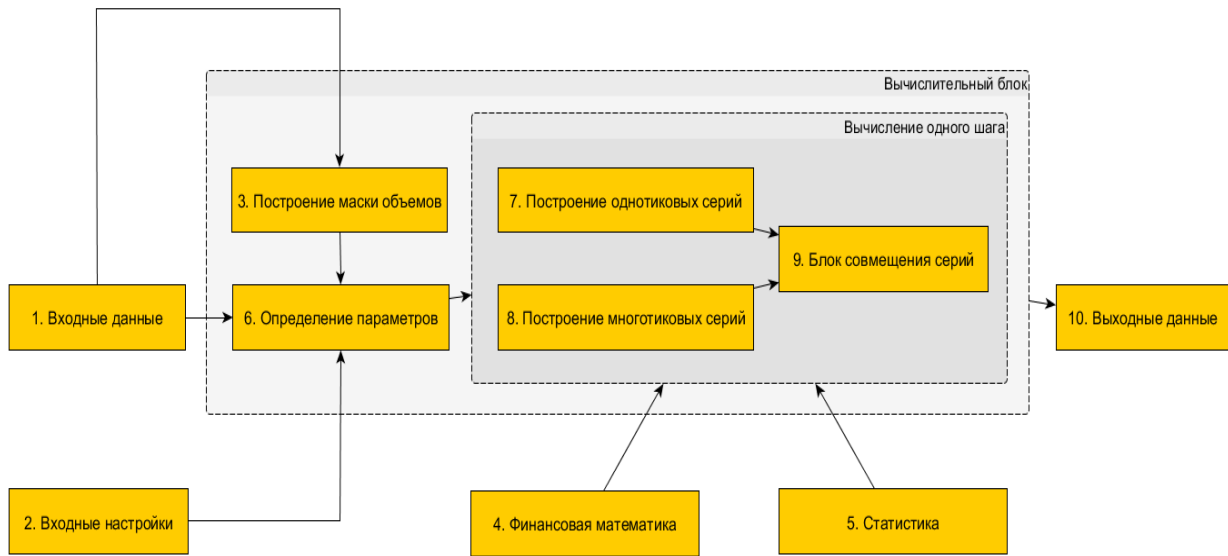


Рис. 1. Схема модуля генерации временных рядов

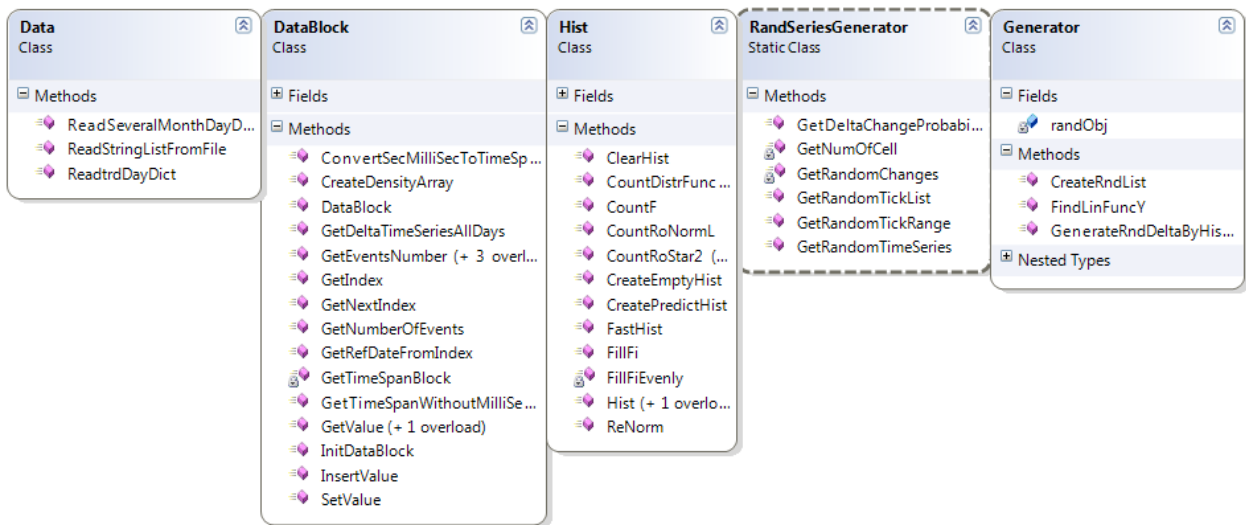


Рис. 2. Перечень функций модуля генерации рядов, предоставленный средой разработки VisualStudio

3.3 Расчет статистик для нестационарного маркированного временного ряда осуществляется в 2 этапа. На первом строится матрица значений статистики, где по одному измерению располагается номер тика, а по другому – количество точек для расчета. На следующем шаге, в предположении, что поток событий является однородным и ординарным, а ряд моментов событий описывается распределением Пуассона, на основании полученной маски плотности тиков осуществляется переход от этой матрицы к результатам в терминах моментов и промежутков времени. На рис.3 изображена схема описываемого модуля, а на рис.4 представлены основные реализованные функции.



Рис. 3. Схема работы модуля расчета статистик

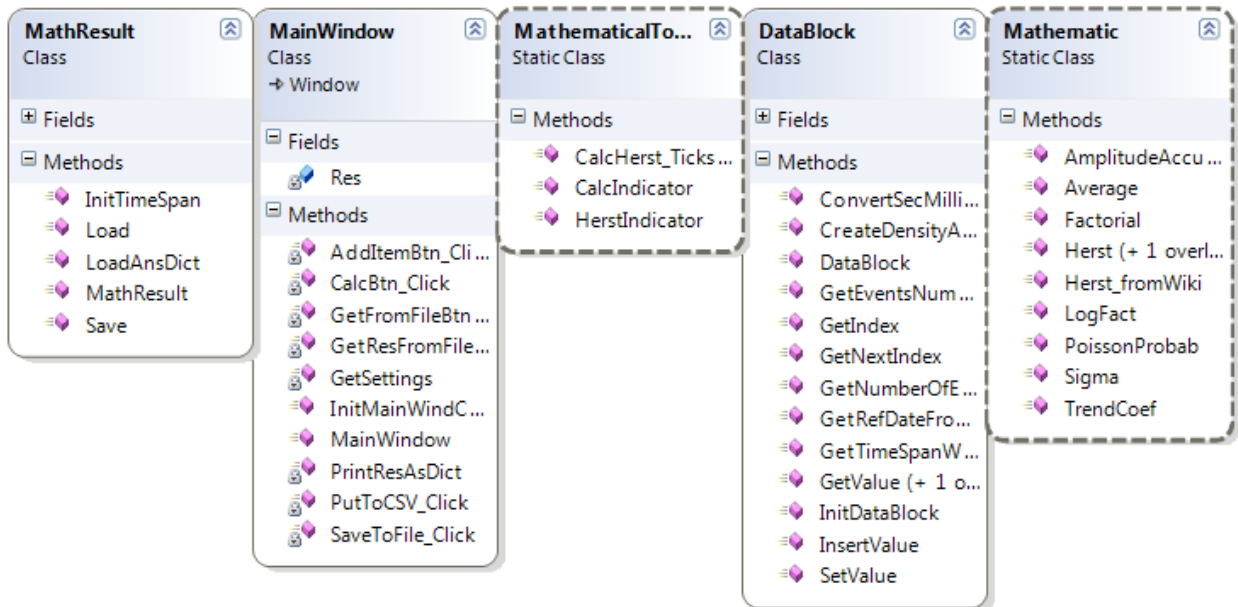


Рис. 4. Перечень функций модуля расчета статистик

4. Входные данные

4.1. Входными данными являются:

- временной ряд, представленный в формате набора записей со значениями случайной величины;
- моменты времени, в которые были зафиксированы эти значения.

4.2 Блок настроек для генерации пучка рядов включает в себя следующие опции:

- целевое количество рядов;

- временной промежуток для генерации;
 - плотность распределения тиковых приростов (опционально);
 - уровень наклона тренда у целевых рядов (опционально).
- 4.3 Для расчета статистик для выбора реализованы следующие индикаторы:
- уровень наклона тренда;
 - волатильность (различные варианты);
 - автокорреляция;
 - показатель Хёрста.

5. Выходные данные

5.1 Выходными данными в модуле генерации является набор полученных временных рядов.

5.2 Выходными данными в блоке расчета статистик является ряд с вычисленными значениями и моментами времени, в которых были эти значения получены.

5.3 Также в выходных данных находятся распределения статистик, посчитанные по полному файлу как функции от длины выборки.

2. Метод генерации неэквидистантного временного ряда

Генерация неэквидистантного нестационарного временного ряда основывается на следующих предположениях относительно структуры потока событий [5, 6]:

- существует некоторый промежуток времени, называемый периодом, внутри которого задана нормированная на единицу функция интенсивности потока;
- существует относительно небольшая часть (первые 10-15 % периода) промежутка времени, позволяющая оценить прогнозное число событий за период, так что фактически модель временного ряда строится на оставшемся участке периода после проведения соответствующих наблюдений за началом реализации процесса;
- абсолютные значения приростов случайной величины имеют распределение с «толстым хвостом», причем нестационарность присуща наиболее вероятному событию;
- рассматриваются серии из одинаковых абсолютных приростов цен последовательных сделок, исключая нулевые приросты: на один условный пункт (первая серия) и остальные (вторая серия), причем длина первой серии нестационарно прерывается второй серией, длительность которой является стационарным случайным процессом;
- распределение трендовых движений по промежуткам времени является стационарным случайным процессом на временных отрезках, превосходящих по длительности 30 минут, собственно тренд реализуется перекосом вероятности приращений цены принимать положительные или отрицательные значения.

Сделанные предположения позволяют построить модель временного ряда, обладающего свойствами, близкими к наблюдаемым на практике.

На первом этапе подготовки данных для моделирования траектории временного ряда собираются статистики:

- функции распределения $F(\theta)$ экспертно отбираемых трендовых движений цены по длительности времени θ общего движения траектории цены вверх или вниз;
- вероятности P^\pm положительного и отрицательного прироста цены на экспертно отбираемых фрагментах трендовых движений, $P^+ + P^- = 1$;
- параметр нестационарного пуассоновского потока событий $\Lambda(t, \tau)$ на промежутке времени $\Delta_t(\tau) = [t - \tau; t]$ внутри периода T (применительно к биржевым рядам это одна торговая сессия);
- функция распределения $G_{x>1}(n)$ серий приращений, абсолютная величина которых x более 1 условного пункта, в зависимости от числа n событий;
- совместная плотность распределения $f_{x=1}(k, k'; K, t)$ длин k и их приращений k' для серий абсолютных приращений цены на один условный пункт по выборке длины K событий в момент времени t .

По собранным статистикам определяется вероятность $p_k(t - \tau, t)$ числа k событий на промежутке времени $\Delta_t(\tau)$ по формуле

$$p_k(t - \tau, t) = \frac{(\Lambda(t, \tau))^k}{k!} \exp(-\Lambda(t, \tau)), \quad \Lambda(t, \tau) = \tau \mu(t - \tau, t). \quad (1)$$

Введенная здесь величина $\mu(t - \tau, t)$ называется интенсивностью потока на промежутке $\Delta_t(\tau)$. Это среднее число событий на указанном промежутке. Оно определяется формулой

$$\mu(t - \tau, t) = \frac{1}{\tau} \sum_{k=1}^{\infty} k p_k(t - \tau, t). \quad (2)$$

Считаем, что события независимы, а поток ординарный. Время агрегирования событий полагаем равным $\tau = 1$ мин.

Далее задается ожидаемое число N событий на временном горизонте T моделирования временного ряда. Оно необходимо для того, чтобы провести нормировку профиля интенсивности $\Lambda(t, \tau)$ на это число событий.

На следующем этапе из распределения $F(\theta)$ генерируется случайный ряд чисел θ_k в единицах измерения времени, принятом в параметре потока, так, что

$$\sum_k \theta_k = T. \quad (3)$$

Условие (3) определяет суммарное количество M макродвижений вверх и вниз и их длительность, на каждом промежутке θ_k задается вероятность P_k^+ движения цены в отдельном событии вверх и тем самым определяется вероятность $P_k^- = 1 - P_k^+$ движения цены вниз.

Затем генерируются случайные целые числа n_j из распределения (1), которые дают числа событий в течение 1 минуты на промежутках $\Delta_1(j)$, где j есть номер текущей минуты. Для данной генерации находится число сделок

$$\tilde{N} = \sum_{j=1}^T n_j. \quad (4)$$

Это число \tilde{N} в общем случае отлично от заданного изначально N , но выборочное среднее числа сделок за период по ансамблю траекторий сходится к числу N при увеличении числа траекторий.

Далее генерируется выборка чисел ± 1 общей длины \tilde{N} из кусочно-стационарного распределения вероятностей P_k^\pm в соответствии со случайным числом макродвижений из (3). Эта выборка определяет знак приращения цены в отдельном событии.

Из плотности функции распределения $f_{x=1}(k, k'; K, t)$ по методике [3] находятся функции

$$\varphi_{x=1}(k; K, t) = \sum_{k'} f_{x=1}(k, k'; K, t), \quad u(k; K, t) \varphi_{x=1}(k; K, t) = \sum_{k'} k' f_{x=1}(k, k'; K, t), \quad (5)$$

которые участвуют в построении уравнения Лиувилля для моделирования эволюции распределения $\varphi_{x=1}(k; K, j)$ из промежутка времени $\Delta_1(j)$ в промежутки $\Delta_1(j+1)$:

$$\begin{aligned} \varphi_{x=1}(k; K, j+1) &= \varphi_{x=1}(k; K, j) + \\ &+ \varphi_{x=1}(k-1; K, j)u(k-1; K, j) - \varphi_{x=1}(k; K, j)u(k; K, j). \end{aligned} \quad (6)$$

Таким образом, из формулы (6) становятся известны нестационарные распределения длин серий приростов на одну условную единицу. Функции $f_{x=1}(k, k'; K, t)$ вычисляются в скользящем окне длины K , так что на их вид влияют и выбранные на предыдущих этапах моделирования параметры потока и промежутки трендов вверх и вниз.

После того как вычисляются функции $\varphi_{x=1}(k; K, j)$, из них как из аналогов генеральных совокупностей строятся выборки длин $k_{1,j}, k_{2,j}, \dots$ в таком количестве, что их сумма равна прогнозируемому числу сделок из (4) либо превосходит их за счет ровно одного слагаемого:

$$\sum_i k_{i,j} = n_j.$$

Длины серий приростов на одну условную единицу прерываются сериями приростов на большие величины. Эти серии, как уже говорилось, имеют стационарное распределение $G_{x>1}(n)$ по длинам. Из этого распределения генерируется случайный набор целых чисел $n_{1,j}, n_{2,j}, \dots$, равных длинам серий указанного второго типа. Далее длины серий $k_{i,j}$ и $n_{i,j}$ чередуются до тех пор, пока их суммарная длина не станет равной n_j или превосходящей это число за счет последнего слагаемого. Затем начинается аналогичное построение в следующем промежутке времени $\Delta_1(j+1)$.

Знаки приростов во всех этих сделках определяются последовательностью случайных знаков ± 1 , которая генерировалась на предыдущих этапах моделирования.

Тем самым, построена модель «тиковой» траектории на протяжении одной торговой сессии.

3. Алгоритм моделирования нестационарного потока событий

На практике параметр потока $\Lambda(t, \tau)$ строится непосредственно по наблюдениям.

Пусть n_j есть число событий в j -ю минуту суток. Тогда число событий за рассматриваемые сутки есть

$$\tilde{N} = \sum_{j=1}^{1440} n_j, \quad (7)$$

а суточный нормированный профиль интенсивности определяется по формуле

$$y_j = \frac{n_j}{\tilde{N}}. \quad (8)$$

После этого среднее число событий (т.е. собственно параметр потока) на промежутке $\Delta_t(\tau)$ может быть введено через построение средневзвешенного профиля суточной активности за некоторый период в T суток. Для этого вводится средняя интенсивность $w(j)$ в j -ю минуту и среднее число тиков N за сутки. Затем определяется средневзвешенный нормированный профиль активности:

$$z(m) = \frac{w(m)}{N}. \quad (9)$$

Пусть, таким образом, ориентировочное число событий за сутки определено и равно N . Тогда среднее число тиков за промежутки τ (мин) до момента времени t (мин) есть

$$\Lambda^{(w)}(t, \tau) = \sum_{m=1}^{\tau} w(t-m+1) = N \sum_{m=1}^{\tau} z(t-m+1). \quad (10)$$

Поскольку профиль $z(m)$ нормирован на единицу, то его можно рассматривать как вероятность интенсивности по минутам в сутках. Его функция распределения есть

$$Z(m) = \sum_{k=1}^m z(k), \quad (11)$$

так что

$$\Lambda^{(w)}(t, \tau) = W \cdot (Z(t) - Z(t - \tau)) = \sum_{k=0}^{\tau-1} \Lambda^{(w)}(t - k, 1). \quad (12)$$

Тем самым определен модельный параметр потока событий для любых моментов времени t и промежутков τ .

4. Алгоритм вычисления показателя Херста

Показатель Херста для неэквилистантного временного ряда требует специального определения, поскольку необходимо статистику в терминах номера события преобразовать в статистику от временного промежутка, что не является однозначным процессом [6, 7].

Рассматривается одномерный временной ряд $r(n)$, где номер n некоторого элементарного события отвечает случайному моменту времени $t(n)$. В качестве ряда $r(n)$ рассматривается ряд первых разностей изучаемого процесса $r(n) = c(n+1) - c(n)$.

Показатель Херста для ряда $c(n)$ определяется следующим образом. Вводится скользящее среднее приростов $r(n)$ на шаге n по выборке длины k :

$$\bar{r}(n, k) = \frac{1}{k} \sum_{i=n-k+1}^n r(i) \quad (13)$$

и вычисляются накопленный размах отклонения от среднего

$$R(n, k) = \max_{j \leq n} \left(\sum_{i=n-k+1}^j (r(i) - \bar{r}(n, k)) \right) - \min_{j \leq n} \left(\sum_{i=n-k+1}^j (r(i) - \bar{r}(n, k)) \right). \quad (14)$$

и скользящая дисперсия по выборке длины k :

$$\sigma_r^2(n, k) = \frac{1}{k} \sum_{i=n-k+1}^n (r(i) - \bar{r}(n, k))^2. \quad (15)$$

Статистика

$$\xi(n, k) = \ln \left(\frac{R(n, k)}{\sigma_r(n, k)} \right) \quad (16)$$

является основной для вычисления показателя Херста $H_N(n)$, который определяется по выборке длины N на шаге n как коэффициент регрессии величины $\xi(n, k)$ на логарифм длины выборки:

$$H_N(n) = \frac{\sum_{k=1}^N (\xi(n, k) - \bar{\xi}_N(n)) \left(\ln k - \frac{\ln N!}{N-1} \right)}{\sum_{m=1}^N \left(\ln m - \frac{\ln N!}{N-1} \right)^2}. \quad (17)$$

В результате при каждой длине выборки N получаем набор показателей Херста на последовательности шагов с номерами n .

Для неэквилистантного временного ряда можно ввести несколько аналогов показателя Херста, совпадающих в случае эквилистантного ряда, но имеющих различное вероятностное содержание в общем случае [7]. В представленном программном продукте реализован следующий вариант.

Считаем, что выборки произвольных длин в (17) вплоть до некоторого фактического N_{\max} реализовывались на промежутке времени длины T ,

последнему событию в которой отвечает время $t(n)$. Тогда аналог показателя Херста определяется пересчетом показателя $H_N(n)$ с использованием вероятности числа событий, произошедших за время T . В соответствии с (1) полагаем, что

$$h_T(t) = \sum_{k=1}^N p_k(t-T, t) H_k(n(t)) / \sum_{m=1}^N p_m(t-T, t). \quad (18)$$

Формула (18) дает значение показателя Херста в пересчете на промежуток времени. Однако теперь его зависимость от T необязательно степенная, а определяется численно в зависимости от параметра потока.

5. Примеры вычислительных экспериментов

На рис. 5 представлен пример работы модуля генерации временных рядов. Синим цветом обозначен исходный ряд, остальные кривые – результаты работы программы.

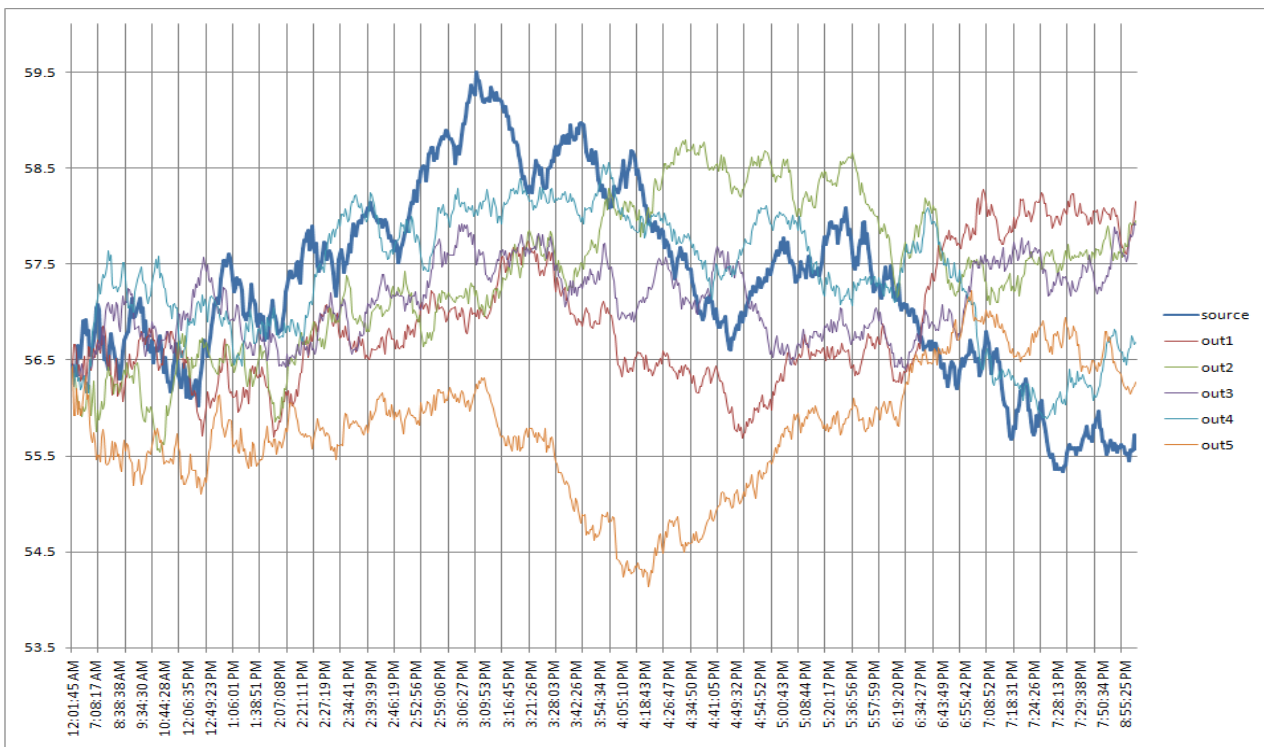


Рис. 5. Пример работы модуля генерации временных рядов

Пример динамики показателя Херста $H_N(n)$ тикового ряда в терминах номеров событий в окне $N = 20$ тыс. тиков, показан на рис. 6.

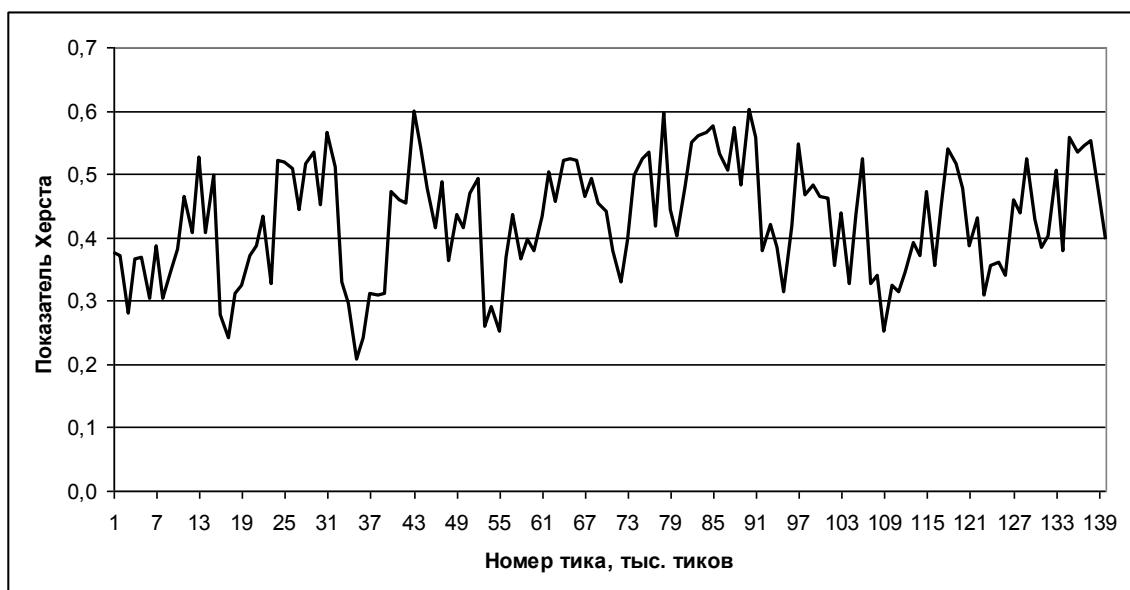


Рис. 6. Динамика показателя Херста $H_N(n)$ для $N = 20$ тыс. тиков

Из рис. 6 видно, что в разные моменты времени, трактуемые как номера событий, показатели Херста существенно отличаются. На большом наборе финансовых инструментов выяснилось, что хотя ряды цен нестационарны, временной ряд показателей Херста оказывается стационарным в узком смысле начиная с выборок длины примерно 300 точек. С увеличением окна N для вычисления $H_N(n)$ этот ряд становится все более стационарным. На длинах выборки показателей Херста более 1 тыс. точек с высокой точностью распределение этого показателя оказывается нормальным. Центр этого распределения находится приблизительно в точке до 0,5. Стандартное отклонение практически не зависит от N и составляет 0,09:

$$f_N(H) = \frac{1}{\sqrt{2\pi}\sigma_H} \exp\left(-\frac{(H-0,5)^2}{2\sigma_H^2}\right), \quad \sigma_H = 0,09. \quad (19)$$

Распределение показателей Херста для окна $N = 100$ тыс. тиков по выборке 20 тыс. значений приведено на рис. 7.

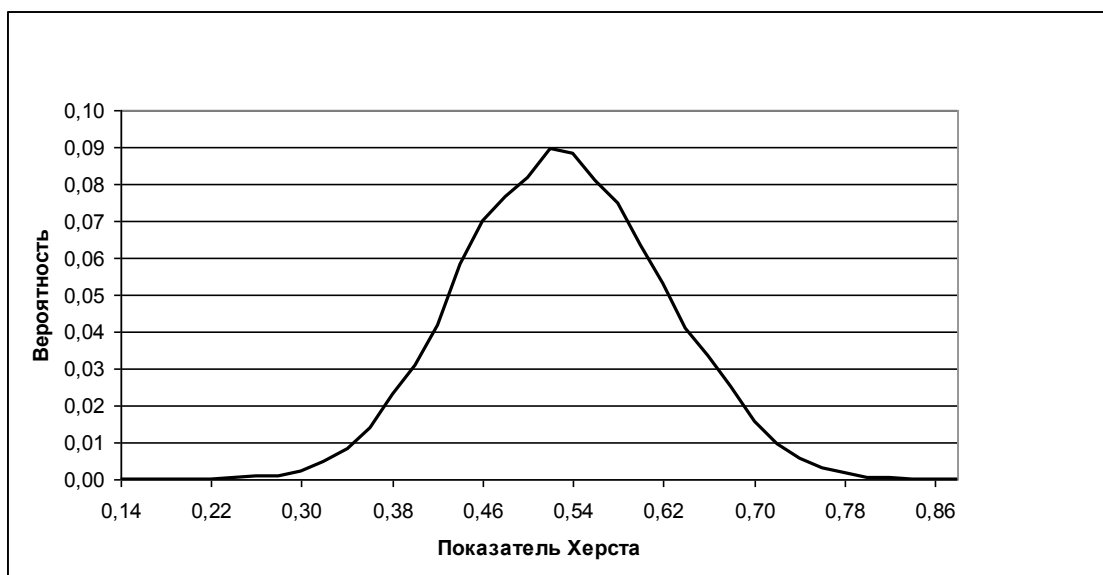


Рис. 7. Распределение показателей Херста $H_N(n)$ для $N = 100$ тыс. тиков

Следовательно, нестационарность многих рядов можно интерпретировать как стационарное броуновское блуждание показателя Херста. Выделение же стационарной составляющей как фильтрации вложения (см. [4]) позволяет построить математическую модель двумерного нестационарного временного ряда, возникающего во многих практических приложениях.

Заключение

Описанный программный комплекс позволяет моделировать нестационарную систему массового обслуживания, когда поток событий и само значение случайной величины являются нестационарными процессами. Помимо биржевых рядов, объектами моделирования могут быть собственно системы массового обслуживания, когда поток телефонных вызовов или запросов на посещение сайта и скачивание определенной информации имеет нестационарные характеристики. Для таких систем построенный комплекс позволяет провести оптимизацию функционала управления. Таким функционалом может быть алгоритм работы торговой системы на бирже, блокирование определенных запросов на сайт и т.п. Кроме того, комплекс позволяет собирать сложные нелинейные статистики по ансамблю траекторий, которые невозможно получить на практике по единственной реализации нестационарного временного ряда.

В дальнейшем предполагается обобщение разработанного метода на многомерный временной ряд, что потребует учета нестационарных выборочных корреляционных функций и исследования возможности распараллеливания алгоритмов.

Литература

1. Босов А.Д., Орлов Ю.Н. Кинетико-гидродинамический подход к прогнозированию нестационарных временных рядов на основе уравнения Фоккера-Планка // Труды МФТИ, 2012. Т. 3. № 4. С. 134-140.
2. Орлов Ю.Н., Федоров С.Л. Моделирование и статистический анализ функционалов, заданных на выборках из нестационарного временного ряда // Препринты ИПМ им. М.В. Келдыша. 2014. № 43. 26 с.
3. Орлов Ю.Н. Кинетические методы исследования нестационарных временных рядов. – М.: МФТИ, 2014. – 276 с. ISBN 978-5-7417-0521-6. ББК 22.172+22.317
4. Орлов Ю.Н., Федоров С.Л. Методы численного моделирования процессов нестационарного случайного блуждания. – М.: МФТИ, 2016. – 112 с.
5. Босов А.Д., Орлов Ю.Н., Федоров С.Л. О распределении рядов абсолютных приростов цен на финансовых рынках // Препринты ИПМ им. М.В. Келдыша. № 96, 2014. – С. 1-15.
6. Кириллов Д.С., Короб О.В., Митин Н.А., Орлов Ю.Н., Плешаков Р.В. Распределения показателя Херста нестационарного маркированного временного ряда // Препринты ИПМ им. М.В. Келдыша. 2013. № 11. 16 с.
7. Нуман Эльшейх М.Х., Огун Д.О., Орлов Ю.Н., Плешаков Р.В., Сакбаев В.Ж. Усреднение случайных полугрупп и неоднозначность квантования гамильтоновых систем // Препринты ИПМ им. М.В. Келдыша. 2014. № 19. 28 с.