

УДК 519.6:519.81

*Н.Д. Панкратова, Н.Г. Зражевская*

## МОДЕЛЬ АВТОКОРРЕЛЯЦИОННОЙ ФУНКЦИИ ВРЕМЕННОГО РЯДА С СИЛЬНОЙ ЗАВИСИМОСТЬЮ

### Введение

Модели временных рядов с сильной зависимостью имеют широкий круг применения: в финансовой математике, области телекоммуникаций, гидрогеологии и др. Наличие сильной зависимости может быть описано с помощью параметра Херста ( $H$ ) [1]. За последние годы разработана фундаментальная теоретическая база для определения этого важного параметра, а также предложено и исследовано достаточное количество практических алгоритмов [2]. Большинство из них основывается на оценке или автокорреляционной функции, или спектральной плотности, или выборочной дисперсии методами математической статистики и применением к ним аппарата регрессионного анализа.

В работе [3] приведены оценки параметра Херста, построенные на основе периодограммных оценок для спектральной плотности. Проведен их анализ на робастность. Результаты применены как к искусственно смоделированным временным рядам, так и к данным по объемам сетевого трафика. Используя оценку параметра Херста на основе дисперсионной модели, проведены исследования по определению сильной зависимости для разных нестационарных процессов, в частности, когда процесс содержит медленный тренд или скачки среднего [4].

Один из классических подходов к оценке параметра Херста основывается на эмпирическом законе Херста. Соответствующее направление в анализе сильнозависимых временных рядов называется  $R/S$ -анализом, общие положения которого описаны, например, в работах [2, 5, 6].

В то же время при реализации существующих моделей для реальных временных рядов при оценке параметра  $H$  часто возникают существенные затруднения. Оценки используемого параметра могут сильно отличаться в зависимости от модели. Таким образом, проблема выбора модели для исследования временного ряда на наличие сильной зависимости остается актуальной.

В данной статье рассмотрено восемь наиболее часто используемых методов оценивания параметра Херста [2] и предложена оптимизационная процедура, которая может существенно улучшить оценку параметра Херста для сильно зависимых волатильных временных рядов. В работе в целях отработки техники анализа построен и проанализирован искусственно сгенерированный, согласно модели FIGARCH, временной ряд с известными характеристиками. Затем предложенная процедура применена для получения оценок характеристик реального временного ряда индекса РТС — официального индикатора Российской фондовой биржи РТС («Российская Торговая Система») за 2014 г.

© Н.Д. ПАНКРАТОВА, Н.Г. ЗРАЖЕВСКАЯ, 2015

### Некоторые сведения о временных рядах с сильной зависимостью и проблемы, связанные с их исследованием

Пусть  $\{X_t, t = 0, 1, 2, \dots\}$  — временной ряд, стационарный в широком смысле. В литературе приведено несколько альтернативных определений сильной зависимости [1, 5]. Рассмотрим два из них.

*Определение 1.* Временной ряд  $\{X_t, t = 0, 1, 2, \dots\}$  с автокорреляционной функцией (АКФ)  $\rho(k) = \text{Cov}(X_t, X_{t+k})$ ,  $k \in N \cup \{0\}$ , подчиняется сильной зависимости (долгосрочной зависимости или зависимости с дальним радиусом), если существуют константы  $0 < \alpha < 1$  и  $c_r > 0$  такие, что

$$\lim_{k \rightarrow \infty} \rho(k)/(c_r k^{-\alpha}) = 1. \quad (1)$$

Таким образом, автокорреляционная функция такого ряда медленно спадает, и это соответственно приводит к расхождению ряда  $\sum_k \rho(k) = \infty$ . В противном случае говорят, что процесс подчиняется слабой зависимости.

От корреляционной функции  $\rho(k)$  с помощью преобразования Фурье можно

перейти к спектральной плотности  $f(\omega) = \frac{1}{2\pi} \sum_{k=1}^{\infty} \rho(k) e^{ik\omega}$  и сформулировать определение сильной зависимости в следующем виде.

*Определение 2.* Временной ряд  $\{X_t, t = 0, 1, 2, \dots\}$  подчиняется сильной зависимости, если существуют параметр  $0 < \beta < 1$  и константа  $c_f > 0$  такие, что

$$\lim_{\omega \rightarrow 0} f(\omega)/(c_f |\omega|^{-\beta}) = 1. \quad (2)$$

В работе [5] доказана эквивалентность определений 1 и 2 ((1) и (2)), а также указывается явный вид связи констант  $c_r$  и  $c_f$ ,  $\alpha$  и  $\beta$ . В частности, показано, что  $\alpha = 2 - 2H$  и  $\beta = 2H - 1$ . Из ограничений на параметры  $\alpha$  и  $\beta$  следует, что параметр Херста для рядов с сильной зависимостью удовлетворяет неравенству  $1/2 < H < 1$ .

Исходя из приведенных определений, понятие сильной зависимости можно рассматривать в двух аспектах. Сильнозависимые ряды имеют достаточно высокую корреляцию между своими значениями в моменты времени  $t_0$  и  $t_0 + T$  для достаточно больших  $T$ . Значит, для их исследования нужно рассматривать автокорреляционные функции с большим временным лагом, что накладывает существенные ограничения на выборки данных. Кроме того, в пространстве параметров Фурье для этих рядов характерны значительные частотные колебания в окрестности нуля, что приводит к появлению больших погрешностей при вычислениях. Поэтому, несмотря на достаточно полное теоретическое исследование сильнозависимых рядов, практический их анализ часто вызывает определенные трудности.

#### Исходная информация (реальные и сгенерированные данные), на которой выполняется апробация моделей

Для генерации временных рядов существует достаточно большое количество моделей. В данной работе используется модель FIGARCH, которая широко применяется при необходимости моделирования долгосрочной зависимости в гетероскедастических рядах.

Модель FIGARCH  $(p, d, q)$  в общем виде запишем следующим образом [7, 8]. Временной ряд  $X_t = u + \varepsilon_t$ , где  $u = \text{const}$ , подчиняется указанной модели, если  $(1 - \alpha(L) - \beta(L))(1 - L)^{-1}(1 - L)^d \varepsilon_t^2 = \omega + (1 - \beta(L))v_t$ , где  $\sigma_t^2$  — условная дисперсия процесса  $\varepsilon_t$ ,  $v_t = \varepsilon_t^2 - \sigma_t^2$  — независимые, одинаково распределенные случайные величины (н.о.р.с.в.),  $L$  — оператор сдвига  $L\varepsilon_t = \varepsilon_{t-1}$ ,  $\alpha(L) = \sum_{i=1}^p \alpha_i L^i$ ,  $\beta(L) = \sum_{i=1}^q \beta_i L^i$ ,  $(1 - L)^d = \sum_{j=1}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} (-L)^j$  — оператор дробной разницы ( $\Gamma(\cdot)$  — гамма-функция). Необходимое условие стационарности модели:  $\omega > 0$ ,  $\sum_{i=1}^p \alpha_i + \sum_{i=1}^q \beta_i < 1$ . Достаточное условие сильной зависимости:  $d \in (0, 1/2)$  (параметр Херста связан с  $d$ :  $H = 1/2 + d$ ).

Для генерации данных рассматривалась модель FIGARCH (1; 0,42; 1) с параметрами  $\omega = 0,1$ ;  $\alpha_1 = 0,1$ ;  $\beta_1 = 0,1$ .

### Реальные данные

В качестве реальных данных выбраны доходы на дневной основе индекса РТС. Индекс РТС — общепризнанный показатель состояния российского фондового рынка, который рассчитывается в режиме реального времени в течение всей торговой сессии биржи РТС при каждом изменении цен акций, которые включены в список для его расчета.

Значение индекса РТС определяется как отношение суммарной рыночной капитализации акций, включенных в список для расчета индекса, к суммарной рыночной капитализации на начальную дату, умноженное на значение индекса на начальную дату и на корректирующий коэффициент. Изменение стоимости акций каждой компании влияет на величину индекса РТС. При этом чем больше доля компании в индексе, тем сильнее ее влияние. В базу индекса РТС входит 50 наиболее капитализированных и ликвидных акций. Среди крупнейших представлены: Газпром (15,00 % в индексе), Лукойл (15,00 %), Сбербанк России (13,59 %), Роснефть (9,03 %).

Индекс РТС используется для отслеживания общего состояния рынка акций страны, вехи его развития и перспективы движения в будущем. Кроме того, индекс составляется таким образом, чтобы максимально адекватно отразить структуру экономики государства. Поэтому вполне логично, что в индексе РТС большая часть приходится на добывающие и нефтегазовые компании.

Заметим, что в дальнейшем реальные данные для краткости обозначаются как «индекс РТС», хотя в действительности используются данные квадратов доходов индекса РТС.

### Оценивание параметра Херста как меры сильной зависимости в наиболее часто применяемых методах

**Метод выборочной дисперсии агрегированного ряда.** Рассмотрим ряд  $\{X_t^{(m)}, t \in T\}$ , полученный из  $\{X_t, t \in T\}$  путем усреднения по блокам длины  $m \in \mathbb{N}$ :  $X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i)$ , где  $k \geq 1$  индексирует блок. Переход к ряду  $\{X_t^{(m)}, t \in T\}$  называется агрегацией по шкале времени с параметром  $m$ . Для агре-

гированных временных рядов  $\{X_t^{(m)}, t \in T\}$  с сильной зависимостью для достаточно больших  $m$  имеет место формула  $\text{var}(X^{(m)}) \sim \frac{\text{var}(X)}{m^\alpha}$ . Параметр  $H = 1 - \frac{\alpha}{2}$  можно отыскать, если построить агрегированный процесс на различных уровнях агрегации  $m$  и вычислить дисперсию каждого уровня.

Итак, для заданного  $m$  разбиваем выборку  $X_1, X_2, \dots, X_N$  на  $\frac{N}{m}$  блоков размерности  $m$ , вычисляем  $X^{(m)}(k)$  для  $k = 1, 2, \dots, N/m$  и выборочную дисперсию  $\text{var} X^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} (X^{(m)}(k))^2 - \left( \frac{1}{N/m} \sum_{k=1}^{N/m} (X^{(m)}(k)) \right)^2$ . Повторяем процедуру

для различных значений  $m$ . Строим график зависимости  $\log(\text{var}(X^{(m)}))$  от  $\log(m)$ , выбирая значения  $m = \{m_i, i \geq 1\}$  так, чтобы  $\frac{m_{i+1}}{m_i} = \text{const}$ . Для найденных значений

строим линейную регрессию. Наклон регрессивной линии на графике будет определять  $\alpha = 2H - 2$ .

**Метод дифференцирования дисперсии.** Для того чтобы отличать стационарные ряды с сильной зависимостью от рядов со слабой зависимостью, но со смещенным средним или с наличием слабо выраженного линейного тренда, в работе [4] предложена модификация рассмотренного выше метода — метода дифференцирования дисперсии. В этом методе вместо выборочной дисперсии используется ее разностная производная  $\text{var} X^{(m_{i+1})} - \text{var} X^{(m_i)}$ . Указанный метод часто используется в комбинации с базовым методом. Если построенный график выборочных дисперсий имеет экспоненциальную форму, то возникает подозрение в нестационарности ряда. В этом случае строят график для разностных производных выборочной дисперсии. Если ряд стационарный, то оба метода дают подобные результаты для всех  $H$ , от  $1/2$  до  $1$ . В случае нестационарности ряда и для  $H$ , близких к  $1/2$ , метод дифференцирования дисперсии будет давать лучший результат. В случае нестационарности и  $H$ , близких к  $1$ , количественно определить параметр Херста невозможно, но можно подтвердить наличие сильной зависимости.

**Метод абсолютных значений агрегированных рядов.** В этом методе вместо выборочных дисперсий  $\text{var} X^{(m)}$  вычисляются суммы абсолютных значений агрегированных рядов, а именно,  $S(m) = \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k)|$ . Если ряд имеет сильную зависимость с параметром  $H$ , то графиком линейной регрессионной зависимости  $\log S(m)$  от  $\log(m)$  будет прямая линия с наклоном  $H - 1$ .

Рассмотренные выше методы популярны, поскольку позволяют получить достаточно точные оценки при небольших объемах вычислений.

**Метод Хигуччи.** Данный метод используется для последовательности частичных сумм временного ряда  $Y(n) = \sum_{i=1}^n X_i$ , вычисляется статистика

$$L(m) = \frac{N-1}{m^3} \sum_{i=1}^m \left[ \frac{N-I}{m} \right]_i^{-1} \sum_{k=1}^{[(N-i)/m]} |Y(i+km) - Y(i+km-m)|,$$

где квадратные скобки обозначают целую часть числа. График  $\log L(m)$  от  $\log m$  — прямая линия с наклоном  $2 - H$ .

**Метод остатков регрессии.** Он включает следующие шаги:

1) разбивается выборка длины  $N$  на блоки длины  $m$ , которые попарно не пересекаются;

2) на каждом блоке находятся частичные суммы  $Y_k(n) = \sum_{i=1}^n X_i$ ,  $k = 1, 2, \dots, m$ ;

3) используя метод наименьших квадратов (МНК), строится по  $Y_k(n)$  линейная регрессия и находится выборочная дисперсия остатков;

4) описанная процедура повторяется на каждом блоке;

5) находится среднее полученных дисперсий  $\sigma^2(m)$ .

6) поскольку для больших  $m$   $\sigma(m) \sim m^{2H}$ , то на графике зависимости  $\log \sigma(m)$  от  $\log m$  прямая будет иметь наклон  $2H$ .

**R/S -метод или метод нормированного размаха.** Это один из первых и наиболее известных методов. Обозначим выборочную дисперсию как  $S^2(n)$ . Функция размаха  $R(n)$  определяется следующим образом:

$$R(n) = \max_{0 \leq t \leq n} \left( Y(t) - \frac{t}{n} Y(n) \right) - \min_{0 \leq t \leq n} \left( Y(t) - \frac{t}{n} Y(n) \right),$$

где  $Y(n) = \sum_{i=1}^n X_i$ .

Для рядов с сильной зависимостью  $E(R(n)/S(n)) \sim Cn^H$ ,  $n \rightarrow \infty$ , где  $C$  — положительная постоянная, не зависящая от  $n$ .

Для определения параметра Херста выборка длины  $N$  разбивается на непесекающиеся блоки длины  $K$ . Вычисляем для каждого  $n$   $R(k_i, n)/S(k_i, n)$ ,  $k_i = \frac{iN}{K} + 1$ ,  $i = 1, 2, \dots$ ,  $k_i + n \leq N$ . Значение  $H$  определяем в соответствии с линейной регрессией, построенной на точках  $(\log(R(k_i, n)/S(k_i, n)), \log n)$  для достаточно больших  $n$ . Основная проблема этого метода — выбор  $n$ . Если выбрать  $n$  недостаточно большим, будут доминировать краткосрочные корреляции, выбор слишком большого  $n$  ограничивается размером выборки.

**Метод периодограмм.** Данный метод использует оценку спектральной плотности вида

$$I(\lambda) = \frac{1}{2\pi N} \left| \sum_{j=1}^N X_j e^{ij\lambda} \right|^2, \text{ где } \lambda \text{ — частота. Поскольку периодограмма}$$

$I(\lambda)$  является оценкой спектральной плотности, то по определению 2 для сильно-зависимых рядов на графике  $\log(I(\lambda))$  от  $\log \lambda$  наклон прямой будет  $1 - 2H$ . Значения периодограммы рассчитываются для последовательности  $\lambda_k = 2\pi k/N$ , где  $k = 1, \dots, j$ , а целое число  $j$  выбирается меньшим, чем  $\frac{1}{2}N$ .

**Модифицированный метод периодограмм.** Учитывается, что большинство частот на графике расположены в правой части и, таким образом, существенно влияют на регрессионную прямую. В модифицированном методе ось частот разбита на одинаковые (в логарифмическом исчислении) блоки и находятся средние значения периодограмм внутри блоков. Значения для низких частот не учитываются.

## Построение моделей АКФ сильнозависимых временных рядов с использованием оценок параметра Херста

Нахождение параметра Херста в контексте сильной зависимости временных рядов может преследовать разные цели. Для построения модели временного ряда (например, FARIMA, FIGARCH [7, 8]) существенное значение имеет точность в пределах лага модели (в этом есть некоторые противоречия, поскольку теоретическое определение сильной зависимости требует бесконечной выборки и параметр  $H$  определяется первым членом асимптотического разложения АКФ). Если цель исследования — прогнозирование значений временного ряда, то существенными являются значения АКФ при больших сдвигах, в то время как модели АКФ, основывающиеся на параметре Херста, учитывают все значения эмпирической АКФ интегральным образом без приоритизации данных.

В настоящей работе сравниваются результаты, полученные с использованием трех различных моделей АКФ на основании регрессионной модели, которая следует из определения сильной зависимости (асимптотическое соотношение (1)):

$$\rho(k) = \alpha_1 H^* (2H^* - 1) k^{2H^* - 2} + \alpha_2 + \varepsilon_k, \quad \varepsilon_k \text{ — н.о.р.с.в., } k_0 \leq k \leq N, \quad (3)$$

где в качестве  $H^*$  берется полученная оценка. Модели различаются способом оценивания  $H^*$ . Линейная регрессия не включает первые  $k_0$  значений автокорреляции, поскольку (1) задает асимптотические свойства. Оценки для коэффициентов  $\alpha_1, \alpha_2$  могут быть найдены по МНК. Сравнение осуществляется на искусственно сгенерированных и реальных данных.

Рассмотрим построение первой модели в виде линейной регрессии АКФ в двойном логарифмическом масштабе:  $\log \rho(k) = b_1 + b_2 \log k + \varepsilon_k$ . Параметры модели  $b_1, b_2$  оцениваются по МНК. Эта модель не учитывает априорной информации о наличии сильной зависимости временного ряда. По полученной оценке  $\hat{b}_2$  можно определить оценку параметра Херста:  $\hat{b}_2 = 2\hat{H}_{\text{corr}} - 2$ . Значения АКФ и регрессионных прямых в двойном логарифмическом масштабе представлены на рис. 1 для искусственно сгенерированных данных и на рис. 2 — для реальных данных. Из рисунков видно, что такая модель не может быть достаточно адекватной для реальных данных, хотя для искусственно сгенерированных данных (т.е. с контролируемыми свойствами) она достаточно достоверна. Оценка параметра Херста этой модели обозначена  $\hat{H}_{\text{corr}}$ . Значения АКФ, полученные по первой модели автокорреляционной функции, представлены на рис. 3 и 4.

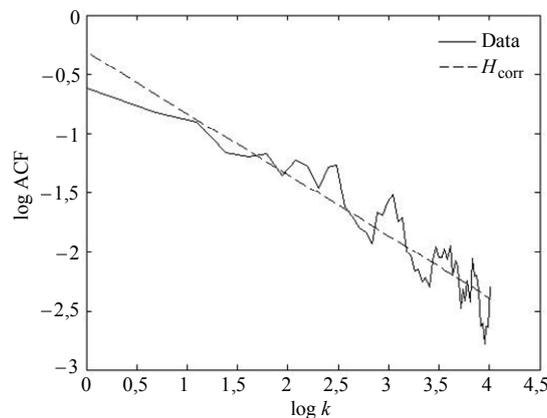


Рис. 1

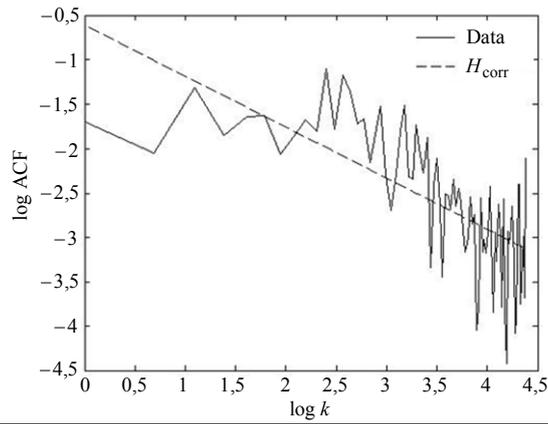


Рис. 2

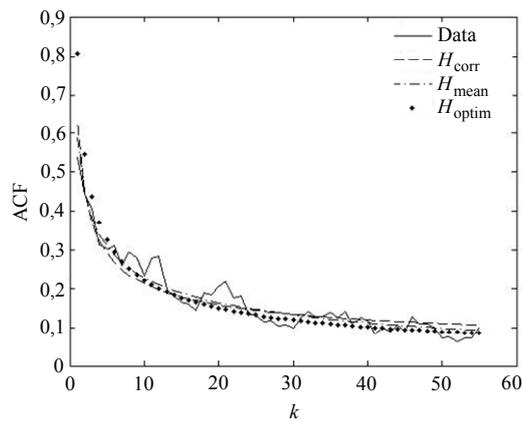


Рис. 3

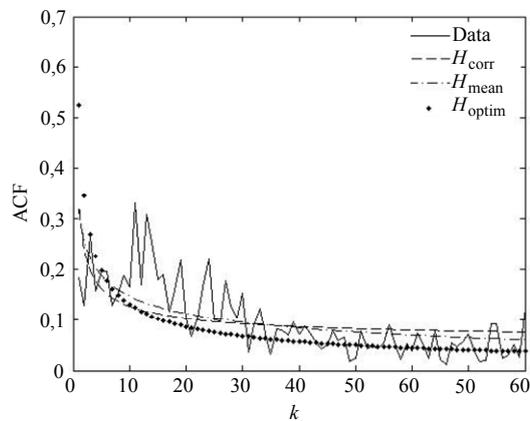


Рис. 4

При построении второй модели учитывается наличие сильной зависимости временного ряда. Эта дополнительная априорная информация позволяет использовать специализированные методы оценивания параметров модели. В данном случае задача моделирования сводится к оцениванию параметра Херста, который определяет модель временного ряда с сильной зависимостью. Как указывалось ранее, известен ряд методов оценивания параметра Херста, которые имеют свои особенности и могут быть более или менее эффективны в зависимости от дополнительных особенностей временных рядов. При построении этой модели для оценивания параметра Херста использовались описанные ранее методы (табл. 1). Из табл. 1 видно, что значения параметра  $H$ , полученные методом Хигуччи, модифицированным методом периодограмм и методом дифференцирования дисперсии,

имеют большой разброс значений ( $\approx 25\%$ ) для реальных данных (кроме того, параметр Херста не может превышать 1) и не могут использоваться для агрегации. Среднее по пяти остальным значениям  $\hat{H}_{\text{mean}} = 0,8692$  используется в (3) для построения модели АКФ. Соответствующие графики изображены на рис. 3 и 4. Как видно, построенная модель АКФ более адекватна, чем та, которая построена без учета сильной зависимости ( $\hat{H}_{\text{сорт}}$ ). Это следует из табл. 2, в которой приведе-

ны ошибки моделирования:  $ME = 1/N \sum_{i=1}^N (y_i - \hat{y}_i)$ ,  $MAE = 1/N \sum_{i=1}^N |y_i - \hat{y}_i|$ ,

$MSE = 1/N \sum_{i=1}^N (y_i - \hat{y}_i)^2$ ,  $MAPE = 1/N \sum_{i=1}^N |1 - y_i / \hat{y}_i| \cdot 100\%$  (средняя ошибка (ME),

средняя абсолютная ошибка (MAE), среднеквадратическая ошибка (MSE), средняя абсолютная процентная ошибка (MAPE),  $y_i$  — фактические значения автокорреляции,  $\hat{y}_i$  — значения автокорреляции модели).

Таблица 1

Методы оценивания	Оценка $H$		Относительная ошибка $H$	
	СД	РТС	СД	РТС
Абсолютных значений агрегированного ряда	0,8700	0,9285	2,3505	4,1750
Выборочной дисперсии агрегированного ряда	0,8053	0,8714	5,2633	2,2337
Модифицированный метод периодограмм	0,8711	1,0989	2,4807	23,2915
Дифференцирования дисперсии	0,8179	0,6768	3,7759	24,0613
Хигуччи	0,9994	1,0085	13,1455	17,5730
Остатков регрессии	0,8600	0,8849	1,1819	0,7173
Периодограмм	0,8692	0,8222	2,2574	7,7483
$R/S$	0,7586	0,8391	10,7536	5,8562

Таблица 2

Модель	ME		MAE		MSE		MAPE (%)	
	С.Д.	РТС	С.Д.	РТС	С.Д.	РТС	С.Д.	РТС
Модель 1	-0,02	0,0276	0,023	0,033	0,0007	0,0014	29,06	116,27
Модель 2	-0,01	-0,016	0,017	0,028	0,0004	0,0010	20,59	90,06
Модель 3	-9e-4	-0,008	0,014	0,021	0,0003	0,0008	16,68	49,64

Предложенная в данной работе модифицированная модель АКФ более гибкая и в то же время более наглядная. Действительно, гибкость второй модели АКФ с учетом наличия сильной зависимости обеспечивается выбором метода построения параметра Херста в силу его применимости и эффективности для конкретно взятых данных. В то же время анализ исходных данных в целях выбора метода зачастую более трудоемок, чем само построение модели. Кроме того, указанные методы (в стандартной формулировке) не предполагают приоритизацию данных в исходной выборке, что ограничивает их специализацию в зависимости от цели построения модели. Эти недостатки устраняются в третьей модели. Эта модель основывается на первых двух и ее качество, в силу построения, гарантировано не хуже качества первых двух моделей. Ее суть состоит в следующем. В соответствии со второй моделью определяется агрегированное значение оценки

параметра Херста и среднеквадратическое значение ошибки этой модели по исходной выборке. Затем на исходной выборке определяются две подвыборки. Способ разбиения может определяться целью моделирования и также является свойством третьей модели. На одной подвыборке строится регрессионная модель, соответствующая первой модели, с использованием МНК. На второй подвыборке определяются ошибки такой регрессионной модели, которые ограничиваются среднеквадратической ошибкой модели, построенной в соответствии со второй моделью. Такой подход удобно реализовать в виде оптимизационной процедуры.

### Модель АКФ, построенная с помощью оптимизационной процедуры

Пусть даны  $\{y_k\}_{k=1}^N$  — исходные значения автокорреляции, по которым с применением второй модели определена агрегированная оценка параметра Херста и средняя дисперсия ошибки этой модели  $\varepsilon^2$ . Определяются на полной выборке две подвыборки:  $\{y_k\}_{k=1}^{k_1}$  и  $\{y_k\}_{k=k_1+1}^{k_2}$ . На второй подвыборке строим регресси-

онную модель (3). Для этого достаточно минимизировать  $\sum_{k=k_1+1}^{k_2} (y_k - \hat{y}_k)^2$ , где  $\hat{y}_k$ ,

$k = \overline{k_1+1, k_2}$ , — значения модели, определяемые по (3). На первой подвыборке определяем квадратичные ошибки значений модели по отношению к фактическим значениям  $(y_k - \hat{y}_k)^2$ ,  $k = \overline{1, k_1}$ , и ограничиваем их средней дисперсией ошибки  $\varepsilon^2$ .

Очевидно, что такая задача минимизации квадратичного функционала с квадратичным ограничением выпуклая. Однако в силу того, что  $\varepsilon^2$  — средняя (а не максимальная) дисперсия, задача может не иметь допустимых решений. Для формулировки задачи с оптимальным решением достаточно ввести в правую часть ограничения релаксационный параметр  $q \geq 0$ , добавление которого в целевую функцию обеспечит минимальное отклонение оптимальной модели от модели, построенной в соответствии со второй моделью. Основным функционал и релаксационный параметр в целевой функции следует взвесить с коэффициентом штрафа  $\lambda$  ( $0 \leq \lambda \leq 1$ ). Варьирование  $\lambda$  позволяет получить желаемый баланс между соответствием модели исходным данным (на второй подвыборке) и соответствием модели условию сильной зависимости (на первой подвыборке).

Таким образом, формулировка оптимизационной задачи имеет вид:

$$\lambda q + (1 - \lambda) \frac{1}{k_2 - k_1} \sum_{k=k_1+1}^{k_2} (y_k - \hat{y}_k)^2 \rightarrow \min,$$

$$(y_k - \hat{y}_k)^2 \leq \varepsilon^2 + q, \quad k = 1, 2, \dots, k_1,$$

$$q \geq 0.$$

Случай  $\lambda = 0$  соответствует первой модели на  $\{y_k\}_{k=k_1+1}^{k_2}$ ,  $\lambda = 1$  — второй модели на  $\{y_k\}_{k=1}^{k_1}$ . Следовательно, третья модель содержит первые две модели как частные случаи и не менее адекватна, но более гибкая.

Третья модель построена на сгенерированных и реальных данных. Результаты решения оптимизационной задачи при различных значениях  $\lambda$  представлены

в табл. 3 (СД — значения для сгенерированных данных, РТС — значения для индекса РТС). В расчетах принято  $k_1 = 40$ ,  $k_2 = 55$ . Как видим из табл. 3, результаты решения оптимизационной задачи на сгенерированных данных существенно зависят от  $\lambda$ , что позволяет регулировать баланс между точностью модели на желаемом временном интервале и подчинением условиям сильной зависимости. В то же время реальные данные удовлетворяют условию сильной зависимости только с некоторой точностью, и, как следствие, результат решения оптимизационной задачи не зависит от  $\lambda$ . Следовательно, построенная для реальных данных модель соответствует некоторому идеальному временному ряду с сильной зависимостью, у которого АКФ близка к АКФ реального временного ряда на второй подвыборке.

Таблица 3

$\lambda$	Целевая функция		$q$		$\sigma^2$		$H$	
	СД	РТС	СД	РТС	СД	РТС	СД	РТС
0,1	0,0673	0,8094	0,0672	0,8645	0,0679	0,3140	0,8416	0,6770
0,2	0,0674	0,7544	0,0672	0,8645	0,0679	0,3140	0,8416	0,6770
0,3	0,0674	0,6993	0,0672	0,8645	0,0679	0,3140	0,8416	0,6770
0,4	0,0675	0,6443	0,0672	0,8645	0,0679	0,3140	0,8416	0,6770
0,5	0,0676	0,5893	0,0672	0,8645	0,0679	0,3140	0,8416	0,6770
0,6	0,0661	0,5342	0,0855	0,8645	0,0532	0,3140	0,8561	0,6770
0,7	0,0613	0,4792	0,1049	0,8645	0,0427	0,3140	0,8701	0,6770
0,8	0,0539	0,4241	0,1230	0,8645	0,0366	0,3140	0,8823	0,6770
0,9	0,0442	0,3691	0,1399	0,8645	0,0336	0,3140	0,8930	0,6770

Графики модели АКФ ( $\lambda = 0,5$ ) изображены на рис. 3 и 4 (кривые  $\hat{H}_{\text{optim}}$ ). Очевидно, они более приемлемы с точки зрения приближения фактических данных на второй подвыборке. Особенно это проявляется для реальных данных (см. рис. 4). Третья модель существенно лучше соответствует средним значениям фактической АКФ для больших сдвигов.

Таким образом, для сравнения предлагается три модели АКФ. Первая учитывает минимум информации и наименее достоверна. Во втором случае агрегируются модели, построенные по пяти наиболее распространенным методам оценки характеристик сильной зависимости. Третья модель, которая введена в данной работе, использует вторую модель, улучшая ее посредством использования процедуры прогноза и коррекции, предоставляет достаточное количество параметров для адаптации модели к использованию в различных задачах.

### Заключение

Настоящая работа посвящена построению модели, основанной на усовершенствованном методе оценивания параметров сильной зависимости временных рядов. Проведен обзор известных методов оценивания параметра Херста. В соответствии с полученными результатами моделирования (см. табл. 2), пять стандартных методов определения параметра Херста демонстрируют хорошие и относительно однородные результаты для искусственно сгенерированных данных, но имеют большой разброс для реальных данных. Основная причина в том, что сильная зависимость является формальным свойством, и ни один реальный процесс не удовлетворяет ей в полной мере. Кроме того, на характеристики сильной зависимости влияют другие неучтенные свойства, например нестационарность. Предложена новая модель, основанная на решении оптимизационной задачи для уточнения (или корректировки) оценки параметра Херста. Уточненная оценка основана на агрегации результатов нескольких стандартных моделей и определяется как результат применения метода прогноза и коррекции. Модель может быть адаптирована в зависимости от конечной цели оценивания, она верифицирована на сгене-

рированных по модели FIGARCH данных и применена к построению оценки параметра Херста для реальных данных индекса РТС. Показано, что для реальных данных предложенная модель является более адекватной для цели прогнозирования реального временного ряда, поскольку она наилучшим образом описывает сильную зависимость в асимптотическом смысле. Результаты работы могут использоваться для построения адаптивных робастных оценок при прогнозировании гетероскедастических временных рядов.

*Н.Д. Панкратова, Н.Г. Зражевська*

### МОДЕЛЬ АВТОКОРЕЛЯЦІЙНОЇ ФУНКЦІЇ ЧАСОВОГО РЯДУ З СИЛЬНОЮ ЗАЛЕЖНІСТЮ

Запропоновано модель автокореляційної функції часового ряду, що задовольняє умові сильної залежності, побудова якої базується на розв'язку оптимізаційної задачі, що дозволяє покращити оцінку параметра Херста. Модель може бути адаптована в залежності від кінцевої мети використання оцінки. Запропонована модель протестована на штучно згенерованих даних з відомими характеристиками і застосована до визначення параметра Херста часового ряду доходів індексу РТС. Розробка нової моделі пов'язана з тим, що в практичних застосуваннях за наявності нестационарності загальноприйняті оцінки параметра Херста можуть мати великий розкид значень.

*N.D. Pankratova, N.G. Zrazhevskaya*

### MODEL OF AUTOCORRELATIVE FUNCTION OF TIME SERIES WITH STRONG DEPENDENCE

The model, based on the optimization problem solving to improve the Hurst parameter estimation for time series with long-range dependence is proposed. The model can be adapted depending on the ultimate goal of the estimation. The proposed model was tested on artificially generated data with known characteristics and applied to the determination of the Hurst parameter of the time series of RTS incomes. Development of the new model is actual because of the fact that traditional Hurst parameter estimations [2] may have a long range of values in practical applications due to nonstationary effects.

1. *Palma W.* Long-memory time series: Theory and methods. — New Jersey : John Wiley & Sons, 2007. — 304 p.
2. *Taqqu M.S., Teverovsky V., Willinger W.* Estimators for long-range dependence: An empirical study // *Fractals*. — 1995. — P. 785–798.
3. *Teverovsky V., Taqqu M.S.* Semi-parametric graphical estimation techniques for long-memory data // *Lecture Notes in Statistics*. — 1996. — **115**. — P. 420–432.
4. *Teverovsky V., Taqqu M.S.* Testing for long-range dependence in the presence of shifting mean or a slowly declining trend, using a variance-type estimator // *Time Series Analysis*. — 1997. — **18**, N 3. — P. 279–304.
5. *Beran J.* Statistics for long-memory processes. — New York : Chapman and Hall, 1994. — 315 p.
6. *Clegg R.G.* A practical guide to measuring the Hurst parameter // *International Journal of Simulation: Systems, Science & Technology*. — 2006. — N 2. — P. 3–14.
7. *Baillie, R.T., Bollerslev, T., Mikkelsen, H.O.* Fractionally integrated generalized autoregressive conditional heteroskedasticity // *Journal of Econometrics*. — 1996. — N 74. — P. 3–30
8. *Бідюк П.І.* Системний підхід до прогнозування на основі моделей часових рядів // Системні дослідження та інформаційні технології. — 2003. — № 3. — С. 88–110.

*Получено 20.05.2015*