**Выявление долгосрочной зависимости изменения курса доллара на основе R/S-анализа**

Ранее я показал, что [колебания курса рубля по отношению к доллару **не** подчиняются нормальному распределению](http://baguzin.ru/wp/?p=1760). Доказательство основывалось на том, что кривая распределения колебаний курса уж никак не может быть описана стандартной гауссианой. Существует однако еще один изящный метод, о котором я узнал из книги Бенуа Мандельброта [(Не)послушные рынки: фрактальная революция в финансах](http://baguzin.ru/wp/?p=1604) – R/S-анализ. Звучит, возможно, несколько пугающе, но не напрягайтесь, для изучения метода математика понадобится, но не бог весть какая… ☺

Напомню, что гауссова статистика основана на одном маленьком, но, как оказывается, очень важном допущении – на независимости изучаемых событий. Если 100 раз подбросить монетку, и она 100 раз упадет орлом, то математик здорово удивится, и начнет считать вероятность столь редкого события. Практик же скажет, что с монетой что-то не то…

R/S-анализ используется для проверки наличия в ряду данных долгосрочной зависимости. Для прояснения названия метода скажем, что буква R обозначает **Р**азмах, а S – **С**тандартное отклонение. О размахе, как одном из параметров статистики, я уже писал ранее в заметке [Контрольные карты Шухарта](http://baguzin.ru/wp/?p=236). Размах – разность максимального и минимального значения в выборке. Размах не так часто используется в статистике, поскольку на его величину значительно влияют разовые «выбросы». Тем не менее, в менеджменте качества это довольно важная характеристика. И вот мы встречаемся с этим понятием повторно. Таким образом, параметр R/S – это размах выборки, нормированный на стандартное отклонение по той же выборке. Если «выбросов» нет, то значение R/S мало, ну а если есть… Это проиллюстрировано на рис. 1. Среднее значение по выборке и стандартное отклонение на обоих графиках одинаково, а вот параметр R/S – разный.



Рис. 1. Влияние «выбросов» на значение параметра R/S

Одно из основных преимуществ R/S-анализа заключается в том, что в отличие от многих широко распространенных статистических критериев, он не основан на каких бы то ни было предположениях об организации исходных данных (о том, какому закону распределения они подчиняются). Это важнейший фактор, когда мы исследуем такие явления, как например, курс акций или валют, для которых явная ошибочность гауссовых подходов подтверждена многочисленными исследованиями. Формула R/S позволяет определить для различных периодов времени, будет ли размах б*о*льшим или меньшим того, какого можно ожидать в случае, когда каждый отдельный элемент исходных данных не зависим от предыдущего. Если разброс отличается от ожидаемого, то важна точная последовательность данных: череда прибыльных или убыточных моментов смещает экстремальные значения дальше, чем в случае их возникновения по чистой случайности.

Для начала выполним **R/S-анализ заведомо случайных данных**. Воспользуемся следующей моделью (см. Excel-файл «R-S-анализ случайных событий»; лист «10»). С помощью функции СЛЧИС(), генерирующей случайные числа в диапазоне от 0 до 1, создадим последовательность из –1 (для СЛЧИС() < 0,5) и +1 (для СЛЧИС() ≥ 0,5). Этим мы создали временной ряд *х = х1, х2, … х10* (столбец А на рис. 2).

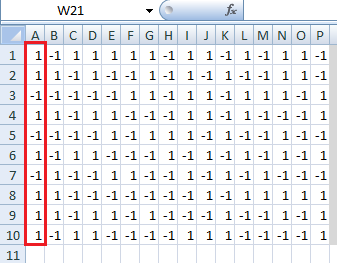


Рис. 2. Временной ряд *х* из 10 последовательных значений *х = х1, х2, … х10*

Среднее значение *хm* временного ряда *х* будет определяться как:

*xm = (х1 + х2 + … + х10)/n* (1)

Стандартное отклонение *Sn* вычисляется по формуле:

*Sn* = , t = 1, 2, … n (2)

Поскольку генератор случайных чисел не дает в точности равное количество отрицательных и положительных единиц (–1 и +1), то следует нормировать наш ряд *х* так, чтобы получить ряд, дающий среднее равное нулю. Для этого из каждого значения *xt* вычтем среднее по выборке *xm*:

*yt = xt – xm* t = 1, 2, … n (3)

Полученный ряд *y* имеет среднее равное нулю. На следующем шаге создадим кумулятивный временной ряд *z*:

*z1 = y1; zt = yt-1 + yt* t = 2, 3 … n (4)

На рис. 3 формула в ячейке Q2 реализует сразу три уравнения, описанные выше: (1), (3) и (4). Заметьте, что последний элемент ряда *z* (в нашем примере *z10*) всегда равно нулю, так как, по определению последний член нормированного кумулятивного ряда равен нулю.

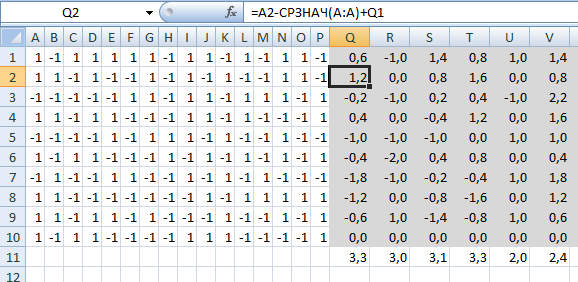


Рис. 3. Кумулятивный нормированный временной ряд

Размах *Rn*:

*Rn = max(z1, z2, … zt) – min(z1, z2, … zt)* (5)

Индекс *n* в последнем уравнении означает, что размах определяется для временного ряда, содержащего *n* элементов. Поскольку кумулятивный ряд *z* был скорректирован к среднему нулю, максимальное значение будет неотрицательным, а минимальное – неположительным. То есть, размах всегда больше нуля. Этот размах *R* является расстоянием, на которое перемещается система за время *n*. Для броуновского движения Эйнштейн обнаружил, что расстояние, которое проходит случайно блуждающая частица, увеличивается пропорционально квадратному корню из времени:

*R = T0,5* (6)

где R – пройденное расстояние, Т – время, в течение которого двигалась частица.

Для систем, которые (в отличие от броуновских) не являются независимыми, Херст предложил более общую формулу:

*(R/S)n = c\*nH* (7)

где n – длина (время) ряда; для нашей модели – число элементов ряда; с – константа; Н – показатель Херста, названный так Мандельбротом в честь автора формула (сам Херст использовал обозначение К). Видно, что для броуновского движения или любого иного процесса из независимых событий должны выполняться равенства с = 1, Н = 0,5. Если Н > 0,5, говорят, что процесс обладает долговременной памятью.

В ячейке Q11 (рис. 3) подсчитан параметр R/S:

=(МАКС(Q1:Q10)-МИН(Q1:Q10))/СТАНДОТКЛОНП(A:A)

Помимо временного ряда в столбце А я создал еще 15 временных рядов (столбцы с В по Р), а затем провел с ними описанные выше преобразования. Усреднение по этим 16 рядам позволило получить среднее значение (R/S)10 ≈ 3,3 ± 0,7.

На следующем шаге я создал аналогичные массивы для 20 элементов в каждом ряду, а затем для 40, 80, … 1280. Получилась следующая зависимость *R/S* от *n*. В таблице ниже каждое значение *R/S* получено путем усреднения по 16 временным рядам одинаковой длины.

|  |  |
| --- | --- |
| *n* | *R/S* |
| 10 | 3,4 |
| 20 | 4,7 |
| 40 | 7,5 |
| 80 | 10,7 |
| 160 | 15,4 |
| 320 | 21,2 |
| 640 | 28,0 |
| 1280 | 44,6 |

Если функцию *(R/S)n = f(n)* построить в логарифмических координатах по обеим осям (рис. 4), то решение уравнения (7) относительно *с* и *Н* находится с помощью линейной аппроксимации (метода наименьших квадратов), так что *с* – расстояние, отсекаемое прямой на оси *y*, а *Н* – угол наклона прямой к оси *х*.

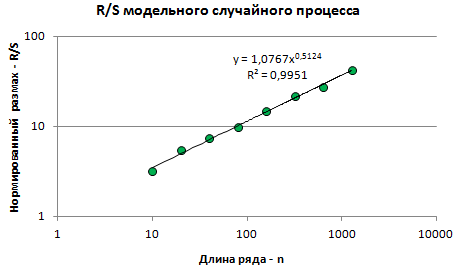


Рис. 4. Вид функции *(R/S)n = c\*nH* для ряда состоящего из случайных событий

На рис. 4 нанесены точки, полученные моделированием и степенная линия тренда, построенная Excel-ем самостоятельно. Видно, что *с* близко к единице (1,0767), а *Н* – к 0,5 (0,5124). К сожалению, ожидаемое для независимых событий Н = 0,5 у меня не получилось (рис. 5). Насколько отклонение в 0,03–0,04 значимо, для меня осталось невыясненным. Я даже решил проверить Excel-евский генератор случайных чисел (рис. 6). Но отклонений от ожидаемого среднего (нуля) не обнаружил (рис. 6). Либо в модели заложена какая-то неточность, либо генератор обладает долговременной памятью (!), либо 0,04 не является значимым отклонением от ожидаемого значения в Н=0,5.

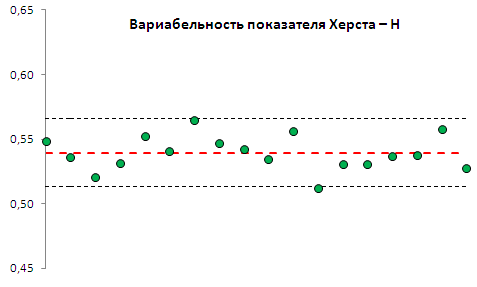


Рис. 5. Вариабельность эмпирически полученного показатель Херста *Н* для временного ряда из независимых событий

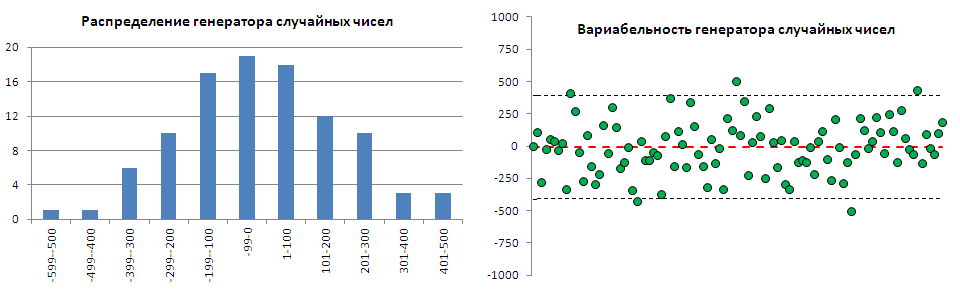


Рис. 6. Проверка генератора случайных чисел

Вооруженные теорией, **проведем теперь R/S-анализ курса доллара** (см. также Excel-файл «R-S-анализ курса доллара»). Все 4238 значений курса доллара за период с 1 января 1992 по 18 октября 2011 года я разбил сначала на интервалы по 10 значений, потом 20, 40, 80, 160, 320, 640, 1280. Последние значения, составляющие лишь часть интервала, я отбросил. Далее для того, чтобы отбросить не слишком много значений для длинных рядов – 640 и 1280 – я уменьшил длину этих рядов до 600 и 1050, соответственно. Проведя описанные выше в уравнениях (1) – (7) преобразования, я получил идеальную прямую для функции *(R/S)n = f(n)* – см. рис. 7

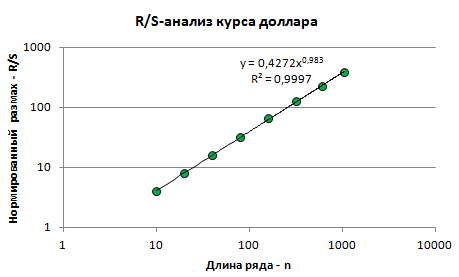


Рис. 7. Вид функции *(R/S)n = c\*nH* для курса доллара

Что можно сказать? Во-первых, точки идеально ложатся на прямую. Величина достоверности аппроксимации R2 = 0,9997 даже выше, чем для модельного ряда (рис. 4) R2 = 0,9951. Во-вторых, показатель Херста Н = 0,983, что существенно больше 0,5 для случайного блуждания. Очень красивый результат! Но что с ним делать? Можно ли найти ему практическое применение? Как и в [первом своем исследовании поведения курса доллара](http://baguzin.ru/wp/?p=1760) ответов у меня нет. Надеюсь, что пока нет…