

Прикладная эконометрика, 2021, т. 62, с. 85–100.

Applied Econometrics, 2021, v. 62, pp. 85–100.

DOI: 10.22394/1993-7601-2021-62-85-100

Р. В. Гарафутдинов<sup>1</sup>

## Исследование влияния некоторых параметров модели ARFIMA на точность прогноза финансовых временных рядов

*В работе проводится анализ влияния параметров модели ARFIMA на точность прогнозирования финансовых временных рядов на примере искусственно сгенерированных рядов с длинной памятью и дневных лог-доходностей индекса РТС. Исследуемые параметры модели: отклонение величины порядка интегрированности от ее «истинного» значения, а также «длина» памяти, учитываемая моделью. По результатам исследования сформулированы некоторые практические рекомендации по моделированию с применением ARFIMA.*

**Ключевые слова:** ARFIMA; финансовые временные ряды; длинная память; показатель Херста; метод детрендированного флуктуационного анализа.

**JEL classification:** C22; C53; G17.

### 1. Введение

С середины XX века и до сих пор при объяснении поведения финансовых рынков в основном используется гипотеза эффективного рынка (Efficient Markets Hypothesis, EMH), выдвинутая Е. Ф. Фاما — нобелевским лауреатом 2013 г. Согласно ей, рынки стохастичны, и ценообразование на них подчиняется вероятностным законам. Хотя EMH входит в противоречие с практикой (которая, как гласит философия науки, является критерием истины), что неоднократно демонстрировалось в исследованиях (Ватрушкин, 2018; Lamouchi, 2020), ее главенство в академической среде сохраняется. В то же время существует альтернативная EMH гипотеза фрактального рынка (Fractal Markets Hypothesis, FMH), предложенная Mandelbrot (1983), Peters (1994). Впрочем, по мнению некоторых исследователей (Song, Liu, 2012), FMH не противоречит и не конкурирует, а дополняет EMH. В соответствии с FMH финансовые рынки представляют собой сложные системы с хаотической динамикой развития. Хаотический характер системы означает, что она, хотя является детерминированной (следовательно, предсказуемой), кажется наблюдателю стохастической (случайной). Характеристики финансовых активов обладают распределениями, отличными от нормального, имеющими «толстые хвосты», а ценовые ряды характеризуются

<sup>1</sup> Гарафутдинов Роберт Викторович — Пермский государственный национальный исследовательский университет, Пермь; rvgarafutdinov@gmail.com.

свойством персистентности — долговременной корреляционной зависимости («длинной памятью», *long memory*). Неудивительно, что в последние десятилетия на фоне повсеместного развития финансовых рынков, а также экспоненциального роста производительности вычислительной техники, позволяющей использовать на практике все более сложный математический аппарат, гипотеза фрактального рынка и применение фрактального анализа к объяснению поведения экономических систем вызывает большой интерес исследователей. Предсказание динамики финансовых показателей с высокой точностью позволило бы повысить доходность и снизить риски от операций на рынках.

Одним из методов моделирования и прогнозирования финансовых рядов, учитывающих их фрактальные свойства, является статистическая модель с длинной памятью ARFIMA — расширение широко применяющейся модели ARIMA, допускающее дробное значение порядка интегрированности ряда. Существуют исследования, в том числе упомянутые далее в тексте, подтверждающие более высокую точность прогнозов модели с длинной памятью в сравнении с ARIMA. В то же время корректная оценка параметров ARFIMA является более сложной, чем для ARIMA, задачей с не всегда очевидным решением. К параметрам модели, вызывающим трудности оценивания, можно отнести величину дробного порядка интегрированности и «длину» памяти. На оценке этих параметров зачастую не фокусируется внимание исследователей, а используются значения, предлагаемые по умолчанию прикладными пакетами статистического анализа, или вовсе заданные произвольно. В этой связи актуальной является задача изучения влияния колебаний указанных параметров на точность прогноза модели. Исследование этой задачи и является целью настоящей работы.

## 2. Обзор литературы

Изучение динамики и прогнозирование экономических показателей с использованием инструментария фрактального анализа, в частности, модели с длинной памятью ARFIMA, осуществлялось разными отечественными и зарубежными авторами. Как отмечено в (Симонов, Гарафутдинов, 2019, с. 273), «в основном фрактальный подход используется авторами для подтверждения гипотезы о ... персистентности финансовых временных рядов». Существует не так много работ, посвященных точечному прогнозированию значений финансовых временных рядов с применением модели ARFIMA. В то же время во всех рассмотренных авторами работах были получены результаты, свидетельствующие об эффективности использования таких моделей. Отметим несколько новейших исследований в данной области.

Балагулой (2020) проведено эконометрическое моделирование временных рядов оптовых цен на электроэнергию по данным торгов на оптовом рынке электроэнергии и мощности Российской Федерации (ОРЭМ РФ) для объединенных энергосистем с применением моделей классов ARFIMA и ARIMA. Показано, что учет длинной памяти в рамках ARFIMA-моделирования улучшает точность прогноза по метрике MAPE на 0.001–0.143 п.п. при выборе наилучших моделей из каждого семейства. Для идентификации параметров моделей автор использовал готовые функции библиотеки *arfima* языка программирования R.

В статье (Симонов, Гарафутдинов, 2019) проведено сравнение моделей четырех классов (ARIMA, GARCH и их модификаций с длинной памятью: ARFIMA, ARFIMA-GARCH) при изучении динамики нескольких торгуемых на ММВБ финансовых инструментов, включая обменные курсы евро и доллара США к рублю. Результаты исследования не противоречат

гипотезе о лучшей прогностической способности моделей с длинной памятью: соответствующие модификации моделей оказались точнее базовых вариантов. Для идентификации параметров модели ARFIMA была использована пользовательская функция на языке R, «длина» памяти не ограничивалась, порядок интегрированности  $d$  оценивался через показатель Херста методами минимального покрытия (на локальном участке моделирования) и R/S-анализа (на всей длине ряда). Выбор техники оценки  $d$  почти не оказал влияния на величину ошибки прогноза.

В работе (Safitri et al., 2019) показано, что цены на рынке золота Индонезии обладают длинной памятью (получено значение коэффициента Херста  $H = 0.77922$ ). Рассмотрено несколько ARFIMA-моделей различных порядков, установлено, что динамика исследуемого показателя наиболее точно описывается моделью ARFIMA(1, 1.05716, 3). Для оценки  $d$  использован метод GPH, реализованный на языке R (по-видимому, с помощью имеющейся в библиотеке функции).

Lamouchi (2020) исследовал ряды доходности и волатильности фондового рынка Саудовской Аравии (на примере индекса TASI) с применением ARFIMA. Полученные им результаты не согласуются с ЕМН и свидетельствуют о наличии длинной памяти в исследуемых рядах (полученные оценки параметра  $d$  модели входят в интервал  $[0; 0.5]$ ). С точки зрения автора, наличие длинной памяти в рядах доходности и волатильности финансовых активов свидетельствует о достаточно слабом развитии национального фондового рынка, его ограниченной интеграции в международные рынки капитала. Для оценки  $d$  использовались методы работ (Geweke, Porter-Hudak, 1983; Robinson, 1995; Sowell, 1992; Andrews, Guggenberger, 2003), результаты которых при анализе рядов доходности и волатильности индекса оказались весьма близкими (программный инструментарий не был указан).

Отметим, что в перечисленных работах не исследовалось влияние точности оценки  $d$  модели ARFIMA и «длины» памяти на качество построенных моделей.

### 3. Методы и данные

#### 3.1. Использованные методы

Интегрированная модель авторегрессии – скользящего среднего ARIMA( $p, d, q$ ) имеет следующий вид (Балагула, 2020):

$$\Phi(L)(1-L)^d Y_t = \mu + \Theta(L)\varepsilon_t, \quad (1)$$

где  $Y_t$  — исследуемый временной ряд;  $L$  — оператор сдвига;  $\Phi(L)$  — полином степени  $p$  от  $L$ , задающий функцию авторегрессии;  $\Theta(L)$  — полином степени  $q$  от  $L$ , задающий функцию скользящего среднего;  $d$  — целочисленный порядок интегрированности ряда  $Y_t$ ;  $\varepsilon_t$  — ошибки модели, соответствующие процессу белого шума;  $\mu$  — константа. Дробно-интегрированная модель авторегрессии — скользящего среднего ARFIMA( $p, d, q$ ) расширяет описанную модель на случай нецелых величин параметра  $d$ . В случае  $d \in (0; 0.5)$  процесс обладает длинной памятью (персистентен), т. е. его автокорреляционная функция убывает гиперболически медленно. Оператор взятия дробных разностей описывается следующей формулой (Балагула, 2020):

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}, \quad (2)$$

где  $\Gamma$  — гамма-функция,  $k$  — лаг. Хотя, согласно формуле (2), каждая разность ряда рассчитывается через бесконечную последовательность лагов, на практике временные ряды имеют ограниченную длину. Заметим, что использование всей предыстории ряда для вычисления каждой разности является трудоемкой задачей в разрезе использования вычислительных мощностей и может существенно увеличивать затраты времени на расчеты. Например, Прудским (2012) число лагов для вычисления каждого элемента ряда разностей дневных лог-доходностей курса доллара США к рублю было ограничено величиной 30, поскольку, как объяснял автор, «было бы неправильно оценивать завтрашний курс с учетом курса десятилетней давности». Данное исследование — редкий пример случая, когда параметр «длины» памяти учитывался при моделировании ARFIMA.

Для оценки порядка дробной интегрированности  $d$  модели ARFIMA часто применяются алгоритмы, реализованные в прикладных статистических пакетах или библиотеках языков программирования. Например, функция `arfima::arfima()` языка R использует точный метод максимального правдоподобия (exact MLE) для одновременного оценивания параметров модели  $p, q, d, \mu$  (Veenstra, 2013). Также известно (Симонов, Гарафутдинов, 2019), что величина  $d$  связана с показателем Херста  $H$  соотношением:

$$d = H - 0.5. \quad (3)$$

В настоящей работе решено применить именно этот подход к определению  $d$ . В качестве способа оценки  $H$  выбран метод детрендрованного флуктуационного анализа (ДФА) как одна из наиболее точных схем оценки монофрактальных характеристик временного ряда (Peng et al., 1994; Кириченко, Чалая, 2014; Гарафутдинов, Куваев, 2021). Приведем краткое описание метода.

Пусть  $Y = \{y_t\}$  — исследуемый временной ряд, а  $Y^{cum} = \{y_t^{cum}\}$  — кумулятивный (накапливаемый) ряд, каждый член которого вычисляется по формуле

$$y_t^{cum} = \sum_{i=1}^t (y_i - \bar{y}_i), \quad (4)$$

где  $\bar{y}_i$  — среднее значение ряда  $Y$ . Кумулятивный ряд разбивается на  $N$  сегментов длины  $\delta$ , и для каждого сегмента вычисляется флуктуационная функция

$$F_i(\delta) = \sqrt{\frac{1}{\delta} \sum_{t=i\delta+1}^{(i+1)\delta} (y_t^{cum} - Y_t^{trend})^2}, \quad i = 0, \dots, N-1, \quad (5)$$

где  $Y_t^{trend}$  — значение в точке  $t$  функции локального линейного тренда, аппроксимирующего динамику данного сегмента ряда. Значение флуктуационной функции, полученное для каждого сегмента, усредняется:

$$\overline{F(\delta)} = \frac{1}{N} \sum_{i=1}^N F_i(\delta), \quad (6)$$

после чего вычисления повторяются для нескольких значений  $\delta$ . Степенная зависимость  $\overline{F(\delta)}$  от  $\delta$  аппроксимируется линейной регрессией в логарифмических координатах:

$$\ln \overline{F(\delta)} = \alpha \ln \delta + b, \quad (7)$$

где  $b$  — свободный член. Показатель Херста определяется по  $\alpha$  следующим образом (Hardstone et al., 2012):

$$H = \begin{cases} \alpha, & \text{если } Y_t \text{ — стационарен,} \\ \alpha - 1, & \text{если } Y_t \text{ — нестационарен.} \end{cases} \quad (8)$$

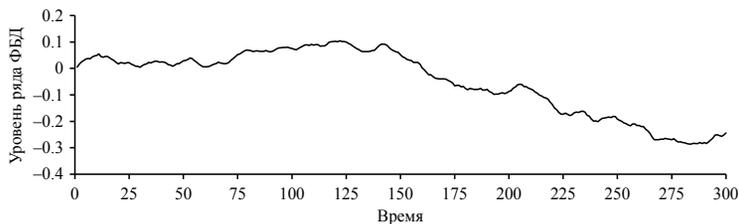
### 3.2. Методика исследования и данные

Перейдем к методике исследования. Проверилось влияние на точность прогноза модели двух ее параметров: порядка дробного интегрирования  $d$  и «длины» памяти  $l_m$ .

Сначала эксперимент проводился на модельных данных. Для генерации искусственных рядов использовалась компьютерная реализация процесса фрактального броуновского движения (ФБД) *fbm.fbm()* на языке Python. Модель ФБД была выбрана, поскольку она достаточно хорошо описывает динамику фондовых рынков (Чичаев, Попов, 2013). Было сгенерировано по 1000 рядов с длиной памяти  $Z_i$  длиной 300 с двумя различными показателями Херста:  $H = 0.7$  и  $H = 0.9$ . На рисунке 1 представлены примеры таких рядов.



а)  $H = 0.7$



б)  $H = 0.9$

**Рис. 1.** Примеры искусственно сгенерированных рядов  $Z_i$ , порожденных процессом ФБД

Каждый из сгенерированных рядов был разделен на обучающую  $Z_i^{train}$  и тестовую  $Z_i^{test}$  выборки в соотношении 299:1. После этого запускался цикл, перебирающий значения

параметра  $d$  в диапазоне  $[d_{true} - 0.5; d_{true} + 0.5]$  с шагом 0.05, где  $d_{true}$  — истинное значение  $d$  ряда (для процессов с  $H = 0.7$  и  $H = 0.9$  оно составляет 0.2 и 0.4 соответственно). Для каждого  $d$  по ряду  $Z_i^{train}$  была обучена модель ARFIMA. Использовалась собственная программная реализация ARFIMA на языке Python, основанная на функции `pmdarima.auto_arima()` с имплементацией алгоритмов получения дробных разностей и восстановления исходного ряда по сгенерированным разностям. Использование собственной функции позволило управлять «длиной» памяти  $l_m$ . При оценке влияния  $d$  на точность прогнозирования величина  $l_m$  во всех случаях была принята равной 100. Обучение модели производилось следующим образом. Порядки  $p, q$  перебирались из диапазона  $[0; 2]$  (более высокие порядки не рассматривались, поскольку, во-первых, это сильно увеличивает вычислительную трудоемкость алгоритма, а во-вторых, «пробные вычисления показали, что такие модели имеют склонность к переобучению» (Балагула, 2020, с. 94)). Для каждой комбинации строилась модель, из всех моделей отбирались прошедшие проверку на адекватность, а из полученных адекватных моделей выбиралась модель с наименьшей величиной информационного критерия BIC. Проверки модели на адекватность состояли в следующем: 1) доля значимых (на уровне 0.05) коэффициентов  $p, q$  модели должна быть более 0.5 (критерий значимых коэффициентов применялся в работе (Симонов, Ахуньянова, 2019) при построении AR-GARCH-моделей; использование критерия доли значимых коэффициентов обусловлено тем, что далеко не на любых данных удается получить модель со всеми значимыми коэффициентами, проходящую и остальные проверки, данный критерий также применялся в (Симонов, Гарафутдинов, 2019)); 2) отсутствует автокорреляция в остатках (на уровне значимости 0.05) согласно  $Q$ -тесту Льюнга–Бокса; 3) остатки нормально распределены (на уровне значимости 0.05) согласно  $W$ -тесту Шапиро–Уилка. Выбранная модель использовалась для прогнозирования на 1 шаг вперед, что обусловлено соображениями о большей достоверности столь краткосрочного прогноза относительно более длинных. Глубина прогноза в 1 шаг при исследовании предсказательной способности ARFIMA-модели применялась, например, в работах (Симонов, Гарафутдинов, 2019; Балагула, 2020). Для оценки точности прогноза использовалась абсолютная ошибка (АЕ):

$$AE = |z_t^* - z_t^{test}|, \quad (9)$$

где  $z_t^*$  — прогнозное значение ряда,  $z_t^{test}$  — истинное значение ряда. В результате всех описанных манипуляций для каждого из процессов (с разным  $H$ ) был получен 21 ряд длиной 1000, состоящий из ошибок АЕ, каждый из которых соответствует той или иной величине  $d$ . Значения абсолютных ошибок по каждому ряду были усреднены, что позволило рассмотреть зависимость средней величины ошибки АЕ от  $\Delta d$  — отклонения величины  $d$  от своего истинного значения.

Следующим этапом в исследовании закономерностей на модельных данных стало изучение зависимости ошибки прогноза от «длины» памяти ARFIMA-модели. Для решения этой задачи применялся описанный выше алгоритм с тем изменением, что переменным фактором выступила не величина  $d$ , а  $l_m$ , принимающая значения 20, 40, 60, 80, 100. Для оценки  $d$  была использована собственная реализация на Python метода ДФА с устранением линейных трендов (детрендрованием). Поскольку реализация ФБД является нестационарным рядом, параметр  $d$  модели оценивался по формуле  $d = \alpha - 1.5$ , где  $\alpha$  — скейлинговый коэффициент, оцениваемый методом ДФА (см. формулы (3), (8)).

Для проверки эмпирических закономерностей, выявленных на модельных данных, те же действия проделывались и на реальных данных, сгенерированных российским финансовым рынком, а именно, был использован ряд дневных цен закрытия индекса РТС за период с 01.09.1995 по 31.12.2009 (3580 наблюдений), преобразованный к лог-доходностям по формуле

$$y_t = \ln(p_t / p_{t-1}), \quad (10)$$

где  $p_t, p_{t-1}$  — цены закрытия индекса текущего и предыдущего дня соответственно. На рисунке 2 приведены графики исходного и логарифмированного рядов.

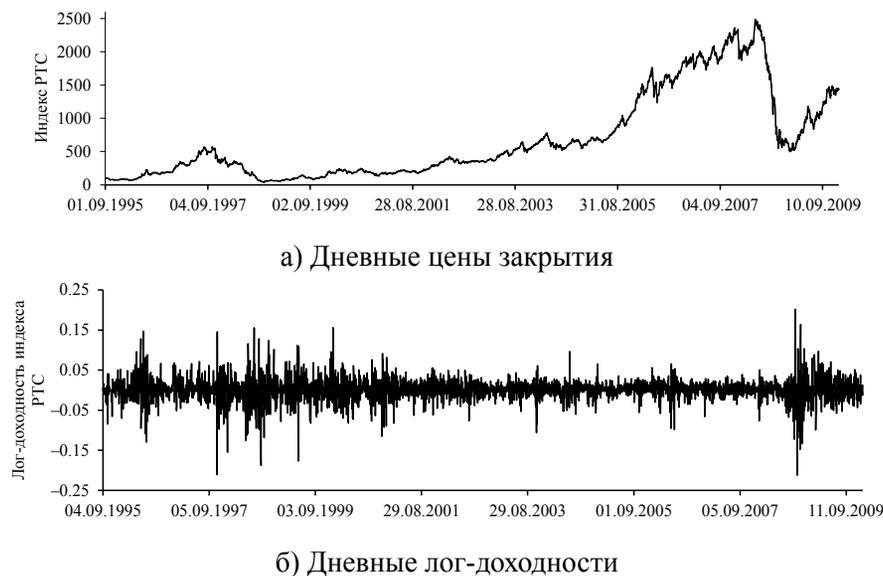


Рис. 2. Показатели индекса РТС

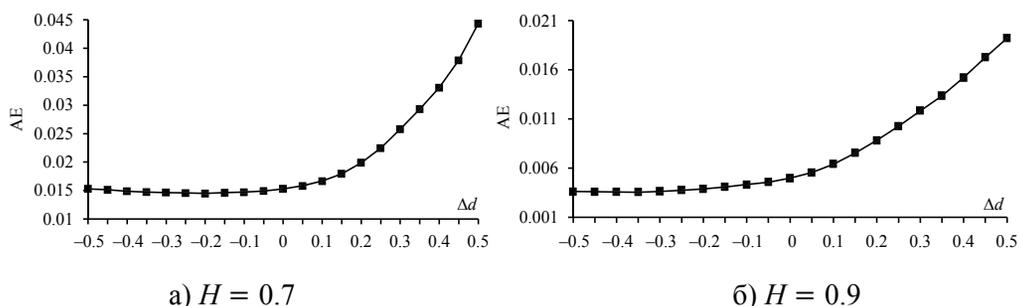
Индекс РТС — ценовой, взвешенный по рыночной капитализации композитный индекс российского фондового рынка, включающий наиболее ликвидные акции крупнейших российских эмитентов, функционирующих в основных секторах экономики РФ. Поведение индекса позволяет судить о состоянии и развитии российского фондового рынка в целом. Динамика этого финансового показателя, а также горизонт ее рассмотрения взяты из работы (Гисин, Марков, 2011), в которой проводилось исследование на персистентность ряда отечественных и зарубежных фондовых индексов. В частности, для лог-доходностей индекса РТС авторами было получено значение показателя Херста  $H = 0.617$ , что соответствует наличию у процесса длинной памяти.

Чтобы повторить массовость эксперимента с модельными данными, генерирование ряда лог-доходностей индекса осуществлялось методом «скользящего окна» шириной 300 с шагом 10. Таким способом было получено 328 выборок. Для оценки «условно-истинного» значения  $d$  был использован метод ДФА (слово «условно» используется, потому что полученное значение показателя — это всего лишь оценка), а параметр дробной интегрированности ARFIMA определялся по формуле  $d = \alpha - 0.5$ , т. к. ряды лог-доходностей стационарны.

В случае с оценкой влияния «длины» памяти алгоритм остался аналогичным, за исключением изменения величины  $l_m$  в диапазоне [20; 100] с шагом 20 при  $d = const$ .

#### 4. Результаты

Обратимся к полученным результатам. На рисунке 3 приведены графики зависимости средней величины ошибки АЕ от  $\Delta d$  — отклонения величины  $d$  от истинного значения (модельные данные).

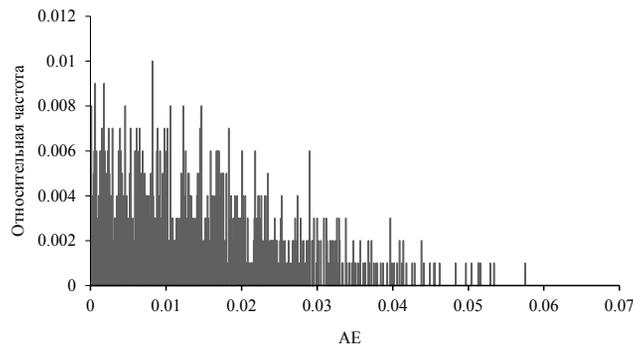
а)  $H = 0.7$ б)  $H = 0.9$ 

**Рис. 3.** Зависимость средней абсолютной ошибки прогноза ARFIMA-модели от  $\Delta d$ , полученная на модельных данных с  $l_m = 100$  и разными  $H$

Для того чтобы убедиться в статистически значимых различиях между средними значениями разных выборок, существуют параметрические критерии, в частности,  $t$ -критерий Стьюдента и ANOVA-тест. Поскольку ошибка прогноза вычислялась по абсолютной величине, распределение ошибок не является нормальным, что показали результаты  $W$ -теста Шапиро–Уилка, а также хорошо видно на графике эмпирической плотности вероятности ошибки (рис. 4).

Для сравнения рядов ошибок между собой использовался непараметрический  $U$ -критерий Манна–Уитни. Применение данного теста на каждой из парных комбинаций выборок, полученных при различных значениях величины  $\Delta d$ , дало следующие результаты. Для процесса с  $H = 0.7$  в 142 случаях из 210 различия рядов статистически значимы на уровне 0.01, для процесса с  $H = 0.9$  различия рядов статистически значимы на уровне 0.01 в 182 случаях из 210, что позволяет сделать вывод о влиянии  $\Delta d$  на точность прогноза.

Как можно заметить, при  $H = 0.7$  (рис. 3а) разница в значении абсолютной ошибки на рассматриваемом диапазоне отклонений  $\Delta d$  достигает троекратной величины (относительно ошибки при истинном значении  $d$ ). Интересно, что минимальное значение ошибки (0.0145) получено не при эталонном  $d$  (ошибка составила 0.0153), а при его отклонении на 0.2 в меньшую сторону, что соответствует прогнозированию с помощью модели ARMA( $p, q$ ), т. е. без учета длинной памяти. На графике хорошо заметно, что в случае эталонного ФБД-процесса занижение  $d$  на величину вплоть до  $-0.5$  почти не влияет на точность прогноза, в то время как изменение  $d$  в противоположную сторону приводит к экспоненциальному росту абсолютной ошибки. В случае процесса с  $H = 0.9$ , обладающего более выраженным свойством длинной памяти, т. е. более трендового (рис. 3б), проявляются похожие закономерности. А именно, занижение  $d$  слабо влияет на величину ошибки, а завышение приводит к ее росту, но не экспоненциальному, а линейному. Минимальное значение ошибки (0.0035) также получено не при эталонном  $d$ , а при  $\Delta d \in (-0.4; -0.35)$ , что соответствует модели ARMA. Это может сигнализировать о том, что модель ARFIMA описывает сгенерированные в данном исследовании ряды, порожденные процессом ФБД, хуже, чем модель ARMA, но этот



**Рис. 4.** Распределение абсолютной ошибки прогноза ARFIMA-модели, полученное на модельных данных,  $\Delta d = 0$  (на примере процесса с  $H = 0.7$ )

момент требует дополнительных исследований. Показательным является тот факт, что при увеличении  $H$  на 0.2 точность прогноза ARFIMA выросла на порядок: при  $H = 0.7$  величина АЕ равна сотым долям единицы, а при  $H = 0.9$  — тысячным долям. Это свидетельствует о том, что при усилении свойства длинной памяти динамика процесса становится более предсказуемой, он лучше моделируется и прогнозируется моделью ARFIMA.

Проанализируем зависимость величины ошибки прогноза от «длины» памяти модели. Формулу оператора взятия дробных разностей (2) можно представить следующим образом: каждый член ряда разностей  $\{y_t^d\}$  вычисляется как

$$y_t^d = \sum_{k=1}^{l_m} (-c_k y_{t-k}^d), \quad (11)$$

где  $k$  — лаг,  $c_k$  — коэффициент при лаге. В силу формулы (2) веса лагов являются экспоненциально убывающими, что должно свидетельствовать об отсутствии практического смысла в использовании лагов высокого порядка. На рисунке 5 представлены графики динамики коэффициентов при лагах, соответствующие различным значениям порядка дробной интегрированности  $d$  модели ARFIMA.

Заметно, что для персистентных процессов с  $d \in (0; 0.5)$  примерно после 20-го лага величина коэффициента начинает стремиться к нулю и, следовательно, все последующие лаги перестают вносить значимый вклад в вычисляемое значение разности  $y_t^d$ . Для антиперсистентных процессов при росте  $|d|$  увеличивается значение предела лагового коэффициента  $\lim_{k \rightarrow \infty} c_k$ , достигая значимо отличной от нуля величины. Исходя из этих данных, можно предположить, что для персистентных и антиперсистентных процессов используемый при моделировании разумный минимум «длины» памяти различен.

На рисунке 6 приведены графики зависимости средней величины ошибки АЕ от величины  $l_m$  (модельные данные).

Согласно  $U$ -тесту Манна–Уитни, в случае рядов с  $H = 0.7$  даже на уровне значимости 0.1 все выборки АЕ, полученные при разных значениях  $l_m$ , не различаются между собой. При  $H = 0.9$  статистически значимые (на уровне 0.1) различия между рядами есть только в 3 случаях из 10. На рисунке 6 также видно, что величина АЕ от «длины» памяти зависит незначительно при обоих значениях  $H$  (диапазон изменений составляет десятитысячные доли), в то же время наименьшая величина ошибки получена при 20 лагах.

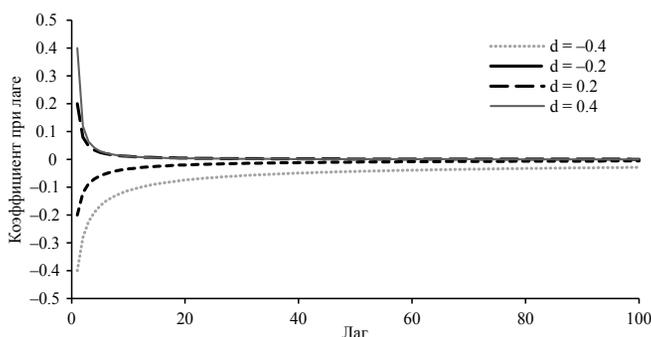


Рис. 5. Динамика коэффициентов при лагах при вычислении дробной разности

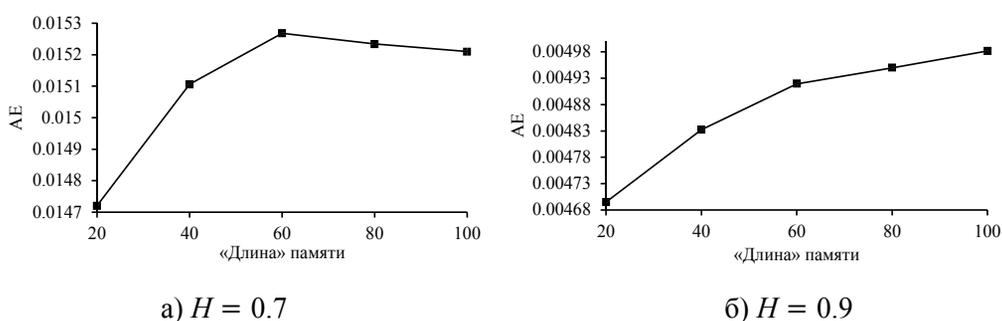


Рис. 6. Зависимость средней абсолютной ошибки прогноза ARFIMA-модели от «длины» памяти, полученная на модельных данных

На рисунке 7 приведен график зависимости времени выполнения алгоритма дробного дифференцирования от величины  $l_m$ . Алгоритм был реализован на языке Python, скрипт запускался в среде PyCharm на платформе Windows, аппаратная конфигурация ПК: CPU Intel Core i5-8400, 64 GB RAM. Входным рядом для алгоритма была реализация ФБД длиной 300 с  $H = 0.7$ ,  $d = 0.2$ , всего производилась генерация 5000 рядов для каждого значения  $l_m$ , затем полученные значения времени выполнения функции усреднялись.

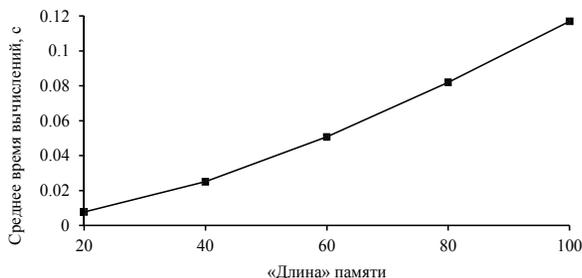


Рис. 7. Время выполнения алгоритма дробного дифференцирования модельного ряда при различных значениях  $l_m$  (на примере процесса с  $H = 0.7$ )

По графику, являющемуся практически линейным, видно, что при пятикратном увеличении «длины» памяти время вычислений возрастает на порядок. Это может быть важным фактором выбора параметра  $l_m$  при множественном обучении моделей с длиной

памятью на большом массиве данных (например, при использовании техники «скользящего окна»).

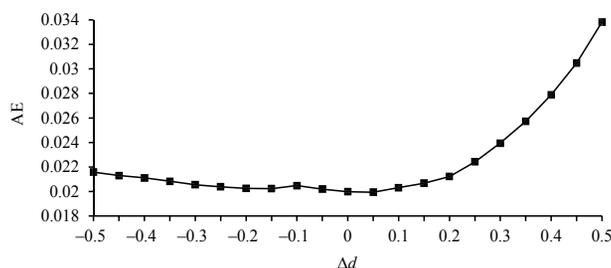
В результате экспериментов с применением компьютерной симуляции рядов с длинной памятью было установлено, что наименьшая величина ошибки прогноза получена при «длине» памяти ряда 20 (наименьшем из проверяемых значений). При увеличении «длины» вычислительная трудоемкость алгоритма оценивания модели резко возрастает, при этом практических преимуществ от данного действия выявлено не было.

Далее проверим обнаруженные закономерности на реальных данных. На рисунке 8 приведен график функции  $H(t)$ , где  $H$  — оценка показателя Херста на временном интервале  $[t; t + 298]$  (в качестве нее использована оценка показателя  $\alpha$ , полученная методом ДФА),  $t$  — левая граница интервала длиной 299 дней.



**Рис. 8.** Зависимость показателя Херста от точки отсчета 299-дневного интервала, на котором он оценивался

По графику  $H$  лог-доходностей индекса РТС видно, что ряд обладает свойством персистентности, показатель Херста принимает значения из диапазона (0.4; 0.8), в основном они выше 0.5. Такой достаточно широкий разброс свидетельствует о переменной фрактальной структуре ряда (попытки анализа рядов с учетом этой особенности проводились, например, в (Загайнов, 2017)). На рисунке 9 представлен график зависимости средней величины ошибки АЕ от  $\Delta d$  — отклонения величины  $d$  от своего «условно-истинного» значения (для ряда дневных лог-доходностей индекса РТС).



**Рис. 9.** Зависимость средней абсолютной ошибки прогноза ARFIMA-модели от  $\Delta d$ , полученная при использовании метода «скользящего окна»,  $l_m = 100$

Согласно  $U$ -тесту Манна–Уитни, на уровне 0.01 различия рядов статистически значимы лишь в 53 случаях из 210; на уровне 0.1 число значимо различных пар выборок возрастает

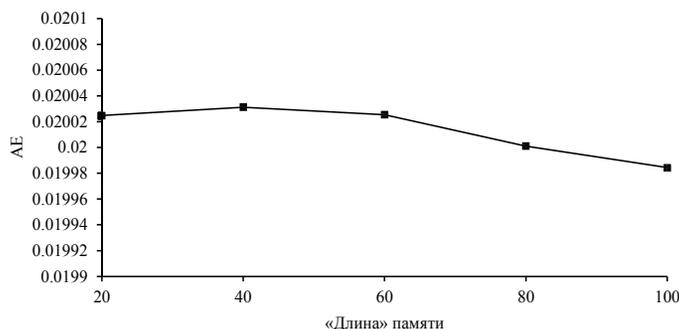
до 90, что составляет меньше половины всех комбинаций. Тем не менее, влияние  $\Delta d$  на величину ошибки прогноза все еще прослеживается. График на рис. 9 подтверждает меньшее влияние отклонения параметра  $d$  от условного «эталона» на величину ошибки в случае реальных данных в сравнении с модельными, что выражается в меньшей разнице значений АЕ на рассматриваемом диапазоне  $\Delta d$  (размах вариации составляет примерно 70% от нижней границы ошибки в отличие от 300% и выше для сгенерированных рядов). Минимум АЕ, равный 0.02, получен при  $\Delta d = 0$ , этот факт свидетельствует о высокой точности метода ДФА как способа оценки параметра длинной памяти ARFIMA-модели. Изменение  $\Delta d$  в диапазоне  $(-0.5; 0.2)$  очень слабо влияет на точность прогноза, колебания не превышают величины 0.002, при дальнейшем увеличении отклонения ошибка возрастает экспоненциально, как и в случае модельных данных с  $H = 0.7$ .

Полученные результаты в целом согласуются с результатами исследования на искусственных рядах. Обнаруженные отличия связаны со следующими факторами (помимо очевидных различий в самой природе исследуемых данных).

1. Вынужденное использование различных методик: если модельные ряды однородны, в некотором смысле стационарны, то фрагменты ряда дневных лог-доходностей индекса РТС, хоть и порождены одним и тем же процессом ценообразования на финансовом рынке, но для широких временных рамок обладают переменной структурой, что заметно по графику на рис. 8. Найти же ведущие себя почти таким же образом реальные финансовые ряды, да еще в большом количестве, невозможно.

2. Число тестовых прогнозов для каждой величины  $\Delta d$  составило не 1000, как в случае ФБД, а только 328, что неизбежно ведет к меньшей статистической надежности полученных результатов.

Далее изучим влияние длины памяти на точность прогноза дневных лог-доходностей индекса РТС. На рисунке 10 представлен график зависимости средней ошибки АЕ от величины  $l_m$ .



**Рис. 10.** Зависимость средней абсолютной ошибки прогноза ARFIMA-модели от «длины» памяти, полученная при использовании метода «скользящего окна», значение  $d$  оценено методом ДФА

Как и в случае модельных данных с  $H = 0.7$ , согласно  $U$ -тесту Манна–Уитни, на уровне значимости 0.1 все выборки АЕ, полученные при разных значениях  $l_m$ , не имеют статистически значимых различий. По графику на рис. 10 заметно, что  $AE = const$  при любом значении «длины» памяти (размах пренебрежимо мал и составляет стотысячные доли).

Также следует отметить, что величина усредненной ошибки прогноза модели ARFIMA на 1 шаг вперед при близких к истинному значениям параметра  $d$  в случае модельных данных

составила в среднем 0.015 для процесса с  $H = 0.7$  и 0.004 для процесса с  $H = 0.9$ , а в случае реальных данных — 0.02. В рамках метода Монте-Карло моделировались нестационарные ряды, а при исследовании реальных рыночных данных — стационарный ряд лог-доходностей индекса, однако, что заметно по рис. 1 и 2б, диапазон моделируемого показателя в обоих случаях оказался сопоставим и составил плюс-минус десятые доли единицы. Это позволяет утверждать, что модель ARFIMA способна достаточно хорошо описывать и предсказывать как однородные модельные ряды (на примере ФБД), так и реальные финансовые ряды (на примере лог-доходностей фондового индекса РТС).

## 5. Заключение

Проведено исследование влияния на точность прогнозирования финансовых временных рядов следующих параметров модели ARFIMA: порядка интегрированности  $d$  (точнее, его отклонения от истинного значения) и «длины» памяти  $l_m$  (числа лагов, используемых при вычислении дробных разностей). Данными для исследования послужили искусственно сгенерированные ряды — реализации процесса фрактального броуновского движения с длинной памятью (проанализированы ряды с показателями Херста  $H = 0.7$  и  $H = 0.9$ ) и реальные данные с фондового рынка — дневные лог-доходности индекса РТС, в ряде которых также было выявлено наличие длинной памяти (большая часть значений показателя Херста сосредоточена в диапазоне  $[0.5; 0.8]$ ). Получены следующие основные результаты.

1. Отклонение величины параметра  $d$  модели ARFIMA в диапазоне от 0 до  $-0.5$  почти не влияет на величину абсолютной ошибки прогноза (изменение составляет тысячные доли единицы). Изменение  $d$  от 0 в сторону увеличения приводит к быстрому росту ошибки (в случае модельных данных с  $H = 0.7$  и реальных данных с  $H \in [0.5; 0.8]$  рост — экспоненциальный, а в случае модельных данных с  $H = 0.9$  — линейный). Визуальный анализ графиков позволяет выдвинуть гипотезу о том, что по мере увеличения  $H$  (усиления свойства длинной памяти процесса) отрицательное влияние завышения параметра  $d$  модели ARFIMA на точность прогноза снижается, однако этот момент требует дополнительных исследований.

2. «Длина» памяти, учитываемая моделью ARFIMA, практически не влияет на точность прогноза. При увеличении этого параметра вычислительная трудоемкость алгоритма оценивания модели резко возрастает.

3. Наименьшая ошибка модели в случае реальных данных получается при величине  $d$ , соответствующей «условно-истинному» значению, в качестве которого была взята оценка этого параметра методом ДФА. Это позволяет сделать вывод о высокой эффективности данного метода фрактального анализа временных рядов как одного из способов оценки параметра дробной интегрированности моделируемого ряда.

По результатам проведенного исследования можно с осторожностью сформулировать следующие рекомендации по применению ARFIMA-модели для моделирования и прогнозирования финансовых временных рядов:

1) для оценки параметра  $d$  следует использовать метод, не склонный завышать его значение, метод ДФА может явиться хорошей альтернативой другим широко распространенным методам;

2) при дробном дифференцировании исходного ряда при вычислении каждой разности не обязательно использовать всю доступную предысторию значений показателя, а можно ограничиться 20 последними лагами.

## Список литературы

- Балагула Ю. М. (2020). Прогнозирование суточных цен на ОПЭМ РФ с помощью модели ARFIMA. *Прикладная эконометрика*, 57, 89–101.
- Ватрушкин С. В. (2018). Оценка временных эффектов на рынках стран БРИКС. *Финансы и кредит*, 24 (4), 913–928.
- Гарафутдинов Р. В., Куваев В. А. (2021). Сравнение двух методов фрактального анализа финансовых временных рядов. *Вестник ПНИПУ. Социально-экономические науки*, 1, 182–193.
- Гисин В. Б., Марков А. А. (2011). Ценообразование производных инструментов европейского типа на фрактальном рынке с транзакционными издержками. *Вестник Финансового университета*, 1, 34–41.
- Загайнов А. И. (2017). Исследование изменения фрактальности хаотических процессов на рынках капитала. *Интеллектуальные технологии на транспорте*, 4, 39–43.
- Кириченко Л. О., Чалая Л. Э. (2014). Комплексный подход к исследованию фрактальных временных рядов. *International Journal «Information Technologies & Knowledge»*, 8 (1), 22–28.
- Прудский М. В. (2012). Фрактальный анализ финансовых рынков. *Информационные системы и математические методы в экономике*, 5. <http://ismme.esrae.ru/pdf/2012/5/331.pdf>.
- Симонов П. М., Ахуньянова С. А. (2019). Сравнительный анализ методик AR-GARCH и  $p$ -адического прогнозирования волатильности финансового рынка. *Вестник Пермского университета. Серия «Экономика»*, 14 (1), 69–92.
- Симонов П. М., Гарафутдинов Р. В. (2019). Моделирование и прогнозирование динамики курсов финансовых инструментов с применением эконометрических моделей и фрактального анализа. *Вестник Пермского университета. Серия «Экономика»*, 14 (2), 268–288.
- Чичаев И. А., Попов В. Ю. (2013). Об одном подходе к вычислению индекса Херста финансовых временных рядов и их аппроксимации фрактальным броуновским движением. *Современные проблемы науки и образования*, 2. <https://www.science-education.ru/ru/article/view?id=8698>.
- Andrews D. W. K., Guggenberger P. (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica*, 71 (2), 675–712.
- Geweke J., Porter-Hudak S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4 (4), 221–238.
- Hardstone R., Poil S.-S., Schiavone G., Jansen R., Nikulin V. V., Mansvelder H. D., Linkenkaer-Hansen K. (2012). Detrended fluctuation analysis: A scale-free view on neuronal oscillations. *Frontiers in Physiology*, 3:450.
- Lamouchi R. A. (2020). Long memory and stock market efficiency: Case of Saudi Arabia. *International Journal of Economics and Financial Issues*, 10 (3), 29–34.
- Mandelbrot B. (1983). *The fractal geometry of nature*. Henry Holt and Company.
- Peng C.-K., Buldyrev S. V., Havlin S., Simons M., Stanley H. E., Goldberger A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49 (2), 1685–1689.
- Peters E. (1994). *Fractal market analysis: Applying chaos theory to investment and economics*. Wiley.
- Robinson P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23 (5), 1630–1661.
- Safitri D., Mustafid, Ispriyanti D., Sugito (2019). Gold price modeling in Indonesia using ARFIMA method. *Journal of Physics: Conference Series*, 1217 (1).

Song G., Liu G. (2012). EMH and FMH: Origin, evolution and tendency. *Fifth International Workshop on Chaos-fractals Theories and Applications*, 308–311.

Sowell F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53 (1–3), 165–188.

Veenstra J. (2013). Persistence and anti-persistence: Theory and software. *Ph. D. Thesis*. Western University.

Поступила в редакцию 14.09.2020;  
принята в печать 29.04.2021.

---

Garafutdinov R. V. Influence of some ARFIMA model parameters on the accuracy of financial time series forecasting. *Applied Econometrics*, 2021, v. 62, pp. 85–100.

DOI: 10.22394/1993-7601-2021-62-85-100

---

### Robert Garafutdinov

Perm State University, Perm, Russian Federation;  
rvgarafutdinov@gmail.com

## Influence of some ARFIMA model parameters on the accuracy of financial time series forecasting

The influence of ARFIMA model parameters on the accuracy of financial time series forecasting on the example of artificially generated long memory series and daily log returns of RTS index is investigated. The investigated parameters are deviation of the integration order value from its «true» value, as well as the memory «length» considered by the model. Based on the research results, some practical recommendations for modeling using ARFIMA have been formulated.

**Keywords:** ARFIMA; financial time series; long memory; Hurst index; detrended fluctuation analysis.

**JEL classification:** C22; C53; G17.

## References

Balagula Yu. M. (2020). Forecasting daily spot prices in the Russian electricity market with the ARFIMA model. *Applied Econometrics*, 57, 89–101 (in Russian).

Chichaev I. A., Popov V. Yu. (2013). About one approach for financial time series Hurst index computation and their approximation using fractal Brownian motion. *Sovremennye problemy nauki i obrazovaniya*, 2. <https://www.science-education.ru/ru/article/view?id=8698> (in Russian).

Garafutdinov R. V., Kuvaev V. A. (2021). Comparison of two methods of fractal analysis applied to financial time series. *PNRPU Sociology and Economics Bulletin*, 1, 182–193 (in Russian).

Gisin V. B., Markov A. A. (2011). Pricing of European-type derivatives on fractal market with transaction costs. *Vestnik Finansovogo universiteta*, 1, 34–41 (in Russian).

Kirichenko L. O., Chalaia L. E. (2014). Integrated approach to the study of fractal time series. *International Journal «Information Technologies & Knowledge»*, 8 (1), 22–28 (in Russian).

Prudskii M. V. (2012). Fractal analysis of financial markets. *Informacionnye sistemy i matematicheskie metody v ekonomike*, 5. <http://ismme.esrae.ru/pdf/2012/5/331.pdf> (in Russian).

Simonov P. M., Akhunyanova S. A. (2019). Comparative analysis of AR-GARCH and  $p$ -adic methods of the prediction of the financial market volatility. *Perm University Herald. ECONOMY*, 14 (1), 69–92 (in Russian).

Simonov P. M., Garafutdinov R. V. (2019). Modeling and forecasting of financial instruments dynamics using econometrics models and fractal analysis. *Perm University Herald. ECONOMY*, 14 (2), 268–288 (in Russian).

Vatrushkin S. V. (2018). Assessment of time effects in BRICS markets. *Finance and Credit*, 24 (4), 913–928 (in Russian).

Zagaynov A. I. (2017). Investigation of the change in the fractality of chaotic processes in the capital markets. *Intellectual Technologies on Transport*, 4, 39–43 (in Russian).

Andrews D. W. K., Guggenberger P. (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica*, 71 (2), 675–712.

Geweke J., Porter-Hudak S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4 (4), 221–238.

Hardstone R., Poil S.-S., Schiavone G., Jansen R., Nikulin V. V., Mansvelder H. D., Linkenkaer-Hansen K. (2012). Detrended fluctuation analysis: A scale-free view on neuronal oscillations. *Frontiers in Physiology*, 3:450.

Lamouchi R. A. (2020). Long memory and stock market efficiency: Case of Saudi Arabia. *International Journal of Economics and Financial Issues*, 10 (3), 29–34.

Mandelbrot B. (1983). *The fractal geometry of nature*. Henry Holt and Company.

Peng C.-K., Buldyrev S. V., Havlin S., Simons M., Stanley H. E., Goldberger A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49 (2), 1685–1689.

Peters E. (1994). *Fractal market analysis: Applying chaos theory to investment and economics*. Wiley.

Robinson P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23 (5), 1630–1661.

Safitri D., Mustafid, Ispriyanti D., Sugito (2019). Gold price modeling in Indonesia using ARFIMA method. *Journal of Physics: Conference Series*, 1217 (1).

Song G., Liu G. (2012). EMH and FMH: Origin, evolution and tendency. *Fifth International Workshop on Chaos-fractals Theories and Applications*, 308–311.

Sowell F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53 (1–3), 165–188.

Veenstra J. (2013). Persistence and anti-persistence: Theory and software. *Ph. D. Thesis*. Western University.

Received 14.09.2020; accepted 29.04.2021.