

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»
Институт математики, физики и информационных технологий

(наименование института полностью)

Кафедра «Прикладная математика и информатика»
(наименование)

01.04.02 Прикладная математика и информатика
(код и наименование направления подготовки)

Математическое моделирование
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему Исследование методов компьютерной реализации фрактальных рядов

Студент

А.В. Цель

(И.О. Фамилия)

(личная подпись)

Научный
руководитель

к. ф-м. н., доцент кафедры «ПМИ», Г.А. Тырыгина

(ученая степень, звание, И.О. Фамилия)

Тольятти 2021

Оглавление

Введение.....	4
Глава 1 Временные ряды	7
1.1 Одномерные и многомерные временные ряды.....	7
1.2 Проблема пробелов данных во временных рядах	8
1.3 Основные компоненты временных рядов	10
1.4 Автокорреляционная функция	12
1.5 Стационарность временных рядов.....	13
1.5.1 Критерий KPSS.....	14
1.5.2 Критерий Дики-Фуллера.....	14
1.6 Дифференцирование временного ряда	15
1.7 Логарифмирование временного ряда.....	17
Глава 2 Фрактальные и мультифрактальные временные ряды.....	20
2.1 Фракталы и самоподобие	20
2.2 История фракталов и самоподобных структур.....	21
2.3 Математическая модель	22
2.4 Фрактальная размерность	22
2.5 Показатель Херста и долговременная память	25
2.6 Методы расчёта показателя Херста	30
2.6.1 Клеточный метод.....	30
2.6.2 Поточечный метод.....	31
2.6.3 Метод R/S анализа.....	33
2.6.4 Оценка показателя Херста	36
2.7 Самоподобные процессы	37
2.7.1 Введение в самоподобные процессы и их классификации.....	37
2.7.2 Свойства самоподобия и долговременной памяти.....	39
2.8 Фрактальное Броуновское движение и Гауссовский шум	42
Глава 3 Алгоритмы и методы.....	47
3.1 Моделирование временного ряда.....	47
3.2 Модель FARIMA.....	48
3.3 Модель HAR-RV	51
3.4 Вейвлет-модель	56

3.5 Модель Average-VAR	66
Глава 4 Исследования и компьютерная реализация	68
4.1 Описание данных	68
4.1.1 Индекс волатильности	68
4.1.2 Данные артериального давления крови.....	70
4.2 Компьютерная реализация.....	71
4.3 Подготовка данных.....	72
4.4 Выбор оптимальных агрегированных уровней для модели HAR- RV и модели Average-VAR.....	76
4.5 Критерии оценки эффективности	78
4.5.1 Показатели ошибки	78
4.5.2 ROC-кривая	79
4.6 Результаты исследований.....	81
4.6.1 Сравнение графиков	82
4.6.2 Численные сравнения.....	84
4.6.2 Сравнение ROC-кривых.....	86
Заключение	89
Список использованной литературы.....	92
Приложение А	96
Показатели ошибок и время выполнения для других сигналов	96
Приложение Б.....	97
Графики ROC-кривых для других сигналов.....	97

Введение

Исторически сложилось, что практически любые объекты и явления можно описать при помощи простейших геометрических форм и привычных математических уравнений, но существует такая область познаний, которая не поддается стандартным инструментам математического анализа. Данную область математики называют хаосом.

Хаос – это состояние или последовательность состояний, имеющих нелинейное или детерминированное поведение. Долгое время считалось, что хаотичные явления – это лишь редкое исключение из непрерывного представления развития природы. Всё изменилось в 1975 году, когда американский учёный с польскими корнями Бенуа Мандельброт выпустил свою первую работу, посвящённую изучению хаоса, под названием «Фрактальная геометрия природы» [1].

Понятие фрактала – это определение, которое придумал сам Мандельброт от латинского «fractus», означающее дробленную, ломанную, изрезанную структуру [1].

Математическое понятие фрактала включает в себя объекты, структура которых схожа при разных масштабах изучения.

Не редко человек встречается с фрактальными структурами, как, например, график фондовой биржи, длина границ береговой линии, граф МРТ, компьютерное моделирование природных объектов (горы, деревья и так далее). Для изучения свойств подобных явлений и выделение зависимостей, благодаря которым можно будет предсказать поведение хаотичного множества на определённом промежутке времени, существуют методы анализа, отличные от привычных методов.

Описание проблемы. Теория хаоса по-прежнему остаётся самым распространённым качественным способом прогнозирования и исследования устойчивости состояний динамических систем. Главной задачей такого анализа является выявление всех стационарных состояний. Если одно из

таких состояний отклоняется от нормы (является не желательным), то можно спрогнозировать переход системы в данное состояние.

При анализе временных рядов используются методы, дающие точечный или интервальный количественный прогноз. Но для рядов, которые не имеют трендовой зависимости (гипотеза о существовании тренда не подтверждается), такие методы крайне непродуктивны. Поэтому для обнаружения тренда и качественного анализа временных рядов предлагается использовать фрактальный анализ.

Многие проблемы любых научных теорий заключаются во влиянии неких внешних нелинейных закономерностей, которые не способны распознать статистические методы.

Обычно динамические системы имеют фрактальные аттракторы, то есть их неустойчивые фазовые траектории со временем стремятся стать фракталами [2]. Реальные фрактальные временные ряды, как отображение реального случайного процесса, трудно прогнозируемы, но по известному поведению в прошлом, можно предположить их поведение в будущем. Но, как правило, в реальной природе чистых фракталов не существует, поэтому говоря о фракталах, подразумеваются некие фрактальные явления. Для анализа их следует рассматривать как некую модель, обладающую фрактальными свойствами.

Актуальность. Фрактальная геометрия, как наука, достаточно молодое направление, которое имеет большой потенциал для изучения. На сегодняшний день для изучения свойств хаотичных множеств широко используются методы статистического анализа случайных функций и величин. Лишь недавно для обработки сигналов наряду с ними стали пользоваться методами анализа, основанными на фрактальных и вейвлет преобразованиях. Преимуществом таких методов заключается в том, что они способны проанализировать не только глобальные свойства хаотических множеств, но и их локальные, что делает такие методы более универсальными. Актуальность данной темы заключается в новизне

исследуемого направления и в его теоретическом превосходстве над стандартными методами статистического анализа.

Объектом исследования являются временные ряды, предметом исследования – фрактальные методы моделирования временных рядов.

Цель данной работы – исследование методов компьютерного моделирования и прогнозирования фрактальных временных рядов.

Гипотеза магистерской работы состоит в том, что для анализа и прогнозирования фрактальных временных рядов можно выбрать подходящий метод моделирования, если будет произведено исследование и последующий сравнительный анализ компьютерных реализаций методов моделирования фрактальных временных рядов.

Задачи магистерской работы:

- Исследовать методы моделирования и прогнозирования фрактальных временных рядов;

- Осуществить компьютерную реализацию методов моделирования и прогнозирования фрактальных временных рядов;

- Произвести сравнительный анализ методов моделирования и прогнозирования фрактальных временных рядов и сделать выводы на основе оценок методов и выбрать наиболее подходящую модель оценки и предсказания фрактальных временных рядов.

Глава 1 Временные ряды

1.1 Одномерные и многомерные временные ряды

Временной ряд – это последовательность наблюдений какого-либо процесса, основной характеристикой является равная периодичность данных.

Временные ряды могут описывать совершенно различные данные: изменение финансовых сдвигов цен на покупку и продажу ценных металлов (рисунок 1), изменение температуры воздуха, и даже изменение сигналов, например, таких как акселерометр (рисунок 2).

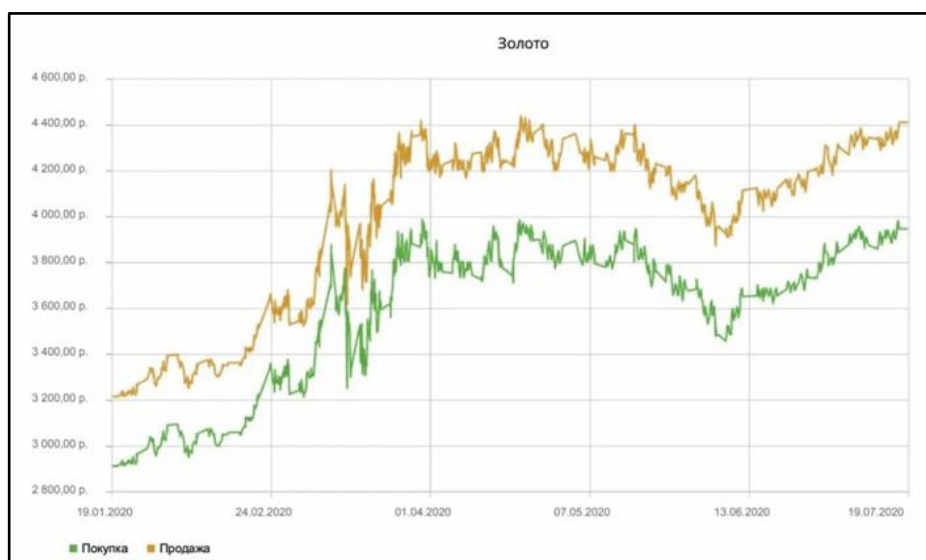


Рисунок 1 – Изменение курса покупок и продаж золота взятые из базы ПАО «Сбербанк» с 19.01.2020 по 19.07.2020

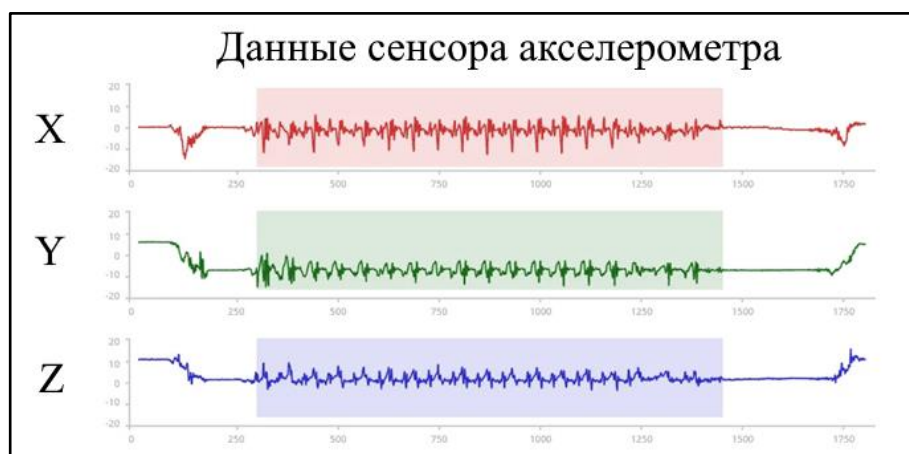


Рисунок 2 – Изменение сигналов акселерометра по трём осям

Существует два глобальных типа временных рядов – это одномерные временные ряды и многомерные временные ряды. По их названиям можно легко понять, что одномерные временные ряды зависят от одного показателя, изменяющегося во времени, а многомерные, соответственно, зависят сразу от нескольких показателей, изменяющихся одновременно во времени.

Для временного ряда не важна единица измерений, это могут быть секунды, часы, дни, года и так далее, всё это характеризует временную шкалу.

1.2 Проблема пробелов данных во временных рядах

Так как важнейшей характеристикой временного ряда является периодичность наблюдений, то исследуемый временной ряд не должен иметь пробелы в данных. Если во временном ряде всё же присутствуют пропуски (рисунок 3), то в таком случае необходимо заполнить эти пропуски для последующего анализа.

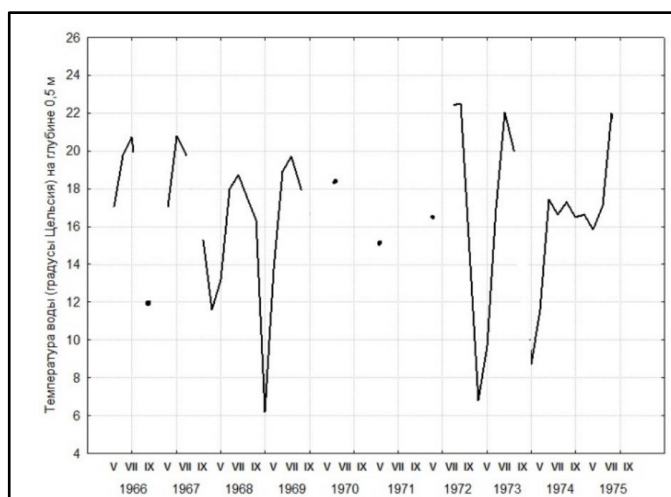


Рисунок 3 – График изменения температуры воды на глубине 0.5 м

Для решения данной проблемы существует множество способов, но в данной работе предлагается выделить основные, которые не требуют дополнительных построений модели данных.

Метод среднего значения – один из самых простых, но действенных способов заполнения пробелов в данных. При таком подходе все недостающие данные заполняются средними арифметическими значениями (рисунок 4). Так же, помимо этого, можно использовать среднее значение не по всем наблюдениям, а только по N ближайших с обеих сторон. Такой подход более приближен к реальным данным, так как ближайшие точки более точно характеризуют друг друга, но в таком случае можно упустить максимумы и минимумы временного ряда.

Метод медианы – это метод, по сути, аналогичный методу среднего значения, за исключением того, что недостающие данные заменяются медианой всех точек. Медиана более точно характеризует данные с несимметричным распределением или с любым распределением, не соответствующим нормальному закону. Так же, как и в методе среднего, можно использовать лишь N соседних точек.

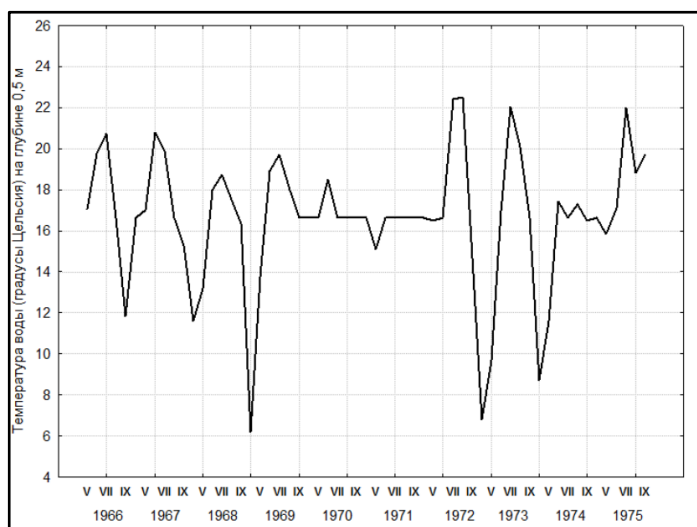


Рисунок 4 – График изменения температуры с заполнением пропусков средними значениями

Заполнение пропусков с помощью расчета процента от знаковой величины. В данном подходе лежит расчёт процента искомой величины от

выбранной знаковой величины (максимума или минимума) (рисунок 5). Этот метод позволяет учитывать как сезонную составляющую, так и тренд.

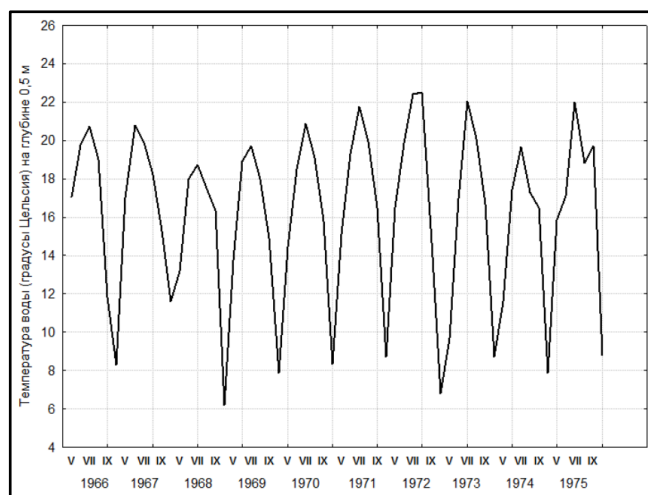


Рисунок 5 – График изменения температуры с заполнением пропусков с помощью расчёта процента от знаковой величины

1.3 Основные компоненты временных рядов

Все временные ряды обладают двумя общими компонентами: трендом и сезонной составляющей.

Тренд – это систематическая линейная или нелинейная компонента, изменяющаяся во времени.

Сезонная составляющая – это компонента, характеризующая периодические колебания уровней временного ряда каждый сезон.

Циклическая составляющая – это компонента, характеризующая периодические колебания уровней временного ряда, выходящие за рамки одного сезона. Промежуток времени между двумя средними соседними вершинами или впадинами определяют как длину цикла.

Все эти компоненты могут встречаться во временном ряде одновременно. К примеру, у некой компании наблюдается ежегодный рост прибыли – это тренд, но при этом на первый квартал каждого года приходится 25% годовой выручки – это сезонная составляющая.

Данные компоненты легко рассмотреть на примере «Ряда G», который предложили Д. Бокс и Г. Дженкинс в 1976 [3] (рисунок 6). Данный

временной ряд представляет график месячных авиаперевозок людей (в тысячах) в течение 12 лет.

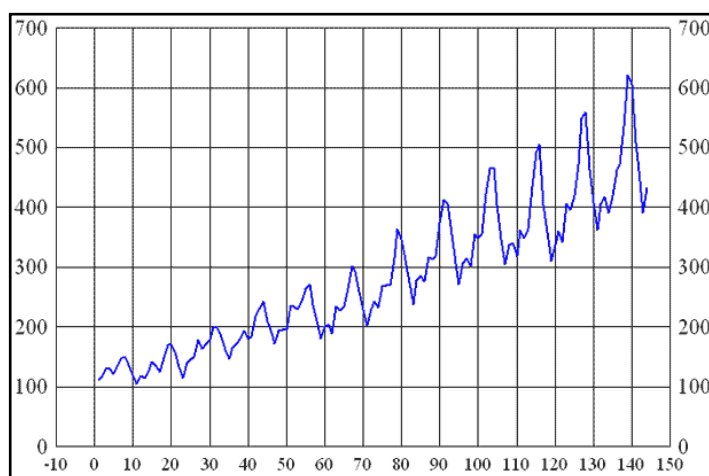


Рисунок 6 – Перевозка пассажиров или «Ряд G»

На графике чётко видно, что временной ряд имеет почти линейный тренд, то есть количество пассажиров, перевозимых каждый год, увеличивается линейно (частное между первым и последним годом примерно равно 4). При этом очевидна сезонность – в отпускной месяц авиаперевозок значительно больше, чем в другие месяцы. Так же данный график показывает временной ряд, у которого амплитуда сезонных изменений увеличивается вместе с трендом. Такой вид компоненты называют мультипликативной сезонностью.

Случайная составляющая (так же называют остатком или нерегулярной компонентой) – это случайный шум или ошибка, которая воздействует на временной ряд нерегулярно. К основным причинам происхождения шумов относят факторы внезапного действия и воздействие текущих факторов. Шум, в основе возникновения которого лежат воздействия текущих факторов, может быть связан также с ошибками наблюдений.

Катастрофическим колебанием называется случайный шум, в основе возникновения которого лежат факторы резкого и внезапного действия.

Временной ряд можно представить следующей математической моделью:

$$y(t) = x(t) + s(t) + z(t), \quad (1)$$

где $x(t)$ – тренд;

$s(t)$ – сезонная составляющая;

$z(t)$ – остаток.

Для временного могут быть применены методы декомпозиции, которые позволяют выделить некоторые компоненты временного ряда для его последующего анализа и предсказания.

1.4 Автокорреляционная функция

Коэффициент корреляции – это значение, характеризующее силу статистической связи между двумя значениями.

Самая распространённая формула расчёта коэффициентов корреляции – это корреляция Пирсона [9].

$$r = \frac{\sum_{i=1}^n (x_n - \bar{x})(y_n - \bar{y})}{n\sigma_x\sigma_y}. \quad (2)$$

В анализе временных рядов автокорреляционная функция – это та же корреляция Пирсона, но она описывает корреляции между исходным временным рядом и его копией, сдвинутой на заданное количество интервалов (это число обычно называют лагом L).

$$f(L) = \sum_{L=0}^n r_{t,t-L}, \quad (3)$$

где n – число уровней временного ряда;

r – коэффициент корреляции.

При увеличении лага коэффициент корреляции уменьшается.

График зависимости коэффициента корреляции от лага называется коррелограммой. Она позволяет эффективно определить некоторые характеристики исследуемого временного ряда. Все значения коррелограммы расположены в диапазоне $[-1; 1]$ и, соответственно, на нулевом лаге значение автокорреляционной функции равно единице.

1.5 Стационарность временных рядов

Стационарность – это свойство процесса, позволяющее ему не менять свои характеристики со временем.

Под стационарными временными рядами подразумевают временные ряды, элементы которых обладают постоянными характеристиками:

- Математическое ожидание и дисперсия элементов временного ряда не изменяются:

$$E[x(t)] = a; D[x(t)] = \sigma^2, \forall t. \quad (4)$$

- Ковариация между различными элементами временного ряда зависит только от сдвига этих элементов во времени относительно друг друга.

$$\text{cov}[x(t), x(t + \tau)] = \gamma(\tau), \forall t. \quad (5)$$

Можно сказать, что математическое ожидание должно быть равно константе, а ковариация двух различных наблюдений временного ряда зависит только от разницы времени между этими значениями.

Для оценки гипотезы стационарности временного ряда существует множество критериев.

1.5.1 Критерий KPSS

Критерий KPSS получил своё название в честь фамилий учёных, описавших его в 1992 году – Квятковский-Филлипс-Шмидт-Шин (Kwiatkowski–Phillips–Schmidt–Shin).

Если рассматриваемый временной ряд имеет вид:

$$\begin{aligned}y_t &= c_t + \delta t + e_t, \\c_t &= c_{t-1} + u_t,\end{aligned}\tag{6}$$

где δ – коэффициент тренда, e_t – некоторый стационарный процесс;

u_t – некоторый независимый и одинаково распределённый процесс с математическим ожиданием равным нулю и дисперсией σ^2 .

Выдвигаются две конкурирующие гипотезы:

H_0 : временной ряд является стационарным ($\sigma^2 = 0$);

H_1 : временной ряд не является стационарным ($\sigma^2 \neq 0$).

Вычисляется статистика:

$$\frac{\sum_{i=1}^T S_i^2}{s^2 T^2},\tag{7}$$

где T – размер выборки;

S – сумма стационарных процессов ($S_t = e_1 + e_2 + \dots + e_t$), s^2 – стационарная ошибка [11].

1.5.2 Критерий Дики-Фуллера

Гипотеза Дики-Фуллера названа в честь Дэвида Дики и Уэйна Фуллера, которые предложили данный критерий в 1979 году. Данный тест является одним из тестов на единичные корни [8].

Временной ряд имеет единичный корень, или порядок интеграции один, если его первые разности образуют стационарный ряд. Это условие так же записывается как $y_t \sim I(1)$ если ряд первых разностей $\Delta y_t = y_t - y_{t-1}$ является стационарным $\Delta y_t \sim I(0)$.

При помощи этого теста проверяют значение коэффициента ρ в авторегрессионном уравнении первого порядка AR(1).

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad (8)$$

где y_t – временный ряд;

ε – ошибка.

В отличие от критерия KPSS, критерий Дики-Фуллера проверяет противоположные гипотезы:

H_0 : временной ряд не является стационарным и не может быть представлен единичным корнем, так как имеет некоторую зависящую от времени структуру;

H_1 : временной ряд является стационарным и не имеет единичного корня.

Результат интерпретируется при помощи значения ρ из теста:

- ρ больше 0.05 – Не удаётся отклонить нулевую гипотезу H_0 , временной ряд имеет единичный корень и является нестационарным;
- ρ меньше или равно 0.05 – Отклоняется нулевая гипотеза H_0 , временной ряд не имеет единичных корней и является стационарным;

1.6 Дифференцирование временного ряда

Дифференцирование временного ряда – это переход от ряда X_t к ряду разностей $X_t - X_{t-1}$.

Данная операция позволяет стабилизировать среднее значение временного ряда, удаляя из него изменения уровня временного ряда и тем самым устраняя тренд и сезонность. Таким образом, можно привести не стационарный временной ряд в стационарный ряд.

Дифференцирование может быть использовано на любое количество лагов.

Инвертирование процесса требуется, когда прогноз должен быть преобразован обратно в исходный масштаб. Получить исходный временной ряд можно добавив наблюдение на предыдущем временном шаге к значению разницы.

$$X_t = (X_t - X_{t-1}) + X_{t-1}. \quad (9)$$

На рисунке 7 представлен пример логарифмирования временного ряда на 1 лаг.

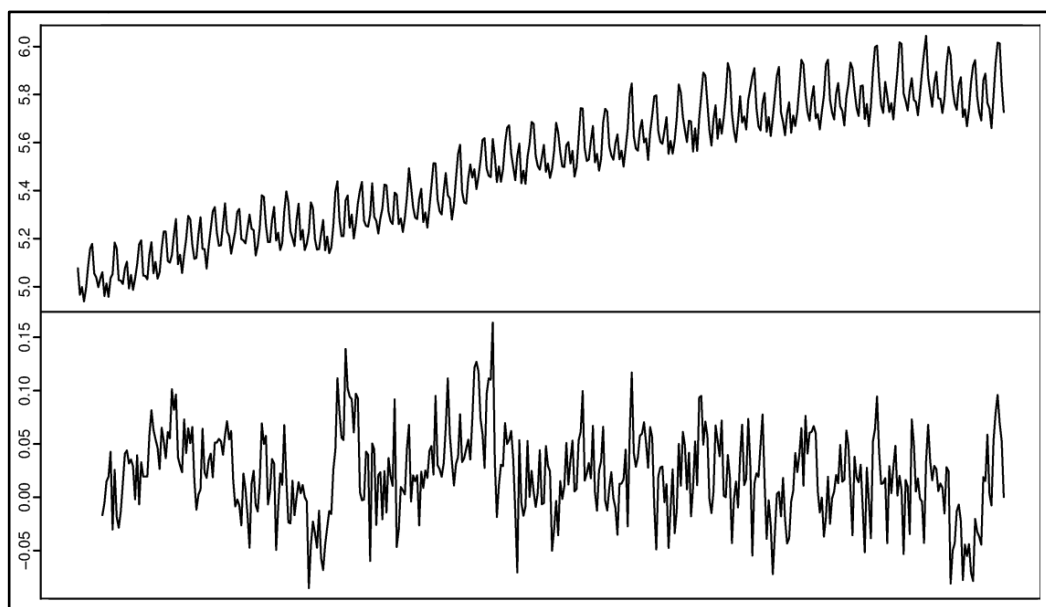


Рисунок 7 – Логарифмирование временного ряда на один лаг

1.7 Логарифмирование временного ряда

Дифференцирование не единственный способ привести временной ряд к стационарному виду.

Данная операция позволяет получить ряд с постоянной дисперсией в тех случаях, когда ошибка в модели носит мультипликативный характер (с ростом уровня ряда растет и дисперсия ошибки) (рисунки 8 и 9).

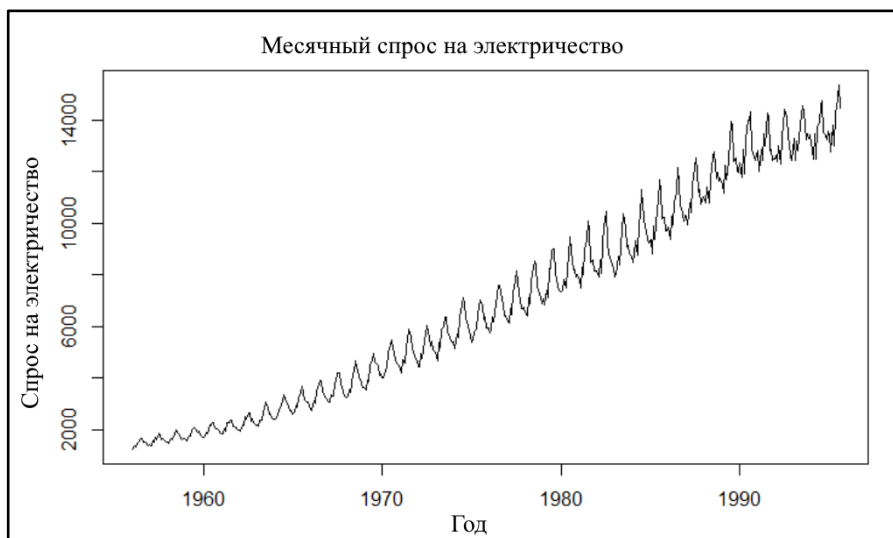


Рисунок 8 – Исходный временной ряд месячного спроса на электричество

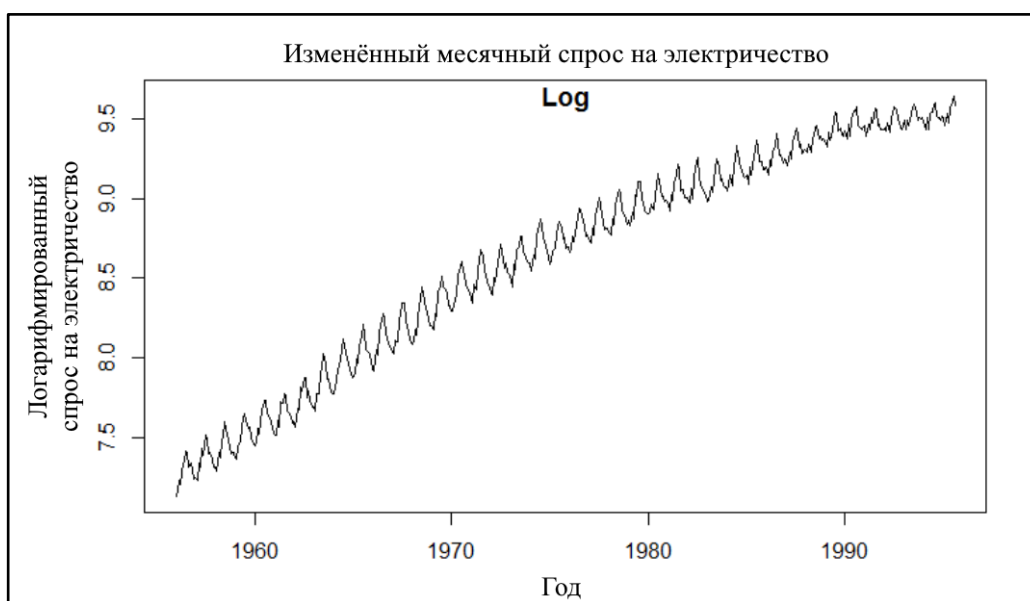


Рисунок 9 – Применение логарифмирования для месячного спроса на электричество

Такое преобразования так же называют преобразованием Бокса-Кокса. Назван он в честь авторов этого метода – известных статистиков Джорджа Эдварда Пелхэм Бокса (George Edward Pelham Box) и Дэвида Роксби Кокса (Sir David Roxbee Cox) [10].

Параметрическое семейство стабилизирующих дисперсию преобразований описано ниже:

$$y'_t = \begin{cases} \ln(y_t), & \lambda = 0, \\ \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases} \quad (10)$$

Параметр λ подбирается так, чтобы минимизировать дисперсию или максимизировать правдоподобие модели (рисунок 10).

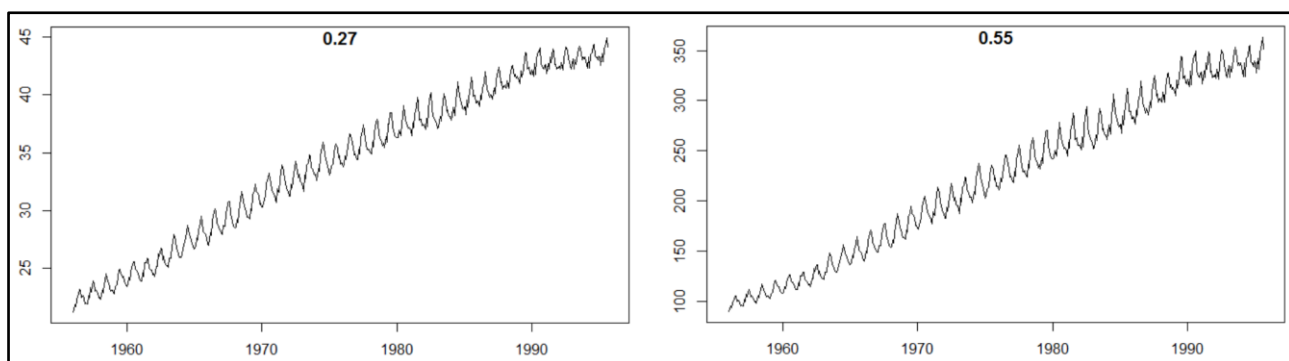


Рисунок 10 – Преобразования Бокса-Кокса для λ равной 0.27 (слева) и λ равной 0.55 (справа)

После построения прогноза для логарифмированного ряда его необходимо преобразовать в прогноз исходного:

$$\hat{y}_t = \begin{cases} \exp(\hat{y}'_t), & \lambda = 0, \\ (\lambda \hat{y}'_t + 1)^{1/\lambda}, & \lambda \neq 0. \end{cases} \quad (11)$$

Стоит отметить, если некоторые значения y_t меньше нуля, преобразования Бокса-Кокса невозможны, и необходимо сдвинуть временной ряд на константу.

Вывод к главе 1

В данной главе были описаны основные компоненты временного ряда, такие как сезонность, тренд и цикличность. Так же было рассмотрено понятие стационарности временных рядов и его критерии, такие как критерий KPSS и критерий Дики-Фуллера. Для достижения стационарности были описаны методы дифференцирования и логарифмирования. Помимо этого, в данной главе исследуется автокорреляционная функция и свойства, которые она отражает.

Глава 2 Фрактальные и мультифрактальные временные ряды

2.1 Фракталы и самоподобие

Фрактал относится к объектам, состоящим из частей повторяющих форму исходного объекта. При наблюдении на более мелком масштабе можно обнаружить ту же структуру, что и на более большом масштабе, поэтому почти невозможно точно узнать в каком масштабе происходит наблюдение [11]. Это свойство называют масштабной инвариантностью, наиболее важной особенностью фракталов, которая также называется самоподобием [11].

Ученые и исследователи не редко находят фракталы в природе. Например, папоротники, снежинки, береговые линии и т. д. [1]. В человеческих телах многие ткани и органы также проявляют фрактальные свойства: легкие, кровеносная система и так далее [12]. Кроме того, фракталы генерируются математическим способом с помощью рекурсии. К примерам таких фракталов можно отнести прокладку Серпинского, множество Кантора, кривую Коха (рисунок 11) и т. д. [13].

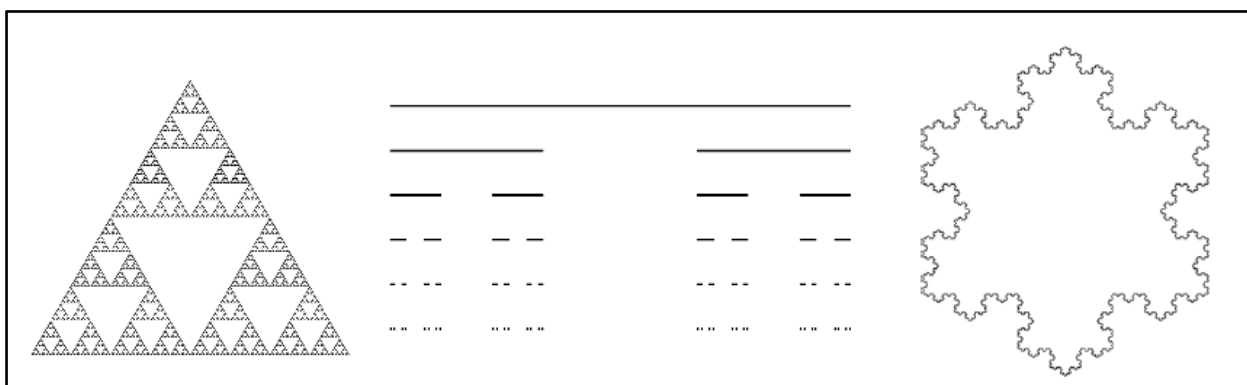


Рисунок 11 – Прокладка Серпинского, множество Кантора и кривая Коха

Фракталы ограничиваются не только природными структурами и геометрическими объектами, но также могут использоваться для описания свойств временного ряда. Некоторыми типичными примерами фрактальных

временных рядов являются интернет-трафик, финансовые временные ряды, физиологические сигналы (рисунок 12) и так далее.

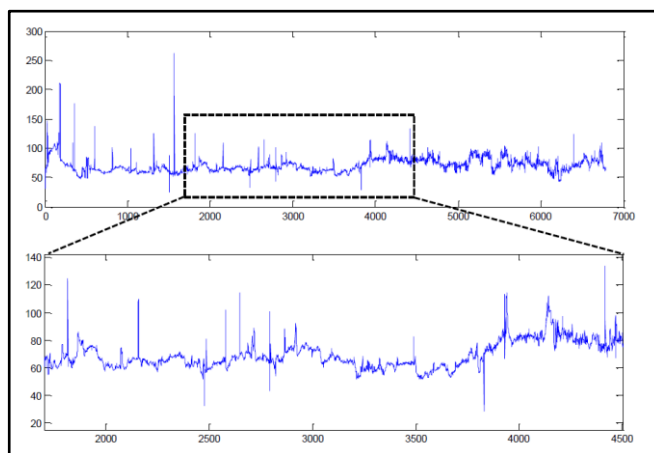


Рисунок 12 – Сигнал артериального давления. Сигнал на меньшей шкале (внизу) аналогичен сигналу на более крупной шкале

2.2 История фракталов и самоподобных структур

Фрактальные явления наблюдались задолго до того, как появился сам термин «фрактал». Когда ученые пытались измерить сложные закономерности, такие как площади островов Эгейского моря, длину британской береговой линии и другое, они часто обнаруживали появление нерегулярных степенных законов [15]. Эти степенные законы, а также размер этих неправильных форм отличаются от геометрии, с которой они обычно имели дело.

Бенуа Мандельброт впервые ввел термин фрактал в свою французскую книгу «Les Objets Fractals: Forme, Hasard et Dimension» в 1975 году, а затем опубликовал знаменитое английское издание «The Fractal Geometry of Nature». В книге он заявил, что природа состоит не только из привычной для нас геометрии, а ещё из некоторой другой более сложной и неустойчивой [2]. Он назвал эту новую геометрию фрактальной. Обычные объекты в природе, включая облака, горы, береговые линии, пути, по которым проходит молния, и многое другое – все это стало частью новой фрактальной концепции. Эти объекты можно описать с помощью рекурсивных уравнений, в которых

похожие элементы повторяются снова и снова, образуя неправильную форму [2].

2.3 Математическая модель

Пусть имеется временной ряд вида:

$$y_t = \eta_t + \chi_t + \varepsilon_t, \quad (12)$$

где $t = t_0, t_1, \dots, t_n$ – моменты времени;

η_t – трендовая составляющая;

χ_t – коэффициент, выражающий меру хаотичности ряда и описывающий характер процесса. Зависит от показателя Херста H_t и фрактальной размерности D_t ;

ε_t – случайный шум.

2.4 Фрактальная размерность

Топологическая размерность объекта – это наиболее часто используемая мера измерения. Например, линия имеет топологическую размерность равную единице, а поверхность равную двум. Однако фракталы имеют свою собственную размерность. В отличие от топологической размерности, которая всегда является целым числом, фрактальная размерность обычно дробная.

Фрактальная размерность тесно связана с измерениями самого объекта. К примеру, отрезок не является фракталом, потому что его длина остается неизменной не зависимо от шкалы, используемой для измерения [10]. Но для фракталов всё совершенно по-другому.

Одним из примеров расчёта длины фрактальной структуры можно привести британскую береговую линию. При увеличении масштаба можно увидеть всё больше мелких отрезков на контуре береговой линии. Длина во многом зависит от масштаба измерения (рисунок 13) [10, 11].

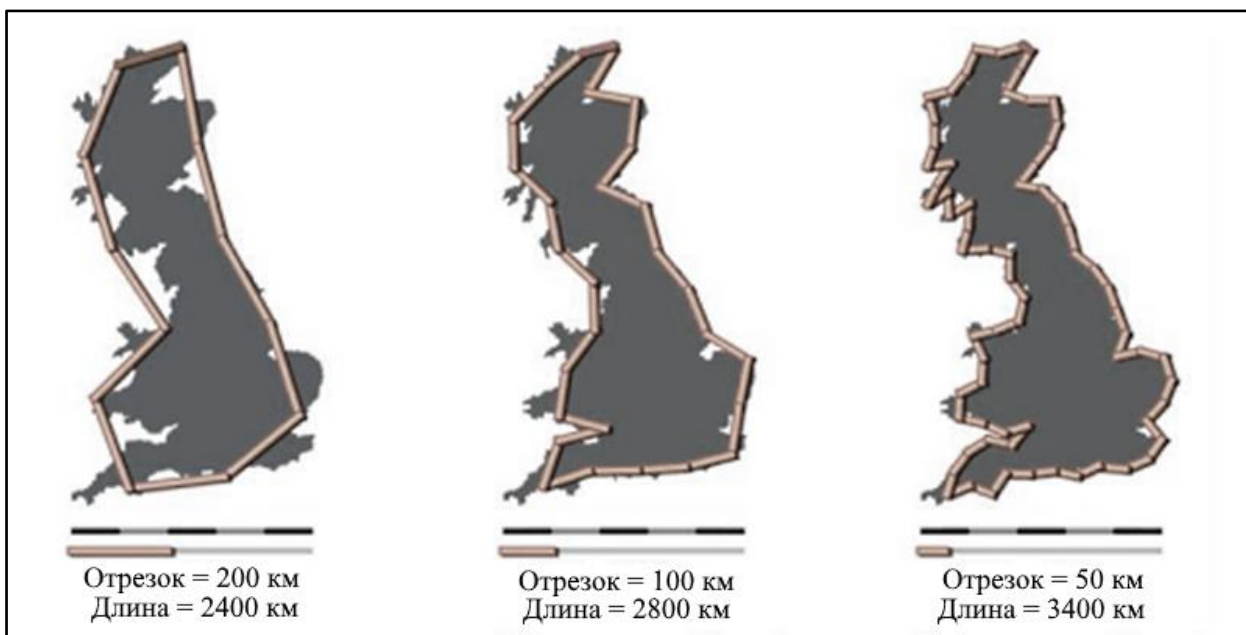


Рисунок 13 – Британская береговая линия

Чем меньше масштаб, тем больше длина береговой линии. При большем масштабе, соответственно, береговая линия будет иметь меньшую длину. На рисунке 13 показано сравнение масштабов измерения и полученной длины. Происходит такой же эффект, если попытаться измерить длину кривой Коха (рисунок 13). Длина между любыми двумя точками на кривой не является фиксированной, а зависит от шкалы измерения. Фактически, длина бесконечна, так как можно применить бесконечное число масштабов для измерения [11]. Эта несогласованность в измерениях также возникает в задачах измерения других фракталов, как, например, поиск площади снежинок, нахождение длины графика фрактального временного ряда и так далее.

Предположим, что r – это единица измерения, которая используется для измерения многомерности объекта (длина отрезков линии, площадь поверхностей и так далее), а соответствующее значение многомерности – $L(r)$, то фрактальная размерность может быть получена при помощи уравнения:

$$D_F = \frac{\log(L(r))}{\log(r)}. \quad (13)$$

Чтобы получить фрактальную размерность определенного объекта, сначала выбираются несколько шкал измерения r , измеряется многомерность объекта – $L(r)$, а затем строится логарифмический график. Наклон графика – это и есть фрактальная размерность объекта.

Объект с фрактальной размерностью, превышающей его топологическую размерность, называется фракталом [15].

Согласно этому определению, к примеру, линия с фрактальной размерностью равной единице не является фракталом, поскольку её фрактальная размерность равна её топологической размерности. То же самое можно сказать про треугольник с фрактальной размерностью равной двум, что соответствует его топологической размерности.

Мандельброт в своих исследованиях обнаружил, что фрактальная размерность для западной британской береговой линии D_F равно 1.25 [10], что строго превышает его топологическую размерность, равную единице, и, следовательно, это фрактал. Дробное измерение указывает на то, что свойство фрактала находится между объектами с более высоким и низким измерением [11]. Фрактальный временной ряд, к примеру, имеет фрактальную размерность в диапазоне от одного до двух. По сути, сам фрактальный временной ряд уже не является линией, так как линия имеет топологическую размерность равную единице, но и не является поверхностью, так как топологическая размерность поверхности равна двум. То есть когда D_F приближается к единице, временной ряд больше похож на линию, а по мере приближения D_F к двум, временной ряд становится все более извилистым и приближается к поверхности [15].

2.5 Показатель Херста и долговременная память

Одним из главных направлений фрактального анализа является изучение фрактальной размерности или показателя скейлинга. Данный показатель описывает фрактальность структуры или раздробленность поверхности. Фрактальная размерность, введённая Хаусдорфом как D -размерность, является основной характеристикой фрактальных структур [4].

Показатель Херста H напрямую связан с топологией исследуемых объектов:

$$H = 2 - D, \quad (14)$$

где D – размерность Хаусдорфа (фрактальная размерность);

H – показатель Херста.

Показатель Херста (H) располагает значениями в промежутке $0 \leq H \leq 1$ и способен определить характеристику финансового временного ряда:

Если показатель Херста расположен в промежутке $0.5 < H \leq 1$, то процесс характеризуется долговременной памятью, или временной ряд является персистентным (инерционным). Данное свойство говорит о независимости от временного масштаба, и является фрактальным. То есть таким временным рядам свойственна последовательность в значениях. Персистентные временные ряды могут быть как восходящими, так и нисходящими. Примером восходящего персистентного временного ряда является график котировок акций «Аэрофлот» (рисунки 14), а нисходящим временным рядом график компании «ВТБ» (рисунок 15).



Рисунок 14 – Котировка акций компании «Аэрофлот»



Рисунок 15 – Котировка акций компании «ВТБ»

Если показатель Херста лежит в диапазоне $0 \leq H < 0.5$, то такой временной ряд антиперсистентный. Данное свойство показывает сильную «зашумлённость» временного ряда, что негативно влияет на предсказуемость последующих значений, а финансовый рынок в данных промежутках называют нестабильным. Так же в случае антиперсистентных процессов, говорят, что движения в нём обратно пропорциональны предыдущему периоду. Примером антиперсистентный финансовых временных рядов являются графики котировок акций компаний «АвтоВАЗ» и «Мечел» (рисунок 16 и рисунок 17).

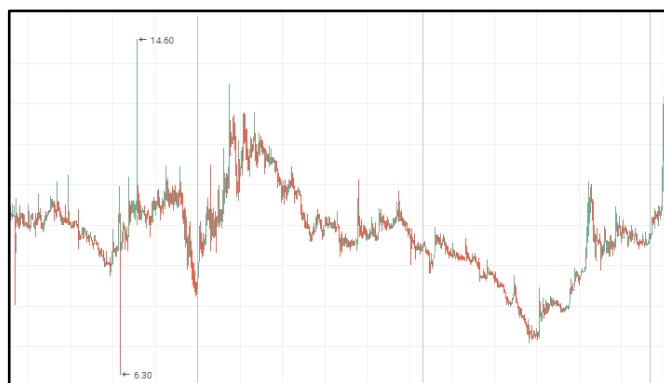


Рисунок 16 – Котировка акций компании «АвтоВАЗ»



Рисунок 17 – Котировка акций компании «Мечел»

Если показатель Херста $H = 0.5$, то можно сказать, что гипотеза эффективного рынка подтверждается. Это означает, что все значения автоматически становятся некоррелированными, а если есть некая внешняя информация, способная оказать влияние на показатели временного ряда, то это влияние уже произведено. Такой временной ряд не будет являться фракталом [6].

Недостатком данной характеристики является тот факт, что для получения наиболее точного показателя Херста необходим достаточно большой объём данных (несколько тысяч значений), иначе показатель может иметь сильную погрешность. Так же различные методы расчёта показателя Херста могут выдавать разные результаты.

Величина фрактальной размерности D_H , основанная на вычислении показателя Херста (H), описывает характеристику всего временного ряда, не

смотря на то, что его динамика может изменяться, тренды могут менять своё направление или переходить в фазы случайных колебаний.

Пусть X_t – стационарный процесс. Если существует действительное число $\alpha \in (0, 1)$ и константа c_p больше нуля, такая, что автокорреляционная функция $\rho(k)$ удовлетворяет условию:

$$\rho(k) \sim c_p k^{-\alpha}, k \rightarrow \infty, \quad (15)$$

то X_t называют стационарным процессом с долговременной памятью [14].

Уравнение 15 показывает, что автокорреляция затухает гиперболически медленнее, чем экспоненциальный спад для процессов с короткой памятью, полученной при помощи автокорреляционной функции [16]. Опираясь на это свойство, процессы с долговременной памятью можно характеризовать как стационарный временной ряд с ковариацией, обозначенной как $\gamma(k)$, и угасающей так медленно, что её сумма расходится [16].

$$\sum_{k=-\infty}^{\infty} \gamma(k) = \infty \text{ или } \sum_{k=0}^{\infty} |\gamma(k)| = \infty. \quad (16)$$

Двумя наиболее важными свойствами долговременной памяти являются стационарность и медленно затухающая автокорреляция. Если мы сравним автокорреляционную функцию сигнала с длительной памятью и автокорреляционную функцию с короткой памятью, то можно увидеть очевидную разницу между двумя автокорреляционными графиками (рисунок 18).

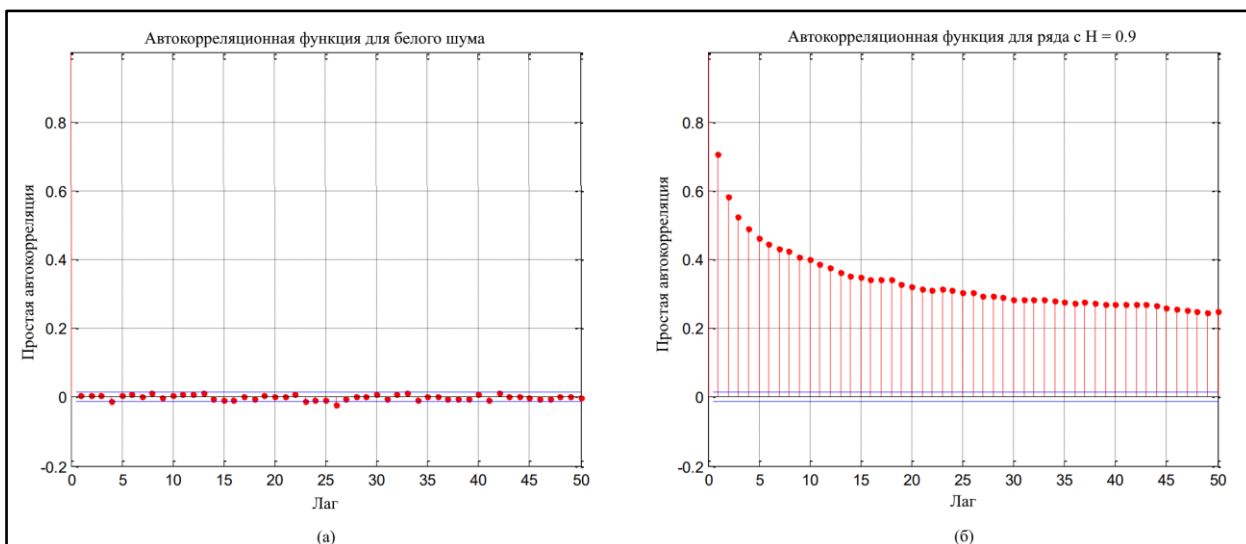


Рисунок 18 – Сравнение автокорреляционной функции для временного ряда короткой памятью и ряда с долговременной памятью. Под пунктом (а) простой белый шум; под пунктом (б) Гауссов шум с показателем Херста H равным 0.9

Для временных рядов с короткой памятью (таких как белый шум) автокорреляционная функция быстро падает до нуля, в то время как для временных рядов с долговременной памятью (таких как Гауссов шум) автокорреляция существует даже после больших лагов. Поэтому для временных рядов с короткой памятью прошлые значения малозначительны для определения поведения будущих значений. Но для временных рядов с длительной памятью предыдущие значения, даже сильно отдалённые, всё ещё коррелированы между собой. Это главная причина, из-за которой неправильно применять один и тот же подход к анализу временных рядов с короткой и долговременной памятью.

Для временных рядов с долговременной памятью можно наблюдать кластеризацию точек данных по одну сторону от среднего значения в течение некоторого времени, а затем с другой стороны. Это явление не наблюдается на временных рядах с кратковременной памятью, и связано с эффектом Джозефа [15, 16]. Из-за этого можно выдвинуть предположение, что сигналы с длительной памятью нельзя анализировать традиционными способами.

2.6 Методы расчёта показателя Херста

2.6.1 Клеточный метод

Для определения фрактальной размерности двумерного множества существует самый простой способ – метод клеточного анализа. В его основе лежит простая идея: всю плоскость, на которой располагается хаотичное множество, делится на квадраты (для первой итерации плоскость обычно делится на один или четыре квадрата). При изменении масштаба полученной сетки, изменяются и количество единичных квадратов, покрывающих множество [5]. По формуле 17 можно вывести фрактальную размерность:

$$D = -\lim_{\delta \rightarrow 0} \frac{\ln(N_\delta)}{-\ln(\delta)}, \quad (17)$$

где δ – площадь единичного квадрата;

N_δ – минимальное количество квадратов, необходимых для покрытия фрактального множества.

Численный расчёт показателя Херста есть линейная регрессия зависимости $\ln(N_\delta)$ от $-\ln(\delta)$. Угловым коэффициентом наклона кривой и есть оценка показателя Херста H .

Таким образом, можно измерять не только двумерные хаотичные множества (анализ на двумерном множестве – это лишь частный случай), но и, например, одномерные (в качестве единицы измерения берётся отрезок), или трёхмерные (в качестве единицы измерения берётся куб).

Данный метод хорош тем, что имеет относительную простоту реализации (блок-схема представлена на рисунке 19).

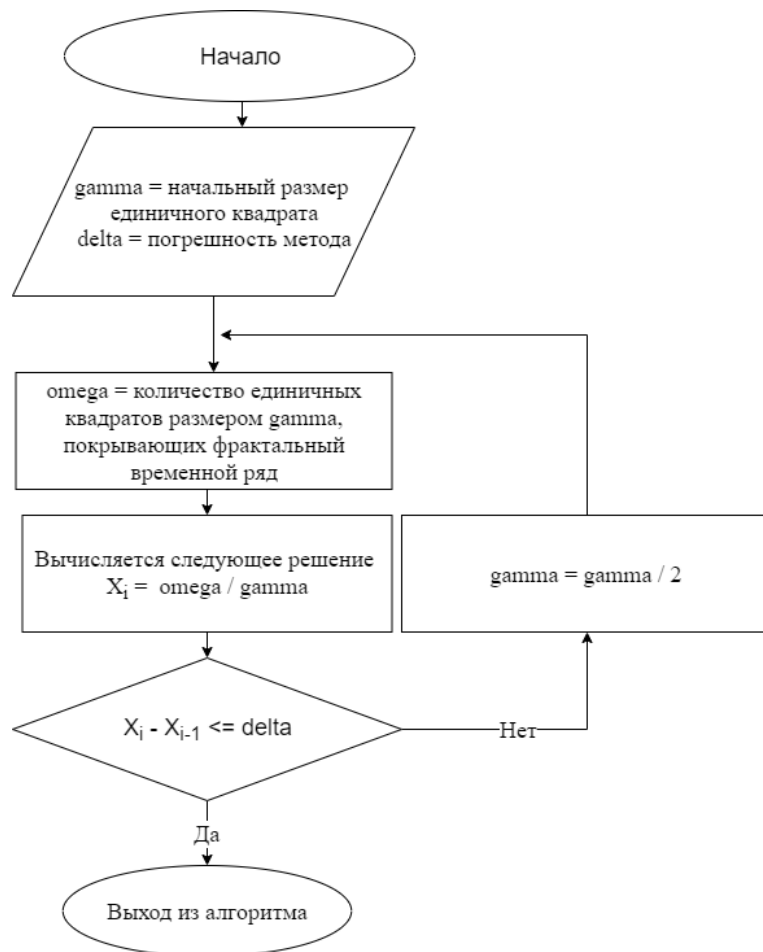


Рисунок 19 – Блок-схема метода клеточного анализа

В качестве недостатков можно выделить большую погрешность вычислений, зависящей исключительно от входного фрактального множества и используемого ЭВМ для вычислений. Так же задача вычислений принимает нетривиальный вид при количестве делений от 2^{15} .

2.6.2 Поточечный метод

Данный метод является исторически одним из самых первых разработанных алгоритмов, поэтому по праву может считаться классическим алгоритмом расчёта размерности Хаусдорфа.

В данном методе временной ряд $N(t)$ разбивается на L отрезков. На каждом отрезке случайным образом отбирается точка x_i и рассчитывается расстояние до остальных точек на отрезке – p_1, p_2, \dots, p_k . Полученные

значения сортируются в порядке возрастания и из них составляется новое множество, необходимое для нахождения предела.

$$D(x_i) = \lim_{p_{k_j} \rightarrow 0} \frac{\log P(x_i, p_{k_j})}{\log p_{k_j}}. \quad (18)$$

Из полученного множества исключаются значения, точки которых лежат на нелинейных участках и вычисляется величина $D(x_i)$ методом линейной аппроксимации. Сумма всех значений $D(x_i)$ на каждом из отрезков, поделённая на их количество и есть значение фрактальной размерности Хаусдорфа.

Блок-схема данного алгоритма представлена на рисунках 20 – 21.

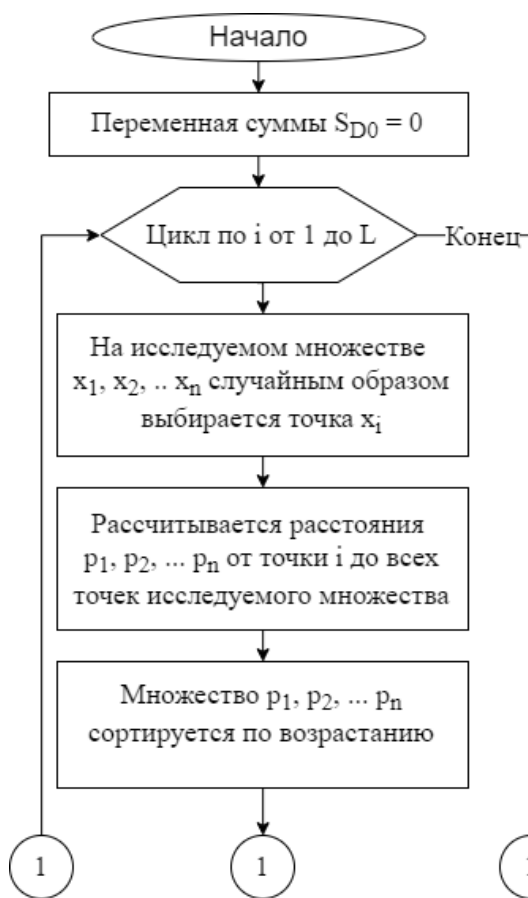


Рисунок 20 – Блок-схема алгоритма поточечного метода

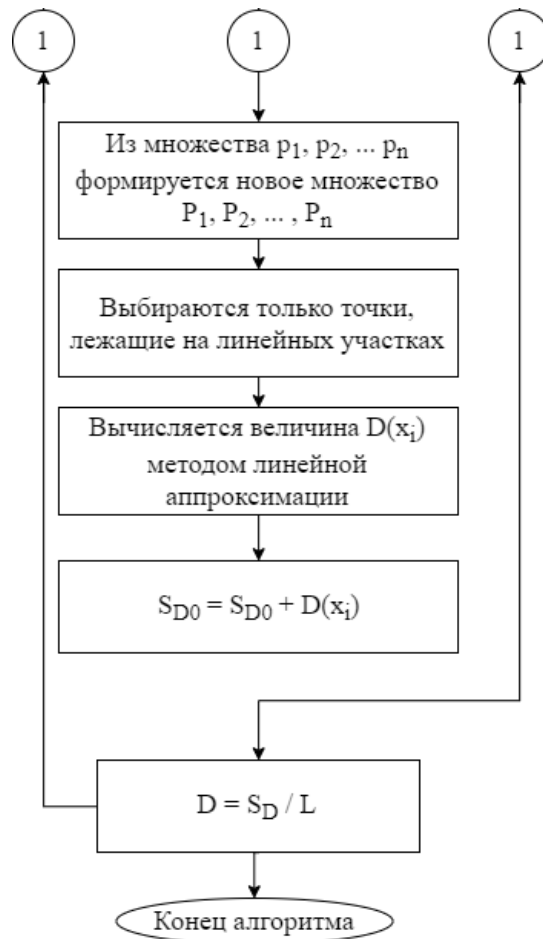


Рисунок 21 – Блок-схема алгоритма поточечного метода

Из достоинств данного метода стоит отметить относительную простоту вычислений, а из недостатков сильную погрешность и неустойчивость к «зашумлённым» данным.

2.6.3 Метод R/S анализа

При увеличении числа испытаний, предельное распределение случайной системы будет нормальным, – утверждает центральная предельная теорема в стандартной гауссовой статистике. Это означает, что все события в рассматриваемой системе должны иметь одинаковую вероятность наступления и не влиять друг на друга. Чтобы произвести статистический анализ, при изучении больших систем, часто предполагают гипотезу о нормальности системы. Но на практике это не всегда так просто. Большинство больших и сложных систем, как например значения выпадения

среднегодовых осадков, финансовые показатели и другие, не являются нормально распределёнными. Кроме того, многие из них даже не близки к тому, чтобы быть таковыми.

Для анализа таких систем Херст предложил метод нормированного размаха или R/S анализ. Данный метод позволяет различать случайные и фрактальные ряды, а так же делать выводы о наличии неперiodических циклов, долговременной памяти и так далее [7].

Алгоритм R/S анализ можно описать следующим образом: изначально берётся выборка из $n + 1$ элементов, при этом n должна быть кратна двум ($n : 2$). Чем больше выборка, тем более качественный получится результат.

Для начала выборка преобразовывается при помощи логарифмических соотношений. В итоге получается временной ряд со значениями уровней длиной n .

$$\log \frac{U_i}{U_{i-1}}, \quad (19)$$

где U_i – элемент временного ряда.

Следующим шагом находится m – наименьший собственный делитель для n , при этом он должен быть не меньше десяти.

Преобразованная выборка делится на группы длиной m . И для каждой группы находится среднее значение.

$$\bar{t}_k = \frac{1}{m} \sum_{i=1}^m t_i, \frac{1}{m} \sum_{i=m+1}^{2m} t_i, \dots, \frac{1}{m} \sum_{i=(k-1)m+1}^n t_i. \quad (20)$$

Затем рассчитывается накопленные отклонения от среднего X_i .

$$\begin{aligned}
X_1 &= t_1 - \frac{1}{m} \sum_{i=1}^m t_i, X_2 = (t_2 - \frac{1}{m} \sum_{i=1}^m t_i) + X_1, \dots, \\
X_m &= (t_m - \frac{1}{m} \sum_{i=1}^m t_i) + X_{m-1}.
\end{aligned}
\tag{21}$$

Таким образом, нормированный размах по каждой группе можно рассчитать формуле:

$$R_k = \max(X_{i,k}) - \min(X_{i,k}). \tag{22}$$

Стандартное отклонение S_k для каждой группы рассчитывается по формуле ниже:

$$S_k = \sqrt{\frac{1}{m} \sum_{i=1}^m (t_i - \bar{t}_k)^2}. \tag{23}$$

Для каждой группы рассчитывается R/S показатель.

$$\overline{R/S}_j = \frac{1}{k} \sum_{i=1}^m R/S_i, \tag{24}$$

Таким образом, получается значение R/S показателя под индексом j . Здесь индекс j соответствует номеру наименьшего собственного делителя m . В итоге по нескольким значениям индекса R/S показателя можно построить уравнение линейной регрессии (формула 25)

$$\log R/S - \log c + H * \log k. \tag{25}$$

2.6.4 Оценка показателя Херста

В 1976 А. Энис и Е. Лойд предложили в своей работе «Ожидаемое значение скорректированного измененного диапазона Херста независимых нормальных слагаемых» [4] термин ожидаемое значение Херста. Они вывели уравнение, выражающее ожидаемое значение:

$$E = \left(n * \frac{\pi}{2}\right)^{-0.5} * \sum_{r=1}^{n-1} \sqrt{\frac{n-r}{r}}, \quad (26)$$

где n – количество наблюдений;

r – целые числа в диапазоне $\overline{1, n - 1}$.

Как отмечалось в работе «Фрактальный анализ финансовых рынков», выведенная формула 26 справедлива только для значений n больших двадцати. Если же n меньше или равно двадцати, то следует использовать её модификацию:

$$E = \frac{(n - 0.5)}{n} * \left(n * \frac{\pi}{2}\right)^{-0.5} * \sum_{r=1}^{n-1} \sqrt{\frac{n-r}{r}}, \quad (27)$$

где n – количество наблюдений;

r – целые числа в диапазоне $\overline{1, n - 1}$.

Для вычисления погрешности показателя Херста необходимо:

- Рассчитать для каждого количества наблюдений показатель Херста (H) и ожидаемый показатель Херста (E);
- Рассчитать ожидаемую дисперсию показателя Херста (H) можно при помощи формулы:

$$\sigma(H) = \frac{1}{N}, \quad (28)$$

где H – показатель Херста;

N – число наблюдений в выборке;

- Проверить значимость рассчитанного коэффициента Херста через оценку количества стандартных отклонений, на которые H превосходит E . Результат можно считать значимым в том случае, если показатель больше двух по модулю.

2.7 Самоподобные процессы

2.7.1 Введение в самоподобные процессы и их классификации

Самоподобие – это наиважнейшая особенность фрактала. Это можно описать как инвариантность при подходящем масштабировании времени или пространства [11].

В целом существует четыре вида самоподобия.

Точное самоподобие означает, что каждая небольшая часть объекта является точной копией целого, и это свойство применяется ко всем масштабам. Кривая Коха – это отличный пример точного самоподобия (рисунок 11). Однако это идеальный случай, который практически не встречается в реальности. Для большинства фракталов в природе структура остается лишь похожей на разных масштабах, а не точно совпадает. Примерами таких фракталов являются папоротники, снежинки и другие объекты.

Квази-самоподобием называют объекты, у которых структура на разных масштабах приближаются к одному и тому же виду. Части такого фрактала на более мелких масштабах могут содержать в себе уменьшенные копии исходного фрактала в ярко выраженном и искажённом виде. В отличие от точного самоподобия, части фрактала являются лишь приближенными к

фракталу, а не точными копиями. Примером квази-самоподобия можно привести спутники множества Мандельброта (рисунок 22).

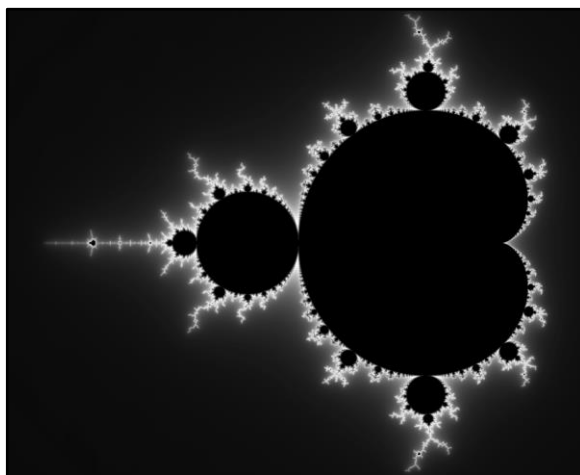


Рисунок 22 – Множество Мандельброта

Мультифрактал – это еще один вид самоподобия, который имеет несколько фрактальных измерений. Можно сказать, что свойство самоподобия различно на разных масштабах.

Статистическое самоподобие – последнее, но наиболее важное свойство. Им обладают фракталы, у которых части повторяют исходную структуру стохастически, то есть стохастические показатели сохраняются на разных масштабах [16]. Примерами таких фракталов являются случайно сгенерированные фракталы, береговая линия (как, например, Британская береговая линия, продемонстрированная на рисунке 13) и временные ряды. Для таких фракталов нельзя ожидать найти точный повторяющийся или похожий сегмент, масштабируемый и повторяющийся так же точно, как и в случае точного самоподобия и квази-самоподобия.

Фрактальные временные ряды статистически самоподобны в том смысле, что сигналы имеют одно и то же статистическое свойство независимо от шкалы времени, в которой они измеряются.

2.7.2 Свойства самоподобия и долговременной памяти

Пусть процесс $\{X(t), t \in \mathbb{R}\}$ является стационарным. Если $\sum_n \rho(t) = +\infty$, то можно сказать, что $X(t)$ является процессом с долговременной памятью. Достаточным условием для этого является существование параметра Херста $H \in (0.5, 1)$, который бы удовлетворял выражению:

$$\lim_{t \rightarrow \infty} \frac{\rho(t)}{t^{2H-2}} > 0. \quad (29)$$

Процесс с долговременной памятью обладает медленно затухающей автокорреляционной функцией.

Процесс $\{X(t), t \geq 0\}$ является самоподобным с параметром $H \in (0, 1)$, если для всех c больше нуля распределение является одинаковым.

$$X(ct) \sim c^H X(t), \quad (30)$$

где H – параметр Херста, который очень важен для характеристики самоподобного объекта.

В идеальном случае уравнение 30 должно выполняться для всех $c, t \in \mathbb{R}$. Однако в действительности это свойство сохраняется только в ограниченных масштабах. Ковариационная функция для самоподобного процесса $X(t)$ представлена ниже:

$$\text{Cov}(X_t, X_s) = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}). \quad (31)$$

Из формул 30 и 31 можно сделать заключение, что самоподобные процессы должны быть нестационарными [17]. Это отличает их от процессов с долговременной памятью, которые требуют стационарности.

На практике сигналы записываются дискретно, а не непрерывно. Если взять дискретную временную выборку процесса $X(t)$, можно получить временной ряд X_n . Среди самоподобных процессов интересна группа, которая имеет стационарные приращения, потому что такие процессы могут создавать стационарные последовательности [17]. Пусть X_n – это самоподобный процесс со стационарным приращением, а Y_n – его процесс приращения, тогда можно описать их взаимосвязь:

$$Y_n = X_{n+1} - X_n. \quad (32)$$

В таком случае временной ряд Y_n стационарный. Определим $\{Y_i^m, i = 1, \dots, n\}$ как совокупный процесс для Y_n , тогда можно получить следующее разложение:

$$Y_i^{(m)} = \frac{1}{m} (Y_{im-m+1} + \dots + Y_{im}). \quad (34)$$

Из этого следует, что масштабная инвариантность для дискретного стационарного временного ряда должна имеет одинаковое распределение.

$$m^{1-H} Y^{(m)} \sim Y, \quad (35)$$

где H – параметр Херста для процесса Y_n , который совпадает с параметром Херста для исходного самоподобного процесса $X(t)$.

Свойство масштабирования, описанное в формуле 35, предполагает, что усреднённая последовательность имеет то же распределение, что и исходный временной ряд после надлежащего масштабирования в соответствии с агрегированными уровнями (масштабами) [17]. Примеры представлены на рисунке 23.

Самоподобие и долговременная память тесно связаны, но это разные понятия. Основное различие между самоподобными процессами и процессами с долговременной памятью заключается в том, что самоподобные процессы нестационарные, в то время как процессы с долговременной памятью являются стационарными по определению.

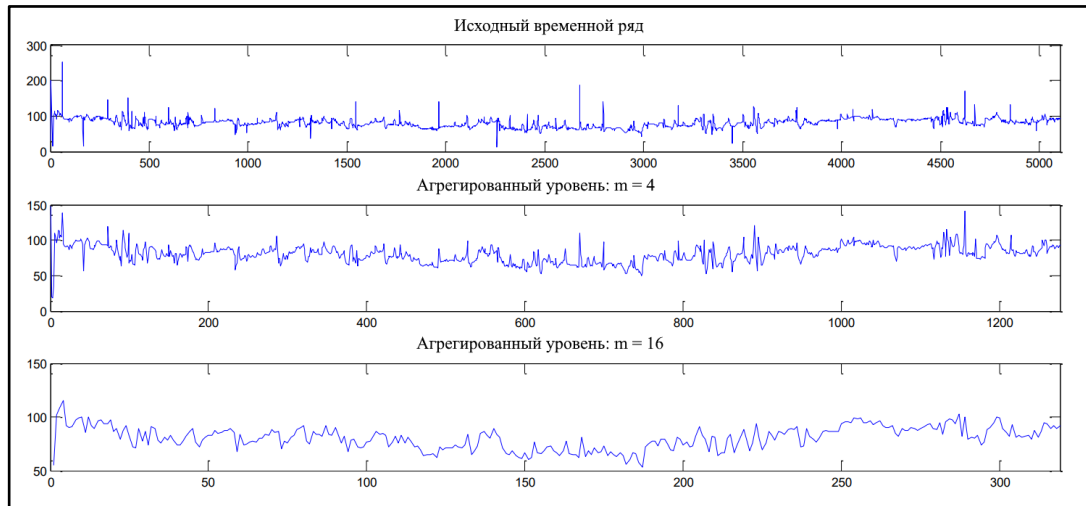


Рисунок 23 – Исходный временной ряд и его агрегированные уровни. На среднем и нижнем графике представлены временные ряды, агрегированные в разных масштабах, m равное 4 и m равное 16, соответственно.

Однако эти два типа процессов связаны единым параметром – показателем Херста, из-за чего один процесс может быть производным от другого. Для самоподобных процессов X_n со стационарными приращениями его процесс приращения Y_n является стационарным. Пусть $\gamma_Y(k)$ обозначает ковариацию для Y_n , тогда $\gamma_Y(k)$ можно описать формулой:

$$\gamma_Y(k) = \frac{\sigma^2}{2} [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]. \quad (36)$$

Если параметр Херста для X_n и Y_n удовлетворяет условию $\frac{1}{2} < H < 1$, то сумма $\gamma_Y(k)$ будет отклоняться или будет выполняться отношение:

$$\sum_{k=-\infty}^{\infty} \gamma_Y(k) = \infty. \quad (37)$$

И этого следует, что Y_n является процессом с долговременной памятью.

Можно сделать следующее заключение: если нестационарный временной ряд самоподобный со стационарными приращениями, а его параметр Херста находится между 0.5 и 1, то соответствующий ему процесс приращения – это и есть процесс с долговременной памятью. И наоборот, если стационарный временной ряд – это процесс с долговременной памятью, то его совокупный процесс нестационарен и самоподобен.

2.8 Фрактальное Броуновское движение и Гауссовский шум

Фрактальное Броуновское движение является типичным примером самоподобным процессом со стационарным приращением. Это обобщение стандартного броуновского движения, которое можно определить с помощью стохастического исчисления [18].

$$\begin{aligned} B_H(t) = & \frac{1}{\Gamma\left(H + \frac{1}{2}\right)} \left(\int_{-\infty}^0 [(t-s)^{(H-\frac{1}{2})} - (-s)^{(H-\frac{1}{2})}] dB(s) + \right. \\ & \left. + \int_0^t (t-s)^{(H-\frac{1}{2})} dB(s) \right), \end{aligned} \quad (38)$$

где $B(t)$ – это стандартное Броуновское движение, которое можно восстановить, взяв значение H равное 0.5.

Фрактальное Броуновское движение – это единственный самоподобный Гауссовский процесс со стационарными приращениями, имеющий параметр Херста в пределах от нуля до единицы [17].

Стандартное Броуновское движение – это частный случай фрактального Броуновского движения с параметром Херста равным 0.5. Он

не является стационарным, но имеет независимые стационарные приращения, что отражает его абсолютную случайность. Это так же говорит о том, что даже если в конкретный момент времени наблюдается возрастание или убывание временного ряда, всё равно не известно в каком направлении он будет развиваться далее. Фрактальное Броуновское движение очень похоже на стандартное Броуновское движение в том плане, что оно так же является нестационарным, но со стационарными приращениями. Однако для фрактального Броуновского движения с параметром Херста не равным 0.5, приращения уже не являются независимыми. Если параметр Херста H больше 0.5, то возрастающий временной ряд с большей вероятностью будет увеличиваться в будущем с течением времени [13]. Таким образом, параметр Херста H описывает резкость фрактального Броуновского движения. По мере увеличения параметра Херста, колебаний становится меньше и кривая временного ряда становится более гладкой. На рисунке 24 представлены показатели фрактального Броуновского движения с разными параметрами Херста H . При увеличении масштаба графика (рисунок 24 (в)) можно наблюдать небольшую часть временного ряда, повторяющую целую часть (рисунок 24 (г)).



Рисунок 24 – Фрактальное Броуновское движение при разных показателях параметрах Херста H . Фрактальное Броуновское движение при H равное 0.1 (а); Стандартное Броуновское движение (оно же фрактальное Броуновское движение при H равное 0.5) (б); Фрактальное Броуновское движение при H равное 0.9; Фрактальное Броуновское движение при H равное 0.9 (приближенная)

Пусть существует итеративный процесс $Y = \{Y_k, k = 1, 2, \dots\}$, который удовлетворяет условию:

$$Y_k = B_H(k+1) - B_H(k), \quad (39)$$

где $B_H(k)$ – это Броуновское движение с показателем Херста H .

В таком случае Y_k называется фрактальным Гауссовским шумом. Так же он является дискретным пошаговым итеративным процессом фрактального Броуновского движения [16]. Стандартное Броуновское движение – это, по своей сути, кумулятивная сумма белого шума. И точно так же фрактальное Броуновское движение является кумулятивной суммой фрактального Гауссовского шума [18].

Фрактальный Гауссовский шум является стационарным, и его ковариационная функция является аналогичной формуле 36. Если показатель Херста H равен 0.5, то ковариационная функция равна нулю для всех $k \neq 0$. Это описание белого шума в случае, когда все наблюдения полностью независимы. Когда показатель Херста H больше 0.5, Y_k является процессом с долговременной памятью. Автокорреляционная функция для фрактального Гауссовского шума с параметром Херста H больше 0.5 затухает гораздо медленнее, чем для обычного белого шума. На рисунке 25 представлено сравнение фрактального Броуновского движения и фрактального Гауссовского шума с долговременной памятью, показатель Херста H которых равен 0.9.

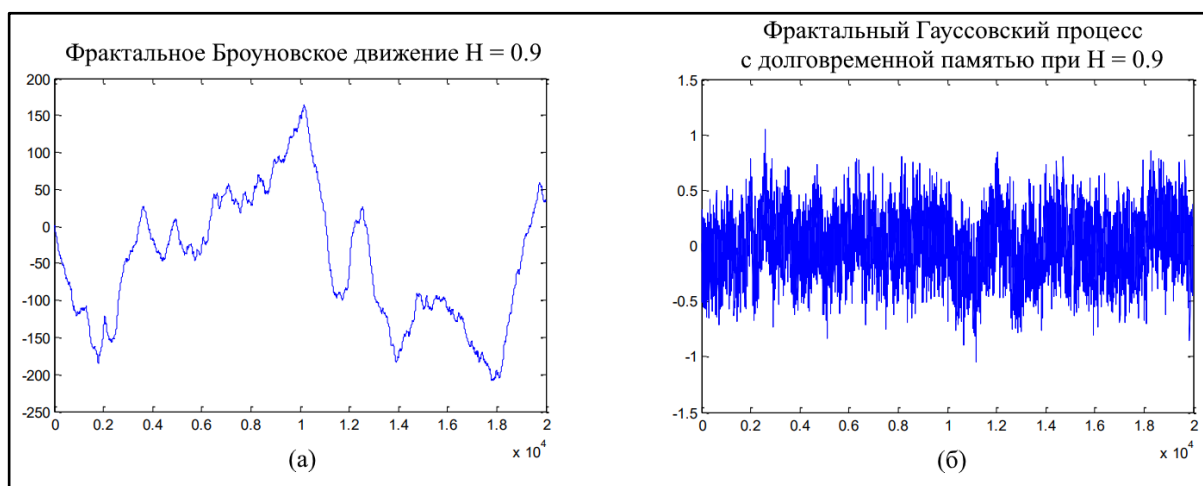


Рисунок 25 – Сравнение фрактального Броуновского движения с параметром Херста H равным 0.9 (а) и фрактального Гауссовского шума с долговременной памятью с показателем Херста H равным 0.9 (б)

Вывод по главе 2

В данной главе были представлены теоретические сведения самоподобных процессов и процессов с долговременной памятью. Помимо этого, были проиллюстрированы некоторые ключевые концепции и статические свойства.

Параметр Херста – это один из самых важных параметров, который может характеризовать самоподобные сигналы и сигналы с

долговременной памятью. Он принимает значение в диапазоне от нуля до единицы. Для изучения интересны процессы с показателем Херста H больше 0.5, потому что они обладают фрактальными свойствами.

Как самоподобные, так и процессы с долговременной памятью обладают свойством масштабной инвариантности, но, как уже упоминалось в данной главе, масштабная инвариантность имеет разные формы для самоподобных процессов и процессов с долговременной памятью. Основная причина заключается в том, что самоподобные процессы и процессы с долговременной памятью теоретически не идентичны, но могут быть пересекаться. В конце данной главы приводятся пример типичного самоподобного процесса, которым является фрактальное Броуновское движение, и ему противопоставляется фрактальный Гауссовский процесс.

Глава 3 Алгоритмы и методы

3.1 Моделирование временного ряда

Моделирование любого временного ряда состоит из следующих шагов: сначала анализируется автокорреляционная функция для определения стационарности временного ряда. Затем подбирается подходящая модель и проводится оценка параметров. И, наконец, применяется модель для составления прогнозов [19]. В приложениях, работающих в реальном времени так же необходимо обновление модели в режиме «онлайн».

Описать математическую модель модели прогнозирования временного ряда можно следующим образом: пусть имеется временной ряд $y_1, \dots, y_T, \dots, y_t \in \mathbb{R}$, который является фиксацией некоторого признака на постоянных промежутках времени. Задача прогнозирования временного ряда сводится к поиску функции f_T .

$$y_{T+d} \approx f_T(y_T, \dots, y_1, \mathbf{d}) \equiv \hat{y}_{T+d|T}, \quad (40)$$

где $d \in \{1, \dots, D\}$ – отсрочка прогноза;

D – горизонт прогнозирования.

Самоподобные сигналы обладают статистическими свойствами отличными от обычных, а наличие долговременной памяти затрудняет анализ таких сигналов. Однако уникальные свойства масштабирования предоставляют другие подходы к анализу и моделированию таких систем.

В данном разделе будет рассмотрена общая методика моделирования временных рядов и некоторые модели, способные описать самоподобные процессы и процессы с долговременной памятью, такие как FARIMA, HAR-RV, вейвлет-преобразования и Average-VAR. Некоторые из этих моделей, как например FARIMA, существуют со времён развития фрактальной

геометрии. А некоторые, такие как Average-VAR, были разработаны совсем недавно.

3.2 Модель FARIMA

Говоря про моделирование временных рядов нельзя игнорировать знаменитую модель ARIMA или метод Бокса-Дженкинса, названный в честь статистиков Джорджа Бокса и Гвилима Дженкинса [19, 20]. Название ARIMA является сокращением от понятия «авторегрессионное интегрированное скользящее среднее» (на английском языке «autoregressive integrated moving average»). Данная модель представляет собой класс линейных моделей, которые могут описывать как стационарные, так и нестационарные временные ряды. Если временной ряд нестационарен, стационарный ряд может быть получен путём дифференцирования исходного ряда, а степень различия зависит от свойств исходного временного ряда. Если исходный временной ряд показывает линейный тренд, достаточно лишь одного дифференцирования. Если же результирующий временной ряд по-прежнему нестационарен, необходимо применить повторное дифференцирование. После надлежащего дифференцирования автокорреляционная функция должна быстро экспоненциально затухать до нуля [19]. Однако степень разности может принимать только целые числа в традиционной модели ARIMA.

В предыдущей главе были описаны группы стационарных процессов, автокорреляционная функция которых медленно убывает, а их сумма расходится. Модель ARIMA не может быть применена к этим фрактальным временным рядам из-за её неспособности описать постоянство. В 1983 году Джонатаном Р. М. Хоскинком было предложено обобщение модели ARIMA, которое было названо FARIMA. Название данной разновидности модели является аббревиатурой от понятия «фрактальное авторегрессионное интегрированное скользящее среднее» [21].

Модель FARIMA успешно описывает свойство долговременной памяти временных рядов и уже не раз применялась для моделирования процессов с долговременной памятью в экономике, гидрологии [17], а так же в информатике [22]. Для этого она вводит дробное дифференцирование, которое позволяет принимать дробные числа степеням дифференцирования [19].

Модель ARIMA, которая обычно используется для моделирования прогнозов временного ряда, может моделировать процессы, для которых автокорреляционная функция сходится к нулю достаточно быстро практически сразу или после надлежащего дифференцирования [20]. Однако для временных рядов с долговременной памятью автокорреляционная функция затухает куда медленнее, что приводит к сильной корреляции между двумя наблюдениями, расположенными далеко друг от друга. Модель FARIMA, предложенная Хоскингом, позволяет определять степень дифференцирования дробными числами, что позволяет уловить эту сильную автокорреляцию [21].

Общая модель FARIMA (p, d, q) может быть определена выражением [21]:

$$\Phi(B)(1 - B)^d X_t = \Theta(B)e_t, \quad (41)$$

где $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots$;

$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots$;

B – оператор обратного сдвига, $d \in [-0.5, 0.5]$ и определяет степень разности;

X_t – моделируемый процесс;

e_t – это ошибка.

Модель ARIMA (p, d, q) имеет тот же вид. Единственное различие между этими двумя моделями заключается в том, что d может принимать

дробные значения в модели FARIMA, в то время как в ARIMA модели d может быть только целым. Исследования показали, что степень разности d связана с показателем Херста H отношением [21]:

$$d = H - \frac{1}{2}. \quad (42)$$

Для $d \in (0, 0.5)$ процесс X_t в уравнении 41 обладает долговременной памятью и может моделировать временной ряд с долговременной памятью [22]. Параметр дробной разности может быть вычислен при помощи формулы:

$$\begin{aligned} (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{1}{2}d(1-d)B^2 - \\ - \frac{1}{6}d(1-d)(2-d)B^3 - \dots \end{aligned} \quad (43)$$

Оценка параметров имеет большое значение при применении модели FARIMA. Учитывая временной ряд с долговременной памятью, первым шагом является оценка параметра Херста H и получение значения d , используя отношение 42. Существует множество методов оценки параметра Херста H , включая R/S анализ, который был описан в главе 2.6.3.

После того, как получен параметр дифференцирования d , необходимо выполнить дробное дифференцирование временного ряда, чтобы удалить постоянство и преобразовать исходный временной ряд с долговременной памятью во временной ряд с кратковременной памятью. Только после этого могут применяться алгоритмы оценки параметров временного ряда, такие как оценка максимального правдоподобия, оценка наименьших квадратов и другие.

Модель FARIMA привлекла к себе внимание сразу, как была впервые предложена. Это обусловлено её способностью моделировать устойчивый временной ряд с долговременной памятью.

Не смотря на очевидное достоинство перед линейными моделями временных рядов с долговременной памятью, FARIMA так же обладает и недостатками:

- Параметр разности d необходимо оценить при помощи параметра Херста H . Однако, существующие методы оценки показателя Херста, к сожалению, не всегда показывают последовательность в своих вычислениях. Различные методы могут давать разные результаты, тем самым вызывая большую ошибку в оценке параметра разности d . Кроме того, оценка параметра Херста H требует использования всех точек данных, что невозможно, например, в приложениях реального времени [29].
- Оператор дробного разложения, представленный в формуле 43, представляет собой бесконечную сумму, которая неприменима в приложениях, работающих в реальном времени. Это может быть критично при анализе сигналов, приходящих в реальном времени. Так же при применении формулы 43 часто выбирается определённый отрезок данных, что приводит к большим ошибкам из-за низкой точности.
- Оценка параметров очень сложна и требует слишком много времени для выполнения всего процесса вычислений.

3.3 Модель HAR-RV

В финансовой сфере понятие волатильности является мерой риска и, следовательно, очень важна для ценообразования активов. Говоря более простыми терминами, волатильность – это изменчивость цены. Резкое падение или рост цены обычно приводит к росту волатильности. А когда,

например, цены долгое время колеблется около одной отметки, то волатильность снижается.

К примеру, 14 мая 2018 года акции компании ООО «Газпром» поднялись на 21% в цене всего за 6 часов (рисунок 26).



Рисунок 26 – Рост цен акций компании ООО «Газпром» на 20.8% за 6 часов за 14 мая 2018 года.

Достаточно большая редкость, когда акции компании меняются так быстро и резко, как правило, это происходит из-за влияния внешних факторов, как, собственно, и произошло в данной ситуации: управление компании рекомендовало выплатить акционерам дивиденды за каждую акцию в 2018 году в два раза больше, чем по итогам 2017 года. Это и послужило причиной того, что многие участники торгов поспешили раскупить акции ООО «Газпрома». Из-за этого спрос на акции превысил предложение, поэтому цена так сильно и резко возросла, а, следовательно, волатильность тоже.

Другой пример – цена акций компании ООО «Яндекс» 11 октября 2019 года, которые упали на 18.5% (рисунок 27).



Рисунок 27 – Падение цен акций компании ООО «Яндекс» на 18.5% за 1 день за 11 октября 2019

Это так же произошло из-за внесения в Госдуму новых законов, которые бы ограничивали доли иностранных компаний и физических лиц в капитале значимых ресурсов до отметки не более 20%.

Широко используется два вида волатильности – это реализованная и подразумеваемая волатильность.

Самым правильным определением волатильности является стандартное отклонение ряда доходности актива. Поскольку в наши дни доступны высокочастотные торговые данные, то дневная реализованная волатильность может быть вычислена как сумма квадратов внутридневной доходности [23].

Наиболее популярным способом подсчёта дневной реализованной волатильности является суммирование всех квадратов пяти минутной доходности за один торговый день [23]. Согласно гипотезе эффективного рынка, доходность практически не коррелирована, а параметр Херста $H \approx 0.5$. Однако исследования показывают, что как абсолютная, так и квадратичная доходность положительно коррелированы [24]. Реализованная волатильность – это сумма квадратов доходности, и это процесс с

долговременной памятью, у которого показатель Херста $H \in (0.5, 1)$. Помимо этого обнаружено, что реализованная волатильность имеет кластерное поведение. Это указывает на то, что большие волатильности имеют тенденцию появляться последовательно.

Модель HAR-RV предложена Ф. Корси в 2009 году [28, 29]. Он обнаружил, что агрегирование нескольких моделей авторегрессии может привести к процессу, который демонстрирует довольно похожее поведение с процессами, обладающими длительной памятью, а функция авторегрессии, созданная моделью HAR-RV, медленно затухает, что близко к процессу с долговременной памятью [28, 29]. Ф. Корси протестировал модель HAR-RV на данных фьючерсов USD/CHF и казначейских облигаций. Для этого он спрогнозировал одно значение на один день [29]. Оказалось, что простая модель HAR-RV может улавливать поведение долговременной памяти и обеспечить аналогичную точность более сложной FARIMA [29].

Модель использует реализованную волатильность, агрегированную на недельной и ежемесячной основе, и представляет будущую волатильность как взвешенную сумму волатильности в той же шкале и на более высоких уровнях [28, 29]. Эта модель может улавливать краткосрочные вариации, а также долгосрочный тренд [29].

Модель HAR-RV можно сформулировать следующим образом. Пусть $p_{t,j}$ – это цена актива в день t на конец внутридневного интервала j длины Δ , $j = 1, \dots, N$ с общим числом таких интервалов за один день, равным N . Тогда $r_{t,j} = \log(p_{t,j}) - \log(p_{t,j-1})$ – внутридневная доходность актива на временном интервале длины Δ .

Реализованную волатильность в день t можно выразить формулой:

$$RV_t = \sum_{j=1}^N r_{t,j}^2. \quad (44)$$

Тогда будущую волатильность можно сформулировать через уравнение:

$$RV_{t+1d}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + e_{t+1d}, \quad (45)$$

где $RV_t^{(d)}$ – суточная реализованная волатильность в момент времени t ;

$RV_t^{(w)}$ – недельная реализованная волатильность в момент времени t ;

$RV_t^{(m)}$ – месячная реализованная волатильность в момент времени t ;

$RV_{t+1d}^{(d)}$ – прогнозируемая суточная реализованная волатильность на время $t + 1$,

e_{t+1d} – случайный шок.

Еженедельная и ежемесячная реализованная волатильность рассчитывается как среднее из последних дневных реализованных волатильностей. На практике считается, что в неделе только пять торговых дней, а в месяце двадцать два торговых дня. Таким образом, недельная и ежемесячная реализованная волатильность суммируется по пять по двадцать два дня соответственно. Коэффициенты можно оценить при помощи метода аппроксимации наименьших квадратов.

Даже простая модель HAR-RV может давать вполне удовлетворительные результаты по сравнению со сложной моделью FARIMA. Однако модель HAR-RV использовалась только в контексте реализованной волатильности, которая имеет естественные агрегационные уровни. Если этот алгоритм применять на других типах временных рядов, то встаёт острый вопрос о выборе агрегационных уровней.

3.4 Вейвлет-модель

Вейвлет-анализ – это частотно-временной анализ, который широко используется в технических областях, таких как сжатие изображений, шумоподавление сигнала и так далее.

Вейвлет – это функция $\psi(t)$, которая удовлетворяет условию [30]:

$$\int_{\mathbb{R}} \psi(t) dt = 0. \quad (46)$$

Вейвлет Хаара является простейшим вейвлетом, предложенным венгерским математиком Альфредом Хааром в 1909 году. Он представляет собой простую кусочно-постоянную функцию, определяемой уравнением:

$$\psi(t) = \begin{cases} 1, & \text{если } 0 \leq t < \frac{1}{2} \\ -1, & \text{если } \frac{1}{2} \leq t < 1 \\ 0 & \text{иначе} \end{cases}. \quad (47)$$

Функция $\psi(t)$ называется материнским вейвлетом, а его дочерний вейвлет $\psi_{j,k}(t)$ можно получить при помощи расширения материнского вейвлета [50]:

$$\psi_{j,k}(t) = \frac{1}{2^{j/2}} \psi(2^{-j}t - k), j, k \in \mathbb{Z}. \quad (48)$$

Дискретное вейвлет преобразование процесса $X(t)$ может быть определено выражением:

$$d_{j,k} = \int_{\mathbb{R}} X(t) \psi_{j,k}(t) dt, j, k \in \mathbb{Z}, \quad (49)$$

где $d_{j,k}$ называются вейвлет-коэффициентами.

Для каждого материнского вейвлета существует соответствующая функция масштабирования $\phi(t)$, позволяющая сгенерировать последовательность пространств V_j , которые могут аппроксимировать функции из пространства $x^2(\mathbb{R})$. Порождённое вейвлет-функцией $\psi(t)$ пространство обозначается как W_i , что так же можно рассматривать как ошибку, возникающую при аппроксимации функций из V_{j+1} функциями из V_j (рисунок 28) [31].

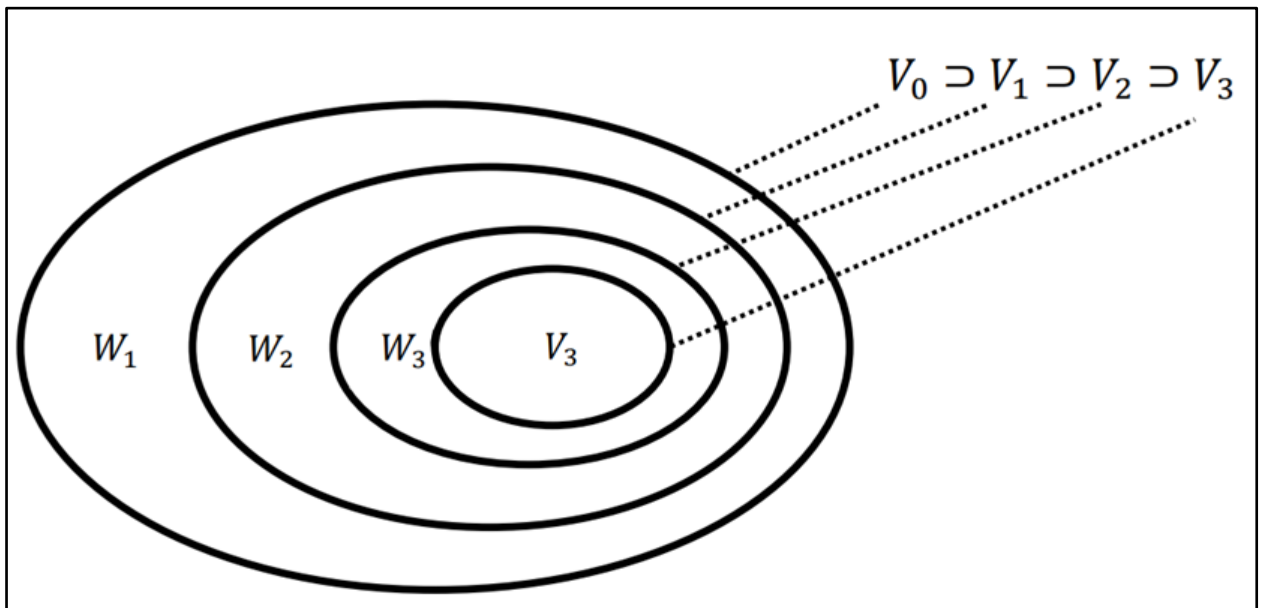


Рисунок 28 – Иллюстрация диаграммы вейвлет-пространств

На рисунке 28 показано следующее отношение:

$$V_0 = V_1 \oplus W_1 = V_2 \oplus W_2 \oplus W_1 = V_3 \oplus W_3 \oplus W_2 \oplus W_1 = \dots \quad (50)$$

Любую функцию $f(t) \in V_0$, можно разложить на две части: первая – это аппроксимация в более грубых масштабах (V_1), а вторая – это детали в более мелких масштабах (W_1). Это называется анализом в нескольких масштабах [31], которое можно выразить через уравнение 51.

$$f(t) = \sum_{k=-\infty}^{\infty} a_{j,k} \phi_{j,k}(t) dt, j, k \in \mathbb{Z}, \quad (51)$$

где $d_{j,k}$ – вейвлет-коэффициенты, которые могут быть получены при помощи формулы 49;

$a_{j,k}$ – это коэффициент аппроксимации и может быть вычислен при помощи формулы:

$$a_{j,k} = \int_{\mathbb{R}} f(t) \phi_{j,k}(t) dt, j, k \in \mathbb{Z}. \quad (52)$$

На рисунке 29 представлена диаграмма многомерного вейвлет-разложения на трёх масштабах.

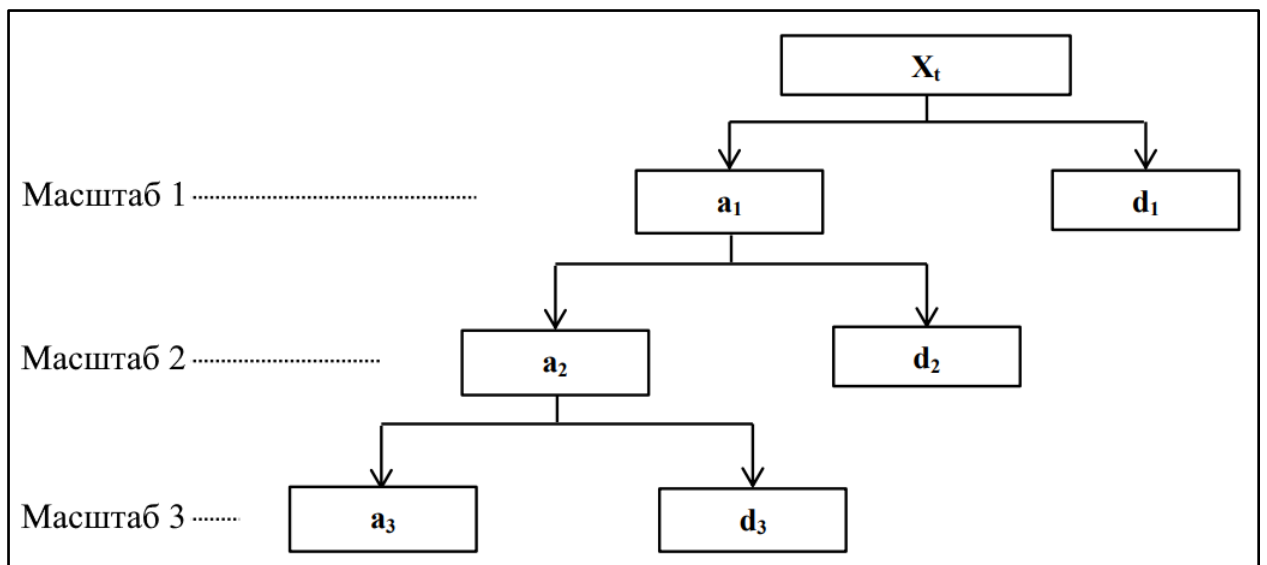


Рисунок 29 – Диаграмма многомерного вейвлет-разложения на трёх масштабах

На рисунке 30 представлено то же разложение на трёх масштабах для артериального давления.

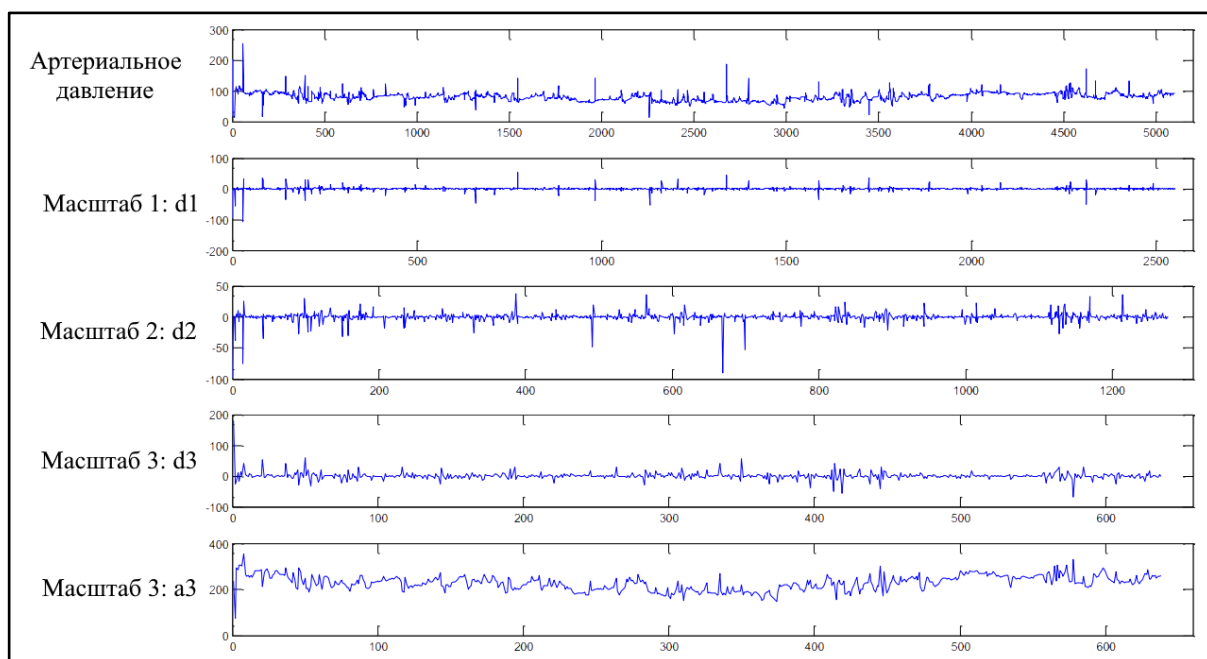


Рисунок 30 – Пример вейвлет-разложения на трёх масштабах артериального давления при помощи Хаар вейвлета

Исчезающие моменты R являются очень важным свойством вейвлет-функции. Вейвлет функция обладает исчезающим моментом R если выполняется условие:

$$\int x^l \psi(x) dx = 0, \quad l = \overline{0, R-1}. \quad (53)$$

Данное условие предполагает, что вейвлет-функция с нулевыми моментами R может генерировать полиномы степени меньше R [30, 31].

Вейвлеты Добеши – это семейство ортогональных вейвлетов с заданными носителями. Они названы в честь математика из США, который сформулировал данное семейство вейвлетов, – Ингрида Добеши [31, 32]. Вейвлет Добеши с исчезающими моментами R имеет опорную ширину $N = 2R - 1$ [32]. Аппроксимация на масштабах R может быть использована для демонстрации вейвлета Добеши с исчезающими моментами R . Вейвлеты Добеши обычно не имеют явно выраженных моментов, за исключением вейвлета Хаара, который является частным случаем вейвлета Добеши с

моментом и шириной опоры равными единице [32]. На рисунке 31 показано приближение некоторых вейвлет-функций Добеши.

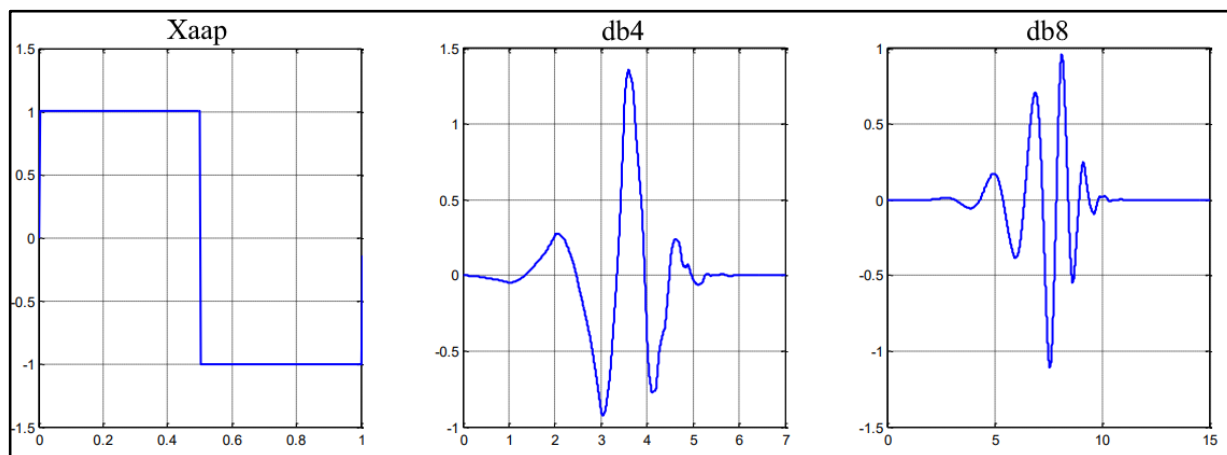


Рисунок 31 – Аппроксимация некоторых вейвлет-функций Добеши

Вейвлет-анализ может применяться и к самоподобным сигналам. Исследования показали, что ковариации вейвлет-коэффициентов для самоподобных сигналов являются стационарными на каждом масштабе [31, 32]. К примеру, автокорреляционная функция для фрактального Броуновского движения затухает медленно, поскольку фрактальное Броуновское движение не стационарно. Однако, после вейвлет-разложения, автокорреляционная функция для вейвлет-коэффициентов на одном масштабе затухает намного быстрее, чем автокорреляция самого фрактального Броуновского движения. На рисунке 32 представлено сравнение автокорреляционных функций для фрактального Броуновского движения и его вейвлет-коэффициентов при использовании вейвлета Хаара.

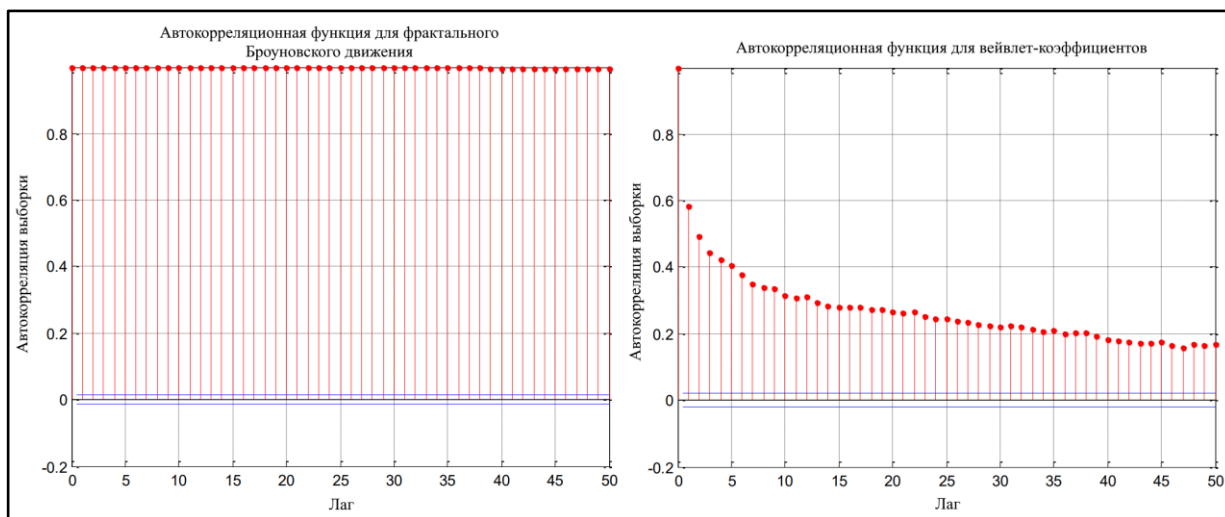


Рисунок 32 – Сравнение автокорреляционной функции для фрактального Броуновского движения до вейвлет-преобразования (слева) и после (справа)

Скорость убывания вейвлет-коэффициентов тесно связана с исчезающими моментами R используемого вейвлета. Большие исчезающие моменты R могут привести к почти некоррелированным вейвлет-коэффициентам [32, 33]. Если R выбрано большим значением, то происходит расхождение автокорреляции вейвлет-коэффициентов [31, 33]. То есть автокорреляционная функция гиперболически затухает. Если используется вейвлет-анализ с большим исчезающим моментом R , автокорреляционная функция вейвлет-коэффициентов может экспоненциально быстро затухать до нуля. На рисунке 33 показано сравнение автокорреляционных функций для вейвлет-коэффициентов при использовании различных вейвлет функций.

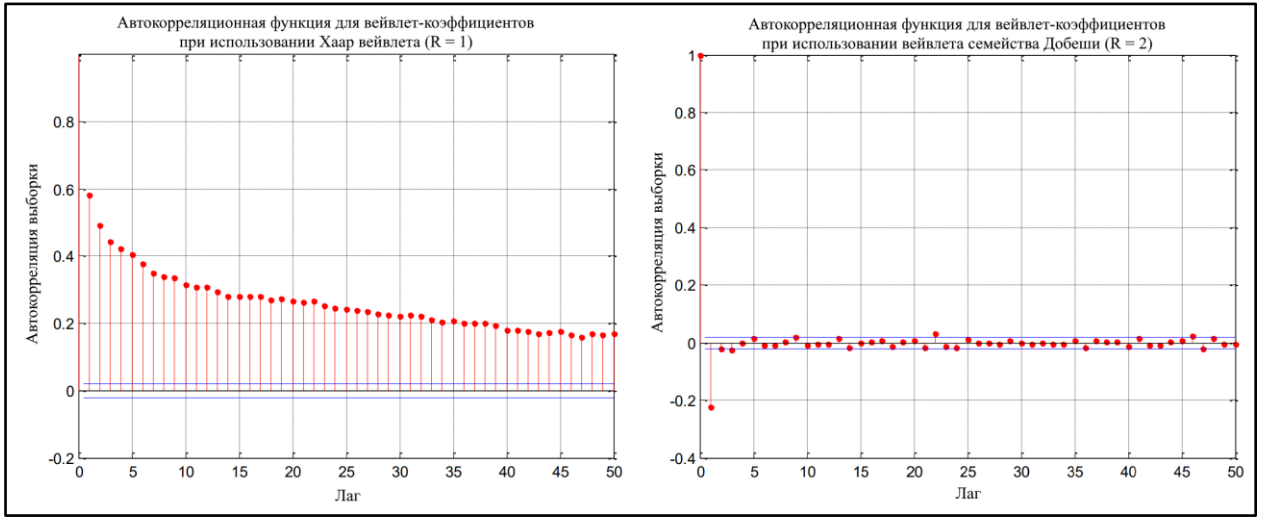


Рисунок 33 – Сравнение автокорреляционных функций для вейвлет-коэффициентов при использовании различных вейвлетов (R равное 1 слева и R равное 2 справа)

Рисунки 32 и 33 показывают, что внутренний масштаб корреляции сильно ослаблен, и нестационарный сигнал можно анализировать стационарно после вейвлет-преобразований. Однако, корреляция между масштабами всё ещё сохраняется [31, 32, 33], что можно описать уравнением:

$$d_{j,k} = 2^{j(H+\frac{1}{2})} d_{0,k}. \quad (54)$$

Доказательство уравнения 54 можно вывести из формул 49, 46 и 30:

$$\begin{aligned}
 d_{j,k} &= \int_{\mathbb{R}} X(t) \psi_{j,k}(t) dt = \int_{\mathbb{R}} X(t) * \frac{1}{2^j} \psi(2^{-j}t - k) dt = \\
 &= \int_{\mathbb{R}} 2^{\frac{j}{2}} * X(t) * 2^{-j} \psi(2^{-j}t - k) dt = \int_{\mathbb{R}} 2^{\frac{j}{2}} * X(2^j u) \psi(u - k) du \\
 &= \\
 &= \int_{\mathbb{R}} 2^{\frac{j}{2}} * (2^j)^H X(u) \psi(u - k) du = \int_{\mathbb{R}} 2^{j(H+\frac{1}{2})} X(u) \psi(u - k) du = \\
 &= 2^{j(H+\frac{1}{2})} * d_{0,k}.
 \end{aligned} \quad (55)$$

Уравнение 55 можно объяснить следующим образом: вейвлет-коэффициенты на разных масштабах демонстрируют одно и то же статическое свойство после надлежащего масштабирования. Сильную корреляцию между масштабами также можно увидеть на графиках вейвлет-коэффициентов на ранее показанном рисунке 30 (масштабы 1-3).

Модель на основе вейвлет-преобразований основана на том, что вейвлет-разложение для самоподобных сигналов может понижать корреляцию внутри шкалы, но при этом сохранять корреляцию между масштабами. Теоретически, чем больше исчезающие моменты R у вейвлет-функции, тем слабее будет внутри масштабная корреляция. Однако вейвлет Хаара выбран из-за того, что он имеет чёткое отсечение и использует только текущие и прошлые наблюдения, в то время как вейвлеты Добеши с исчезающими моментами больше единицы имеют длинные остаточные «хвосты» (их можно увидеть на рисунке 31) и используют будущие наблюдения, что неприемлемо в данной модели. Для самоподобного сигнала алгоритм вейвлет-анализа может быть описан следующим образом: сначала сигнал раскладывается при помощи вейвлета Хаара, а затем применяется модель векторной авторегрессии, чтобы зафиксировать корреляцию между масштабами [33]. Прогнозируемые вейвлет-коэффициенты затем преобразуются обратно, чтобы получить желаемый прогноз для исходного временного ряда.

При применении модели векторной авторегрессии к вейвлет-коэффициентам было обнаружено, что количество вейвлет-коэффициентов не одинаково на разных уровнях. Однако модель векторной авторегрессии требует одинакового количества вейвлет-коэффициентов на всех уровнях. Д. Ванг и С. Д. Лии [32] использовали в своей работе метод понижающей дискретизации для решения этой проблемы (рисунок 34).

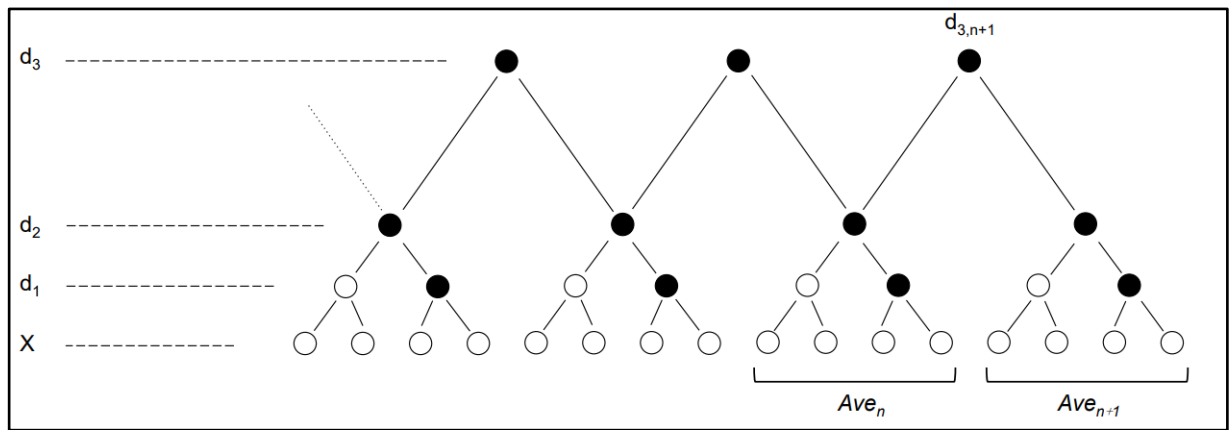


Рисунок 34 – Подход понижающей дискретизации для алгоритма масштабного прогнозирования

Сначала временной ряд группируется в блоки по четыре числа, а затем создаются векторы вейвлет-коэффициентов на основе трёх уровневого вейвлет-разложения. Для получения прогноза среднего значения будущих четырех периодов времени используется вейвлет-коэффициенты на третьем уровне. Для вейвлета Хаара вейвлет-коэффициенты на масштабе $j + 1$ удовлетворяют уравнению:

$$d_{j+1,n} = 2^{\frac{j-1}{2}} \left(X_{2n-1}^{(2^j)} - X_{2n}^{(2^j)} \right), \quad (56)$$

где $X_{2n}^{(2^j)}$ – это совокупный процесс, определяемый уравнением 2.23.

После получения вейвлет-коэффициентов на третьем масштабе разложения, прогноз среднего значения для четырех будущих периодов времени может быть вычислен при помощи формулы:

$$Ave_{n+1} = Ave_n - \frac{d_{3,n+1}}{2^{\frac{3}{2}}}. \quad (57)$$

Однако если исследуемые временные ряды обладают долговременной памятью, необходимо сначала аккумулировать значения, прежде чем

использовать вейвлет модель. Каждая точка накопленного ряда важна и должна быть спрогнозирована для получения дифференциала на последнем шаге, поэтому знать только средние значения точек не достаточно. Для этого используется избыточное вейвлет преобразование (рисунок 35).

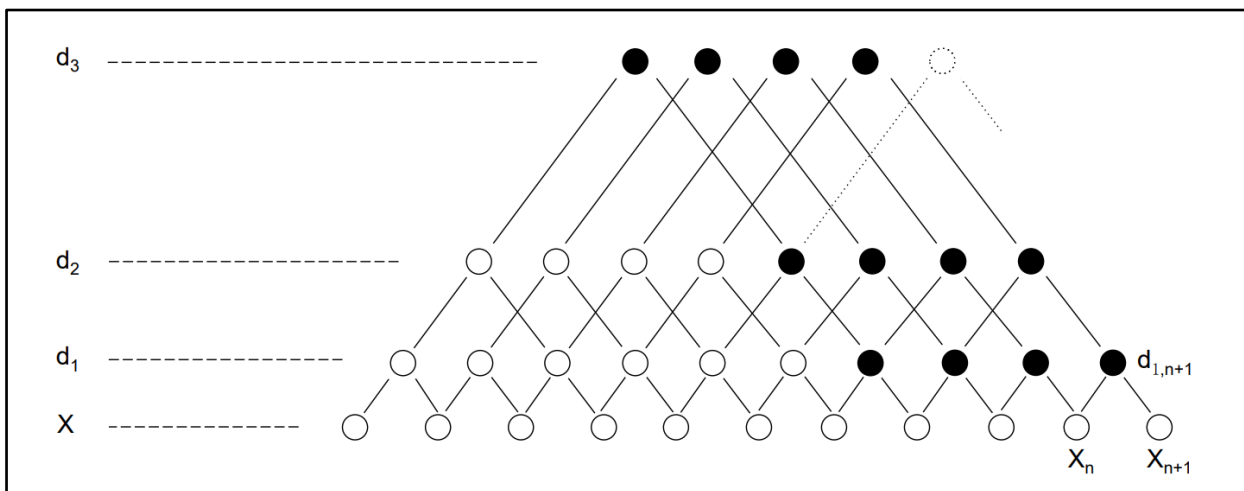


Рисунок 35 – Пример избыточного вейвлет преобразования

Создаются векторы вейвлет-коэффициентов, а затем используется векторная авторегрессионная модель для прогнозирования следующего вектора. Вместо использования коэффициентов верхнего уровня, используется коэффициент нижнего уровня $d_{1,n+1}$, который позволяет составить прогноз на следующий период времени.

$$X_{n+1} = X_n - \frac{d_{1,n+1}}{2^{-1/2}} \quad (58)$$

И таким образом прогноз для исходного временного ряда может быть получен путём дифференцирования прогнозируемого накопленного ряда (формула 59).

$$Y_n = X_{n+1} - X_n \quad (59)$$

3.5 Модель Average-VAR

Согласно уравнению 2.24 процесс приращения Y_n является стационарным и обладает свойством долговременной памяти. На основе этого свойства масштабирования, Р. Ванг в своей работе [34] предложил модель средней векторной авторегрессии. Агрегированный процесс получается путём усреднения временного ряда. Между процессами, агрегированными в разных масштабах, должна быть сильная корреляция, поэтому для описания этой масштабной корреляции применяется модель векторной авторегрессии (VAR). На рисунке 36 представлена трёх уровневая модель среднего VAR.

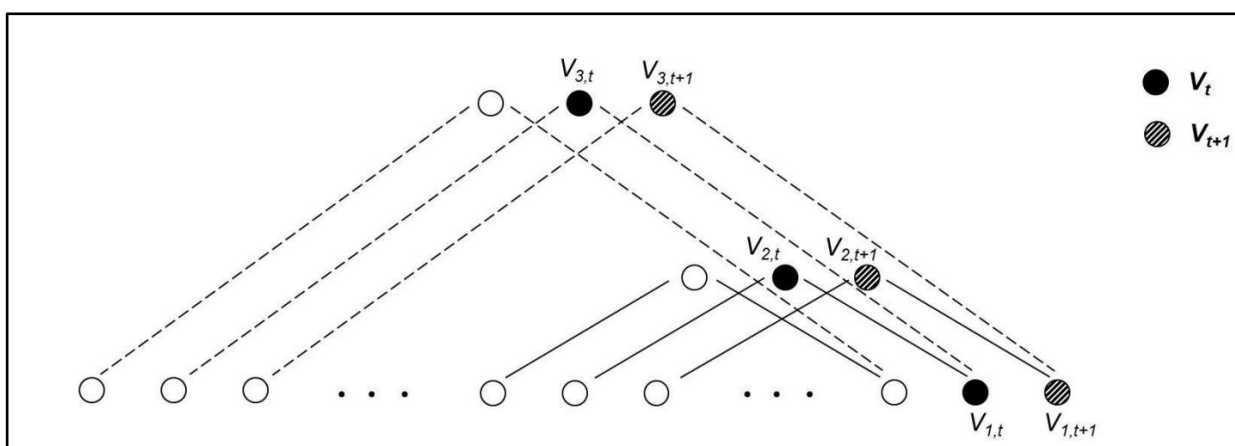


Рисунок 36 – Модель трёх уровневого разложения Average-VAR

Данная модель среднего-VAR довольно проста теории и в использовании. Для стационарного временного ряда с долговременной памятью общие шаги в реализации этой модели можно описать следующим образом: сначала усреднить временной ряд на трёх различных временных масштабах, чтобы учесть как краткосрочный, так и долгосрочный тренд. Затем применяется модель VAR, описываемая в формуле 60. Значение $V_{1,t+1}$ и будет искомым прогнозом.

$$\begin{aligned}
V_{t+1} &= c + AV_t + e_t \Rightarrow \\
\Rightarrow \begin{bmatrix} V_{1,t+1} \\ V_{2,t+1} \\ V_{3,t+1} \end{bmatrix} &= \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} V_{1,t} \\ V_{2,t} \\ V_{3,t} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}. \tag{60}
\end{aligned}$$

Вывод к главе 3

В данной главе были рассмотрены четыре модели для моделирования фрактальных временных рядов: FARIMA, HAR-RV, вейвлет-модель и Average-VAR. Все они могут быть применены для стационарных временных рядов с долговременной памятью, но только вейвлет-модель может быть использована для самоподобных нестационарных временных рядов. Если есть необходимость в применении остальных моделей к самоподобному временному ряду, то необходимо его предварительно преобразовать в стационарный процесс.

Глава 4 Исследования и компьютерная реализация

В этой главе будут представлены численные исследования нескольких реальных сигналов, затем будут применены четыре модели, описанные в предыдущей главе. Но перед этим все данные будут приведены к стационарному виду.

Так как все описанные модели могут спрогнозировать будущие значения временного ряда, в исследовании более важны точность и скорость каждой модели, которые можно сравнить.

4.1 Описание данных

Многие сигналы и процессы в реальной жизни проявляют свойства самоподобия или долговременной памяти. К таким примерам можно отнести интернет трафик, изменения в финансовых значениях, физиологические сигналы и многое другое. Однако такие данные действительно сложно собрать из-за отсутствия их в открытом доступе, либо они описаны в достаточно малом промежутке времени. В данной диссертации используются только те временные ряды, которые можно найти исключительно в открытом доступе, а их количество наблюдений больше четырёх тысяч.

4.1.1 Индекс волатильности

Как упоминалось в главе 3.3.1, существует два вида волатильности: реализованная волатильность и подразумеваемая волатильность. В качестве тестовых данных взяты индексы подразумеваемой волатильности.

Подразумеваемая волатильность часто связана с ценообразованием опционов. Она может быть вычислена как волатильность, при которой цена, полученная с помощью модели ценообразования опционов, равна ее текущей рыночной стоимости [26]. Поскольку вычисление реализованной волатильности требует большого количества ежедневных высокочастотных торговых данных, которые достаточно сложно получить, было принято

решение в данном исследовании использовать подразумеваемую волатильность.

Данные взяты из портала «СВОЕ» (Чикагская биржа опционов). В качестве данных взяты показатели «VIX», являющиеся тиккерами индекса биржи «S&P 500», в которую входят пятьсот пять избранных торгуемых на фондовых биржах США публичных компаний, имеющих наибольшую капитализацию. Сигналы «GVZ», являющиеся тиккерами индекса волатильности золота биржи «СВОЕ». А так же показатели «VXO», являющимися тиккерами индекса биржи «S&P 100», которая так же является фондовой биржей США. Примеры этих данных можно увидеть на рисунке 37.

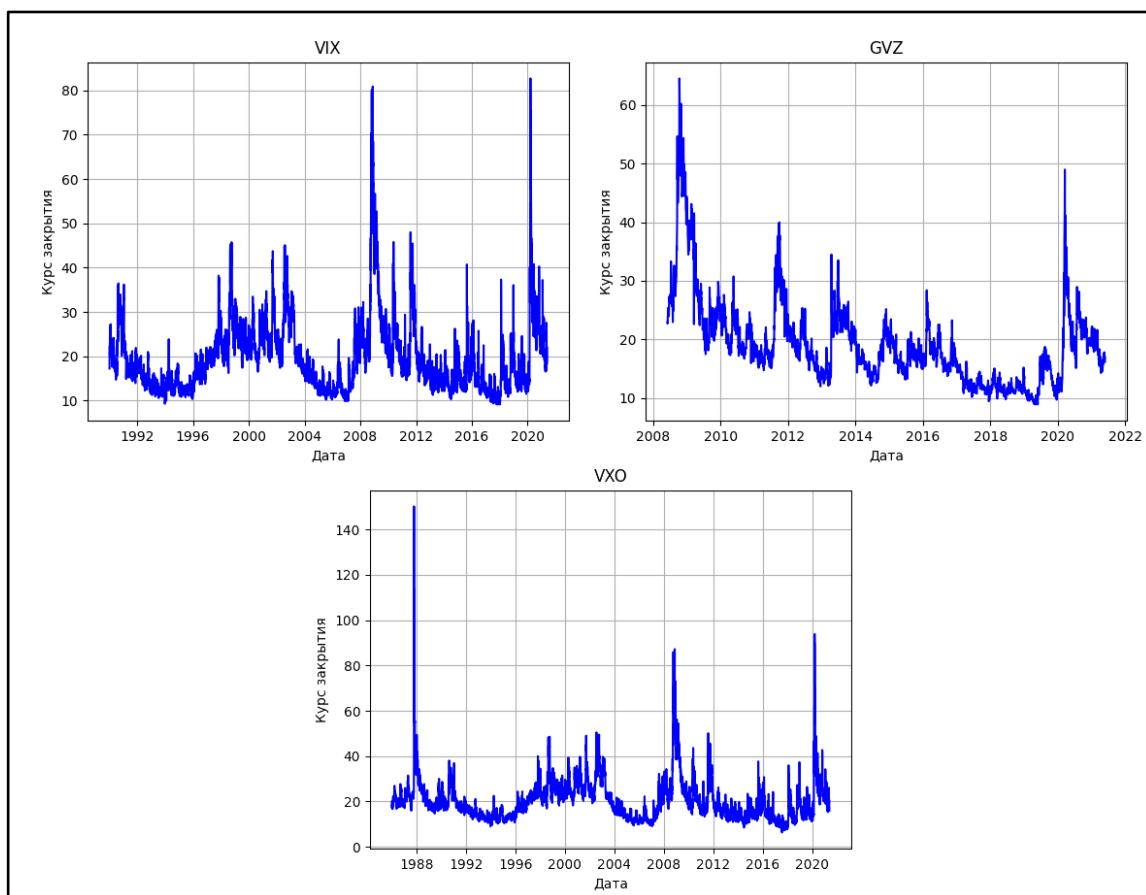


Рисунок 37 – Визуализация индексов «VIX», «GVZ» и «VXO»

Периоды выборки различны для каждого временного ряда. В таблице 1 показаны рассматриваемые промежутки для выбранных индексов.

Таблица 1 – Периоды используемых индексов волатильности

Индекс	Период
VIX	02.01.1990 – 27.05.2021
GVZ	03.06.2008 – 27.05.2021
VXO	01.02.1986 – 27.05.2021

4.1.2 Данные артериального давления крови

На открытом онлайн форуме «PhysioNet» [35] можно бесплатно получить совершенно разные биомедицинские сигналы, такие как: сигналы сердечно сосудистой системы, сигналы ЭКГ, физиологические сигналы людей с различными заболеваниями и другие.

В качестве тестовых данных были взяты три различных сигнала артериального давления крови, которые были помечены как «ABP1», «ABP2» и «ABP3» соответственно. Визуальное представление этих данных можно увидеть на рисунке 38.

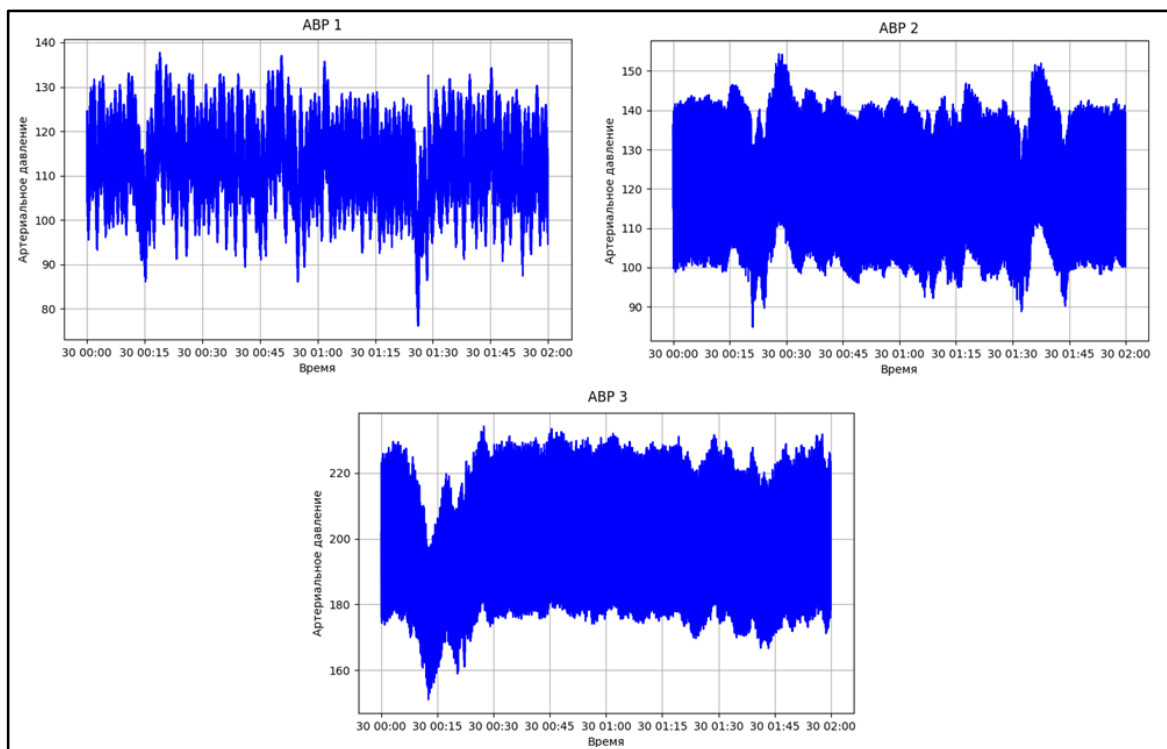


Рисунок 38 – Визуализация сигналов «ABP1», «ABP2» и «ABP3»

Измерения артериального давления на портале «PhysioNet» фиксируются каждые несколько секунд, а данные представлены в диапазоне двух часов, что составляет двенадцать тысяч наблюдений для каждого сигнала [27].

4.2 Компьютерная реализация

Для компьютерной реализации был выбран язык программирования Python версии 3.9.5. Данный язык программирования позволяет достаточно просто визуализировать все исследования, проводимые в данной диссертации. Помимо этого, данный язык предоставляет некоторые готовые решения в виде библиотек, как например: «matplotlib», «pandas», «numpy», «statsmodel», «scipy» и другие.

Для реализации модели FARIMA была взята готовая модель из пакета «statsmodel». Для реализаций моделей Average-VAR и вейвлет-модели была использована та же библиотека «statsmodel» для выполнения векторной авторегрессии. А модель HAR-RV была реализована с нуля.

Помимо этого были реализованы остальные алгоритмы, как например, расчёт показателя Херста, расчёт ошибок и другое.

4.3 Подготовка данных

Прежде чем моделировать фрактальные временные ряды, сначала необходимо подготовить данные.

В первую очередь необходимо заполнить все пропуски данных значениями. Так, например, во всех выбранных индексах волатильности существуют пропуски данных (рисунок 39).

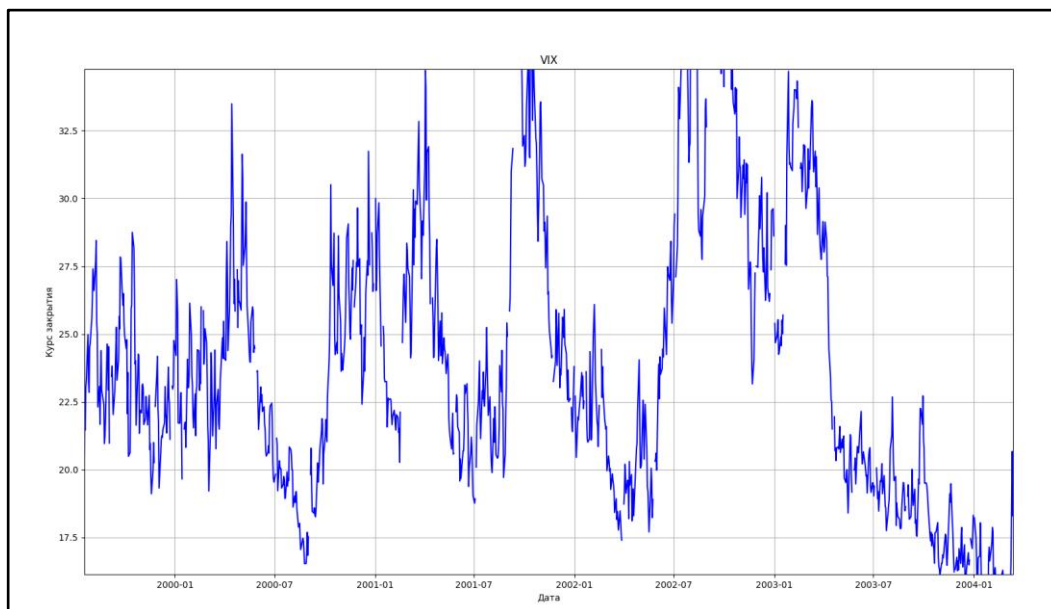


Рисунок 39 – Индекс волатильности «VIX» до заполнения пропущенных значений

Для решения данной проблемы будет использован метод среднего значения по N рядом лежащих значений, описанного во главе 1.2. После такого преобразования временной ряд выглядит уже цельно (рисунок 40).

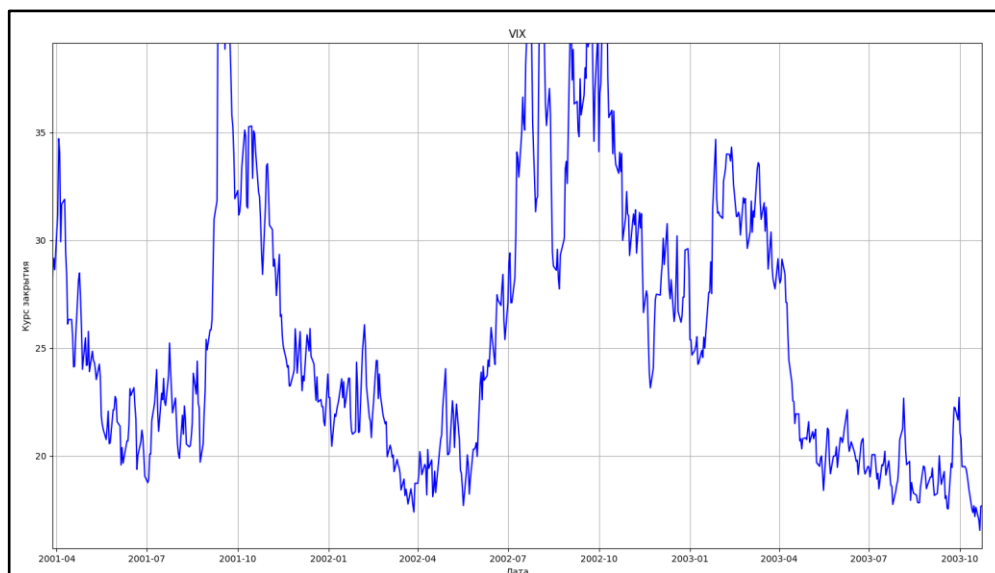


Рисунок 40 – Индекс волатильности «VIX» после заполнения пропущенных значений

Теперь можно оценить временные ряды при помощи показателя Херста. Для расчёта этого показателя будет использоваться R/S анализ, описанный в главе 2.6.3, так как данный метод является одним из самых точных и проверенных методов.

Так же важно понимать, что R/S анализ может определять параметр Херста только для стационарного временного ряда. Если результат, полученный в результате расчёта показателя Херста, находится в диапазоне $H \in [0, 1]$, то этот параметр можно считать показателем временного ряда. Иначе, если параметр Херста H больше единицы, значит, что анализируемый временной ряд является нестационарным и его необходимо провести дифференцирование перед повторным использованием R/S анализа и получения истинного показателя для временного ряда. Весь алгоритм анализа вынесен в блок-схему на картинке 41.

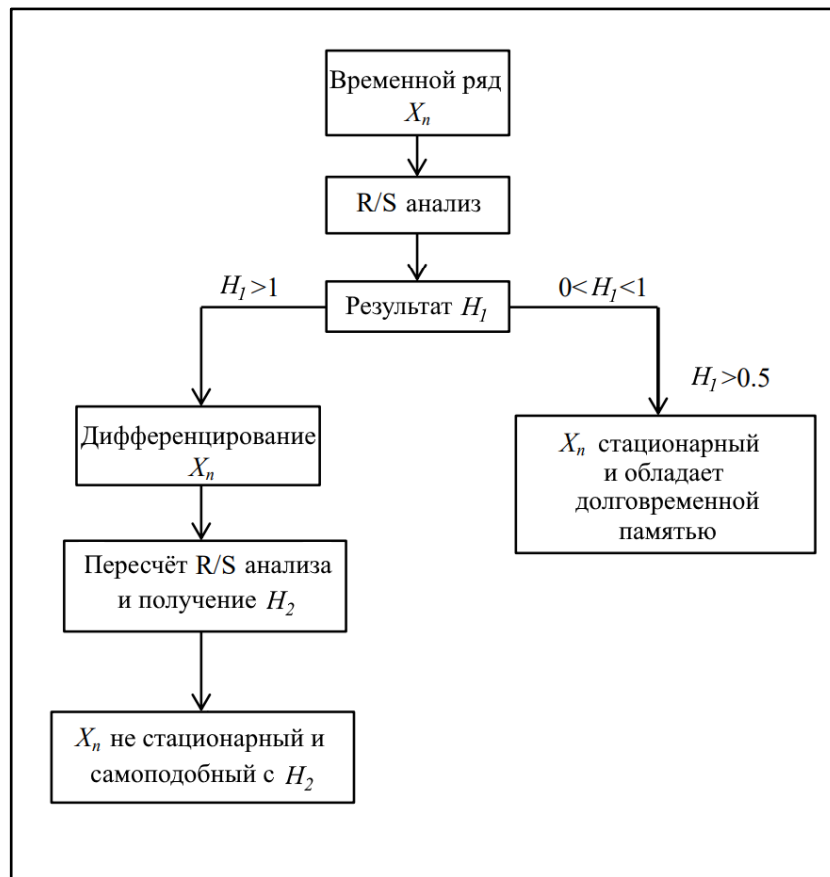


Рисунок 41 – Блок схема анализа исходных временных рядов

Примеры графиков расчёта показателя Херста представлены на рисунке 42. Слева расчёт параметра Херста для индекса волатильности «VIX» и сигнала артериального давления «ABP1».

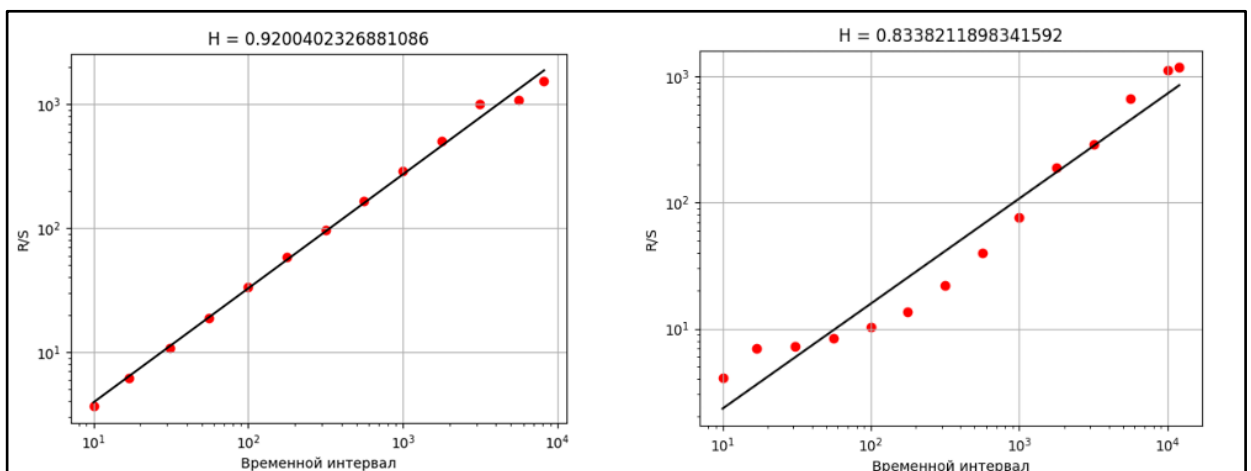


Рисунок 42 – Расчёт параметра Херста для волатильности «VIX» (слева) и сигнала артериального давления «ABP1» (справа)

В таблице 2 приведены показатели Херста для всех временных рядов, рассчитанные при помощи R/S анализа.

Таблица 2 – Расчёт показателя Херста для тестовых временных рядов методом R/S анализа

Название сигнала	R/S показатель
VIX	0.92004
GVZ	0.951147
VXO	0.916559
ABP1	0.833821
ABP2	0.73211
ABP3	0.7724951

Автокорреляционные функции для индексов волатильности «VXO» и сигналов артериального давления «ABP2» представлены на рисунках 43. По графикам автокорреляционных функций видно, что значения обоих сигналов коррелируют на достаточно длительном промежутке. Это может свидетельствовать о наличии стационарности и долговременной памяти.

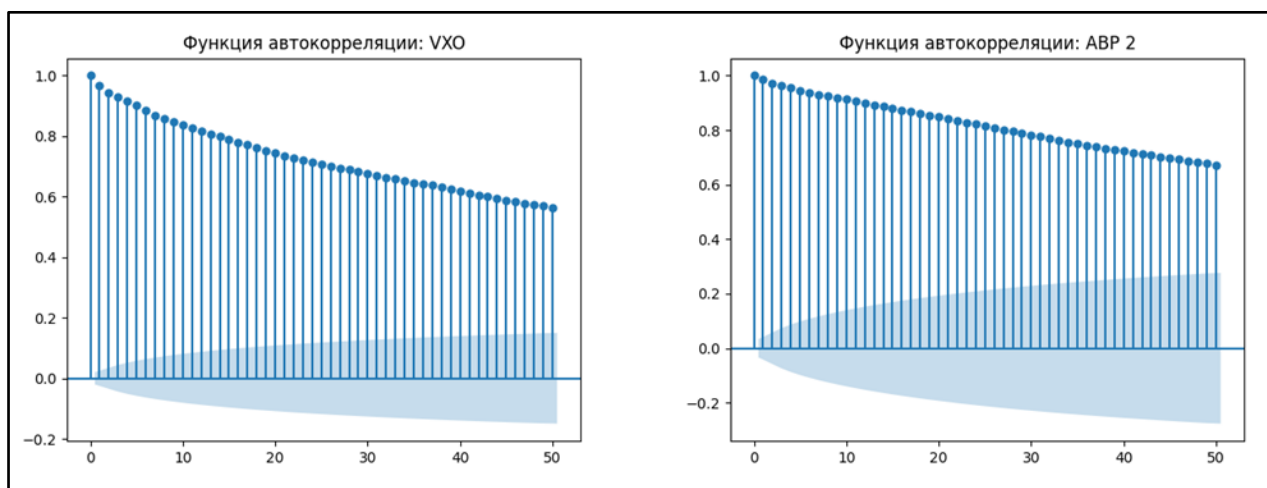


Рисунок 43 – Автокорреляционные функции для индексов волатильности «VXO» (слева), сигналов артериального давления «ABP2» (справа)

Для подтверждения гипотезы стационарности исследуемых временных рядов, был использован критерий Дики-Фуллера, который был описан во главе 1.4.2. В данном критерии необходимо оценить параметр p : если он

будет меньше или равен 0.05, значит можно утверждать, что временной ряд является стационарным. В таблице 3 приведены параметры p для исследуемых временных рядов.

Таблица 3 – Оценка стационарности критерием Дики-Фуллера

Название сигнала	Параметр p
VIX	0.0011
GVZ	0.008606
VXO	0.0011
ABP1	0.010019
ABP2	0.000394
ABP3	0.0001

Из таблицы 3 можно сделать вывод, что все исследуемые временные ряды являются стационарными. А так как показатель Херста для данных временных рядов находится в диапазоне (0.5, 1) (таблица 2), то все они обладают долговременной памятью. Это позволяет использовать описанные ранее методы моделирования без предварительной модификации данных.

4.4 Выбор оптимальных агрегированных уровней для модели HAR-RV и модели Average-VAR

В модели HAR-RV Ф. Корси агрегировал реализованную волатильность на основе недельных и ежемесячных приростов волатильности, а затем спрогнозировал будущую суточную реализованную волатильность как взвешенную сумму прошлых дневных, недельных и месячных волатильностей [26]. В качестве агрегированных уровней он выбрал значения 1, 5 и 22, что разумно, так как они соответствуют временным промежуткам обновления – день, неделя и месяц.

Для индексов волатильности использование таких уровней агрегации оправдано, однако для сигналов артериального давления нет чётких практических значений. Использование тех же агрегированных уровней, как и для индексов волатильности логически не верно. Поэтому необходимо

численно отобрать оптимальные агрегированные уровни [28]. Для учёта краткосрочных и долгосрочных эффектов используются три критерия:

- Критерий 1: наиболее детализированный масштаб (краткосрочный процесс), $s_1 = 1$
- Критерий 2: средний масштаб, $s_2 \in [4, 12]$
- Критерий 3: самый большой масштаб, $s_3 \in [12, 25]$

Для определённого сигнала артериального давления применяется модель HAR-RV с различными комбинациями агрегированных уровней. При этом уровни выбираются так, чтобы их значения удовлетворяли условию:

$$s_3 - s_2 \geq 4. \quad (61)$$

Таким образом, обеспечивается гарантия того, что обе шкалы s_2 и s_3 будут существенно различаться.

Среднеквадратичные ошибки (MSE) для различных пар комбинаций уровней масштабирования сравниваются так, чтобы отыскать минимальное значение ошибки. Для сигналов артериального давления «ABP1», «ABP2» и «ABP3» среднеквадратичные ошибки комбинаций агрегированных уровней равны (5, 24), (8, 23) и (5, 25) соответственно. Можно заметить, что оптимальные агрегированные уровни не всегда одинаковы для разных сигналов. Но выполнение подбора оптимальных уровней для каждого временного ряда будет занимать достаточно длительное время, из-за чего время выполнения прогнозирования одной моделью для разных типов временных рядов будет слишком различаться. А так как среднеквадратичная ошибка не слишком различается на разных уровнях агрегирования, то будут использоваться одинаковые уровни для всех сигналов артериального давления: 1, 5 и 25.

Как и модель HAR-RV, модель Average-VAR так же требует агрегирования на первом этапе. Поэтому для более разумного сравнения этих

двух моделей, будут использоваться одинаковые уровни агрегирования сигналов.

4.5 Критерии оценки эффективности

4.5.1 Показатели ошибки

Предположим что $\{\hat{Y}_i\}$ – прогнозируемая последовательность, $\{Y_i\}$ – существующие значения временного ряда, а n – длина последовательности. Для сравнения эффективности прогноза используются следующие показатели ошибок:

- 1) Среднеквадратичная ошибка: $MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$;
- 2) Средняя процентная абсолютная ошибка: $MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| * 100\%$;
- 3) Средняя абсолютная ошибка: $MAE = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i|$.

Описанные выше ошибки обычно используются при оценке эффективности прогнозирования.

Кроме того, можно определить ещё одну меру, позволяющую учитывать точность предсказания направления будущих значений. В реальных предсказательных моделях точное значение следующего предсказания менее важный фактор, чем направление этого предсказания. Процент предсказаний правильного направления (PRD) рассчитывается как количество предсказаний правильных направлений по отношению ко всем значениям.

$$PRD = \frac{\text{количество правильных предсказаний направления}}{\text{количество всех предсказаний направлений}}. \quad (62)$$

4.5.2 ROC-кривая

Рабочая характеристика ROC (с английского «receiver operating characteristic») используется в качестве инструмента определения наилучшего порога, а так же оценки производительности двоичного классификатора [36]. Этот график зависимости истинно положительного результата (TPR) от частоты ложных положительных результатов (FPR) при изменении порогового значения T [25].

Чтобы проиллюстрировать принцип работы ROC-кривой, далее будет рассмотрен пример диагностики гипотетического заболевания.

Есть две группы испытуемых: в первой находятся здоровые люди (отсутствие болезни – это негативный результат), а во второй заболевшие (наличие заболевания – положительный результат). Задача состоит в том, чтобы разделить эти группы между собой и отправлять сигнал тревоги в случае возникновения новых заболеваний. Для этого измеряется физиологический сигнал, как например артериальное давление, и устанавливается порог T . Люди, имеющие кровяное давление выше порога T рассматриваются как заболевшие, и при обнаружении таких людей отправляется сигнал тревоги. А люди с артериальным давлением ниже порогового значения T считаются здоровыми.

Если человек действительно болеет и тест вызывает тревогу, это называется истинно положительным сигналом (TP) [36]. Напротив, если человек здоров, но тест ложно подаёт сигнал тревоги, это называется ложноположительным сигналом (FP) [36]. Тогда количество истинно положительных результатов (TPR) может быть получено путём деления количества истинных положительных результатов на общее количество фактических положительных результатов. А количество ложноположительных результатов (FPR) можно рассчитать через деление количества ложных срабатываний на общее количество фактических отрицательных результатов [36]. Каждому порогу T соответствует пара FPR и TPR. Изменяя порог T , можно получить набор пар FPR и TPR. Таким

образом, ROC-кривая может быть получена путем построения графика зависимости TPR от FPR [36]. Таблица 4 и рисунки 44 - 45 демонстрируют описанный выше процесс подсчёта.

Таблица 4 – Таблица расчёта TPR и FPR

		Фактическое состояние	
		Положительное (P)	Отрицательное (N)
Результат теста	Тест положительный	Истинно положительный (TP)	Ложноположительный (FP)
	Тест отрицательный	Ложноотрицательный	Истинно отрицательный
Формула расчёта		$TPR = \frac{TP}{P}$	$FPR = \frac{FP}{N}$

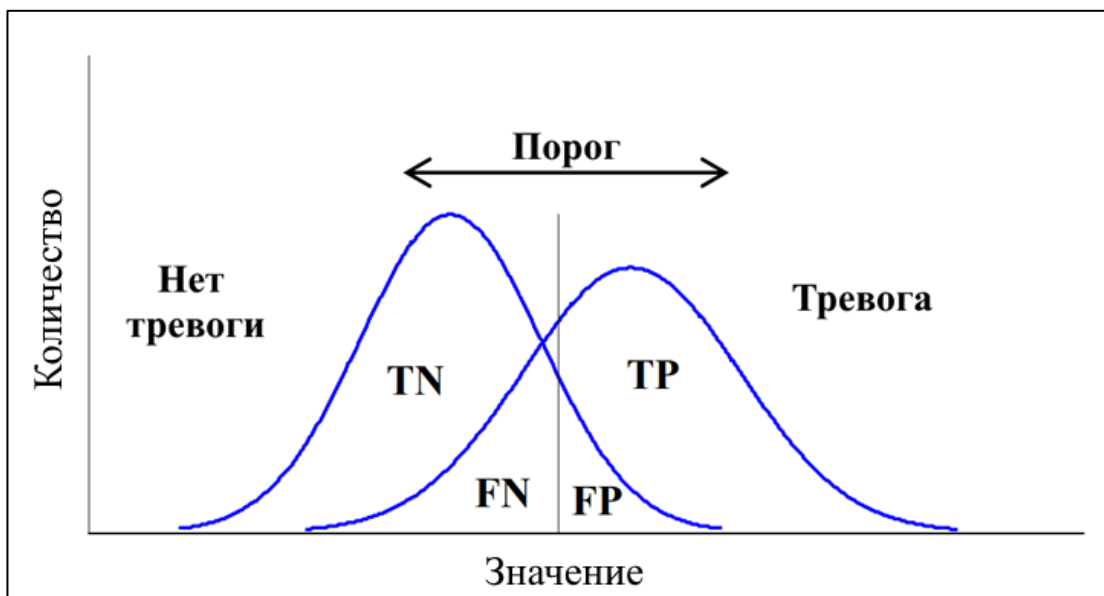


Рисунок 44 – Роль порогового значения в получении ROC-кривой

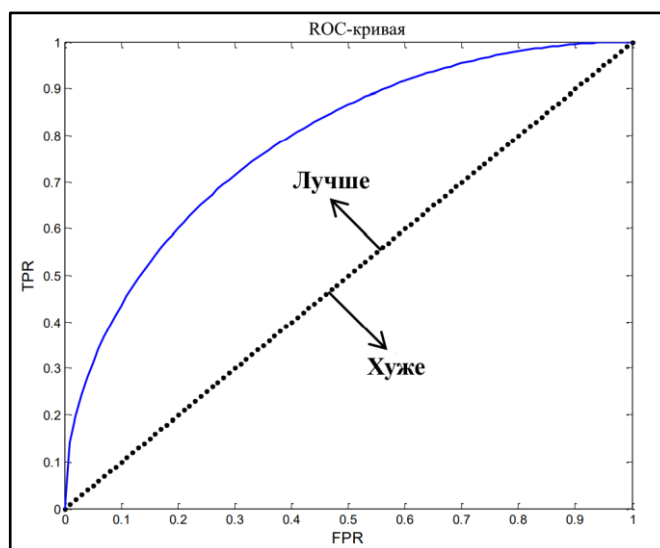


Рисунок 45 – Пример ROC-кривой

На рисунке 45 представлен пример ROC-кривой. Чем ближе точка находится к левому верхнему углу графика, тем лучше. Случайное предположение будет находиться на прямой $x = y$ [36]. Площадь под кривой линией (AUC), представляющая из себя значения в диапазоне $[0; 1]$, часто используется в качестве оценки эффективности модели. Чем больше площадь, тем выше эффективность. Случайное предположение, лежащее на диагонали, даст AUC равное 0.5. Если модель имеет AUC менее 0.5, то необходимо инвертировать принятие решений [36]. Например, если модель имеет AUC менее 0.5, и при этом модель классифицирует состояние как положительное, то необходимо рассматривать решение как отрицательное и не отправлять сигнал тревоги.

4.6 Результаты исследований

Как уже говорилось ранее, всего было использовано четыре модели: FARIMA, HAR-RV, вейвлет-преобразования и Average-VAR. Каждый временной ряд содержит минимум четыре тысячи наблюдений, поэтому можно взять первую тысячу точек и использовать их в качестве исторических данных для создания модели, которая может быть использована для прогнозирования на один шаг вперёд. Когда становится доступна новая точка данных, она добавляется к историческим данным и

используется как обновлённые данные для будущей модели. И прогнозируемые результаты, и фактические значения сохраняются и сравниваются для подтверждения работоспособности модели.

В данной главе будут показаны численные сравнения критериев эффективности MSE, MAPE, MAE, PRD, а так же сравнение ROC-кривых, которые уже обсуждались в предыдущей главе.

4.6.1 Сравнение графиков

Рисунки 46 – 49 представляют собой результаты прогноза сигнала индекса волатильности «VIX» с использованием всех четырёх моделей. На каждом рисунке красная кривая представляет результат прогноза, а синяя – исходный временной ряд. Так же на правой части рисунков представлен одинаковый временной промежуток, чтобы сравнение было более наглядным.

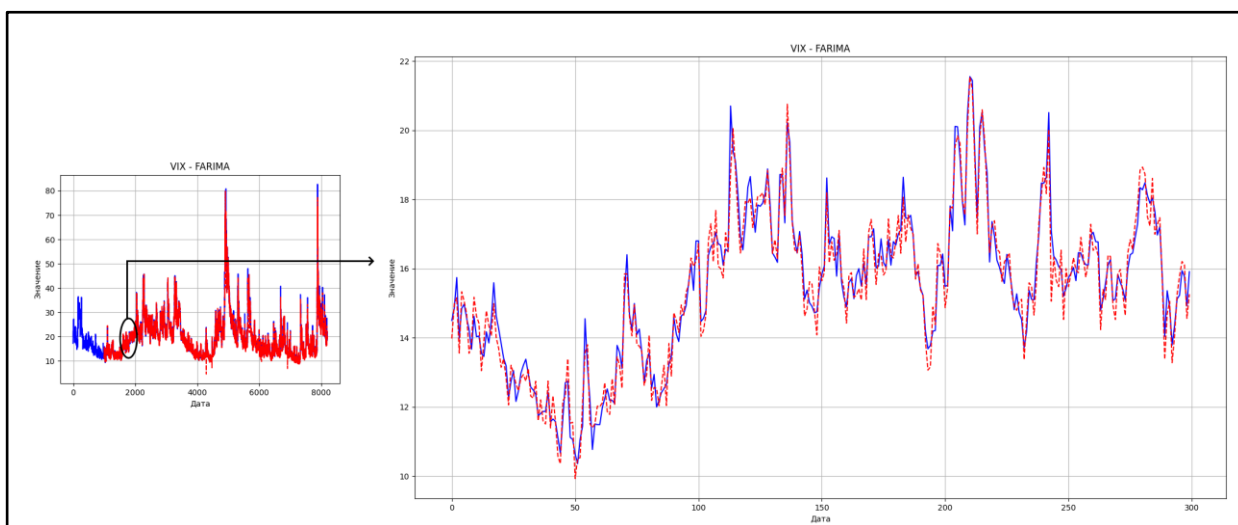


Рисунок 46 – Результаты модели FARIMA для сигнала «VIX»

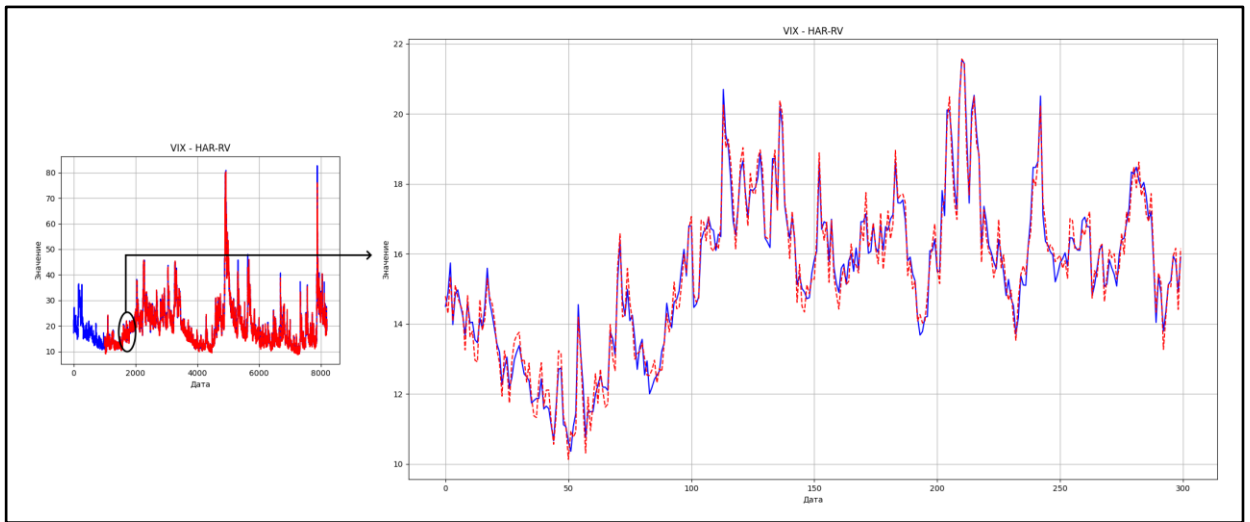


Рисунок 47 – Результаты модели HAR-RV для сигнала «VIX»

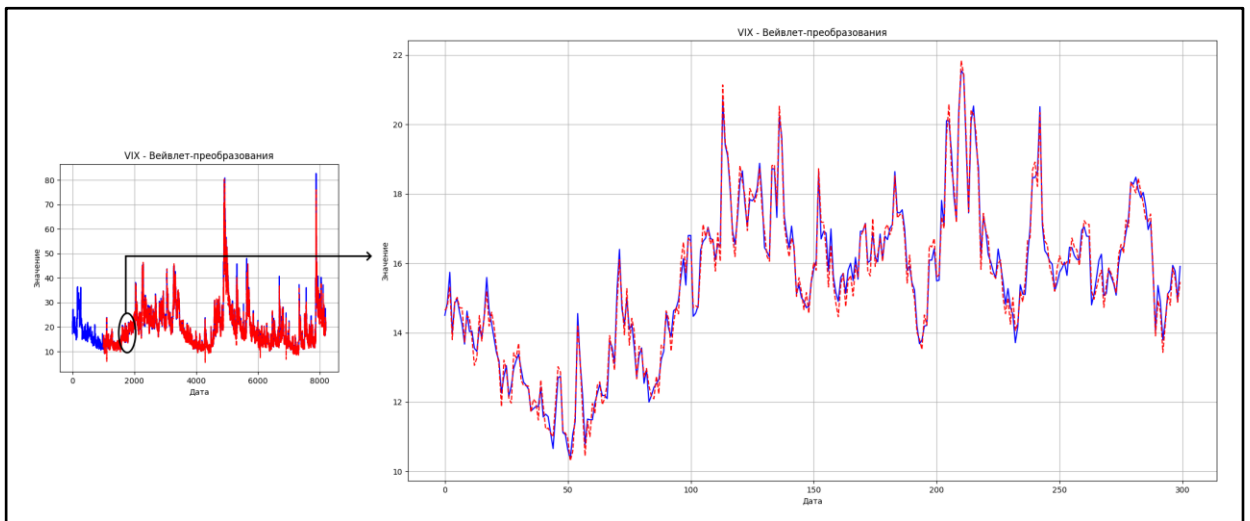


Рисунок 48 – Результаты модели на основе вейвлет-преобразований для сигнала «VIX»

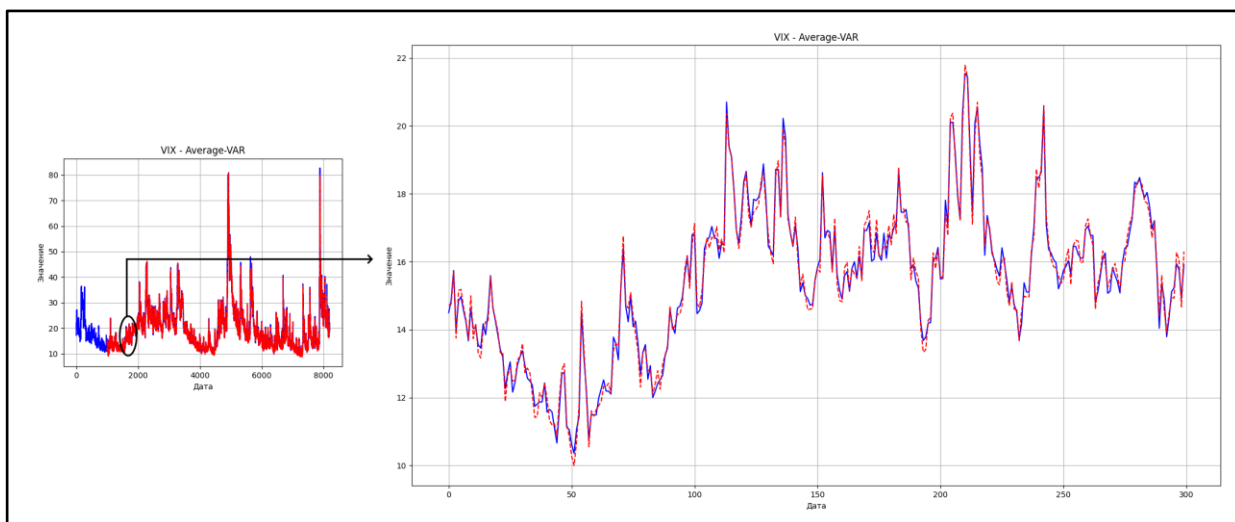


Рисунок 49 – Результаты модели Average-VAR для сигнала «VIX»

Как видно на рисунках, существует некоторое смещение между реальными и прогнозируемыми величинами. Прогнозируемое значение всегда представляет собой сдвиг предыдущего значения сложенного с небольшой погрешностью. Если текущее наблюдение показывает увеличение по сравнению с предыдущим значением, то прогноз на следующий период так же имеет тенденцию к увеличению, и наоборот. Такое поведение нормально, поскольку последнее наблюдение и тренд должны иметь самое сильное влияние на будущие наблюдения, поэтому прогноз будущего значения должен быть близок к самому последнему наблюдению.

4.6.2 Численные сравнения

В качестве меры сложности модели используется общее время вычислений в секундах, которое потребовалось для выполнения всего процесса моделирования и прогнозирования временного ряда. Показатели ошибок (MSE, MAPE, MAE и PRD) и время выполнения реализации для сигналов «VIX», «VXO» и «ABP1» представлены в таблицах 5 – 7 соответственно. Численные результаты для других сигналов представлены в приложении А.

Таблица 5 – Показатели ошибок и время выполнения для сигнала «VIX»

	MSE	MAPE (%)	MAE	PRD	Время (сек)
FARIMA	2.1543	4.28	0.9140	0.4846	7036.155
HAR-RV	2.1044	4.28	0.9107	0.4828	131.097
Вейвлет	2.1235	4.27	0.9096	0.4835	412.187
Average-VAR	2.1041	4.28	0.9105	0.4828	226.974

Таблица 6 – Показатели ошибок и время выполнения для сигнала «VXO»

	MSE	MAPE (%)	MAE	PRD	Время (сек)
FARIMA	25.5494	2.89	2.0749	0.4876	18224.264
HAR-RV	24.5877	2.86	2.0561	0.4901	165.405
Вейвлет	27.9540	2.86	2.0523	0.4799	488.376
Average-VAR	24.6106	2.86	2.0570	0.4899	292.168

Таблица 7 – Показатели ошибок и время выполнения для сигнала «ABP1»

	MSE	MAPE (%)	MAE	PRD	Время (сек)
FARIMA	61.4231	5.24	3.9933	0.4688	5143.031
HAR-RV	59.5287	5.21	3.9776	0.4569	160.821
Вейвлет	60.7234	5.09	3.9156	0.4552	299.964
Average-VAR	56.0381	5.19	3.9429	0.4569	166.085

Из таблиц 5 – 7 можно понять, что характеристики четырех моделей с точки зрения ошибок MSE, MAPE, MAE и PRD очень схожи. Но FARIMA, безусловно, требует наибольшего времени для моделирования и прогнозирования временного ряда, тогда как модель HAR-RV и модель average-VAR наоборот. Поэтому модель FARIMA крайне не рекомендуется для приложений в реальном времени.

В то же время, модель FARIMA чаще остальных моделей показывает самое высокое значение характеристики PRD. Это означает, что данная модель лучше остальных предсказывает направление следующего значения.

Показатели MAPE и PRD используются при сравнении точности прогнозов для разных сигналов, поскольку другие показатели ошибок основаны на абсолютных значениях. Они помогут выяснить, какие свойства сигнала могут привести к более точному прогнозу. Для сигнала с долговременной памятью параметр Херста и сильная корреляция с

прошлыми значениями являются двумя наиболее важными характеристиками.

Из приведенных выше показателей, а также соответствующих параметров Херста легко обнаружить, что большее значение параметра Херста не обязательно приводит к более точному прогнозу. Так, например, результаты прогноза для сигнала «ABP1» являются лучшими среди всех трех сигналов с точки зрения MAPE, в то время как параметр Херста для данного временного ряда является не самым большим, а значит, показатель Херста не влияет на качество прогноза.

4.6.2 Сравнение ROC-кривых

Для того чтобы использовать ROC-кривую в качестве инструмента оценки в данном исследовании, сначала необходимо определить положительные и отрицательные условия. Для положительных условий будет выбрана ситуация, когда предсказанное значение отклоняется от реального сигнала более чем на 10% в большую или меньшую степень.

Теперь можно построить ROC-кривые и сравнить их AUC. На рисунках 50 – 4.19 представлены кривые ROC и соответствующие AUC для сигналов «VIX», «VXO» и «ABP1». Графики ROC-кривых для остальных сигналов можно найти в приложении Б.

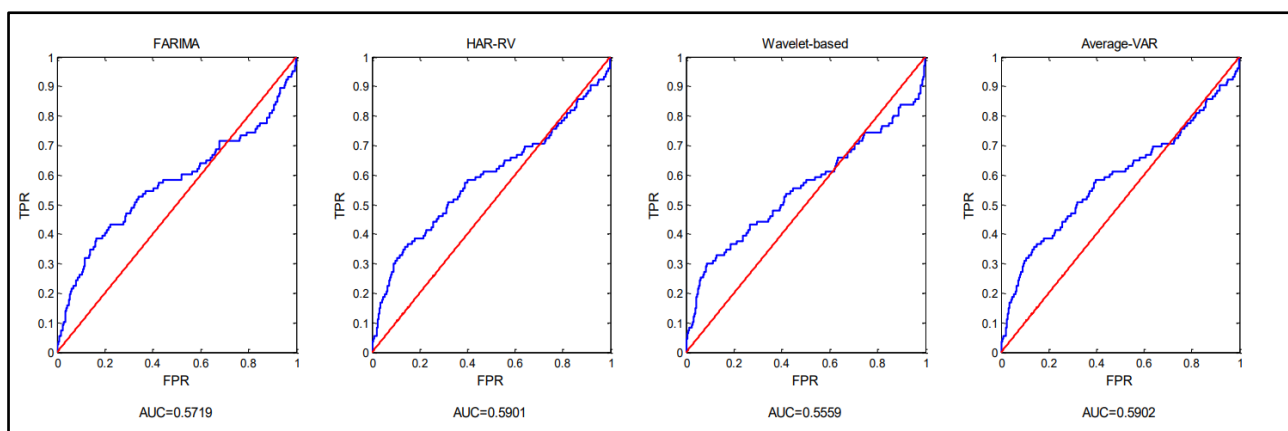


Рисунок 50 – ROC-кривая для сигнала «VIX»

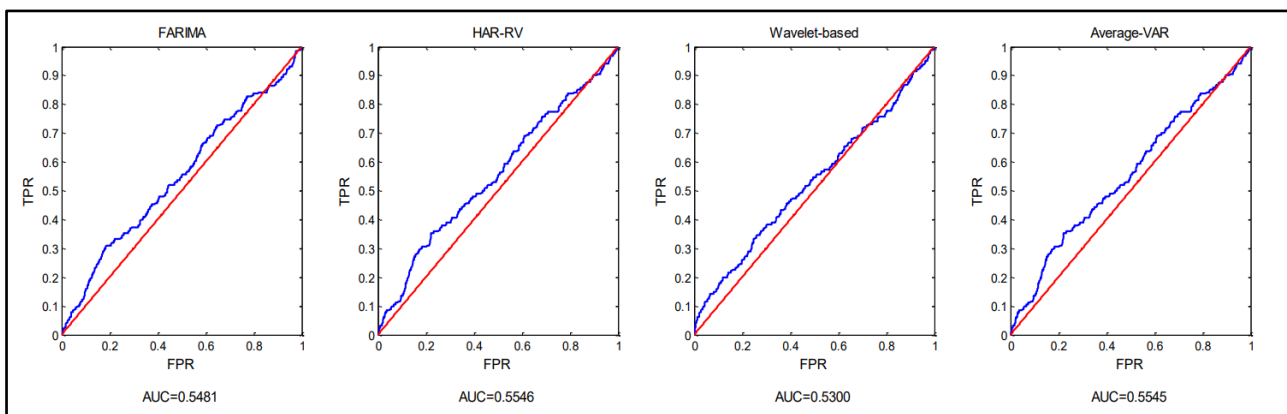


Рисунок 51 – ROC-кривая для сигнала «VXO»

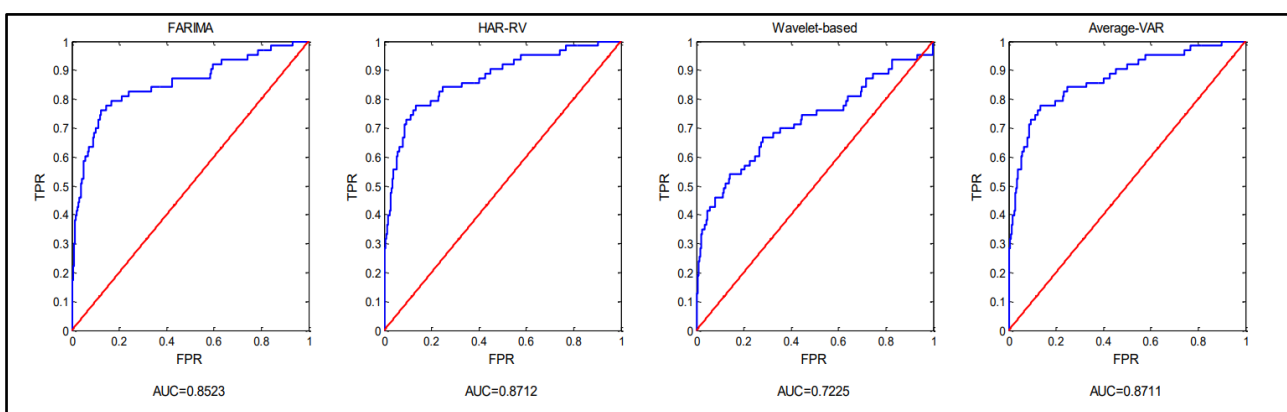


Рисунок 52 – ROC-кривая для сигнала «ABP1»

Из кривых ROC и соответствующих AUC видно, что модель HAR-RV и модель среднего VAR последовательно превосходят две другие модели.

Так же и данных графиков видно, что в случае обнаружения уменьшения будущего значения временного ряда, AUC сильно превышает значение 0.5, в то время как в случае обнаружения увеличения будущего значения, AUC едва превышает значение 0.5. Это указывает на то, что обнаружение уменьшения будущего значения временного ряда является более точным, чем обнаружение увеличения значения.

Вывод к главе 4

В данной главе была описана компьютерная реализация, критерии эффективности, такие как MSE, MAPE, MAE, PRD, а так же способ сравнения ROC-кривых.

Для проведения исследования были собраны два типа временных рядов: индексы волатильности финансовых бирж и сигналы артериального давления в крови. Было использовано три различных временных ряда для каждого типа. Все исследуемые временные ряды имеют количество наблюдений не менее четырёх тысяч значений.

Были произведены визуальные и численные сравнения, а так же сравнения ROC-кривых для каждого временного ряда с использованием всех четырёх моделей: FARIMA, HAR-RV, вейвлет-модель и Average-VAR.

Заключение

В данной диссертационной работе исследуются самоподобные процессы и процессы с долговременной памятью, а также исследуются некоторые из их основных свойств. Одним из основных вкладов данной диссертационной работы является предоставление концепций самоподобия и долговременной памяти. Эти две концепции тесно связаны, но теоретически различаются.

Помимо этого, были исследованы алгоритмы компьютерного моделирования фрактальных рядов. Всего было исследовано четыре модели: FARIMA, HAR-RV, вейвлет-модель и Average-VAR.

Модель FARIMA достаточно популярная, описанная во многих источниках и лучше всех позволяет учесть параметр долговременной памяти. Но она очень сложна в реализации и использовании, а так же требует много времени для анализа. Её точность не особо превосходит остальные модели, но она способна лучше всех уловить моменты изменения направления будущих значений.

Модель HAR-RV была первоначально предложена для моделирования процессов с долговременной памятью при анализе реализованной волатильности. Но при расширении данной модели для моделирования физиологических сигналов, производительность оказалась достаточно хорошей, по сравнению с другими моделями предсказаний, и при этом модель обладает небольшим временем вычислений. Однако проблема заключается в численном подборе масштабов агрегации для каждого временного ряда. Это порядком усложняет реализацию модели и её время выполнения.

Модель на основе вейвлетов была впервые предложена для моделирования самоподобных процессов. К сожалению, на практике многие фрактальные временные ряды оказываются стационарными с долговременной памятью, а не самоподобными. Из-за этого появилась необходимость в усложнении модели для её обобщения и использования для

временных рядов с долговременной памятью. Хотя и результаты работы такой модели кажутся не столь удовлетворительными, по сравнению с результатами других моделей, они дают некоторую информацию для будущих исследований, в которых можно будет пересмотреть вейвлет-модель и учесть ошибки модели, описанной в данной диссертации.

Модель Average-VAR – это достаточно новый подход, предложенный совсем недавно как альтернатива более громоздким моделям, как например FARIMA. Она использует свойство масштабирования сигналов с долговременной памятью. Эта модель теоретически проста и может обеспечить неплохую производительность по сравнению с другими моделями, описанными в данной диссертации. Но Average-VAR, как и модель HAR-RV, требует численного определения уровней масштабирования, а так же её показатели ошибок немного хуже, чем у остальных моделей.

Таки образом можно подвести итоги, что самой быстрой моделью является HAR-RV и Average-VAR, если подобрать подходящие уровни масштабирования заранее, но при этом результаты будут иметь большую погрешность по сравнению с остальными моделями. Если необходимы более точные прогнозы за чуть большее время работы, то лучшим решением будет использование моделей вейвлет-модели. А если же необходимы самые точные результаты, а время работы модели не критично, то стоит использовать модель FARIMA.

Так же сравнение кривых-ROC показало, что обнаружение уменьшения будущего значения временного ряда более точно, чем обнаружение увеличения будущего значения, что стоит учитывать при интерпретировании результатов моделирования временного ряда.

В ходе выполнения магистерской диссертационной работы были решены следующие задачи:

– Исследованы методы моделирования и предсказания фрактальных временных рядов, такие как FARIMA, HAR-RV, вейвлет-модель и Average-VAR;

– Была осуществлена компьютерная реализация исследованных алгоритмов моделирования и прогнозирования фрактальных временных рядов;

– Был произведён сравнительный анализ реализованных алгоритмов для выявления сильных и слабых сторон каждой из моделей и сделаны соответствующие выводы по каждой модели.

Список использованной литературы

1. Андриенко В.М. Фрактальные свойства фондовых рынков / В.М. Андриенко, О.В. Варчук // Одесский национальный политехнический университет. – Одесса, Украина. – 2010, 4 с.
2. Бокс Д. Анализ временных рядов / Дженкинс Г – М.: Мир, 1974.
3. Божокин, С.В. Фракталы и мультифракталы / С.В. Божокин, Д.А. Паршин. – Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. – 126 – 128 с.
4. Боровской И.Г. Исследование прогнозной способности показателя Херста применительно к российскому фондовому рынку / И.Г. Боровской, А.О. Жучков // Доклады ТУСУР. – 2017, №2. – 3 с.
5. Мандельброт Б. Фрактальная геометрия природы. – М.: Институт компьютерных исследований, 2002. – 656 с.
6. Малинецкий Г.Г. Нелинейная динамика. Подходы, результаты, надежды / Г.Г. Малинецкий, А.В. Потапов, А.В. Подлазов. – М.: Комкнига, 2006. – 216 с.
7. Масловская А. Г. Применение фрактальных методов для анализа динамических данных / А.Г. Масловская, Т.Р. Осокина, Т.К. Барабаш // Вестник Амурского государственного университета, №12, 2017, с. 51 – 59.
8. Ньюи У.К. Простая, положительно полуопределенная оценка асимптотической матрицы ковариаций, состоятельная при наличии гетероскедастичности и автокорреляции / У.К. Ньюи, Д.В. Кеннет // № 33 (1), 2014
9. Наследов А. IBM SPSS 20 Statistics и AMOS профессиональный статистический анализ данных // Москва – 2013.
10. Порунов А.Н. Mathcad в руках экономиста: Бокс-Кокс преобразование и иллюзия «нормальности» макроэкономического ряда // Бизнес-информатика №2(12) – 2010.

11. Скроботов А.А. Тестирование единичных корней в панельных данных против неоднородной альтернативы с приложением к региональным индексам потребительских цен РФ // Том 18, Номер 2 , 2017

12. Федер Е. Фракталы // М.: Мир, 1991. – 262 с

13. Цель А. В. Компьютерная реализация фрактального ряда // Тольятти: Тольяттинский государственный университет, Институт математики, физики и информационных технологий, Кафедра Прикладная математика и информатика, 2019, 55 стр.

14. Шайдук А.М. Влияние фрактальной размерности сигнала на распределение энергии в его спектре / А.М. Шайдук, С.А. Останин // Журнал радиоэлектроники. - 2016. - №2.

15. Яковлев А. Н. Введение в вейвлет-преобразования // Новосибирск: Министерство образования Российской Федерации Новосибирский Государственный Технический Университет, 2003, 104 стр.

16. Andersen T. G. "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility," / T. G. Andersen, T. Bollerslev, and F. X. Diebold // The Review of Economics and Statistics, vol. 89, pp. 701-720 2007.

17. Andersen T. G. "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility," / T. G. Andersen, T. Bollerslev, and F. X. Diebold // The Review of Economics and Statistics, vol. 89, pp. 701-720 2007.

18. Burrus C. S., "Introduction to wavelets and wavelet transforms: a primer," / C. S. Burrus, R. A. Gopinath, and H. Guo // 1997.

19. Box G. E. Time series analysis: forecasting and control / G. E. Box, G. M. Jenkins, G. C. Reinsel // John Wiley & Sons, 2013.

20. Bollerslev T. "Modeling and pricing long memory in stock market volatility," / T. Bollerslev, H. Ole Mikkelsen // Journal of Econometrics, vol. 73, pp. 151-184, 1996.

21. Corsi F. "A simple long memory model of realized volatility," Manuscript, University of Southern Switzerland, 2004
22. Corsi F. "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, vol. 7, pp. 174-196, 2009.
23. Doukhan P. *Theory and applications of long-range dependence* / P. Doukhan, G. Oppenheim, and M. S. Taqqu // Springer, 2003.
24. Embrechts P. "An introduction to the theory of self-similar stochastic processes," / P. Embrechts and M. Maejima // *International Journal of Modern Physics B*, vol. 14, pp. 1399-1420, 2000.
25. Fawcett T., "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, pp. 861-874, 2006.
26. Gao R. X. *Theory and Applications for manufacturing* / R. X. Gao, R. Yan, *Wavelets* // Springer, 2010
27. Goldberger A. L. "Physiobank, physiokit, and physionet components of a new research resource for complex physiologic signals," / A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark // *Circulation*, vol. 101, pp. e215-e220, 2000.
28. Goldberger A. L. "Science in pictures: Chaos and fractals in human physiology," / A. L. Goldberger, D. R. Rigney, and B. J. West // *Scientific American*, vol. 262, pp. 42-49, 1990
29. Giron-Nava A. *Circularity in fisheries data weakens real world prediction* / Giron-Nava, A., Munch, S. B., Johnson, A. F., Deyle, E., James, C. C., Saberski, E., Sugihara, G. // 2020
30. Hastings H. M. "Fractals. A user's guide for the natural sciences," / H. M. Hastings and G. Sugihara // Oxford Science Publications, Oxford, New York: Oxford University Press, | c1993, vol. 1, 1993.
31. Hausdorff F. *Dimension und A usseres Mass* / F. Hausdorff // *Matematische Annalen*. – 1919. – 79. – P.157-179.

32. Provata A. Abstract Phase-space Networks Describing Reactive Dynamics / A. Provata, E.Panagakou // National Center for Scientific Research «Demokritos». – Greece. – 2018, 10 c.

33. Reider R. "Volatility forecasting I: GARCH models," New York, 2009

34. Tewfik A. H. "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," / A. H. Tewfik, M. Kim // IEEE transactions on information theory, vol. 38, pp. 904-909, 1992

35. Wang R. Self-Similar Based Time Series Analysis and Prediction // University of Toronto, 2016

Приложение А

Показатели ошибок и время выполнения для других сигналов

Таблица А.1 – Показатели ошибок и время выполнения для сигнала «GVZ»

	MSE	MAPE (%)	MAE	PRD	Время (сек)
FARIMA	25.382	2.88	2.7875	0.5186	26785.252
HAR-RV	24.758	2.87	2.6834	0.4972	309.629
Вейвлет	25.243	2.88	2.6817	0.4976	629.394
Average-VAR	24.788	2.87	2.6837	0.4974	386.321

Таблица А.2 – Показатели ошибок и время выполнения для сигнала «ABP2»

	MSE	MAPE (%)	MAE	PRD	Время (сек)
FARIMA	55.3460	3.25	2.6240	0.4433	10425.935
HAR-RV	57.2400	3.31	2.6682	0.4281	271.586
Вейвлет	60.8650	3.01	2.4944	0.4374	425.150
Average-VAR	57.2812	3.31	2.6693	0.4280	272.875

Таблица А.3 – Показатели ошибок и время выполнения для сигнала «ABP3»

	MSE	MAPE (%)	MAE	PRD	Время (сек)
FARIMA	59.1668	3.17	2.3945	0.4477	6068.3
HAR-RV	57.5167	3.20	2.4061	0.4435	207.085
Вейвлет	148.6356	3.17	2.5647	0.4455	1737.454
Average-VAR	57.8482	3.13	2.3969	0.4436	229.592

Приложение Б

Графики ROC-кривых для других сигналов

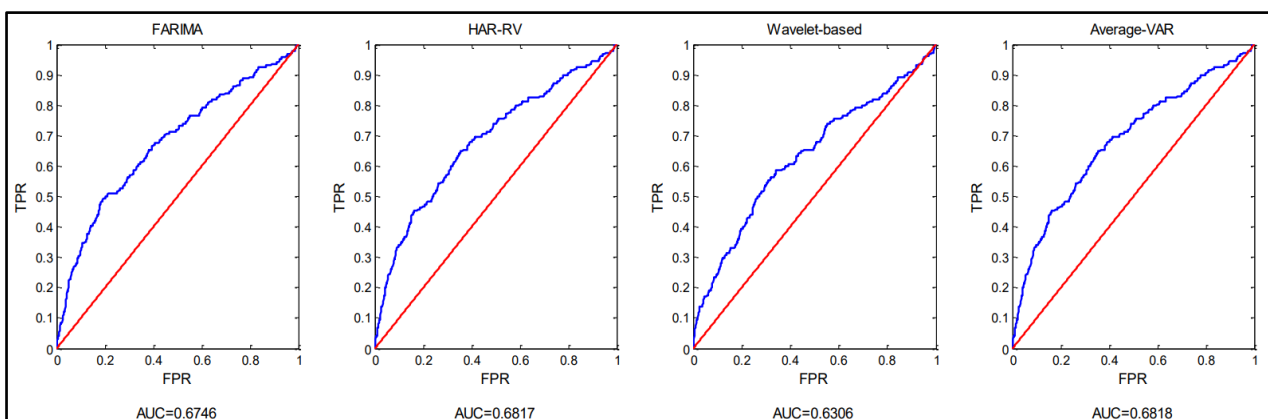


Рисунок Б.1 – ROC-кривая для сигнала «GVZ»

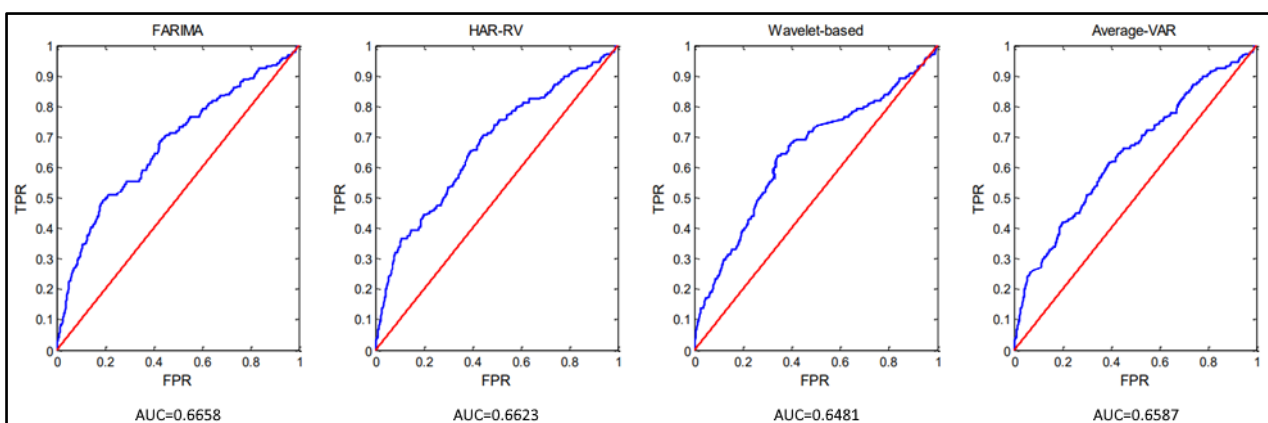


Рисунок Б.2 – ROC-кривая для сигнала «ABP2»

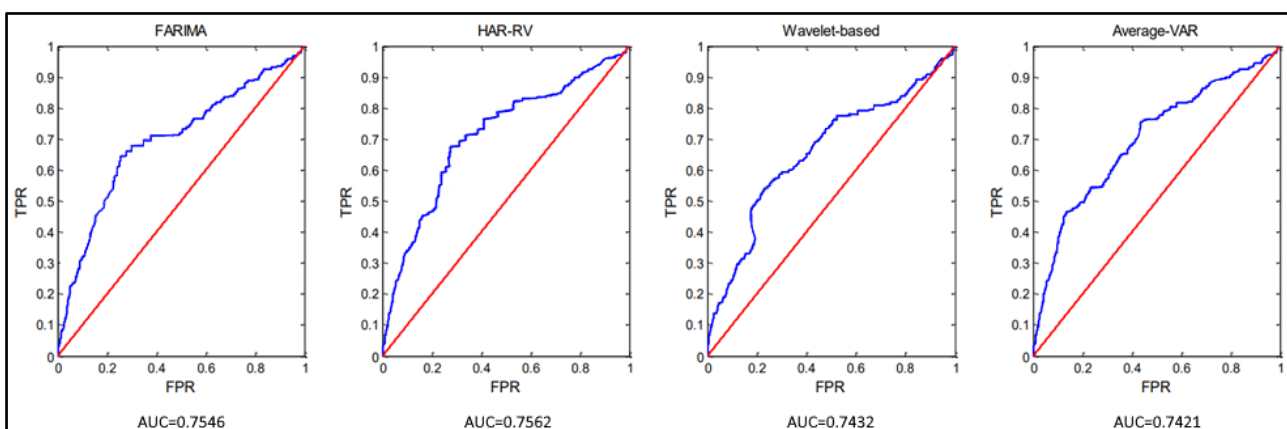


Рисунок Б.3 – ROC-кривая для сигнала «ABP3»