

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и  
стохастического анализа

ВЕЙВЛЕТ-АНАЛИЗ КАК ИНСТРУМЕНТ ИССЛЕДОВАНИЯ РЫНКА  
АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студентки 4 курса 412 группы  
направления 01.03.02 — Прикладная математика и информатика  
механико-математического факультета  
Балдаевой Анастасии Олеговны

Научный руководитель

к. ф.-м. н.

\_\_\_\_\_

М. Г. Плешаков

Заведующий кафедрой

д. ф.-м. н.

\_\_\_\_\_

С. П. Сидоров

Саратов 2017

## СОДЕРЖАНИЕ

1	Определения .....	3
	ВВЕДЕНИЕ .....	5
1	Задачи исследования финансовых временных рядов .....	8
2	Вейвлет анализ .....	11
	2.0.1 Быстрое вейвлетное преобразование .....	18
	2.1 Использование вейвлет-анализа для финансовых временных рядов	21
3	Практическая часть .....	24
	ЗАКЛЮЧЕНИЕ .....	27

## 1 Определения

- Переменные, влияние которых характеризуется некоторым запаздыванием во времени, называются лаговыми переменными.
- Инерционность экономических показателей — привычка людей к определенному образу жизни, потреблению, а также внутренняя инерционность механизмов формирования экономических показателей;
- Институциональные причины — например контракты, трудовые договоры подразумевают определенное постоянство механизмов формирования цен, заработной платы и т.п.;
- Технологические причины — замена оборудования, внедрение новых технологий, безусловно, требует определенного времени.
- Тренд  $u_t$ , представляющий собой плавно меняющуюся компоненту, отражающую влияние долговременных, систематических факторов, основную тенденцию в формировании рассматриваемого показателя. Для моделирования тренда используют плавно меняющиеся, гладкие функции;
- Сезонная компонента  $s_t$ , отражающая повторяемость экономических процессов в течение не слишком длительного периода (например, года, месяца, недели). Причины сезонных колебаний могут быть связаны с природно-климатическими условиями, могут носить социальный характер (например, увеличение закупок в предпраздничные дни, увеличение платежей в конце квартала и т.д.) Для описания сезонной компоненты используют периодические функции;
- Циклическая компонента  $v_t$ , отражающая повторяемость экономических процессов в течении длительных периодов (например волны экономической активности, демографические, инвестиционные циклы);
- Случайная компонента  $\epsilon_t$ , отражающая влияние случайных, а также неучтенных факторов.
- Стационарный временной ряд, у которого математическое ожидание равно нулю, а возмущения  $\epsilon_t$  некоррелированы называется «белым шумом» (таким образом, в классической линейной регрессионной модели возмущения образуют белый шум, а в случае их нормального распределения — нормальный (гауссовский) белый шум).

— Функция

$$\hat{f} : R \rightarrow C, \quad \alpha \rightarrow \hat{f}(\alpha)$$

называется преобразованием Фурье, или спектральной функцией функции  $f$ .

## ВВЕДЕНИЕ

Вейвлеты: определение и история возникновения

Слово «вейвлет» произошло в результате копирования произношения английского аналога «wavelet», переводящегося как «небольшая волна» или «волны, идущие друг за другом». Вейвлет представляет собой некоторую функцию, с помощью которой можно анализировать различные частотные компоненты данных. График вейвлета выглядит как волнообразные колебания, амплитуда которых уменьшается до нуля с увеличением расстояния от начала координат, то есть вейвлет имеет компактный носитель. Вейвлетный базис – семейство функций, каждая из которых ограничена по времени и частоте. Все функции пространства, базисом которого он является, получаются посредством операций сдвига и растяжения базисных вейвлетов. Вейвлет-анализ может быть охарактеризован как альтернатива классическому взвешенному анализу Фурье. Строительными блоками взвешенного анализа Фурье являются синусы и косинусы (волны), умноженные на скользящее окно. Они обычно называются частотно-временными атомами. В вейвлет-анализе окно изначально является осциллирующим и называется "материнским вейвлетом". Вместо умножения на синусы и косинусы этот вейвлет произвольно сдвигается и растягивается по временной оси.

Таким образом, генерирующий (или материнский) вейвлет образует другие вейвлеты, которые являются строительными блоками вейвлет-анализа. Эти растяжения соответствуют различным степеням увеличения, а вейвлеты выполняют роль масштабно-временных атомов вейвлет-анализа.

Вейвлет функции базиса позволяют локализовать особенности анализируемых процессов, которые не могут быть выявлены с помощью традиционного преобразования Фурье. Вейвлеты имеют возможность анализировать нестационарные сигналы (в нашем случае временные ряды) с изменением компонентного содержания во времени или в пространстве. При анализе нестационарных сигналов, а именно - временных рядов, за счет свойства локальности вейвлеты получают существенное преимущество перед своим прямым конкурентом – преобразованием Фурье, которое дает нам только глобальные сведения о частотах исследуемого сигнала, поскольку используемая при этом система функций (синусы, косинусы и комплексные экспоненты) определена на бесконечном интервале. По сравнению с разложением сигналов

на ряды Фурье, вейвлеты с гораздо более высокой точностью представляют локальные особенности сигналов, вплоть до разрывов 1-го рода (скачков). В отличие от преобразований Фурье, вейвлет-преобразование одномерных сигналов обеспечивает двумерную развертку, при этом частота и координата рассматриваются как независимые переменные, что дает возможность анализа сигналов сразу в двух пространствах. Вейвлет представление сигналов на различных уровнях декомпозиции (разложения) заключается в разделении функций приближения к сигналу на две группы: аппроксимирующую - грубую, с достаточно медленной временной динамикой изменений, и детализирующую - с локальной и быстрой динамикой изменений на фоне плавной динамики, с последующим их дроблением и детализацией на других уровнях декомпозиции сигналов. Это возможно как во временной, так и в частотной областях представления сигналов вейвлет разложениями. Вейвлеты выступают самым лучшим инструментом при сжатии большого диапазона данных, что также необходимо в анализе временных рядов, в малый набор коэффициентов вейвлет-преобразования. При этом огромное количество данных, имеющих в абсолютном выражении небольшое отклонение от ярких значений колебаний функции динамики временного ряда, обобщается в один коэффициент без потери важной информации. В данной работе рассматривается анализ финансового временного ряда по средствам вейвлетов, выбор теории вейвлетов обоснован выше на примере сравнения её с теорией Фурье.

Так как целью данной работы является исследование финансовых временных рядов по средствам вейвлетов, необходимо понимать «природу» финансовых временных рядов.

#### Финансовые временные ряды

Временной ряд (ВР) – это последовательность значений, измеренных в последовательные моменты времени (обычно через равные промежутки), которые описывают протекающий во времени процесс. Данные типа временных рядов широко распространены в самых разных областях человеческой деятельности. В экономике это ежедневные цены на акции, курсы валют, еженедельные и месячные объемы продаж, годовые объемы производства и т.п. Эти значения обычно получают как результаты измерений в некоторой физической системе. Если нас интересуют зависимости между текущими и прошлыми значениями, то нужно рассматривать также вектор задержки в пространстве

сдвинутых во времени значений.

Целью практики является исследование основных понятий теории вейвлетов, исследование финансовых временных рядов по средствам вейвлетов, анализ временных рядов на примере изучения и анализа динамики цен на золото в России в период с 2005 по 2016 год.

В данной работе обосновано применение вейвлет-анализа в исследовании временного ряда, показаны его преимущества перед классическими методами, в частности преобразованием Фурье, уже на стадии предварительной обработки данных.

- В результате прохождения практики должны быть отработаны навыки
- методов обработки и анализа экономических данных с помощью вейвлетов;
  - методов прогнозирования финансовых временных рядов;
  - сравнения анализа Фурье и вейвлет-анализа экономических данных на примере конкретной задачи;
  - документирования результата работы программы и обоснования оптимального выбора вида анализа.

Практическая значимость исследования заключается в том, что сформулированные выводы и предложения, разработанные модели и алгоритмы могут быть использованы финансовыми учреждениями, частными инвесторами, разработчиками информационно-аналитических систем, другими субъектами рынка ценных бумаг в качестве инструментария для получения дополнительной информации, способствующей повышению степени обоснованности инвестиционных решений.

## 1 Задачи исследования финансовых временных рядов

Повседневная практика финансовых рынков находится в противоречии с теорией технического анализа, аксиомой которого является то, что изменения цен финансовых активов происходят мгновенно, эффективно отражая всю доступную информацию. Любой фактор, влияющий на цену, — экономический, политический, психологический, — заранее учтен и отражен в её временном ряде. Основная цель анализа финансовых временных рядов - извлечение из данного ряда полезной информации для того, чтобы были возможны: предсказание будущих значений на основе прошлых наблюдений; подбор статистической модели, описывающей финансовый временной ряд; краткое описание характерных особенностей ряда и управление процессом, порождающим временной ряд.

Предсказание будущих значений, то есть прогнозирование является важнейшим элементом любой инвестиционной деятельности. Закономерности развития финансовых рядов постоянно меняются, а эти изменения могут происходить очень быстро или же постепенно.

В работе [1] приводится необходимая информация из теории временных рядов для проведения исследования, а также описываются причины использования лаговых переменных, рассматривается классификация динамических моделей, описываются компоненты уровней временных рядов.

Типичной задачей при исследовании временного ряда является выявление и оценка неслучайных компонент, а также изучение статистических свойств случайной компоненты. При выборе модели детерминированной составляющей прежде всего учитываются содержательные составляющие, экономические закономерности, обуславливающие ее формирование. В качестве основных этапов исследования временных рядов можно выделить следующие:

1. Графическое представление временного ряда;
2. Выявление и оценка закономерных (неслучайных) составляющих, удаление их из рассматриваемого ряда;
3. Сглаживание и удаление низко- и высокочастотных составляющих (так называемая фильтрация);
4. Исследование случайной составляющей, проверка адекватности построенной модели;
5. Прогнозирование на основе построенной модели.



## Авторегрессия и автокорреляция

При анализе временных рядов достаточно широкое распространение получили регрессионные модели с лаговыми переменными, а также авторегрессионные модели. В общем случае авторегрессионная модель  $p$ -го порядка (или модель  $AR(p)$ ) имеет вид  $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \epsilon_t$ , здесь  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  - константы. Таким образом, эта модель описывает изучаемый показатель в момент  $t$  в зависимости от его значений в предыдущие моменты  $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-p}$ . Кроме значений показателя в предыдущие моменты времени в качестве регрессоров могут быть и другие факторы, оказывающие влияние на изучаемый показатель, пример подобной смешанной модели:  $y_t = \beta_0 + \beta x_t + \gamma y_{t-1} + \epsilon_t$ . Авторегрессионная модель 1-го порядка (то есть модель  $AR(1)$ ), имеет вид:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t,$$

эта модель представляет собой марковский случайный процесс, согласно которому значения процесса в момент  $t$  определяется его значениями только в предыдущий момент  $t-1$ .

Степень тесноты связи между уровнями временного ряда, сдвинутых относительно друг друга на  $\tau$  единиц (с лагом  $\tau$ ) можно оценить, вычисляя коэффициент корреляции между последовательными наблюдениями  $y_1, y_2, \dots, y_{n-\tau}$  и — так называемый коэффициент автокорреляции  $\rho(\tau)$ . Статической оценкой  $\rho(\tau)$  является выборочный коэффициент автокорреляции  $r(\tau)$ , определяемый по формуле:

$$r(\tau) = \frac{(n - \tau) \sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n - \tau) \sum_{t=1}^{n-\tau} y_t^2 - (\sum_{t=1}^{n-\tau} y_t)^2} \sqrt{(n - \tau) \sum_{t=1}^{n-\tau} y_{t+\tau}^2 - (\sum_{t=1}^{n-\tau} y_{t+\tau})^2}} \quad (1)$$

Если полученное значение коэффициента автокорреляции окажется меньше табличного, то это свидетельствует об отсутствии во временном ряде существенной автокорреляции.

Факторы, обуславливающие колебания уровней временного ряда, как правило, объективны, что предопределяет самостоятельное исследование колеблемости. Типы колебаний статистических показателей весьма разнообразны. Три основных типа колебаний: пилообразная или маятниковая колебле-

мость, циклическая долгопериодическая колеблемость и случайно распределенная во времени колеблемость.

В нашей работе выведены правила, по которым строится анализ структуры временного ряда с помощью коэффициентов автокорреляции.

## 2 Вейвлет анализ

Так как в нашей работе обосновано применение вейвлет-анализа для конкретной задачи, в которой мы рассматриваем финансовый временной ряд, необходимо упомянуть и такой мощный инструмент как анализ Фурье. Также, необходимо привести сравнительную характеристику вейвлет-анализа и анализа Фурье, где будет подробно описано преимущество выбранного инструмента для данной задачи.

Стоит отметить, что при исследовании информации предыдущих отрезков временного ряда, чем дальше мы будем удаляться в историю ряда, тем меньшее влияние эта информация будет производить на поведение этого ряда в будущем. Именно вейвлеты помогают нам найти такое представление динамики ряда, которое бы имело избирательную точность, то есть: чем дальше в прошлое, тем меньше деталей, при сохранении общего вида кривой.

Ряды Фурье Ряды Фурье касаются  $2\pi$  периодических функций

$$f : \mathbb{R} \rightarrow \mathbb{C}, f(t + 2\pi) = f(t),$$

что может быть записано и иначе  $f : \frac{\mathbb{R}}{2\pi} \rightarrow \mathbb{C}$ . «Естественной» областью определения такой функции является единичная окружность  $S^1$  на комплексной  $z$ -плоскости (рис.2). На  $S^1$  бесконечное число точек  $t + 2k\pi$ ,  $k \in \mathbb{Z}$ , эквивалентных по модулю  $2\pi$ , представлено одной точкой  $z = e^{it}$ .

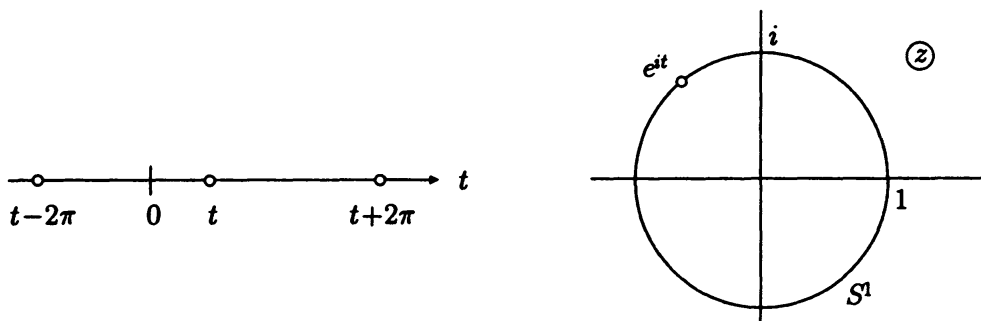


Рисунок 2

«Преобразование Фурье», которое ставит в соответствие  $2\pi$ -периодической функции  $f$  её массив коэффициентов Фурье  $(c_k | k \in \mathbb{Z})$ , рассматривает  $f$  в качестве "глобального объекта". В частности, в этом случае нет никакой ло-

кализации по временной оси. В массиве  $(y_k | 0 \leq k < N)$ ,

$$y_k := f\left(\frac{2k\pi}{N}\right) \quad (0 \leq k < N),$$

т.е. в простой таблице значений  $f$ , информация об  $f$  сохраняется в таком виде, что становится возможной непосредственная и точная локализация индивидуальных деталей (например: локальных экстремумов, точек поворота и т.д.) на временной оси. В отличие от этого свойства таблицы  $(y_k | 0 \leq k < N)$ , каждый отдельный коэффициент Фурье  $c_k$  содержит информацию об  $f$ , относящуюся ко всей области определения  $f$ , т.е. невозможно определить, например, где  $f$  имеет скачок или максимум.

Пример: Разрывная функция

$$f(t) = \begin{cases} \frac{1}{2}(\pi - t) & (0 < t < 2\pi), \\ 0 & (t = 0), \\ f(t + 2\pi) & \forall t. \end{cases}$$

График функции представлен на рисунке 4.

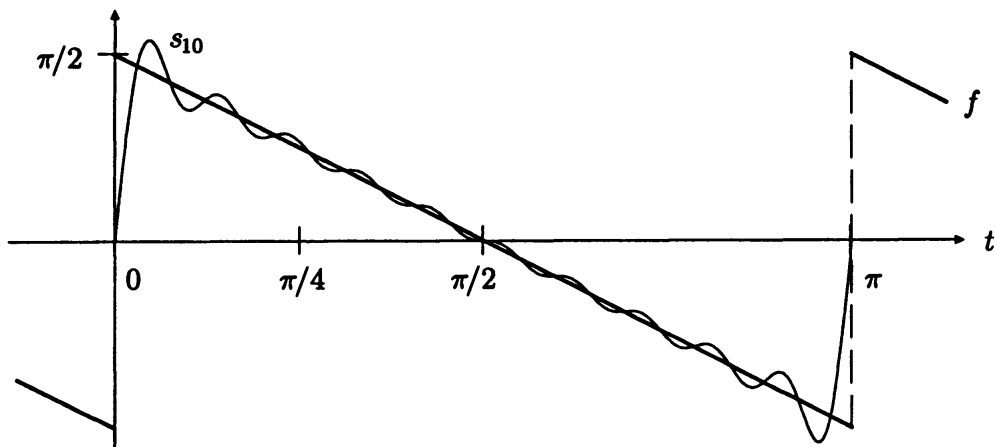


Рисунок 4

Имеет следующее представление в виде ряда Фурье:

$$f(t) = \sum_{k=1}^{\infty} \frac{1}{k} \sin(kt).$$

Данный ряд действительно представляет  $f$  во всех точках  $t$ , но не сходится он «равномерно плохо»: так как коэффициенты  $\frac{1}{k}$  убывают очень медленно при

$k \rightarrow \infty$ , то в каждой точке  $t \neq 0 \pmod{2\pi}$  сходимость зависит от осцилляций  $k \rightarrow \sin(kt)$ .

Более того: проявляется явление Гиббса, а именно - любая частичная сумма  $s_N$  ряда Фурье превосходит максимальное значение функции, равное  $\frac{\pi}{2}$ , в некоторой точке  $t_N$  около нуля приблизительно на 18 %.

Таким образом, если нам необходимо провести анализ Фурье функции  $g$ , представленной на рисунке 5, то сразу надо смириться с тем, что вследствие скачка непрерывности в точке  $t_0$  эта функция имеет ряд Фурье плохо сходящийся всюду.

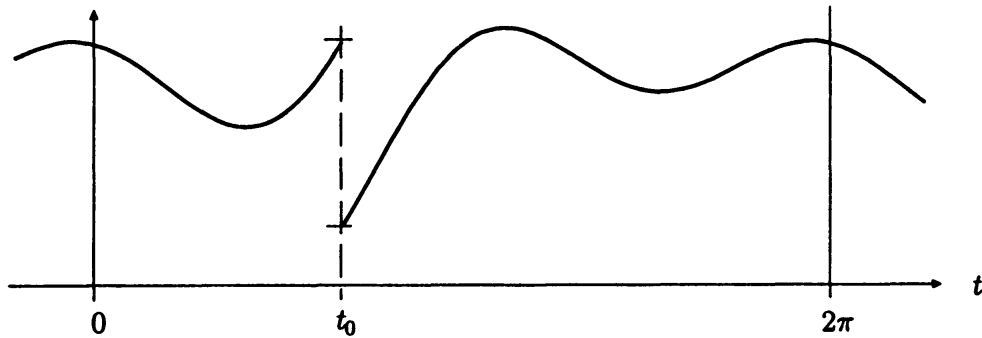


Рисунок 5

Более того, рассматривая только  $c_k$ , нельзя установить расположение скачка, хотя это может быть самой интересной деталью  $g$ .

Если же функция  $f$  аппроксимируется с помощью вейвлетов, то в этом случае совершенно определенно наличествует некоторый тип локализации; более того, эта локализация скроена по мерке: детали переходных процессов (кратковременные детали) функции  $f$ , такие, например, как скачки или отмеченные пики могут быть легко локализованы при непосредственном рассмотрении вейвлетных коэффициентов, в то время как более медленные изменения  $f$  запоминаются в более глубоких слоях иерархии коэффициентов и автоматически представляются в более мелком масштабе; как следствие этого, они менее точно локализуются на временной оси.

### Взвешенное преобразование Фурье

Необходимо получить такой «тип данных», который позволял бы легко находить или извлекать как временную(или пространственную), так и частотную информацию о сигнале  $f : R \rightarrow C$ .

Так называемое взвешенное преобразование Фурье (ВПФ) представляет непрерывную версию такого типа данных. Однако, одновременная локализа-

ция (в пределах фундаментальных ограничений) по отношению к временной и частотной переменным осуществляется за счёт введения огромной избыточности по той причине, что в этом случае множество индексов получающегося вектора данных  $(Gf(\alpha, s) | (\alpha, s) \in R \times R)$  является двумерным, в то время как с его помощью кодируется функция одного действительного переменного  $t$ .

ВПФ описывается следующим образом: сначала выбирается и фиксируется весовая функция  $g : R \rightarrow R_{\geq 0}$ . Функция  $g$  должна иметь «общую массу» 1 и быть более или менее сконцентрированной в окрестности  $t = 0$ , что означает, что она должна иметь компактный носитель, содержащий 0 (Рис. 7) или по крайней мере иметь максимум при  $t = 0$  и быстро убывать при  $|t| \rightarrow \infty$ .

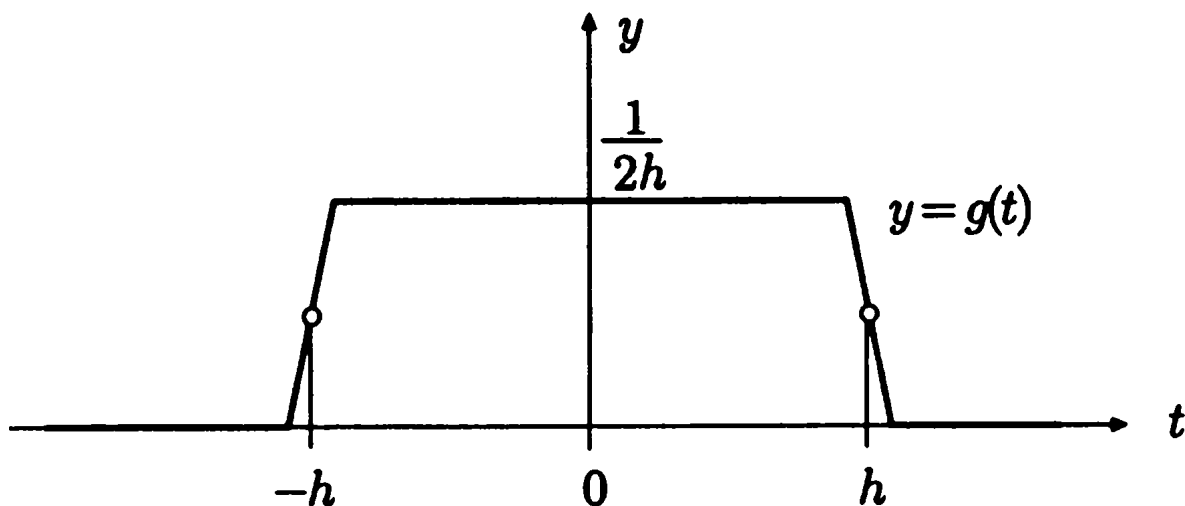


Рисунок 7

Часто используемая весовая функция (или функция окна) имеет вид

$$g(t) := N_{\sigma,0}(t) := \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad (12)$$

где  $\sigma$ - фиксированный параметр. Соответствующее преобразование часто называют преобразованием Габора, так как Деннис Габор (Нобелевский лауреат по физике, 1971) был одним из первых, кто систематически использовал ВПФ. Для заданного  $s \in R$ , функция  $g_s : t \rightarrow g(t - s)$  представляет собой весовую функцию  $g$ , сдвинутую на  $s$  (вправо, если  $s > 0$ ). Определяем взвешенное преобразование

$$Gf : R \times R \rightarrow C, \quad (\alpha, s) \rightarrow Gf(\alpha, s)$$

функции  $f$  как:

$$Gf(\alpha, s) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)g(t-s)e^{-it\alpha} dt. \quad (13)$$

Формула (13) будет интерпретирована для весовой функции  $g$ , изображенной на предыдущем рисунке, так: величина  $Gf(\alpha, s)$  представляет в некотором смысле комплексную амплитуду, с которой чистая гармоника  $e_\alpha$  присутствует в  $f$  в течение временного интервала  $[s-h, s+h]$ . Если в течение этого интервала, среди прочих, прозвучит «нота»  $\alpha$ , то  $|Gf(\alpha, s)|$  будет большой.

Так как информация об  $f$  в  $Gf$  представлена с избыточностью, то существует несколько формул обращения для взвешенного преобразования Фурье  $f \rightarrow Gf$ . Для практических целей следует применять дискретный вариант ВПФ, используя равномерное подразбиение как оси  $t$ , так и оси  $\alpha$ .

Рассмотрим первый случай:

Для постоянного весового окна ширины  $2h$  (соответственно,  $\sim 2\sigma$  в случае (12)) при  $|\alpha| \gg \frac{1}{h}$  «основной волновой пакет»  $t \rightarrow g(t-s)e^{-it\alpha}$  имеет форму изображенную на рисунке 8.

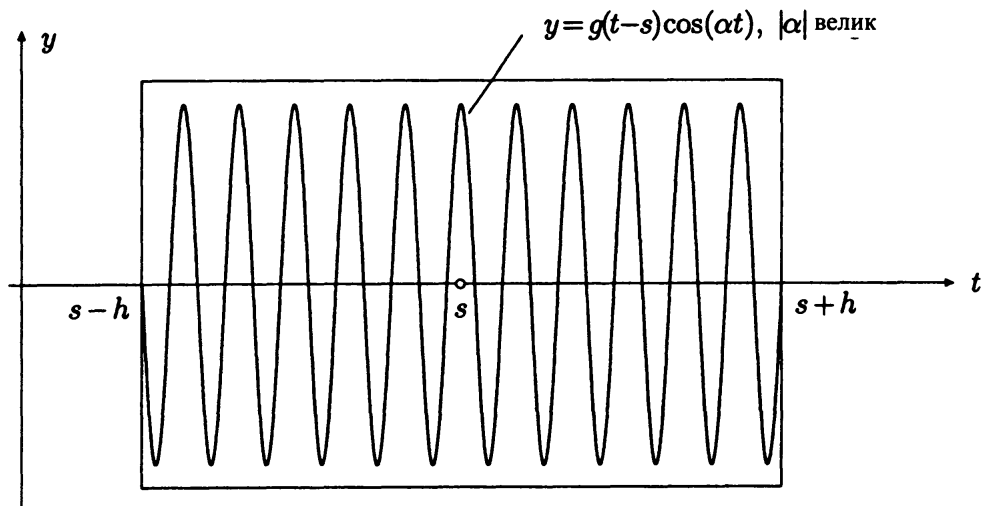


Рисунок 8

Пусть теперь заданный сигнал содержит всего несколько колебаний частоты  $\alpha$  на интервале  $[s-h, s+h]$ , которые занимают очень небольшую часть интервала. Значение  $Gf(\alpha, s)$  при этом будет достаточно велико, но «основной волновой пакет» на представленном выше рисунке не даёт возможностей найти локализацию таких коротких импульсов с достаточной точностью.

Рассмотрим второй случай:

При  $|\alpha| \ll \frac{1}{h}$ , т.е в самом низу диапазона, дела обстоят хуже: «основной волновой пакет» имеет форму изображенную на рисунке 9.

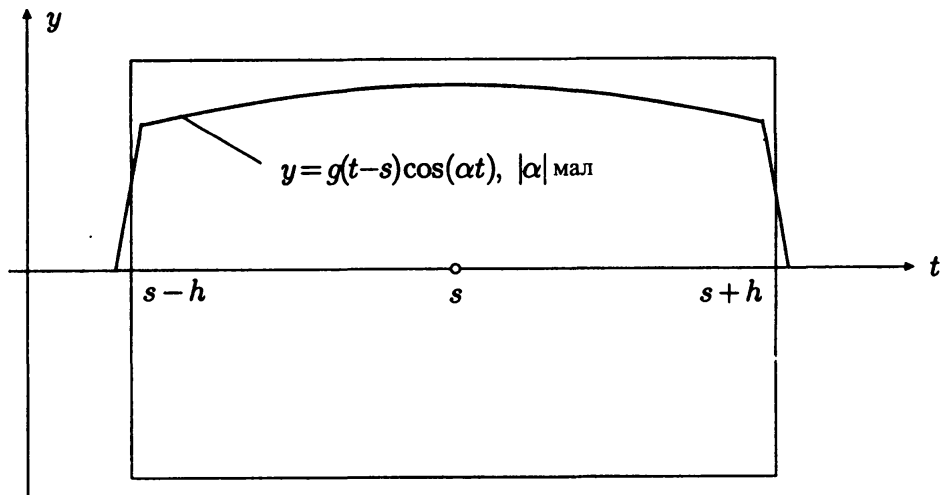


Рисунок 9

и если сигнал  $f$  содержит (возможно, представляющую наибольший интерес) колебательную компоненту характеристической частоты  $|\alpha| \ll \frac{1}{h}$ , то преобразование  $G$  не обнаружит её, потому что весовое окно будет слишком узким, чтобы охватить хотя бы один полный период такой низкой частоты. Основная модель вейвлетного преобразования действует на комплексно-значных временных сигналах  $f : R \rightarrow C$ . Сначала выбирается подходящий анализирующий вейвлет, который также называется материнским вейвлетом или просто вейвлетом,  $x \rightarrow \psi(x)$ .

На рисунке 10 изображен  $\psi$ , имеющий компактный носитель  $[0, L]$ .

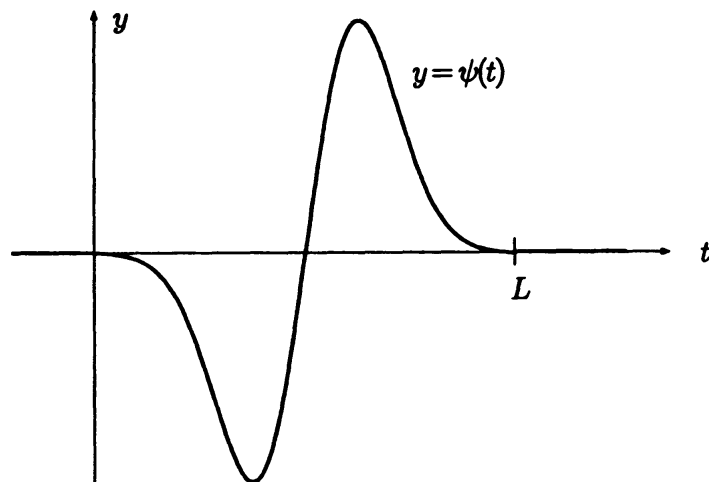


Рисунок 10



Растянутые и сдвинутые копии вейвлета  $\psi$  называют вейвлетными функциями. «Основные волновые пакеты», используемые для анализа временных сигналов  $f$ , собственно и являются такими вейвлетными функциями, обозначаются они таким образом:

$$\psi_{a,b} : R \rightarrow C, t \rightarrow \frac{1}{|a|^{\frac{1}{2}}} \psi \left( \frac{t-b}{a} \right) \quad (14).$$

Двойной индекс  $(a, b)$ , появляющийся здесь, принадлежит множеству  $R^* \times R$  или  $R_{>0} \times R$ . Переменная  $a$  называется масштабирующим параметром, а  $b$  - параметр сдвига. Множитель  $\frac{1}{|a|^{\frac{1}{2}}}$  в (14) не принципиален и носит технический характер, он используется для того, чтобы гарантировалось  $\|\psi_{a,b}\| = 1$ .

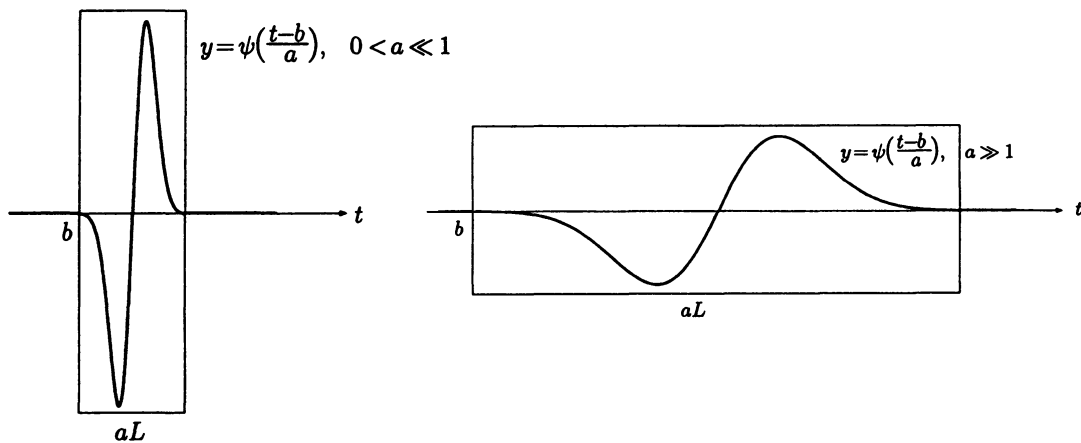


Рисунок 11

Как становится очевидным из рассмотрения Рисунка 11, что ширина «основного волнового пакета» или «основного окна» растет пропорционально  $|a|$ , и для всех значений  $a$  и  $b$  это окно представляет единственную и полную копию анализирующего вейвлета.

Значения масштабирующего параметра  $a$  по модулю  $0 < |a| \ll 1$  соответствуют очень узким окнам и служат для точной локализованной регистрации высокой частоты соответствующих соответствующих переходных процессов, имеющих в сигнале  $f$ .

Значения масштабирующего параметра  $a$  по модулю  $|a| \gg 1$  соответствуют очень широким окнам и служат для регистрации медленных процессов или длинноволновых колебательных составляющих  $f$ .

Благодаря всему вышесказанному становится ясным, что вейвлетное преобразование

$$Wf : R^* \times R \rightarrow C, (a, b) \rightarrow Wf(a, b)$$

временного сигнала  $f$  определяется следующим образом:

$$Wf(a, b) := \langle f, \psi_{a,b} \rangle = \frac{1}{|a|^{\frac{1}{2}}} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt.$$

Более корректно писать  $W_\psi f$  вместо  $Wf$ , так как получающийся массив данных  $(Wf(a, b) | (a, b) \in R^* \times R)$  зависит от вейвлета  $\psi$ , выбранного в самом начале. Во всех тех случаях, когда рассматривается только один вейвлет, можно и не использовать обозначение  $W_\psi f$ .

Областью определения преобразования  $Wf$  является плоскость  $(a, b)$ , так как переменная  $b$  описывает сдвиг по временной оси, то в теории вейвлетов принято  $b$ -ось располагать горизонтально, а  $a$ -ось располагать вертикально в отличие от обычного расположения осей, соответствующих первому и второму множителям декартового произведения.

Для вейвлетного преобразования также существует формула обращения. Эта формула представляет исходный сигнал  $f$  в виде «линейной комбинации» базисных функций  $\psi_{a,b}$  с коэффициентами, равными значениям  $Wf(a, b)$  вейвлетного преобразования. Для того, чтобы установить такую формулу, необходимо определить "элемент объёма" на множестве индексов  $R^* \times R$ .

Если функции  $\psi_{a,b}$  заданы в виде (14), то получаем

$$f = \frac{1}{C_\psi} \int_{R^* \times R} \frac{dad b}{|a|^2} Wf(a, b) \psi_{a,b}$$

с константой  $C_\psi$ , зависящей только от выбранного  $\psi$ .

### 2.0.1 Быстрое вейвлетное преобразование

Принципиально важным обстоятельством в описанной выше ситуации является то, что по оси масштабирования (вейвлетного аналога частотной оси) преимущественно используется логарифмический масштаб. Подобное положение в реальной действительности может быть знакомо из акустики или музыки: равные шаги тональности соответствуют равным частотным от-

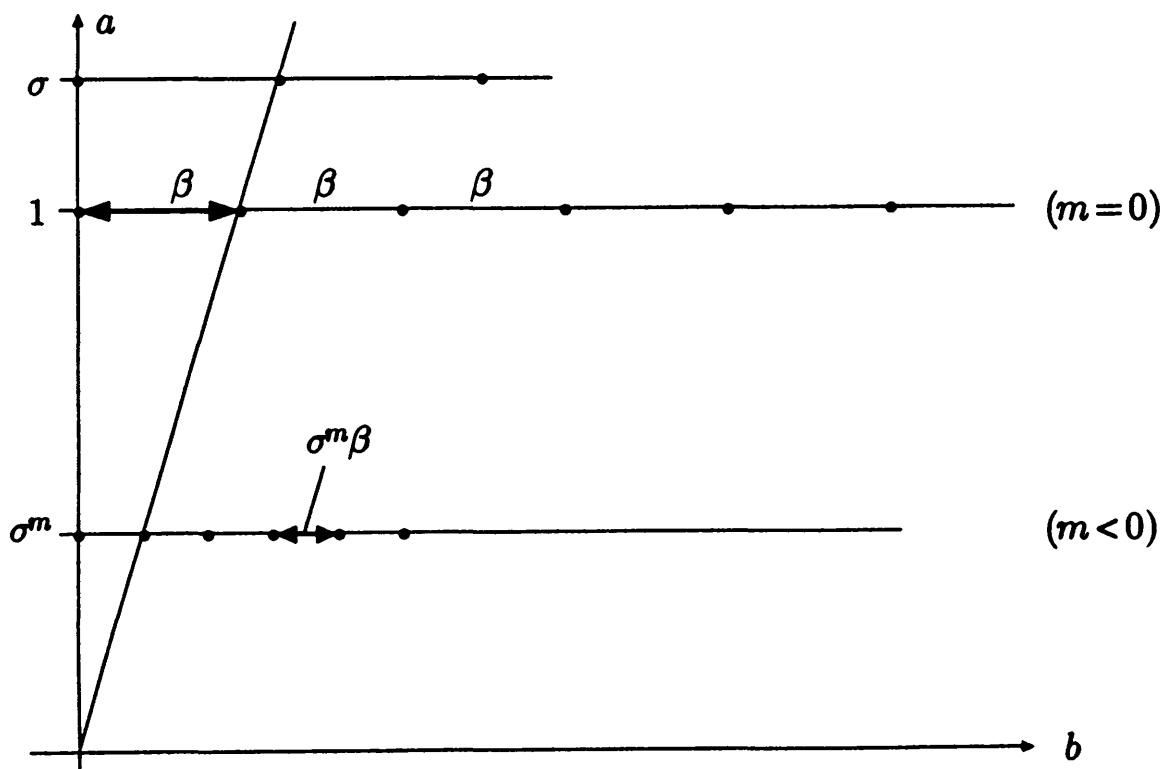
ношениям  $\frac{\omega_2}{\omega_1}$  (например,  $5 : 4$ ) для основной трети), а не равным частотным разностям  $\omega_2 - \omega_1$ .

Этот факт становится особенно очевидным, когда в качестве следующего шага мы дискретизируем множество индексов  $R_{>0} \times R$ : мы выбираем шаг растяжения  $\sigma > 1$  (обычно используется  $\sigma = 2$ ) и рассматриваем далее только дискретное множество множителей растяжения

$$a_r := \sigma^r \quad (r \in Z).$$

Отметим, что большим числам  $r \in Z$  соответствуют большие множители растяжения  $a_r > 0$ . Что же касается параметра сдвига, то мы не можем просто выбрать базовый шаг  $\beta > 0$  и затем использовать единственную решетку значений сдвига  $b_k := k\beta (k \in Z)$ , как в случае преобразования Фурье. Происходит следующее: при более тонких масштабах при меньших значениях  $r$  необходимо использовать соответственно меньший шаг сдвига, для того, чтобы всё было в порядке. Например, на уровне  $a_r$  на плоскости  $(a, b)$  ( $a$  располагается вертикально,  $b$  - горизонтально!) мы выбираем в качестве значений решетки числа

$$b_{r,k} := k\sigma^r \beta \quad (k \in Z)$$



(Рис. 12) Это означает, что последовательные  $b_{r,k}$  располагаются на расстоянии  $\sigma^r \beta$  друг от друга. Такой выбор, как показывает простое наблюдение, является вполне естественным. В частности, такой выбор решетки соответствует оптимальной точной локализации высоких частот и \и или переходных процессов в анализируемом временном сигнале  $f$ .

В таком духе мы устанавливаем существование дискретной группы самоподобий, связывающей  $\mathbb{R}$ , с одной стороны, и  $\psi$  и его промасштабированные версии, с другой стороны. Систематическое использование этой группы приводит к так называемому мультиразрешающему анализу и связанным с ним быстрому алгоритму. Последний, обычно именуемый быстрым вейвлетным преобразованием (БВП, используется для вычисления вейвлетных коэффициентов

$$c_{r,k} := Wf(a_r, b_{r,k})$$

а также для реконструкции (то есть синтеза) сигнала  $f$  из сохраненных данных  $c_{r,k}$ .

В отличие от жестких рамок фурье-анализа, в выборе анализирующего вейвлета  $\psi$  мы располагаем большой свободой. В сущности, достаточно быть уверенным, что  $\psi$  принадлежит  $L^1 \cap L^2$  и что  $\int_{-\infty}^{\infty} \psi(t) dt = 0$ . В зависимости от обстоятельств и требований, всегда можно добиться того, что

- $\psi$  имеет компактный носитель,
- вейвлетные функции («основные волновые пакеты»)

$$\psi_{r,k}(t) := 2^{-\frac{r}{2}} \psi \left( \frac{t - k \cdot 2^r}{2^r} \right),$$

принадлежащие к описанной дискретизации, являются ортонормированными,

- существуют быстрые алгоритмы,
- $\psi$  дифференцируем требуемое число раз,
- вейвлетные коэффициенты обладают оптимальным убыванием при  $r \rightarrow -\infty$ ,
- и так далее.

Так как в нашей работе мы пользуемся вейвлетами Добеши, то подробная теория о них, также, содержится в главе 2.

## 2.1 Использование вейвлет-анализа для финансовых временных рядов

В данной главе подробно описан функционал для работы с сигналами в Matlab и построение вейвлет анализа для рассматриваемой задачи. Пакет расширения Wavelet Toolbox системы Matlab позволяет использовать вейвлет-ный анализ и преобразование данных в самых различных областях науки и техники. Программное обеспечение пакета позволяет выполнять вейвлет-преобразования как в командном режиме (и готовить специализированные программы), так и в диалоговом режиме по интерфейсу GUI (включение командой "wavemenu" или из окна редактора, Wavelet Toolbox -> Main Menu).

Непрерывное Одномерное Преобразование.

Основные функции. Непрерывное одномерное вейвлет-преобразование (НВП-1D) уже само по себе, без реконструкции сигналов, используется для анализа формы сигналов и выявления их локальных особенностей. Преобразование выполняется функцией `cwt` в следующих форматах:

- $C = \text{cwt}(S, \text{SCALES}, 'wname')$  – возвращает коэффициенты 'с' прямого НВП вещественного или комплексного сигнала  $S$  вейвлетом 'wname' в шкале масштабирования  $\text{SCALES}$ . Стандартное задание  $\text{SCALES} = \text{начало} : \text{шаг} : \text{конец}$  (по значениям коэффициента масштабирования "a").
- $C = \text{cwt}(S, \text{SCALES}, 'wname', 'plot')$  – то же, плюс строит график коэффициентов.
- $C = \text{cwt}(S, \text{SCALES}, 'wname', \text{plotmode})$  – то же, с заданием для построения графика коэффициентов следующих настроек цвета `plotmode`:
  - 'lvl' – окраска шаг за шагом,
  - 'glb' – окраска с учетом всех коэффициентов,
  - 'abslvl' или 'lvlabs' – окраска шаг за шагом с использованием абсолютных значений коэффициентов,
  - 'absglb' или 'glbabs' - окраска с масштабированием и с использованием абсолютных значений коэффициентов.
- $C = \text{cwt}(S, \text{SCALES}, 'wname', \text{plotmode}, \text{xlim})$  – то же, с дополнительной настройкой цвета  $\text{xlim} = [\text{xmin}, \text{xmax}]$ , где значениями в квадратных скобках устанавливаются номера точек векторов коэффициентов  $C$  (значения сдвига  $b_{min}$  и  $b_{max}$ ), по интервалу которых определяются

значения  $C_{min}$  и  $C_{max}$  интервала изменения цветовой окраски.

Интерфейс GUI удобен для анализа данных в диалоговом режиме. Демонстрация сигналы загружаются в окно из меню File -> Example Analysis, устанавливается тип вейвлета, параметры анализа и нажимается кнопка 'Analyze', после чего в графической части окна появляется сигнал и результаты его разложения в трех представлениях (полное, сечение по среднему уровню разложения и линии локальных максимумов). Графическое представление можно изменять ниже расположенными кнопками и переключателями, а также используя типовые возможности оконного меню. Коэффициенты разложения можно записать на диск в виде mat-файла (File -> Save Coefficients) и считать затем в рабочую область Matlab для детального изучения.

При выполнении непрерывного разложения комплексными вейвлетами в 'Wavelet Toolbox Main Menu' используется кнопка 'Complex Continuous Wavelet 1-D'.

При обработке данных, записанных в других (не .mat) форматах, следует сначала перевести данные в mat-формат, что можно выполнить из основного окна Matlab (Файл -> Импорт данных), или из окна команд. При обработке одномерных сигналов второе предпочтительнее, так как одновременно дает возможность подготовить для GUI однострочные векторные массивы.

### ДИСКРЕТНОЕ ОДНОМЕРНОЕ ПРЕОБРАЗОВАНИЕ.

Главным достоинством дискретного вейвлет-преобразования (ДВП) является возможность быстрого преобразования (БВП) с пирамидальным алгоритмом вычислений, что позволяет выполнять анализ больших выборок данных. Однако возможности БВП реализуются не для всех типов вейвлетов. Тем не менее, при обработке данных БВП используется весьма интенсивно и в пакете Wavelet Toolbox представлено большим количеством специальных функций. Многоуровневое вейвлет-разложение сигналов (декомпозиция) выполняется функцией `wavedec`, которая используется в двух формах:

? [C,L] = `wavedec`(S, N, 'wname').

? [C,L] = `wavedec`(S, N, LD, HD).

Функция возвращает векторы разложения сигнала S на уровне N с использованием вейвлета 'wname' или его низкочастотного (LD) и высокочастотного (HD) фильтров декомпозиции. N должно быть целым числом и определяет длину вектора L ( $\text{length}(L)=N+2$ ). Значение N при K-точках сигнала должно

быть не более  $\frac{K}{2} > 2^N \leq K$ . Состав векторов C и L поясняет нижеследующий пример.

```
Пример вычисления коэффициентов декомпозиции сигнала (рис. 1.2). dwtmode('save
load sumsin; S=sumsin(1:200); [C,L]=wavedec(S,3,'db4');
subplot(131); plot(S); grid; title('signal S');
subplot(132); plot(C); grid; title('signal C');
subplot(133); plot(L); grid; title('signal L');
```

Первой строкой в приведенном примере стоит включение метода задания граничных условий свертки. В данном случае включен метод 'per' - периодизации продолжения сигналов. Возможные методы задания граничных условий будут рассмотрены ниже.

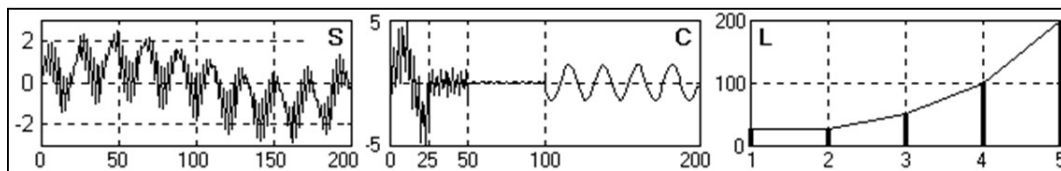


Рисунок 1.2

### 3 Практическая часть

В данной главе мы исследуем финансовый временной ряд с помощью вейвлет-анализа, в качестве входного сигнала нами были рассмотрены данные, отражающие изменение цен на золото в России в период с 2005 по 2016 год. Мы использовали многоуровневый одномерный вейвлет-анализ сигнала, получая при этом аппроксимирующие коэффициенты  $c_{A_N}$  и детализирующие коэффициенты  $c_{D_N}$ , выше была изложена теория для алгоритмов Малла[8]. Результаты разложения на коэффициенты графически представлены на рисунке 1.3 в форме дерева, также на данном рисунке присутствует исходный сигнал и очищенный от шума, но это более наглядно видно на рисунке 1.5. На рисунке 1.4 детально изображена декомпозиция сигнала на рассматриваемом уровне 5.

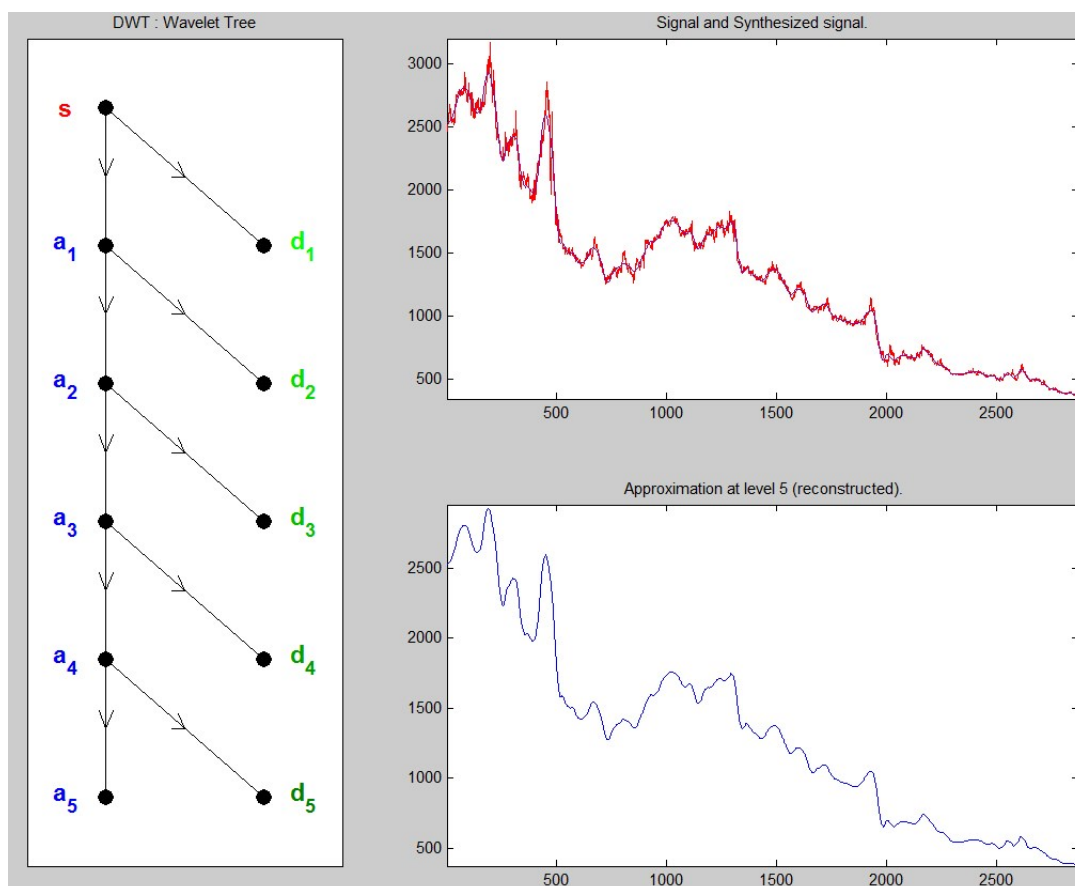


Рисунок 1.3



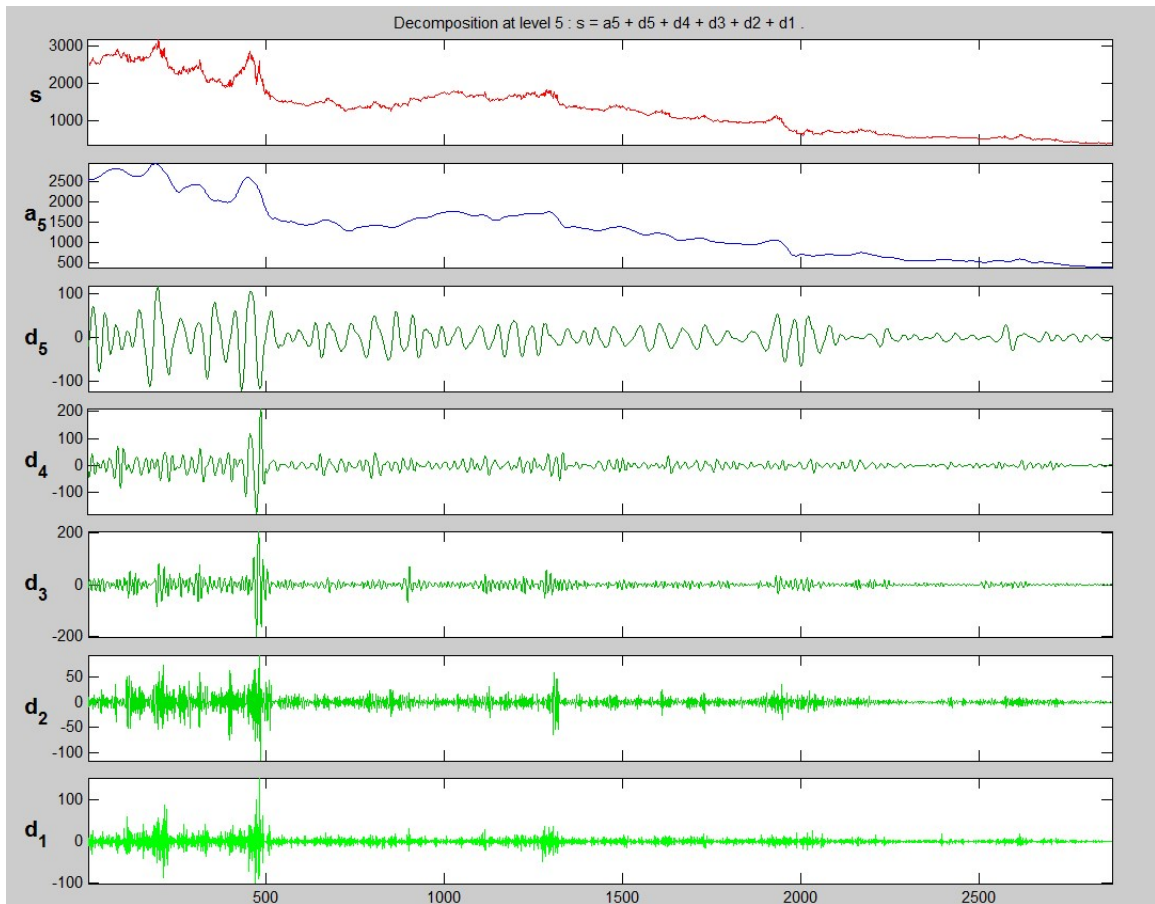


Рисунок 1.4

Полученные при разложении аппроксимирующие коэффициенты представляют сглаженный сигнал, а детализирующие коэффициенты описывают колебания. Следовательно, шумовая компонента больше отражается в детализирующих коэффициентах ( $c_{D_N}$ ), из-за этого при удалении шума обрабатывают обычно детализирующие коэффициенты. Удаление шума реализуется с помощью метода пороговой обработки коэффициентов (трешолдинг) и заключается в обнулении значений коэффициентов, меньших некоторого порогового значения. Полученный сигнал имеет шумовую компоненту, поэтому сначала вычислим порог для удаления шума, удалим этот шум. Для удаления шума используем вейвлет db4, уровень разложения 5 и минимаксный порог 2.771. Исходный и очищенный сигнал изображены на рис. 1.5. Красной сплошной линией изображен график исходного сигнала, черной жирной линией изображен график очищенного от шума сигнала.

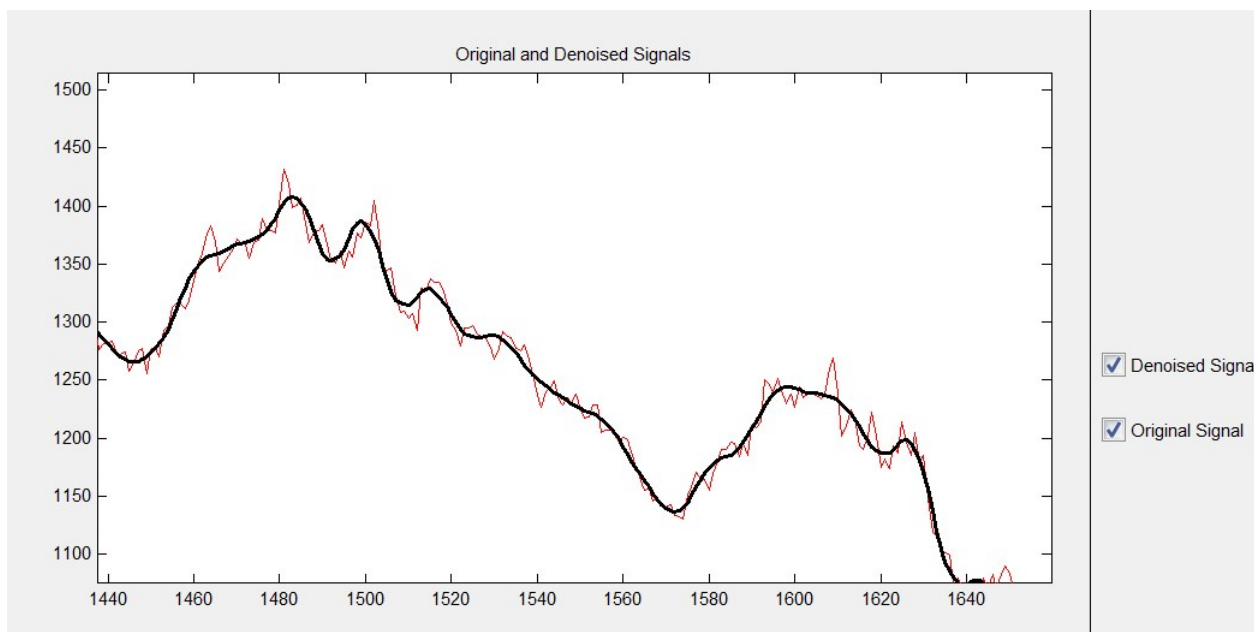


Рисунок 1.5

В результирующем сигнале удалено 87.4573% коэффициентов. Это позволяет эффективно сжимать такие сигналы для их хранения. При этом в очищенном сигнале колебания цены видны более явно, хорошо отражена их форма волн, что позволяет четче идентифицировать фигуры технического анализа на графике. Очищенные от шума паттерны фигур удобно использовать для прогнозирования тренда. Можно также получить всю удаленную шумовую компоненту в виде разности исходного сигнала и очищенного. Кроме подавления шума, вейвлет-анализ позволяет из непомерно большого числа входных переменных извлекать наиболее значимые для предсказания признаки, что является плюсом для предобработки больших массивов значений финансовых временных рядов (тот же дискретный сигнал).

## ЗАКЛЮЧЕНИЕ

В ходе практики был использован пакет расширения Matlab - Wavelet Toolbox с целью закрепления навыков анализа, прогнозирования временных рядов, а также, показана сравнительная простота использования уже готовых функциональных пакетов, что сокращает время первичной обработки огромного объема информации. При выполнении заданий были использованы источники [8],[9].

В представленной работе был проведен анализ финансовых временных рядов на основе вейвлет-преобразования. В частности рассмотрена возможность подавления шумовой составляющей и флуктуаций в рядах, извлечение наиболее значимых для предсказания признаков. На основе данной работы можно сделать вывод об эффективности применения вейвлет-анализа в качестве математического аппарата для обработки финансовых данных.

В каждой реальной задаче присутствует шум, и необходимо уметь справляться с ним. В особенности это относится к задачам обработки временных рядов, в которых переменные получены в результате измерений в некоторой физической системе, причем в самой системе и/или в механизме измерений шум присутствует естественным образом. В финансовых приложениях данные зашумлены особенно сильно. Пропуск значений или неполную информацию также иногда рассматривают как шум: в таких случаях берется среднее или наилучшее значение, и это приводит к зашумлению базы данных. Поэтому для успешного прогнозирования необходима эффективная обработка входных данных, в частности, минимизация случайных флуктуаций и шума. Понизить шумовую составляющую можно, прибегнув к вейвлет-анализу, что показано в нашей работе.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Айвазян С.А Мхитарян, В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.
2. Бородич, С.А. Вводный курс эконометрики. Учебное пособие. - Мн.: БГУ, 2000.
3. Доугерти, К. Введение в эконометрику: учебник. 2-е изд. М.: ИН-ФРА-М, 2004.
4. Добеши И., Десять лекций по вейвлетам, Ижевск: НИЦ "Регулярная и хаотическая динамика 2001.
5. Малла С. Вейвлеты в обработке сигналов – М.: Мир, 2005.
6. Чуи К. Введение в вейвлеты – М.: Мир, 2001.
7. Новиков Л.В. Основы вейвлет-анализа сигналов – Спб: МОДУС+, 1999.
8. К.Блаттер, Вейвлет-анализ. Основы теории, перевод с немецкого Т.Э. Кренкеля под редакцией А.Г.Кюркчана, ТЕХНОСФЕРА Москва, 2004.
9. Смоленцев, Н.К. Основы теории вейвлетов. Вейвлеты в Matlab. ДМК Москва, 2001.