

Санкт-Петербургский государственный университет

На правах рукописи

Александров Фёдор Игоревич

Разработка программного комплекса автоматического  
выделения и прогноза аддитивных компонент  
временных рядов в рамках подхода “Гусеница”-SSA

Специальность 05.13.18 — математическое моделирование,  
численные методы и комплексы программ

Диссертация на соискание ученой степени  
кандидата физико-математических наук

Научный руководитель:  
доктор физ.-мат. наук,  
профессор Сергей Михайлович Ермаков

Санкт-Петербург

2006 г.

# Содержание

|  |           |
|--|-----------|
| Введение   | 4         |
| <b>Глава 1. Подход “Гусеница”-SSA для анализа и прогноза временных рядов</b>                               | <b>8</b>  |
| 1.1 Базовый алгоритм   | 10        |
| 1.1.1 Разложение   | 10        |
| 1.1.2 Восстановление   | 11        |
| 1.1.3 Комментарии к алгоритму  | 13        |
| 1.1.4 Выбор параметра $L$ — длины окна   | 14        |
| 1.2 Разделимость рядов   | 15        |
| 1.2.1 Приближённая и асимптотическая разделимость  | 16        |
| 1.3 Ряды конечного ранга   | 17        |
| 1.3.1 Примеры рядов конечного ранга  | 18        |
| 1.4 Прогноз аддитивной составляющей  | 21        |
| 1.4.1 Вычисление коэффициентов ЛРФ порядка $L - 1$   | 21        |
| 1.4.2 Минимизация линейной рекуррентной формулы  | 23        |
| <b>Глава 2. Автоматический метод выделения тренда</b>  | <b>25</b> |
| 2.1 Вводные данные   | 25        |
| 2.2 Описание метода низких частот для идентификации трендовых компонент                                    | 27        |
| 2.3 Выбор параметра $\omega_0$   | 28        |
| 2.4 Проверка метода для модели с известной трендовой составляющей  | 34        |
| 2.4.1 Расчёт ошибки АИ при наилучшем в среднем $\mathcal{C}_0$ с помощью моделирования                     | 36        |
| 2.4.2 Поведение автоматической процедуры с $\mathcal{C}_0^{opt}$ при изменении параметров $\alpha, \sigma$ | 39        |
| 2.4.3 Зависимость ошибки от $\mathcal{C}_0$  | 41        |
| 2.4.4 Общие соображения по выбору $\mathcal{C}_0$  | 42        |
| 2.5 Оценка качества выделения тренда   | 43        |
| 2.5.1 Требования к мере качества выделения тренда  | 43        |
| 2.5.2 $\mathcal{R}$ -мера качества выделения тренда  | 44        |
| 2.6 Выбор порогового значения $\mathcal{C}_0$ на основе меры $\mathcal{R}$                                 | 45        |

|        |   |    |
|--------|---|----|
| 2.6.1  | $C_{max}$ — пороговое значение, при котором идентифицируются все трендовые собственные тройки . . . . . | 45 |
| 2.6.2  | Вычисление $C_{max}$ с помощью меры $\mathcal{R}$ . . . . .   | 46 |
| 2.6.3  | Примеры поиска $C_{max}$ с помощью меры $\mathcal{R}$ . . . . .   | 48 |
| 2.6.4  | Сравнение $C_0^{(T)}$ , $C_{max}$ и $C_0^{opt}$ для экспоненциального тренда                            | 52 |
| 2.6.5  | Описание процедуры TRRMEAS автоматического выделения тренда ряда . . . . .                              | 55 |
| 2.7    | Случай известной модели шума . . . . .  | 56 |
| 2.8    | Примеры выделения трендов реальных рядов . . . . .  | 61 |
| 2.8.1  | Исследование уровня экспрессии гена . . . . .   | 61 |
| 2.8.2  | Выделение тренда различной детализации из данных по уровню безработицы . . . . .                        | 62 |
| 2.8.3  | Сравнение процедуры TRRMEAS с другими методами выделения тренда . . . . .                               | 64 |
| 2.9    | Применение процедуры автоматического выделения тренда для обработки множества рядов . . . . .           | 66 |
| 2.9.1  | Проблема проверки качества процедуры TRRMEAS . . . . .  | 66 |
| 2.9.2  | Математическая постановка задачи, оценка ошибки автоматической процедуры и её свойства . . . . .        | 67 |
| 2.9.3  | Схема применения TRRMEAS к множеству рядов . . . . .  | 71 |
| 2.10   | Пример применения процедуры TRRMEAS к множеству рядов . . . . .   | 73 |
| 2.11   | Прогноз тренда . . . . .  | 76 |
| 2.11.1 | Схема прогноза тренда ряда . . . . .  | 76 |
| 2.11.2 | Проблема выбора порогового значения при прогнозе . . . . .  | 76 |
| 2.11.3 | Моделирование прогноза экспоненциального тренда . . . . .   | 77 |

### **Глава 3. Автоматический метод выделения периодической составляющей** **81**

|       |  |    |
|-------|--|----|
| 3.1   | Отличие задачи выделения тренда от задачи выделения периодической составляющей . . . . .   | 82 |
| 3.2   | Описание метода Фурье для идентификации $\varepsilon$ -м гармонических компонент . . . . . | 82 |
| 3.2.1 | Первая часть метода Фурье . . . . .  | 83 |
| 3.2.2 | Вторая часть метода Фурье . . . . .  | 88 |
| 3.3   | Проверка процедуры PER для модели с известной периодической составляющей . . . . .         | 98 |

|  |  |            |
|--|--|------------|
| 3.3.1  | Расчёт ошибки АИ при наилучшем в среднем $\rho_0$ с помощью моделирования . . . . .                              | 99         |
| 3.4  | Подходы к выбору порогового значения $\rho_0$ . . . . .  | 107        |
| 3.4.1  | Аналитическое вычисление $\rho_0$ для известного $\alpha$ и при $\omega$ :<br>$L\omega \in \mathbb{N}$ . . . . . | 108        |
| 3.5  | Эмпирический подход к выбору $\rho_0$ . . . . .  | 110        |
| 3.5.1  | Результаты численного исследования . . . . .   | 113        |
| 3.5.2  | Выделение гармоник в присутствии тренда . . . . .  | 118        |
| 3.6  | Оценка частоты выделенной $\varepsilon$ -м гармоник . . . . .  | 121        |
| 3.7  | Пример выделения периодической составляющей реального ряда . . . . .   | 122        |
| 3.8  | Прогноз периодических компонент . . . . .  | 125        |
| 3.8.1  | Моделирование прогноза $\varepsilon$ -м гармоник . . . . .   | 125        |
| <b>Глава 4. Оценка коэффициентов линейной рекуррентной формулы порядка 2</b> |  | <b>134</b> |
| 4.1  | Методы оценки коэффициентов ЛРФ . . . . .  | 135        |
| 4.1.1  | Метод, основанный на подходе “Гусеница”-SSA . . . . .  | 136        |
| 4.1.2  | Регрессионный метод . . . . .  | 137        |
| 4.2  | Сравнение методов с помощью моделирования . . . . .  | 138        |
| 4.2.1  | Значения параметров . . . . .  | 138        |
| 4.2.2  | Результаты . . . . .   | 139        |
| <b>Заключение</b>  |  | <b>143</b> |
| <b>Библиография</b>  |  | <b>145</b> |

# Введение

Одномерным вещественнозначным *временным рядом*  $F$  длины  $N$  ( $N > 2$ ) будем называть упорядоченную последовательность вещественных чисел  $F = (f_0, \dots, f_{N-1})$ ,  $f_n \in \mathbb{R}$ . Временной ряд чаще всего является результатом последовательных замеров некоторого показателя в равноотстоящие моменты времени, поэтому естественной является следующая интерпретация его элементов:  $f_n = f(t_n)$ , где  $f(t)$  — некоторая функция, а  $t_n$  соответствует времени регистрации измерений, хотя это может быть и функция от другого физического параметра, например пространственной координаты [7, 9, 16, 41].

Будем говорить, что ряд  $F$  состоит из суммы *аддитивных составляющих*  $F^{(1)}$ ,  $F^{(2)}$  и обозначать это как  $F = F^{(1)} + F^{(2)}$ , если

$$f_n = f_n^{(1)} + f_n^{(2)}, \quad 0 \leq n \leq N - 1.$$

Проблема выделения некоторой аддитивной составляющей ряда — одна из самых общих и часто встречающихся проблем анализа временных рядов.

Среди задач, связанных с выделением аддитивной составляющей, можно назвать: определение глобального поведения ряда (выделение тренда), выделение периодической составляющей или её удаление из ряда (например, так называемое *seasonal adjustment* [36, 48]), сглаживание ряда (выделение низкочастотной составляющей), отделение детерминированной составляющей ряда от шума (выделение сигнала).

Будем называть *периодической составляющей* с периодом  $T$  ( $T \geq 2$ ) ряд,  $n$ -ый элемент которого задаётся следующим выражением:

$$f_n^{(P)} = \sum_{k=1}^{\lfloor T/2 \rfloor} A_k e^{\alpha_k n} \cos(2\pi nk/T + \phi_k), \quad (0.1)$$

где  $A_k, \alpha_k \in \mathbb{R}$ ,  $A_k > 0$ , а  $\phi_k \in [0, 2\pi)$ .

Существуют много подходов к определению тренда [56]. Иногда под трендом понимают некоторую параметрическую составляющую ряда, на которую накладываются определённые ограничения, например гладкость или, для полиномиального тренда — его порядок [7, глава 3]. При стохастическом подходе к исследованию временных рядов тренд понимается как нестационарная медленно меняющаяся аддитивная составляющая [52]. В монографии [32] трендом называется медленное изменение среднего уровня<sup>1</sup>.

<sup>1</sup>В оригинале — long term change in the mean level

Будем считать *трендом* медленно меняющуюся составляющую, описывающую глобальное поведение ряда. Для определённости будем полагать, что она не является периодической. При таком подходе тренд определяется неоднозначно, в главе 2 приводятся примеры выделения трендов одного ряда, которые отличаются детализацией, или, как говорят, *разрешением*.

Будем предполагать, что ряд содержит следующие аддитивные составляющие: тренд  $F^{(T)}$ , периодическую составляющую  $F^{(P)}$  и шум случайной природы  $F^{(\varepsilon)}$ . Тогда в общем случае рассматриваемая модель временного ряда имеет вид:

$$F = F^{(T)} + F^{(P)} + F^{(\varepsilon)}, \quad \text{т. е.} \quad f_n = f_n^{(T)} + f_n^{(P)} + f_n^{(\varepsilon)},$$

где  $(f_0^{(\varepsilon)}, \dots, f_{N-1}^{(\varepsilon)})$  — реализация последовательности случайных величин.

Основная задача данной работы — это выделение из известного ряда  $F$  неизвестных составляющих  $F^{(T)}$  и  $F^{(P)}$ . Будем использовать для этого общий подход к исследованию временных рядов “Гусеница”-SSA (см. главу 1). В рамках этого подхода уже существуют методы для решения данных задач, однако они могут быть применены только в интерактивном режиме, так как строятся на визуальном исследовании информации, предоставляемой о ряде.

Целью работы будет разработка и исследование методов выделения и прогноза тренда и периодических составляющих ряда на базе подхода “Гусеница”-SSA, которые позволяют решать поставленные задачи не только интерактивно, но и при пакетной обработке данных. Для решения этой задачи выясним, как построены интерактивные методы выделения аддитивных составляющих [5, 10, 41], и попробуем автоматизировать процедуру визуального исследования.

Кроме того, исследуем свойства метода “Гусеница”-SSA при оценке коэффициентов линейной рекуррентной формулы, управляющей сигналом ряда.

**Первая глава** содержит информацию о методе “Гусеница”-SSA, которая потребуется для решения поставленных задач. В ней описываются уже существующие интерактивные способы решения задач выделения тренда и периодических составляющих в рамках этого подхода, а также рассматриваются условия, при которых возможно решение этих задач.

Во **второй главе** предлагается и исследуется метод автоматического выделения тренда. Проводится статистическое исследование качества работы метода и рассматривается способ выбора его параметров. Построена математическая модель применения метода выделения тренда к множеству рядов. Представлены примеры обработки временных и пространственных рядов.

**Третья глава** посвящена автоматическому методу выделения периодической составляющей ряда. Как и в главе 2, проводится статистическое исследование качества работы метода и предлагаются способы выбора его параметров. Приводятся примеры исследования реальных временных рядов.

**Четвертая глава** посвящена задаче оценки коэффициентов линейной рекуррентной формулы сигнала ряда  $F$  для случая  $F = F^{(P)} + F^{(\varepsilon)}$ . В ней рассматривается задача оценки коэффициентов ЛРФ порядка два для периодического сигнала. Предлагаются два метода её решения: один — регрессионный, другой — в рамках подхода “Гусеница”-SSA. Выполнено статистическое исследование по сравнению их качества.

Автор приносит искреннюю благодарность своему научному руководителю С.М. Ермакову, а также пользуется случаем выразить признательность Н.Э. Голяндиной за бесценную помощь, оказанную при работе над диссертацией и в процессе написания совместных работ.



# Глава 1

## Подход “Гусеница”-SSA для анализа и прогноза временных рядов

Подход “Гусеница”-SSA зародился в 80-х–90-х годах в различных областях как теоретической, так и прикладной науки. Аббревиатура SSA расшифровывается как *Singular Spectrum Analysis*. Этот подход может быть использован для решения разнообразных задач, связанных с исследованием одномерных временных рядов. Существуют его обобщения для двумерных временных рядов [61], а также для случайных полей [15], но в данной работе ограничимся задачами анализа и прогноза только одномерных временных рядов.

В основе подхода “Гусеница”-SSA лежит представление временного ряда в виде матрицы, которая затем раскладывается в сумму матриц с помощью сингулярного разложения<sup>1</sup> [62], каждой из которых сопоставляется аддитивная составляющая исходного временного ряда. Таким образом, получаем разложение ряда в сумму составляющих, причём информация о каждой из них содержится в сингулярных числах и векторах.

Подход независимо появился в России под названием “Гусеница” и в других странах, под разными названиями — Singular Spectrum Analysis, Eigen Decomposition. Первые упоминания о нём относятся к 60-м – 70-м годам прошлого века. Санкт-Петербургский государственный университет (в то время Ленинградский государственный университет) с самого начала принимал активное участие в его разработке, большой вклад в развитие метода внесли сотрудники кафедры статистического моделирования СПбГУ под руководством С.М. Ермакова.

К ранним упоминаниям о методе “Гусеница” относятся работы О.М. Калинина, М.Д. Белониной [8], М.М. Кислицина [18]. В 1997 году вышел сборник [15], содержащий статьи о истории развития метода, описание базового алгоритма, теоретические вопросы, а также примеры применения. В нём были опубликованы работы Д.Л. Данилова, А.А. Жиглявского, В.Н. Солнцева, Н.Э. Голяндиной, В.В. Некруткина, М.М. Кислицина. Стоит отдельно отметить статьи В.М. Бухштабера [1, 30]. В 2001 г. на английском языке вышла монография N. Golyandina, V. Nekrutkin, A. Zhigljavsky “Analysis of Time Series Structure: SSA and Related Techniques” [41], которая на сегодняшний день является ос-

<sup>1</sup>В англоязычной литературе — *Singular Value Decomposition*, или *SVD*

новным трудом по подходу “Гусеница”. В настоящее время ведётся работа над обобщением подхода для двумерных рядов [12, 21, 61], а также для рядов с пропусками [13, 42]. Существует программный комплекс, который позволяет интерактивно решать задачи анализа и прогноза временных рядов с помощью подхода “Гусеница” [64].

В других странах подход получил название SSA (Singular Spectrum Analysis), которое претерпевало изменения (например, в [29] используется название Singular System Analysis) и закрепилось с работами [67, 68]. История развития подхода SSA описана в [40]. Его идея возникла из применения метода главных компонент к анализу цифровых сигналов [49], одновременно появившись в прикладных исследованиях, например, в океанологии [33]. Ещё одной областью, в которой был сформулирован метод SSA, была теория динамических систем.

Среди ранних работ, посвящённых основам SSA, стоит отметить статьи [54, 63, 69]. В более поздних работах [28, 55] подход получает название Singular System Analysis. Работы [29, 38] по нахождению аттракторов нелинейных динамических систем и оценке их размерности внесли большой вклад в его развитие. Позже, в конце 80-х гг., вышла статья [67], в которой подход используется для улучшения отношения сигнал/шум и сжатия данных. С этого времени за ним закрепилось название Singular Spectrum Analysis.

Существует много модификаций SSA: Nonlinear Multichannel SSA [45], Random-lag SSA [65], Random-lag Cross-spectrum SSA [66], Monte Carlo SSA [24, 25], SSA для рядов с пропусками [47, 58]. На интернет-сайте [60] доступен программный пакет для применения SSA к задачам анализа климатических временных рядов, см. [68].

В настоящее время метод используется в различных областях прикладных исследований: для исследования интернет-траффика [26], в биометрике [23, 39], для скрытой передачи данных<sup>2</sup> [53]. По-прежнему актуально его применение в метеорологии [27] и геологии [57].

В диссертации в качестве основного источника сведений используется монография, посвящённая подходу “Гусеница” [41], и учебные пособия [10, 11], доступные на интернет-сайте [64]. Также разрабатываются идеи, представленные в статье о SSA [68]. Поэтому далее будем использовать название “Гусеница”-SSA, чтобы подчеркнуть общий характер развития подходов “Гусеница” и SSA.

---

<sup>2</sup>В англоязычной литературе — watermarking

## 1.1 Базовый алгоритм

Основой метода “Гусеница”-SSA является алгоритм, который применяется к некоторой числовой последовательности — временному ряду. Алгоритм состоит из четырёх шагов: вложение, сингулярное разложение, группировка и диагональное усреднение. Первые два в совокупности называются *разложением*, последние — *восстановлением*. Алгоритм раскладывает временной ряд на аддитивные компоненты и предоставляет информацию, на основе которой среди них можно найти те, которые в сумме дают некоторую искомую составляющую ряда (например, тренд или периодическую составляющую).

### 1.1.1 Разложение

Пусть  $F = (f_0, \dots, f_{N-1})$  — одномерный вещественнозначный временной ряд длины  $N$ ,  $N \geq 3$ . Выберем значение параметра  $L$ ,  $L \in \mathbb{N}$ ,  $1 < L < N$ , который называется *длиной окна*<sup>3</sup>. Результат работы алгоритма во многом зависит от значения  $L$ , вопрос его выбора рассматривается позже, в разделе 1.1.4.

Построим по ряду  $F$  матрицу  $\mathbf{X} \in \mathbb{R}^{L \times K}$ ,  $K = N - L + 1$ , называемую *L-траекторной* (или просто *траекторной*) матрицей ряда:

$$\mathbf{X} = \begin{bmatrix} f_0 & f_1 & \dots & f_{N-L} \\ f_1 & f_2 & \dots & f_{N-L+1} \\ \vdots & \ddots & \ddots & \vdots \\ f_{L-1} & f_L & \dots & f_{N-1} \end{bmatrix}. \quad (1.1)$$

Заметим, что так как  $x_{ij} = f_{i+j-2}$ , то матрица  $\mathbf{X} = \{x_{ij}\}_{i,j=1}^{i=L,j=K}$  — *ганкелева*, т. е. для любых индексов  $i, j$ ,  $1 \leq i \leq L$ ,  $1 \leq j \leq K$ , таких, что  $i + j = k$ ,  $x_{ij} = c_k$ .

Обозначим  $\mathbf{S} = \mathbf{X}\mathbf{X}^T \in \mathbb{R}^{L \times L}$ . Матрица  $\mathbf{X}\mathbf{X}^T$  симметричная и неотрицательно определённая, а значит её собственные числа  $\{\lambda_k\}_{k=1}^L$  вещественны и неотрицательны. Запишем их следующим образом:  $\lambda_1 \geq \dots \geq \lambda_L \geq 0$ . Пусть  $U_1, \dots, U_L$  — соответствующие им ортонормированные собственные вектора. Будем называть  $d = \max\{k : \lambda_k > 0\}$  *порядком сингулярного разложения*. Обозначим

$$V_k = \frac{1}{\sqrt{\lambda_k}} \mathbf{X}^T U_k, \quad k = 1, \dots, d.$$

<sup>3</sup>В ранних источниках, например, в [15], этот параметр называется ещё “*длиной гусеницы*”

Тогда *сингулярным разложением* матрицы  $\mathbf{X}$  называется её представление в виде суммы

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d, \quad \mathbf{X}_k = \sqrt{\lambda_k} U_k V_k^T. \quad (1.2)$$

Вектор  $U_k$  называют *k-м левым сингулярным вектором* или просто *k-м собственным вектором*,  $V_k$  — *k-м правым сингулярным вектором* или *k-м факторным вектором*<sup>4</sup>. Набор  $(\sqrt{\lambda_k}, U_k, V_k)$  будем называть *k-ой собственной тройкой*, которая иногда обозначается как ET, от английского *eigentriple*. Интересующие нас свойства сингулярного разложения описаны в [41, глава 4].

### 1.1.2 Восстановление

Разобьём множество собственных троек с номерами  $\{1, \dots, d\}$  на некоторое количество непересекающихся подмножеств,  $\mathcal{I}_j: \{1, \dots, d\} = \bigcup_{j=1}^m \mathcal{I}_j$ . Обозначим  $\mathbf{X}_{\mathcal{I}_j} = \sum_{k \in \mathcal{I}_j} \mathbf{X}_k$ . С помощью такой группировки собственных троек получаем разложение матрицы  $\mathbf{X}$ :

$$\mathbf{X} = \mathbf{X}_{\mathcal{I}_1} + \mathbf{X}_{\mathcal{I}_2} + \dots + \mathbf{X}_{\mathcal{I}_m}.$$

Построим теперь по каждой из матриц  $\mathbf{X}_{\mathcal{I}_j}$  ряд  $F^{(j)}$  так, чтобы получить разложение исходного ряда  $F = \sum_j F^{(j)}$ . В случае, когда матрица  $\mathbf{X}_{\mathcal{I}_j}$  — ганкелева, это можно сделать процедурой, обратной к построению траекторной матрицы. Но полученная  $\mathbf{X}_{\mathcal{I}_j}$  не всегда ганкелева (см. раздел 1.2) и поэтому необходимо предварительно провести усреднение матрицы по её диагоналям вида  $\{x_{pq} : p + q = \text{const}\}$ .

Для произвольной матрицы  $\mathbf{Y}$  процедура приведения её к ганкелевому виду и последующее преобразование в ряд (обозначим его как  $G$ ) выглядит следующим образом. Пусть матрица  $\mathbf{Y} \in \mathbb{R}^{L \times K}$ ,  $\mathbf{Y} = \{y_{ij}\}_{i,j=1}^{i=L, j=K}$ . Введём обозначения:  $\tilde{L} = \min\{L, K\}$ ,  $\tilde{K} = \max\{L, K\}$  и

$$\tilde{y}_{ij} = \begin{cases} y_{ij}, & \text{если } L \leq K, \\ y_{ji}, & \text{если } L > K. \end{cases}$$

<sup>4</sup>Эта терминология пришла из метода главных компонент

Элементы ряда  $G = (g_0, \dots, g_{N-1})$ ,  $N = L + K - 1$ , выражаются следующим образом:

$$g_n = \begin{cases} \frac{1}{n+1} \sum_{m=1}^{n+1} \tilde{y}_{m,n-m+2}, & 0 \leq n < \tilde{L} - 1, \\ \frac{1}{\tilde{L}} \sum_{m=1}^{\tilde{L}} \tilde{y}_{m,n-m+2}, & \tilde{L} - 1 \leq n < \tilde{K}, \\ \frac{1}{N-n} \sum_{m=n-\tilde{K}+2}^{N-\tilde{K}+1} \tilde{y}_{m,n-m+2}, & \tilde{K} \leq n < N. \end{cases} \quad (1.3)$$

Таким образом, результатом этапа восстановления, а значит и результатом работы всего алгоритма “Гусеница”-SSA, является разложение исходного ряда  $F$  в сумму аддитивных составляющих

$$F = F^{(1)} + \dots + F^{(m)}, \quad (1.4)$$

где, напомним,  $m$  — это количество  $\mathcal{I}_j$ , которые формируют разбиение множества всех собственных троек  $\{1, \dots, d\}$ .

**АЛГОРИТМ 1.1. Процедура SSADEC: сингулярное разложения ряда**

**Входные данные:**  $F$  — ряд длины  $N$  ( $N \geq 3$ )

**Результат:**  $\{(\sqrt{\lambda_k}, U_k, V_k)\}_{k=1}^d$  — собственные тройки

**Параметры:**  $L$  — длина окна ( $1 < L < N$ )

**Процедура:**

1. Построим по  $F$  траекторную матрицу  $\mathbf{X} \in \mathbb{R}^{L \times K}$  по (1.1)
2. Посчитаем собственные числа  $\{\lambda_k\}_{k=1}^L$ , собственные  $\{U_k\}_{k=1}^L$  и факторные  $\{V_k\}_{k=1}^L$  вектора матрицы  $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ , упорядочив их так, чтобы  $\lambda_1 \geq \dots \geq \lambda_L \geq 0$
3.  $d = \max\{k : \lambda_k > 0\}$

**АЛГОРИТМ 1.2. Процедура SSAREC: восстановление составляющей ряда**

**Входные данные:**  $\{(\sqrt{\lambda_k}, U_k, V_k)\}_{k=1}^d$  — собственные тройки

**Результат:**  $G$  — составляющая ряда  $F$

**Параметры:**  $\mathcal{I}$  — номера собственных троек для восстановления

**Процедура:**

1. Посчитаем  $\mathbf{X}_{\mathcal{I}} = \sum_{k \in \mathcal{I}} \mathbf{X}_k$ ,  $\mathbf{X}_k = \sqrt{\lambda_k} U_k V_k^T$
2. Восстановим по  $\mathbf{X}_{\mathcal{I}}$  ряд  $G$  пользуясь (1.3)

### 1.1.3 Комментарии к алгоритму

Траекторная матрица ряда состоит из столбцов, представляющих собой отрезки ряда длиной  $L$ , взятые последовательно с единичным сдвигом. Именно способу построения траекторной матрицы метод обязан названием “Гусеница”. В работах [29, 69] предлагается также брать сдвиг, отличный от единицы, но это связано с инженерными вопросами цифрового представления реальных данных и к описанию метода отношения не имеет. Тем не менее, эта идея нашла воплощение в модификации метода Random-lag SSA [65, 66], в которой предлагается рассматривать отрезки ряда, взятые со случайной задержкой с тем, чтобы уменьшить количество  $K$  столбцов матрицы  $\mathbf{X}$  и тем самым ускорить дальнейшие вычисления.

Что касается матрицы  $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ , то в случае, когда ряд  $F$  — стационарный,  $\mathbf{S}$  представляет собой оценку ковариационной матрицы. Существуют другие способы выбора  $\mathbf{S}$ , см. [35], [40, раздел 2.2], [41, раздел 1.7.2].

Видно, что алгоритм “Гусеница”-SSA после формирования  $\mathbf{X}$  похож на алгоритм *метода главных компонент* (*Principal Component Analysis*). Собственные числа имеют здесь ту же интерпретацию, что и в PCA. Неформально говоря, отношение

$$\frac{\sum_{i \in \mathcal{I}} \lambda_i}{\sum_{i=1}^d \lambda_i}, \quad \mathcal{I} \subset \{1, \dots, d\}$$

отражает вклад составляющей ряда, восстановленной по группе собственных троек  $\mathcal{I}$ , в вид исходного ряда, см. [41, раздел 4.3]. Отсюда понятна простая идея сглаживания ряда — восстановление ряда по первым компонентам [51].

Алгоритм метода управляется, во-первых, заданием параметра  $L$ , а во-вторых, выбором  $\mathcal{I}$  на этапе группировки. Процесс выбора  $\mathcal{I}$  при решении зада-

чи выделения некоторой аддитивной составляющей назовём *идентификацией*. Как уже было сказано, существуют интерактивные подходы к идентификации собственных троек, соответствующих тренду или периодической составляющей. Основной целью данной работы является разработка методов неинтерактивной идентификации.

#### 1.1.4 Выбор параметра $L$ — длины окна

Свойства сингулярного разложения приводят к тому, что выбор длины окна  $L' = K = N - L + 1$  порождает факторные вектора, равные собственным векторам разложения с длиной окна  $L$ . Вследствие этого будем выбирать  $L$  из диапазона  $2 \leq L \leq \lceil N/2 \rceil$ .

До последнего времени не существовало математически обоснованного решения проблемы выбора  $L$ , и различные источники предлагали лишь выработанные на опыте рекомендации. Broomhead и King в [29, раздел 3.4] дают рекомендации по выбору  $L$  для рядов ограниченного спектра<sup>5</sup>, но отмечают, что “с теоретической точки зрения эта проблема всё ещё открыта”<sup>6</sup>. Vautard развил их идеи в [68, 2.2], но лишь на общем уровне. Он отмечает сложность данной проблемы и косвенно подтверждает это, делая ошибочное утверждение, что метод не позволяет выделять периодические составляющие с длиной больше  $L$ . Что это утверждение — ошибочно, будет видно впоследствии.

Математическая теория, которая позволила дать ответ на вопрос выбора длины окна при решении разнообразных задач, была предложена В.В. Некруткиным в [20]. Обобщение метода “Гусеница”-SSA на языке функционального анализа позволило вывести аналогии между разделением суммы рядов  $F^{(1)} + F^{(2)}$  и ортогональностью их траекторных пространств<sup>7</sup>  $\mathcal{L}_L^{(1)}$ ,  $\mathcal{L}_L^{(2)}$ . Возможность разделения суммы рядов  $F = F^{(1)} + F^{(2)}$  — необходимое условие выделения аддитивной составляющей  $F^{(j)}$  ( $j = 1, 2$ ). В дальнейшем изложении покажем, что разделимость может существовать для разных рядов, и условия её выполнения накладывают ограничения на  $L$ . Суть этих ограничений во многом даёт ответ на вопрос выбора  $L$ . Соответствующая теория и примеры наиболее полно рассмотрены в [41, разделы 1.5, 6.1], см. также [10, раздел 2].

Вопрос выбора  $L$  хорошо, хоть и вкратце, освещён в [10, раздел 4.2], где сделан структурированный обзор решения проблемы для разных задач (выде-

<sup>5</sup>Коэффициенты разложения Фурье которых равны нулю вне определённого интервала

<sup>6</sup>В оригинале — “theoretically the problem is still open”

<sup>7</sup>Определение траекторного пространства см. дальше, в разделе 1.2

ление тренда, периодических составляющих) и с учётом различных аспектов (длина ряда, вид собственных чисел).

## 1.2 Разделимость рядов

Понятие разделимости рядов важно не только тем, что условия её выполнения дают рекомендации по выбору  $L$ . Разделимость аддитивных составляющих ряда влечёт за собой существование разложения и возможность группировки компонент сингулярного разложения для выделения любой из этих составляющих. Поэтому разделимость — это необходимое условие возможности точного выделения аддитивной составляющей.

Пусть  $F^{(1)}$  и  $F^{(2)}$  — ряды длины  $N$  и  $F = F^{(1)} + F^{(2)}$ . Зафиксируем длину окна  $L$  и построим по каждому из рядов траекторную матрицу:  $\mathbf{X}$ ,  $\mathbf{X}^{(1)}$ ,  $\mathbf{X}^{(2)}$ .

**Определение 1.1**  $L$ -траекторным пространством (или просто траекторным пространством) для ряда  $F$  будем называть линейное пространство, натянутое на вектора его траекторной матрицы  $\mathbf{X}$ :

$$\mathcal{L}_L(F) = \mathcal{L}_L = \text{span}(X_1, \dots, X_K), \quad \mathbf{X} = [X_1 : \dots : X_K].$$

Обозначим как  $\mathcal{L}_L^{(1)}$ ,  $\mathcal{L}_L^{(2)}$   $L$ -траекторные пространства рядов  $F^{(1)}$ ,  $F^{(2)}$ .

**Определение 1.2** Будем говорить, что ряды  $F^{(1)}$  и  $F^{(2)}$  слабо  $L$ -разделимы, если  $\mathcal{L}_L^{(1)} \perp \mathcal{L}_L^{(2)}$  и  $\mathcal{L}_K^{(1)} \perp \mathcal{L}_K^{(2)}$ .

**Утверждение 1.1** Пусть ряды  $F^{(1)}$  и  $F^{(2)}$  слабо  $L$ -разделимы. Рассмотрим некоторые их сингулярные разложения

$$\mathbf{X}^{(1)} = \sum_j \sqrt{\lambda^{(1)}} U_j^{(1)} V_j^{(1)\top}, \quad \mathbf{X}^{(2)} = \sum_j \sqrt{\lambda^{(2)}} U_j^{(2)} V_j^{(2)\top}.$$

Тогда представление матрицы  $\mathbf{X}$  в виде суммы

$$\mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)} = \sum_j \sqrt{\lambda^{(1)}} U_j^{(1)} V_j^{(1)\top} + \sum_j \sqrt{\lambda^{(2)}} U_j^{(2)} V_j^{(2)\top}$$

будет её сингулярным разложением (см. [10, Предложение 2.2]).

**Следствие 1.1** Пусть ряды  $F^{(1)}$  и  $F^{(2)}$  слабо  $L$ -разделимы и  $F = F^{(1)} + F^{(2)}$ . Тогда существует такое сингулярное разложение траекторной матрицы  $\mathbf{X}$  ряда  $F$ , что его можно разбить на две части, которые являются сингулярными разложениями траекторных матриц  $\mathbf{X}^{(1)}$ ,  $\mathbf{X}^{(2)}$ .

Из-за неоднозначности сингулярного разложения траекторной матрицы необходимо усилить понятие разделимости.



**Определение 1.3** Будем называть ряды  $F^{(1)}$  и  $F^{(2)}$  *сильно  $L$ -разделимыми*, если они слабо  $L$ -разделимы и множества собственных чисел их  $L$ -траекторных матриц не пересекаются.

**Утверждение 1.2** Пусть ряды  $F^{(1)}$  и  $F^{(2)}$  сильно  $L$ -разделимы и  $F = F^{(1)} + F^{(2)}$ . Тогда любое сингулярное разложение траекторной матрицы  $\mathbf{X}$  ряда  $F$  можно разбить на две части, которые являются сингулярными разложениями траекторных матриц  $\mathbf{X}^{(1)}$ ,  $\mathbf{X}^{(2)}$  (см. [10, Предложение 2.3]).

**Замечание 1.1** Если ряды  $F^{(1)}$  и  $F^{(2)}$  (сильно) разделимы,  $F = F^{(1)} + F^{(2)}$ ,  $\mathbf{X}$  —  $L$ -траекторная матрица  $F$ , а  $L$  выбрано так, чтобы были выполнены условия разделимости, то после соответствующей группировки матрицы  $\mathbf{X}_{\mathcal{I}_1}$ ,  $\mathbf{X}_{\mathcal{I}_2}$ :  $\mathbf{X} = \mathbf{X}_{\mathcal{I}_1} + \mathbf{X}_{\mathcal{I}_2}$  будут ганкелевыми.

Введём здесь определение ряда, который играет важную роль в применении подхода “Гусеница”-SSA.

**Определение 1.4** Ряд с  $n$ -ым членом, задаваемым уравнением

$$f_n = Ae^{\alpha n} \cos(2\pi\omega n + \phi), \quad A, \alpha \in \mathbb{R}, \quad \omega \in (0, 0.5], \quad \phi \in [0, 2\pi), \quad (1.5)$$

будем называть *экспоненциально-модулированным (э-м) гармоническим рядом* или просто *э-м гармоникой* с частотой  $\omega$  и периодом  $T = 1/\omega$ .

Как будет показано далее, по своим свойствам э-м гармоника в подходе “Гусеница”-SSA не отличается от немодулированной гармоника, именно поэтому периодическая составляющая (0.1) была определена, как сумма э-м гармоник.

**Примеры разделимости** Существуют условия разделимости гармоника от гармоника и гармоника от константы, а также э-м гармоника от э-м гармоника и э-м гармоника от экспоненты. Во всех этих случаях выполнение условий разделимости накладывает на  $L$  ограничения, связанные с периодами рассматриваемых гармоник (например, требование  $L\omega \in \mathbb{N}$  для гармоника с частотой  $\omega$ ). Подробное изложение см. в [2], [10, раздел 2.2], [41, раздел 6.1.1].

### 1.2.1 Приближённая и асимптотическая разделимость

Определённая выше (точная) разделимость накладывает сильные ограничения на составляющие ряда, поэтому была введена мера разделимости и с помощью неё определена приближённая разделимость, а также асимптотическая разделимость (при  $N, L \rightarrow \infty$ ), которые выполняются для гораздо более широкого класса составляющих. Достаточно заметить, что вместо упомянутых

четырёх случаев выполнения точной разделимости, асимптотическая разделимость выполняется для 12 сочетаний составляющих, среди которых полиномы, гармоники и э-м гармоники. Примеры асимптотической разделимости см. в [10, раздел 2.3], [41, раздел 6.1.1].

Для практического применения метода “Гусеница”-SSA, когда обрабатываемые данные содержат шум (как от ошибки наблюдений, так и другого рода), необходимым для применимости метода является возможность разделения детерминированного сигнала и шума. В этой области было получено два асимптотических результата. Во-первых, о разделимости стационарного сигнала определённого вида и стационарного шума [41, раздел 6.4.4], а во-вторых, о разделимости любого сигнала и гауссовского шума [41, раздел 6.1.3].

Эти результаты позволяют работать с рядами, содержащими в себе тренд, периодическую составляющую и шум<sup>8</sup>, так как при наличии разделимости собственные тройки, соответствующие разным аддитивным составляющим ряда, могут быть сгруппированы так, чтобы получить в результате применения требуемое разложение ряда.

### 1.3 Ряды конечного ранга

Если  $F = F^{(1)} + F^{(2)}$ , то разделимость  $F^{(1)}$  и  $F^{(2)}$  — это необходимое условие для возможности выделения  $F^{(1)}$ . Заметим, что если ряд  $F^{(1)}$  отделим от  $F^{(2)}$ , то его траекторная матрица — матрица неполного ранга. Опишем теперь класс рядов, которые могут быть выделены.

Рассмотрим первый шаг процедуры SSADEC — построение траекторной матрицы  $\mathbf{X}$  по ряду  $F$ . Обозначим  $k$ -ый столбец матрицы  $\mathbf{X}$  как  $X_k$ . Введём следующее определение, следуя [10, Определение 3.2].

**Определение 1.5** Если для ряда  $F$  существует некоторое  $d \geq 0$ , такое что при любом  $L: d \leq \min(L, K)$  выполняется равенство  $\dim \mathcal{L}_L(F) = d$ , то будем говорить, что *ряд  $F$  имеет ранг  $d$* , что обозначается как  $\text{rank}(F) = d$ . Если такое  $d$  существует, то ряд  $F$  будем называть *рядом конечного ранга*.

**Утверждение 1.3** Ранг  $F$  равен порядку сингулярного разложения траекторной матрицы  $\mathbf{X}$  (см. [10, Предложение 3.2]).

---

<sup>8</sup>Напомним, что под шумовой составляющей ряда понимается реализация случайной последовательности

Точно выделить составляющую  $F^{(1)}$  можно только тогда, когда она является рядом конечного ранга. Для того, чтобы описать класс рядов конечного ранга, приведём следующий факт (см. [10, Предложение 3.3]).

**Утверждение 1.4** Любой ряд, являющийся линейной комбинацией произведений полиномов, экспонент и косинусов, является рядом конечного ранга.

### 1.3.1 Примеры рядов конечного ранга

Приведём примеры рядов конечного ранга и вид соответствующих им собственных векторов, следуя [10, раздел 3.2], [41, раздел 5]. Эта информация будет использоваться при идентификации собственных троек, соответствующих тренду и периодическим составляющим.

**Экспоненциально-модулированный гармонический ряд** с  $0 < \omega < 0.5$ . Рассмотрим пример, когда  $F$  — э-м гармонический ряд (1.5) и  $0 < \omega < 0.5$ . Для любых  $L \geq 2$  и  $N \geq L + 1$  такой ряд имеет ранг 2. Оба собственных (и факторных) вектора имеют такую же форму с теми же  $\alpha$  и  $\omega$ .

В работе [2] автором было доказано следующее утверждение.

**Утверждение 1.5** Если  $\omega \in (0, 0.5)$ ,  $L\omega \in \mathbb{N}$ , то элементы собственных векторов  $U_1, U_2$ , соответствующих ряду (1.5), имеют вид

$$u_n^{(1)} = B_1 e^{\alpha n} \cos(2\pi\omega n + \tilde{\phi}_1), \quad u_n^{(2)} = B_2 e^{\alpha n} \cos(2\pi\omega n + \tilde{\phi}_2) \quad (1.6)$$

при  $1 \leq n \leq L$ , где  $\tilde{\phi}_j \in [0, 2\pi)$ . Причём если  $L \rightarrow \infty$ ,  $\alpha \rightarrow 0$ ,  $L\alpha \rightarrow \gamma \in \mathbb{R}$ , то

$$|\tilde{\phi}_1 - \tilde{\phi}_2| \rightarrow \pi/2. \quad (1.7)$$

Аналогичное утверждение верно для элементов факторных векторов, только требования накладываются на  $K$ , а не на  $L$ .

*Доказательство.* Известно, что собственные вектора  $U_1, U_2$  образуют ортонормированный базис траекторного пространства  $\mathcal{L}_L(F)$  [10, раздел 3.1], [41, раздел 1.2.2]. Отсюда следует, что их элементы представляются в виде (1.6).

Докажем теперь, что верно (1.7). Исходя из вида векторов  $U_1, U_2$ , запишем условие их ортогональности в следующем виде

$$\sum_{n=0}^{L-1} e^{2\alpha n} \cos(2\pi\omega n + \psi_1) \cos(2\pi\omega n + \psi_2) = 0,$$

где  $\psi_j = \tilde{\phi}_j + 2\pi\omega$ ,  $j = 1, 2$ , заменив нумерацию при суммировании для удобства доказательства. Будем доказывать теперь, что для  $\psi_1 = \psi_1(\alpha, L)$  и

$\psi_2 = \psi_2(\alpha, L)$  выполняется  $|\psi_1 - \psi_2| \rightarrow \pi/2$ . Очевидно, что это равносильно утверждению (1.7). Очевидным образом получаем

$$\sum_{n=0}^{L-1} e^{2\alpha n} \cos(4\pi\omega n + \psi_1 + \psi_2) + \cos(\psi_1 - \psi_2) \sum_{n=0}^{L-1} e^{2\alpha n} = 0, \quad (1.8)$$

Воспользуемся [14, формулы 1.353.1, 1.353.3] ( $p \in \mathbb{R}, x \in \mathbb{R}$ ):

$$\sum_{n=1}^{L-1} p^n \sin(nx) = \frac{p \sin x - p^L \sin(Lx) + p^{L+1} \sin((L-1)x)}{1 - 2p \cos x + p^2},$$

$$\sum_{n=0}^{L-1} p^n \cos(nx) = \frac{1 - p \cos x - p^L \cos(Lx) + p^{L+1} \cos((L-1)x)}{1 - 2p \cos x + p^2}.$$

При  $x: xL = 4\pi k, k \in \mathbb{N}$ , получаем:

$$\sum_{n=0}^{L-1} p^n \sin(nx) = \frac{p \sin x - p^{L+1} \sin x}{1 - 2p \cos x + p^2},$$

$$\sum_{n=0}^{L-1} p^n \cos(nx) = \frac{1 - p \cos x - p^L + p^{L+1} \cos x}{1 - 2p \cos x + p^2}.$$

Соответственно, для  $\psi \in [0, 2\pi)$ :

$$\sum_{n=0}^{L-1} p^n \cos(4\pi\omega n + \psi) = \frac{(1 - p^L)(\cos \psi - p \cos(4\pi\omega - \psi))}{1 - 2p \cos(4\pi\omega) + p^2}. \quad (1.9)$$

Возьмём  $p = e^{2\alpha}$ ,  $\psi = \psi_1 + \psi_2$ . Так как  $0 < \omega < 0.5$ , а, следовательно,  $\cos(4\pi\omega) < 1$ , то первое слагаемое суммы (1.8) имеет конечный предел при  $\alpha \rightarrow 0, L \rightarrow \infty, L\alpha \rightarrow \gamma$ :

$$\sum_{n=0}^{L-1} e^{2\alpha n} \cos(4\pi\omega n + \psi_1 + \psi_2) \xrightarrow[L\alpha \rightarrow \gamma]{\alpha \rightarrow 0, L \rightarrow \infty} \frac{(1 - e^{2\gamma})(\cos(\psi_1 + \psi_2) - \cos(4\pi\omega - \psi_1 - \psi_2))}{2 - 2 \cos(4\pi\omega)}.$$

Так как

$$\sum_{n=0}^{L-1} e^{2\alpha n} = \frac{e^{2\alpha L} - 1}{e^{2\alpha} - 1} \xrightarrow[L\alpha \rightarrow \gamma]{\alpha \rightarrow 0, L \rightarrow \infty} \infty,$$

то для того, чтобы равенство (1.8) выполнялось в пределе, необходимо, чтобы  $\cos(\psi_1 - \psi_2) \rightarrow 0$ , что равносильно (1.7). ■

**Замечание 1.2** Если  $\alpha = 0$ , то  $F$  принимает вид гармонике с частотой  $\omega$ . При  $L\omega \in \mathbb{N}$  ( $K\omega \in \mathbb{N}$ ) элементы собственных (факторных) векторов имеют вид  $B \cos(2\pi\omega n + \tilde{\phi})$ ,  $B \sin(2\pi\omega n + \tilde{\phi})$ . Оба собственных числа совпадают (собственные числа  $\varepsilon$ -м гармонического ряда с экспоненциальным показателем  $\alpha \leq 0$  асимптотически совпадают при  $L, K \rightarrow \infty$ ).

Из этих свойств собственных чисел можно заключить, что собственные тройки, соответствующие гармонике (а также  $\varepsilon$ -м гармонике — при больших  $L, K$ ), будут иметь номера вида  $k, k + 1$ .

**Замечание 1.3** Вид элементов собственных (факторных) векторов, соответствующих гармонике с частотой  $\omega$ , будет  $B \cos(2\pi\omega n + \tilde{\phi})$ ,  $B \sin(2\pi\omega n + \tilde{\phi})$  и при более слабом требовании на  $\omega$ , когда  $2L\omega \in \mathbb{N}$  ( $2K\omega \in \mathbb{N}$ ), см. [2].

*Таким образом, при идентификации собственных троек, соответствующих  $\varepsilon$ -м гармоническому ряду с  $0 < \omega < 0.5$ , будем искать два собственных вектора  $U_k, U_{k+1}$ , элементы которых задаются  $\varepsilon$ -м гармониками с одинаковыми  $\omega$  и  $\alpha$ .*

**Экспоненциально-модулированный гармонический ряд с  $\omega = 0.5$**  Если  $\cos(\phi) \neq 0$ , то  $f_n$  пропорционально  $(-e^\alpha)^n$ . При любых допустимых  $L, N$ :  $\text{rank}(F) = 1$ . Собственный вектор будет иметь такую же пилообразную форму, как и ряд  $F$ , причём с тем же  $\alpha$ .

*При идентификации собственных троек, соответствующих  $\varepsilon$ -м гармоническому ряду с  $\omega = 0.5$ , будем искать один собственный вектор, элементы которого имеют  $\varepsilon$ -м пилообразный вид.*

**Экспоненциальный ряд** Ряд  $F$  (1.5) при  $\omega = 0$  и  $\cos(\phi) \neq 0$  становится экспоненциальным рядом ( $f_n$  пропорционально  $e^{\alpha n}$ ). Для любых допустимых  $L, N$ :  $\text{rank}(F) = 1$ . Собственный вектор имеет экспоненциальную форму с тем же  $\alpha$ .

**Полиномиальный ряд** Рассмотрим ряд  $F$ , который задаётся уравнением

$$f_n = \sum_{k=0}^m a_k n^k, \quad a_m \neq 0.$$

При  $L \geq m + 1$ ,  $N \geq L + m$ :  $\text{rank}(F) = m + 1$ , и все собственные вектора также являются полиномами степени, не превосходящей  $m$  (см. [10, раздел 3.2.2]).

Задача выделения тренда может быть представлена как задача выделения медленно меняющихся (не осциллирующих и не шумовых) составляющих конечного ранга (в частности — экспоненциальных и полиномиальных рядов), вклад которых в вид ряда велик.

*При идентификации собственных троек, соответствующих тренду, будем искать собственные вектора, элементы которых медленно меняются.*

Традиционно выделение аддитивных составляющих ряда в рамках подхода “Гусеница”-SSA выполняется с помощью визуального изучения сингулярных чисел и векторов. В качестве правил для группировки собственных троек используются сформулированные выше требования к собственным векторам, см. [2, раздел 3], [5, раздел 3], [10, раздел 4.1.2], [41, раздел 1.6].

## 1.4 Прогноз аддитивной составляющей

**Определение 1.6** Следуя [11, раздел 2.1], будем говорить, что временной ряд  $F$  длины  $N$  имеет *конечно-разностную размерность, не большую чем  $d$*  ( $\text{fdim}(F) \leq d$ ), если  $1 \leq d \leq N - 1$  и найдутся такие  $a_1, \dots, a_d \in \mathbb{R}$ , что

$$f_{i+d} = \sum_{k=1}^d a_k f_{i+d-k}, \quad 0 \leq i \leq N - d - 1, \quad a_d \neq 0. \quad (1.10)$$

**Определение 1.7** Число  $d = \min\{k : \text{fdim}(F) \leq k\}$  называется *конечно-разностной размерностью* ряда  $F$  и записывается как  $\text{fdim}(F) = d$ . Для  $F: f_n = 0$  положим  $\text{fdim}(F) = 0$ .

**Определение 1.8** Формула (1.10) называется *линейной рекуррентной формулой (ЛРФ)* порядка  $d$ . ЛРФ порядка, равного  $\text{fdim}(F)$ , называется *минимальной ЛРФ*. Если для элементов ряда  $F$  справедливо (1.10), то будем говорить, что  $F$  *управляется ЛРФ* (1.10).

**Утверждение 1.6** Согласно [41, пример 5.4],  $\alpha$ -м гармонический ряд  $F = (f_0, \dots, f_{N-1})$  с частотой  $\omega$  и экспоненциальным показателем  $\alpha$  имеет конечно-разностную размерность 2 ( $\text{fdim}(F) = 2$ ) и задаётся следующей ЛРФ:

$$f_{n+2} = 2e^\alpha \cos(2\pi\omega) f_{n+1} - e^{2\alpha} f_n, \quad 0 \leq n \leq N - 3. \quad (1.11)$$

### 1.4.1 Вычисление коэффициентов ЛРФ порядка $L - 1$

Рассмотрим ряд  $F$  конечного ранга  $d$ , управляемый ЛРФ порядка  $d$ . Воспользуемся теоремой [11, теорема 2.2], [41, теорема 5.2] для поиска коэффициентов ЛРФ порядка  $L - 1$ , которая управляет рядом  $F$ .

Введём следующие обозначения для вектора  $P \in \mathbb{R}^L$ ,  $P = (p_1, \dots, p_L)^T$ :  
 $P^\nabla = (p_1, \dots, p_{L-1})^T$ ,  $\pi = p_L$ .

Пусть  $\{P_1, \dots, P_d\}$  — некоторый базис  $L$ -траекторного пространства  $\mathcal{L}_L(F)$ . Если  $L > d$ ,  $e_L \notin \mathcal{L}_L(F)$ , то существует ЛРФ порядка  $L - 1$ , управляющая рядом  $F$ , с коэффициентами  $b_1, \dots, b_{L-1}$ , которые задаются равенством:

$$(b_{L-1}, \dots, b_1)^T = \frac{\sum_{j=1}^d \pi_j P_j^\nabla}{1 - \sum_{j=1}^d \pi_j^2}. \quad (1.12)$$

Таким образом, чтобы построить ЛРФ порядка  $L - 1$  для ряда  $F$ , необходимо сначала провести сингулярное разложение траекторной матрицы ряда  $F$ . После этого, взяв собственные вектора  $U_1, \dots, U_d$  в качестве базиса траекторного подпространства ряда (обоснование этому см. в [10, раздел 3.1], [41, раздел 1.2.2]), посчитаем коэффициенты ЛРФ по формуле (1.12).

Если  $F = F^{(1)} + F^{(2)}$  и  $F^{(1)}$  — ряд ранга  $d$ , отделимый от  $F^{(2)}$ , то для получения коэффициентов ЛРФ, управляющей  $F^{(1)}$ , нужно использовать собственные вектора, соответствующие  $F^{(1)}$ .

### АЛГОРИТМ 1.3. Процедура SSAFORE: прогноз составляющей ряда с помощью метода “Гусеница”-SSA

**Входные данные:**  $F$ :  $F = F^{(1)} + F^{(2)}$ ,  $F^{(1)} = (f_1^{(1)}, \dots, f_{N-1}^{(1)})$

**Результат:**  $\tilde{f}$  — прогноз значения  $f_N^{(1)}$

**Параметры:**  $L$  — длина окна ( $1 < L \leq \lceil N/2 \rceil$ ),  $\mathcal{I}$  — номера собственных троек, соответствующих  $F^{(1)}$

#### Процедура:

1. Применим к ряду  $F$  процедуру SSADEC с параметром  $L$ , получив собственные тройки  $\{(\sqrt{\lambda_k}, U_k, V_k)\}_{k=1}^d$  (алгоритм 1.1)
2. По ним восстановим  $\tilde{F}^{(1)}$  с помощью процедуры SSAREC, взяв  $\mathcal{I}$  в качестве параметра (алгоритм 1.2)
3. Найдём коэффициенты ЛРФ для  $\tilde{F}^{(1)}$  по (1.12), взяв  $\{U_j\}_{j \in \mathcal{I}}$  в качестве  $\{P_j\}$  в (1.12)
4. Найдём  $\tilde{f}$ , применив найденные коэффициенты ЛРФ к последним элементам  $\tilde{F}^{(1)}$  по (1.10)

### 1.4.2 Минимизация линейной рекуррентной формулы

Ряду  $F$  конечно-разностной размерности  $d$  ( $\text{fdim}(F) = d$ ) соответствует много ЛРФ порядка больше  $d$  и только одна минимальная ЛРФ порядка  $d$  (см. [11, раздел 2.1]). Для бесконечного ряда  $F$  конечного ранга  $d$  порядок минимальной ЛРФ будет равен  $d$  (см. [10, предложение 3.4], [41, следствие 5.1]).

Таким образом, имея для ряда ранга  $d$  ( $d < L - 1$ ) ЛРФ порядка  $L - 1$ , можно поставить задачу её минимизации, т. е. нахождения по ней ЛРФ порядка  $d$ . Покажем, как это может быть сделано.

Рассмотрим характеристический полином, соответствующий ЛРФ порядка  $L - 1$  с коэффициентами  $b_1, \dots, b_{L-1}$ :

$$P_{L-1}(x) = x^{L-1} - \sum_{k=1}^{L-1} b_k x^{L-k-1}, \quad b_{L-1} \neq 0. \quad (1.13)$$

Обозначим за  $k_m$  ( $1 \leq m \leq p$ ) кратности несовпадающих корней  $x_1, \dots, x_p$  полинома и пусть  $h_n(m, j) = n^j x_m^n$ ,  $1 \leq m \leq p$ ,  $0 \leq j \leq k_m - 1$ . Известно, что ряд  $F$  удовлетворяет заданной ЛРФ тогда и только тогда, когда

$$f_n = \sum_{m=1}^p \sum_{j=0}^{k_m-1} c_{mj} h_n(m, j), \quad (1.14)$$

где комплексные коэффициенты  $c_{mj}$  определяются по первым значениям ряда  $f_0, \dots, f_{L-2}$  (см. [41, теорема 5.3]).

Таким образом, установлена связь между коэффициентами ЛРФ и видом ряда: коэффициенты ЛРФ задают характеристический полином, корни которого определяют вид ряда. Минимальная ЛРФ порождает корни, которые называются главными. Набор корней для неминимальной ЛРФ состоит из этих корней, к которым добавляются несущественные корни [11, раздел 2.5]. Поэтому для того, чтобы минимизировать ЛРФ произвольного большого порядка, необходимо среди всех корней ее характеристического полинома отобрать главные и по ним восстановить коэффициенты минимальной ЛРФ, которые связаны с коэффициентами полинома уравнением (1.13).

**Замечание 1.4** Поиск главных корней основывается на представлении (1.14). Нетрудно увидеть, что линейному ряду соответствует два главных корня  $x_1 = x_2 = 1$ . Э-м гармоническому ряду с частотой  $\omega$ ,  $0 < \omega < 0.5$ , и экспоненциальным показателем  $\alpha \in \mathbb{R}$  соответствуют два комплексно-сопряженных главных корня  $x_1 = e^{\alpha+i2\pi\omega}$ ,  $x_2 = e^{\alpha-i2\pi\omega}$  (см. [41, пример 5.10]). Соответственно, для немодулированной гармонике (при  $\alpha = 0$ )  $|x_1| = |x_2| = 1$ .



Опишем принцип выбора главных корней, которые соответствуют э-м гармонике. Он будет использоваться в главах 3, 4 для оценки параметров выделенной э-м гармоники.

**Пример 1.1** Из представления (1.14) следует, что для ряда — э-м гармоники с частотой  $\omega$ :  $0 < \omega < 0.5$  необходимо в качестве главных корней брать два комплексно сопряжённых корня.

Опыт поиска главных корней для модельных примеров показывает, что главные корни отличаются от остальных корней (побочных) бóльшим модулем. Таким образом, будем считать, что *главные корни для э-м гармоники с частотой  $0 < \omega < 0.5$  — это два сопряжённых корня с максимальным (среди всех корней) модулем.*

Такой способ поиска главных корней э-м гармоники продемонстрировал хорошее качество при статистическом исследовании.

При вычислении корней характеристического полинома для некоторой составляющей ряда расчёт коэффициентов ЛРФ (а, значит, и коэффициентов характеристического полинома) должен осуществляться по собственным векторам, которые соответствуют этой составляющей.

Этот способ позволяет также найти главные корни э-м гармонической составляющей по немного искажённым собственным векторам, что случается из-за приближённой разделимости э-м гармоники с некоторой другой составляющей (например, с шумом).

# Глава 2

## Автоматический метод выделения тренда

В данной главе предложим метод автоматической идентификации (АИ) трендовых собственных троек, который позволяет реализовать автоматическое выделение тренда в рамках подхода “Гусеница”-SSA.

Напомним, что интерактивная (визуальная) идентификация (ВИ) трендовых компонент опирается на тот факт, что последовательность элементов собственного вектора, соответствующего тренду, тоже медленно изменяется (см. раздел 1.3.1). Построим метод, который будет проверять для каждой собственной тройки, медленно ли изменяется последовательность  $u_1, \dots, u_L$  элементов её собственного вектора  $U = (u_1, \dots, u_L)^T$ .

**Замечание 2.1** Можно использовать не собственные, а факторные вектора, что не меняет сути метода, так как факторный вектор при длине окна  $L$  совпадает с собственным вектором при длине окна  $N - L + 1$ . Таким образом, переход от рассмотрения собственных векторов к факторным равносильно смене  $L$  на  $N - L + 1$ .

При разработке метода идентификации трендовых составляющих были испробованы три критерия для проверки медленной изменяемости последовательности элементов  $U$ , основанные на использовании: рангового коэффициента корреляции Кендалла [68, раздел 4.1.1], количестве “нулей” кусочно-линейной функции, значениями которой являются элементы вектора [68, раздел 4.1.2], а также на разложении Фурье последовательности элементов вектора [2, 5]. Последний критерий оказался наилучшим как по качеству идентификации, так и по интерпретируемости [2, раздел 7.1]. Удобным оказалось также то, что метод для идентификации периодических составляющих, описанный далее в главе 3, тоже использует частотный подход.

### 2.1 Вводные данные

**Определение 2.1** Будем называть *периодограммой ряда*  $F = (f_0, \dots, f_{N-1})$  длины  $N$  функцию, определённую следующим образом на множестве частот  $\{k/N\}_{k=0}^{\lfloor N/2 \rfloor}$ :

$$\Pi_F^N(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{-i2\pi\omega n} f_n \right|^2, \quad \omega \in \{k/N\}_{k=0}^{\lfloor N/2 \rfloor}. \quad (2.1)$$

Запишем представление Фурье ряда  $F$  в обычном виде (см., например, [32, раздел 7.3]):

$$f_n = c_0 + \sum_{1 \leq k \leq \frac{N-1}{2}} (c_k \cos(2\pi nk/N) + s_k \sin(2\pi nk/N)) + (-1)^n c_{N/2},$$

где  $k \in \mathbb{N}$ ,  $0 \leq n \leq N-1$ , а  $c_{N/2} = 0$ , если  $N$  — нечётное. Тогда периодограмма ряда  $F$  выражается следующим образом:

$$\Pi_F^N(k/N) = \frac{N}{2} \begin{cases} 2c_0^2, & k = 0, \\ c_k^2 + s_k^2, & 1 \leq k \leq (N-1)/2, \\ 2c_{N/2}^2, & \text{если } N \text{ чётное и } k = N/2. \end{cases} \quad (2.2)$$

Периодограмма была введена А. Шустером в 1898 г. [59], см. также [9, раздел 5.2].

**Утверждение 2.1** В [41, раздел 1.4.1] доказано следующее равенство:

$$\sum_{1 \leq k \leq (N-1)/2} \Pi_F^N(k/N) = \|F\|_2^2,$$

где под  $\|F\|_2$  понимается норма последовательности элементов ряда  $F$  в  $l_2$ -пространстве, такая что  $\|F\|_2^2 = \sum_{n=0}^{N-1} f_n^2$  (см. также [10, Предложение 2.5]).

Значение периодограммы  $\Pi_F^N(\omega)$  отражает относительный вклад гармоники с частотой  $\omega$  в разложение Фурье ряда  $F$ , поэтому она может служить характеристикой медленной изменяемости ряда: чем медленнее изменяется ряд, тем больше значения периодограммы в области около нуля.

Будем понимать под периодограммой вектора  $U \in \mathbb{R}^L$  периодограмму последовательности его элементов  $u_1, \dots, u_L$  и обозначать её как  $\Pi_U^L(\omega)$ .

**Интервал низких частот**  $[0, \omega_0]$  Периодограмма позволяет формализовать понятие медленной изменяемости последовательности чисел (например, элементов ряда). Зададим некоторую частоту  $\omega_0 \in (0, 0.5)$ , такую что гармоники с частотами из интервала  $[0, \omega_0]$  будут рассматриваться как медленно меняющиеся. *Будем считать, что ряд (или его составляющая) изменяется медленно, если значения периодограммы ряда (или периодограммы этой составляющей) на частотном интервале  $[0, \omega_0]$  достаточно велики по отношению к значениям на  $(\omega_0, 0.5]$ .* Этот естественный с точки зрения представления Фурье подход<sup>1</sup> позволяет построить следующий метод для идентификации трендовых собственных троек.

<sup>1</sup>Например, в низкочастотном фильтре Фурье тоже есть параметр, соответствующий границе низкочастотного интервала — это частота, начиная с которой коэффициенты Фурье обнуляются

## 2.2 Описание метода низких частот для идентификации трендовых компонент

Назовём предложенный далее метод *методом низких частот (НЧ)*. Опишем его в применении к собственному вектору  $U \in \mathbb{R}^L$ .

Фиксируем частоту  $\omega_0$  — границу интервала низких частот  $[0, \omega_0]$ .

**Определение 2.2** Определим  $\mathcal{C}(U)$ :

$$\mathcal{C}(U) = \frac{\sum_{0 \leq \omega \leq \omega_0} \Pi_U^L(\omega)}{\sum_{0 \leq \omega \leq 0.5} \Pi_U^L(\omega)}, \quad \omega \in \{k/N\}_{k=0}^{\lfloor N/2 \rfloor}. \quad (2.3)$$

Назовём  $\mathcal{C}(U)$  *вкладом гармоник с низкими частотами в разложение Фурье элементов вектора  $U$* .

Методом низких частот будем называть метод идентификации трендовых составляющих, который использует критерий на основе  $\mathcal{C}(U)$ . Будем идентифицировать собственную тройку, если её собственный вектор удовлетворяет неравенству:

$$\mathcal{C}(U) \geq \mathcal{C}_0,$$

где  $\mathcal{C}_0 \in [0, 1]$  — заранее заданное пороговое значение.

Метод управляется заданием параметра  $\omega_0$  и порогового значения  $\mathcal{C}_0$ . Очевидно, что чем больше  $\mathcal{C}_0$ , тем строже критерий.

**АЛГОРИТМ 2.1. Метод низких частот: автоматическая идентификация трендовых собственных троек**

**Входные данные:**  $\{U_k\}_{k=1}^d$  — собственные вектора сингулярного разложения траекторной матрицы некоторого ряда

**Результат:**  $\mathcal{I}_A$  — номера собственных троек

**Параметры:**  $\omega_0 \in (0, 0.5)$ ,  $\mathcal{C}_0 \in [0, 1]$

**Процедура:**

1.  $\mathcal{I}_A = \{k : \mathcal{C}(U_k) \geq \mathcal{C}_0\}$ , где  $\mathcal{C}(U_k)$  вычисляется при  $\omega_0$  по (2.3)

**АЛГОРИТМ 2.2. Процедура TREND: выделение тренда ряда с помощью метода низких частот**

**Входные данные:**  $F$  — ряд длины  $N$ , содержащий тренд  $F^{(T)}$

**Результат:**  $F_A$  — аппроксимация  $F^{(T)}$

**Параметры:**  $L$  ( $1 < L \leq \lceil N/2 \rceil$ ),  $\omega_0 \in (0, 0.5)$ ,  $\mathcal{C}_0 \in [0, 1]$

**Процедура:**

1. Получим собственные тройки, применив процедуру SSADEC с параметром  $L$  к ряду  $F$ :  $\{(\sqrt{\lambda_k}, U_k, V_k)\}_{k=1}^d$  (алгоритм 1.1)
2. Применим метод низких частот к  $\{U_k\}_{k=1}^d$  с параметрами  $\omega_0$ ,  $\mathcal{C}_0$ , получив  $\mathcal{I}_A$  (алгоритм 2.1)
3. Восстановим  $F_A$ , применив к  $\{(\sqrt{\lambda_k}, U_k, V_k)\}_{k=1}^d$  процедуру SSAREC с параметром  $\mathcal{I}_A$  (алгоритм 1.2)

**Замечание 2.2** Заметим, что хотя метод НЧ основан на разложении Фурье, процедура TREND не является простым частотным фильтром. Основанная на методе “Гусеница”-SSA, она будет обладать такими возможностями, как обработка нестационарных рядов, отделение сигнала от стационарного шума, выделение тренда при наличии периодической составляющей (см. главу 1).

### 2.3 Выбор параметра $\omega_0$

Выбор параметра  $\omega_0 \in (0, 0.5)$  должен осуществляться исходя из условий поставленной задачи и дополнительной информации о ряде. Значение  $\omega_0$  управляет детализацией результата: чем меньше  $\omega_0$ , тем медленнее будет изменяться выделенный тренд. Можно сказать, что значение  $\omega_0$  отражает частотную разницу между трендовыми и нетрендовыми составляющими.

Значение  $\omega$  надо выбирать так, чтобы аргументы больших значений периодограммы трендового (медленно меняющегося) вектора попали в интервал  $[0, \omega_0]$ . Нарушение этого условия может привести к ошибке первого рода — неидентификации компоненты, соответствующей тренду.

В то же время слишком большим  $\omega_0$  не может быть, ведь это правая граница интервала низких частот, и при выборе слишком большого  $\omega_0$  будут идентифицироваться слишком быстро изменяющиеся собственные вектора.

Необходимо также помнить, что при отсутствии точной разделимости собственные вектора могут быть искажены, что может привести к некоторому искажению их периодограмм, поэтому  $\omega_0$  нужно выбирать с запасом.

### Выделение тренда в присутствии периодической составляющей

В случае выделения тренда в присутствии периодической составляющей с известным периодом  $T$ , значение  $\omega_0$  должно выбираться так, чтобы частотные интервалы тренда и периодической составляющей не пересекались, т. е. чтобы  $\omega_0 < 1/T$ .

Под частотным интервалом понимается интервал, на котором сосредоточен основной вклад значений периодограммы. Для тренда это интервал низких частот, для гармоник это её частота при условии, что она попадает на решётку  $\{k/L\}_{k=0}^{\lfloor L/2 \rfloor}$ , а для  $\varepsilon$ -м гармоник с частотой  $\omega$  — это некоторая окрестность вокруг  $\omega$ , на которой сосредоточены большие значения её периодограммы.

Например, для ряда с ежемесячными данными, который содержит сезонную составляющую (периодическую с периодом 12), при выделении тренда необходимо выбирать  $\omega_0 < 1/12$ , чтобы не были ошибочно идентифицированы  $\varepsilon$ -м гармоник, которые дают в сумме сезонную составляющую.

### Выбор $\omega_0$ на основе периодограммы ряда

В общем случае, при отсутствии сведений о наличии в наблюдениях периодической составляющей или когда её период неизвестен, следует выбирать  $\omega_0$  на основании периодограммы исходного ряда, принимая во внимание следующий факт.

**Утверждение 2.2** Пусть  $G = (g_0, g_1, \dots, g_{N-1})$ ,  $H = (h_0, h_1, \dots, h_{N-1})$  — два ряда длины  $N$ , тогда для любого  $k: 0 \leq k \leq \lfloor N/2 \rfloor$  верно, что

$$|\Pi_{G+H}^N(k/N) - \Pi_G^N(k/N) - \Pi_H^N(k/N)| \leq 2\sqrt{\Pi_G^N(k/N)\Pi_H^N(k/N)}. \quad (2.4)$$

*Доказательство.* Согласно определению (2.1), периодограмма ряда  $G + H$  записывается следующим образом:

$$\Pi_{G+H}^N(k/N) = \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{-i2\pi nk/N} (g_n + h_n) \right|^2, \quad 0 \leq k \leq \lfloor N/2 \rfloor.$$

Введём для рядов  $G$  и  $H$  следующие  $z_k^{(1)}$  и  $z_k^{(2)}$ :

$$z_k^{(1)} = \sum_{n=0}^{N-1} e^{-i2\pi nk/N} g_n, \quad z_k^{(2)} = \sum_{n=0}^{N-1} e^{-i2\pi nk/N} h_n.$$

Если обозначить для числа  $z \in \mathbb{C}$  его вещественную и мнимую части как  $\Re z$ ,  $\Im z$ , то для любых двух  $z_1, z_2 \in \mathbb{C}$  выполняется следующее равенство

$$|z_1 + z_2|^2 = \Re^2(z_1 + z_2) + \Im^2(z_1 + z_2) = |z_1|^2 + |z_2|^2 + 2\Re z_1 \Re z_2 + 2\Im z_1 \Im z_2.$$

Поэтому верно следующее утверждение

$$\begin{aligned} \Pi_{G+H}^N(k/N) &= \frac{1}{N} \left| z_k^{(1)} + z_k^{(2)} \right|^2 = \\ &= \Pi_G^N(k/N) + \Pi_H^N(k/N) + \frac{2}{N} \Re z_k^{(1)} \Re z_k^{(2)} + \frac{2}{N} \Im z_k^{(1)} \Im z_k^{(2)}. \end{aligned}$$

Пользуясь выражением значения периодограммы через  $c_k$ ,  $s_k$  (2.2), и, считая, что  $c_k^{(1)}$ ,  $s_k^{(1)}$  — коэффициенты, соответствующие ряду  $G$ , а  $c_k^{(2)}$ ,  $s_k^{(2)}$  — соответствующие ряду  $H$ , получаем, что

$$\begin{aligned} \Pi_{G+H}^N(k/N) - \Pi_G^N(k/N) - \Pi_H^N(k/N) &= \frac{2}{N} \Re z_k^{(1)} \Re z_k^{(2)} + \frac{2}{N} \Im z_k^{(1)} \Im z_k^{(2)} = \\ &= \frac{N}{2} \left( c_k^{(1)} c_k^{(2)} + s_k^{(1)} s_k^{(2)} \right). \end{aligned}$$

Так как  $\forall a, b \in \mathbb{R} : |a| \leq \sqrt{a^2 + b^2}$ , то

$$\begin{aligned} \left| c_k^{(1)} c_k^{(2)} + s_k^{(1)} s_k^{(2)} \right| &\leq \left| c_k^{(2)} + s_k^{(2)} \right| \sqrt{(c_k^{(1)})^2 + (s_k^{(1)})^2} \leq \\ &\leq 2 \sqrt{(c_k^{(1)})^2 + (s_k^{(1)})^2} \sqrt{(c_k^{(2)})^2 + (s_k^{(2)})^2} = \frac{4}{N} \sqrt{\Pi_G^N(k/N) \Pi_H^N(k/N)}. \end{aligned}$$

Следовательно

$$\left| \Pi_{G+H}^N(k/N) - \Pi_G^N(k/N) - \Pi_H^N(k/N) \right| \leq \frac{N}{2} \cdot \frac{4}{N} \sqrt{\Pi_G^N(k/N) \Pi_H^N(k/N)}.$$

■

**Определение 2.3** Назовём носителем периодограммы  $\Pi_F^N$  ряда  $F$  подмножество  $\{k/N\}_{k=0}^{\lfloor N/2 \rfloor}$ , такое что  $\Pi_F^N(k/N) \neq 0$ .

Из утверждения 2.2 следует, что в случае, когда носители периодограмм для  $G$  и  $H$  не пересекаются,  $\Pi_{G+H}^N(k/N) = \Pi_G^N(k/N) + \Pi_H^N(k/N)$ .

При пересекающихся носителях равенство не выполняется, однако в случае, когда значения периодограмм сильно отличаются,  $\Pi_{G+H}^N \cong \Pi_G^N + \Pi_H^N$ . Продемонстрируем это.

То, что периодограммы в точке  $k/N$  сильно отличаются, можно записать как

$$\Pi_G^N(k/N) / \Pi_H^N(k/N) = d \gg 1,$$

(не умаляя общности, считаем, что  $\Pi_G^N(k/N) > \Pi_H^N(k/N)$ ).

Воспользовавшись утверждением 2.4, получаем, что

$$\begin{aligned} & \left| \Pi_{G+H}^N(k/N) - \Pi_G^N(k/N) - \Pi_H^N(k/N) \right| = \\ & = 2\sqrt{\Pi_G^N(k/N)\Pi_H^N(k/N)} = \frac{2}{\sqrt{d}}\Pi_G^N(k/N) \ll \Pi_G^N(k/N) \ll \Pi_H^N(k/N), \end{aligned}$$

т. е. разница  $\left| \Pi_{G+H}^N(k/N) - \Pi_G^N(k/N) - \Pi_H^N(k/N) \right|$  много меньше значений периодограмм.

**Замечание 2.3** Существует связь между разделимостью и непересечением носителей периодограмм отрезков рядов длины  $L$ ,  $K$ , см. [41, раздел 1.5.3], [10, раздел 2.5].

Таким образом, в случае, когда ряд состоит из тренда, большие значения периодограммы которого лежат в некотором низкочастотном интервале  $[0, \omega_0]$ , и остатка, большие значения периодограммы которого лежат в интервале  $(\omega_0, 0.5]$ , периодограмма их суммы (исходного ряда) будет близка к сумме периодограмм составляющих. Это позволяет по форме периодограммы исходного ряда увидеть, есть ли в ряде низкочастотная составляющая (тренд), и оценить  $\omega_0$  как границу низкочастотного интервала, на котором содержатся большие значения периодограммы.

Построим выбор  $\omega_0$  следующим образом. Рассмотрим периодограмму  $\Pi_F^N$  исходного ряда. Если граница трендового частотного интервала чётко видна, т. е.

- $\Pi_F^N$  около нуля имеет значения существенно большие, чем в среднем на всём интервале  $[0, 0.5]$ ,
- если вдалеке от нуля  $\Pi_F^N$  принимает сравнительно большие значения (соответствующие узкополосным<sup>2</sup> составляющим), то эти значения отделены от больших значений в окрестности нуля,

то в этом случае *возьмём  $\omega_0$ , которое соответствует этой границе.*

**Замечание 2.4** Опыт показывает, что при отсутствии больших значений периодограммы вдали от нуля хорошим является выбор такого  $\omega_0$ , что сумма значений периодограммы на интервале  $[0, \omega_0]$  составляет больше 80-90% от суммы всех значений периодограммы.

**Замечание 2.5** Надо учитывать изменение разрешения периодограммы при переходе от рассмотрения ряда  $F$  к собственным векторам  $U_j$ . В первом случае оно составляет  $1/N$ , а во втором —  $1/L$ . Если по  $\Pi_F^N$  было определено  $\omega_0 = \omega_0^{(N)}$  (как это предложено выше), то надо брать  $\omega_0 = \lceil L\omega_0^{(N)} \rceil / L$ . Если  $\omega_0^{(N)}$

<sup>2</sup>Узкополосным в частотном смысле



не попадает на решётку  $\{k/L\}_{k=0}^{\lfloor L/2 \rfloor}$ , то стоит взять  $\omega_0$  чуть больше  $\lceil L\omega_0^{(N)} \rceil / L$  (например,  $\lceil L\omega_0^{(N)} \rceil / L + 1/L$ ), так как в этом случае может наблюдаться эффект растекания максимума периодограммы по соседним точкам решётки  $\{k/L\}_{k=0}^{\lfloor L/2 \rfloor}$ .

Применение данного подхода на модельных примерах показало, что во многих случаях он позволяет определить  $\omega_0$ .

**Пример выбора  $\omega_0$  на основе периодограммы исходного ряда для экспоненциального тренда** Рассмотрим выбор  $\omega_0$  для ряда, состоящего из экспоненциального тренда и белого нормального шума, и покажем, как изменяется периодограмма экспоненциального ряда при добавлении к нему шума. Приведём также границы для  $|\Pi_{G+H}^N - \Pi_G^N - \Pi_H^N|$ , рассчитанные по (2.4).

**Пример 2.1** Пусть  $N = 120$ ,  $F = G + H$ ,  $g_n = Ae^{0.01n}$ ,  $h_n = B\varepsilon_n$ ,  $\varepsilon_n \sim NID(0, 1)$ , а нормирующие константы  $A$  и  $B$  выбраны таким образом, что  $\|G_N\|_2^2 = \|H_N\|_2^2 = \sum_{n=0}^N g_n^2 = \sum_{n=0}^N h_n^2 = 1$ . Это было сделано с тем, чтобы  $\sum_{k=0}^{60} \Pi_G^N(k/120) = \sum_{k=0}^{60} \Pi_H^N(k/120) = 1$ . Судя по графику  $\Pi_F$ , границе интервала низких частот соответствует 0.05. Принимая во внимание замечание 2.5, стоит взять  $\omega_0 = \lceil 0.05L \rceil / L + 1/L$ . Например, при  $L = 60$  это составляет  $\omega_0 = 0.067$ .

Графики ряда, его периодограммы и периодограмм его составляющих вместе с границами, рассчитанными по (2.4), приведены на рис. 2.1. Видно, что предложенная оценка для  $|\Pi_{G+H}^N - \Pi_G^N - \Pi_H^N|$  достаточно хороша и позволяет оценить форму  $\Pi_G^N$  по  $\Pi_{G+H}^N$  и выбрать по ней правильное  $\omega_0$ .

**Связь  $\omega_0$  с детализацией выделяемого тренда, выбор  $\omega_0$  в сложном случае** В случае, когда выделенный тренд недостаточно детален (это означает, что были взяты слишком медленно меняющиеся компоненты ряда), можно увеличить детализацию тренда, увеличив  $\omega_0$ .

При этом следует руководствоваться следующим эмпирическим правилом, которое надлежит использовать также и для выбора  $\omega_0$  в сложном случае (когда нельзя выбрать  $\omega_0$  по периодограмме и отсутствует информация о периодических составляющих).

*Возьмём исходный ряд и определим, гармоника какой максимальной частоты<sup>3</sup> должна быть отнесена к тренду. Отступим немного от этой частоты вправо (с учётом замечания 2.5) и возьмём полученное значение за  $\omega_0$ .*

---

<sup>3</sup>На языке разложения Фурье

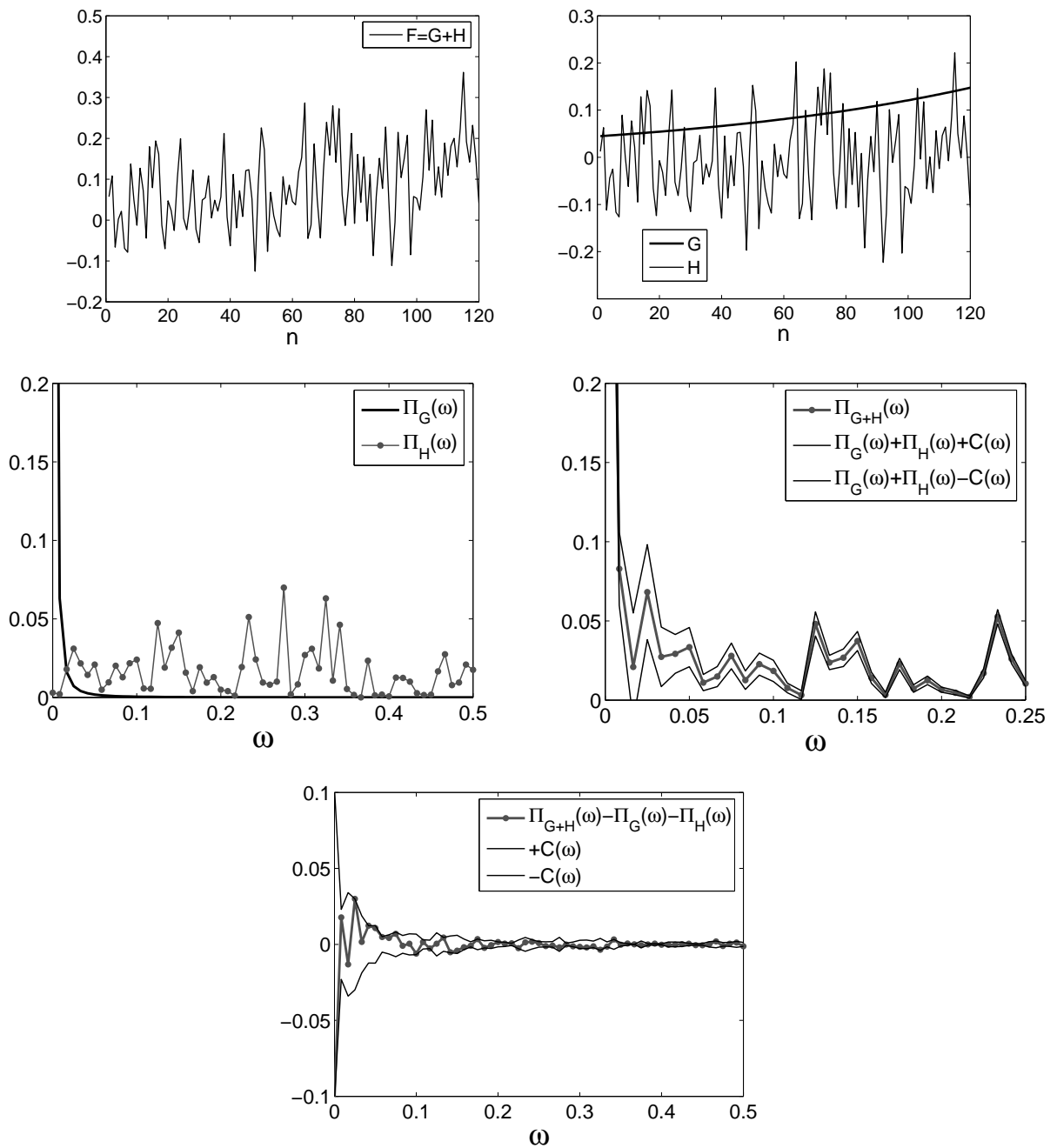


Рис. 2.1: Изменение периодограммы экспоненциального ряда  $G$  при добавлении к нему белого нормального шума  $H$ , использованное в обозначениях  $C(\omega) = 2\sqrt{\Pi_G^N(\omega)\Pi_H^N(\omega)}$ .

**Выбор  $\omega_0$  для данных с различной частотой замеров** Рассмотрим ряд  $F = (f_0, \dots, f_{N-1})$  как последовательность замеров функции  $f(\cdot)$  в равностоящих точках  $t_0, \dots, t_{N-1}$  ( $f_n = f(t_n)$ ), назовём  $r = (t_n - t_{n+1})^{-1}$  *разрешением*. Выберем некоторым образом  $\omega_0$ .

Рассмотрим теперь последовательность временных точек с большим разрешением  $r' = mr$ ,  $m \in \mathbb{N}$ :  $t'_0, \dots, t'_{(N-1)m+1}$ , тем самым увеличив длину ряда:  $N' = m(N-1) + 1$ . Для того, чтобы получить тот же тренд, но с большим разрешением, необходимо взять  $\omega'_0 = \omega_0/m$ . Аналогичным образом при уменьшении разрешения в  $m$  раз<sup>4</sup> надо брать  $\omega'_0 = m\omega_0$ .

## 2.4 Проверка метода для модели с известной трендовой составляющей

Прежде чем перейти к вопросу выбора  $\mathcal{C}_0$ , рассмотрим пример применения процедуры TREND к рядам следующей модели:  $F$  — ряд, состоящий из суммы экспоненциального тренда и нормального белого шума.

Будем моделировать ряды, вычислять лучшее в среднем пороговое значение  $\mathcal{C}_0$  и рассматривать средние характеристики процедуры TREND при этом  $\mathcal{C}_0$ .

Покажем с помощью моделирования, что для такой модели ряда при наилучшем в среднем  $\mathcal{C}_0$  метод НЧ выполняет поставленную задачу — автоматически идентифицирует собственные тройки, соответствующие тренду, не сильно ухудшая качество аппроксимации тренда по сравнению с ВИ. Изучим также среднюю ошибку аппроксимации: вычислим её значение при наилучшем в среднем  $\mathcal{C}_0$ , а также изучим её зависимость от  $\mathcal{C}_0$ .

**Схема проверки** Возьмём ряд  $F = (f_0, \dots, f_{N-1})$ , являющийся суммой экспоненциального тренда и реализации нормального белого шума с изменяющейся дисперсией:

$$f_n = f_n^{(T)} + f_n^{(\varepsilon)}, \quad f_n^{(T)} = e^{\alpha n}, \quad f_n^{(\varepsilon)} = \sigma e^{\alpha n} \varepsilon_n, \quad (2.5)$$

где  $\varepsilon_n$  соответствует белому нормальному шуму с единичной дисперсией, т. е.  $(\varepsilon_0, \dots, \varepsilon_{N-1})$  — это реализация последовательности независимых случайных величин, распределённых по нормальному закону с нулевым средним и единичной дисперсией. Будем обозначать<sup>5</sup> это как  $\varepsilon_n \sim NID(0, 1)$ .

<sup>4</sup>В обработке сигналов это называется *децимацией*

<sup>5</sup>Обозначение *NID* — от английского *normally and independently distributed*

Для различных реализаций шума при фиксированном  $\mathcal{C}_0$  осуществим выделение тренда с помощью процедуры TREND, оценивая качество АИ по сравнению с ВИ для каждого  $\mathcal{C}_0$ . Поскольку настоящая визуальная идентификация при моделировании не реализуема, то обратимся к её имитации. Согласно теории из главы 1, экспоненциальному тренду  $F^{(T)}$  при не слишком малом отношении сигнал/шум соответствует первая компонента разложения  $F$ . Рассмотрим такие  $N, \alpha, \sigma$ , при которых это выполняется (проверка осуществлялась визуально для нескольких реализаций ряда). Проведём имитацию ВИ, взяв в качестве соответствующей тренду первую собственную тройку.

Покажем, что при наилучшем в среднем  $\mathcal{C}_0$  для рядов рассматриваемой модели:

- АИ близка к ВИ (точнее к имитируемой процедуре ВИ),
- АИ обеспечивает хорошее качество аппроксимации.

**Комментарии к выбору модели ряда** Данная модель тренда была выбрана потому, что для неё при всех рассматриваемых значениях параметров легко имитировать ВИ. Другие достаточно простые тренды (например, полиномиальный) этим свойством не обладают. Статистическое исследование применения процедуры TREND к рядам с такими трендами было проведено, но его результаты сложнее интерпретируются (так как необходимо учитывать изменение результатов визуальной идентификации) и поэтому здесь не приведены. Скажем лишь, что сделанные в конце данного раздела выводы о качестве процедуры TREND распространяются на многие рассмотренные случаи полиномиальных и экспоненциально-модулированных полиномиальных трендов.

Шум такого рода встречается в реальных рядах. Уровень шума изменяется как в мультипликативной модели. Вообще говоря, уравнение (2.5) можно рассматривать как аддитивную запись мультипликативной модели.

**Оценка качества АИ** Качество автоматической идентификации по сравнению с визуальной идентификацией будем оценивать двумя способами. Во-первых, сравнивая тренды, восстановленные после проделанной идентификации и, во-вторых, сравнивая номера идентифицированных собственных троек.

Обозначим номера собственных троек — результат АИ, как  $\mathcal{I}_A$ , тренд, выделенный по  $\mathcal{I}_A$ , как  $F_A = (f_0^{(A)}, \dots, f_{N-1}^{(A)})$ , и сравним его с  $F_V = (f_0^{(V)}, \dots, f_{N-1}^{(V)})$ , выделенным по  $\mathcal{I}_V = \{1\}$ , взятой для имитирования ВИ. Если надо, то будем указывать  $\mathcal{C}_0$ , при котором проводилась идентификация:  $\mathcal{I}_A(\mathcal{C}_0)$ ,  $F_A(\mathcal{C}_0)$ .

Так как стандарт  $n$ -го элемента шума изменяется согласно изменению  $n$ -го элемента тренда, то естественно при сравнении  $F_{\mathcal{A}}$  и  $F_{\mathcal{V}}$  рассчитывать ошибку, как взвешенную среднеквадратичную, используя в качестве веса в  $n$ -й точке ряда значение  $e^{-\alpha n}$ . Обозначим такую ошибку как  $\mathcal{D}$ . Тогда ошибку АИ по сравнению с ВИ обозначим как  $\mathcal{D}_{\mathcal{AV}}$  и будем вычислять её следующим образом:

$$\mathcal{D}_{\mathcal{AV}} = \mathcal{D}(F_{\mathcal{A}}, F_{\mathcal{V}}) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-2\alpha n} (f_n^{(\mathcal{A})} - f_n^{(\mathcal{V})})^2.$$

Введём также абсолютные ошибки автоматической и визуальной идентификации (имеется в виду — вычисляемые независимо друг от друга):

$$\mathcal{D}_{\mathcal{A}} = \mathcal{D}(F_{\mathcal{A}}, F^{(\text{T})}) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-2\alpha n} (f_n^{(\mathcal{A})} - f_n^{(\text{T})})^2,$$

$$\mathcal{D}_{\mathcal{V}} = \mathcal{D}(F_{\mathcal{V}}, F^{(\text{T})}) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-2\alpha n} (f_n^{(\mathcal{V})} - f_n^{(\text{T})})^2$$

и будем вычислять их для сравнения с  $\mathcal{D}_{\mathcal{AV}}$ .

Заметим, что ошибка, вычисляемая с помощью АИ, зависит от  $\mathcal{C}_0$ .

#### 2.4.1 Расчёт ошибки АИ при наилучшем в среднем $\mathcal{C}_0$ с помощью моделирования

**Схема оценки ошибки автоматической идентификации** Вычислим ошибку АИ по следующей схеме.

Обозначим  $\mathcal{D}_{\mathcal{AV}}(\mathcal{C}_0) = \mathcal{D}(F_{\mathcal{A}}(\mathcal{C}_0), F_{\mathcal{V}})$ ,  $\mathcal{D}_{\mathcal{AV}}(\mathcal{C}_0)$  описывает зависимость от  $\mathcal{C}_0$  ошибки АИ по сравнению с ВИ.

Будем моделировать  $R$  раз ( $R \in \mathbb{N}$ ) ряды  $F^{(r)}$  согласно модели (2.5), перебирать  $\mathcal{C}_0$  в некотором интервале  $[\mathcal{C}_0^l, \mathcal{C}_0^h]$  с шагом  $\Delta\mathcal{C}$  и для каждого  $\mathcal{C}_0$  по выборке рядов  $\{F^{(r)}\}_{r=1}^R$  оценивать среднюю ошибку АИ по отношению к ВИ, взяв выборочное среднее:

$$\bar{\mathcal{D}}_{\mathcal{AV}}(\mathcal{C}_0) = \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F_{\mathcal{A}}^{(r)}(\mathcal{C}_0), F_{\mathcal{V}}^{(r)}).$$

**Определение 2.4** Обозначим значение  $\mathcal{C}_0$ , при котором достигается минимум средней ошибки, за  $\mathcal{C}_0^{opt}$ :

$$\mathcal{C}_0^{opt} = \arg \min_{\mathcal{C}_0} \bar{\mathcal{D}}_{\mathcal{AV}}(\mathcal{C}_0). \quad (2.6)$$

В случае, если (2.6) определяет  $\mathcal{C}_0^{opt}$  неоднозначно, будем считать

$$\mathcal{C}_0^{opt} = \max \left\{ \arg \min_{\mathcal{C}_0} \overline{\mathcal{D}}_{\mathcal{AV}}(\mathcal{C}_0) \right\}.$$

Далее рассмотрим средние ошибки автоматической процедуры с этим пороговым значением, введя следующие обозначения:

$$\begin{aligned} \overline{\mathcal{D}}_{\mathcal{AV}}^{opt} &= \overline{\mathcal{D}}_{\mathcal{AV}}(\mathcal{C}_0^{opt}), & \overline{\mathcal{D}}_{\mathcal{A}}^{opt} &= \overline{\mathcal{D}}_{\mathcal{A}}(\mathcal{C}_0^{opt}) = \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F_{\mathcal{A}}^{(r)}(\mathcal{C}_0^{opt}), F^{(T)}), \\ \overline{\mathcal{D}}_{\mathcal{V}} &= \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F_{\mathcal{V}}^{(r)}, F^{(T)}). \end{aligned} \quad (2.7)$$

Сравним также непосредственный результат АИ — группы идентифицированных компонент  $\{\mathcal{I}_{\mathcal{A}}^{(r)}\}_{r=1}^R$  и  $\mathcal{I}_{\mathcal{V}} = \{1\}$ , оценивая средний размер  $\mathcal{I}_{\mathcal{A}}^{(r)}$  и вероятность ошибки первого рода  $\mathcal{E}_1$  (когда  $\mathcal{I}_{\mathcal{A}}^{(r)}(\mathcal{C}_0^{opt}) \not\subseteq \{1\}$ ) при  $\mathcal{C}_0^{opt}$  следующими значениями:

$$\begin{aligned} \overline{\#\mathcal{I}_{\mathcal{A}}}^{opt} &= \overline{\#\mathcal{I}_{\mathcal{A}}}(\mathcal{C}_0^{opt}) = \frac{1}{R} \sum_{r=1}^R \#\mathcal{I}_{\mathcal{A}}^{(r)}(\mathcal{C}_0^{opt}), \\ \tilde{\mathcal{E}}_1^{opt} &= \tilde{\mathcal{E}}_1(\mathcal{C}_0^{opt}) = \#\{r : \mathcal{I}_{\mathcal{A}}^{(r)}(\mathcal{C}_0^{opt}) \not\subseteq \{1, 2\}\} / R, \end{aligned} \quad (2.8)$$

где  $\#$  обозначает мощность множества.

Приведём результаты вычислений  $\mathcal{C}_0^{opt}$  и  $\overline{\mathcal{D}}_{\mathcal{AV}}^{opt}$ ,  $\overline{\mathcal{D}}_{\mathcal{A}}^{opt}$ ,  $\overline{\mathcal{D}}_{\mathcal{V}}$ ,  $\overline{\#\mathcal{I}_{\mathcal{A}}}^{opt}$  и  $\tilde{\mathcal{E}}_1^{opt}$  для различных  $\alpha$ ,  $\sigma$ , полученные с помощью схемы, описанной в алгоритме 2.3.

**Значения параметров** Рассмотрим ряды с  $N = 47$ . Так как разделимость тренда с шумом — асимптотическая, то выбор большого  $L$  позволяет достичь лучшего качества разделимости, поэтому возьмём  $L = \lceil N/2 \rceil = 24$ . Получим оценки на  $R = 10^4$  реализациях.

В качестве  $\alpha$  выберем следующие значения: 0, 0.01, 0.02, что соответствует увеличению значений тренда от  $f_0 = 1$  до  $f_{N-1}$  в 1, 1.6, 2.5 раз соответственно. Тренд с таким возрастанием встречается в исследованиях реальных рядов.

Значение  $\sigma$  следует выбирать так, чтобы предложенная имитация ВИ имела смысл. Визуальное исследование несколько смоделированных рядов  $F^{(r)}$  показало, что при приблизительно  $\sigma > 1.6$  уровень шума слишком велик в том смысле, что нет чёткой разницы между вкладом (измеряемым собственным числом) собственных троек соответствующих тренду и шуму. Это может привести к тому, что трендовая тройка будет иметь номер, отличный от единицы, и имитация ВИ неоправданна. Поэтому станем рассматривать  $0 < \sigma \leq 1.6$ , взяв значения  $\sigma$  с шагом 0.2.

**АЛГОРИТМ 2.3. Оценка характеристик процедуры TREND при наилучшем в среднем  $\mathcal{C}_0$  ( $\mathcal{C}_0^{opt}$ ) для экспоненциального тренда в присутствии шума**

**Результат:**  $\mathcal{C}_0^{opt}$  — наилучшее в среднем  $\mathcal{C}_0$ , оценки характеристик процедуры TREND:  $\overline{\mathcal{D}}_{AV}^{opt}$ ,  $\overline{\mathcal{D}}_A^{opt}$ ,  $\overline{\mathcal{D}}_V$ ,  $\overline{\#\mathcal{I}}_A^{opt}$ ,  $\tilde{\mathcal{E}}_1^{opt}$

**Параметры:**

- Основные параметры:  $R$ ,  $\mathcal{C}_0^l$ ,  $\mathcal{C}_0^r$ ,  $\Delta\mathcal{C}$
- Параметры модели ряда:  $N$ ,  $\alpha$ ,  $\sigma$
- Параметры процедуры TREND:  $L$ ,  $\omega_0$

**Процедура:**

1. Смоделируем  $R$  реализаций  $\{F^{(r)}\}_{r=1}^R$  по (2.5)
2.  $\Omega = \{\mathcal{C}_0^l + k\Delta\mathcal{C}\}_{k=0}^M$ ,  $M = \left\lceil \frac{\mathcal{C}_0^r - \mathcal{C}_0^l}{\Delta\mathcal{C}} \right\rceil$
3. Для каждого смоделированного ряда  $F^{(r)}$  и  $\forall \mathcal{C}_0 \in \Omega$ :
  - 1) применим процедуру TREND с  $L$ ,  $\omega_0$ ,  $\mathcal{C}_0$  к  $F^{(r)}$ , получив  $\mathcal{I}_A^{(r)}(\mathcal{C}_0)$ ,  $F_A^{(r)}(\mathcal{C}_0)$  (см. алгоритм 2.2)
  - 2) получим собственные тройки ряда  $F^{(r)}$  с помощью SSADEC и восстановим по ним  $F_V^{(r)}$  с помощью SSAREC с параметром  $\mathcal{I}_V = \{1\}$  (алгоритмы 1.1, 1.2)
  - 3) вычислим ошибку  $\mathcal{D}(F_A^{(r)}(\mathcal{C}_0), F_V^{(r)})$
4. Для каждого  $\mathcal{C}_0 \in \Omega$  вычислим

$$\overline{\mathcal{D}}_{AV}(\mathcal{C}_0) = \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F_A^{(r)}(\mathcal{C}_0), F_V^{(r)})$$

5. Оптимальным пороговым значением  $\mathcal{C}_0^{opt}$  тогда является

$$\mathcal{C}_0^{opt} = \max \left\{ \arg \min_{\mathcal{C}_0} \overline{\mathcal{D}}_{AV}(\mathcal{C}_0) \right\}$$

6. Рассчитаем  $\overline{\mathcal{D}}_{AV}^{opt}$ ,  $\overline{\mathcal{D}}_A^{opt}$ ,  $\overline{\mathcal{D}}_V$ ,  $\overline{\#\mathcal{I}}_A^{opt}$  и  $\tilde{\mathcal{E}}_1^{opt}$  по (2.7), (2.8)

Для того чтобы выбрать  $\omega_0$ , рассмотрим несколько примеров периодограммы ряда  $F^{(r)}$  с  $\alpha = 0, 0.2$  и  $\sigma = 1.6$  (такое  $\sigma$  соответствует худшему случаю из рассмотренных, при  $\sigma < 1.6$  граница интервала низких частот видна ещё лучше). Два из них изображены на рис. 2.2. Видно, что значение в нуле значительно превышает значения в остальных точках. Принимая во внимание замечание 2.5, возьмём  $\omega_0 = 0.05 \approx \lceil 0L \rceil / L + 1/L$ .

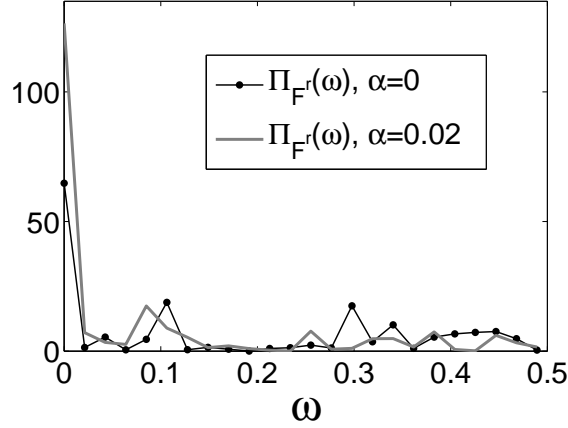


Рис. 2.2: Примеры  $\Pi_{F^{(r)}}$  для  $\alpha = 0$  и  $\alpha = 0.02$ .

#### 2.4.2 Поведение автоматической процедуры с $\mathcal{C}_0^{opt}$ при изменении параметров $\alpha, \sigma$

Результаты вычисления  $\mathcal{C}_0^{opt}$  и  $\overline{\mathcal{D}}_{AV}^{opt}$ ,  $\overline{\mathcal{D}}_A^{opt}$ ,  $\overline{\mathcal{D}}_V$ ,  $\overline{\#\mathcal{I}}_A^{opt}$  и  $\tilde{\mathcal{E}}_1^{opt}$  для разных  $\sigma$  при  $\alpha = 0$  приведены на рис. 2.3, на рис. 2.4 представлены для сравнения результаты при  $\alpha = 0.02$  (результаты для  $\alpha = 0.01$  опущены, так как они не меняют общей картины выводов). Сравним ошибку  $\mathcal{D}_{AV}$  с  $\mathcal{D}_V$ . Видно, что АИ близка к ВИ по сравнению с ошибкой, которую даёт ВИ процедура, т. е. выполняется  $\overline{\mathcal{D}}_{AV} \ll \overline{\mathcal{D}}_V$  (особенно это заметно при малом  $\sigma$ ).

На графике оценки вероятности ошибки первого рода  $\tilde{\mathcal{E}}_1^{opt}$  видно, что при  $\sigma < 1.6$  эта вероятность не превышает 0.05, что является хорошим показателем.

Всё это даёт основания утверждать, что для данной модели (экспоненциальный тренд плюс шум) автоматическая процедура TREND при наилучшем в среднем  $\mathcal{C}_0^{opt}$  близка к имитируемой визуальной.

Среднеквадратичное качество автоматического выделения (оцениваемое с помощью среднего  $\overline{\mathcal{D}}_A$ ) также довольно высоко, если сравнивать  $\sqrt{\overline{\mathcal{D}}_A}$  (среднее отклонение) с значением  $\sigma$  (стандартом шума), при котором оно измеряется.



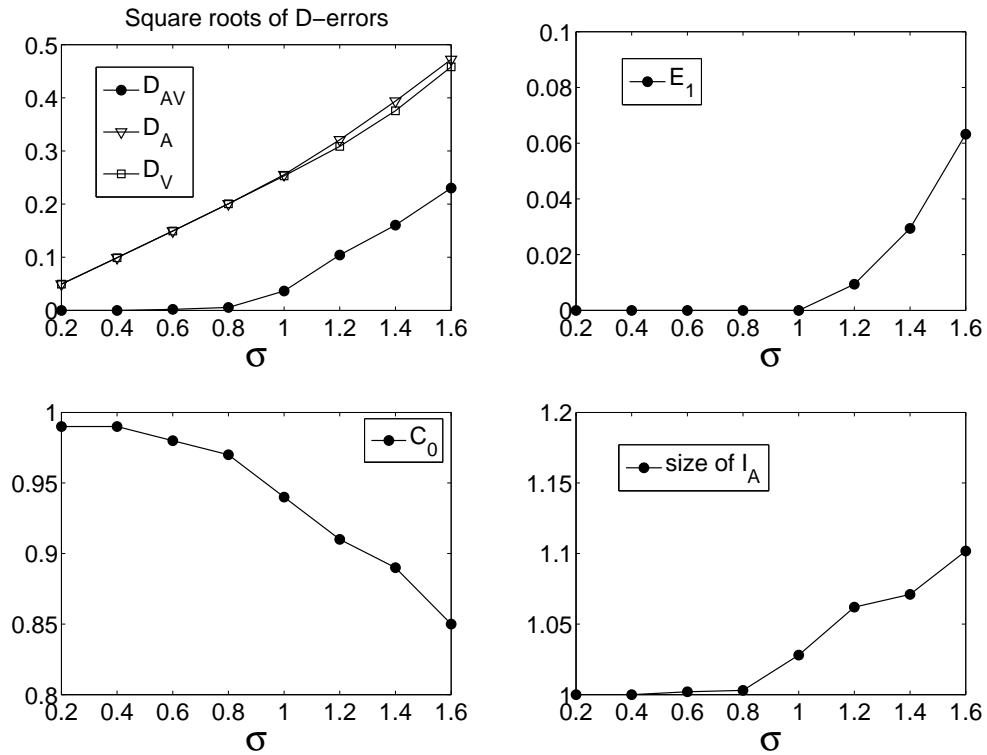


Рис. 2.3: Зависимость  $\overline{D}_{AV}^{opt}$ ,  $\overline{D}_A^{opt}$ ,  $\overline{D}_V$  (на графике изображены квадратные корни этих величин),  $\tilde{\mathcal{E}}_1^{opt}$ ,  $C_0^{opt}$  и  $\#\mathcal{I}_A^{opt}$  от  $\sigma$  при  $\alpha = 0$ .

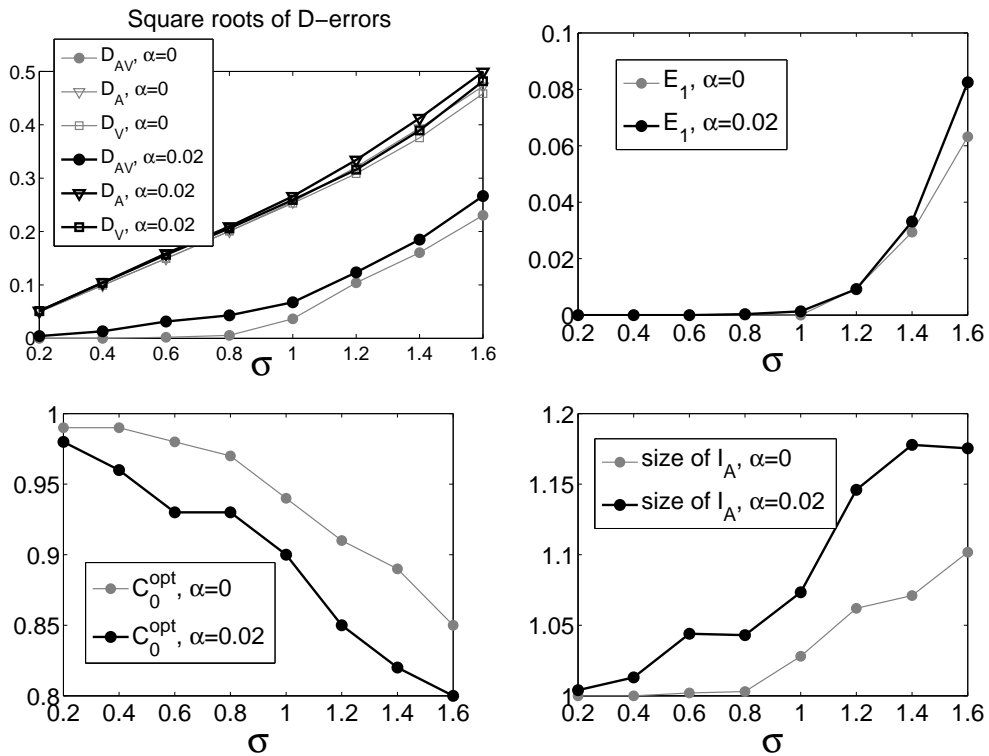


Рис. 2.4: Зависимость  $\overline{D}_{AV}^{opt}$ ,  $\overline{D}_A^{opt}$ ,  $\overline{D}_V$  (на графике изображены квадратные корни этих величин),  $\overline{D}_A$ ,  $\overline{D}_V$ ,  $\tilde{\mathcal{E}}_1^{opt}$ ,  $C_0^{opt}$  и  $\#\mathcal{I}_A^{opt}$  от  $\sigma$  при  $\alpha = 0$ ,  $\alpha = 0.02$ .

Видно, что для разных  $\alpha$  отличаются  $\mathcal{C}_0^{opt}$ ,  $\overline{\#\mathcal{I}}_A^{opt}$ , но характеристики качества  $(\overline{\mathcal{D}}_{AV}^{opt}, \tilde{\mathcal{E}}_1^{opt})$  практически не меняются. Это говорит о том, что процедура TREND успешно подстраивается к изменению параметров модели.

Конечно, отсюда нельзя сделать вывод о том, что процедура TREND работает для любого тренда и остатка. Поэтому далее, в разделе 2.9, будет представлена методика проверки процедуры TREND при обработки множества похожих рядов.

### 2.4.3 Зависимость ошибки от $\mathcal{C}_0$

Изучим теперь зависимость  $\overline{\mathcal{D}}_{AV}$  от  $\mathcal{C}_0$ . На рис. 2.5 изображен график зависимости  $\overline{\mathcal{D}}_{AV}(\mathcal{C}_0)$  от  $\mathcal{C}_0$  при  $\alpha = 0$ ,  $\sigma = 1.4$ . На нём видно, что справа от  $\mathcal{C}_0^{opt}$  ( $\mathcal{C}_0 > \mathcal{C}_0^{opt}$ ) значения  $\overline{\mathcal{D}}_{AV}(\mathcal{C}_0)$  резко увеличиваются. Это происходит из-за систематической неидентификации трендовой компоненты, что подтверждают графики оценки ошибки первого рода (рис. 2.6). При уменьшении  $\mathcal{C}_0$  ( $\mathcal{C}_0 < \mathcal{C}_0^{opt}$ ) ошибка увеличивается постепенно из-за малого вклада каждой из компонент шума, которые начинают идентифицироваться при ослаблении критерия АИ.

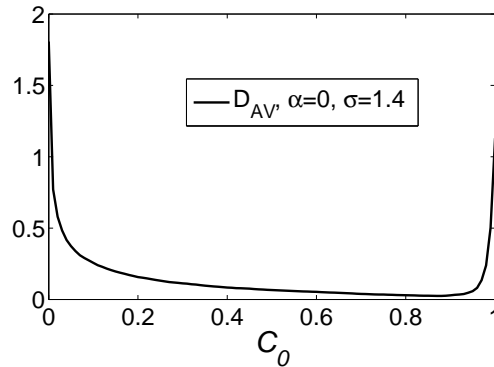


Рис. 2.5:  $\overline{\mathcal{D}}_{AV}(\mathcal{C}_0)$  при  $\alpha = 0$ ,  $\sigma = 1.4$  ( $\mathcal{C}_{max} \approx 0.87$ ).

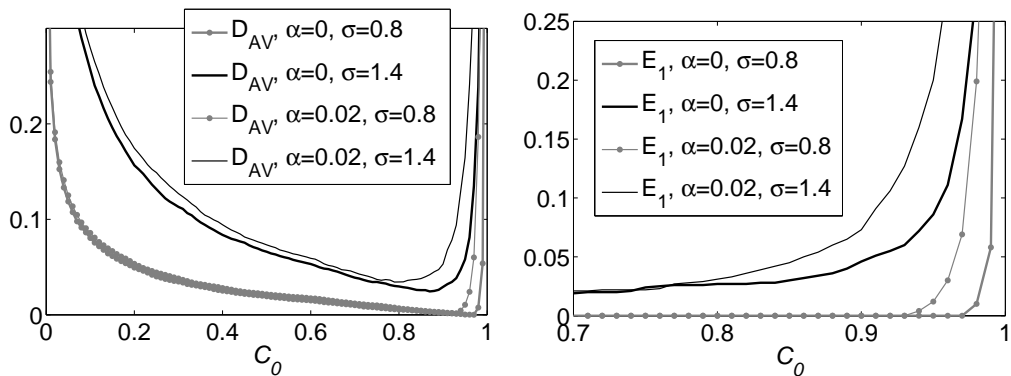


Рис. 2.6: Зависимость  $\overline{\mathcal{D}}_{AV}$  (слева) и  $\tilde{\mathcal{E}}_1$  (справа) от  $\mathcal{C}_0$  для разных параметров.

Из этого можно сделать вывод, что *выбор всех трендовых компонент обязателен для качественного выделения тренда, тогда как выбор лишних шумовых компонент сравнительно несильно ухудшает качество.* Это будет верно и для других типов тренда с большим вкладом в вид ряда.

#### 2.4.4 Общие соображения по выбору $\mathcal{C}_0$

Основное и естественное требование к АИ уже было сформулировано — идентификация всех трендовых компонент. В реальности, при отсутствии точной разделимости, при конечном  $N$ , иногда невозможно выбрать все трендовые собственные тройки и не выбрать лишних.

**Определение 2.5** Определим  $\mathcal{C}_0^{(T)}$ , *пороговое значение для тренда  $F^{(T)}$* , как

$$\mathcal{C}_0^{(T)} = \mathcal{C}_0^{(T)}(F^{(T)}) = \max\{\mathcal{C}_0 : \mathcal{I}^{(T)}(\mathcal{C}_0) = \{1, \dots, d\}\}, \quad \text{где } d = \text{rank } F^{(T)},$$

а  $\mathcal{I}^{(T)}(\mathcal{C}_0)$  — это номера идентифицированных с  $\mathcal{C}_0$  компонент разложения  $F^{(T)}$ .

То есть  $\mathcal{C}_0^{(T)}$  — это наибольшее пороговое значение из таких, что если взять  $F^{(T)}$  и применить к нему процедуру TREND с пороговым значением  $\mathcal{C}_0$ , то будут идентифицированы все компоненты. Здесь имеется в виду, что  $F^{(T)}$  — тренд конечного ранга, но это естественно, потому что подход “Гусеница”-SSA предполагает точное выделение только рядов конечного ранга.

Очевидно, что при  $\sigma = 0$  выполняется равенство  $\mathcal{C}_0^{opt} = \mathcal{C}_0^{(T)}$ .

В случае, когда  $F = F^{(T)} + F^{(1)}$ , и  $F^{(T)}$  и  $F^{(1)}$  точно разделимы, можно брать  $\mathcal{C}_0 = \mathcal{C}_0^{(T)}$ , но в реальности разделимость обычно приближённая. При приближённой разделимости соответствующие  $F^{(T)}$  собственные вектора немного искажаются, поэтому надо брать  $\mathcal{C}_0$  меньше  $\mathcal{C}_0^{(T)}$ . В то же время чем меньше  $\mathcal{C}_0$ , тем больше компонент не соответствующих тренду будет идентифицировано. *Поэтому  $\mathcal{C}_0$  должно выбираться так, чтобы с одной стороны, все трендовые компоненты были точно идентифицированы ( $\mathcal{C}_0 < \mathcal{C}_0^{(T)}$ ), а с другой стороны, чтобы было идентифицировано не слишком много компонент, не соответствующих тренду.*

К сожалению, выбор  $\mathcal{C}_0$  на основе применения правила  $\mathcal{C}_0 < \mathcal{C}_0^{(T)}$  не реализуем даже при точной разделимости, так как рассчитать зависимость  $\mathcal{C}_0^{(T)}$  от параметров модели тренда можно только в отдельных случаях. Для экспоненциального тренда это будет сделано в разделе 2.6.4. Для другого простого тренда, полиномиального, аналитический вид зависимости  $\mathcal{C}_0^{(T)}$  от параметров (коэффициентов) и  $\omega_0$  не может быть получен из-за сложности процедуры представления ряда в виде собственных троек.

## 2.5 Оценка качества выделения тренда

В предыдущем разделе было показано на примере выделения экспоненциального тренда, что процедура TREND при  $\mathcal{C}_0^{opt}$  выполняет поставленную задачу — автоматически идентифицирует трендовые компоненты, обеспечивая при этом хорошее качество аппроксимации тренда. Однако выбор порогового значения  $\mathcal{C}_0^{opt}$ , при котором оценивались характеристики TREND, был основан на использовании имитации ВИ ( $\mathcal{I}_Y = \{1\}$ ), что возможно для экспоненциального тренда.

В этом и следующем разделах будет изучен вопрос выбора  $\mathcal{C}_0$ .

Подход к выбору  $\mathcal{C}_0$  построим следующим образом. Предположим, что существует мера качества аппроксимации тренда, которая не зависит от неизвестного настоящего тренда  $F^{(T)}$ . В этом случае выбрать  $\mathcal{C}_0$  просто: переберём возможные  $\mathcal{C}_0$  и возьмём то, при котором достигается наилучшее качество.

**Замечание 2.6** В разделе 2.4 в качестве такой меры качества была взята взвешенная среднеквадратичная ошибка тренда, выделенного с помощью процедуры TREND по сравнению с трендом, выделенным по  $\mathcal{I}_Y = \{1\}$ .

В данном разделе изучим вопросы качества выделения тренда с помощью процедуры TREND и предложим искомую меру, которая может быть использована для выбора  $\mathcal{C}_0$ .

### 2.5.1 Требования к мере качества выделения тренда

Стандартной мерой близости двух рядов  $F$ ,  $G$  является среднеквадратичная ошибка, СКО, введём для её обозначения специальную норму:

$$\text{СКО}(F, G) = \|F - G\|^2 = \frac{1}{N} \|F - G\|_2^2 = \frac{1}{N} \sum_{n=0}^{N-1} (f_n - g_n)^2.$$

В гипотетическом случае, когда известен тренд  $F^{(T)}$ ,  $\text{СКО}(F_A, F^{(T)})$  будет естественной мерой качества выделения тренда или, иначе, мерой близости полученного результата  $F_A$  к настоящему тренду  $F^{(T)}$ .

**Замечание 2.7** В некоторых случаях бывает естественнее использовать взвешенную среднеквадратичную ошибку, которая рассчитывается в каждой временной точке с коэффициентом, обратно пропорциональным значению тренда в этой точке, как это было сделано в разделе 2.4.

$\text{СКО}(F_A, F^{(T)})$  является хорошей стандартной мерой близости двух рядов, однако в случае обработки реального ряда  $F^{(T)}$  неизвестен. Необходима

другая мера, зависящая только от  $F_A$ . Потребуем от неё согласованности с  $СКО(F_A, F^{(T)})$ , а именно, чтобы в случае, когда известен тренд  $F^{(T)}$ , минимум этой меры и минимум  $СКО$  (по  $C_0$ , с которым выделяется  $F_A$ ) достигались при одном и том же  $C_0$  или, хотя бы, чтобы по значениям меры можно было определить аргумент минимума  $СКО$ .

В случае, когда тренд заранее не известен и нет информации, что он принадлежит определённому классу, будем считать тренд  $F_{\gamma}$ , выделенный вручную, с помощью ВИ (см. раздел 1.3.1), идеальным результатом, и требовать, чтобы критерий был согласован с среднеквадратичной ошибкой по отношению к  $F_{\gamma}$ , т. е. в некотором смысле  $F^{(T)}$  заменяется на  $F_{\gamma}$ . Таким образом в общем случае мера должна отражать близость автоматической процедуры к визуальной, чтобы можно было оценить качество выделения тренда для любого ряда, для которого возможно извлечь тренд визуально.

Сформулированные требования “близости результата АИ к настоящему тренду” и “близости результатов АИ и ВИ” должны совпадать для рассматриваемых типов трендов (медленно изменяющиеся ряды конечного ранга). Это выполняется, так как асимптотически метод “Гусеница”-SSA сколь угодно точно выделяет такой тренд. Значит результат выделения тренда с помощью “Гусеница”-SSA асимптотически совпадает с настоящим трендом и, следовательно, не возникает расхождения в предъявленных только что требованиях.

### 2.5.2 $\mathcal{R}$ -мера качества выделения тренда

Пусть есть исходный ряд  $F$ , а  $F_A$  — тренд, выделенный с помощью процедуры TREND.

**Определение 2.6** Определим  $\mathcal{R}(F, F_A)$ :

$$\mathcal{R} = \mathcal{R}(F, F_A) = \frac{\mathcal{C}(F - F_A)}{\mathcal{C}(F)}, \quad (2.9)$$

где для ряда  $G$  под  $\mathcal{C}(G)$  понимаем вклад в его разложение Фурье гармоник с низкими (из интервала  $[0, \omega_0]$ ) частотами, см. (2.3)<sup>6</sup>. Назовём функцию  $\mathcal{R}(F, F_A)$   $\mathcal{R}$ -мерой качества аппроксимации тренда.

Критерий просто интерпретируется: это нормированный вклад гармоник с низкими частотами в разложение Фурье остатка  $F - F_A$ . Этот критерий хорошо согласуется с общим пониманием тренда как медленно меняющейся со-

<sup>6</sup>В (2.3) функция  $\mathcal{C}$  вводилась для последовательности элементов вектора, для последовательности элементов ряда вычислим её таким же образом

ставляющей, а также с самим методом НЧ (поскольку метод идентифицирует компоненты с большим  $\mathcal{C}(U)$ ).

Значения  $\mathcal{R}$  всегда неотрицательны (0 достигается при нулевом остатке) и для любой осмысленной процедуры выделения тренда они не превышают единицы, поэтому возьмём  $\min\{1, \mathcal{R}\}$  вместо  $\mathcal{R}$  и далее будем считать, что всегда выполняется  $\mathcal{R} \in [0, 1]$ .

В разделе 2.6 предложим процедуру выбора  $\mathcal{C}_0$  на основе меры  $\mathcal{R}$  и в разделе 2.6.3 представим примеры, демонстрирующие её согласованность с СКО.

## 2.6 Выбор порогового значения $\mathcal{C}_0$ на основе меры $\mathcal{R}$

В разделе 2.4.4 было показано, что при идентификации тренда необходимо брать такое  $\mathcal{C}_0$ , чтобы обязательно идентифицировались все трендовые собственные тройки. При этом оно должно быть не слишком мало, иначе будет идентифицировано слишком много лишних собственных троек.

В данном разделе введём  $\mathcal{C}_{max}$  — такое  $\mathcal{C}_0$ , при котором идентифицируются все компоненты, соответствующие тренду, предложим способ его поиска на основе меры  $\mathcal{R}$  и опишем процедуру выделения тренда на его основе.

### 2.6.1 $\mathcal{C}_{max}$ — пороговое значение, при котором идентифицируются все трендовые собственные тройки

Рассмотрим ряд  $F$ , содержащий тренд  $F^{(T)}$ . Проведём его разложение с помощью процедуры SSADEC (алгоритм 1.1). Пусть  $\mathcal{I}^{(T)}$  — группа компонент, соответствующая  $F^{(T)}$ .

**Определение 2.7** Обозначим через  $\mathcal{C}_{max}$  такое наибольшее  $\mathcal{C}_0$ , при котором идентифицируются все собственные тройки, соответствующие тренду  $F^{(T)}$ :

$$\mathcal{C}_{max} = \max\{\mathcal{C}_0 : \mathcal{I}_A \supset \mathcal{I}^{(T)}\}.$$

Отличие  $\mathcal{C}_{max}$  от  $\mathcal{C}_0^{(T)}$  состоит в том, что  $\mathcal{C}_0^{(T)}$  было определено только для тренда ( $F = F^{(T)}$ ), не принимая во внимание влияние остатка на тренд. При точной разделимости  $F^{(T)}$  и остатка  $F - F^{(T)}$  выполняется  $\mathcal{C}_{max} = \mathcal{C}_0^{(T)}$ . При приближённой разделимости обычно выполняется  $\mathcal{C}_{max} < \mathcal{C}_0^{(T)}$  (из-за искажения трендовых собственных векторов).

**Замечание 2.8** Отличие  $\mathcal{C}_{max}$  от  $\mathcal{C}_0^{opt}$  в том, что  $\mathcal{C}_0^{opt}$  рассчитывается в среднем (для модели), тогда как  $\mathcal{C}_{max}$  определяется для одного ряда.

Очевидным образом надо выбирать такое  $\mathcal{C}_0$ , что  $\mathcal{C}_0 \leq \mathcal{C}_{max}$ , так как нарушение этого условия приведёт к ошибке идентификации первого рода (неидентификации трендовых собственных троек). В разделе 2.4.3 было показано, что если брать  $\mathcal{C}_0$  в некоторой небольшой окрестности  $\mathcal{C}_0^{opt}$ , то при  $\mathcal{C}_0 > \mathcal{C}_0^{opt}$  ошибка будет много больше ошибки при  $\mathcal{C}_0 < \mathcal{C}_0^{opt}$ . Это объясняется именно неидентификацией трендовых компонент (т. е. взяли  $\mathcal{C}_0 > \mathcal{C}_{max}$ ). Таким образом вычисление  $\mathcal{C}_{max}$  является важнейшей проблемой при выборе  $\mathcal{C}_0$ .

К сожалению, получение аналитического вида  $\mathcal{C}_{max}$  невозможно в большинстве случаев из-за сложности процедуры сингулярного разложения траекторной матрицы ряда и необходимо использовать другие подходы для расчёта порогового значения  $\mathcal{C}_{max}$ .

### 2.6.2 Вычисление $\mathcal{C}_{max}$ с помощью меры $\mathcal{R}$

**Утверждение 2.3** Пусть  $(\sqrt{\lambda}, U, V)$  — некоторая собственная тройка разложения ряда  $F$  и  $\tilde{F}^{(1)}$  — ряд, восстановленный по этой тройке,  $U = (u_1, \dots, u_L)^T$ ,  $V = (v_1, \dots, v_L)^T$ . Если верно, что

$$\exists \delta_1, \delta_2 \in \mathbb{R} : \forall k, 1 \leq k \leq L-1 : |u_{k+1} - u_k| < \delta_1, |v_{k+1} - v_k| < \delta_2,$$

то для элементов восстановленного ряда  $\tilde{F}^{(1)} = (\tilde{f}_0^{(1)}, \dots, \tilde{f}_{N-1}^{(1)})$  выполняется следующее:

$$\exists \epsilon(\delta_1, \delta_2) : \forall n, 0 \leq n \leq N-1 : \left| \tilde{f}_{n+1}^{(1)} - \tilde{f}_n^{(1)} \right| < \epsilon(\delta_1, \delta_2).$$

*Доказательство.* Данное утверждение несложно доказывается, зная, что  $\tilde{F}^{(1)}$  получается из  $(\sqrt{\lambda}, U, V)$  следующим образом. Во-первых, рассчитывается матрица  $\mathbf{X}^{(1)} = \sqrt{\lambda}UV^T$ , во-вторых, проводится её диагональное усреднение.

Покажем, как вычислить  $\epsilon$  по  $\delta_1, \delta_2$  при  $L \leq K$  для  $\tilde{f}_n^{(1)}$ ,  $L < n < K$  (см. раздел 1.1.2). Для других случаев  $\epsilon(\delta_1, \delta_2)$  вычисляется аналогично.

$$\begin{aligned} \left| \tilde{f}_{n+1}^{(1)} - \tilde{f}_n^{(1)} \right| &= \frac{\sqrt{\lambda}}{L} \left| \sum_{m=1}^L (u_m v_{n-m+3} - u_m v_{n-m+2}) \right| < \\ &< \frac{\sqrt{\lambda}}{L} \sum_{m=1}^L |u_m| |v_{n-m+3} - v_{n-m+2}| < \\ &< \frac{\sqrt{\lambda}}{L} \delta_2 \sum_{m=1}^L |u_m| < \delta_2 \frac{\sqrt{\lambda}}{L} (u_1 + (L-1)\delta_1). \end{aligned}$$

■

Данное утверждение, а также свойства процедуры формирования матрицы  $\mathbf{X}$  по  $U$  и  $V$  позволяют утверждать следующее. Если собственный и факторный вектора медленно изменяются, то и ряд, восстановленный по ним, тоже будет медленно изменяться. И наоборот, если они изменяются быстро (хаотически или периодически, см. [2, раздел 2.3]), то восстановленный по ним ряд будет обладать такими же свойствами.

*Принимая во внимание утверждение 2.2, естественно ожидать, что если в ряде есть тренд, то восстановленные по отдельным трендовым собственным тройкам ряды дают большой вклад ряда в низкочастотном интервале  $[0, \omega_0]$ .*

Таким образом, значения  $\mathcal{C}_0$ , при которых происходит идентификация трендовых собственных троек, будут отмечены резким изменением  $\mathcal{R}(\mathcal{C}_0)$ . Воспользуемся этим для поиска  $\mathcal{C}_{max}$ .

Найдём значения  $\mathcal{C}_0$ , в которых  $\mathcal{R}(\mathcal{C}_0)$  резко увеличивается (при увеличении  $\mathcal{C}_0$ ). Переберём значения  $\mathcal{C}_0$  из некоторого интервала  $[\mathcal{C}_0^l, \mathcal{C}_0^h]$  с заданным шагом  $\Delta\mathcal{C}$  и для каждого из них применим к ряду  $F$  процедуру TREND с  $\mathcal{C}_0$  и посчитаем  $\mathcal{R}(\mathcal{C}_0) = \mathcal{R}(F, F_A(\mathcal{C}_0))$ . Получим зависимость  $\mathcal{R}(\mathcal{C}_0)$  от  $\mathcal{C}_0$  с разрешением  $\Delta\mathcal{C}$ .

Выберем минимальное  $\mathcal{C}_0$ , при котором прирост  $\mathcal{R}(\mathcal{C}_0 + \Delta\mathcal{C}) - \mathcal{R}(\mathcal{C}_0)$  будет больше некоторого заданного значения  $\Delta\mathcal{R}$ :

$$\mathcal{C}_0^{\mathcal{R}} = \min\{\mathcal{C}_0 : \mathcal{C}_0^l \leq \mathcal{C}_0 \leq \mathcal{C}_0^h, \mathcal{R}(\mathcal{C}_0 + \Delta\mathcal{C}) - \mathcal{R}(\mathcal{C}_0) \geq \Delta\mathcal{R}\}. \quad (2.10)$$

Это  $\mathcal{C}_0^{\mathcal{R}}$  соответствует минимальному  $\mathcal{C}_0$  из тех, при которых идентифицируются собственные тройки тренда, а это и есть  $\mathcal{C}_{max}$ .

Схема процедуры для вычисления  $\mathcal{C}_{max} = \mathcal{C}_0^{\mathcal{R}}$  приведена в алгоритме 2.4.

**Выбор  $\Delta\mathcal{C}$**  Напомним,  $\Delta\mathcal{C}$  используется для приближённого построения  $\mathcal{R}(\mathcal{C}_0)$  с тем, чтобы найти значения  $\mathcal{C}_0$ , при которых происходит идентификация отдельных собственных троек. Чем меньше оно будет, тем точнее будут найдены искомые  $\mathcal{C}_0$ . Ещё одна причина выбора малого  $\Delta\mathcal{C}$  состоит в том, что рассматривается приращение  $\mathcal{R}(\mathcal{C}_0)$  на интервале шириной  $\Delta\mathcal{C}$ . В случае, когда в него попадают несколько значений  $\mathcal{C}_0$ , при которых происходит идентификация некоторых собственных троек, это приращение будет отражать суммарный прирост, что может повлечь за собой ошибку (выбор слишком малого  $\mathcal{C}_{max}$ ).

Опыт применения данного метода оценки  $\mathcal{C}_{max}$  показывает, что достаточным является выбор  $\Delta\mathcal{C} = 0.01$ .



**АЛГОРИТМ 2.4. Вычисление  $\mathcal{C}_{max} = \mathcal{C}_0^{\mathcal{R}}$  на основе меры  $\mathcal{R}$**

**Входные данные:**  $F$ , содержащий тренд  $F^{(T)}$

**Результат:**  $\mathcal{C}_{max}$

**Параметры:**

- Основные параметры:  $\mathcal{C}_0^l, \mathcal{C}_0^h, \Delta\mathcal{C}, \Delta\mathcal{R}$ , см. (2.10)
- Параметры процедуры TREND:  $L, \omega_0$

**Процедура:**

1.  $\Omega = \{\mathcal{C}_0^l + k\Delta\mathcal{C}\}_{k=0}^M, \quad M = \left\lceil \frac{\mathcal{C}_0^h - \mathcal{C}_0^l}{\Delta\mathcal{C}} \right\rceil$
2.  $\forall \mathcal{C}_0 \in \Omega$ :
  - 1) Получим  $F_{\mathcal{A}}(\mathcal{C}_0)$ , применив процедуру TREND с пороговым значением  $\mathcal{C}_0$  к ряду  $F$  (алгоритм 2.2)
  - 2) Сосчитаем  $\mathcal{R}(\mathcal{C}_0) = \mathcal{R}(F, F_{\mathcal{A}}(\mathcal{C}_0))$  по (2.9)
3.  $\mathcal{C}_{max} = \mathcal{C}_0^{\mathcal{R}}$ , где  $\mathcal{C}_0^{\mathcal{R}}$  рассчитывается по (2.10)

**Выбор  $\Delta\mathcal{R}$**  Напомним, что  $\mathcal{R} \in [0, 1]$  для любого ряда. Выбор  $\Delta\mathcal{R}$  зависит от соотношения сигнал/шум, а также от предполагаемого ранга тренда  $\text{rank}(F^{(T)})$ . Чем меньше соотношение сигнал/шум или чем больше  $\text{rank}(F^{(T)})$ , тем меньше  $\Delta\mathcal{R}$  стоит выбирать.

Соотношение сигнал/шум влияет на выбор  $\Delta\mathcal{R}$ , потому что при большом шуме искажаются трендовые собственные вектора.

Значение  $\Delta\mathcal{R}$  зависит от ранга тренда  $\text{rank}(F^{(T)}) = d$ , так как приращение  $\mathcal{R}(0) - \mathcal{R}(\mathcal{C}_{max})$  в общем случае дробится на  $d$  изменений меры  $\mathcal{R}$  в точках  $\mathcal{C}_0$ , при которых происходит идентификация каждой из  $d$  собственных троек тренда.

Опыт применения данного метода оценки  $\mathcal{C}_{max}$  показывает, что данный способ выбора  $\mathcal{C}_{max}$  даёт хорошие результаты при  $0.05 \lesssim \Delta\mathcal{R} \lesssim 0.1$ .

### 2.6.3 Примеры поиска $\mathcal{C}_{max}$ с помощью меры $\mathcal{R}$

Приведём примеры поиска  $\mathcal{C}_{max}$  с помощью меры  $\mathcal{R}$ , как это описано в алгоритме 2.4. Эти примеры также демонстрируют схожесть меры  $\mathcal{R}$  и СКО. Зафиксируем следующие параметры:  $\mathcal{C}_0^l = 0, \mathcal{C}_0^h = 1, \Delta\mathcal{C} = 0.01, \Delta\mathcal{R} = 0.1$ .

**Пример 2.2** Возьмём ряд  $F$ :  $F = F^{(T)} + F^{(\varepsilon)}$ ,  $f_n^{(T)} = e^{0.01n}$ ,  $f_n^{(\varepsilon)} \sim NID(0, 5^2)$ ,  $N = 300$ . Возьмём  $L = 150$ . На графике периодограммы  $\Pi_F^N$  (рис. 2.7) видна граница интервала низких частот, которая равна приблизительно 0.025. Учитывая замечание 2.5, возьмём  $\omega_0 = 0.034 \approx \lceil 0.025L \rceil / L + 1/L$ . На рис. 2.7 изображены ряд, его периодограмма, а также графики  $\mathcal{R}(\mathcal{C}_0)$  и  $\text{СКО}(F_{\mathcal{A}}(\mathcal{C}_0), F^{(T)})$  (в логарифмическом масштабе) для двух взятых реализаций шума.

Видно, что для обеих реализаций ряда резкое изменение  $\mathcal{R}(\mathcal{C}_0)$  соответствует резкому изменению СКО и возникает из-за идентификации при этом  $\mathcal{C}_0$  единственной собственной тройки тренда. Мера  $\mathcal{R}$  позволяет найти это  $\mathcal{C}_0$ , которое составляет для одной реализации шума 0.98, а для другой 0.97.

**Пример 2.3**  $F$  :  $f_n^{(T)} = (n - 10)(n - 70)(n - 160)^2(n - 290)^2/10^{11}$ ,  $f_n^{(\varepsilon)} \sim NID(0, 5^2)$ ,  $N = 300$ ,  $L = 150$ ,  $\omega_0 = 0.034 \approx \lceil 0.025L \rceil / L + 1/L$ , см. рис. 2.8.

**Пример 2.4**  $F$  :  $f_n^{(T)} = e^{0.01n} + (n - 10)(n - 70)(n - 160)^2(n - 290)^2/10^{11}$ ,  $f_n^{(\varepsilon)} \sim NID(0, 5^2)$ ,  $N = 300$ ,  $L = 150$ ,  $\omega_0 = 0.03$ , см. рис. 2.9.

**Пример 2.5**  $F$  :  $f_n^{(T)} = e^{0.01n} + (n - 10)(n - 70)(n - 120)^2(n - 280)^2/10^{11}$ ,  $f_n^{(\varepsilon)} \sim NID(0, 2^2)$ ,  $N = 300$ . Рассмотрим результаты для одной реализации шума, но для разных  $L$ :  $L = 50$  и  $L = 150$ . Граница интервала НЧ равно приблизительно 0.025, что при  $L = 50$  соответствует  $\omega_0 = 0.034 \approx \lceil 0.025L \rceil / L + 1/L$ , при  $L = 150 - \omega_0 = 0.06$ . Результаты приведены на рис. 2.10.

**Пример 2.6** Рассмотрим теперь пример выделения тренда в присутствии периодической составляющей (э-м гармоники),  $F = F^{(T)} + F^{(P)} + F^{(\varepsilon)}$ ,  $f_n^{(T)} = (n - 10)(n - 70)(n - 160)^2(n - 290)^2/10^{11}$ ,  $f_n^{(P)} = e^{0.01n} \sin(2\pi n/12)$ ,  $f_n^{(\varepsilon)} \sim NID(0, 5^2)$ ,  $N = 300$ . Выберем  $L = 144$  (кратным периоду 12),  $\omega_0 = 0.035 \approx \lceil 0.025L \rceil / L + 1/L$ , см. рис. 2.11.

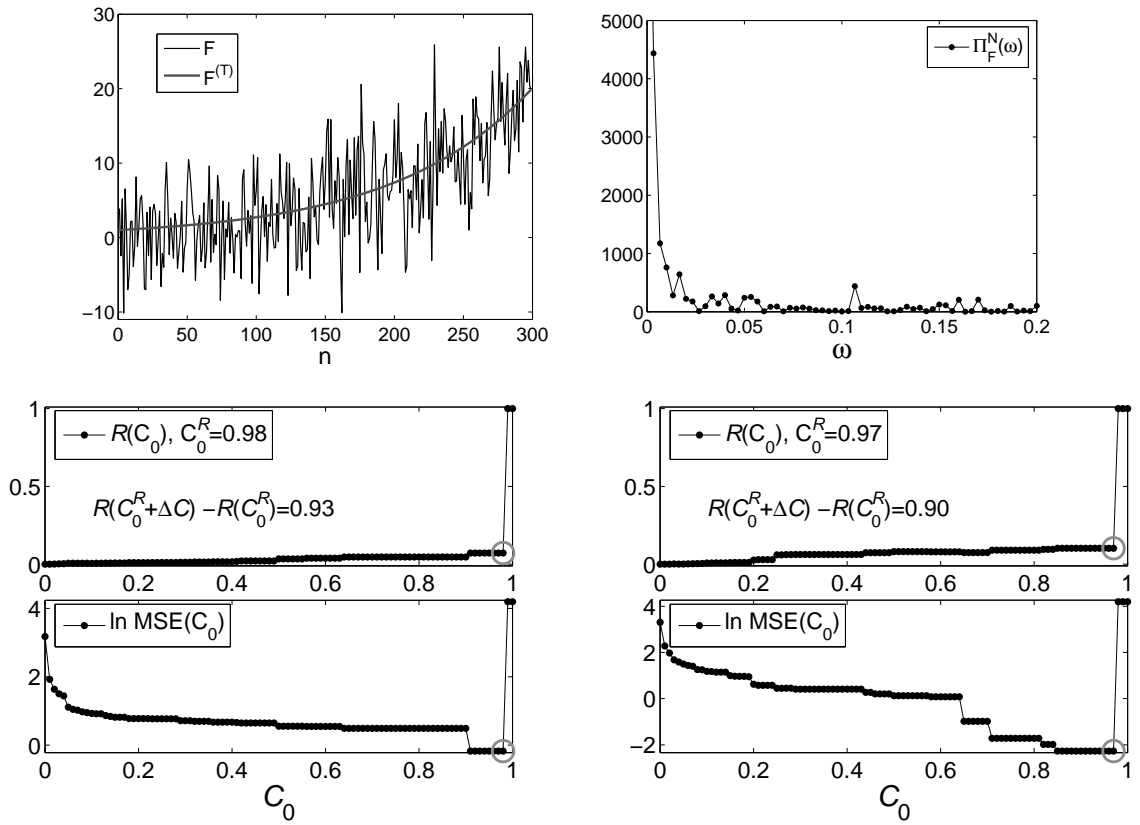


Рис. 2.7: Пример 2.2, значения  $\mathcal{R}(C_0)$  приведены для двух реализаций шума.

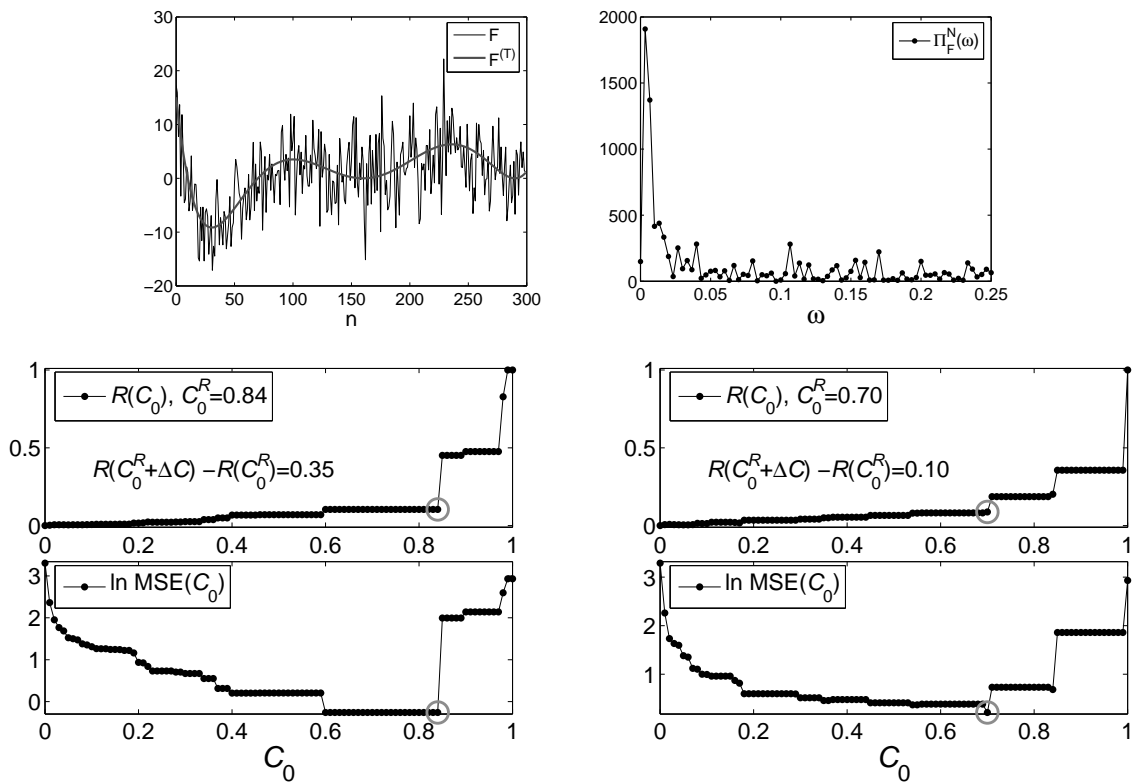


Рис. 2.8: Пример 2.3, значения  $\mathcal{R}(C_0)$  приведены для двух реализаций шума.

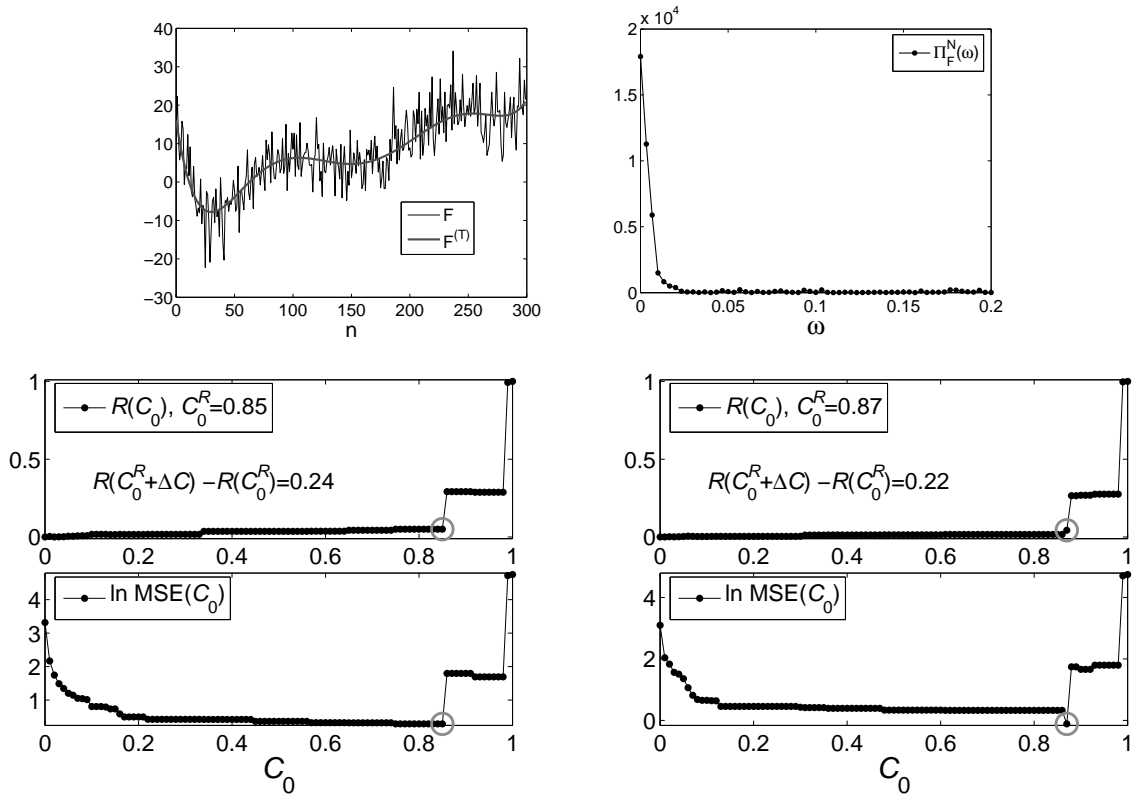


Рис. 2.9: Пример 2.4, значения  $\mathcal{R}(C_0)$  приведены для двух реализаций шума.

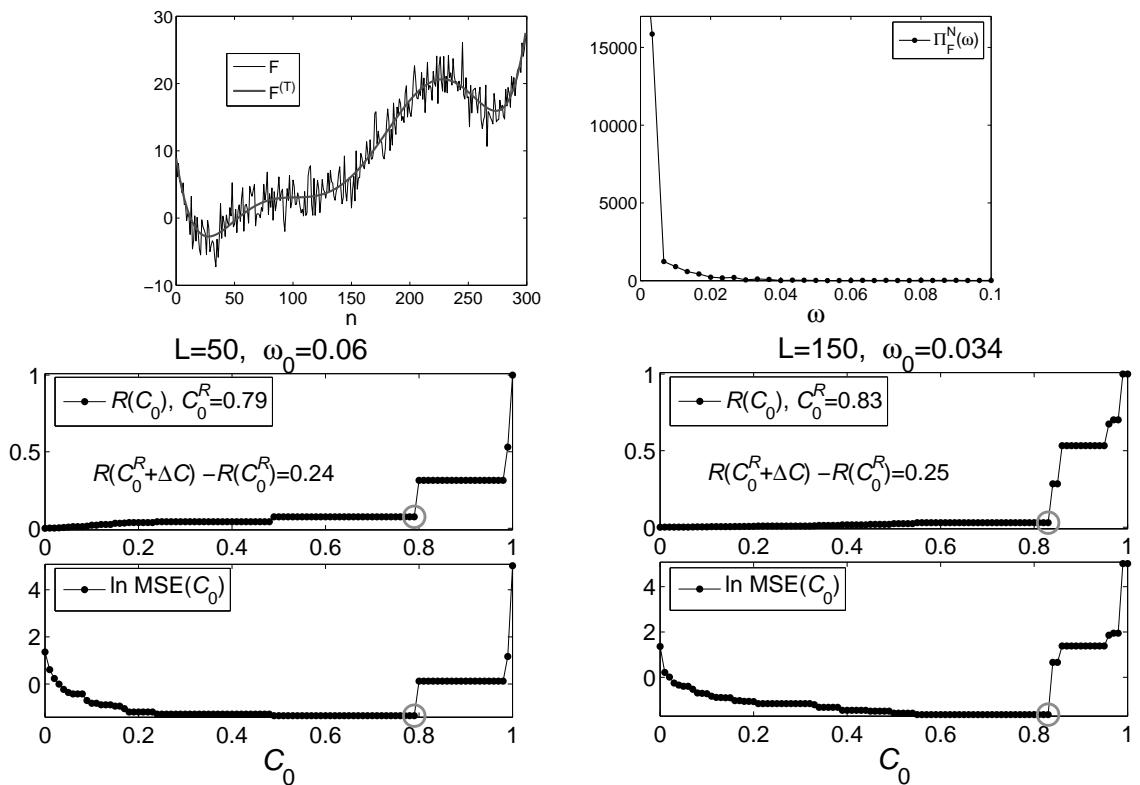


Рис. 2.10: Пример 2.5, значения  $\mathcal{R}(C_0)$  приведены для разных  $L$ .

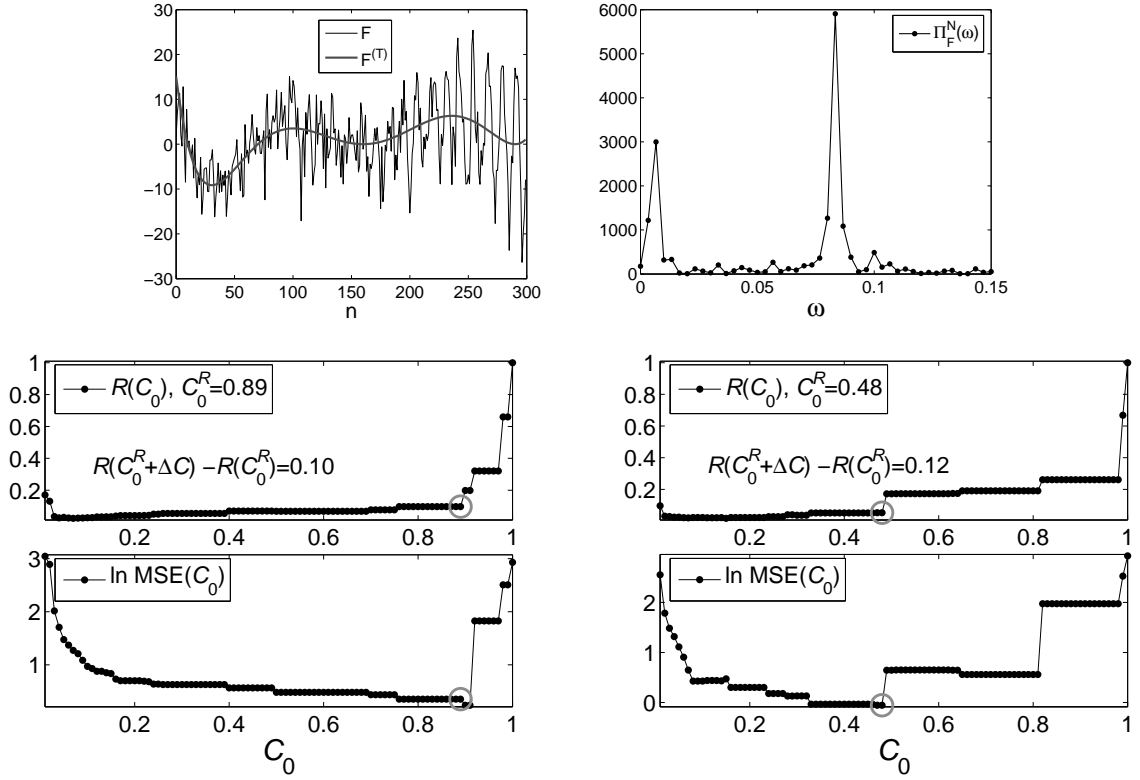


Рис. 2.11: Пример 2.6, выделения тренда в присутствии периодической составляющей, значения  $\mathcal{R}(C_0)$  приведены для двух реализаций шума.

#### 2.6.4 Сравнение $C_0^{(T)}$ , $C_{max}$ и $C_0^{opt}$ для экспоненциального тренда

Экспоненциальный тренд является одним из немногих примеров, для которого возможен аналитический расчёт  $C_0^{(T)}$ . Сделаем это, а после возьмём пример ряда из раздела 2.4, посчитаем для него предложенным способом  $C_0^{(T)}$  и сравним результаты со средним  $C_0^R$  и  $C_0^{opt}$ .

**Утверждение 2.4** Пусть  $F = (f_0, \dots, f_{N-1})$  — экспоненциальный ряд,  $f_n = e^{\alpha n}$ ,  $\alpha > 0$ , тогда коэффициенты Фурье последовательности элементов его собственного вектора  $U \in \mathbb{R}^L$  равны:

$$c_0 = \frac{C(e^{\alpha L} - 1)}{L(1 - e^{-\alpha})}, \quad c_{L/2} = \frac{C((-e)^{\alpha L} - 1)}{L(e^{-\alpha} + 1)}, \quad \text{где } C = \sqrt{\frac{1 - e^{-2\alpha}}{e^{2\alpha L} - 1}},$$

$$\forall k, 0 < k < L/2: \quad c_k = A_k (\cos \beta_k - e^\alpha), \quad s_k = A_k \sin \beta_k, \quad (2.11)$$

$$\text{где } \beta_k = 2\pi k/L, \quad A_k = \frac{C(1 - e^{\alpha L})}{L(e^\alpha + e^{-\alpha} - 2 \cos \beta_k)}.$$

*Доказательство.* Единственный собственный вектор  $U = (u_1, \dots, u_L)^T$ , соответствующий  $F$ , тоже имеет экспоненциальную форму с тем же  $\alpha$  (см. 1.3.1), т. е.  $u_n = Ce^{\alpha n}$ ,  $\alpha > 0$ . Следуя требованию  $\|U\| = 1$ , посчитаем по  $\sum_{n=1}^L (Ce^{\alpha n})^2 = 1$  требуемое значение для  $C$ .

Получим теперь выражения для коэффициентов Фурье последовательности  $u_1, \dots, u_L$ . Коэффициенты  $c_0$  и  $c_{L/2}$  (последний определён только для чётного  $L$ ) имеют простой вид:

$$\begin{aligned} c_0 &= \frac{1}{L} \sum_{n=1}^L u_n = \frac{C}{L} \sum_{n=1}^L (e^\alpha)^n = \frac{C(e^{\alpha L} - 1)}{L(1 - e^{-\alpha})}, \\ c_{L/2} &= \frac{C}{L} \sum_{n=1}^L (-e^\alpha)^n = \frac{C((-e)^\alpha - 1)}{L(e^{-\alpha} + 1)}. \end{aligned} \quad (2.12)$$

Получим теперь выражения для  $c_k$  и  $s_k$ ,  $0 < k < L/2$ . Вычислим при  $\beta_k = 2\pi k/L$  следующую сумму:

$$\begin{aligned} \sum_{n=0}^{L-1} (e^{\alpha+i\beta_k})^n &= \frac{e^{(\alpha+i\beta_k)L} - 1}{e^{\alpha+i\beta_k} - 1} = \frac{e^{\alpha L} - 1}{e^\alpha (\cos \beta_k + i \sin \beta_k) - 1} = \\ &= \frac{(e^{\alpha L} - 1)(\cos \beta_k - e^{-\alpha} - i \sin \beta_k)}{(e^\alpha (\cos \beta_k + i \sin \beta_k) - 1)(\cos \beta_k - e^{-\alpha} - i \sin \beta_k)} = \\ &= \frac{e^{\alpha L} - 1}{e^\alpha + e^{-\alpha} - 2 \cos \beta_k} [\cos \beta_k - e^{-\alpha} - i \sin \beta_k]. \end{aligned} \quad (2.13)$$

Таким образом, пользуясь (2.13), получаем, что:

$$\begin{aligned} \frac{C}{L} \sum_{n=1}^L (e^{\alpha+i\beta_k})^n &= \frac{C}{L} e^{\alpha+i\beta_k} \sum_{n=0}^{L-1} (e^{\alpha+i\beta_k})^n = \\ &= \frac{CL^{-1}(1 - e^{\alpha L})}{e^\alpha + e^{-\alpha} - 2 \cos \beta_k} [\cos \beta_k - e^\alpha + i \sin \beta_k]. \end{aligned} \quad (2.14)$$

Получим коэффициенты  $c_k$ ,  $s_k$ , как комплексную и вещественную части суммы (2.14):

$$\begin{aligned} c_k &= A_k (\cos \beta_k - e^\alpha), \quad s_k = A_k \sin \beta_k, \\ \text{где } A_k &= \frac{C(1 - e^{\alpha L})}{L(e^\alpha + e^{-\alpha} - 2 \cos \beta_k)}. \end{aligned} \quad (2.15)$$

■

Таким образом, для собственного вектора  $U$  экспоненциального ряда значение  $\mathcal{C}(U)$  выражается через  $L$ ,  $\alpha$ ,  $\omega_0$  следующим образом:

$$\mathcal{C}(U) = \frac{2c_0^2 + \sum_{k/L \leq \omega_0} (c_k^2 + s_k^2)}{2c_0^2 + \sum_k (c_k^2 + s_k^2) + 2c_{L/2}^2}, \quad (2.16)$$

где коэффициенты  $c_k$ ,  $s_k$  рассчитываются по (2.11).

Для экспоненциального тренда, которому соответствует собственный вектор  $U$ ,  $\mathcal{C}_0^{(T)} = \mathcal{C}(U)$ . В случае, когда  $F = F^{(T)}$ , справедливо следующее равенство:  $\mathcal{C}_{max} = \mathcal{C}_0^{(T)} = \mathcal{C}(U)$ .

**Численный пример** Рассмотрим ряд  $F = F^{(T)} + F^{(\varepsilon)}$ :  $f_n = f_n^{(T)} + f_n^{(\varepsilon)}$ ,  $f_n^{(T)} = e^{\alpha n}$ ,  $f_n^{(\varepsilon)} \sim NID(0, \sigma^2)$  при  $N = 47$ ,  $L = 24$ ,  $\omega_0 = 0.05$ . Рассчитаем значение  $\mathcal{C}_0^{(T)}$  для тренда  $F^{(T)}$  с помощью формулы (2.16), предполагая, что тренду соответствует первая собственная тройка. Результаты см. в табл. 2.1.

Для разных уровней шума  $\sigma$  смоделируем  $R = 10^3$  реализаций шума  $F^{(\varepsilon)}$ , получив таким образом  $R$  рядов  $F = F^{(T)} + F^{(\varepsilon)}$ , и сосчитаем для каждого их них  $\mathcal{C}_{max} = \mathcal{C}_0^{\mathcal{R}}$ , пользуясь процедурой TRRMEAS с параметрами  $\mathcal{C}_0^l = 0.5$ ,  $\mathcal{C}_0^j = 1$ ,  $\Delta\mathcal{C} = 0.001$  (алгоритм 2.5). После этого по полученной выборке значений  $\mathcal{C}_0^{\mathcal{R}}$  посчитаем среднее  $\bar{\mathcal{C}}_0^{\mathcal{R}}$ . В качестве  $\Delta\mathcal{R}$  было взято 0.2, так как известно, что тренд содержит только одну компоненту, и уровень шума достаточно мал. Результаты представлены в табл. 2.1.

Приведём также значение  $\mathcal{C}_0^{opt}$ , найденное при тех же параметрах с тем же шагом  $\Delta\mathcal{C} = 0.001$  и  $R = 10^3$  с помощью алгоритма 2.3.

Таблица 2.1:  $\mathcal{C}_0^{(T)}$ ,  $\bar{\mathcal{C}}_0^{\mathcal{R}}$  и  $\mathcal{C}_0^{opt}$  для экспоненциального тренда при разных  $\alpha$

| $\alpha$  | 0.01    | 0.02    | 0.03    | 0.04    | 0.05    |
|---|---------|---------|---------|---------|---------|
| $\mathcal{C}_0^{(T)}$                             | 0.99815 | 0.99272 | 0.98400 | 0.97246 | 0.95864 |
| $\mathcal{C}_0^{opt}, \sigma = 0$                 | 0.998   | 0.992   | 0.984   | 0.972   | 0.958   |
| $\mathcal{C}_0^{opt}, \sigma = 0.1$               | 0.996   | 0.989   | 0.979   | 0.965   | 0.948   |
| $\mathcal{C}_0^{opt}, \sigma = 0.8$               | 0.950   | 0.932   | 0.898   | 0.871   | 0.831   |
| $\bar{\mathcal{C}}_0^{\mathcal{R}}, \sigma = 0.1$ | 0.998   | 0.992   | 0.983   | 0.971   | 0.958   |
| $\bar{\mathcal{C}}_0^{\mathcal{R}}, \sigma = 0.8$ | 0.995   | 0.990   | 0.982   | 0.971   | 0.957   |
| $\bar{\mathcal{C}}_0^{\mathcal{R}}, \sigma = 1.6$ | 0.979   | 0.983   | 0.978   | 0.968   | 0.955   |

Видно, что при  $\sigma = 0$  с точностью до второго знака выполняется  $\mathcal{C}(U) = \mathcal{C}_0^{opt}$ , а при малом  $\sigma$  верно  $\bar{\mathcal{C}}_0^{\mathcal{R}} \cong \mathcal{C}_0^{(T)}$ . Это подтверждает тот факт, что  $\mathcal{C}_{max}$ , найденный с помощью меры  $\mathcal{R}$ , аппроксимирует  $\mathcal{C}_0^{(T)}$ .

Значение  $\mathcal{C}_0^{\mathcal{R}}$  при  $\sigma = 0$  не рассчитывалось, так как применение процедуры TRRMEAS осмысленно только при наличии остатка.

Стоит отметить, что при  $\sigma > 0$  выполняется  $\mathcal{C}_0^{opt} < \bar{\mathcal{C}}_0^{\mathcal{R}}$ , что объясняется тем, что  $\mathcal{C}_0^{opt}$  подбирается так, чтобы аппроксимация  $F^{(T)}$  была наилучшей. Вследствие приближённой разделимости происходит систематическое искажение трендовых собственных векторов, из-за чего аппроксимация по группе  $\mathcal{I}^{(T)} = \{1\}$  может быть улучшена добавлением в группу  $\mathcal{I}_A$  каких-то компонент шума, что достигается при меньшем  $\mathcal{C}_0$ . В то же время,  $\mathcal{C}_0^{\mathcal{R}}$  для этих рядов выбирается с тем, чтобы идентифицировать только трендовые компоненты  $\mathcal{I}^{(T)}$ .

### 2.6.5 Описание процедуры TRRMEAS автоматического выделения тренда ряда

Опишем теперь полностью процедуру автоматического выделения тренда ряда  $F$  (обозначим её TRRMEAS), которая использует метод низких частот, оценивая  $\mathcal{C}_0$  с помощью меры  $\mathcal{R}$ .

**АЛГОРИТМ 2.5. Процедура TRRMEAS: автоматическое выделение тренда ряда с выбором порогового значения с помощью меры  $\mathcal{R}$**

**Входные данные:**  $F$  — ряд, содержащий тренд

**Результат:**  $\mathcal{C}_0^{\mathcal{R}}$ ,  $\mathcal{I}_A$  — группа идентифицированных компонент,  $F_A$  — полученная аппроксимация тренда

**Параметры:**

- Параметры процедуры TREND:  $L$ ,  $\omega_0$
- Параметры перебора  $\mathcal{C}_0$ :  $\mathcal{C}_0^l$ ,  $\mathcal{C}_0^h$ ,  $\Delta\mathcal{C}$
- $\Delta\mathcal{R}$

**Процедура:**

1. Найдём  $\mathcal{C}_0^{\mathcal{R}}$  с помощью алгоритма 2.4, взяв в качестве параметров  $\mathcal{C}_0^l$ ,  $\mathcal{C}_0^h$ ,  $\Delta\mathcal{C}$ ,  $\Delta\mathcal{R}$
2. Вычислим  $\mathcal{I}_A$ ,  $F_A$  с помощью процедуры TREND с параметрами  $L$ ,  $\omega_0$  и пороговым значением  $\mathcal{C}_0 = \mathcal{C}_0^{\mathcal{R}}$  (алгоритм 2.2)



## 2.7 Случай известной модели шума

Рассмотрим случай, когда остаток представляет собой реализацию случайной последовательности, распределение которой известно. Тогда можно оценить такое минимальное значение  $\mathcal{C}_0$ , при котором выделяется не слишком много нетрендовых собственных троек и сузить область поиска  $\mathcal{C}_0$ , рассматривая только бóльшие значения.

Пусть  $F = F^{(T)} + F^{(\varepsilon)}$ ,  $F^{(T)}$  — ряд конечного ранга,  $F^{(\varepsilon)}$  — реализация случайной последовательности (шум) и известно её распределение  $\mathcal{E}$ . Если предполагать разделимость  $F^{(T)}$  и  $F^{(\varepsilon)}$ , то часть сингулярного разложения траекторной матрицы  $F$ , которая соответствует составляющей  $F^{(\varepsilon)}$ , будет равна сингулярному разложению траекторной матрицы ряда  $F^{(\varepsilon)}$ , что можно выразить как (см. утверждение 1.2):

$$\text{SVD ряда } F = \text{SVD тренда } F^{(T)} + \text{SVD остатка } F^{(\varepsilon)}.$$

Пусть есть сингулярное разложение остатка, применим к нему процедуру идентификации с  $\mathcal{C}_0$ , обозначим набор идентифицированных компонент как  $\mathcal{I}^{(\varepsilon)}(\mathcal{C}_0)$ . Обозначим через  $F^{(\varepsilon)}(\mathcal{C}_0)$  ряд, восстановленный по  $\mathcal{I}^{(\varepsilon)}(\mathcal{C}_0)$ . В данном разделе изучим следующую ошибку идентификации:

$$\mathcal{M}(F^{(\varepsilon)}(\mathcal{C}_0)) = \|F^{(\varepsilon)}(\mathcal{C}_0)\|^2. \quad (2.17)$$

Предполагая разделимость  $F^{(T)}$  и  $F^{(\varepsilon)}$ , можно утверждать, что если применять процедуру TREND с  $\mathcal{C}_0 \leq \mathcal{C}_{max}$  к ряду  $F = F^{(T)} + F^{(\varepsilon)}$ , то собственные тройки, соответствующие составляющей  $F^{(\varepsilon)}$ , совпали бы с собственными тройками разложения ряда  $F^{(\varepsilon)}$  (результат применения SSADEC к  $F^{(\varepsilon)}$ ). Группа ошибочно идентифицированных троек совпала бы с  $\mathcal{I}^{(\varepsilon)}(\mathcal{C}_0)$ , восстановленный по ним ряд был бы равен  $F^{(\varepsilon)}(\mathcal{C}_0)$  и ошибка восстановления была бы равна  $\|F_{\mathcal{A}}(\mathcal{C}_0) - F^{(T)}\|^2 = \mathcal{M}(F^{(\varepsilon)}(\mathcal{C}_0))$ .

Конечно, при конечном  $N$  точная разделимость чаще всего не наблюдается, но для широкого класса трендов и остатков выполняются условия асимптотической разделимости (при  $N, L \rightarrow \infty$ ). При конечном  $N$  для них наблюдается приближённая разделимость, при которой  $\|F_{\mathcal{A}}(\mathcal{C}_0) - F^{(T)}\|^2 \approx \mathcal{M}(F^{(\varepsilon)}(\mathcal{C}_0))$ .

Определим значение  $\mathcal{C}_{min}(\mathcal{E})$  для модели шума  $\mathcal{E}$  следующим образом. Рассмотрим процедуру TREND с  $\mathcal{C}_0$  как функцию  $\mathfrak{T}_{\mathcal{C}_0}$ :

$$\mathfrak{T}_{\mathcal{C}_0} : \mathbb{R}^N \rightarrow \mathbb{R}^N, \quad \mathfrak{T}_{\mathcal{C}_0}(F) = F_{\mathcal{A}}.$$

Введём для  $\mathcal{E}$  случайную величину  $\mathcal{M}_0^\mathcal{E}(\mathcal{C}_0)$ , как функцию от случайной величины  $\xi$ , зависящую от параметра  $\mathcal{C}_0$ :

$$\mathcal{M}^\mathcal{E}(\mathcal{C}_0) = \mathcal{M}(\mathfrak{T}_{\mathcal{C}_0}(\xi)), \quad \xi \sim \mathcal{E}.$$

Это корректно, так как функция  $\mathfrak{T}_{\mathcal{C}_0}$  измерима. Тогда

$$\mathcal{C}_{min}(\mathcal{E}) = \min\left\{\mathcal{C}_0 : P(\mathcal{M}^\mathcal{E}(\mathcal{C}_0) \geq \mathcal{M}_0) = 1 - \gamma\right\}, \quad (2.18)$$

где  $\mathcal{M}_0$  — максимально-допустимый уровень ошибки  $\mathcal{M}^\mathcal{E}(\mathcal{C}_0)$ , а  $\gamma$  — уровень надёжности.

$\mathcal{C}_{min}(\mathcal{E})$  — это такое значение  $\mathcal{C}_0$ , при котором невелика вероятность того, что ошибка, возникающая из-за неправильно идентифицированных компонент шума, превышает заданный уровень  $\mathcal{M}_0$ .

Зная значение  $\mathcal{C}_{min}(\mathcal{E})$ , можно сузить область поиска  $\mathcal{C}_0$  (и, соответственно, область поиска  $\mathcal{C}_{max}$ ) до  $[\mathcal{C}_{min}, 1]$ . Например, в процедуре TRRMEAS (см. алгоритм 2.5) следует брать  $\mathcal{C}_0^l = \mathcal{C}_{min}$ .

### Вычисление $\mathcal{C}_{min}$ с помощью моделирования

Для того, чтобы сосчитать  $\mathcal{C}_{min}(\mathcal{E})$  для модели шума  $\mathcal{E}$ , необходимо вычислить зависимость  $\mathcal{M}$  от  $\mathcal{C}_0$ . Получение аналитического вида  $\mathcal{C}_{min}(\mathcal{E})$  является сложной задачей даже для простого случая, когда  $\mathcal{E} = NID$  (остаток — белый нормальный шум). Однако возможно оценить  $\mathcal{C}_{min}(\mathcal{E})$  с помощью статистического моделирования. Покажем, как это может быть сделано.

Зная  $\mathcal{E}$ , рассчитаем эмпирическую функцию распределения случайной величины  $\mathcal{M}^\mathcal{E}(\mathcal{C}_0)$ , взяв  $R$  реализаций  $\{F^{(\varepsilon,r)}\}_{r=1}^R$  случайной величины  $\xi \sim \mathcal{E}$ :

$$\widehat{\mathcal{F}}(x, \mathcal{C}_0) = \frac{Q}{R}, \quad Q = \#\{F^{(\varepsilon,r)} : \mathcal{M}(F^{(\varepsilon,r)}(\mathcal{C}_0)) < x\}, \quad (2.19)$$

где  $\#$  обозначает мощность множества. Причём

$$\widehat{\mathcal{F}}(x, \mathcal{C}_0) \xrightarrow{R \rightarrow \infty} \mathcal{F}(x, \mathcal{C}_0).$$

где  $\mathcal{F}(x, \mathcal{C}_0)$  — функция распределения случайной величины  $\mathcal{M}^\mathcal{E}(\mathcal{C}_0)$ , параметром которой выступает  $\mathcal{C}_0$ . Тогда

$$P(\mathcal{M}^\mathcal{E}(\mathcal{C}_0) \geq x) \approx 1 - \widehat{\mathcal{F}}(x, \mathcal{C}_0), \quad \widehat{\mathcal{F}}(\mathcal{M}_0, \mathcal{C}_0) = \gamma.$$

Рассчитаем предварительно  $\widehat{\mathcal{F}}^{-1}(\gamma, \mathcal{C}_0)$ , где  $\widehat{\mathcal{F}}^{-1}(\widehat{\mathcal{F}}(x, \mathcal{C}_0), \mathcal{C}_0) = x$ . Это такое значение, что  $\mathcal{M}^\mathcal{E}(\mathcal{C}_0)$  не превосходит его с вероятностью  $\gamma$ . Рассчитаем его

эмпирически, по выборке  $\{\mathcal{M}(F^{(\varepsilon,r)}(\mathcal{C}_0))\}_{r=1}^R$ . Отбросим  $R(1 - \gamma)$  наибольших значений и возьмём максимальное из оставшихся. Тогда

$$\hat{\mathcal{C}}_{min}(\mathcal{E}) = \min\{\mathcal{C}_0 : \hat{\mathcal{F}}^{-1}(\gamma, \mathcal{C}_0) \leq \mathcal{M}_0\}. \quad (2.20)$$

Пользуясь (2.20), построим процедуру поиска  $\mathcal{C}_{min}(\mathcal{E})$ , её описание приведено в алгоритме 2.6. Параметрами процедуры будут параметры процедуры TREND, а также  $\mathcal{M}_0$  и  $\gamma$ .

**Выбор  $\mathcal{M}_0$**  Стоит выбирать  $\mathcal{M}_0$ , как долю от  $\hat{\mathcal{F}}^{-1}(\gamma, \mathcal{C}_0)$  при  $\mathcal{C}_0 = 0$ , так как при этом результаты не будут зависеть от уровня шума.

Например, выбор  $\mathcal{M}_0 = \hat{\mathcal{F}}^{-1}(\gamma, 0) \cdot 10\%$  соответствует тому, что очень мала вероятность<sup>7</sup> того, что выделенный из исходного  $F$  ряд  $F_{\mathcal{A}} = F^{(T)} + F^{(\varepsilon)}(\mathcal{C}_{min})$ , будет ухудшен (в среднеквадратичном) больше чем на 10% по сравнению со случаем, когда все компоненты шума идентифицированы, т. е.  $F_{\mathcal{A}} = F$ .

**Замечание 2.9** Ещё раз заметим, что оценивать ошибку идентификации, рассматривая остаток, можно только при наличии точной делимости тренда и остатка, а когда остаток — реализация случайного шума, делимость при конечном  $N$  будет только приближённой. Тем не менее, на практике, для достаточно большого  $N$ , можно хорошо разделить сумму тренда и шума.

**Пример вычисления  $\hat{\mathcal{C}}_{min}(\mathcal{E})$  для белого нормального шума** Возьмём модель  $\mathcal{E} = NID(0, 1)$  и при следующих параметрах оценим  $\mathcal{C}_{min}(\mathcal{E})$ , пользуясь алгоритмом 2.6:  $N = 47$ ,  $L = 24$ ,  $\omega_0 = 0.05$ ,  $R = 10^4$ ,  $\mathcal{M}_0 = \hat{\mathcal{F}}^{-1}(\gamma, 0) \cdot 10\%$ . Рассматриваем шум только при  $\sigma = 1$ , так как при таком выборе  $\mathcal{M}_0$  значение  $\sigma$  не существенно при вычислении  $\hat{\mathcal{C}}_{min}(\mathcal{E})$ .

График для  $\hat{\mathcal{F}}^{-1}(0.95, \mathcal{C}_0)$ , нормированный так чтобы  $\hat{\mathcal{F}}^{-1}(0.95, 0) = 1$ , приведен на рис. 2.12. Видно, что при таких параметрах  $\hat{\mathcal{C}}_{min}(\mathcal{E}) \approx 0.54$ .

<sup>7</sup>Эта вероятность не превышает  $1 - \gamma$

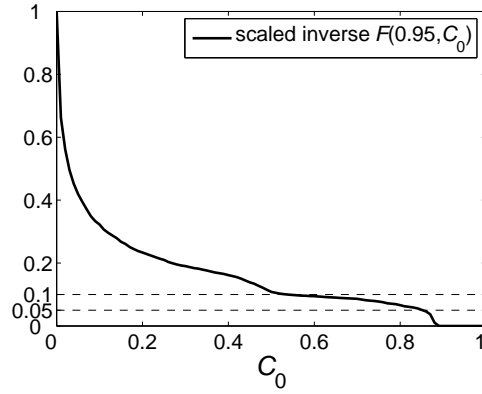


Рис. 2.12:  $\widehat{\mathcal{F}}^{-1}(0.95, \mathcal{C}_0)$  для  $NID(0, 1)$  с  $N = 47$  при  $L = 24$ ,  $\omega_0 = 0.05$ ,  $R = 10^4$  (значения нормированы так, чтобы  $\widehat{\mathcal{F}}^{-1}(0.95, 0) = 1$ ).

#### АЛГОРИТМ 2.6. Оценка $\mathcal{C}_{min}(\mathcal{E})$ для модели шума $\mathcal{E}$

**Входные данные:**  $N$  — длина ряда,  $\mathcal{E}$  — закон распределения шума

**Результат:**  $\widehat{\mathcal{C}}_{min}(\mathcal{E})$

**Параметры:**

- Параметры TREND:  $L$  ( $1 < L \leq \lceil N/2 \rceil$ ),  $\omega_0 \in (0, 0.5)$
- $\mathcal{C}_0^l, \mathcal{C}_0^r, \Delta\mathcal{C}, R$  — количество реализаций шума,  $\mathcal{M}_0$  — максимально допустимая ошибка,  $\gamma$  — уровень надёжности

**Процедура:**

1. Смоделируем  $R$  раз шум  $F^{(\varepsilon)}$ :  $\{F^{(\varepsilon, r)}\}_{r=1}^R$
2.  $\Omega = \{\mathcal{C}_0^l + k\Delta\mathcal{C}\}_{k=0}^M$ ,  $M = \left\lceil \frac{\mathcal{C}_0^r - \mathcal{C}_0^l}{\Delta\mathcal{C}} \right\rceil$
3. Для каждого  $r = 1, \dots, R$ ,  $\mathcal{C}_0 \in \Omega$ 
  - применим процедуру TREND с параметрами  $L, \omega_0, \mathcal{C}_0$  к ряду  $F^{(\varepsilon, r)}$ , получив  $F^{(\varepsilon, r)}(\mathcal{C}_0)$  (алгоритм 2.2)
  - посчитаем  $\mathcal{M}(F^{(\varepsilon, r)}(\mathcal{C}_0))$  по (2.17)
4. Для каждого  $\mathcal{C}_0 \in \Omega$  по выборке  $\{\mathcal{M}(F^{(\varepsilon, r)}(\mathcal{C}_0))\}_{r=1}^R$  посчитаем  $\widehat{\mathcal{F}}^{-1}(\gamma, \mathcal{C}_0)$
5. Вычислим по  $\widehat{\mathcal{F}}^{-1}(\gamma, \mathcal{C}_0)$  оценку  $\widehat{\mathcal{C}}_{min}(\mathcal{E})$  пользуясь (2.20)

**Пример вычисления  $\widehat{\mathcal{C}}_{min}(\mathcal{E})$  для белого нормального шума с изменяющейся дисперсией** Вычислим теперь  $\widehat{\mathcal{C}}_{min}(\mathcal{E})$  модели шума  $\mathcal{E}$  с изменяющейся дисперсией:  $F^{(\varepsilon)} = (f_0^{(\varepsilon)}, \dots, f_{N-1}^{(\varepsilon)})$ ,  $f_n^{(\varepsilon)} = e^{\alpha n} \varepsilon_n$ , где  $\varepsilon_0, \dots, \varepsilon_{N-1}$  — реализация белого нормального шума  $NID(0, 1)$ , т. е. стандарт шума меняется пропорционально экспоненциальной функции с показателем  $\alpha$ . Рассмотрим два примера:  $F_{47}$  и  $F_{100}$  длиной 47 и 100, соответственно. Для  $F_{47}$  возьмём  $L = 24$ ,  $\alpha = 0, 0.02, 0.05$ . Для  $F_{100}$  возьмём  $L = 50$  и такие значения  $\alpha$ , что равны последние элементы рядов  $F_{47}$ ,  $F_{100}$ :  $f_{99}^{100} = f_{46}^{47}$ , что составляет  $\alpha = 0, 0.0093, 0.0232$ . Это позволяет определить влияние на результаты увеличения  $N, L$  за вычетом изменения поведения ряда (как если бы был взят тот же ряд, но с большим разрешением).

Графики для  $\widehat{\mathcal{F}}^{-1}(0.95, \mathcal{C}_0)$ , нормированные так чтобы  $\widehat{\mathcal{F}}^{-1}(0.95, 0) = 1$  приведены на рис. 2.13.

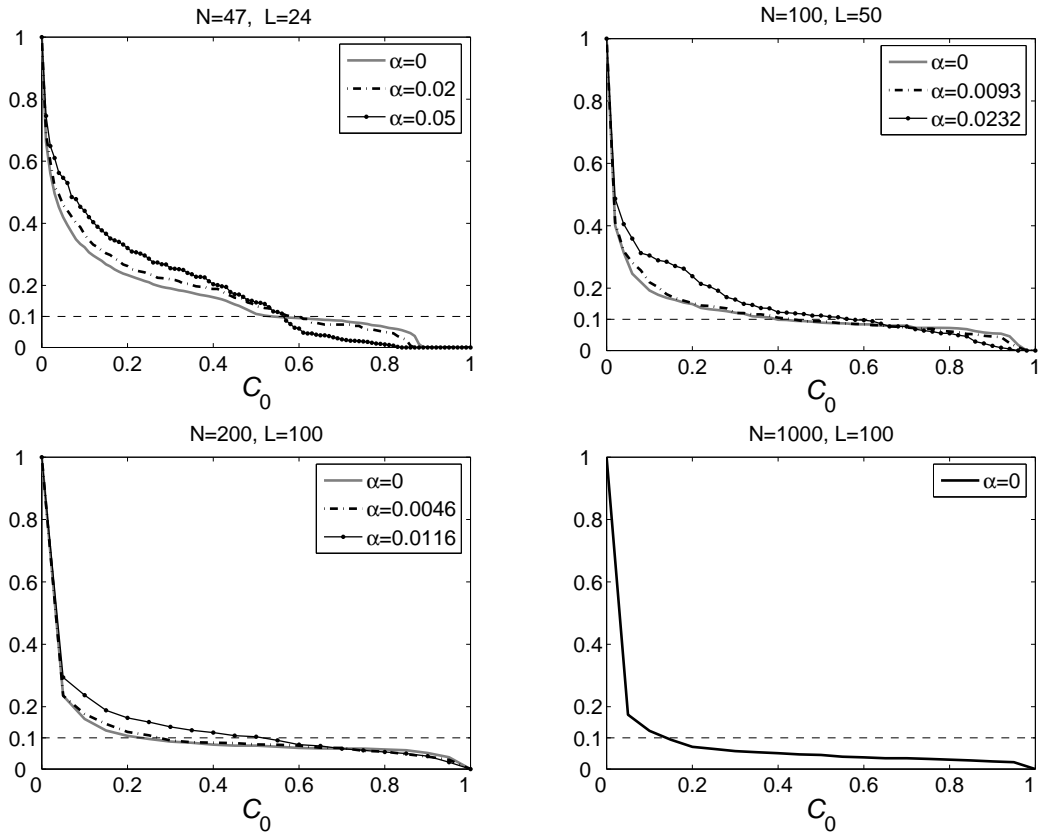


Рис. 2.13:  $\widehat{\mathcal{F}}^{-1}(0.95, \mathcal{C}_0)$  для белого нормального шума различной длины  $N$ ,  $\sigma$  которого меняется пропорционально экспоненте с  $\alpha$  (значения нормированы так, чтобы  $\max_{\mathcal{C}_0} \widehat{\mathcal{F}}^{-1} = 1$ ).

На представленных графиках видно, что, во-первых, при увеличении длины ряда  $N$  уменьшается рассчитанное значение  $\mathcal{C}_{min}(\mathcal{E})$ .

Во-вторых, изменение  $\alpha$  незначительно влияет на значение  $\mathcal{C}_{min}(\mathcal{E})$ . Это

позволяет предположить, что незначительное изменение дисперсии шума<sup>8</sup> не меняет  $\hat{C}_{min}(\mathcal{E})$ , что подтверждается и другими экспериментами. Слабое изменение дисперсии шума часто наблюдается в действительности, а сделанный вывод позволяет не принимать это во внимание при расчёте  $C_{min}(\mathcal{E})$ .

## 2.8 Примеры выделения трендов реальных рядов

Приведём в данном разделе примеры применения процедуры TRRMEAS к некоторым реальным рядам. Везде далее будем использовать следующие параметры:  $C_0^l = 0.5$ ,  $C_0^r = 1$ ,  $\Delta\mathcal{C} = 0.01$ .

### 2.8.1 Исследование уровня экспрессии гена

Данный раздел содержит пример из совместного исследования Санкт-Петербургского государственного университета и Stony Brook University, New York, при поддержке грантов CRDF № RUB1-1556-ST-05, № RUB1-1643-ST-06. Исследовалось пространственное распределение уровня экспрессии различных генов в эмбрионе плодовой мушки дрозофилы.

Данные и способ их получения описаны в [44]. Отметим, что это пространственные ряды, содержащие некоторую медленно меняющуюся составляющую и сложный шум с изменяющейся дисперсией. Рассмотрим сейчас пример одного ряда (рис. 2.14, слева сверху), соответствующего гену Hunchback.

Длина ряда  $N = 1679$ . Возьмём  $L = 300$ , такое  $L$  позволяет достаточно хорошо отделить тренд от шума и отразить в трендовых компонентах все особенности поведения тренда.

Рассмотрим периодограмму ряда  $\Pi_F^N$ , интересующая часть графика которой изображена на рис. 2.14 (справа снизу). Видна граница интервала низких частот, приходящаяся примерно на 0.004. Принимая во внимание замечание 2.5, возьмём  $\omega_0 = \lceil 0.004L \rceil / L + 1/L = 0.01$ .

Применим к ряду процедуру TRRMEAS с  $L = 300$ ,  $\omega_0 = 0.01$  и разными  $\Delta\mathcal{R}$ : 0.05, 0.1. Результаты при таких  $\Delta\mathcal{R}$  одинаковы:  $C_0^R = 0.84$  (см. рис. 2.14, справа снизу) и  $\mathcal{I}_A = \{1, 2, 3\}$ , что совпадает с результатом визуальной идентификации. Выделенный тренд также изображён на рис. 2.14 (справа сверху).

<sup>8</sup>Для ряда  $F_{47}$   $\alpha = 0.02$  соответствует увеличению стандарта шума приблизительно в 2.5 раза

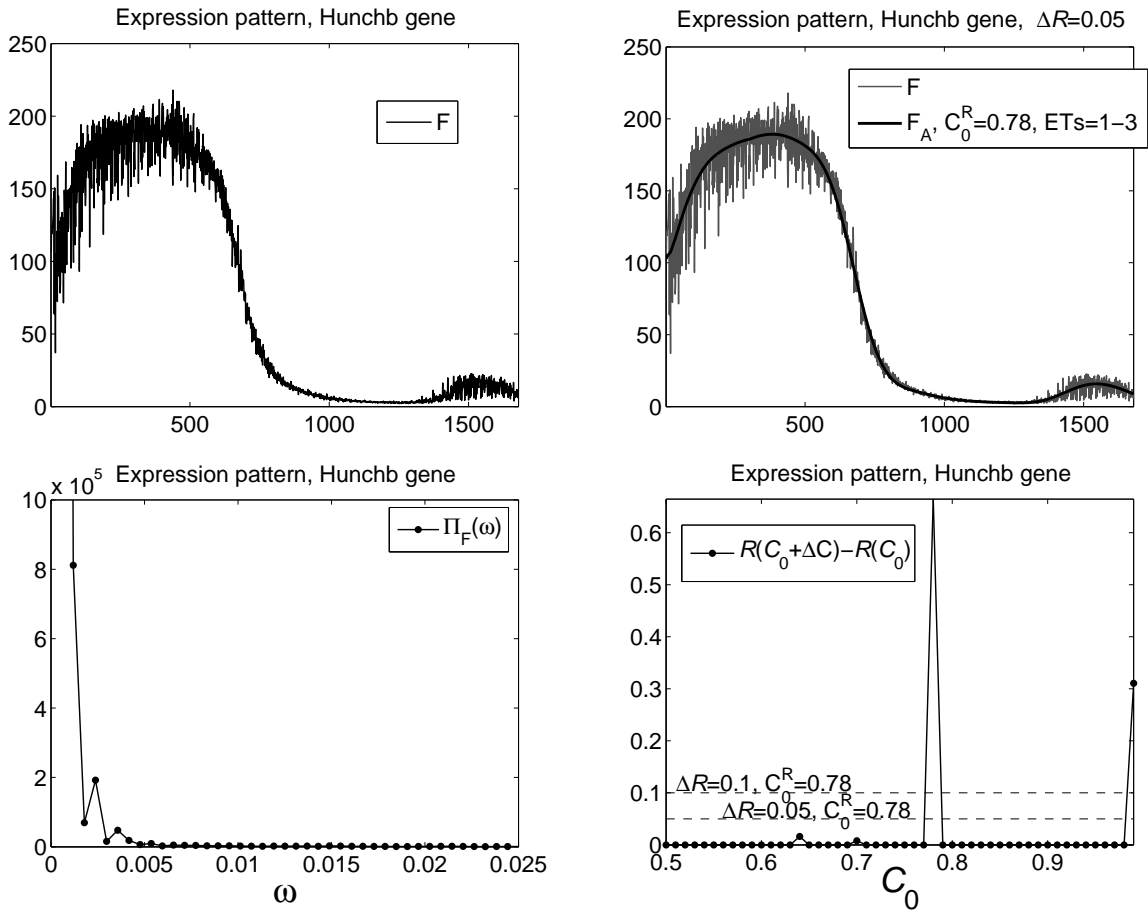


Рис. 2.14: Выделение тренда из данных об уровне экспрессии гена.

## 2.8.2 Выделение тренда различной детализации из данных по уровню безработицы

Рассмотрим ряды, взятые из [34], отражающие ежемесячные измерения уровня безработицы в отдельных штатах США с января 1976 г. по сентябрь 2006 г. Данные ряды имеют сложную структуру, основной задачей в данном разделе является не столько выделение некоторого определённого тренда, сколько изучение зависимости детализации получаемого тренда от  $\omega_0$ .

Длина рассматриваемых рядов одинаковая:  $N = 369$ . Выберем  $L = 180$ , чтобы улучшить разделимость с остатком, который может содержать сезонную составляющую с периодом  $T = 12$ .

**Ряд Florida** Возьмём ряд Florida, содержащий информацию о штате Флорида (рис. 2.15). На периодограмме  $\Pi_F^N$  видна область низких частот, граница которой приходится приблизительно на 0.02. Принимая во внимание замечание 2.5, возьмём  $\omega_0 = 0.03 \approx [0.02L]/L + 1/L$ . Применим процедуру TRRMEAS с параметрами  $L = 180$ ,  $\omega_0 = 0.03$  к ряду  $F$ .

На графике, отражающем выбор  $C_0^R$  (рис. 2.15, справа снизу) видно, что результаты получаются одинаковые при  $\Delta R = 0.05, 0.1$ :  $C_0^R = 0.58$ .

Применим теперь процедуру TRRMEAS с тем же  $L$ , с  $\Delta R = 0.05$ , но с разными  $\omega_0$ :  $\omega_0 = 0.05$  и  $\omega_0 = 0.08$ . Результаты изображены на рис. 2.16. Видно, что значение  $\omega_0$  управляет детализацией получаемого тренда — чем больше  $\omega_0$ , тем более детальным будет результат процедуры TRRMEAS.

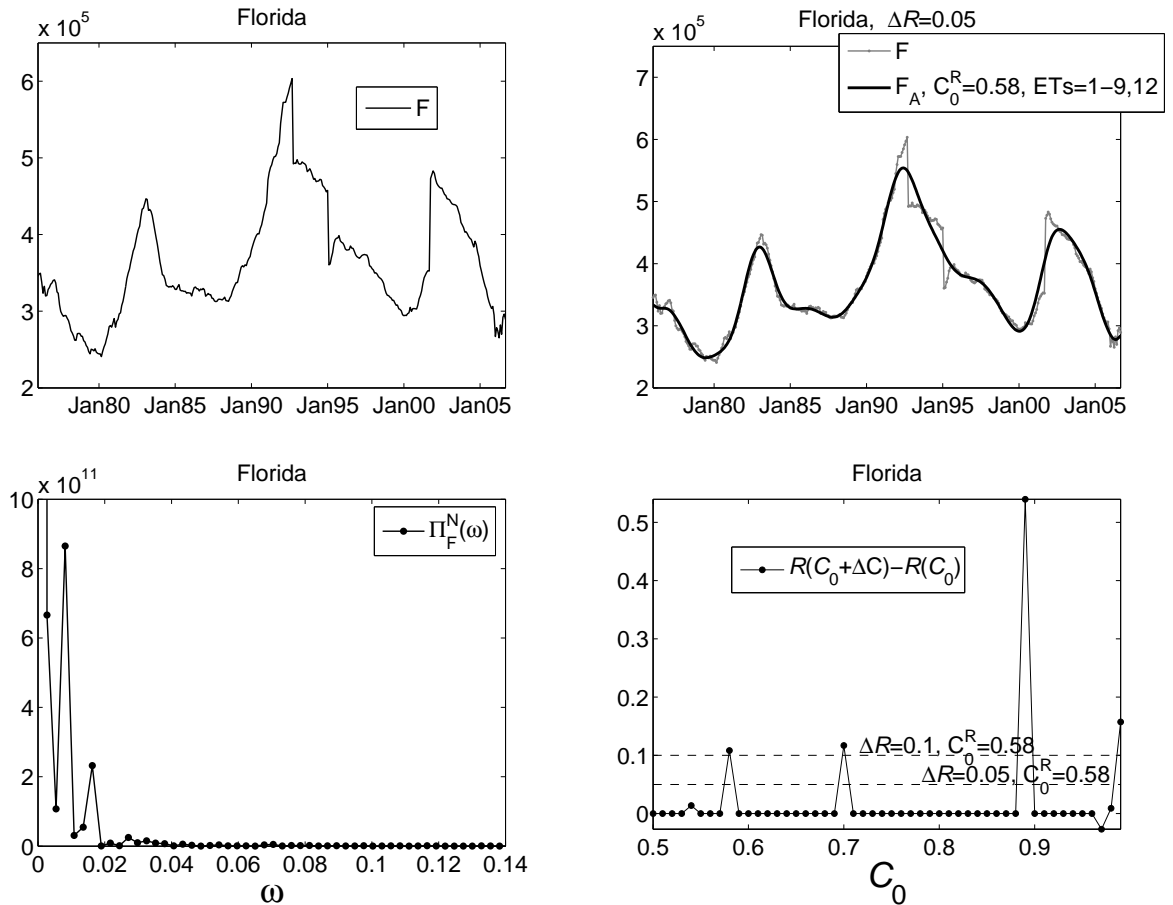


Рис. 2.15: Выделение тренда ряда Florida,  $\omega_0 = 0.03$ .

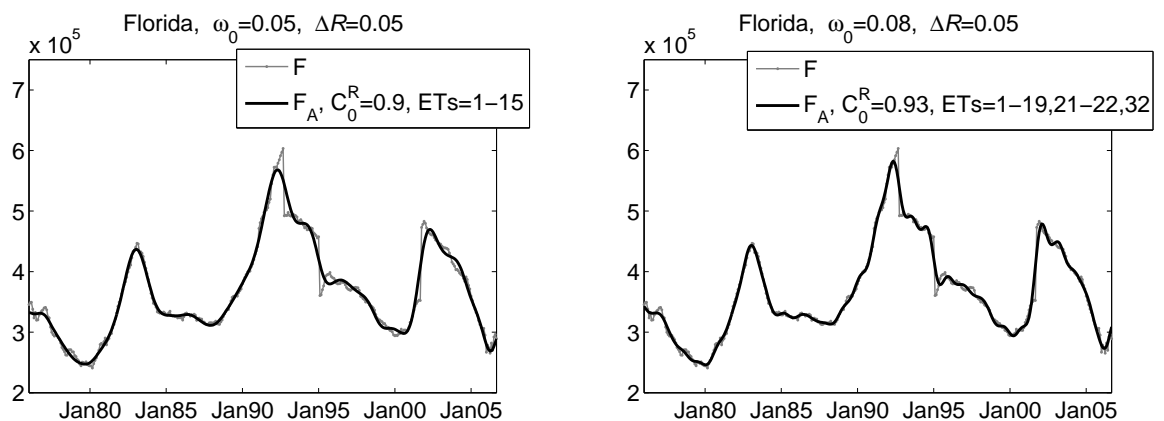


Рис. 2.16: Ряд Florida, более детальные тренды ( $\omega_0 = 0.05$  и  $\omega_0 = 0.08$ ).



**Ряд Alaska** Продемонстрируем ещё на одном примере зависимость детализации получаемого автоматически тренда от  $\omega_0$ . Возьмём ряд Alaska, соответствующий штату Аляска. Применим к нему процедуру TRRMEAS с  $L = 180$ , разными  $\omega_0$  и  $\Delta\mathcal{R} = 0.05$  (полученные при  $\Delta\mathcal{R} = 0.05$  и  $\Delta\mathcal{R} = 0.1$  значения  $C_0^R$  совпадают). Результаты для  $\omega_0 = 0.02, 0.05, 0.08$  приведены на рис. 2.17.

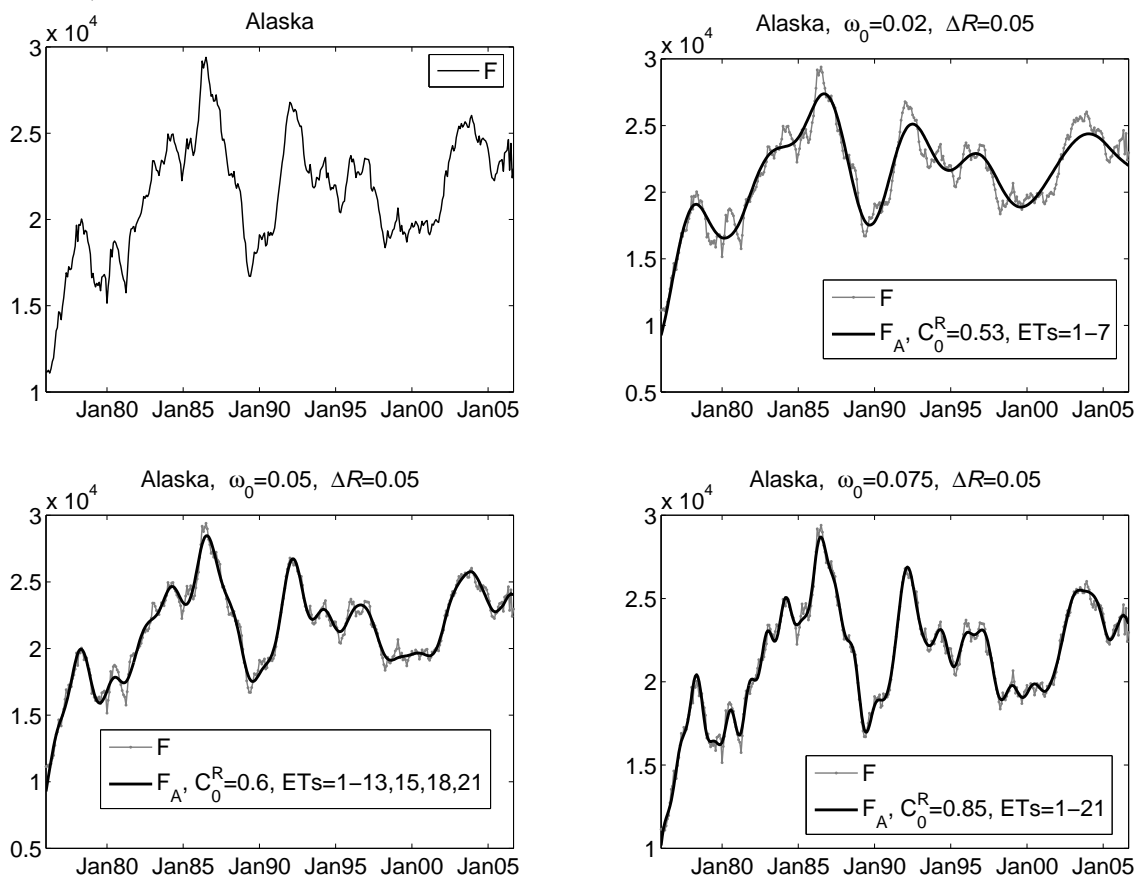


Рис. 2.17: Ряд Alaska и результаты процедуры TRRMEAS при разных  $\omega_0$ : 0.02, 0.05, 0.08.

### 2.8.3 Сравнение процедуры TRRMEAS с другими методами выделения тренда

В данном разделе на примере сравнения процедуры TRRMEAS с другим методом выделения тренда продемонстрируем одну важную особенность метода “Гусеница”-SSA, которой обладает и процедура TRRMEAS.

Пусть  $F = F^{(T)} + F^{(R)} + F^{(\varepsilon)}$  и тренд  $F^{(T)}$  представляет собой ряд конечного ранга, а  $F^{(\varepsilon)}$  соответствует шуму. Тогда, при условии того, что  $F^{(T)}$  делим с  $F^{(R)}$ , наличие  $F^{(R)}$  не влияет на результат процедур TREND и TRRMEAS. При приближённой делимости — чем лучше качество делимости, тем меньше влияние.

Продemonстрируем это, сравнив процедуру TRRMEAS с методом выделения тренда с помощью фильтра Hodrick-Prescott (HP) [43], применение которого в экономике стало популярным в последние годы [46]. Данный метод имеет только один параметр,  $w \in \mathbb{N}$ .

Рассмотрим ряды  $F^{(1)} = F^{(T)} + F^{(\varepsilon)}$  и  $F^{(2)} = F^{(1)} + F^{(P)}$ :

$$f_n^{(T)} = \frac{1}{33000}(n+10)(n-10)(n-30)(110-n), \quad f_n^{(P)} = 10 \cos(2\pi n/6),$$

$F^{(\varepsilon)} \sim NID(0, 1)$ ,  $N = 100$ . Ряды изображены на рис. 2.18 (слева). Выделим тренд из  $F^{(1)}$  и  $F^{(2)}$ , сравним результаты. Разница между ними отражает влияние  $F^{(P)}$ .

Возьмём следующие параметры процедуры TRRMEAS:  $L = 50$ ,  $\omega_0 = 0.1$ ,  $C_0^l = 0.5$ ,  $C_0^r = 1$ ,  $\Delta C = 0.01$ ,  $\Delta R = 0.1$ . Обозначим результат её применения к  $F^{(1)}$ ,  $F^{(2)}$ , как  $F_A^{(1)}$ ,  $F_A^{(2)}$ .

Для HP-фильтра переберём значения параметра  $w$  из  $\Omega_w = 10(10)2000$ . Обозначим результат его применения с параметром  $w$  к рядам  $F^{(1)}$ ,  $F^{(2)}$ , как  $F_{HP}^{(1)}(w)$ ,  $F_{HP}^{(2)}(w)$ .

**Результаты** Применим процедуры к рядам  $F^{(1)}$  и  $F^{(2)}$  ( $F^{(2)} = F^{(1)} + F^{(P)}$ ). Результаты вычисления  $SKO(F^{(T)}, F_A^{(j)})$  и  $SKO(F^{(T)}, F_{HP}^{(j)}(w))$  ( $j = 1, 2$ ) приведены на рис. 2.18 (справа).

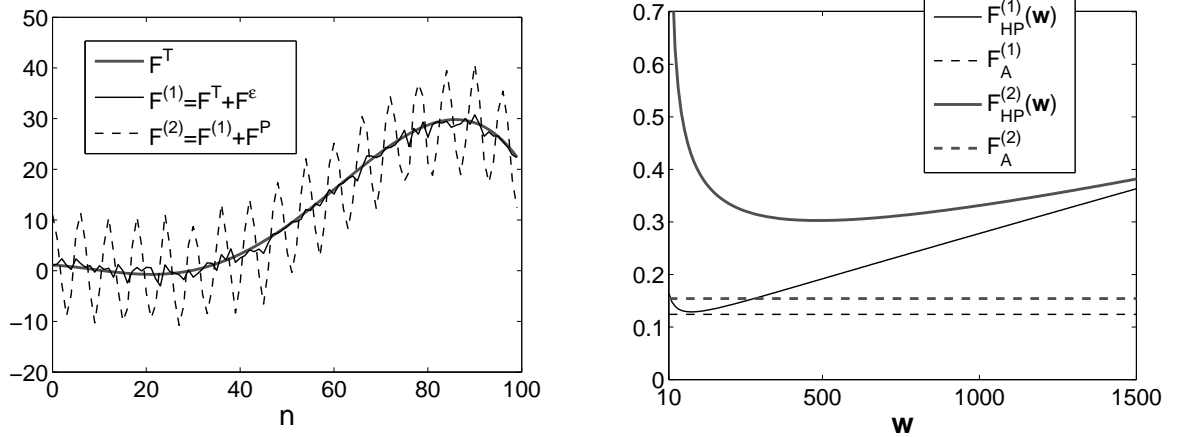


Рис. 2.18: Тренд  $F^{(T)}$  и ряды  $F^{(1)}$ ,  $F^{(2)}$  (слева) и влияние  $F^{(P)}$  на среднеквадратичную ошибку выделения тренда с помощью процедуры TRRMEAS и HP-фильтра с параметром  $w$  (справа).

Видно, что для процедуры TRRMEAS составляющая  $F^{(P)}$  оказывает гораздо меньшее влияние на результат, чем для HP-фильтра.

На рис. 2.19 продемонстрировано, что это вызвано неточностью выделения тренда с помощью фильтра НР на концах ряда, где крайние значения  $F^{(P)}$  влияют на результат.

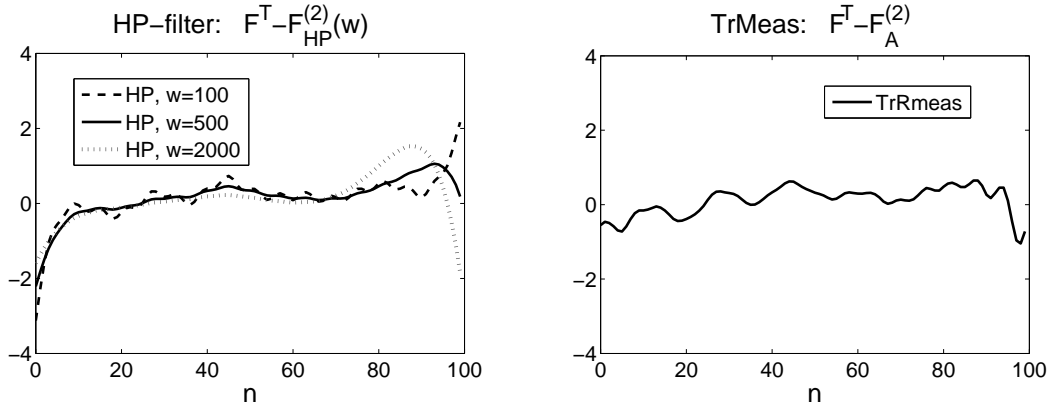


Рис. 2.19: Ряд  $F^{(T)} - F_{\mathcal{HP}}^{(2)}(w)$  для фильтра НР с разными  $w$  (слева) и ряд  $F^{(T)} - F_{\mathcal{A}}^{(2)}$  (справа).

## 2.9 Применение процедуры автоматического выделения тренда для обработки множества рядов

### 2.9.1 Проблема проверки качества процедуры TRRMEAS

Для обоснованного применения процедуры TRRMEAS (см. алгоритм 2.5) при решении задачи выделения тренда ряда необходимо проверить следующее:

- Существуют такое  $L$  и такой набор собственных троек  $\mathcal{I}_{\mathcal{V}}$ , которые можно визуально идентифицировать, что тренд ряда  $F^{(T)}$  аппроксимируется с хорошим качеством составляющей  $F_{\mathcal{V}}$ , восстановленной по  $\mathcal{I}_{\mathcal{V}}$ . Для этого необходимо, чтобы тренд хорошо аппроксимировался рядом конечного ранга и был приближённо отделен от остатка.
- Существуют такие параметры автоматической процедуры TREND  $(\omega_0, \mathcal{C}_0)$ , что  $F_{\mathcal{V}}$  будет хорошо аппроксимирован рядом  $F_{\mathcal{A}}$ , полученным применением процедуры TREND с этими параметрами при том же  $L$ .
- Параметры процедуры TREND  $(\omega_0, \mathcal{C}_0)$  могут быть найдены с помощью предложенных выше методов, в частности,  $\mathcal{C}_0$  может быть определён посредством меры  $\mathcal{R}(\mathcal{C}_0)$ .

**Замечание 2.10** Если оценивалось  $\mathcal{C}_{min}$ , то при этом использовали ещё одно предположение: выбранная стохастическая модель для остатка хорошо отражает поведение настоящего остатка. Однако это предположение наименее важно

из сформулированных, так как выполнение условий принципиальной выделенности тренда и  $C_0 \leq C_{max}$  являются более важными, чем  $C_0 \geq C_{min}$ .

Для того чтобы проверить, выполняются ли эти предположения (особенно относительно меры  $\mathcal{R}$ ) для ряда  $F$ , можно пойти по такому пути: задать некоторым образом модель ряда и проверить, выполняются ли предположения для рядов этой модели (точно, асимптотически, в среднем). Однако существует постановка задачи, которая позволяет проверить сформулированные предположения, но не требует задания модели ряда, что позволяет сохранить присущую подходу “Гусеница”-SSA “немодельность” — это пакетная обработка множества рядов. При решении такой задачи можно проверить процедуру TREND следующим образом. Возьмём тестовое подмножество и для каждого ряда  $F$  из подмножества вручную выделим тренд  $F_{\mathcal{V}}$  и сравним  $F_{\mathcal{V}}$  с  $F_{\mathcal{A}}$ , полученным автоматически. По среднему результату (на тестовом подмножестве) оценим среднее качество процедуры TREND на всём множестве.

Конечно, для такой оценки необходимо некоторым образом ограничить разнообразие рядов в обрабатываемом множестве, иначе оценка по тестовому множеству будет бессмысленна. Также должен быть сравнительно большим размер тестового (а значит и обрабатываемого) множества рядов.

Опишем подобную постановку задачи на математическом языке и покажем, как строится оценка и от чего зависит её точность и размер необходимого тестового множества.

### 2.9.2 Математическая постановка задачи, оценка ошибки автоматической процедуры и её свойства

Обозначим множество рядов длины  $N$ , которые необходимо обработать, за  $\Theta$ . Рассмотрим каждый ряд  $F$  из  $\Theta$ , как реализацию  $N$ -мерной случайной величины  $\xi$ .

Рассмотрим случай, когда разнообразие трендов для рядов из  $\Theta$  ограничено, что можно записать, как:

$$\forall F \in \Theta : F^{(T)} \in \{T_j\}_{j=1}^M, \quad T_j \in \mathbb{R}^N. \quad (2.21)$$

Определим  $p_j$ ,  $j = 1, \dots, M$ , следующим образом:

$$p_j = \frac{\#\{F : F^{(T)} = T_j\}}{\#\Theta},$$

где  $\#$  обозначает мощность множества. В случае, когда  $\Theta$  достаточно большое,

$p_j$  имеет смысл вероятности выпадения тренда  $T_j$ , поэтому определим  $\xi_1$  —  $N$ -мерную дискретную случайную величину, такую что

$$P(\xi_1 = T_j) = p_j.$$

Введём  $\xi_2 = \xi - \xi_1$  —  $N$ -мерную случайную величину. Потребуем, чтобы она не зависела от  $\xi_1$ .

Пусть есть некоторые функции  $\mathcal{A}, \mathcal{V} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ , определим функцию  $g : \mathbb{R}^N \rightarrow \mathbb{R}$ , такую что

$$g(x, y) = \|\mathcal{A}(x + y) - \mathcal{V}(x + y)\|^2, \quad x, y \in \mathbb{R}^N.$$

Если считать случайные величины  $\xi_1, \xi_2$  соответствующими тренду и шуму, функции  $\mathcal{A}(\cdot), \mathcal{V}(\cdot)$  — соответствующими автоматической и визуальной процедурам выделения тренда, то видно, что данная модель описывает задачу выделения тренда для каждого ряда из множества  $\Theta$ . Функция  $g(\xi_1, \xi_2)$  отражает ошибку автоматической процедуры относительно визуальной, а  $\mathbf{E}g(\xi_1, \xi_2)$  — среднюю ошибку (где  $\mathbf{E}$  обозначает математическое ожидание).

Так как  $\xi_1, \xi_2$  — независимы,  $g \geq 0$ , то, предполагая борелевость  $g$ , получаем, что

$$\mathbf{E}g(\xi_1, \xi_2) = \sum_{j=1}^M p_j \mathbf{E}_2 g(T_j, \xi_2), \quad \text{где } \mathbf{E}_2 \zeta = \int_{\mathbb{R}^N} x d\mathcal{F}(x),$$

где  $\mathcal{F}(\cdot)$  обозначает функцию распределения  $\xi_2$ . Это верно при условии существования  $\mathbf{E}_2 g(T_j, \xi_2)$  для  $j = 1, \dots, M$ . Тогда очевидно, что оценку  $\widehat{\mathbf{E}}g(\xi_1, \xi_2)$  можно получить следующим образом.

Возьмём случайным образом  $\Gamma \subset \Theta$ . Такое  $\Gamma$  является выборкой из закона распределения  $\xi = \xi_1 + \xi_2$ , поэтому

$$\widehat{\mathbf{E}}g(\xi_1, \xi_2) = \frac{1}{\#\Gamma} \sum_{F \in \Gamma} \|\mathcal{A}(F) - \mathcal{V}(F)\|^2. \quad (2.22)$$

Схема оценки средней ошибки автоматической процедуры представлена в алгоритме 2.7. В нём также вычисляется эмпирический доверительный интервал  $[E_{min}, E_{max}]$  для данной оценки.

**АЛГОРИТМ 2.7. Оценка средней ошибки процедуры TRRMEAS при обработке множества рядов**

**Входные данные:**  $\Theta$  — множество рядов для обработки, тестовое подмножество  $\Gamma \subset \Theta$

**Результат:**  $\{F_{\mathcal{A}}\}$  — выделенные тренды рядов из  $\Gamma$ ,  $[E_{min}, E_{max}]$  — эмпирический доверительный интервал для оценки  $\widehat{\mathbf{E}}g(\xi_1, \xi_2)$

**Параметры:**

- Параметры процедуры TRRMEAS:  $L, \omega_0, \mathcal{C}_0^[, \mathcal{C}_0^], \Delta\mathcal{C}, \Delta\mathcal{R}$
- $\gamma$  — надёжность эмпирического интервала  $[E_{min}, E_{max}]$

**Процедура:**

1.  $\forall F \in \Gamma$ :

- 1) Выделим  $F_{\mathcal{V}}$  с помощью SSADEC, SSAREC (алгоритмы 1.1, 1.2), используя визуальную идентификацию
- 2) Применив к  $F$  процедуру TRRMEAS с заданными параметрами, получим  $F_{\mathcal{A}}$  (алгоритм 2.5)
- 3) Посчитаем  $\mathcal{D}_{\mathcal{AV}}(F) = \|F_{\mathcal{A}} - F_{\mathcal{V}}\|^2$

2. Вычислим  $\widehat{\mathbf{E}}g(\xi_1, \xi_2)$  по (2.22), взяв  $\|\mathcal{A}(F) - \mathcal{V}(F)\|^2 = \mathcal{D}_{\mathcal{AV}}(F)$

3. По выборке  $\{\mathcal{D}_{\mathcal{AV}}(F)\}_{F \in \Gamma}$  посчитаем границы эмпирического доверительного интервала для оценки  $\widehat{\mathbf{E}}g(\xi_1, \xi_2)$ , откинув  $\#\Gamma(1 - \gamma)/2$  наибольших и  $\#\Gamma(1 - \gamma)/2$  наименьших значений и взяв минимум/максимум оставшихся

**Дисперсия ошибки** Исследуем теперь свойства дисперсии  $\mathbf{D}$  величины  $g(\xi_1, \xi_2)$ , так как от  $\mathbf{D}g(\xi_1, \xi_2)$  зависит дисперсия оценки  $\widehat{\mathbf{E}}g(\xi_1, \xi_2)$ , которая в свою очередь определяет размер  $\Gamma$ . Известно, что дисперсия совокупности, состоящей из нескольких групп, состоит из суммы межгрупповой дисперсии и внутригрупповых дисперсий. Покажем, что это означает.

Для каждого  $j = 1, \dots, M$  определим

$$\zeta_j = g(T_j, \xi_2), \quad \mathbf{E}\zeta_j = \mathbf{E}_2g(T_j, \xi_2),$$

а также введём  $\eta$  — вещественную дискретную случайную величину, такую что

$$P(\eta = E\zeta_j) = p_j, \quad 1 \leq j \leq M.$$

Посчитаем  $\mathbf{D}g(\xi_1, \xi_2)$ :

$$\mathbf{D}g(\xi_1, \xi_2) = \mathbf{E}g^2(\xi_1, \xi_2) - \mathbf{E}^2g(\xi_1, \xi_2).$$

Так как  $\xi_1, \xi_2$  независимы и  $g \geq 0$ , то, предполагая борелевость  $g$ , получаем, что

$$\begin{aligned} \mathbf{D}g(\xi_1, \xi_2) &= \sum_{j=1}^M p_j \mathbf{E}\zeta_j^2 - \left( \sum_{j=1}^M p_j \mathbf{E}\zeta_j \right)^2 = \\ &= \sum_{j=1}^M p_j (\mathbf{E}\zeta_j^2 - \mathbf{E}^2\zeta_j) + \sum_{j=1}^M p_j \mathbf{E}^2\zeta_j - \left( \sum_{j=1}^M p_j \mathbf{E}\zeta_j \right)^2. \end{aligned}$$

Очевидно, что

$$\mathbf{D}\eta = \mathbf{E}\eta^2 - \mathbf{E}^2\eta = \sum_{j=1}^M p_j \mathbf{E}^2\zeta_j - \left( \sum_{j=1}^M p_j \mathbf{E}\zeta_j \right)^2.$$

Следовательно,

$$\mathbf{D}g(\xi_1, \xi_2) = \mathbf{D}\eta + \sum_{j=1}^M p_j \mathbf{D}\zeta_j.$$

Таким образом, размер  $\Gamma$  зависит от  $\mathbf{D}\eta$  и  $\sum_{j=1}^M p_j \mathbf{D}\zeta_j$ . Значение  $\mathbf{D}\eta$  отражает “стабильность” автоматического выделения тренда, т. е. насколько одинаково для разных трендов качество автоматической процедуры по сравнению с процедурой визуальной. Значение  $\sum_{j=1}^M p_j \mathbf{D}\zeta_j$  зависит от точности выделения тренда для ряда модели:  $T_j$ +шум. Причём чем больше рядов с трендом  $T_j$  в  $\Theta$ , тем большее влияние оказывает эта точность на дисперсию ошибки. Теперь ограничение на  $\Theta$  можно сформулировать следующим образом: чем меньше  $\mathbf{D}\eta$ , тем больше будет точность оценки ошибки.

Итак,  $\mathbf{D}\eta$  характеризует “однородность”  $\Theta$ , а  $\sum_{j=1}^M p_j \mathbf{D}\zeta_j$  — качество автоматической процедуры для каждого из  $T_j$ .

**Пример 2.7** Пусть все ряды из  $\Theta$  имеют один и тот же тренд,  $T$ , тогда  $\mathbf{D}\eta = 0$  и, что совершенно естественно,  $\mathbf{D}g(\xi_1, \xi_2) = \mathbf{D}g(T, \xi_2)$ .

В случае различных  $T_j$  минимальная  $\mathbf{D}\eta$  достигается в случае, когда  $\mathbf{E}\zeta_j$  постоянно по  $j$ . Тогда  $\mathbf{D}\eta = 0$ , и  $\mathbf{D}g(\xi_1, \xi_2)$  зависит только от  $\sum_{j=1}^M p_j \mathbf{D}\zeta_j$ . Этот случай можно охарактеризовать как одинаковость (в среднем) поведения процедуры автоматического выделения тренда для разных моделей  $T_j$ +шум (при фиксированной модели шума), что для процедуры TREND выполняется при следующих ограничениях:

- Все  $T_j$  — ряды конечного ранга.
- Для всех  $j$  разделимость  $T_j$  с реализациями  $\xi_2$  одинаково хороша.

Разделимость  $T_j$  и шума определяется собственными векторами  $T_j$ , так что приблизительно одинаковая разделимость для всех  $j$  означает похожее представление трендов  $T_j$  через собственные вектора. Что должно выполняться, если тренды  $T_j$  похожи по своим свойствам и при этом ненамного отличаются друг от друга.

Таким образом, чем более похожи между собой тренды  $T_j$ , тем меньше будет  $\mathbf{D}\eta$  и тем меньший размер тестового множества  $\Gamma$  необходимо брать.

### 2.9.3 Схема применения TRRMEAS к множеству рядов

**Выбор размера тестового множества** Выбор размера тестового множества построим на рассмотрении ширины эмпирического доверительного интервала для  $\widehat{\mathbf{E}}g(\xi_1, \xi_2)$ . Если согласно общим представлениям она слишком велика, то надо увеличить  $\Gamma$ . Если при последовательном увеличении  $\Gamma$  размер доверительного интервала не уменьшается, то возможно надо попробовать применять процедуру TRRMEAS с другими параметрами. Если это не помогает, то данная схема не может быть обоснованно применена (возможно,  $T_j$  слишком разные или процедура TRRMEAS не подходит для таких рядов).

**Схема** Последовательность действий при обработке множества рядов  $\Theta$  излагается в алгоритме 2.8.



**АЛГОРИТМ 2.8. Использование процедуры TRRMEAS для обработки множества рядов  $\Theta$**

**Входные данные:**  $\Theta$  — множество рядов для обработки

**Результат:** Выделенные тренды рядов из  $\Theta$

**Параметры:**

- Параметры процедуры TRRMEAS:  $L, \omega_0, \mathcal{C}_0^l, \mathcal{C}_0^r, \Delta\mathcal{C}, \Delta\mathcal{R}$
- $S$  — начальный размер тестового множества,  $\gamma$  — надёжность рассчитываемого доверительного интервала

**Процедура:**

1. Выберем случайное тестовое подмножество  $\Gamma \subset \Theta$ ,  $\#\Gamma = S$
2. Оценим  $\Gamma$  ошибку  $\mathbf{E}g(\xi_1, \xi_2)$  и её доверительный интервал с заданной надёжностью  $\gamma$ :  $[E_{min}, E_{max}]$  (алгоритм 2.7)
3. Если  $[E_{min}, E_{max}]$  слишком велик, возьмём большее  $\Gamma$
4. Если при последовательном увеличении размера  $\Gamma$  ширина доверительного интервала не уменьшается до приемлемого, то данная схема не подходит
5. Применим к оставшимся рядам из  $\Theta$  процедуру TRRMEAS

## 2.10 Пример применения процедуры TRRMEAS к множеству рядов

Рассмотрим множество рядов  $\Theta$ ,  $\#\Theta = 38$ , которые содержат данные о пространственном распределении экспрессии гена Krüppel в различных эмбрионах плодовой мушки дрозофилы (см. раздел 2.8.1). Ряды отличаются друг от друга длиной, шумовой составляющей и деталями тренда. Выберем случайным образом тестовое множество  $\Gamma_5$  из 5-ти рядов с номерами  $\{6, 15, 17, 24, 30\}$ . Графики некоторых рядов из  $\Gamma_5$  приведены на рис. 2.20.

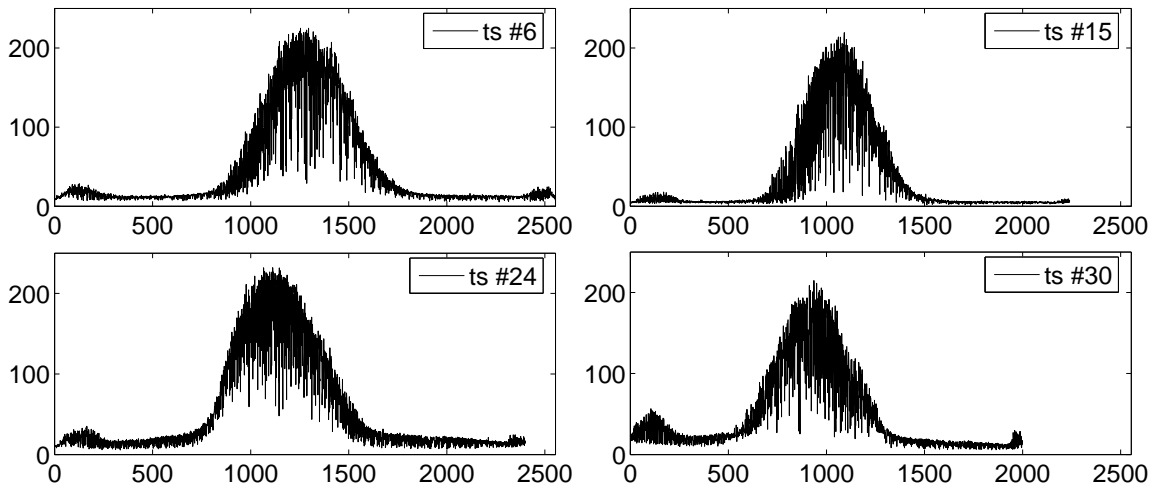


Рис. 2.20: Примеры рядов из обрабатываемого множества.

Возьмём  $L = 500$ . Визуальное исследование рядов из  $\Gamma_5$  показало, что такое  $L$  позволяет хорошо выделить тренд ряда. Результаты визуальной идентификации представлены в табл. 2.2.

Таблица 2.2: Результаты визуальной идентификации

| Номер ряда | $\mathcal{I}_\nu$ |
|------------|-------------------|
| 6          | {1-4, 15}         |
| 15         | {1-5, 8}          |
| 17         | {1-4}             |
| 24         | {1-4, 7}          |
| 30         | {1-4, 19}         |

Выберем  $\omega_0$ . Рассмотрим графики периодограмм тех же рядов из  $\Gamma_5$  (рис. 2.21). Видно, что граница области низких частот составляет приблизительно 0.005–0.008. Возьмём  $\omega_0$  с небольшим запасом, так как оно выбирается для всего множества, в котором ряды могут немного отличаться:  $\omega_0 = 0.012 \approx [0.008L]/L + 2/L$ .

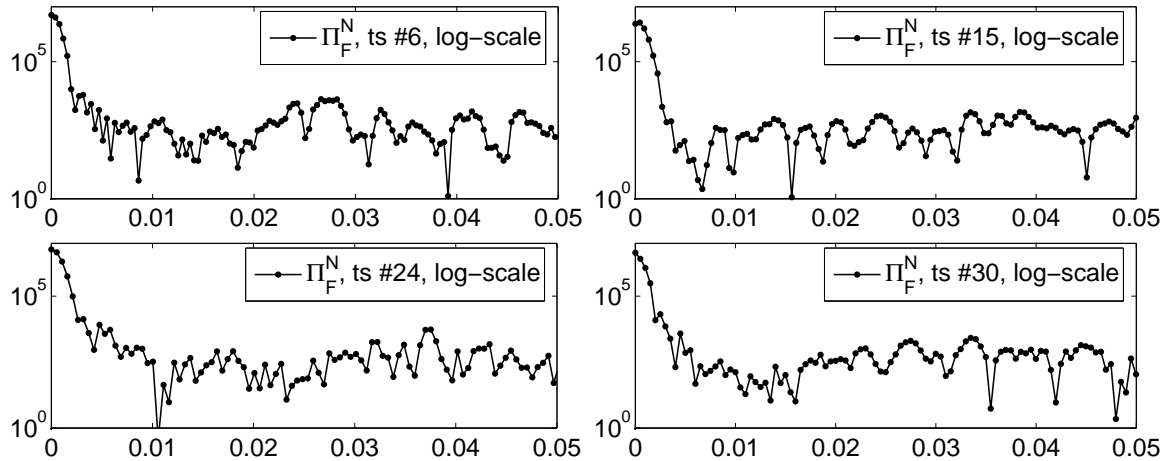


Рис. 2.21: Периодограммы некоторых рядов.

В качестве остальных параметров процедуры TRRMEAS возьмём  $\mathcal{C}_0^l = 0.5$ ,  $\mathcal{C}_0^r = 1$ ,  $\Delta\mathcal{C} = 0.01$ ,  $\Delta\mathcal{R} = 0.1$ .

Применим процедуру TRRMEAS (алгоритм 2.5) с выбранными параметрами к рядам из  $\Gamma_5$ , результаты приведены в табл. 2.3. Видно, что, во-первых,

Таблица 2.3: Результаты процедуры TRRMEAS,  $\mathcal{C}_0 = \mathcal{C}_0^R$ .

| Номер ряда | $\mathcal{I}_V$ | $\mathcal{I}_A$ | $\mathcal{I}_V \setminus \mathcal{I}_A$ | $\mathcal{D}_{AV}$ |
|------------|-----------------|-----------------|---|--------------------|
| 6          | {1-4, 15}       | {1-3}           | {4, 15}                                 | 3.261              |
| 15         | {1-5, 8}        | {1-3, 5}        | {4, 8}                                  | 3.953              |
| 17         | {1-4}           | {1-3}           | {4}                                     | 1.507              |
| 24         | {1-4, 7}        | {1-3, 7}        | {4}                                     | 2.534              |
| 30         | {1-4, 19}       | {1-3}           | {4, 19}                                 | 4.683              |

хотя происходит систематическая неидентификация собственной тройки ЕТ4, среднеквадратичная разница между АИ и ВИ невелика. Во-вторых, процедура TRRMEAS показывает стабильный результат на всех рядах из  $\Gamma_5$ .

Визуальное исследование характеристик рядов из  $\Gamma$  показало, что если брать при некотором  $\mathcal{C}_\varepsilon$ :  $\mathcal{C}_0 = \mathcal{C}_{max} - \mathcal{C}_\varepsilon$ , где  $\mathcal{C}_{max} = \mathcal{C}_0^R$ , то результат АИ будет намного лучше. Было опробованы  $\mathcal{C}_\varepsilon = 0.1, 0.2$  и взято  $\mathcal{C}_\varepsilon = 0.2$ . Результаты приведены в табл. 2.4.

Таблица 2.4: Результаты процедуры TRRMEAS,  $\mathcal{C}_0 = \mathcal{C}_0^R - 0.02$ 

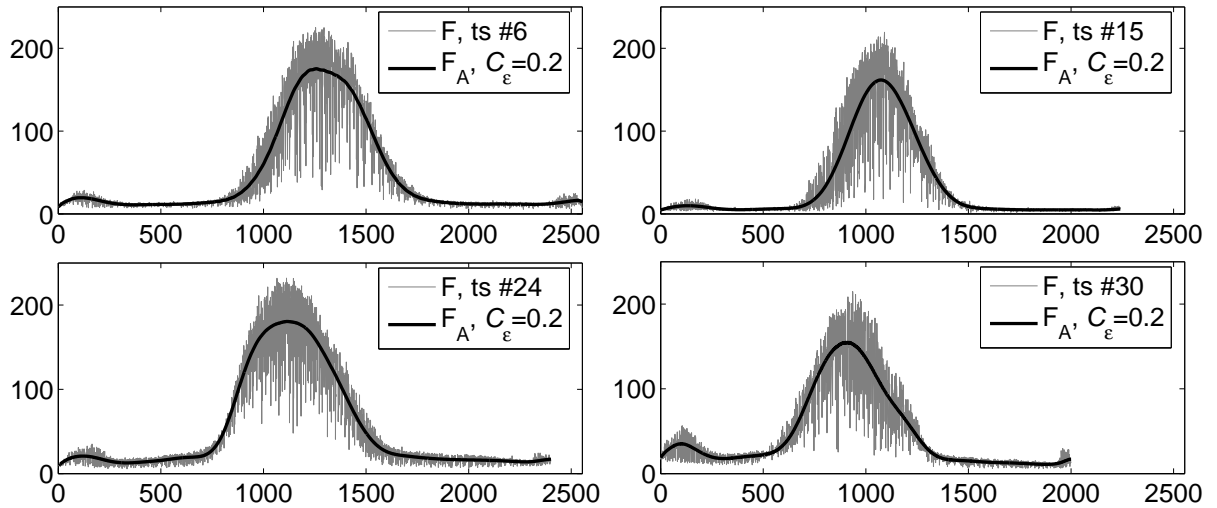
| Номер ряда | $\mathcal{I}_V$ | $\mathcal{I}_A$ | $\mathcal{I}_V \setminus \mathcal{I}_A$ | $\mathcal{D}_{AV}$ |
|------------|-----------------|-----------------|---|--------------------|
| 6          | {1-4, 15}       | {1-4, 15}       | $\emptyset$                             | 0                  |
| 15         | {1-5, 8}        | {1-5, 8}        | $\emptyset$                             | 0                  |
| 17         | {1-4}           | {1-4}           | $\emptyset$                             | 0                  |
| 24         | {1-4, 7}        | {1-4, 7}        | $\emptyset$                             | 0                  |
| 30         | {1-4, 19}       | {1-4, 19}       | $\emptyset$                             | 0                  |

Видно, что при таком выборе  $\mathcal{C}_0$  процедура TRRMEAS показывает идеальное качество — для всех рядов из  $\Gamma_5$  результат АИ равен результату ВИ.

Увеличим тестовое множество, взяв случайным образом ещё два ряда,  $\Gamma_7 = \Gamma_5 \cup \{28, 35\}$ . Результаты для ВИ и процедуры TRRMEAS с  $\mathcal{C}_0 = \mathcal{C}_0^R - 0.2$  приведены в табл. 2.5. Графики примеров трендов изображены на рис. 2.22.

Таблица 2.5: Результаты процедуры TRRMEAS к  $\Gamma_7 \setminus \Gamma_5$ ,  $\mathcal{C}_0 = \mathcal{C}_0^R - 0.02$ 

| Номер ряда | $\mathcal{I}_V$ | $\mathcal{I}_A$ | $\mathcal{I}_V \setminus \mathcal{I}_A$ | $\mathcal{D}_{AV}$ |
|------------|-----------------|-----------------|---|--------------------|
| 28         | {1-5}           | {1-5}           | $\emptyset$                             | 0                  |
| 35         | {1-6}           | {1-6}           | $\emptyset$                             | 0                  |

Рис. 2.22: Тренды, выделенные с помощью TRRMEAS с  $\mathcal{C}_0 = \mathcal{C}_0^R - 0.02$ .

**Вывод** Процедура TRRMEAS с выбранными параметрами и  $\mathcal{C}_0 = \mathcal{C}_0^R - 0.02$  показала идеальные результаты на случайно выбранных тестовых множествах  $\Gamma_5$ ,  $\Gamma_7$  и поэтому у нас есть основания для того, чтобы применять её ко всем рядам из  $\Theta$ .

## 2.11 Прогноз тренда

### 2.11.1 Схема прогноза тренда ряда

На основании процедуры SSAFORE для прогноза ряда с помощью метода “Тусеница”-SSA (см. раздел 1.4, алгоритм 1.3) построим процедуру прогноза ряда с идентификацией по методу НЧ.

Пусть у нас есть ряд  $F$ , содержащий тренд  $F^{(T)}$ , и необходимо получить прогноз составляющей  $F^{(T)}$  на одну или несколько точек вперед. Выделим тренд с помощью предложенной в разделе 2.2 процедуры TREND (алгоритм 2.2) и используем идентифицированные собственные вектора для вычисления коэффициентов ЛРФ порядка  $L - 1$ . После этого, применив ЛРФ к последним значениям выделенного тренда, получим требуемый прогноз.

**АЛГОРИТМ 2.9. Процедура TRFORE: прогноз тренда ряда с помощью процедуры TREND**

**Входные данные:**  $F$  — ряд длины  $N$ , содержащий  $F^{(T)}$

**Результат:**  $\tilde{f}$  — прогноз значения  $f_N^{(T)}$

**Параметры:**  $L$  ( $1 < L \leq \lceil N/2 \rceil$ ),  $\omega_0 \in (0, 0.5)$ ,  $C_0 \in [0, 1]$

**Процедура:**

1. Применим процедуру TREND с параметрами  $L$ ,  $\omega_0$ ,  $C_0$  к  $F$ , получив  $\mathcal{I}_A$  (алгоритм 2.2)
2. Применим процедуру SSAFORE с параметрами  $L$ ,  $\mathcal{I}_A$  к ряду  $F$ , получив прогноз  $\tilde{f}$  (алгоритм 1.3)

### 2.11.2 Проблема выбора порогового значения при прогнозе

При использовании процедуры TRFORE возникает проблема выбора порогового значения  $C_0$  для метода НЧ. С помощью статистического моделирования покажем на примере выделения экспоненциального тренда, что  $C_0$ , наилучшее в среднем для прогноза, совпадает с  $C_0^{opt}$ , при котором достигается наилучшее в среднем качество аппроксимации. Поэтому при прогнозе можно руководствоваться уже предложенными способами выбора  $C_0$ .

### 2.11.3 Моделирование прогноза экспоненциального тренда

Рассмотрим тот же ряд  $F = (f_0, \dots, f_{N-1})$ , что и в разделе 2.4:

$$F = F^{(T)} + F^{(\varepsilon)}, \quad f_n^{(T)} = e^{\alpha n}, \quad f_n^{(\varepsilon)} = \sigma e^{\alpha n} \varepsilon_n, \quad 0 \leq n \leq N-1,$$

где  $(f_0^{(\varepsilon)}, \dots, f_{N-1}^{(\varepsilon)})$  — реализация белого нормального шума  $NID(0, 1)$ . Оценим по  $R$  реализациям наилучшее в среднем пороговое значение  $\mathcal{C}_0$  для задачи прогноза тренда на одну точку вперёд и сравним его с  $\mathcal{C}_0^{opt}$ , рассчитанным для аппроксимации (раздел 2.4).

Обозначим аппроксимацию тренда, полученную визуальным образом как  $F_{\mathcal{V}} = (f_0^{(\mathcal{V})}, \dots, f_{N-1}^{(\mathcal{V})})$ , а тренд, выделенный автоматически как  $F_{\mathcal{A}} = (f_0^{(\mathcal{A})}, \dots, f_{N-1}^{(\mathcal{A})})$ . Обозначим как  $f_N^{(\mathcal{V})}$  и  $f_N^{(\mathcal{A})}$  значения прогноза тренда на одну точку вперёд, полученные на основе  $F_{\mathcal{V}}$  и  $F_{\mathcal{A}}$  соответственно. Если необходимо указать  $\mathcal{C}_0$ , при котором проводилась идентификация, то будем использовать обозначения  $F_{\mathcal{A}}(\mathcal{C}_0)$ ,  $f_n^{(\mathcal{A})}(\mathcal{C}_0)$ ,  $f_N^{(\mathcal{A})}(\mathcal{C}_0)$ .

Определим ошибку прогноза  $\mathcal{D}_{\mathcal{AV}}^{(\mathcal{F})}(\mathcal{C}_0)$  тренда как взвешенную квадратичную (как ранее, в разделе 2.4):

$$\mathcal{D}_{\mathcal{AV}}^{(\mathcal{F})}(\mathcal{C}_0) = \mathcal{D}(f_N^{(\mathcal{A})}(\mathcal{C}_0), f_N^{(\mathcal{V})}) = e^{-2\alpha N} \left( f_N^{(\mathcal{A})}(\mathcal{C}_0) - f_N^{(\mathcal{V})} \right)^2.$$

Аналогичным образом будем вычислять для  $f_N^{(\mathcal{A})}$  значения ошибок  $\mathcal{D}_{\mathcal{A}}^{\mathcal{F}}(\mathcal{C}_0)$ ,  $\mathcal{D}_{\mathcal{V}}^{\mathcal{F}}$ .

Обозначим как  $\mathcal{C}_0^{optA}$  и  $\mathcal{C}_0^{optF}$  наилучшие пороговые значения при аппроксимации и прогнозе соответственно:

$$\mathcal{C}_0^{optA} = \arg \min_{\mathcal{C}_0} \mathcal{D}_{\mathcal{AV}}(\mathcal{C}_0), \quad \mathcal{C}_0^{optF} = \arg \min_{\mathcal{C}_0} \mathcal{D}_{\mathcal{AV}}^{(\mathcal{F})}(\mathcal{C}_0),$$

где  $\mathcal{D}_{\mathcal{AV}}(\mathcal{C}_0)$  — это ошибка аппроксимации  $F^{(T)}$ , введённая в разделе 2.4, в котором  $\mathcal{C}_0^{optA}$  оценивалось при разных  $\alpha$  и  $\sigma$  (там оно обозначалось как  $\mathcal{C}_0^{opt}$ ). Оценим теперь  $\mathcal{C}_0^{optF}$  при тех же параметрах. Процедура оценки  $\mathcal{C}_0^{optF}$  совпадает с процедурой оценки  $\mathcal{C}_0^{optA}$  (алгоритм 2.3), за исключением того, что после выделения  $F_{\mathcal{A}}$  делается ещё прогноз с помощью процедуры TrFORE и квадратичные ошибки рассчитываются только для спрогнозированного элемента  $f_N^{(\mathcal{A})}$ .

**Параметры** Как и в разделе 2.4, возьмём следующие параметры:  $N = 47$ ,  $L = 24$ ,  $R = 10^4$ ,  $\alpha = 0, 0.01, 0.02$ ,  $0 < \sigma \leq 1.6$  с шагом 0.2,  $\omega_0 = 0.05$ . Поиск  $\mathcal{C}_0^{optF}$  проведём при  $\mathcal{C}_0^l = 0.5$ ,  $\mathcal{C}_0^r = 1$ ,  $\Delta\mathcal{C} = 0.01$ .

**Замечание 2.11** При прогнозе с помощью процедуры TrFORE надо брать  $\mathcal{C}_0 > 0$ , так как при  $\mathcal{C}_0 = 0$  будут идентифицированы все собственные тройки

( $\mathcal{I}_A = \{1, \dots, L\}$ ) и, следовательно,  $1 - \sum_{j \in \mathcal{I}_A} \pi_j^2 = 0$  (см. (1.12)), что приводит к невозможности определения коэффициентов ЛРФ<sup>9</sup>.

Опустим результаты, полученные при  $\alpha = 0.01$ , так как они не меняют сделанных далее выводов. Результаты при  $\alpha = 0$  изображены на рис. 2.23, на рис. 2.24 к ним добавлены для сравнения результаты при  $\alpha = 0.02$ . На рис. 2.25 для сравнения приведены результаты для процедуры TRFORE (прогноз) и полученные в разделе 2.4 результаты для процедуры TREND (аппроксимация) при  $\alpha = 0$ .

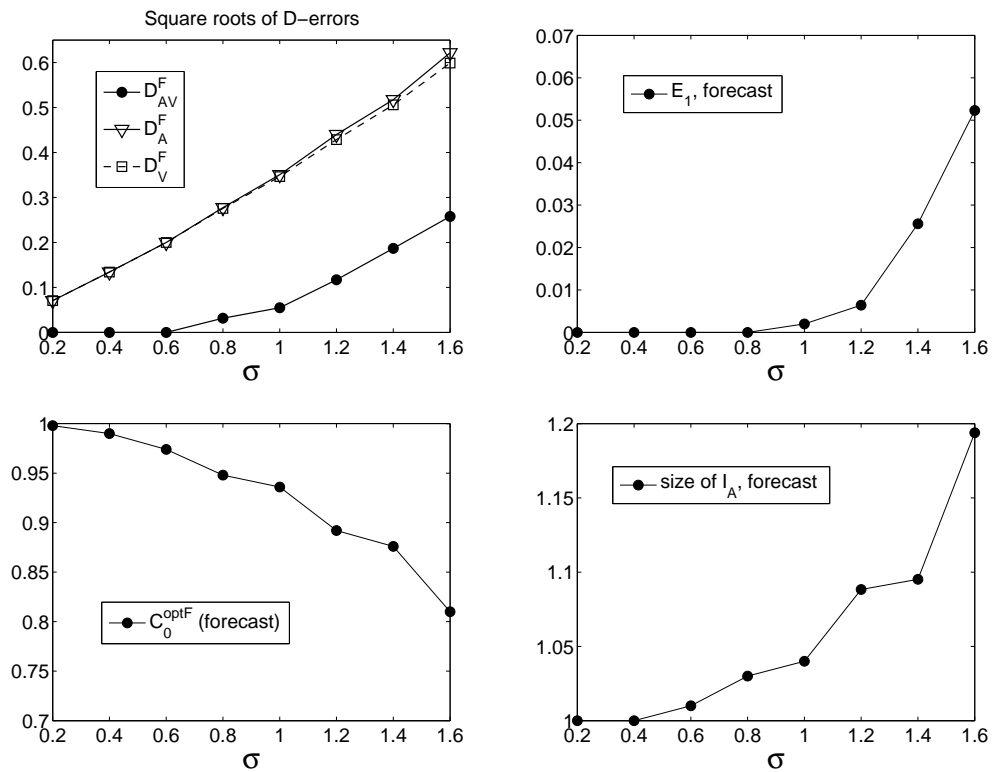


Рис. 2.23: Средние характеристики процедуры TRFORE при  $\alpha = 0$ .

<sup>9</sup>Более подробно это описано в [11, раздел 2.2], [41, раздел 2.1]

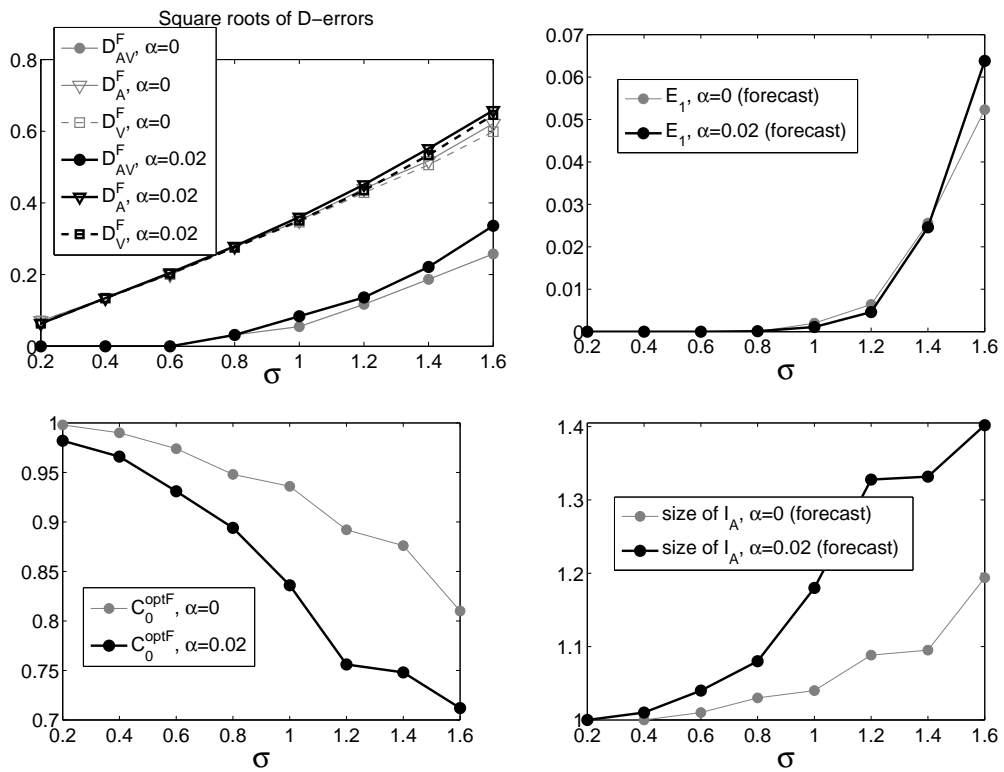


Рис. 2.24: Средние характеристики процедуры TrFORE при  $\alpha = 0$  (серые линии) и  $\alpha = 0.02$  (чёрные жирные линии).

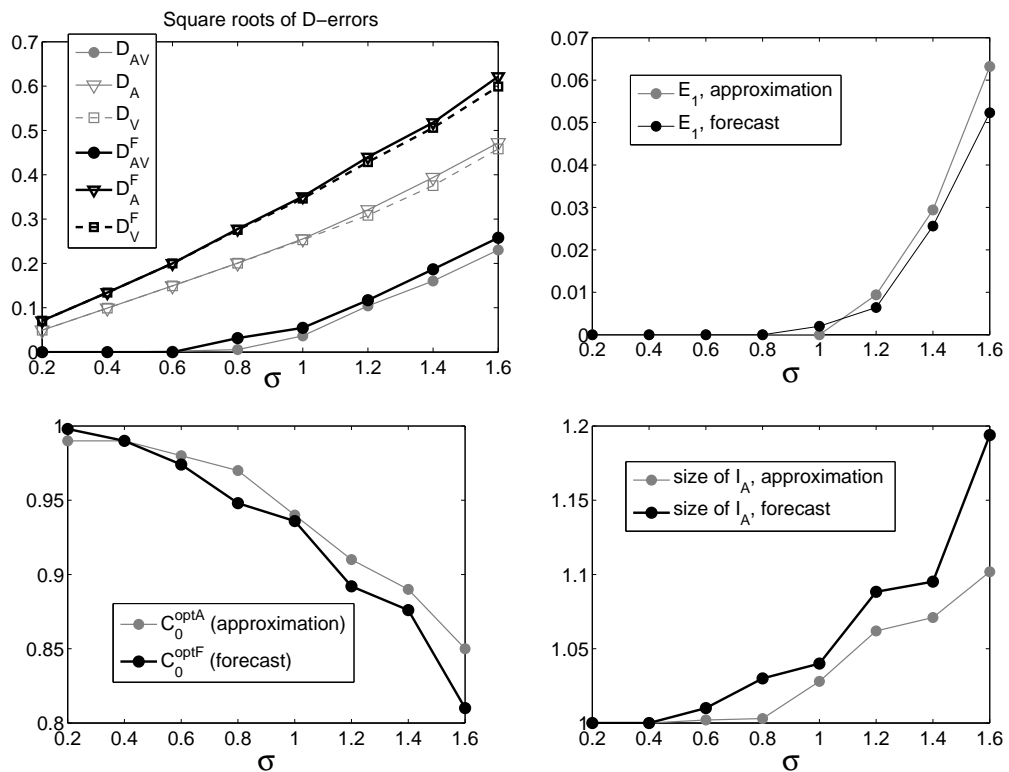


Рис. 2.25: Средние характеристики при  $\alpha = 0$  для процедур TREND (аппроксимация, серые линии) и TrFORE (прогноз, чёрные жирные линии).



**Выводы** Во-первых, видно, что АИ при прогнозе выполняет поставленную задачу — её результаты близки к ВИ и выполненный с её помощью прогноз достаточно близок к настоящему значению  $f_N^{(T)}$  в точке  $N$ . Это подтверждается также малыми значениями ошибки первого рода  $\tilde{\mathcal{E}}_1^{opt\mathcal{F}}$ .

Во-вторых, на графиках рис. 2.24, 2.25, видно, что  $\tilde{\mathcal{E}}_1(\mathcal{C}_0^{opt\mathcal{F}})$  (вероятность ошибки первого рода) и  $\overline{\mathcal{D}}_{AV}^{(\mathcal{F})}(\mathcal{C}_0^{opt\mathcal{F}})$  (квадратичная ошибка АИ по сравнению с ВИ) мало зависят от  $\alpha$  (т. е. процедура TRFORE адаптируется к изменению  $\alpha$ ). Разница в  $\overline{\mathcal{D}}_{\mathcal{A}}(\mathcal{C}_0^{opt\mathcal{A}})$  и  $\overline{\mathcal{D}}_{\mathcal{A}}^{\mathcal{F}}(\mathcal{C}_0^{opt\mathcal{F}})$  обусловлена тем, что при прогнозе (по сравнению с аппроксимацией) даже визуальная процедура даёт бóльшую ошибку.

В-третьих, было обнаружено, что  $\mathcal{C}_0^{opt\mathcal{F}} \approx \mathcal{C}_0^{opt\mathcal{A}}$ . Это означает, что *при прогнозе можно пользоваться той же методикой выбора  $\mathcal{C}_0$ , что и при аппроксимации* (например, выбирать  $\mathcal{C}_0$  с помощью меры  $\mathcal{R}(\mathcal{C}_0)$ , см. раздел 2.5.2).

## Глава 3

### Автоматический метод выделения периодической составляющей

В данной главе опишем метод автоматической идентификации (АИ)  $\varepsilon$ -м гармонических компонент ряда, на базе которого в рамках подхода “Гусеница”-SSA можно построить процедуру выделения как одной  $\varepsilon$ -м гармоники, присутствующей в ряде, так и периодической составляющей, которая задаётся уравнением (0.1).

В главе 1 (раздел 1.3.1) были описаны следующие принципы, на основе которых проводится визуальная идентификация (ВИ) собственных троек, соответствующих периодической составляющей:

- $\varepsilon$ -м гармонике с периодом  $T = 2$  соответствует один собственный вектор, последовательность элементов которого описывается тем же уравнением, что и последовательность элементов самой  $\varepsilon$ -м гармоники (с точностью до фазы),
- $\varepsilon$ -м гармонике с периодом  $T > 2$  соответствуют два собственных вектора, последовательности элементов которых при определённых условиях описываются формулами  $\varepsilon$ -м синуса и  $\varepsilon$ -м косинуса с теми же  $\alpha$ ,  $\omega$ , что и у самой  $\varepsilon$ -м гармоники.

В методе, который будет предложен в данной главе, используется разложение Фурье последовательности элементов собственного вектора. Идея метода была изложена в [68], однако без достаточной проработки: не была решена проблема выбора порогового значения метода, не было проделано исследование качества метода (аналитическое или с помощью моделирования), к тому же метод предназначался для идентификации только немодулированных гармонических составляющих. В данной главе представим доработанный вариант метода, а также устраним эти пробелы и предложим способы выбора порогового значения метода: аналитический при известных ограничениях на параметры  $\varepsilon$ -м гармонических составляющих и эмпирический, который позволяет выбрать пороговое значение во многих случаях, когда это невозможно сделать с помощью теоретически обоснованного метода.

### 3.1 Отличие задачи выделения тренда от задачи выделения периодической составляющей

Задача выделения периодической составляющей принципиально отличается от задачи выделения тренда. Напомним, трендом ряда называется медленно меняющаяся составляющая ряда, которая описывает глобальное поведение ряда и является (или хорошо аппроксимируется) рядом конечного ранга. Ряды конечного ранга — это в некотором смысле параметрическое семейство с неограниченным числом параметров, см. раздел 1.3, а также [10, глава 3], [41, глава 5]. В отличие от тренда, периодическая составляющая  $F^{(P)} = (f_0, \dots, f_{N-1})$  с периодом  $T$  ( $T \geq 2$ ) задаётся параметрической моделью с параметрами  $\{A_k, \alpha_k, \phi_k\}_{k=1}^{\lfloor T/2 \rfloor}$  и  $T$ .

$$f_n = \sum_{k=1}^{\lfloor T/2 \rfloor} A_k e^{\alpha_k n} \cos(2\pi n k / T + \phi_k), \quad A_k, \alpha_k \in \mathbb{R}, \quad \phi_k \in [0, 2\pi). \quad (3.1)$$

Для идентификации компонент тренда используется критерий, который основан не на параметрическом задании тренда, а на свойстве его медленной изменяемости. Плюсом такого подхода является гибкость, поскольку его можно применять для выделения тренда любого параметрического вида, лишь бы это был ряд конечного ранга. Но при этом следует для каждого типа тренда проверять, работает ли для него предложенная процедура TREND (что можно сделать, например, при обработке множества похожих рядов, см. раздел 2.9).

Метод идентификации компонент периодической составляющей может быть построен как метод выделения известных параметрически заданных составляющих —  $\varepsilon$ -м гармоник, что упрощает задачу АИ. Разработаем процедуру, которая будет идентифицировать только собственные вектора, имеющие  $\varepsilon$ -м гармоническую форму или близкую к ней. Естественно, эта процедура будет работать только при достаточно хорошей отделимости периодической составляющей от остатка, но ведь поставленной задачей является автоматизация визуальной процедуры идентификации, которая накладывает эти же ограничения на модель ряда.

### 3.2 Описание метода Фурье для идентификации $\varepsilon$ -м гармонических компонент

Пусть  $F$  — ряд длины  $N$ :  $F = F^{(P)} + F^{(R)}$ , где  $F^{(P)}$  — периодическая функция, заданная уравнением (0.1), а  $F^{(R)}$  — некоторый остаток.

Проведём сингулярное разложение с длиной окна  $L$ . Представим решение проблемы идентификации всех собственных троек, соответствующих  $F^{(P)}$ .

Предположим, что при взятом значении  $L$  остаток  $F^{(R)}$  приближённо разделим с  $F^{(P)}$ . Это предположение выполняется для широкого спектра различных рядов  $F^{(R)}$ : стохастический шум, детерминированная составляющая, тренд (примеры и условия отделимости  $\varepsilon$ -м гармонической компоненты см. в [10, раздел 2], [41, разделы 1.5, 6.1]).

Описанный далее метод автоматической идентификации собственных троек, соответствующих  $\varepsilon$ -м гармоническим составляющим, проверяет

- для каждой пары собственных векторов тот факт, что последовательности их элементов описываются  $\varepsilon$ -м гармоническими функциями с одинаковой частотой,
- для каждого собственного вектора — что последовательность его элементов описывается  $\varepsilon$ -м гармоникой с периодом 2.

**Замечание 3.1** Рассматриваются только пары соседних собственных векторов, с номерами вида  $i, i + 1$  (см. замечание 1.2).

Метод состоит из двух частей, на первом этапе проводится предварительная проверка и собственные тройки, идентифицированные на этом этапе, проходят дополнительную проверку на втором этапе метода.

В виду того, что метод основан на представлении Фурье, назовём его *методом Фурье*.

### 3.2.1 Первая часть метода Фурье

Эта часть метода Фурье основана на том факте, что для гармоники в случае, когда её частота  $\omega$  и длина окна  $L$  такие, что  $L\omega \in \mathbb{N}$ , периодограммы её обоих собственных векторов равны  $\Pi_{U_j}^L(k/L) = \mathcal{X}_\omega(k/L)$  (где  $\mathcal{X}_\omega(\cdot)$  это индикатор множества  $\{\omega\}$ ). Поэтому при идентификации *будем проверять, что аргументы максимумов периодограмм соседних собственных векторов совпадают (что очевидным образом выполняется для гармоники при  $L\omega \in \mathbb{N}$ ) или, в более общем случае, что они близки.*

#### Схема первой части метода Фурье

Для периодограммы  $\Pi_{U_j}^L$   $j$ -го собственного вектора  $U_j$  обозначим аргумент её максимума как  $\theta_j$ :  $\theta_j = \arg \max_{0 \leq k \leq L/2} \{\Pi_{U_j}^L(k/L)\}$ .

Выберем из всех пар собственных троек вида  $i, i + 1$  такие, для которых аргументы максимумов периодограмм их собственных векторов  $U_i, U_{i+1}$  больше 0 (так как рассматриваются гармоники только с  $\omega > 0$ ) и близки, т. е. отличаются друг от друга не более чем на  $s_0$  шагов размером  $1/L$  (где  $s_0 \in \mathbb{Z}_+$  — фиксированный параметр метода Фурье):

$$\mathcal{I}_1^{(P)} = \left\{ (i, i + 1) : \theta_1 \theta_2 > 0, \quad L|\theta_i - \theta_{i+1}| \leq s_0, \quad 1 \leq i \leq L - 1 \right\}. \quad (3.2)$$

Аналогично проверим каждую собственную тройку, не соответствует ли она гармонике с периодом 2:

$$\mathcal{I}_2^{(P)} = \left\{ i : \quad L|\theta_i - 0.5| \leq s_0, \quad 1 \leq i \leq L \right\}. \quad (3.3)$$

Результатом первой части метода Фурье являются множества собственных троек с номерами из  $\mathcal{I}_1^{(P)}, \mathcal{I}_2^{(P)}$ .

**АЛГОРИТМ 3.1. Метод Фурье для автоматической идентификации собственных троек, соответствующих периодической составляющей. Часть 1**

**Входные данные:**  $\{U_j\}_{j=1}^d$  — собственные вектора сингулярного разложения траекторной матрицы некоторого ряда,  $U_j \in \mathbb{R}^L$

**Результат:**

- $\mathcal{I}_1^{(P)}$  — пары номеров собственных троек, идентифицированных как соответствующие  $\varepsilon$ -м гармоникам с  $0 < \omega < 0.5$
- $\mathcal{I}_2^{(P)}$  — номера собственных троек, идентифицированных как соответствующие  $\varepsilon$ -м гармонике с  $\omega = 0.5$

**Параметры:**  $s_0 \in \mathbb{Z}_+$

**Процедура:**

1. Вычислим  $\mathcal{I}_1^{(P)}$  по (3.2)
2. Вычислим  $\mathcal{I}_2^{(P)}$  по (3.3)

### Поведение тестовой функции для гармоника с $\omega: L\omega \notin \mathbb{N}$

Изучим, как ведёт себя тестовая функция  $L|\theta_i - \theta_{i+1}|$  для гармоника с  $L, \omega: L\omega \notin \mathbb{N}$ .

Пусть  $F: f_n = \sin(2\pi\omega n + \phi)$ ,  $1 < L < N$ ,  $0 < \omega < 0.5$ , тогда ему соответствуют два собственных вектора  $U_1, U_2$  ( $U_j = (u_1^{(j)}, \dots, u_L^{(j)})^T$ ,  $j = 1, 2$ ), элементы которых задаются следующими выражениями:

$$u_k^{(1)} = \sin(2\pi\omega n + \psi), \quad u_k^{(2)} = \cos(2\pi\omega n + \psi), \quad k = 1, \dots, L. \quad (3.4)$$

Проведём численный эксперимент, в ходе которого промоделируем последовательности  $U_1, U_2$  по формулам (3.4) при различных  $L, \omega, \psi$  и вычислим

$$\max_{L, \omega, \psi} L|\theta_i - \theta_{i+1}|. \quad (3.5)$$

Были рассмотрены следующие области поиска. Для  $L$ :  $\Omega_L = 2(1)1000$  ( $2, 3, \dots$  и так далее до 1000 с шагом 1), для  $\omega$ :  $\Omega_\omega = 0.01(0.01)0.49$ , для  $\psi$ :  $\Omega_\psi = 0(0.02\pi)0.48\pi$ . Численное вычисление максимума (3.5) для  $L \in \Omega_L, \omega \in \Omega_\omega, \psi \in \Omega_\psi$  дало результат:

$$\max_{L \in \Omega_L, \omega \in \Omega_\omega, \psi \in \Omega_\psi} L|\theta_i - \theta_{i+1}| = 1.$$

### Поведение тестовой функции для э-м гармоника с $\omega: L\omega \in \mathbb{N}$

Периодограмма э-м гармоника отличается от периодограммы немодулированной гармоника тем, что хотя она по прежнему имеет один пик, но он менее выражен, периодограмма имеет ненулевые значения в окрестности пика. Этот эффект называют “растеканием” периодограммы, причём чем больше экспоненциальный показатель  $\alpha$ , тем сильнее он выражен. При этом максимум периодограммы э-м гармоника с частотой  $\omega$  может не совпадать с  $\omega$ .

Поэтому для э-м гармоника даже при  $L\omega \in \mathbb{N}$  не обязательно  $\theta_i = \theta_{i+1}$  и естественно предположить, что чем сильнее модуляция, тем большее отклонение  $\theta_i$  от  $\theta_{i+1}$  будет иметь место.

Возьмём  $U_1, U_2$  :  $u_n^{(1)} = e^{\alpha n} \cos(2\pi\omega n + \psi)$ ,  $u_n^{(2)} = e^{\alpha n} \sin(2\pi\omega n + \psi)$  ( $n = 1, \dots, L$ ) и посчитаем для этих векторов значение  $\max_{\psi} L|\theta_i - \theta_{i+1}|$  при различных  $\alpha, \omega, L$ . Возьмём значения  $\psi \in \Omega_\psi = 0(0.02\pi)0.48\pi$ . Результаты представлены в табл. 3.1. Видно, что  $\max_{\psi \in \Omega_\psi} L|\theta_i - \theta_{i+1}| > 0$  наблюдается при невозможных в реальности значениях  $\alpha, L$ : при  $\alpha = 0.05$ ,  $L = 729$  возрастание  $u_L^{(j)}/u_1^{(j)}$  ( $j = 1, 2$ ) для векторов  $U_1, U_2$  имеют порядок  $10^{15}$ , а при  $\alpha = 0.1$ ,  $L = 243$  — порядок  $10^{10}$ .

Таблица 3.1: Значение  $\max_{\psi} L|\theta_i - \theta_{i+1}|$  для  $\alpha$ -м гармоники с  $\alpha$ ,  $L\omega \in \mathbb{N}$ 

| $\alpha \leq 0.04$ |      | L  |    |     |     | $\alpha \leq 0.07$ |      | L  |    |     |     |
|--------------------|------|----|----|-----|-----|--------------------|------|----|----|-----|-----|
|                    |      | 27 | 81 | 243 | 729 |                    |      | 27 | 81 | 243 | 729 |
| $\omega$           | 1/3  | 0  | 0  | 0   | 0   | $\omega$           | 1/3  | 0  | 0  | 0   | 0   |
|                    | 1/9  | 0  | 0  | 0   | 0   |                    | 1/9  | 0  | 0  | 0   | 0   |
|                    | 1/27 | 0  | 0  | 0   | 0   |                    | 1/27 | 0  | 0  | 0   | 2   |

| $\alpha \leq 0.1$ |      | L  |    |     |     |
|-------------------|------|----|----|-----|-----|
|                   |      | 27 | 81 | 243 | 729 |
| $\omega$          | 1/3  | 0  | 0  | 0   | 0   |
|                   | 1/9  | 0  | 0  | 0   | 2   |
|                   | 1/27 | 0  | 0  | 2   | 5   |

### Поведение тестовой функции для $\alpha$ -м гармоники с $\omega$ : $L\omega \notin \mathbb{N}$

Рассмотрим теперь случай  $\alpha$ -м гармоники, когда  $L$  и  $\omega$  такие, что необязательно выполняется  $L\omega \in \mathbb{N}$ . Посчитаем для различных  $0.01 \leq \alpha \leq 0.06$  значение  $\max_{L,\omega,\psi} L|\theta_i - \theta_{i+1}|$  при тех же  $\omega, \psi$ , что и раньше для немодулированной гармоники:  $\omega \in \Omega_{\omega}$ ,  $\psi \in \Omega_{\psi}$ . Рассмотрим разные  $\Omega_L$ . Результаты приведены в табл. 3.2.

Таблица 3.2: Значение  $\max_{L \in \Omega_L, \omega \in \Omega_{\omega}, \psi \in \Omega_{\psi}} L|\theta_i - \theta_{i+1}|$  для  $\alpha$ -м гармоники с  $\alpha$  при разных  $\Omega_L$ 

|            |           | $\alpha$ |      |      |      |      |      |
|------------|-----------|----------|------|------|------|------|------|
|            |           | 0.01     | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 |
| $\Omega_L$ | 2...100   | 1        | 1    | 1    | 1    | 1    | 1    |
|            | 101...200 | 1        | 1    | 1    | 1    | 2    | 2    |
|            | 201...400 | 1        | 1    | 1    | 2    | 3    | 4    |
|            | 401...600 | 1        | 1    | 2    | 3    | 4    | 5    |

## Выбор $s_0$

Необходимость задания  $s_0 > 0$  определяется следующими тремя факторами: во-первых, тем, что при  $L\omega \notin \mathbb{N}$  аргументы максимумов периодограмм собственных векторов для гармонике не совпадают. Во-вторых, “растеканием” периодограмм собственных векторов, соответствующих э-м гармонике с  $\alpha \neq 0$ .

В предыдущих разделах были рассмотрены эти случаи. Проинтерпретируем результаты табл. 3.2. Видно, при широком спектре значений параметров  $L$ ,  $\alpha$  выполняется  $\max_{L \in \Omega_L, \omega \in \Omega_\omega, \psi \in \Omega_\psi} L|\theta_i - \theta_{i+1}| = 1$ . Примем во внимание, что каждое  $L$  соответствует рядам длины  $N \geq 2L$ . Рассмотрим, насколько возможны в реальности такие сочетания  $N$ ,  $\alpha$ , при которых значение тестовой функции больше единицы. Порядок возрастания амплитуды э-м гармонике при таких параметрах составляет: при  $N = 2L = 802$ ,  $\alpha = 0.03 - 10^{10}$ , при  $N = 2L = 402$ ,  $\alpha = 0.04 - 10^6$ , при  $N = 2L = 202$ ,  $\alpha = 0.05 - 10^4$ . Такое возрастание амплитуды э-м гармонике в реальных рядах не встречается.

Также необходимо учесть, что из-за приближённого характера делимости (как это обычно бывает при реальном исследовании) соответствующие э-м гармонике собственные вектора будут искажены, и значение  $s_0 > 0$  необходимо чтобы учесть искажение их периодограмм.

Из приведённых численных примеров видно, что для выделения э-м гармонике можно выбирать  $s_0 = 1$  и практика показывает, что, несмотря на приближённый характер делимости, выбор такого  $s_0 = 1$  позволяет применять метод Фурье с хорошим качеством.

Далее по умолчанию будет использоваться значение  $s_0 = 1$ .

## Необходимость второй части метода Фурье

Описанный принцип позволяет идентифицировать собственные тройки э-м гармонике, но возможны случаи, когда для собственных векторов, не соответствующих э-м гармонике, разница аргументов максимумов их периодограмм может быть меньше  $s_0$  и множество идентифицированных компонент будет содержать лишние компоненты. Для того чтобы их отбросить, проверим результат первой части метода Фурье во второй части.

**Пример 3.1** Продемонстрируем на примере необходимость дополнительной проверки после идентификации компонент с помощью первой части метода Фурье. Рассмотрим ряд Wheat, содержащий ежегодные данные о ценах на пшеницу, с 1264 г. по 1996 г. (в ценах на 1996 г.), взятый из [50].



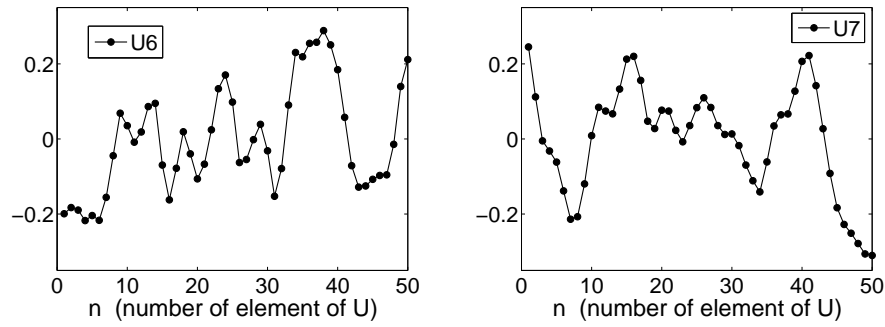


Рис. 3.1: Пример 3.1, графики последовательностей элементов векторов  $U_6$ ,  $U_7$ .

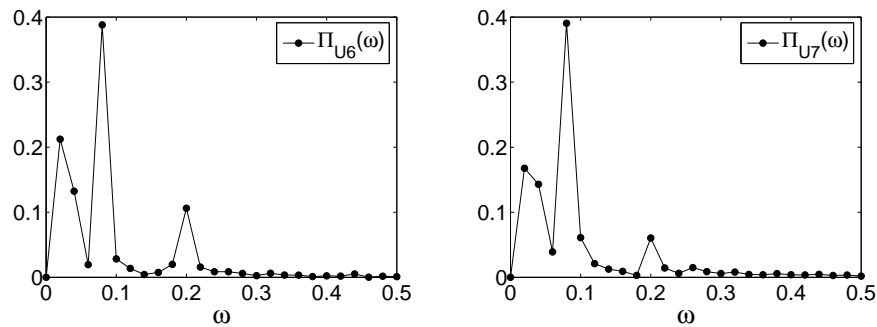


Рис. 3.2: Пример 3.1, периодограммы собственных векторов  $U_6$ ,  $U_7$ .

Возьмём в качестве  $F$  участок ряда с 1661 г. по 1760 г. ( $N = 100$ ) и проведём разложение с  $L = 50$ . Рассмотрим собственные вектора  $U_6$ ,  $U_7$ , изображённые на рис. 3.1. Видно, что их форма далека от формы  $\varepsilon$ -м гармоники, что видно и по их периодограммам (рис. 3.2). Тем не менее, аргументы максимумов этих периодограмм совпадают:

$$\arg \max_k \Pi_{U_6}(k/L) = \arg \max_k \Pi_{U_7}(k/L) = 4/50.$$

### 3.2.2 Вторая часть метода Фурье

В первой части метода Фурье учитывались только аргументы максимумов периодограмм, чего может быть недостаточно. Во второй части проверим

- для каждой пары собственных векторов с номерами из  $\mathcal{I}_1^{(P)}$  — похожи ли их периодограммы на периодограмму некоторой  $\varepsilon$ -м гармоники (с неизвестной частотой),
- для каждого собственного вектора из  $\mathcal{I}_2^{(P)}$  — похожа ли его периодограмма на периодограмму  $\varepsilon$ -м гармоники с частотой  $\omega = 0.5$ .

Идея проверки состоит в том, чтобы учитывать значение максимума периодограммы.

**Утверждение 3.1** Пусть  $F : f_n = A \cos(2\pi\omega n + \phi)$ ,  $\omega \in (0, 0.5)$ . Возьмём длину окна  $L : L\omega \in \mathbb{N}$ . Обозначим собственные вектора траекторной матрицы ряда  $F$ , как  $U_1, U_2 \in \mathbb{R}^L$ . Тогда

$$\max_{0 \leq k \leq L} \Pi_{U_1}(k/L) = \max_{0 \leq k \leq L} \Pi_{U_2}(k/L) = \Pi_{U_1}(\omega) = \Pi_{U_2}(\omega) = 1,$$

причём для любого собственного вектора  $W_j$  ( $j = 1, \dots, L$ ) любого другого ряда  $G \neq F$  будет выполняться

$$\max_{0 \leq k \leq L} \Pi_{W_j}(k/L) < 1.$$

Доказательство следует из утверждения 2.1, из которого следует, что  $\sum_{k=0}^{\lfloor L/2 \rfloor} \Pi_{U_i}^L(k/L) = \|U_i\|_2^2 = 1$ .

Таким образом, для того чтобы проверить, что собственные вектора с номерами  $i, i+1$  соответствуют гармонике, необходимо убедиться, что максимумы их периодограмм достигаются в одной точке и каждый из них при этом равен единице. Э-м гармонике и гармонике с частотой  $\omega : L\omega \notin \mathbb{N}$  тоже будет соответствовать большое значение максимума периодограммы (хоть и меньше 1), так как их периодограммы имеют один чётко выраженный пик.

Этот принцип проверки был предложен в работе [68], однако сама идея не нова, ещё в 1929 г. Фишер в статье [37] предложил использовать высоту пика периодограммы для проверки наличия гармоник в ряде.

Рассмотрим ряд  $F$ , выберем длину окна  $L$  и проведём его сингулярное разложение. Ему соответствуют  $d$  собственных троек ( $d \leq \min(L, K)$ ).

**Определение 3.1** Пусть  $\{U_j\}_{j=1}^L$ ,  $U_j \in \mathbb{R}^L$  — собственные вектора разложения некоторого ряда. Для множества  $\mathcal{A} \subset \{1, \dots, L\}$  мощности  $\#\mathcal{A}$  введём функцию  $\rho_{\mathcal{A}}$ , определённую на множестве  $\{k/L\}_{k=0}^{\lfloor L/2 \rfloor}$ :

$$\rho_{\mathcal{A}}(k/L) = \frac{1}{\#\mathcal{A}} \sum_{j \in \mathcal{A}} \Pi_{U_j}^L(k/L).$$

### Схема второй части метода Фурье

Переберём все собственные тройки и пары собственных троек, найденные в первой части метода и проверим их с помощью следующих критериев, основанных на максимуме периодограммы:

$$\max_{0 \leq k \leq L/2} \rho_{\{i,j\}}(k/L) \geq \rho_0 \quad \text{и} \quad \rho_{\{i\}}(0.5) \geq \rho_0,$$

где  $\rho_0 \in [0, 1]$  — это заранее заданное пороговое значение.

Результатом идентификации с помощью метода Фурье является следующее множество компонент  $I^{(P)}$ :

$$\begin{aligned} \mathcal{I}^{(P)} = & \left\{ (i, j) \in \mathcal{I}_1^{(P)} : \max_{0 \leq k \leq L/2} \rho_{\{i,j\}}(k/L) \geq \rho_0 \right\} \cup \\ & \cup \left\{ i \in \mathcal{I}_2^{(P)} : \rho_{\{i\}}(0.5) \geq \rho_0 \right\}. \end{aligned} \quad (3.6)$$

При выделении гармоник с  $\omega : L\omega \in \mathbb{N}$  пороговое значение  $\rho_0$  следует выбирать близким к единице. В случае точной отделимости можно брать  $\rho_0 = 1$ , но при приближённой разделимости необходимо учесть искажение собственных векторов соответствующих гармонике и выбрать  $\rho_0$  несколько меньше единицы. Вопросу выбора  $\rho_0$  посвящён раздел 3.4.

Изучим, как ведёт себя тестовая функция  $\rho_{\{1,2\}}(k/L)$  для  $\alpha$ -м гармоники с  $\alpha > 0$ , а также для гармоники с частотой  $\omega : L\omega \notin \mathbb{N}$ .

### Поведение критерия для $\alpha$ -м гармоники

Рассмотрим теперь, как ведёт себя  $\rho_{\{1,2\}}(\theta)$  для  $\alpha$ -м гармоники вида (1.5), с  $\alpha > 0$ . Так как для любого собственного вектора  $U$  верно, что  $\sum_{k=0}^{\lfloor L/2 \rfloor} \Pi_U^L(k/L) = 1$  (см. утверждение 2.1), то из-за растекания периодограммы значение  $\rho_{\{1,2\}}(k/L)$  ( $0 \leq k \leq L/2$ ) будет меньше единицы. Попробуем оценить его зависимость от  $\alpha$  и  $L$ , которые задают уровень модуляции.

Рассмотрим  $\alpha$ -м гармонический ряд  $F = (f_0, \dots, f_{N-1})$  длины  $N$  с экспоненциальным показателем  $\alpha$ . Если длина ряда  $N$  и длина окна  $L$  такие, что  $L\omega \in \mathbb{N}$ ,  $L \geq 2$ ,  $K\omega \in \mathbb{N}$ ,  $K \geq 2$  ( $K = N - L + 1$ ), то элементы двух первых собственных векторов  $U_1, U_2 \in \mathbb{R}^L$  разложения ряда  $F$  имеют вид:

$$\begin{aligned} u_n^{(1)} &= A_1(\alpha, \omega, L) e^{\alpha n} \cos(2\pi\omega n + \phi), \\ u_n^{(2)} &= A_2(\alpha, \omega, L) e^{\alpha n} \sin(2\pi\omega n + \phi), \end{aligned} \quad (3.7)$$

где  $n = 0, \dots, L - 1$ . Примем для простоты  $\phi = 0$ . Воспользуемся тем, что  $\|U_1\|_2^2 = \|U_2\|_2^2 = 1$ . Тогда  $A_1 = A_1(\alpha, \omega, L)$  и  $A_2 = A_2(\alpha, \omega, L)$  равны

$$A_1 = \left( \sum_{n=0}^{L-1} e^{2\alpha n} \cos^2(2\pi\omega n) \right)^{-1/2}, \quad A_2 = \left( \sum_{n=0}^{L-1} e^{2\alpha n} \sin^2(2\pi\omega n) \right)^{-1/2}. \quad (3.8)$$

Согласно определению 3.1,  $\rho_{\{1,2\}}(\omega) = \frac{1}{2} (\Pi_{U_1}^L(\omega) + \Pi_{U_2}^L(\omega))$ . Введём обозначение  $\rho(\alpha, L) = \rho_{\{1,2\}}(\omega)$ .

Сосчитаем значения периодограмм векторов  $U_1$  и  $U_2$  в точке  $\omega$ .

Обозначим через  $H_0 = (h_0^{(0)}, \dots, h_{L-1}^{(0)})$  и  $H_1 = (h_0^{(1)}, \dots, h_{L-1}^{(1)})$  ряды, задаваемые следующими функциями:

$$h_n^{(0)} = \cos(2\pi n\omega), \quad h_n^{(1)} = \sin(2\pi n\omega), \quad 0 \leq n \leq L-1.$$

Для вектора  $U \in \mathbb{R}^L$ ,  $U = (u_0, \dots, u_{L-1})^T$  и ряда  $H = (h_0, \dots, h_{L-1})$  длины  $L$  обозначим как  $(U, H)$  скалярное произведение вектора  $U$  с вектором  $(h_0, \dots, h_{L-1})^T$ . Из вида разложения Фурье (2.2) и так как  $\sum_{n=0}^{L-1} f_n \cos(2\pi nk/L) = \frac{L}{2} c_k$ , видно, что при  $\omega: L\omega \in \mathbb{N}$  значение периодограммы вектора  $U_j$  в  $\omega$  выражается следующим образом:

$$\Pi_{U_j}^L(\omega) = \frac{2}{L} \left\{ (U_j, H_0)^2 + (U_j, H_1)^2 \right\}, \quad j = 1, 2.$$

Тогда для  $\rho(\alpha, L)$  верно:

$$\rho(\alpha, L) = \frac{1}{L} \left\{ (U_1, H_0)^2 + (U_1, H_1)^2 + (U_2, H_0)^2 + (U_2, H_1)^2 \right\}. \quad (3.9)$$

Введём следующие обозначения:

$$\begin{aligned} \Sigma_1(\alpha, L) &= \sum_{n=0}^{L-1} e^{\alpha n} \cos^2(2\pi\omega n), \\ \Sigma_2(\alpha, L) &= \sum_{n=0}^{L-1} e^{\alpha n} \cos(2\pi\omega n) \sin(2\pi\omega n), \quad \Sigma_3(\alpha, L) = \Sigma_1(2\alpha, L), \\ \Sigma_4(\alpha, L) &= \sum_{n=0}^{L-1} e^{\alpha n} \sin^2(2\pi\omega n), \quad \Sigma_5(\alpha, L) = \Sigma_4(2\alpha, L). \end{aligned}$$

Тогда (3.9) можно записать в следующем виде:

$$\rho(\alpha, L) = \frac{1}{L} \left\{ \frac{\Sigma_1^2(\alpha, L) + \Sigma_2^2(\alpha, L)}{\Sigma_3(\alpha, L)} + \frac{\Sigma_2^2(\alpha, L) + \Sigma_4^2(\alpha, L)}{\Sigma_5(\alpha, L)} \right\}. \quad (3.10)$$

Оценим суммы в (3.10) с помощью интегралов. Обозначим за  $J_j(\alpha, L)$ ,  $j = 1, \dots, 5$ , следующие интегралы:

$$\begin{aligned} J_1(\alpha, L) &= \int_0^L e^{\alpha t} \cos^2(2\pi t\omega) dt, \\ J_2(\alpha, L) &= \int_0^L e^{\alpha t} \cos(2\pi t\omega) \sin(2\pi t\omega) dt, \quad J_3(\alpha, L) = J_1(2\alpha, L), \\ J_4(\alpha, L) &= \int_0^L e^{\alpha t} \sin^2(2\pi t\omega) dt, \quad J_5(\alpha, L) = J_4(2\alpha, L). \end{aligned}$$

Введём

$$J(\alpha, L) = \frac{1}{L} \left\{ \frac{J_1^2(\alpha, L) + J_2^2(\alpha, L)}{J_3(\alpha, L)} + \frac{J_2^2(\alpha, L) + J_4^2(\alpha, L)}{J_5(\alpha, L)} \right\}. \quad (3.11)$$

**Утверждение 3.2** Пусть  $\omega \in (0, 0.5)$ ,  $L \in \mathbb{N}$ :  $L\omega \in \mathbb{N}$ ,  $\alpha > 0$ . Если  $L \rightarrow \infty$ ,  $\alpha \rightarrow 0$ :  $L\alpha \rightarrow \gamma \in \mathbb{R}$ , ( $\gamma \geq 0$ ), тогда

$$J(\alpha, L) \rightarrow \begin{cases} \frac{2(e^\gamma - 1)}{\gamma(e^\gamma + 1)}, & \gamma > 0, \\ 1, & \gamma = 0. \end{cases} \quad (3.12)$$

*Доказательство.* Так как [14, формула 2.663.4]

$$\int e^{ax} \cos^2(bx) dx = \frac{e^{ax}}{2a} + \frac{e^{ax}}{a^2 + 4b^2} \left( \frac{a}{2} \cos(2bx) + b \sin(2bx) \right),$$

то получаем, что

$$\int e^{\beta t} \cos^2(ct) dt = \frac{e^{\beta t}}{2\beta(\beta^2 + 4c^2)} (\beta^2 \cos(2ct) + 2\beta c \sin(2ct) + \beta^2 + 4c^2). \quad (3.13)$$

Соответственно, при  $c = 2\pi\omega$ :

$$J_1(\alpha, \omega, L) = \frac{\alpha^2 + 2c^2}{\alpha(\alpha^2 + 4c^2)} (e^{\alpha L} - 1), \quad J_3(\alpha, L) = \frac{2\alpha^2 + c^2}{4\alpha(\alpha^2 + c^2)} (e^{2\alpha L} - 1),$$

$$J_4(\alpha, L) = \frac{2c^2}{\alpha(\alpha^2 + 4c^2)} (e^{\alpha L} - 1), \quad J_5(\alpha, L) = \frac{c^2}{4\alpha(\alpha^2 + c^2)} (e^{2\alpha L} - 1).$$

Из того что  $J_2(\alpha, L)$  и  $J_4(\alpha, L)$  связаны следующим соотношением:

$$J_2(\alpha, L) = -\frac{\alpha}{2c} J_4(\alpha, L),$$

находим, что

$$J_3(\alpha, L) = \frac{-c^2}{c(\alpha^2 + 4c^2)} (e^{\alpha L} - 1).$$

Подставив полученные значения для  $J_j(\alpha, L)$ ,  $j = 1, \dots, 5$ , в (3.11) и устремив  $L \rightarrow \infty$ ,  $\alpha \rightarrow 0$ ,  $L\alpha \rightarrow \gamma$ , получаем искомый предел:

$$J(\alpha, L) \rightarrow \frac{2(e^\gamma - 1)}{\gamma(e^\gamma + 1)}.$$

■

Таким образом, с помощью приближения интегралами получили оценку для  $\rho_{\{1,2\}}(\omega)$  для э-м гармонического ряда с частотой  $\omega \in (0, 0.5)$  и экспоненциальным показателем  $\alpha > 0$  при  $L$ :  $L\omega \in \mathbb{N}$ :

$$\rho_{\{1,2\}}(\omega) \cong \frac{2(e^\gamma - 1)}{\gamma(e^\gamma + 1)}, \quad \text{где } \gamma = L\alpha. \quad (3.14)$$

На рис. 3.3 приведены графики зависимости функции (3.14) от  $\gamma$ , и, в качестве примеров — графики её зависимости от  $\alpha$  при различных  $L$ .

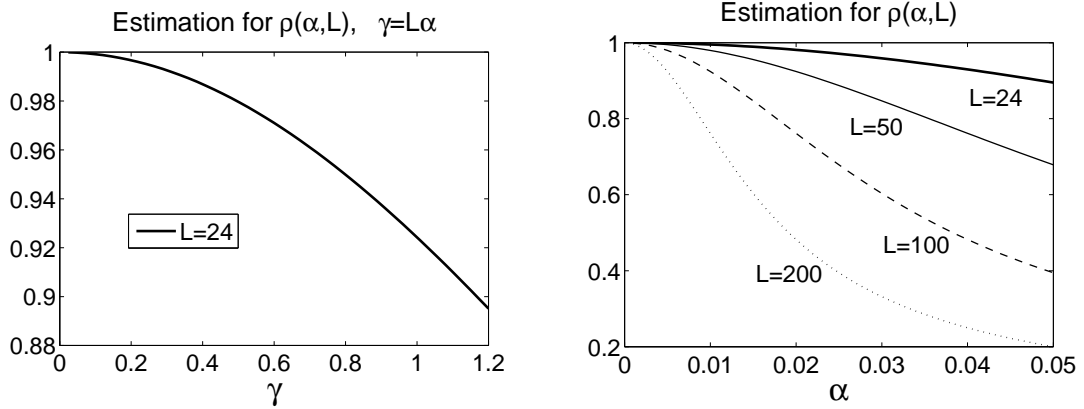


Рис. 3.3: Графики зависимости оценки (3.14) от  $\gamma$  (слева) и от  $\alpha$  при различных  $L$  (справа).

### Поведение тестовой функции для гармоник с $\omega: L\omega \notin \mathbb{N}$

Гармонике с  $\omega: L\omega \notin \mathbb{N}$  соответствуют две собственные тройки с собственными векторами, элементы которых задаются гармониками с частотой  $\omega$ .

Как было замечено ранее, периодограммы этих собственных векторов похожи на  $\mathcal{X}_\omega$  — характеристическую функцию множества  $\{\omega\}$ , но отличаются от неё тем, что имеет ненулевые значения в точках, близких к  $\omega$ . Вследствие этого, а также принимая во внимание утверждение 2.1, значение рассматриваемой тестовой функции  $\max_k \rho_{\{1,2\}}(k/L)$  будет меньше единицы.

Для того, чтобы выяснить, насколько будет меньше значение тестовой функции для такой гармоник, приведём численно полученный график зависимости от  $\omega$  тестовой функции  $\max_k \rho_{\mathcal{I}}(k/L)$  для ряда  $F_N(\omega) = (f_0, \dots, f_{N-1}) : f_n = \cos(2\pi\omega n)$ .

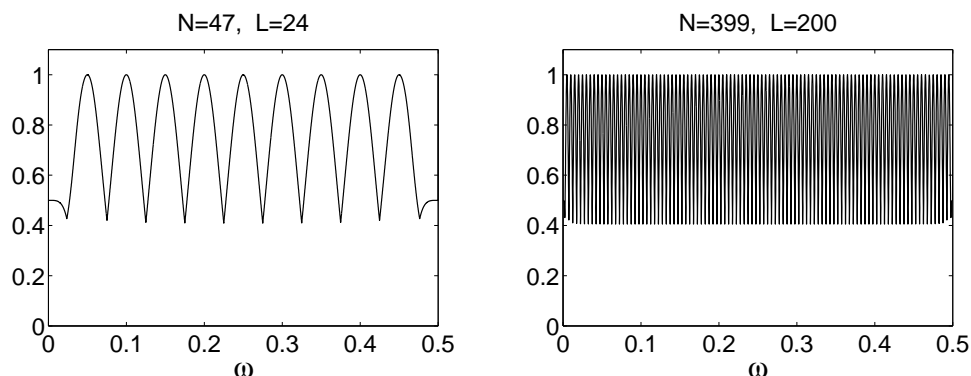


Рис. 3.4: Зависимость от  $\omega$  значения  $\max_k \rho_{\mathcal{I}}(k/L)$  для гармоник с частотой  $\omega$  при  $N = 39, L = 20$  (слева),  $N = 399, L = 200$  (справа).

Были рассмотрены ряды  $F_N(\omega)$  с  $N = 39$  и  $N = 399$ , и перебиралось  $\omega \in (0, 0.5]$  с шагом  $10^{-4}$ . Для каждого ряда проводилось сингулярное разложение с  $L = 20$  для  $N = 39$  и  $L = 200$  для  $N = 399$ , и рассчитывалось значение  $\max_k \rho_{\mathcal{I}}(k/L)$ . Для гармоник с частотами  $\omega < 0.5$  использовалось  $\mathcal{I} = \{1, 2\}$ , а с частотой  $\omega = 0.5$  —  $\mathcal{I} = \{1\}$ .

График зависимости  $\max_k \rho_{\mathcal{I}}(k/L)$  от  $\omega$  приведён на рис. 3.4. Видно, что значение  $\max_k \rho_{\mathcal{I}}(k/L)$  равно 1 только для гармоник с частотой  $\omega$  вида  $k/L$ ,  $k \in \mathbb{Z}$ . Для гармоник же с  $\omega : L\omega \notin \mathbb{N}$  значение  $\max_k \rho_{\mathcal{I}}(k/L)$  может быть значительно меньше единицы, вплоть до 0.4 (независимо от  $N, L$ ).

Это плохо, так как в случае выделения гармоники с неизвестной частотой необходимо выбирать  $\rho_0 \cong 0.4$ , а для выделения  $\varepsilon$ -м гармоники  $\rho_0$  должно быть ещё меньше. При таких значениях  $\rho_0$  критерий, применяемый во второй части метода Фурье, будет слабым и вырастет ошибка за счёт ошибочно идентифицируемых собственных троек.

### Вариант критерия для случая, когда частота $\omega$ неизвестна

Предложим здесь модифицированный вариант тестовой функции, который не имеет данного недостатка и построим на его основе вторую часть метода Фурье. Далее, в разделе 3.3, с помощью статистического моделирования будет показано, что предложенная модифицированная тестовая функция лучше используемой в (3.6).

Идея решения проблемы малого максимума периодограммы основана на том факте, что для  $\varepsilon$ -м последовательности длины  $L$  при  $\omega : L\omega \notin \mathbb{N}$  растекание периодограммы происходит в основном по двум соседним с  $\omega$  частотам (точкам вида  $k/L$ ,  $k \in \mathbb{Z}$ ). Поэтому для идентификации собственных троек, соответствующих  $\varepsilon$ -м гармоникам с частотой  $\omega \in (0, 0.5)$ , будем вместо предложенной ранее тестовой функции  $\max_k \rho_{\{i,j\}}(k/L)$  использовать следующую

$$\max_k \left\{ \rho_{\{i,j\}}(k/L) + \rho_{\{i,j\}}((k+1)/L) \right\}. \quad (3.15)$$

Тестовая функция для идентификации собственной тройки  $\varepsilon$ -м гармоники с частотой  $\omega = 0.5$  выглядит, соответственно, как

$$\rho_{\{i\}}(\lfloor L/2 \rfloor) + \rho_{\{i\}}(\lfloor L/2 \rfloor + 1). \quad (3.16)$$

Приведём графики зависимости от  $\omega$  данной тестовой функции для гармонического ряда с частотой  $\omega$ , с теми же параметрами, что ранее:  $N = 39$ ,  $L = 20$  и  $N = 399$ ,  $L = 200$ ,  $\omega \in (0, 0.5]$  изменяется с шагом  $10^{-4}$ ,  $\mathcal{I} = \{1, 2\}$

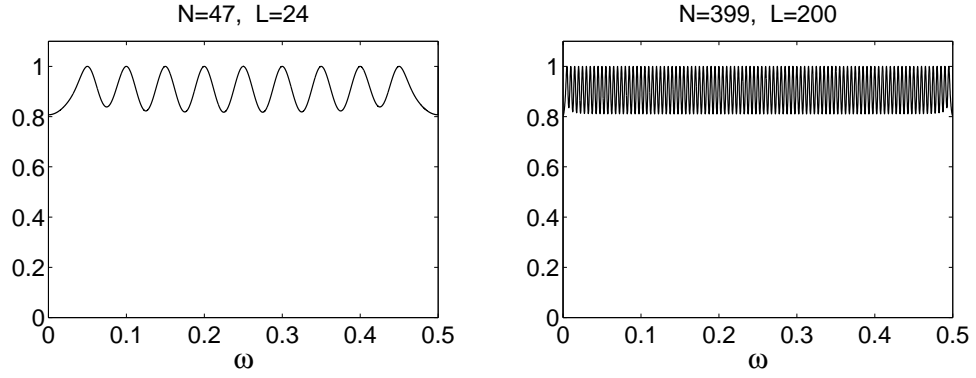


Рис. 3.5: Зависимость от  $\omega$  значения  $\max_k \{ \rho_{\mathcal{I}}(k/L) + \rho_{\mathcal{I}}((k+1)/L) \}$  для гармоника с частотой  $\omega$  при  $N = 39$ ,  $L = 20$  (слева),  $N = 399$ ,  $L = 200$  (справа).

при  $0 < \omega < 0.5$  и  $\mathcal{I} = \{1\}$  при  $\omega = 0.5$ . Видно, что минимум данной функции достигает приблизительно 0.8. Это лучше для выбора  $\rho_0$ , чем 0.4 (можно брать  $\rho_0 \cong 0.8$ , а не  $\rho_0 \cong 0.4$ ).

Суммировать больше чем по двум соседним  $\rho_{\mathcal{I}}(k/L)$  нецелесообразно, так как это увеличивает вероятность ошибки второго рода (идентификации лишних собственных троек), а изменение минимума тестовой функции не столь значительно (0.86 против 0.8), как при суммировании по двум соседним (0.4 против 0.8), что демонстрируют графики на рис. 3.6.

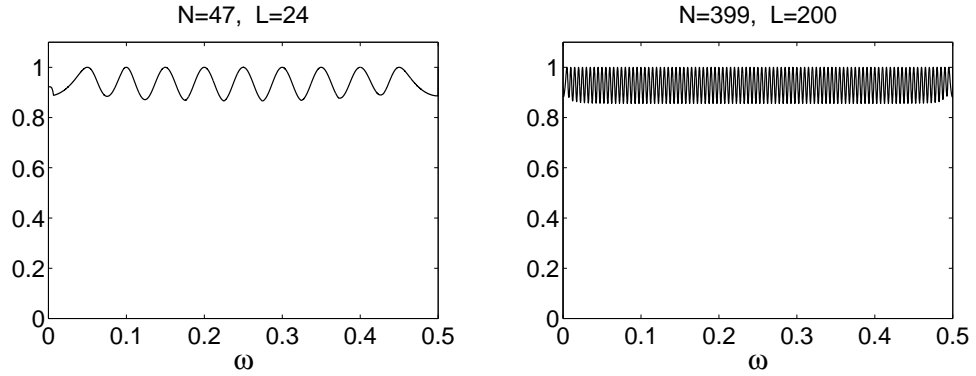


Рис. 3.6: Зависимость от  $\omega$  значения  $\max_k \{ \rho_{\mathcal{I}}(k/L) + \rho_{\mathcal{I}}((k+1)/L) + \rho_{\mathcal{I}}((k+2)/L) \}$  для гармоника с частотой  $\omega$  при  $N = 39$ ,  $L = 20$  (слева),  $N = 399$ ,  $L = 200$  (справа).

Предложим тестовые функции (3.15), (3.16) для использования в методе Фурье при неизвестном  $\omega$ :

$$\mathcal{I}^{(P)} = \left\{ (i, j) \in \mathcal{I}_1^{(P)} : \max_{0 \leq k < L/2} \left\{ \rho_{\{i,j\}}(k/L) + \rho_{\{i,j\}}((k+1)/L) \right\} \geq \rho_0 \right\} \cup \left\{ i \in \mathcal{I}_2^{(P)} : \rho_{\{i\}}(\lfloor L/2 \rfloor) + \rho_{\{i\}}(\lfloor L/2 \rfloor + 1) \geq \rho_0 \right\}. \quad (3.17)$$



При использовании критерия (3.17) вместо (3.6) количество ошибочно идентифицированных компонент должно возрасти. Для того чтобы решить, какой вариант обеспечивает лучшее качество, сравним их с помощью моделирования позже, в разделе 3.3.

**Замечание 3.2** Заметим, что идентифицированные пары собственных троек при анализе реальных рядов, вообще говоря, могут пересекаться.

**АЛГОРИТМ 3.2. Метод Фурье для автоматической идентификации компонент, соответствующих периодической составляющей. Часть 2**

**Входные данные:**  $\mathcal{I}_1^{(P)}$ ,  $\mathcal{I}_2^{(P)}$  — результат алгоритма 3.1,  
 $\{U_j\}_{j \in \mathcal{I}_1^{(P)} \cup \mathcal{I}_2^{(P)}}$  — собственные вектора с номерами из  $\mathcal{I}_1^{(P)}$ ,  $\mathcal{I}_2^{(P)}$

**Результат:**  $\mathcal{I}^{(P)}$  — множество собственных троек

**Параметры:**  $\rho_0 \in [0, 1]$

**Процедура:**

1. Вычислим  $\mathcal{I}^{(P)}$  по  $\{U_j\}_{j \in \mathcal{I}_1^{(P)} \cup \mathcal{I}_2^{(P)}}$  с помощью (3.17)

**АЛГОРИТМ 3.3. Процедура PER: выделение периодической составляющей ряда с помощью метода Фурье**

**Входные данные:**  $F$  — ряд длины  $N$ , содержащий периодическую составляющую  $F^{(P)}$

**Результат:**  $\mathcal{I}_A$  — номера идентифицированных собственных троек,  $F_A$  — полученная аппроксимация периодической составляющей  $F^{(P)}$

**Параметры:**  $L$  — длина окна ( $1 < L \leq \lceil N/2 \rceil$ ), пороговые значения  $s_0 \in \mathbb{Z}_+$ ,  $\rho_0 \in [0, 1]$

**Процедура:**

1. Применим процедуру SSADEC к ряду  $F$ , получив собственные вектора  $\{U_j\}_{j=1}^d$  (алгоритм 1.1)
2. Применим к  $\{U_j\}_{j=1}^d$  первую часть метода Фурье, получив, получив  $\mathcal{I}_1^{(P)}$   $\mathcal{I}_2^{(P)}$  (алгоритм 3.1)
3. По  $\mathcal{I}_1^{(P)}$ ,  $\mathcal{I}_2^{(P)}$  и  $\{U_j\}_{j=d}^d$  вычислим  $\mathcal{I}_A = \mathcal{I}^{(P)}$  с помощью второй части метода Фурье (алгоритм 3.2)
4. Восстановим с помощью процедуры SSAREC составляющую  $F_A$  по  $\mathcal{I}_A$  (алгоритм 1.2)

### 3.3 Проверка процедуры PER для модели с известной периодической составляющей

Проведём статистическое исследование процедуры PER. Рассмотрим ряды следующей модели:  $\varepsilon$ -м гармоника плюс шум. Для разных параметров  $\varepsilon$ -м гармоники получим реализации зашумлённых рядов, вычислим лучшее в среднем пороговое значение  $\rho_0$  и рассчитаем средние характеристики процедуры PER при этом  $\rho_0$ .

Покажем, что при таком  $\rho_0$  метод Фурье выполняет поставленную задачу — идентифицирует собственные тройки, соответствующие  $\varepsilon$ -м гармонике. Кроме того, изучим зависимость средней ошибки выделения  $\varepsilon$ -м гармоники от  $\rho_0$  и сравним критерии (3.17) и (3.6), предложенные для использования во второй части метода Фурье.

**Рассматриваемая модель ряда** Рассмотрим ряд  $F = (f_0, \dots, f_{N-1})$  длины  $N$ , являющийся суммой  $\varepsilon$ -м гармонического ряда с  $\omega$ :  $0 < \omega < 0.5$  и реализации белого нормального шума с изменяющейся дисперсией:

$$f_n = f_n^{(P)} + f_n^{(\varepsilon)}, \quad f_n^{(P)} = e^{\alpha n} \cos(2\pi\omega n), \quad f_n^{(\varepsilon)} = \sigma e^{\alpha n} \varepsilon_n, \quad (3.18)$$

где  $\varepsilon_n \sim NID(0, 1)$ , т. е. соответствует белому нормальному шуму с единичной дисперсией. Это опять же аддитивная модель, но со стандартом шума, который меняется пропорционально амплитуде гармоники. Поэтому (3.18) можно рассматривать в некотором роде как аддитивную запись мультипликативной модели ( $f_n = e^{\alpha n}(\cos(2\pi\omega n) + \sigma\varepsilon_n)$ ).

Данная модель изменения дисперсии шума достаточно естественна, поскольку отражает поведение процессов, наблюдаемых в реальной жизни.

**Оценка качества АИ** Получим с помощью моделирования реализации ряда  $F$  и для каждой из них при фиксированном  $\rho_0$  выделим  $\varepsilon$ -м гармонику с помощью процедуры PER с пороговым значением  $\rho_0$  и оценим среднее качество автоматической идентификации.

Естественно сравнивать автоматическую идентификацию (АИ) с визуальной (ВИ), но ВИ нереализуема при моделировании и поэтому будем её имитировать, взяв первые две собственные тройки. Это имеет смысл при широком спектре параметров, задающих уровень сигнал/шум. Рассмотрим такие параметры модели, при которых не только наблюдается достаточное качество разделимости, но и вклад каждой из собственных троек  $\varepsilon$ -м гармоники больше вклада

собственных троек шума, что делает используемую имитацию ВИ обоснованной.

Покажем, что при наилучшем в среднем  $\rho_0$  для рядов модели (3.18):

- используемый в методе Фурье критерий (3.17) обеспечивает лучшее качество идентификации, чем критерий (3.6),
- результаты АИ близки к результатам имитируемой процедуры ВИ,
- АИ обеспечивает хорошее качество аппроксимации.

Результаты данного раздела были опубликованы в работах [6, 22].

Оценим качество АИ так же, как это делали при выделении тренда, см. раздел 2.4. Обозначим номера компонент — результат АИ, как  $\mathcal{I}_A$ . Как было сказано выше, для имитации ВИ возьмём  $\mathcal{I}_V = \{1, 2\}$ .

Напомним, что для ряда  $F$  качество АИ определяется малостью ошибки АИ по сравнению с ВИ, потому что главное требование к автоматической процедуре — близость к визуальной процедуре. Эта ошибка рассчитывается следующим образом:

$$\mathcal{D}_{AV} = \mathcal{D}(F_A, F_V) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-2\alpha n} (f_n^{(A)} - f_n^{(V)})^2,$$

где  $F_A = (f_0^{(A)}, \dots, f_{N-1}^{(A)})$  — составляющая ряда  $F$ , восстановленная по  $\mathcal{I}_A$  — результату автоматической идентификации, а  $F_V = (f_0^{(V)}, \dots, f_{N-1}^{(V)})$  — составляющая ряда  $F$ , восстановленная по  $\mathcal{I}_V$ .

Получим также абсолютные ошибки автоматической и визуальной идентификаций ( $\mathcal{D}_A$  и  $\mathcal{D}_V$  соответственно) по тем же формулам, что и для тренда, см. раздел 2.4.

### 3.3.1 Расчёт ошибки АИ при наилучшем в среднем $\rho_0$ с помощью моделирования

**Схема оценки ошибки автоматической идентификации** Схема оценки ошибки АИ совпадает с той, которая была использована в примере с экспоненциальным трендом.

Для ряда  $F$  обозначим через  $\mathcal{I}_A(\rho_0)$  группу собственных троек, идентифицированных с пороговым значением  $\rho_0$ ,  $F_A(\rho_0)$  — составляющую, восстановленную по  $\mathcal{I}_A(\rho_0)$ . Значение  $\mathcal{D}(F_V, F_A(\rho_0))$  описывает зависимость ошибки процедуры АИ от  $\rho_0$ . Обозначим множество значений  $\rho_0$ , на котором производится перебор, за  $\Omega$ .

Получим  $R$  ( $R \in \mathbb{N}$ ) рядов  $F^{(r)}$  — реализаций ряда  $F$  согласно модели (3.18) и для каждого  $\rho_0 \in \Omega$  по выборке рядов  $\{F^{(r)}\}_{r=1}^R$  оценим среднюю ошибку АИ по отношению к ВИ, взяв выборочное среднее:

$$\bar{\mathcal{D}}_{AV}(\rho_0) = \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F_A^{(r)}(\rho_0), F_V^{(r)}).$$

Обозначим значение  $\rho_0$ , при котором достигается минимум средней ошибки, за  $\rho_0^{opt}$  (естественно считать это значение наилучшим в среднем для модели ряда с взятыми параметрами):

$$\rho_0^{opt} = \arg \min_{\rho_0} \bar{\mathcal{D}}_{AV}(\rho_0).$$

Вычислим следующие средние характеристики процедуры PER с пороговым значением  $\rho_0^{opt}$ :

$$\bar{\mathcal{D}}_{AV}^{opt} = \bar{\mathcal{D}}_{AV}(\rho_0^{opt}), \quad \bar{\mathcal{D}}_A^{opt} = \bar{\mathcal{D}}_A(\rho_0^{opt}) = \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F_A^{(r)}(\rho_0^{opt}), F^{(P)}). \quad (3.19)$$

Как и для тренда, сравним также непосредственный результат АИ и ВИ — группы идентифицированных компонент  $\{\mathcal{I}_A^{(r)}\}_{r=1}^R$  и  $\mathcal{I}_V = \{1, 2\}$ . Для этого вычислим  $\overline{\#\mathcal{I}_A}^{opt}$  — среднее количество идентифицированных собственных троек при  $\rho_0 = \rho_0^{opt}$  и  $\tilde{\mathcal{E}}_1^{opt}$  — вероятность ошибки первого рода при  $\rho_0 = \rho_0^{opt}$  (когда  $\mathcal{I}_A^{(r)}(\rho_0^{opt}) \not\supset \{1, 2\}$ ):

$$\begin{aligned} \overline{\#\mathcal{I}_A}^{opt} &= \overline{\#\mathcal{I}_A}(\rho_0^{opt}) = \frac{1}{R} \sum_{r=1}^R \#\mathcal{I}_A^{(r)}(\rho_0^{opt}), \\ \tilde{\mathcal{E}}_1^{opt} &= \tilde{\mathcal{E}}_1(\rho_0^{opt}) = \frac{\#\{r : \mathcal{I}_A^{(r)}(\rho_0^{opt}) \not\supset \{1, 2\}\}}{R}, \end{aligned} \quad (3.20)$$

где  $\#$  обозначает мощность множества.

Приведём результаты вычислений  $\rho_0^{opt}$  и  $\bar{\mathcal{D}}_{AV}^{opt}$ ,  $\bar{\mathcal{D}}_A^{opt}$ ,  $\bar{\mathcal{D}}_V$ ,  $\overline{\#\mathcal{I}_A}^{opt}$  и  $\tilde{\mathcal{E}}_1^{opt}$  для различных  $\alpha$ ,  $\sigma$ , полученные с помощью схемы, описанной в алгоритме 3.4.

**АЛГОРИТМ 3.4. Оценка характеристик процедуры PER при наилучшем в среднем  $\rho_0$  ( $\rho_0^{opt}$ ) для  $\varepsilon$ -м гармоники в присутствии шума**

**Результат:**  $\rho_0^{opt}$  — наилучшее в среднем  $\rho_0$ , оценки характеристик процедуры PER:  $\overline{\mathcal{D}}_{AV}^{opt}$ ,  $\overline{\mathcal{D}}_A^{opt}$ ,  $\overline{\mathcal{D}}_V$ ,  $\overline{\#\mathcal{I}}_A^{opt}$ ,  $\tilde{\mathcal{E}}_1^{opt}$

**Параметры:**

- Основные параметры:  $R$ ,  $\rho_0^l$ ,  $\rho_0^h$ ,  $\Delta\rho$
- Параметры модели ряда:  $N$ ,  $\alpha$ ,  $\omega$ ,  $\sigma$
- Параметры процедуры PER:  $L$ ,  $s_0$

**Процедура:**

1. Смоделируем  $R$  реализаций  $\{F^{(r)}\}_{r=1}^R$  по (3.18)
2.  $\Omega = \{\rho_0^l + k\Delta\rho\}_{k=0}^M$ ,  $M = \left\lceil \frac{\rho_0^h - \rho_0^l}{\Delta\rho} \right\rceil$
3. Для каждого смоделированного ряда  $F^{(r)}$  и каждого  $\rho_0 \in \Omega$ :
  - 1) применим процедуру PER с параметрами  $L$ ,  $s_0$ ,  $\rho_0$  к ряду  $F^{(r)}$ , получив  $\mathcal{I}_A^{(r)}(\rho_0)$ ,  $F_A^{(r)}(\mathcal{C}_0)$  (алгоритм 3.3)
  - 2) получим составляющую ряда  $F$  по  $\mathcal{I}_V = \{1\}$  с помощью процедур SSADEC (с параметром  $L$ ), SSAREC (с параметром  $\mathcal{I}_V$ ) (алгоритмы 1.1, 1.2)
  - 3) вычислим ошибку  $\mathcal{D}(F_A^{(r)}(\rho_0), F_V^{(r)})$
4. Для каждого  $\rho_0 \in \Omega$  вычислим
 
$$\overline{\mathcal{D}}_{AV}(\rho_0) = \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F_A^{(r)}(\rho_0), F_V^{(r)})$$
5. Найдём  $\rho_0^{opt} = \max \{ \arg \min_{\rho_0} \overline{\mathcal{D}}_{AV}(\rho_0) \}$
6. Рассчитаем  $\overline{\mathcal{D}}_{AV}^{opt}$ ,  $\overline{\mathcal{D}}_A^{opt}$ ,  $\overline{\mathcal{D}}_V$ ,  $\overline{\#\mathcal{I}}_A^{opt}$  и  $\tilde{\mathcal{E}}_1^{opt}$  по (3.19), (3.20)

**Значения параметров** Зафиксируем  $N = 47$ ,  $L = 24$ . Получим оценки на  $R = 10^4$  реализациях. В качестве  $\alpha$  возьмём следующие значения: 0, 0.01, 0.02, это соответствует увеличению значений амплитуды на протяжении ряда (от  $n = 0$  до  $n = N - 1$ ) в 1, 1.6, 2.5 раз соответственно. Такое возрастание амплитуды встречается в исследованиях реальных рядов.

Как уже говорилось, значение  $\sigma$  следует выбирать так, чтобы предложенная имитация ВИ имела смысл. Исследование нескольких смоделированных рядов  $F^{(r)}$  показало, что при  $\sigma > 1$  уровень шума можно считать слишком большим. Это характеризуется в частности тем, что основной вклад в вид ряда не сосредоточен в первых двух собственных тройках и восстановленный по ним ряд слишком далёк от исходной э-м гармонике. Поэтому рассмотрим  $\sigma \leq 1$ , взяв следующие значения: 0.2, 0.4, 0.6, 0.8, 1.

### Проверка модифицированного метода Фурье

Проверим, что используемый в методе Фурье критерий (3.17) лучше предложенного сначала критерия (3.6). Рассмотрим кроме метода Фурье (МФ) ещё и метод, основанный на критерии (3.6), назовём его *упрощённым методом Фурье* (УМФ). Напомним, что переход от критерия (3.6) к критерию (3.17) был предложен для того, чтобы улучшить качество в случае, когда  $L\omega \notin \mathbb{N}$ .

Проверим, что

- МФ значительно лучше, чем УМФ при  $L\omega \notin \mathbb{N}$ ,
- МФ не хуже (или ненамного хуже), чем УМФ при  $L\omega \in \mathbb{N}$ .

Оценим характеристики  $\mathcal{D}_{\mathcal{A}\nu}(\rho_0^{opt})$ ,  $\mathcal{D}_{\mathcal{A}}(\rho_0^{opt})$  процедуры PER, использующей для идентификации метод Фурье, обозначим оценки как  $\overline{\mathcal{D}}_{\mathcal{A}\nu}^{opt}(\text{МФ})$  и  $\overline{\mathcal{D}}_{\mathcal{A}}^{opt}(\text{МФ})$ .

Оценим те же характеристики процедуры PER, но в которой на шаге 2 алгоритма 3.3 используется для идентификации критерий (3.6) вместо (3.17). Обозначим их как  $\overline{\mathcal{D}}_{\mathcal{A}\nu}^{opt}(\text{УМФ})$  и  $\overline{\mathcal{D}}_{\mathcal{A}}^{opt}(\text{УМФ})$ .

Рассмотрим два случая:  $L\omega \notin \mathbb{N}$  и  $L\omega \in \mathbb{N}$ . Для демонстрации первого случая возьмём  $\omega = 1/7$ , для второго —  $\omega = 1/12$ .

Рассчитаем  $\overline{\mathcal{D}}_{\mathcal{A}\nu}^{opt}(\text{МФ})$  и  $\overline{\mathcal{D}}_{\mathcal{A}}^{opt}(\text{МФ})$ , а также  $\Delta\overline{\mathcal{D}}_{\mathcal{A}\nu}^{opt} = \overline{\mathcal{D}}_{\mathcal{A}\nu}^{opt}(\text{УМФ}) - \overline{\mathcal{D}}_{\mathcal{A}\nu}^{opt}(\text{МФ})$ ,  $\Delta\overline{\mathcal{D}}_{\mathcal{A}}^{opt} = \overline{\mathcal{D}}_{\mathcal{A}}^{opt}(\text{УМФ}) - \overline{\mathcal{D}}_{\mathcal{A}}^{opt}(\text{МФ})$ .

Результаты вычисления  $\overline{\mathcal{D}}_{AV}^{opt}$  и  $\overline{\mathcal{D}}_A^{opt}$  для УМФ и  $\Delta\overline{\mathcal{D}}_{AV}^{opt}$  и  $\Delta\overline{\mathcal{D}}_A^{opt}$  приведены в табл. 3.3. Видно, что при  $L\omega \notin \mathbb{N}$  метод Фурье лучше, чем упрощённый метод Фурье ( $\Delta\overline{\mathcal{D}}_{AV}^{opt} > 0$ ,  $\Delta\overline{\mathcal{D}}_A^{opt} > 0$ ), а при  $L\omega \in \mathbb{N}$  метод Фурье ненамного хуже упрощённого метода Фурье.

Таким образом, для идентификации собственных троек  $\varepsilon$ -м гармоникой с неизвестной частотой  $\omega$  метод Фурье лучше, чем упрощённый метод Фурье.

Таблица 3.3: Средние ошибки для упрощённого метода Фурье (УМФ) и их сравнение со средними ошибками метода Фурье (МФ) при разных  $\alpha$ ,  $\sigma$  для случаев  $L\omega \notin \mathbb{N}$ ,  $L\omega \in \mathbb{N}$

| $\alpha = 0, L\omega \notin \mathbb{N}$ |                                     |                                  |   |  | $\alpha = 0, L\omega \in \mathbb{N}$ |                                     |                                  |   |  |
|---|-------------------------------------|----------------------------------|---|--|--------------------------------------|-------------------------------------|----------------------------------|---|--|
|   | УМФ                                 |                                  | УМФ-МФ                                    |  |                                      | УМФ                                 |                                  | УМФ-МФ                                    |  |
| $\sigma$                                | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\Delta\overline{\mathcal{D}}_{AV}^{opt}$ | $\Delta\overline{\mathcal{D}}_A^{opt}$ | $\sigma$                             | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\Delta\overline{\mathcal{D}}_{AV}^{opt}$ | $\Delta\overline{\mathcal{D}}_A^{opt}$ |
| 0.2                                     | 0.01                                | 0.02                             | 0.01                                      | 0.01                                   | 0.2                                  | 0.000                               | 0.005                            | 0.000                                     | 0.000                                  |
| 0.4                                     | 0.06                                | 0.08                             | 0.03                                      | 0.03                                   | 0.4                                  | 0.002                               | 0.02                             | 0.000                                     | 0.000                                  |
| 0.6                                     | 0.14                                | 0.19                             | 0.06                                      | 0.06                                   | 0.6                                  | 0.01                                | 0.05                             | -0.002                                    | -0.003                                 |
| 0.8                                     | 0.25                                | 0.34                             | 0.09                                      | 0.09                                   | 0.8                                  | 0.04                                | 0.12                             | -0.003                                    | -0.005                                 |
| 1                                       | 0.39                                | 0.52                             | 0.12                                      | 0.12                                   | 1                                    | 0.12                                | 0.24                             | 0.006                                     | 0.008                                  |

| $\alpha = 0.01, L\omega \notin \mathbb{N}$ |                                     |                                  |   |  | $\alpha = 0.01, L\omega \in \mathbb{N}$ |                                     |                                  |   |  |
|--|-------------------------------------|----------------------------------|---|--|---|-------------------------------------|----------------------------------|---|--|
|  | УМФ                                 |                                  | УМФ-МФ                                    |  |   | УМФ                                 |                                  | УМФ-МФ                                    |  |
| $\sigma$                                   | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\Delta\overline{\mathcal{D}}_{AV}^{opt}$ | $\Delta\overline{\mathcal{D}}_A^{opt}$ | $\sigma$                                | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\Delta\overline{\mathcal{D}}_{AV}^{opt}$ | $\Delta\overline{\mathcal{D}}_A^{opt}$ |
| 0.2  | 0.01                                | 0.02                             | 0.01                                      | 0.01                                   | 0.2                                     | 0.000                               | 0.005                            | 0.000                                     | 0.000                                  |
| 0.4  | 0.06                                | 0.08                             | 0.02                                      | 0.03                                   | 0.4                                     | 0.002                               | 0.02                             | 0.000                                     | 0.001                                  |
| 0.6  | 0.14                                | 0.18                             | 0.05                                      | 0.05                                   | 0.6                                     | 0.01                                | 0.06                             | -0.004                                    | -0.003                                 |
| 0.8  | 0.25                                | 0.34                             | 0.08                                      | 0.08                                   | 0.8                                     | 0.04                                | 0.13                             | -0.012                                    | -0.010                                 |
| 1  | 0.38                                | 0.52                             | 0.11                                      | 0.11                                   | 1                                       | 0.12                                | 0.25                             | 0.002                                     | 0.000                                  |

| $\alpha = 0.02, L\omega \notin \mathbb{N}$ |                                     |                                  |   |  | $\alpha = 0.02, L\omega \in \mathbb{N}$ |                                     |                                  |   |  |
|--|-------------------------------------|----------------------------------|---|--|---|-------------------------------------|----------------------------------|---|--|
|  | УМФ                                 |                                  | УМФ-МФ                                    |  |   | УМФ                                 |                                  | УМФ-МФ                                    |  |
| $\sigma$                                   | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\Delta\overline{\mathcal{D}}_{AV}^{opt}$ | $\Delta\overline{\mathcal{D}}_A^{opt}$ | $\sigma$                                | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\Delta\overline{\mathcal{D}}_{AV}^{opt}$ | $\Delta\overline{\mathcal{D}}_A^{opt}$ |
| 0.2  | 0.01                                | 0.02                             | 0.01                                      | 0.01                                   | 0.2                                     | 0.000                               | 0.006                            | 0.000                                     | 0.000                                  |
| 0.4  | 0.06                                | 0.08                             | 0.03                                      | 0.03                                   | 0.4                                     | 0.003                               | 0.03                             | -0.001                                    | 0.000                                  |
| 0.6  | 0.14                                | 0.19                             | 0.06                                      | 0.05                                   | 0.6                                     | 0.02                                | 0.07                             | -0.005                                    | -0.003                                 |
| 0.8  | 0.26                                | 0.35                             | 0.09                                      | 0.08                                   | 0.8                                     | 0.05                                | 0.14                             | -0.013                                    | -0.012                                 |
| 1  | 0.40                                | 0.53                             | 0.13                                      | 0.12                                   | 1                                       | 0.13                                | 0.26                             | -0.003                                    | -0.007                                 |



## Зависимость характеристик автоматической процедуры с $\rho_0^{opt}$ от параметров ряда $\alpha, \sigma$

Изучим зависимость от  $\sigma$  следующих характеристик процедуры PER:  $\overline{\mathcal{D}}_{AV}^{opt}$ ,  $\overline{\mathcal{D}}_A^{opt}$ ,  $\overline{\mathcal{D}}_V^{opt}$ ,  $\overline{\#I}_A^{opt}$ ,  $\tilde{\mathcal{E}}_1^{opt}$ , полученных при различных  $\alpha$  и  $\omega$ .

Чтобы выяснить, близка ли АИ к ВИ, сравним ошибку  $\mathcal{D}_{AV}$  с  $\mathcal{D}_V$ . Покажем, что при “хорошем”  $\omega$ , когда  $L\omega \in \mathbb{N}$ , АИ близка к ВИ по сравнению с ошибкой, которую даёт АИ или ВИ процедура. Покажем также, что хотя при “плохом”  $\omega$  ( $L\omega \notin \mathbb{N}$ ) ошибка АИ становится больше, качество автоматической процедуры по-прежнему остаётся хорошим.

**Случай хорошего  $\omega$**  Рассмотрим случай, когда  $L\omega \in \mathbb{N}$ , возьмём  $\omega = 1/12$ . Результаты расчётов при  $\omega = 1/12$  приведены в табл. 3.4. На рис. 3.7 изображена зависимость  $\sqrt{\overline{\mathcal{D}}_{AV}^{opt}}$ ,  $\sqrt{\overline{\mathcal{D}}_A^{opt}}$  и  $\sqrt{\overline{\mathcal{D}}_V^{opt}}$  от  $\sigma$  при  $\alpha = 0$ . Изображены квадратные корни (отклонения), а не сами ошибки, чтобы продемонстрировать близкий к линейному характер зависимости отклонений от  $\sigma$  (стандарта шума).

При других  $\alpha$  характер зависимости сохраняется, поэтому график приведён только для  $\alpha = 0$ .

Видно, что  $\mathcal{D}_{AV} < \mathcal{D}_V$ , а при малом шуме  $\mathcal{D}_{AV} \ll \mathcal{D}_V$ , что означает, что по сравнению с ошибкой визуальной процедуры АИ незначительно отличается от ВИ. Это подтверждают значения вероятности ошибки первого рода, которые при уменьшении  $\sigma$  уменьшаются до 0 и при  $\sigma < 1$  не превышают 0.072.

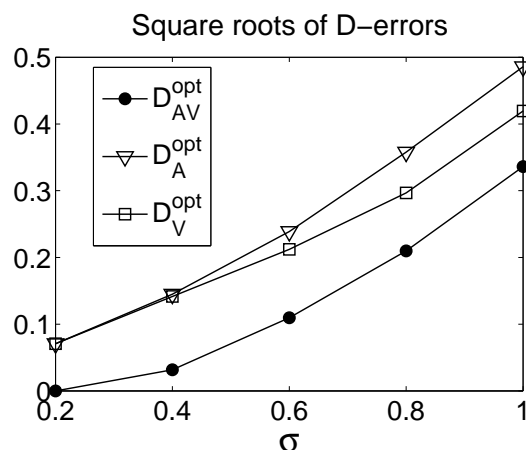


Рис. 3.7: Зависимость квадратных корней  $\overline{\mathcal{D}}_{AV}^{opt}$ ,  $\overline{\mathcal{D}}_A^{opt}$  и  $\overline{\mathcal{D}}_V^{opt}$  от  $\sigma$  при  $\alpha = 0$ .

Таблица 3.4: Зависимость характеристик процедуры PER от  $\sigma$  при разных  $\alpha$ ,  $L\omega$  — целое

| $\alpha = 0$ |                |                                     |                                  |                            |                                    |                               |
|--------------|----------------|-------------------------------------|----------------------------------|----------------------------|------------------------------------|-------------------------------|
| $\sigma$     | $\rho_0^{opt}$ | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\overline{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2          | 0.99           | 0.000                               | 0.005                            | 0.005                      | 2.00                               | 0.000                         |
| 0.4          | 0.97           | 0.001                               | 0.021                            | 0.020                      | 2.08                               | 0.000                         |
| 0.6          | 0.96           | 0.012                               | 0.057                            | 0.045                      | 2.11                               | 0.016                         |
| 0.8          | 0.92           | 0.044                               | 0.128                            | 0.088                      | 2.37                               | 0.044                         |
| 1.0          | 0.91           | 0.113                               | 0.236                            | 0.176                      | 2.15                               | 0.174                         |

| $\alpha = 0.01$ |                |                                     |                                  |                            |                                    |                               |
|-----------------|----------------|-------------------------------------|----------------------------------|----------------------------|------------------------------------|-------------------------------|
| $\sigma$        | $\rho_0^{opt}$ | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\overline{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2             | 0.98           | 0.000                               | 0.005                            | 0.005                      | 2.02                               | 0.000                         |
| 0.4             | 0.97           | 0.002                               | 0.021                            | 0.020                      | 2.06                               | 0.001                         |
| 0.6             | 0.93           | 0.018                               | 0.063                            | 0.046                      | 2.38                               | 0.006                         |
| 0.8             | 0.92           | 0.056                               | 0.138                            | 0.092                      | 2.33                               | 0.072                         |
| 1.0             | 0.88           | 0.114                               | 0.246                            | 0.176                      | 2.61                               | 0.125                         |

| $\alpha = 0.02$ |                |                                     |                                  |                            |                                    |                               |
|-----------------|----------------|-------------------------------------|----------------------------------|----------------------------|------------------------------------|-------------------------------|
| $\sigma$        | $\rho_0^{opt}$ | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\overline{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2             | 0.96           | 0.001                               | 0.006                            | 0.005                      | 2.11                               | 0.000                         |
| 0.4             | 0.94           | 0.005                               | 0.025                            | 0.020                      | 2.26                               | 0.000                         |
| 0.6             | 0.92           | 0.021                               | 0.068                            | 0.050                      | 2.34                               | 0.020                         |
| 0.8             | 0.88           | 0.062                               | 0.153                            | 0.100                      | 2.67                               | 0.060                         |
| 1.0             | 0.88           | 0.131                               | 0.268                            | 0.194                      | 2.26                               | 0.210                         |

**Случай плохого  $\omega$**  Приведём в табл. 3.5 результаты для  $\omega = 1/7$ , т. е.  $\omega : L\omega \notin \mathbb{N}$ . Видно, что хотя ошибка  $\mathcal{D}_V$  не отличается от полученной при хорошем  $\omega$ , ошибки  $\mathcal{D}_A$  и  $\mathcal{D}_{AV}$  стали существенно больше. Это объясняется тем, что  $\rho_0^{opt}$  при плохом  $\omega$  становится меньше и идентифицируются лишние компоненты шума, что подтверждается значительным увеличением  $\overline{\#\mathcal{I}}_A^{opt}$  и незначительным увеличением  $\tilde{\mathcal{E}}_1^{opt}$ .

Таблица 3.5: Зависимость характеристик процедуры PER от  $\sigma$  при разных  $\alpha$ ,  $L\omega$  — не целое

| $\alpha = 0$ |                |                                     |                                  |                            |                                    |                               |
|--------------|----------------|-------------------------------------|----------------------------------|----------------------------|------------------------------------|-------------------------------|
| $\sigma$     | $\rho_0^{opt}$ | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\overline{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2          | 0.79           | 0.006                               | 0.011                            | 0.005                      | 4.26                               | 0.000                         |
| 0.4          | 0.75           | 0.031                               | 0.051                            | 0.019                      | 5.15                               | 0.001                         |
| 0.6          | 0.74           | 0.084                               | 0.130                            | 0.045                      | 5.33                               | 0.029                         |
| 0.8          | 0.72           | 0.160                               | 0.246                            | 0.087                      | 5.70                               | 0.065                         |
| 1.0          | 0.72           | 0.271                               | 0.403                            | 0.169                      | 5.31                               | 0.192                         |

| $\alpha = 0.01$ |                |                                     |                                  |                            |                                    |                               |
|-----------------|----------------|-------------------------------------|----------------------------------|----------------------------|------------------------------------|-------------------------------|
| $\sigma$        | $\rho_0^{opt}$ | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\overline{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2             | 0.79           | 0.006                               | 0.011                            | 0.005                      | 4.21                               | 0.000                         |
| 0.4             | 0.75           | 0.032                               | 0.053                            | 0.019                      | 5.04                               | 0.005                         |
| 0.6             | 0.73           | 0.085                               | 0.135                            | 0.045                      | 5.42                               | 0.032                         |
| 0.8             | 0.72           | 0.169                               | 0.259                            | 0.089                      | 5.51                               | 0.101                         |
| 1.0             | 0.71           | 0.267                               | 0.410                            | 0.173                      | 5.48                               | 0.188                         |

| $\alpha = 0.02$ |                |                                     |                                  |                            |                                    |                               |
|-----------------|----------------|-------------------------------------|----------------------------------|----------------------------|------------------------------------|-------------------------------|
| $\sigma$        | $\rho_0^{opt}$ | $\overline{\mathcal{D}}_{AV}^{opt}$ | $\overline{\mathcal{D}}_A^{opt}$ | $\overline{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2             | 0.76           | 0.006                               | 0.012                            | 0.005                      | 4.75                               | 0.000                         |
| 0.4             | 0.74           | 0.033                               | 0.056                            | 0.021                      | 5.14                               | 0.012                         |
| 0.6             | 0.71           | 0.084                               | 0.136                            | 0.050                      | 5.81                               | 0.031                         |
| 0.8             | 0.70           | 0.169                               | 0.261                            | 0.101                      | 5.93                               | 0.103                         |
| 1.0             | 0.71           | 0.273                               | 0.412                            | 0.187                      | 5.28                               | 0.230                         |

**Замечание 3.3** В данных расчётах использовалось  $s_0 = 1$ . Результаты для других  $s_0$  не приведены, заметим только, что взятие большего  $s_0$  в данном случае не уменьшает ошибки  $\mathcal{D}_{AV}$ ,  $\mathcal{D}_A$ .

**Выводы** Таким образом можно сделать вывод, что для модели зашумлённой  $\varepsilon$ -м гармонике при хорошем  $\omega$  ( $L\omega \in \mathbb{N}$ ) предложенная АИ очень близка к ВИ (по сравнению с ошибкой, даваемой самой ВИ).

При плохом  $\omega$  ( $L\omega \notin \mathbb{N}$ ), несмотря на то, что вероятность ошибки первого рода почти не увеличивается, качество АИ ухудшается за счёт того, что значение  $\rho_0^{opt}$  значительно меньше и процедура АИ ошибочно идентифицирует

больше шумовых компонент. Это приводит к тому, что АИ даёт ошибку  $\mathcal{D}_A$  до двух раз большую ошибки  $\mathcal{D}_V$ . Однако эта ошибка ( $\mathcal{D}_A$ ) по прежнему невелика (не превышает 0.261 при  $\sigma < 1$ ) и можно сделать вывод о том, что предложенная процедура АИ выполняет поставленную задачу.

### Поведение автоматической процедуры при изменении $\rho_0$

Рассмотрим зависимость результата автоматической процедуры от порогового значения  $\rho_0$  ( $0 \leq \rho_0 \leq 1$ ). Это поможет сформулировать критерий выбора  $\rho_0$ .

Очевидным образом, чем меньше взятое пороговое значение  $\rho_0$ , тем больше идентифицируется компонент, в то время как при выборе слишком большого  $\rho_0$  необходимые компоненты могут быть не идентифицированы (см. график изменения вероятности ошибки первого рода, рис. 3.8).

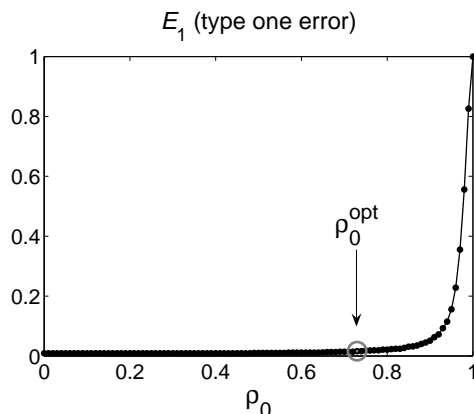


Рис. 3.8: Зависимость от  $\rho_0$  ошибки  $\tilde{\mathcal{E}}_1$  при  $\alpha = 0.01$ ,  $\sigma = 0.8$  для случая  $L\omega \in \mathbb{N}$  ( $\rho_0^{opt} = 0.92$ ,  $\tilde{\mathcal{E}}_1^{opt} = 0.072$ ).

Это определяет поведение ошибок  $\mathcal{D}_{AV}$ ,  $\mathcal{D}_A$ , графики оценок которых при  $\omega = 1/12, 1/7$  изображены на рис. 3.9.

## 3.4 Подходы к выбору порогового значения $\rho_0$

В данном разделе будет предложено два подхода к выбору порогового значения  $\rho_0$  процедуры PER.

Первый основан на аналитическом вычислении максимума периодограммы  $\varepsilon$ -м гармоники с известным  $\alpha$  и с частотой  $\omega$ :  $L\omega \in \mathbb{N}$ .

Второй основан на предположении, что при идентификации с помощью метода Фурье идентифицированные собственные тройки  $\varepsilon$ -м гармоники имеют

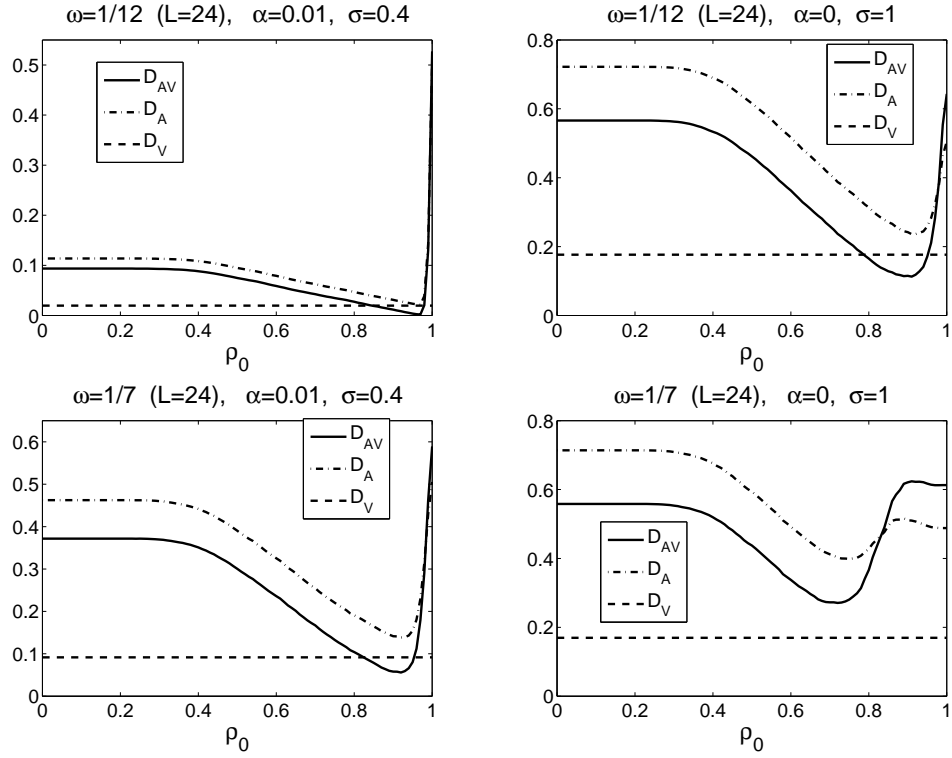


Рис. 3.9: Зависимость характеристик  $\overline{D}_{AV}$ ,  $\overline{D}_A$  и  $\overline{D}_V$  от  $\rho_0$  при разных  $\alpha$ ,  $\sigma$  для случаев  $L\omega \in \mathbb{N}$  и  $L\omega \notin \mathbb{N}$ .

большой вклад в вид ряда, чем ошибочно идентифицированные тройки остатка ряда. Это позволяет построить процедуру выделения  $\varepsilon$ -м гармоники, параметром которой является ограничение снизу на амплитуду искомой  $\varepsilon$ -м гармоники.

### 3.4.1 Аналитическое вычисление $\rho_0$ для известного $\alpha$ и при $\omega$ : $L\omega \in \mathbb{N}$

Опишем принцип выбора порогового значения  $\rho_0$  процедуры PER при решении задачи выделения  $\varepsilon$ -м гармоники с экспоненциальным показателем  $\alpha$  и частотой  $\omega$ . Пусть есть ряд  $F$ , содержащий  $F^{(P)}$ :

$$f_n^{(P)} = Ae^{\alpha n} \cos(2\pi\omega n + \phi).$$

Предполагаем, что  $F^{(P)}$  достаточно хорошо отделим от остатка  $F - F^{(P)}$ .

**Определение 3.2** Определим  $\rho_0^{(P)}$  — пороговое значение метода Фурье, такое что если взять ряд  $F^{(P)}$ , провести его разложение на собственные тройки количеством  $d$ , то метод Фурье идентифицирует их все

$$\rho_0^{(P)} = \rho_0^{(P)}(F^{(P)}) = \max\{\rho_0 : \mathcal{I}^{(P)}(\rho_0) = \{1, \dots, d\}\},$$

где  $\mathcal{I}^{(P)}(\rho_0)$  — номера идентифицированных с  $\rho_0$  компонент разложения  $F^{(P)}$ .

В случае, когда  $F^{(P)}$  —  $\varepsilon$ -м гармоника с частотой  $\omega$ ,  $d = 2$  для  $0 < \omega < 0.5$  и  $d = 1$  для  $\omega = 0.5$ .

В отличие от случая выделения тренда, где вычисление  $\mathcal{C}_0^{(T)}$  было невозможно, при выделении  $\varepsilon$ -м гармоники с параметрами  $\alpha$ ,  $\omega$  можно оценить  $\rho_0^{(P)}$  по известным параметрам  $\alpha$ ,  $L$  при условии, что  $L\omega \in \mathbb{N}$ .

**Расчёт  $\rho_0^{(P)}$**  В случае, когда  $L, \omega$ :  $L\omega \in \mathbb{N}$ , оценим  $\rho_0^{(P)}$  для  $\varepsilon$ -м гармоники с параметрами  $\alpha$ ,  $L$  с помощью (3.14),  $\rho_0^{(P)} \approx \widehat{\rho}_0^{(P)}(\alpha, L)$ , где

$$\widehat{\rho}_0^{(P)}(\alpha, L) = \frac{2(e^\gamma - 1)}{\gamma(e^\gamma + 1)}, \quad \gamma = L\alpha. \quad (3.21)$$

Проверим, насколько достигнутая оценка  $\widehat{\rho}_0^{(P)}(\alpha, L)$  близка к настоящему значению  $\rho_0^{(P)}$ . Приведём в табл. 3.6 значения  $\widehat{\rho}_0^{(P)}(\alpha, L)$  и результаты численного вычисления  $\rho_0^{(P)}$  для ряда  $F$  с разными  $\alpha$  и  $N = 47$ ,  $L = 24$  при  $\omega = 1/12$  ( $L\omega \in \mathbb{N}$ ). Поиск  $\rho_0^{(P)}$  проводился с шагом  $10^{-3}$ .

Таблица 3.6: Численно полученные значения  $\rho_0^{(P)}$  и их оценки  $\widehat{\rho}_0^{(P)}(\alpha, L)$  для разных  $\alpha$  ( $L = 24$ )

| $\alpha$                            | 0.01  | 0.02  | 0.03  | 0.04  | 0.05  |
|-------------------------------------|-------|-------|-------|-------|-------|
| $\rho_0^{(P)}$                      | 0.996 | 0.988 | 0.973 | 0.955 | 0.932 |
| $\widehat{\rho}_0^{(P)}(\alpha, L)$ | 0.995 | 0.981 | 0.959 | 0.930 | 0.900 |

Видно, что с помощью формулы (3.21) для малых  $\gamma = L\alpha$  достигается хорошее качество приближения  $\rho_0^{(P)}$  при выделении  $\varepsilon$ -м гармоники с параметрами  $\alpha$ ,  $\omega$ , где  $L$  — используемая длина окна.

Рассмотрим, возможно ли применение оценки (3.21) для расчёта  $\rho_0$ .

При выделении периодического ряда  $F^{(P)}$  в присутствии некоторого остатка  $F^{(R)}$ , надо брать  $\rho_0$  меньше, чем  $\rho_0^{(P)}$ . Это связано с тем, что в реальности имеет место приближённая разделимость  $F^{(P)}$  и  $F^{(R)}$ . Из-за этого собственные вектора, соответствующие  $\varepsilon$ -м гармонике, искажаются, что влечёт уменьшение значений тестовой функции, используемой в критерии (3.17) второй части метода Фурье.

Поэтому при реальном исследовании нужно брать  $\rho_0 < \rho_0^{(P)}$  и, как и для тренда, встаёт вопрос, насколько меньше  $\rho_0^{(P)}$  нужно брать пороговое значение  $\rho_0$ .

Ещё одна проблема состоит в том, что данный способ выбора  $\rho_0$  требует знания  $\alpha$ . В случае, когда ряд  $F$  представляет собой зашумлённую  $\varepsilon$ -м гармонику, значение  $\alpha$  можно оценить по поведению элементов ряда. Если же

ряд  $F^{(P)}$  является суммой э-м гармоник с разными  $\alpha_i$ , необходимо выбирать  $\rho_0 = \rho_0^{(P)}(\alpha_{max}, L)$ , где  $\alpha_{max} = \max_i |\alpha_i|$ . Однако оценивание  $\alpha_{max}$  по ряду  $F$ , содержащему такую составляющую  $F^{(P)}$ , является трудной задачей и во многих случаях невозможно.

Всё это, а также необходимость знания частоты  $\omega$  с тем, чтобы выбирать  $L$ :  $L\omega \in \mathbb{N}$ , делает этот подход выбора  $\rho_0$  применимым только в отдельных случаях.

### 3.5 Эмпирический подход к выбору $\rho_0$

Представим другой, *эмпирический*, подход к выбору  $\rho_0$ . Опишем его сначала в применении к ряду следующей модели:

$$F = F^{(P)} + F^{(\varepsilon)}, \quad f_n^{(P)} = Ae^{\alpha n} \cos(2\pi n\omega + \phi), \quad (3.22)$$

где  $A, \alpha \in \mathbb{R}$ ,  $A > 0$ , а  $\phi \in [0, 2\pi)$ , а  $F^{(\varepsilon)}$  — ряд, являющийся реализацией случайного белого шума. Получим собственные тройки, соответствующие ряду, с помощью процедуры SSADEC с некоторым  $L$  (алгоритм 1.1). Обозначим множество пар номеров собственных троек, соответствующих  $F^{(P)}$  — за  $\mathcal{I}^{(P)}$ , а номера собственных троек, соответствующих  $F^{(\varepsilon)}$  — за  $\mathcal{I}^{(\varepsilon)}$ .

Эмпирический подход к выбору  $\rho_0$ , который описывается дальше в этом разделе, основан на предположении, что

$$\forall \{i, j\} \in \mathcal{I}^{(P)}, \forall k \in \mathcal{I}^{(\varepsilon)} \text{ выполняется } \|F_{\{i,j\}}\|^2 \gg \|F_k\|^2, \quad (3.23)$$

где  $F_{\{i,j\}}$  — это ряд, восстановленный по собственным тройкам с номерами  $i, j$ , а  $F_k$  — ряд, восстановленный по  $k$ -ой тройке. Это предположение основано на том факте, что вклад каждой из тройки из  $\mathcal{I}^{(\varepsilon)}$  в вид ряда  $F$  (измеряемый сингулярным числом) достаточно равномерно распределён между всеми тройками из  $\mathcal{I}^{(\varepsilon)}$ .

**Идея эмпирического подхода к выбору  $\rho_0$**  *Идея выбора  $\rho_0$  заключается в том, чтобы найти все значения  $\rho_0$ , при которых идентифицируются пары собственных троек из  $\mathcal{I}^{(P)}$ , и взять в качестве  $\rho_0$  минимальное из них.*

Найдём сначала такие значения  $\rho_0$ , что при  $\rho_0$  идентифицируется некоторая собственная тройка или пара собственных троек. Восстановим по ней ряд и посчитаем его норму. Любой ряд, восстановленный по некоторой тройке (или паре собственных троек) из  $\mathcal{I}^{(P)}$ , будут иметь гораздо большую норму, чем ряд, восстановленный по любой из троек из  $\mathcal{I}^{(\varepsilon)}$ .

**Схема подхода** Переберём значения  $\rho_0$  из заданного интервала  $[\rho_0^l, \rho_0^h]$  с заданным шагом  $\Delta\rho$ . Обозначим как  $\mathcal{I}(\rho_0)$  группу троек, идентифицированных с пороговым значением  $\rho_0$ .

Фиксируем некоторое  $\mathcal{G}_0 > 0$ . Найдём такое минимальное  $\rho_0$ , при котором норма ряда, восстановленного по тройкам с номерами из  $\mathcal{I}(\rho_0) \setminus \mathcal{I}(\rho_0 + \Delta\rho)$ , больше  $\mathcal{G}_0$ . Обозначим результат за  $\rho_E$ :

$$\rho_E = \min \left\{ \rho_0 : \|F(\mathcal{J}(\rho_0))\|^2 \geq \mathcal{G}_0 \right\},$$

где  $\mathcal{J}(\rho_0) = \mathcal{I}(\rho_0) \setminus \mathcal{I}(\rho_0 + \Delta\rho)$ , а  $F(\mathcal{J}(\rho_0))$  — ряд, восстановленный по  $\mathcal{J}(\rho_0)$ .

С учётом сделанного предположения (3.23), при правильно выбранном  $\mathcal{G}_0$  это будет минимальное  $\rho_0$  при котором идентифицируются собственные тройки из  $\mathcal{I}^{(P)}$ .

Шаг  $\Delta\rho$  надо выбирать как можно меньше, так как он используется для поиска значений  $\rho_0$ , при которых идентифицируются отдельные собственные тройки. Опыт показывает, что в большинстве случаев достаточно выбирать  $\Delta\rho = 0.01$ .

Основной проблемой применения этого подхода является выбор  $\mathcal{G}_0$ .

**Выбор  $\mathcal{G}_0$  для гармоник с известной амплитудой** Пусть ряд  $F$ :  $F = F^{(P)} + F^{(\varepsilon)}$ ,

$$f_n = f_n^{(P)} + f_n^{(\varepsilon)}, \quad f_n^{(P)} = A \sin(2\pi\omega n + \phi), \quad (3.24)$$

и известны амплитуда  $A$  и частота  $\omega$ :  $N\omega \in \mathbb{N}$ .

Несложно посчитать, что для  $F^{(P)}$  выполняется

$$\|F^{(P)}\|^2 = A^2 \frac{1}{N} \sum_{n=0}^{N-1} \cos^2(2\pi\omega n + \phi_j) = \frac{A^2}{2}.$$

Таким образом, для выделения гармоник с амплитудой  $A$  и такой  $\omega$ , что  $N\omega \in \mathbb{N}$ , необходимо взять  $\mathcal{G}_0$ :  $\mathcal{G}_0 \leq A^2/2$ . Выберем  $\mathcal{G}_0$ , равное части от  $A^2/2$ , выражаемой в процентах, обозначим такое значение как  $\mathcal{G}_0(A, \mathcal{P}\%)$ :

$$\mathcal{G}_0(A, \mathcal{P}\%) = \mathcal{P}\% \cdot A^2/2, \quad (3.25)$$

что, например, при  $A = 1$ ,  $\mathcal{P}\% = 75\%$  даёт  $\mathcal{G}_0 = 0.375$ .



**Выбор  $\mathcal{G}_0$  для э-м гармоник с известной модуляцией** Пусть ряд  $F$  представляет собой зашумлённую э-м гармоническую гармонику,  $F = F^{(P)} + F^{(\varepsilon)}$ :

$$f_n = f_n^{(P)} + f_n^{(\varepsilon)}, \quad f_n^{(P)} = Ae^{\alpha n} \sin(2\pi\omega n + \phi), \quad T = 1/\omega, \quad (3.26)$$

и известны  $A$ ,  $\alpha$  и  $\omega$ , причём  $T = 1/\omega$ :  $T \in \mathbb{N}$ ,  $N/T \in \mathbb{N}$ . Тогда можно оценить  $\mathcal{G}_0$  следующим образом.

В случае, когда  $\alpha > 0$ , оценим снизу значения  $F^{(P)}$  на каждом из интервалов вида  $[kT, k(T+1))$ , где  $k \in \mathbb{Z}_+$ , выражением  $M_k = e^{\alpha T(k-1)}$ , тогда

$$\begin{aligned} \|F^{(P)}\|^2 &= \frac{1}{N} \sum_{n=0}^{N-1} f_n^2 \geq \frac{A^2}{N} \sum_{k=1}^{N/T} M_k^2 \sum_{n=kT}^{k(T+1)-1} \sin^2(2\pi n/T + \phi) = \\ &= \frac{A^2 T}{N} \sum_{k=1}^{N/T} M_k^2 = \frac{A^2 T (e^{2\alpha N} - 1)}{2N (e^{2\alpha T} - 1)}. \end{aligned}$$

Это будет верно и при  $\alpha < 0$ , только вместо  $A$  надо брать  $Ae^{\alpha(N-1)}$ .

Как и для немодулированной гармоники (см. (3.25)), будем брать  $\mathcal{G}_0$ , равное процентной части от полученной оценки. Введём обозначение  $\mathcal{G}_0(A, \mathcal{P}\%, \alpha, T)$ :

$$\mathcal{G}_0(A, \mathcal{P}\%, \alpha, T) = \mathcal{P}\% \cdot \frac{B^2(A, \alpha, N)T(e^{2\alpha N} - 1)}{2N(e^{2\alpha T} - 1)}, \quad (3.27)$$

$$\text{где } B(A, \alpha, N) = \min\{A, Ae^{\alpha(N-1)}\},$$

и будем выбирать  $\mathcal{G}_0 = \mathcal{G}_0(A, \mathcal{P}\%, \alpha, T)$ . Например, для  $\alpha = 0.02$ ,  $N = 48$ ,  $T = 12$  и  $\mathcal{P}\% = 75\%$  выполняется  $\mathcal{G}_0(A, \mathcal{P}\%, \alpha, T) \approx 0.557$ .

Таким образом, было показано, как выбрать  $\mathcal{G}_0$  для выделения одной гармоники и как улучшить выбор  $\mathcal{G}_0$  для э-м гармоник с известными  $A$ ,  $\alpha$ ,  $\omega$ .

**Выбор  $\mathcal{G}_0$  для периодической составляющей ряда с ограничениями на параметры** Рассмотрим теперь случай, когда в ряде присутствует периодическая составляющая  $F^{(P)}$  с периодом  $T$ :  $T \in \mathbb{N}$ ,  $N/T \in \mathbb{N}$ , определённая согласно (0.1):

$$f_n^{(P)} = \sum_{k=1}^{\lfloor T/2 \rfloor} A_k e^{\alpha_k n} \cos(2\pi nk/T + \phi_k),$$

где  $A_k, \alpha_k \in \mathbb{R}$ ,  $A_k > 0$ , а  $\phi_k \in [0, 2\pi)$ . Пусть известно либо  $\min_k A_k$ , либо  $\{A_k, \alpha_k\}_{k=1}^{\lfloor T/2 \rfloor}$  и  $T$ .

Пользуясь (3.27), запишем выражение для  $\mathcal{G}_0$ , которое необходимо выбрать, чтобы были идентифицированы все компоненты периодической составляющей:

$$\mathcal{G}_0 = \begin{cases} \min_k \left\{ \mathcal{G}_0(A_k, \mathcal{P}\%, \alpha_k, T) \right\} & \text{в случае известных } A_k, \alpha_k, T, \\ \mathcal{G}_0(A_{min}, \mathcal{P}\%), \quad A_{min} = \min_k A_k, & \text{в противном случае.} \end{cases} \quad (3.28)$$

**Выбор  $\mathcal{P}\%$**  Выбор  $\mathcal{P}\%$  зависит от многих факторов, таких как длина ряда, уровень шума, а также степень искажённости собственных векторов, соответствующих периодической составляющей. Однако опыт применения эмпирического подхода к выбору  $\rho_0$  и проведённые статистические исследования показывают, что выбор  $\mathcal{P}\% \approx 50\%$  позволяет найти  $\rho_0$  в большинстве случаев.

Построим на основе изложенного в данном разделе эмпирического подхода к выбору  $\rho_0$  процедуру PEREMP, см. алгоритм 3.5.

### 3.5.1 Результаты численного исследования

**Результаты вычисления  $\rho_E$  для э-м гармоника без шума** Рассмотрим ряд модели (3.26) при отсутствии шума ( $f_n^{(\varepsilon)} = 0$ ). Фиксируем параметры  $N = 47$ ,  $\omega = 1/12$ ,  $\phi = 0$ . Взяв  $L = 24$ , вычислим  $\rho_E$  для различных значений параметра  $\alpha$  с помощью процедуры PEREMP (см. алгоритм 3.5) и сравним результаты с  $\rho_0^{opt}$  при  $\sigma = 0$ , рассчитанным ранее, см. раздел 3.3. Результаты вычислений приведены в табл. 3.7; видно, что они совпадают.

Таблица 3.7:  $\rho_0^{opt}$  и  $\rho_E$  для э-м гармоника без шума при разных  $\alpha$

| $\alpha$       | 0.01  | 0.02  | 0.03  | 0.04  | 0.05  |
|----------------|-------|-------|-------|-------|-------|
| $\rho_0^{opt}$ | 0.996 | 0.988 | 0.973 | 0.955 | 0.932 |
| $\rho_E$       | 0.996 | 0.988 | 0.973 | 0.955 | 0.932 |

**АЛГОРИТМ 3.5. Процедура PEREMP: автоматическое выделение периодической составляющей с известными параметрами**

**Входные данные:**  $F$  — ряд, содержащий периодическую составляющую  $F^{(P)}$

**Результат:**  $\rho_E$ ,  $\mathcal{I}_A$  — группа идентифицированных компонент,  $F_A$  — полученная аппроксимация периодической составляющей

**Основные параметры:**

- Используемая информация о периоде:  $A_{min} = \min_k A_k$  или множество параметров  $\{A_k, \alpha_k\}_k$  и  $T$ , см. (3.28)
- Параметры процедуры PER:  $L, s_0 \in \mathbb{Z}_+$
- Параметры выбора  $\rho_0$ :  $\rho_0^l, \rho_0^r, \Delta\rho, \mathcal{P}\%$

**Процедура:**

1. Сосчитаем  $\mathcal{G}_0$  по (3.28), если известны  $\{A_k, \alpha_k\}_k$  и  $T$  или по (3.25), используя  $\min_k A_k$
2.  $\Omega = \{\rho_0^l + k\Delta\rho\}_{k=0}^M, \quad M = \left\lceil \frac{\rho_0^r - \rho_0^l}{\Delta\rho} \right\rceil$
3.  $\forall \rho_0 \in \Omega$  с помощью процедуры PER (см. алгоритм 3.3) с параметрами  $L, s_0, \rho_0$  получим  $\mathcal{I}_A(\rho_0) = \mathcal{I}_A$
4.  $\forall \rho_0 \in \Omega$  вычислим  $\mathcal{J}(\rho_0) = \mathcal{I}_A(\rho_0) \setminus \mathcal{I}_A(\rho_0 + \Delta\rho)$  и восстановим по ней составляющую  $F(\mathcal{J}(\rho_0))$  ряда  $F$  с помощью процедур SSADEC, SSAREC (по группе  $\mathcal{J}(\rho_0)$ ) (алгоритмы 1.1, 1.2)
5. Найдём  $\rho_E = \min_{\rho_0 \in \Omega} \{\rho_0 : \|F(\mathcal{J}(\rho_0))\|^2 \geq \mathcal{G}_0\}$
6.  $\mathcal{I}_A = \mathcal{I}_A(\rho_E), F_A = F(\mathcal{I}_A)$

**Результаты вычисления  $\rho_E$  для э-м гармоника с шумом** Рассмотрим  $F = F^{(P)} + F^{(\varepsilon)}$ ,  $F^{(\varepsilon)}$  — реализация белого нормального шума с нулевым средним и дисперсией  $\sigma^2$ . Получим с помощью статистического моделирования реализации  $\{F^{(r)}\}_{r=1}^R$  ряда  $F$  с заданными параметрами, к каждому ряду  $F^{(r)}$  применим процедуру PEREMP, получив  $\rho_E^{(r)}$ ,  $\mathcal{I}_A^{(r)}$ ,  $F_A^{(r)}$  и, как и в разделе 3.3.1, рассчитаем следующие характеристики процедуры PEREMP:

$$\begin{aligned}\bar{\rho}_E &= \frac{1}{R} \sum_{r=1}^R \rho_E^{(r)}, & \bar{\mathcal{D}}_{AV} &= \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F_A^{(r)}(\rho_E^{(r)}), F_V^{(r)}), \\ \bar{\mathcal{D}}_A &= \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F^{(P)}, F_A^{(r)}(\rho_E^{(r)})), & \bar{\mathcal{D}}_V &= \frac{1}{R} \sum_{r=1}^R \mathcal{D}(F^{(P)}, F_V^{(r)}), \\ \overline{\#\mathcal{I}}_A &= \frac{1}{R} \sum_{r=1}^R \mathcal{I}_A^{(r)}(\rho_E^{(r)}), & \tilde{\mathcal{E}}_1 &= \frac{1}{R} \#\{r : \mathcal{I}_A^{(r)} \not\subseteq \{1, 2\}\},\end{aligned}\tag{3.29}$$

где  $F_V^{(r)}$  — составляющая ряда  $F^{(r)}$ , восстановленная по группе компонент  $\mathcal{I}_V = \{1, 2\}$ . Используем для получения этих характеристик схему, описанную в алгоритме 3.6.

С помощью алгоритма 3.6 получим оценки характеристик процедуры PEREMP для следующих параметров:  $R = 10^3$ ,  $N = 47$ ,  $\phi = 0$ ,  $L = 24$ ,  $\Delta\rho = 0.01$ ,  $\rho_0^{\downarrow} = 0$ ,  $\rho_0^{\uparrow} = 1$ ,  $\mathcal{P}\% = 50\%$ .

Рассмотрим две модели э-м гармоника, с  $\omega = 1/12$  и  $\omega = 1/7$ , которые соответствуют случаям, когда  $L\omega \in \mathbb{N}$  и  $L\omega \notin \mathbb{N}$ . Результаты приведены в табл. 3.8, 3.9, 3.10 (для  $\omega = 1/12$ ), и в табл. 3.11, 3.12, 3.13 (для  $\omega = 1/7$ ), которые находятся в конце данной главы.

Для сравнения в этих таблицах приведены также средние характеристики для процедуры PER при наилучшем в среднем  $\rho_0^{opt}$  (полученные в разделе 3.3 с помощью алгоритма 3.4).

Результаты для  $\alpha = 0.02$ ,  $\omega = 1/12$  изображены на рис. 3.10, 3.11.

**АЛГОРИТМ 3.6. Оценка характеристик процедуры PEREMP для модели э-м гармоника плюс шум**

**Результат:**  $\bar{\rho}_E$ ,  $\bar{\mathcal{D}}_{AV}$ ,  $\bar{\mathcal{D}}_A$ ,  $\bar{\mathcal{D}}_V$ ,  $\overline{\#\mathcal{I}}_A$ ,  $\tilde{\mathcal{E}}_1$  — оценки характеристик процедуры PEREMP для заданной модели ряда

**Параметры:**

- $R$  — количество реализаций
- Параметры модели ряда:  $N$ ,  $A$ ,  $\alpha$ ,  $\phi$ ,  $\sigma$
- Параметры процедуры PEREMP:
  - $L$ ,  $s_0$ ,  $\rho_0^{\lceil}$ ,  $\rho_0^{\lfloor}$ ,  $\Delta\rho$ ,  $\mathcal{P}\%$
  - Используемая информация о э-м гармонике:  $A$  (расчёт  $\mathcal{G}_0$  как для гармоника) или  $A$ ,  $\alpha$ ,  $T$  (расчёт  $\mathcal{G}_0$  с поправкой на экспоненциальную модуляцию)

**Процедура:**

1. Смоделируем ряды  $\{F^{(r)}\}_{r=1}^R$  согласно модели (3.26)
2. Для каждого ряда  $F^{(r)}$  вычислим:
  - $\rho_E^{(r)}$ ,  $\mathcal{I}_A^{(r)}$  и  $F_A^{(r)}$ , пользуясь процедурой PEREMP (алгоритм 3.5)
  - $F_V^{(r)}$ , восстановив составляющую ряда  $F^{(r)}$  по группе компонент  $\mathcal{I}_V = \{1, 2\}$  с помощью процедур SSADEC, SSAREC (алгоритмы 1.1, 1.2)
3. По полученным выборкам  $\{\rho_E^{(r)}\}_{r=1}^R$ ,  $\{F_A^{(r)}\}_{r=1}^R$ ,  $\{F_V^{(r)}\}_{r=1}^R$ ,  $\{\mathcal{I}_A^{(r)}\}_{r=1}^R$  оценим необходимые средние характеристики по формулам (3.29)

Видно, что при известных  $\alpha$ ,  $T$ , когда можно посчитать  $\mathcal{G}_0$  с помощью формулы (3.27), предложенная процедура PEREMP показывает очень хорошее качество, почти во всех рассмотренных случаях она не хуже чем процедура PER при  $\rho_0^{opt}$  (наилучшем в среднем  $\rho_0$ ). Поэтому можно использовать выводы, сделанные по отношению к процедуре PER при  $\rho_0^{opt}$  (см. раздел 3.3) о том, что АИ, реализованная в процедуре PEREMP, обеспечивает хорошее качество, как по отношению к ВИ, так и независимо от ВИ.

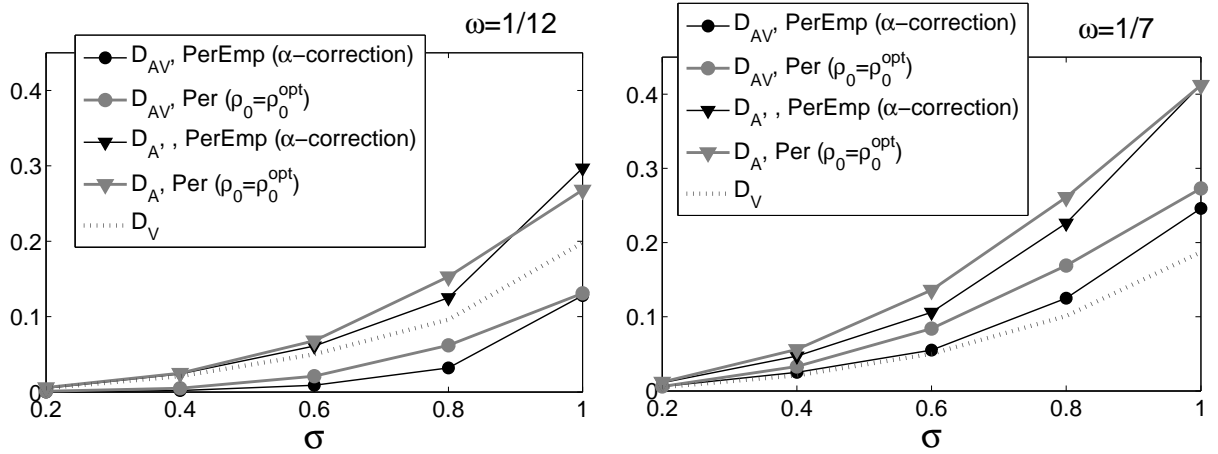


Рис. 3.10: Результаты для процедур PEREMP (с поправкой на  $\alpha$ ) и PER при  $\rho_0^{opt}$ ,  $\alpha = 0.02$  для  $\omega = 1/12$  (слева) и  $\omega = 1/7$  (справа).

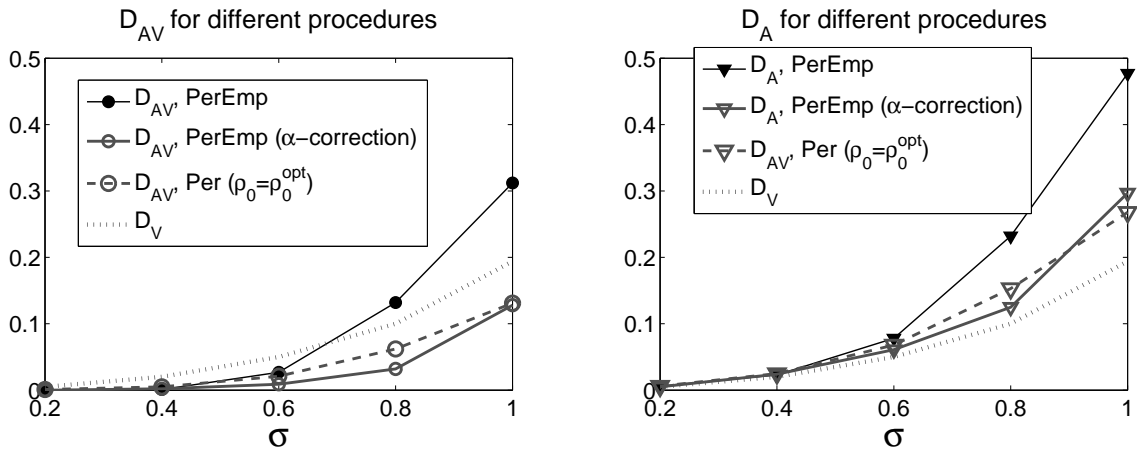


Рис. 3.11: Результаты для процедур PEREMP, PEREMP с поправкой на  $\alpha$  и PER при  $\rho_0^{opt}$ ,  $\alpha = 0.02$ ,  $\omega = 1/12$ :  $\bar{D}_{AV}$  (слева) и  $\bar{D}_A$  (справа).

**Замечание 3.4** На рис. 3.10, 3.11 видно, что процедура PEREMP в среднем показывает результаты лучше чем процедура PER при наилучшем в среднем  $\rho_0^{opt}$ . Это объясняется тем, что в процедуре PEREMP подбор  $\rho_0$  происходит для каждого ряда в отдельности, тогда как для процедуры PER приведены средние результаты обработки всех рядов при одном и том же  $\rho_0^{opt}$ .

Проведённые статистические исследования также показали, что при выделении  $\varepsilon$ -м гармоника использование формулы (3.27) для вычисления  $\mathcal{G}_0$  (учитывающей  $\alpha$ ), обеспечивает лучшее качество, чем применение формулы (3.25) (учитывающей только  $A$ ).

### 3.5.2 Выделение гармоника в присутствии тренда

Проверим теперь, будет ли работать процедура PEREMP в присутствии тренда, когда ряд представляется в виде суммы  $F = F^{(T)} + F^{(P)} + F^{(\varepsilon)}$ :

$$f_n = f_n^{(T)} + f_n^{(P)} + f_n^{(\varepsilon)}, \quad f_n^{(P)} = Ae^{\alpha n} \sin(2\pi\omega n + \phi),$$

а  $F^{(\varepsilon)}$  — ряд, являющийся реализацией случайного белого шума  $NID(0, \sigma^2)$ . Для этого рассмотрим примеры оценки характеристик процедуры PEREMP при выделении  $\varepsilon$ -м гармоника в присутствии

- точно отделимого от  $\varepsilon$ -м гармоника тренда,
- приближённо отделимого от  $\varepsilon$ -м гармоника тренда.

**Случай точно отделимого тренда** Рассмотрим примеры выделения  $\varepsilon$ -м гармоника из ряда  $F = F^{(T)} + F^{(P)} + F^{(\varepsilon)}$ , где  $F^{(T)}$  точно отделим от  $F^{(P)}$ .

Оценим характеристики процедуры PEREMP с помощью моделирования.

Возьмём те же самые значения параметров, что и для случая без тренда:  $N = 47$ ,  $\omega = 1/12$ ,  $\phi = 0$ ,  $L = 24$ ,  $\rho_0^{\uparrow} = 0$ ,  $\rho_0^{\downarrow} = 1$ ,  $\Delta\rho = 0.01$ ,  $R = 10^3$ ,  $\mathcal{P}\% = 75\%$ ,  $F^{(\varepsilon)} \sim NID(0, \sigma^2)$ ,  $0 < \sigma \leq 1$ . Приведём результаты только для  $\varepsilon$ -м гармоника с  $\alpha = 0.02$ , результаты для остальных взятых  $\alpha$  (0, 0.01) качественно не отличаются (приводят к тем же выводам), поэтому они здесь не приведены.

**Пример 3.2** Пусть  $F^{(T)}$ ,  $F^{(P)}$ :

$$f_n^{(T)} = 20, \quad f_n^{(P)} = \cos(2\pi\omega n + \phi).$$

Имитировать ВИ  $\varepsilon$ -м гармоника можно, взяв  $\mathcal{I}_\nu = \{2, 3\}$ , потому что в данной модели первая компонента соответствует тренду. Рассчитаем  $\mathcal{G}_0$  по формуле (3.25), учитывающей  $A$ :  $\mathcal{G}_0 = 0.375$ . Оценки характеристик приведены на рис. 3.12 вместе с оценками, полученными ранее, при отсутствии тренда. Видно, что оценки рассмотренных средних характеристик процедуры PEREMP при добавлении к ряду тренда практически не изменяются.

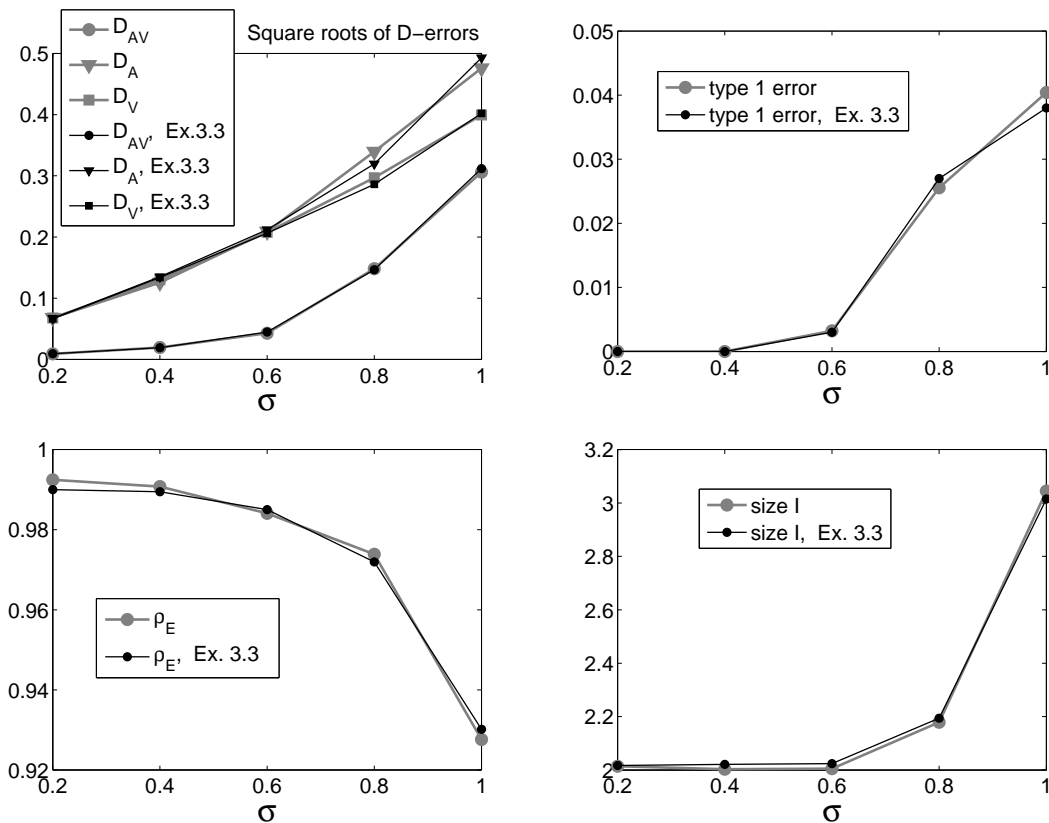


Рис. 3.12: Пример 3.2 (точно отделимый тренд): сравнение оценок характеристик процедуры PEREMP для моделей  $F = F^{(P)} + F^{(\varepsilon)}$  (серые линии) и  $F = F^{(T)} + F^{(P)} + F^{(\varepsilon)}$ , где тренд — константа (черные линии).

**Пример 3.3** Пусть  $F^{(T)}, F^{(P)}$ :

$$f_n^{(T)} = 10e^{-0.02n}, \quad f_n^{(P)} = e^{0.02n} \cos(2\pi n/12). \quad (3.30)$$

Возьмём  $\mathcal{I}_Y = \{2, 3\}$ . Значение  $\mathcal{G}_0$  рассчитаем по формуле (3.27) (учитывая  $A, \alpha$  и  $T$ ):  $\mathcal{G}_0 = 0.86$ . Результаты приведены на рис. 3.13. Видно, что оценки рассмотренных характеристик процедуры PEREMP ( $\bar{D}_{AV}, \bar{D}_A, \bar{D}_V, \tilde{\mathcal{E}}_1$ ) почти не изменяются при добавлении тренда.

**Случай приближённо отделимого тренда**

**Пример 3.4** Пусть  $F^{(T)}, F^{(P)}$ :

$$f_n^{(T)} = 5e^{0.03n} + 10, \quad f_n^{(P)} = e^{0.02n} \cos(2\pi/12n). \quad (3.31)$$

Возьмём  $\mathcal{I}_Y = \{2, 3\}$ ,  $\mathcal{G}_0 = 0.86$ . Результаты приведены на рис. 3.14. Видно, что из-за приближённого характера разделимости ошибка ВИ  $\bar{D}_V$  становится больше. Тем не менее уровень ошибки  $\bar{D}_{AV}$  (АИ по сравнению с ВИ) сохраняется, что говорит о том, что в данном примере приближённый характер разде-



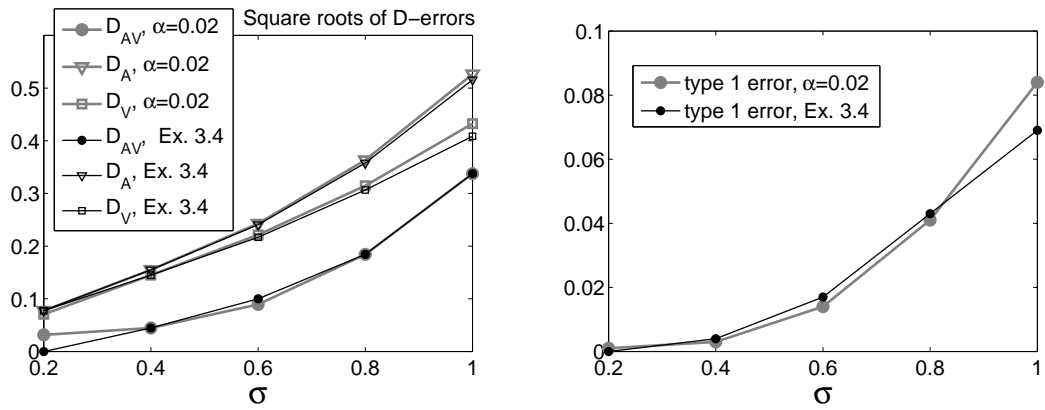


Рис. 3.13: Пример 3.3 (точно отделимый тренд): сравнение оценок характеристик процедуры PEREMP для моделей  $F = F^{(P)} + F^{(\varepsilon)}$  (серые линии) и  $F = F^{(T)} + F^{(P)} + F^{(\varepsilon)}$ , где тренд — экспонента (черные линии).

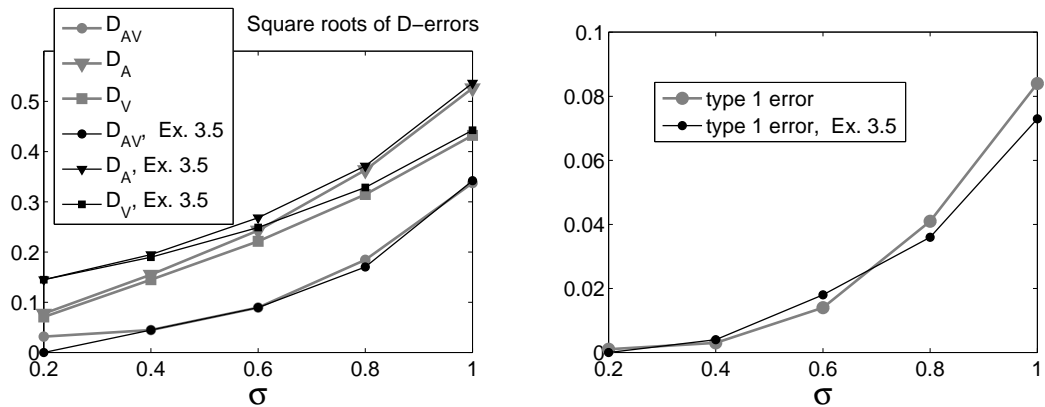


Рис. 3.14: Пример 3.4 (приближённо отделимый тренд): сравнение оценок характеристик процедуры PEREMP для моделей  $F = F^{(P)} + F^{(\varepsilon)}$  (серые линии) и  $F = F^{(T)} + F^{(P)} + F^{(\varepsilon)}$ , где тренд — константа + экспонента (черные линии).

лимости не сказался на качестве автоматической идентификации. Это подтверждается графиком оценки ошибки первого рода  $\tilde{\mathcal{E}}_1$ , на котором видно, что её отличия от случаев без тренда и точно отделимого тренда несущественны.

Для того чтобы продемонстрировать сложность поставленной задачи, приведём график ряда  $F$  — реализации из примера 3.4 без шума и с шумом при  $\sigma = 1$ , см. рис. 3.15. Видно, что уровень зашумлённости велик настолько, что сложно визуально выделить  $\varepsilon$ -м гармоническую составляющую.

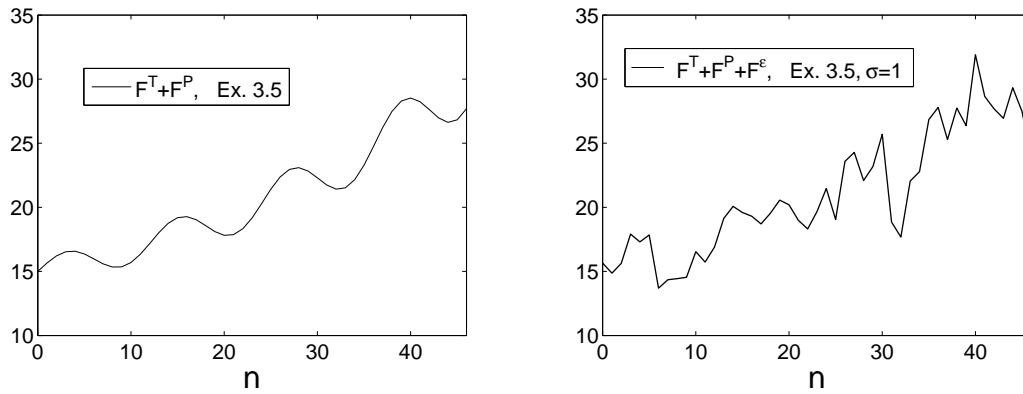


Рис. 3.15: График ряда-реализации из примера 3.4 без шума (слева) и с шумом (справа).

### 3.6 Оценка частоты выделенной $\varepsilon$ -м гармоники

Для того, чтобы решить задачу выделения периодической составляющей ряда  $F$  с некоторым известным периодом  $T$  (которая определена согласно (0.1)), необходимо выделить все  $\varepsilon$ -м гармоники с частотами из множества  $\{k/T\}_{k=1}^{\lfloor T/2 \rfloor}$ . Для этого среди всех собственных троек, идентифицированных с помощью метода Фурье, необходимо отобрать те, которые соответствуют гармоникам с требуемыми частотами. Опишем метод оценки частоты выделенной гармоники.

В работе [2, раздел 5] нами было предложено три метода оценки частоты  $\varepsilon$ -м гармоники.

Первый основан на идее, используемой при визуальной идентификации — представление собственных векторов  $\varepsilon$ -м гармоники с помощью парных диаграмм собственных векторов. Второй метод использует разложение Фурье выделенной  $\varepsilon$ -м гармоники, частота которой оценивается аргументом максимума её периодограммы.

Третий метод основан на использовании корней характеристического полинома ЛРФ, управляющей выделенной  $\varepsilon$ -м гармоникой. Этот метод показал лучшее качество, опишем его более подробно.

Пусть  $F^{(P)}$  —  $\varepsilon$ -м гармонический ряд с частотой  $\omega$  и экспоненциальным показателем  $\alpha$ . Найдём по его собственным векторам соответствующую ему ЛРФ порядка  $L - 1$  с помощью (1.12), см. раздел 1.4.1:

$$f_{i+L-1}^{(P)} = \sum_{k=1}^{L-1} a_k f_{i+L-1-k}^{(P)}. \quad (3.32)$$

После того, как вычислены коэффициенты  $a_1, \dots, a_{L-1}$ , найдём корни характеристического полинома, соответствующего этой ЛРФ:

$$x^{L-1} = \sum_{k=1}^{L-1} a_k x^{L-1-k}.$$

Среди всех корней найдём главные корни  $x_1, x_2$ , пользуясь сведениями из раздела 1.4.2 (пример 1.1). Они имеют вид  $x_1 = e^{\alpha+i2\pi\omega}$ ,  $x_2 = e^{\alpha-i2\pi\omega}$  и частота  $\omega$  выражается следующим образом:

$$\omega = \frac{1}{2\pi} \arccos\left(\frac{\Re x_1}{|x_1|}\right), \quad (3.33)$$

где  $\Re z$  для  $z \in \mathbb{C}$  обозначает вещественную часть  $z$ .

### 3.7 Пример выделения периодической составляющей реального ряда

Рассмотрим пример применения процедуры PEREMP к реальному ряду.

Возьмём ряд ElecPower, содержащий ежемесячные данные об уровне потребления электроэнергии в США с января 1972 г. по октябрь 2005 г. Длина ряда  $N = 406$ , ряд изображён на рис. 3.16 (слева сверху). Видно, что ряд содержит как тренд, так и периодическую составляющую с периодом  $T \approx 12$ , отражающую сезонное поведение ряда.

Разложим ряд на тренд, сезонную составляющую и остаток. Выделим сначала тренд.

Возьмём длину окна  $L = 198$  (близкую к  $N/2$  и делящуюся на предполагаемый период периодики). По периодограмме ряда ElecPower (рис. 3.16) определим границу низких частот, равную приближённо 0.05, и возьмём  $\omega_0 = 0.056 \approx \lceil 0.05L \rceil / L + 1/L$ .

Выделим тренд с помощью процедуры TRRMEAS с параметрами  $L = 198$ ,  $\omega_0 = 0.56$ ,  $C_0^l = 0.5$ ,  $C_0^h = 1$ ,  $\Delta C = 0.01$  и  $\Delta R = 0.05$ . Результат процедуры (выделенный тренд  $F_A^{(T)}$ ) изображён на рис. 3.16 (внизу).

Рассмотрим теперь ряд  $F - F_A^{(T)}$ , это исходный ряд, из которого вычтен выделенный тренд, см. рис. 3.17 (слева сверху). На основании вида ряда предположим, что минимальная амплитуда для  $n$ -м гармоник, составляющих периодическую составляющую, равна  $A_{min} = 1$ .

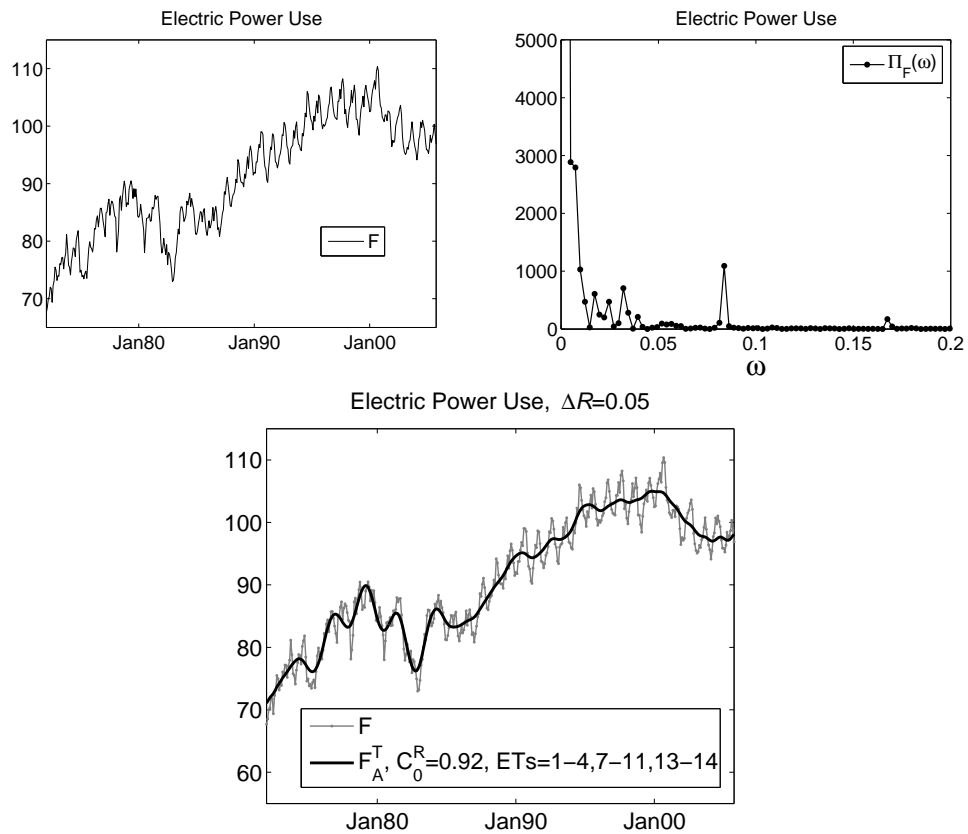


Рис. 3.16: Ряд ElecPower, выделение тренда.

Рассчитаем  $\mathcal{G}_0$  по формуле (3.25) (случай неизвестных  $\alpha_k$ ), взяв  $\mathcal{P}\% = 50\%$ :  $\mathcal{G}_0 = \mathcal{G}_0(1, 50\%) = 0.25$ . Применим к ряду  $F - F_A^{(T)}$  процедуру PEREMP с параметрами  $A_{min} = 1$ ,  $L = 198$ ,  $s_0 = 1$ ,  $\rho_0^l = 0.5$ ,  $\rho_0^r = 1$ ,  $\Delta\rho = 0.01$ ,  $\mathcal{P}\% = 50\%$ .

С помощью метода оценки частоты (раздел 3.6) оценим частоты всех выделенных э-м гармоник и возьмём только те, для которых выполняется:  $1/13 \leq \hat{\omega} \leq 1/11$ . Результат (выделенная сезонная составляющая) изображён на рис. 3.17 (внизу). На рис. 3.18 изображено полученное разложение ряда ElecPower на тренд, сезонную составляющую и остаток.

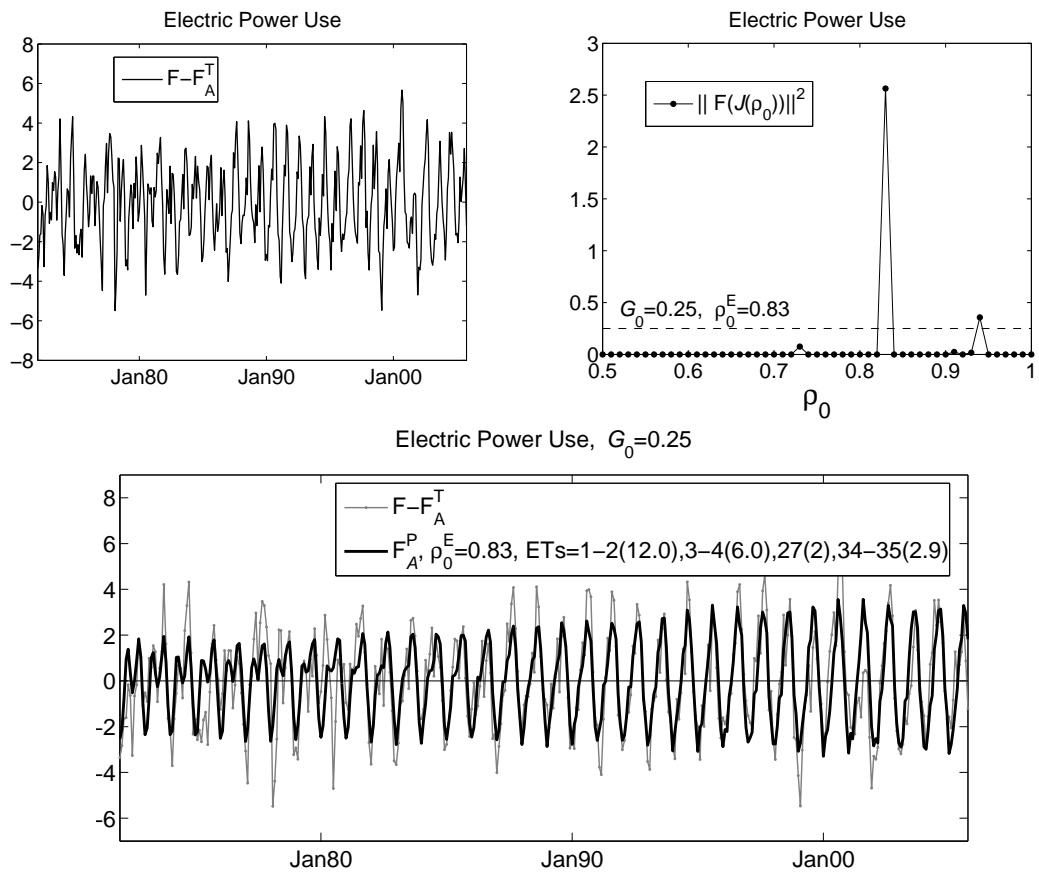


Рис. 3.17: Ряд ElecPower: выделение сезонной составляющей из  $F - F_A^{(T)}$  (в скобках на нижнем графике — оценки периодов э-м гармоник).

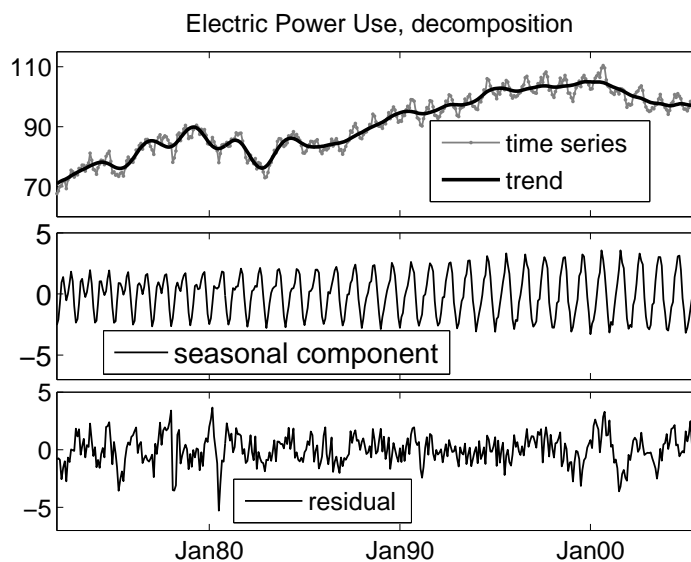


Рис. 3.18: Ряд ElecPower: результат разложения ряда на тренд, сезонную составляющую (периоду с периодом 12) и остаток.

### 3.8 Прогноз периодических компонент

Как и в разделе 2.11 для тренда, рассмотрим проблему автоматического прогноза периодической составляющей. В разделе 2.11.2 было показано, что при использовании подхода “Гусеница”-SSA эта проблема сводится к проблеме идентификации нужных компонент, которая при использовании процедуры PER, сводится к выбору пороговых значений метода Фурье.

**АЛГОРИТМ 3.7. Процедура PERFORE: прогноз тренда ряда с помощью процедуры PER**

**Входные данные:** Ряд  $F$  длины  $N$ , содержащий  $F^{(P)}$

**Результат:**  $\tilde{f}$  — прогноз значения  $f_N^{(P)}$

**Параметры:**  $L$  ( $1 < L \leq \lceil N/2 \rceil$ ),  $s_0 \in \mathbb{Z}_+$ ,  $\rho_0 \in [0, 1]$

**Процедура:**

1. Применим процедуру PER с параметрами  $L$ ,  $s_0$ ,  $\rho_0$  к  $F$ , получив  $\mathcal{I}_A$  (алгоритм 3.3)
2. Применим процедуру SSAFORE с параметрами  $L$ ,  $\mathcal{I}_A$  к ряду  $F$ , получив прогноз  $\tilde{f}$  (алгоритм 1.3)

#### 3.8.1 Моделирование прогноза $\varepsilon$ -м гармоники

Изучим в данном разделе пример прогноза  $\varepsilon$ -м гармоники с помощью процедуры PERFORE (алгоритм 3.7) при наилучшем в среднем пороговом значении (обозначим его  $\rho_0^{opt\mathfrak{F}}$ ) для уже рассмотренной в разделе 3.3 модели ряда  $F = (f_0, \dots, f_{N-1})$ :

$$f_n = f_n^{(P)} + f_n^{(\varepsilon)}, \quad f_n^{(P)} = Ae^{an} \cos(2\pi\omega n), \quad f_n^{(\varepsilon)} = \sigma e^{an} \varepsilon_n.$$

С помощью процедуры PERFORE будем прогнозировать  $\varepsilon$ -м гармонику на одну точку вперёд, оценивая тем самым  $f_N^{(P)}$ . Обозначим полученное значение за  $\tilde{f}_A$ .

Сравним результат процедуры PERFORE с результатом визуального прогноза, который можно имитировать, взяв в качестве группы собственных троек, соответствующих  $\varepsilon$ -м гармонике,  $\mathcal{I}_V = \{1, 2\}$ . Результат такого прогнозирования обозначим как  $\tilde{f}_V$ .

Исследуем, как ведёт себя процедура PERFORE при наилучшем в среднем  $\rho_0^{opt\mathfrak{F}}$ .

Рассчитаем те же характеристики процедуры PERFORE, что и в разделе 3.3, с той разницей, что квадратичная ошибка  $\mathcal{D}$  рассчитывается только в точке прогнозирования  $N$ :

$$\mathcal{D}_{AV} = e^{-\alpha N} (\tilde{f}_A - \tilde{f}_V)^2, \quad \mathcal{D}_A = e^{-\alpha N} (f_N^{(P)} - \tilde{f}_A)^2, \quad \mathcal{D}_V = e^{-\alpha N} (f_N^{(P)} - \tilde{f}_V)^2.$$

**Параметры** Возьмём те же самые параметры, что и в разделе 3.3.1, при расчёте средних характеристик процедуры PER (аппроксимация) при наилучшем в среднем  $\rho_0^{opt}$ :  $N = 27$ ,  $L = 24$ ,  $\omega = 1/12$ ,  $R = 10^4$ .

Приведём результаты для  $\alpha = 0$ ,  $\alpha = 0.02$  (рис. 3.19, 3.20). На этих рисунках для сравнения изображены результаты для процедуры PER при  $\rho_0^{optA} = \rho_0^{opt}$  (для аппроксимации) взятые из раздела 3.3. Для  $\mathcal{D}$ -ошибок на графиках изображены корни их значений, чтобы подчеркнуть их линейную зависимость от  $\sigma$ .

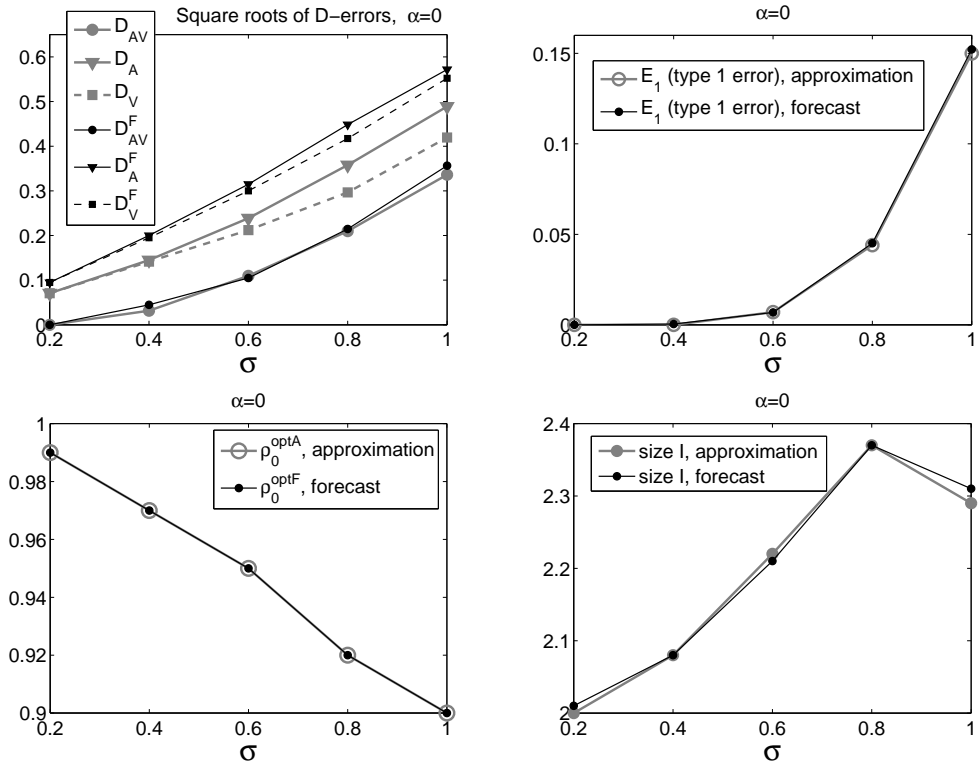


Рис. 3.19:  $\rho_0^{opt}$ ,  $\rho_0^{optF}$  и оценки характеристик процедур PER при  $\rho_0^{opt}$  (аппроксимация) и PERFORE при  $\rho_0^{optF}$  (прогноз),  $\alpha = 0$ .

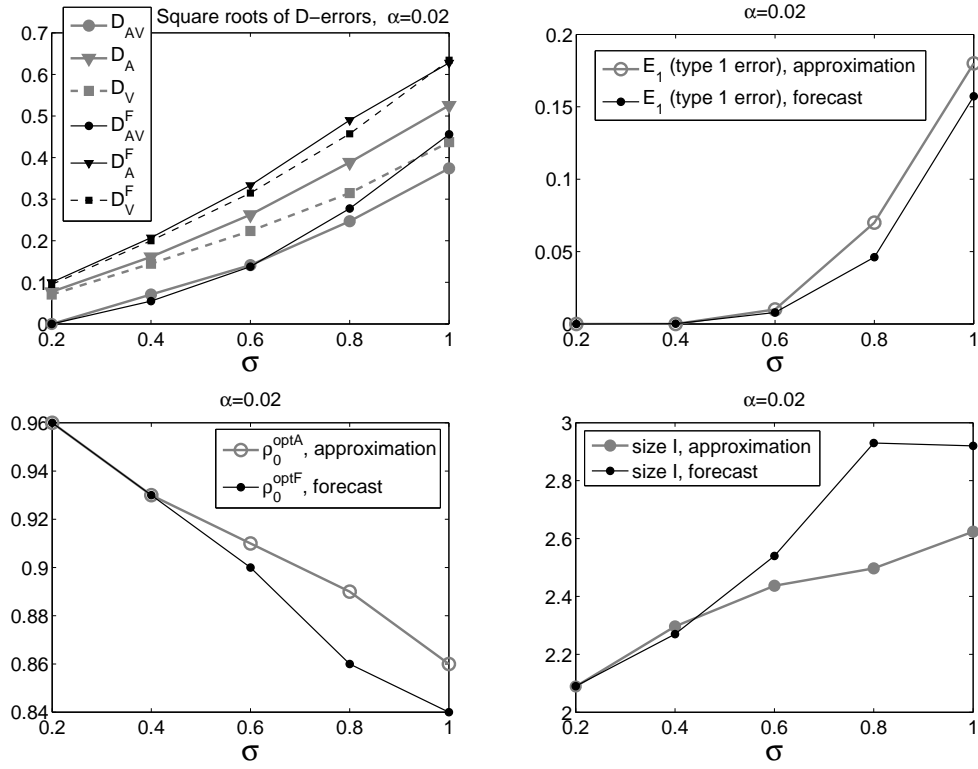


Рис. 3.20:  $\rho_0^{opt}$ ,  $\rho_0^{opt\mathfrak{F}}$  и оценки характеристик процедур PER при  $\rho_0^{opt}$  (аппроксимация) и PERFORE при  $\rho_0^{opt\mathfrak{F}}$  (прогноз),  $\alpha = 0.02$ .

**Выводы** Видно, что, во-первых, при прогнозе э-м гармоники процедура PERFORE, основанная на методе Фурье при наилучшем в среднем  $\rho_0^{opt\mathfrak{F}}$ , решает поставленную задачу — автоматическую идентификацию, близкую к визуальной. Это подтверждается как выполнением

$$\overline{D}_{AV}^{opt\mathfrak{F}} \ll \overline{D}_V,$$

(т. е. АИ вносит небольшую ошибку по сравнению с ошибкой визуальной), так и малыми значениями ошибки первого рода  $\tilde{\mathcal{E}}_1^{opt\mathfrak{F}}$ . Также видно, что уровень  $\sqrt{\overline{D}_A}$  достаточно мал по сравнению с  $\sigma$ .

Во-вторых, при сравнении графиков на рис. 3.19, 3.20, видно, что при изменении  $\alpha$  значения  $\tilde{\mathcal{E}}_1(\rho_0^{opt\mathfrak{F}})$  и  $\overline{D}_{AV}^{(F)}(\rho_0^{opt\mathfrak{F}})$  мало изменяются (т. е. процедура TRFORE адаптируется к изменению  $\alpha$ ).

В-третьих, значения характеристик процедуры PERFORE (прогноз) при  $\rho_0^{opt\mathfrak{F}}$  и процедуры PER (аппроксимация) при  $\rho_0^{opt}$  близки (при  $\alpha = 0$  они практически совпадают, при увеличении  $\alpha$  отличия вырастают, но похожесть по-прежнему сохраняется). Это позволяет при прогнозе периодической составляющей пользоваться теми же принципами выбора  $\rho_0$  что и при аппроксимации (например использовать эмпирический подход к выбору  $\rho_0$  из раздела 3.5).



Таблица 3.8: Характеристики процедур PEREMP (с эмпирическим расчётом  $\rho_0$  по имеющейся информации о э-м гармонической составляющей) и PER при  $\rho_0^{opt}$ ,  $\alpha = 0$ ,  $L\omega \in \mathbb{N}$  (раздел 3.5.1)

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math></b> |                |                          |                       |                       |                              |                         |
|--|----------------|--------------------------|-----------------------|-----------------------|------------------------------|-------------------------|
| $\sigma$   | $\bar{\rho}_E$ | $\bar{\mathcal{D}}_{AV}$ | $\bar{\mathcal{D}}_A$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A$ | $\tilde{\mathcal{E}}_1$ |
| 0.2  | 0.99           | 0.000                    | 0.005                 | 0.005                 | 2.01                         | 0.000                   |
| 0.4  | 0.99           | 0.012                    | 0.036                 | 0.019                 | 1.95                         | 0.034                   |
| 0.6  | 0.99           | 0.032                    | 0.088                 | 0.045                 | 1.85                         | 0.092                   |
| 0.8  | 0.98           | 0.045                    | 0.142                 | 0.091                 | 1.88                         | 0.120                   |
| 1.0  | 0.96           | 0.075                    | 0.218                 | 0.170                 | 2.14                         | 0.136                   |

| <b>PER с пороговым значением <math>\rho_0 = \rho_0^{opt}</math></b> |                |                                |                             |                       |                                    |                               |
|---|----------------|--------------------------------|-----------------------------|-----------------------|------------------------------------|-------------------------------|
| $\sigma$  | $\rho_0^{opt}$ | $\bar{\mathcal{D}}_{AV}^{opt}$ | $\bar{\mathcal{D}}_A^{opt}$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2   | 0.99           | 0.000                          | 0.005                       | 0.005                 | 2.00                               | 0.000                         |
| 0.4   | 0.97           | 0.001                          | 0.021                       | 0.020                 | 2.08                               | 0.000                         |
| 0.6   | 0.96           | 0.012                          | 0.057                       | 0.045                 | 2.11                               | 0.016                         |
| 0.8   | 0.92           | 0.044                          | 0.128                       | 0.088                 | 2.37                               | 0.044                         |
| 1.0   | 0.91           | 0.113                          | 0.236                       | 0.176                 | 2.15                               | 0.174                         |

Таблица 3.9: Характеристики процедур PEREMP (с эмпирическим расчётом  $\rho_0$  по имеющейся информации о э-м гармонической составляющей) и PER при  $\rho_0^{opt}$ ,  $\alpha = 0.01$ ,  $L\omega \in \mathbb{N}$  (раздел 3.5.1)

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math></b> |                |                          |                       |                       |                              |                         |
|--|----------------|--------------------------|-----------------------|-----------------------|------------------------------|-------------------------|
| $\sigma$   | $\bar{\rho}_E$ | $\bar{\mathcal{D}}_{AV}$ | $\bar{\mathcal{D}}_A$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A$ | $\tilde{\mathcal{E}}_1$ |
| 0.2  | 0.99           | 0.000                    | 0.005                 | 0.005                 | 2.01                         | 0.000                   |
| 0.4  | 0.99           | 0.000                    | 0.020                 | 0.020                 | 2.03                         | 0.000                   |
| 0.6  | 0.98           | 0.003                    | 0.050                 | 0.047                 | 2.07                         | 0.001                   |
| 0.8  | 0.95           | 0.035                    | 0.124                 | 0.091                 | 2.51                         | 0.018                   |
| 1.0  | 0.90           | 0.137                    | 0.286                 | 0.161                 | 3.69                         | 0.041                   |

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math>, <math>\alpha</math>, <math>T</math></b> |                |                          |                       |                       |                              |                         |
|---|----------------|--------------------------|-----------------------|-----------------------|------------------------------|-------------------------|
| $\sigma$  | $\bar{\rho}_E$ | $\bar{\mathcal{D}}_{AV}$ | $\bar{\mathcal{D}}_A$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A$ | $\tilde{\mathcal{E}}_1$ |
| 0.2   | 0.99           | 0.000                    | 0.005                 | 0.005                 | 2.01                         | 0.000                   |
| 0.4   | 0.99           | 0.003                    | 0.024                 | 0.020                 | 2.01                         | 0.008                   |
| 0.6   | 0.98           | 0.014                    | 0.065                 | 0.046                 | 1.98                         | 0.040                   |
| 0.8   | 0.97           | 0.032                    | 0.125                 | 0.090                 | 2.04                         | 0.076                   |
| 1.0   | 0.94           | 0.086                    | 0.234                 | 0.177                 | 2.47                         | 0.109                   |

| <b>PER с пороговым значением <math>\rho_0 = \rho_0^{opt}</math></b> |                |                                |                             |                       |                                    |                               |
|---|----------------|--------------------------------|-----------------------------|-----------------------|------------------------------------|-------------------------------|
| $\sigma$  | $\rho_0^{opt}$ | $\bar{\mathcal{D}}_{AV}^{opt}$ | $\bar{\mathcal{D}}_A^{opt}$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2   | 0.98           | 0.000                          | 0.005                       | 0.005                 | 2.02                               | 0.000                         |
| 0.4   | 0.97           | 0.002                          | 0.021                       | 0.020                 | 2.06                               | 0.001                         |
| 0.6   | 0.93           | 0.018                          | 0.063                       | 0.046                 | 2.38                               | 0.006                         |
| 0.8   | 0.92           | 0.056                          | 0.138                       | 0.092                 | 2.33                               | 0.072                         |
| 1.0   | 0.88           | 0.114                          | 0.246                       | 0.176                 | 2.61                               | 0.125                         |

Таблица 3.10: Характеристики процедур PEREMP (с эмпирическим расчётом  $\rho_0$  по имеющейся информации о  $\varepsilon$ -м гармонической составляющей) и PER при  $\rho_0^{opt}$ ,  $\alpha = 0.02$ ,  $L\omega \in \mathbb{N}$  (раздел 3.5.1)

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math></b> |                |                          |                       |                       |                              |                         |
|--|----------------|--------------------------|-----------------------|-----------------------|------------------------------|-------------------------|
| $\sigma$   | $\bar{\rho}_E$ | $\bar{\mathcal{D}}_{AV}$ | $\bar{\mathcal{D}}_A$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A$ | $\tilde{\mathcal{E}}_1$ |
| 0.2  | 0.98           | 0.000                    | 0.006                 | 0.005                 | 2.03                         | 0.000                   |
| 0.4  | 0.98           | 0.001                    | 0.023                 | 0.022                 | 2.07                         | 0.000                   |
| 0.6  | 0.94           | 0.027                    | 0.078                 | 0.050                 | 3.00                         | 0.001                   |
| 0.8  | 0.85           | 0.132                    | 0.232                 | 0.100                 | 5.12                         | 0.009                   |
| 1.0  | 0.76           | 0.312                    | 0.477                 | 0.187                 | 7.26                         | 0.055                   |

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math>, <math>\alpha</math>, <math>T</math></b> |                |                          |                       |                       |                              |                         |
|---|----------------|--------------------------|-----------------------|-----------------------|------------------------------|-------------------------|
| $\sigma$  | $\bar{\rho}_E$ | $\bar{\mathcal{D}}_{AV}$ | $\bar{\mathcal{D}}_A$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A$ | $\tilde{\mathcal{E}}_1$ |
| 0.2   | 0.98           | 0.000                    | 0.005                 | 0.005                 | 2.04                         | 0.000                   |
| 0.4   | 0.98           | 0.002                    | 0.024                 | 0.021                 | 2.08                         | 0.003                   |
| 0.6   | 0.97           | 0.009                    | 0.061                 | 0.050                 | 2.08                         | 0.018                   |
| 0.8   | 0.95           | 0.032                    | 0.125                 | 0.096                 | 2.42                         | 0.027                   |
| 1.0   | 0.89           | 0.128                    | 0.297                 | 0.198                 | 3.47                         | 0.086                   |

| <b>PER с пороговым значением <math>\rho_0 = \rho_0^{opt}</math></b> |                |                                |                             |                       |                                    |                               |
|---|----------------|--------------------------------|-----------------------------|-----------------------|------------------------------------|-------------------------------|
| $\sigma$  | $\rho_0^{opt}$ | $\bar{\mathcal{D}}_{AV}^{opt}$ | $\bar{\mathcal{D}}_A^{opt}$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2   | 0.96           | 0.001                          | 0.006                       | 0.005                 | 2.11                               | 0.000                         |
| 0.4   | 0.94           | 0.005                          | 0.025                       | 0.020                 | 2.26                               | 0.000                         |
| 0.6   | 0.92           | 0.021                          | 0.068                       | 0.050                 | 2.34                               | 0.020                         |
| 0.8   | 0.88           | 0.062                          | 0.153                       | 0.100                 | 2.67                               | 0.060                         |
| 1.0   | 0.88           | 0.131                          | 0.268                       | 0.194                 | 2.26                               | 0.210                         |

Таблица 3.11: Характеристики процедур PEREMP (с эмпирическим расчётом  $\rho_0$  по имеющейся информации о э-м гармонической составляющей) и PER при  $\rho_0^{opt}$ ,  $\alpha = 0$ ,  $L\omega \notin \mathbb{N}$  (раздел 3.5.1)

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math></b> |                |                          |                       |                       |                              |                         |
|--|----------------|--------------------------|-----------------------|-----------------------|------------------------------|-------------------------|
| $\sigma$   | $\bar{\rho}_E$ | $\bar{\mathcal{D}}_{AV}$ | $\bar{\mathcal{D}}_A$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A$ | $\tilde{\mathcal{E}}_1$ |
| 0.2  | 0.82           | 0.007                    | 0.011                 | 0.005                 | 4.26                         | 0.001                   |
| 0.4  | 0.82           | 0.026                    | 0.047                 | 0.020                 | 4.36                         | 0.002                   |
| 0.6  | 0.81           | 0.059                    | 0.108                 | 0.045                 | 4.52                         | 0.005                   |
| 0.8  | 0.81           | 0.114                    | 0.208                 | 0.091                 | 4.67                         | 0.030                   |
| 1.0  | 0.79           | 0.213                    | 0.374                 | 0.179                 | 5.10                         | 0.074                   |

| <b>PER с пороговым значением <math>\rho_0 = \rho_0^{opt}</math></b> |                |                                |                             |                       |                                    |                               |
|---|----------------|--------------------------------|-----------------------------|-----------------------|------------------------------------|-------------------------------|
| $\sigma$  | $\rho_0^{opt}$ | $\bar{\mathcal{D}}_{AV}^{opt}$ | $\bar{\mathcal{D}}_A^{opt}$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2   | 0.79           | 0.006                          | 0.011                       | 0.005                 | 4.26                               | 0.000                         |
| 0.4   | 0.75           | 0.031                          | 0.051                       | 0.019                 | 5.15                               | 0.001                         |
| 0.6   | 0.74           | 0.084                          | 0.130                       | 0.045                 | 5.33                               | 0.029                         |
| 0.8   | 0.72           | 0.160                          | 0.246                       | 0.087                 | 5.70                               | 0.065                         |
| 1.0   | 0.72           | 0.271                          | 0.403                       | 0.169                 | 5.31                               | 0.192                         |

Таблица 3.12: Характеристики процедур PEREMP (с эмпирическим расчётом  $\rho_0$  по имеющейся информации о э-м гармонической составляющей) и PER при  $\rho_0^{opt}$ ,  $\alpha = 0.01$ ,  $L\omega \notin \mathbb{N}$  (раздел 3.5.1)

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math></b> |                |                          |                       |                       |                              |                         |
|--|----------------|--------------------------|-----------------------|-----------------------|------------------------------|-------------------------|
| $\sigma$   | $\bar{\rho}_E$ | $\bar{\mathcal{D}}_{AV}$ | $\bar{\mathcal{D}}_A$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A$ | $\tilde{\mathcal{E}}_1$ |
| 0.2  | 0.82           | 0.006                    | 0.012                 | 0.005                 | 4.27                         | 0.001                   |
| 0.4  | 0.82           | 0.025                    | 0.046                 | 0.020                 | 4.31                         | 0.002                   |
| 0.6  | 0.81           | 0.060                    | 0.109                 | 0.047                 | 4.58                         | 0.002                   |
| 0.8  | 0.78           | 0.144                    | 0.241                 | 0.095                 | 5.62                         | 0.019                   |
| 1.0  | 0.74           | 0.291                    | 0.459                 | 0.180                 | 6.91                         | 0.063                   |

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math></b> |      |       |       |       |      |       |
|--|------|-------|-------|-------|------|-------|
| 0.2  | 0.82 | 0.006 | 0.011 | 0.005 | 4.18 | 0.001 |
| 0.4  | 0.82 | 0.025 | 0.046 | 0.020 | 4.27 | 0.003 |
| 0.6  | 0.81 | 0.059 | 0.108 | 0.048 | 4.42 | 0.008 |
| 0.8  | 0.80 | 0.115 | 0.215 | 0.095 | 4.69 | 0.038 |
| 1.0  | 0.77 | 0.222 | 0.380 | 0.166 | 5.55 | 0.068 |

| <b>PER с пороговым значением <math>\rho_0 = \rho_0^{opt}</math></b> |                |                                |                             |                       |                                    |                               |
|---|----------------|--------------------------------|-----------------------------|-----------------------|------------------------------------|-------------------------------|
| $\sigma$  | $\rho_0^{opt}$ | $\bar{\mathcal{D}}_{AV}^{opt}$ | $\bar{\mathcal{D}}_A^{opt}$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2   | 0.79           | 0.006                          | 0.011                       | 0.005                 | 4.21                               | 0.000                         |
| 0.4   | 0.75           | 0.032                          | 0.053                       | 0.019                 | 5.04                               | 0.005                         |
| 0.6   | 0.73           | 0.085                          | 0.135                       | 0.045                 | 5.42                               | 0.032                         |
| 0.8   | 0.72           | 0.169                          | 0.259                       | 0.089                 | 5.51                               | 0.101                         |
| 1.0   | 0.71           | 0.267                          | 0.410                       | 0.173                 | 5.48                               | 0.188                         |

Таблица 3.13: Характеристики процедур PEREMP (с эмпирическим расчётом  $\rho_0$  по имеющейся информации о э-м гармонической составляющей) и PER при  $\rho_0^{opt}$ ,  $\alpha = 0.02$ ,  $L\omega \notin \mathbb{N}$  (раздел 3.5.1)

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math></b> |                |                          |                       |                       |                              |                         |
|--|----------------|--------------------------|-----------------------|-----------------------|------------------------------|-------------------------|
| $\sigma$   | $\bar{\rho}_E$ | $\bar{\mathcal{D}}_{AV}$ | $\bar{\mathcal{D}}_A$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A$ | $\tilde{\mathcal{E}}_1$ |
| 0.2  | 0.82           | 0.005                    | 0.011                 | 0.005                 | 4.19                         | 0.000                   |
| 0.4  | 0.81           | 0.023                    | 0.044                 | 0.021                 | 4.31                         | 0.000                   |
| 0.6  | 0.77           | 0.081                    | 0.135                 | 0.052                 | 5.72                         | 0.000                   |
| 0.8  | 0.71           | 0.203                    | 0.303                 | 0.100                 | 7.66                         | 0.021                   |
| 1.0  | 0.66           | 0.372                    | 0.543                 | 0.191                 | 9.39                         | 0.065                   |

| <b>PEREMP, расчёт <math>\mathcal{G}_0</math> по <math>A</math></b> |      |       |       |       |      |       |
|--|------|-------|-------|-------|------|-------|
| 0.2  | 0.82 | 0.006 | 0.011 | 0.005 | 4.16 | 0.001 |
| 0.4  | 0.81 | 0.025 | 0.047 | 0.022 | 4.27 | 0.004 |
| 0.6  | 0.81 | 0.055 | 0.106 | 0.049 | 4.46 | 0.005 |
| 0.8  | 0.78 | 0.125 | 0.226 | 0.098 | 5.27 | 0.022 |
| 1.0  | 0.75 | 0.246 | 0.413 | 0.187 | 6.09 | 0.094 |

| <b>PER с пороговым значением <math>\rho_0 = \rho_0^{opt}</math></b> |                |                                |                             |                       |                                    |                               |
|---|----------------|--------------------------------|-----------------------------|-----------------------|------------------------------------|-------------------------------|
| $\sigma$  | $\rho_0^{opt}$ | $\bar{\mathcal{D}}_{AV}^{opt}$ | $\bar{\mathcal{D}}_A^{opt}$ | $\bar{\mathcal{D}}_V$ | $\overline{\#\mathcal{I}}_A^{opt}$ | $\tilde{\mathcal{E}}_1^{opt}$ |
| 0.2   | 0.76           | 0.006                          | 0.012                       | 0.005                 | 4.75                               | 0.000                         |
| 0.4   | 0.74           | 0.033                          | 0.056                       | 0.021                 | 5.14                               | 0.012                         |
| 0.6   | 0.71           | 0.084                          | 0.136                       | 0.050                 | 5.81                               | 0.031                         |
| 0.8   | 0.70           | 0.169                          | 0.261                       | 0.101                 | 5.93                               | 0.103                         |
| 1.0   | 0.71           | 0.273                          | 0.412                       | 0.187                 | 5.28                               | 0.230                         |

## Глава 4

### Оценка коэффициентов линейной рекуррентной формулы порядка 2

Рассмотрим одномерный вещественнозначный временной ряд  $F = (f_0, \dots, f_{N-1})$  длины  $N$  ( $N > 2$ ). Если под моделью ряда понимать способ задания его элементов  $f_n$ , то можно выделить два типа моделей: *авторегрессионные* (AR) и “сигнал плюс шум”.

В AR-модели  $n$ -ый элемент ряда  $F$  задаётся следующим образом:

$$f_n = \mathcal{A}(f_0, \dots, f_{n-1}) + \varepsilon_n,$$

а в модели “сигнал плюс шум”:

$$f_n = s_n + \varepsilon_n, \quad s_n = \mathcal{S}(s_0, \dots, s_{n-1}),$$

где  $\mathcal{A}$ ,  $\mathcal{S}$  — некоторые функции, а  $\varepsilon_n$  соответствует случайной составляющей.

Разница между ними состоит в том, что в AR-модели  $n$ -ый элемент  $\varepsilon_n$  воздействует не только на  $f_n$ , но и на все последующие элементы  $f_{n+1}, f_{n+2}, \dots$

Ряд модели “сигнал плюс шум” может быть представлен как детерминированный сигнал  $(s_0, \dots, s_{N-1})$ ,  $n$ -ое наблюдение которого ( $f_n$ ) было получено с ошибкой  $\varepsilon_n$ .

Обычно рассматриваются AR-модели с линейной функцией  $\mathcal{A}$  и случайной последовательностью  $\varepsilon_0, \varepsilon_1, \dots$ , элементы которой имеют нулевое среднее и одинаковую дисперсию  $\sigma^2$  [32, раздел 3.4.4]. В случае, если  $\mathcal{A}$  имеет порядок  $p \in \mathbb{N}$ , параметрами ряда  $F$ , кроме стандарта  $\sigma$ , будут коэффициенты  $a_1, \dots, a_p$  функции  $\mathcal{A}$ :

$$f_n = a_1 f_{n-1} + \dots + a_p f_{n-p} + \varepsilon_n.$$

Подобный подход используется в моделях типа “сигнал плюс шум” следующим образом: предполагается, что сигнал ряда задается линейной рекуррентной формулой (ЛРФ) порядка  $p$ :

$$f_n = s_n + \varepsilon_n, \quad s_n = a_1 s_{n-1} + \dots + a_p s_{n-p}. \quad (4.1)$$

Такая модель тоже имеет  $p$  неизвестных параметров  $a_1, \dots, a_p$  вдобавок к параметрам распределения шума.

**Постановка задачи** Данная глава посвящена исследованию методов получения оценок параметров модели вида (4.1), где сигнал представляет собой гармонический ряд:

$$f_n = A \sin(2\pi\omega n + \phi) + \varepsilon_n, \quad (4.2)$$

а последовательность  $(\varepsilon_0, \dots, \varepsilon_{N-1})$  является реализацией белого нормального шума с нулевым средним и единичной дисперсией. В рекуррентном представлении сигнал ряда (4.2) задаётся ЛРФ второго порядка (см. раздел 1.4), т. е.:

$$f_n = s_n + \varepsilon_n, \quad \text{где} \quad s_n = a_1 s_{n-1} + a_2 s_{n-2}. \quad (4.3)$$

Будем оценивать параметры  $a_1$ ,  $a_2$  и  $\omega$ .

**Замечание 4.1** Пользуясь выражением для корней характеристического полинома, соответствующего минимальной ЛРФ для  $\varepsilon$ -м гармонике (см. пример 1.1), несложно посчитать, что её экспоненциальный показатель  $\alpha$  выражается через коэффициенты ЛРФ  $a_1$ ,  $a_2$  следующим образом:

$$\alpha = \ln(\sqrt{-a_2}).$$

Следовательно, для рассматриваемого случая немодулированной гармонике ( $\alpha = 0$ ) выполняется  $a_2 = -1$ , и поэтому далее результаты по оценке  $a_2$  будут опущены.

Значение частоты  $\omega$  можно выразить через  $a_1$  и  $a_2$ , пользуясь утверждением 1.6:

$$\omega(a_1, a_2) = \frac{1}{2\pi} \arccos\left(\frac{a_1}{2\sqrt{-a_2}}\right), \quad \hat{\omega} = \omega(\hat{a}_1, \hat{a}_2). \quad (4.4)$$

Рассмотрим результаты оценок  $a_1$  и  $\omega$  двумя методами. Один из них использует подход “Гусеница”-SSA, а идея второго состоит в использовании метода наименьших квадратов (МНК) для расчёта коэффициентов линейной регрессии. Качество методов будем оценивать через близость оценки параметров к их реальным значениям. Сравним методы с помощью статистического моделирования.

## 4.1 Методы оценки коэффициентов ЛРФ

Первый метод, назовём его SSALRF, основан на подходе “Гусеница”-SSA. С его помощью оцениваются коэффициенты некоторой ЛРФ, управляющей сигналом ряда, рассматривается её характеристический полином и, используя значения “главных” корней этого полинома, строится минимальная ЛРФ (порядка 2). Во втором методе, назовём его REGR, рассмотрим стандартную модель



линейной регрессии, подставим вместо зависимых и независимых переменных значения ряда  $F$  таким образом, что они будут связаны соотношением (4.3), и посчитаем  $\hat{a}_1$  и  $\hat{a}_2$  с помощью формулы МНК для коэффициентов линейной регрессии.

В обоих методах оценка  $\hat{\omega}$  частоты  $\omega$  рассчитывается по полученным  $\hat{a}_1$ ,  $\hat{a}_2$  по формуле (4.4).

#### 4.1.1 Метод, основанный на подходе “Гусеница”-SSA

Используя введённые в главе 1 обозначения, сформулируем метод оценки коэффициентов ЛРФ порядка 2.

Для ряда  $F$  проведём его сингулярное разложение с заданным  $L$ , после чего, так как известно, что ЛРФ сигнала имеет порядок 2, возьмём два первых собственных вектора в качестве базиса траекторного пространства сигнала. Вычислим по ним коэффициенты ЛРФ порядка  $L - 1$  по (1.12), как это было описано в разделе 1.4.1. Выберем среди корней характеристического полинома главные корни, пользуясь сведениями из раздела 1.4.2 (пример 1.1). Найдём коэффициенты полинома степени 2, корнями которого они являются. Это и будут коэффициенты линейной рекуррентной формулы порядка 2, управляющей сигналом, т. е.  $\hat{a}_1$ ,  $\hat{a}_2$ .

Процедура SSALRF для оценки коэффициентов ЛРФ и частоты зашумлённой гармоники, использующая данный метод, описана в алгоритме 4.1.

Рассмотрим также модификацию метода SSALRF, которая использует идею, лежащую в основании итеративной процедуры Cadzow [31] и состоит из двух частей. Сначала с помощью подхода “Гусеница”-SSA выделим сигнал ряда  $F$ . После этого применим метод SSALRF, но уже к выделенному ряду. Выделять сигнал будем с помощью процедур SSADEC (с параметром  $L$ ), SSAREC (с параметром  $\mathcal{I} = \{1, 2\}$ ) (алгоритмы 1.1, 1.2).

**АЛГОРИТМ 4.1. Процедура SSA\_LRF: оценка коэффициентов ЛРФ и частоты зашумлённой гармоник**

**Входные данные:**  $F$  — ряд длины  $N$  ( $N > 2$ )

**Параметр:**  $L$  — длина окна ( $1 < L \leq \lceil N/2 \rceil$ )

**Результат:**  $\hat{a}_1, \hat{a}_2, \hat{\omega}$

**Процедура:**

1. Применим к  $F$  процедуру SSADEC (с параметром  $L$ ), получив собственные тройки  $\{(\sqrt{\lambda_k}, U_k, V_k)\}_{k=1}^d$  (алгоритм 1.1)
2. По собственным векторам  $U_1, U_2$  вычислим  $b_1, \dots, b_{L-1}$  — коэффициенты ЛРФ порядка  $L - 1$ , пользуясь (1.12)
3. Найдём  $\{z_j\}_{j=1}^{L+1}$  — корни полинома с коэффициентами  $1, -b_1, \dots, -b_{L-1}$ , см. (1.13)
4. Найдём  $z$  — один из двух главных корней:  $|z| = \max_{1 \leq j \leq L+1} |z_j|$
5.  $\hat{a}_1 = 2\Re z$ ,  $\hat{a}_2 = -\Re^2 z - \Im^2 z$ , где  $\Re z$ ,  $\Im z$  обозначают вещественную и комплексную части  $z$ , а  $\hat{\omega} = \omega(\hat{a}_1, \hat{a}_2)$ , см. (4.4)

### 4.1.2 Регрессионный метод

Рассмотрим стандартную модель линейной регрессии:  $Y = \mathbf{Z}^T \beta + \xi$ , где  $\mathbf{Z}$  — матрица независимых переменных,  $Y$  — вектор зависимых переменных,  $\xi$  — случайная ошибка, а  $\beta$  — вектор коэффициентов регрессии. Наложим стандартные ограничения:  $\mathbf{A} = \mathbf{Z}\mathbf{Z}^T$  — неособенная,  $\mathbf{E}\xi = 0$ ,  $\mathbf{D}\xi = \sigma^2 \mathbb{E}$ , где  $\mathbb{E}$  — единичная матрица. Обозначим оценку метода наименьших квадратов для коэффициентов линейной регрессии [17, раздел 5.2] через

$$\hat{\beta} = \mathbf{A}^{-1} \mathbf{Z} Y. \quad (4.5)$$

Определим  $Y$  таким образом, чтобы его элементами были значения  $f_2, \dots, f_{N-1}$  ряда  $F$ , обозначим коэффициенты регрессии как  $a_1, a_2$ , а элементы матрицы  $\mathbf{Z}$  определим таким образом, чтобы  $(n + 1)$ -ый элемент  $Y$  (т. е.  $f_n$ ) задавался следующим уравнением:

$$f_n = \begin{pmatrix} f_{n-1} & f_{n-2} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \varepsilon_n = a_1 f_{n-1} + a_2 f_{n-2} + \varepsilon_n.$$

Заметим, что при таких подстановках данная модель совпадает с авторегрессионной моделью для членов ряда  $F$ . Получим  $a_1, a_2$ , воспользовавшись формулой МНК для оценки коэффициентов линейной регрессии.

Метод REGR состоит в вычислении оценок  $\hat{a}_1$  и  $\hat{a}_2$  по формуле (4.5):  $(\hat{a}_1, \hat{a}_2)^T = \mathbf{A}^{-1} \mathbf{Z} \mathbf{Y}$ , где  $\mathbf{Z} = [Z_1 : \dots : Z_{N-1}]$ ,  $Z_i = (f_{i-1}, f_i)^T$ ,  $\mathbf{Y} = (f_2, \dots, f_{N-1})^T$ .

**Замечание 4.2** Метод REGR больше подходит для AR-модели, тогда как подход “Гусеница”-SSA, а, следовательно, и метод SSALRF, предназначены для обработки рядов модели “сигнал плюс шум”.

## 4.2 Сравнение методов с помощью моделирования

Сравним представленные методы с помощью статистического моделирования. Получим  $R$  реализаций  $\{F^{(r)}\}_{r=1}^R$  ряда согласно модели (4.2), для каждой реализации  $F^{(r)}$  вычислим с помощью методов SSALRF и REGR оценки для  $a_1, \omega$ :  $\hat{a}_1^{(r)}, \hat{\omega}^{(r)}$ . По полученным выборкам значений  $\{\hat{a}_1^{(r)}\}_{r=1}^R, \{\hat{\omega}^{(r)}\}_{r=1}^R$  рассчитаем точность оценок (абсолютное отклонение выборочного среднего от настоящего значения и выборочную дисперсию):

$$|\bar{\omega} - \omega|, \quad |\bar{a}_1 - a_1|, \quad \mathbf{D}_s \omega, \quad \mathbf{D}_s a_1, \quad \text{где}$$

$$\bar{\omega} = \frac{1}{R} \sum_{r=1}^R \hat{\omega}^{(r)}, \quad \bar{a}_1 = \frac{1}{R} \sum_{r=1}^R \hat{a}_1^{(r)},$$

$$\mathbf{D}_s \omega = \frac{1}{R} \sum_{r=1}^R (\hat{\omega}^{(r)})^2 - (\bar{\omega})^2, \quad \mathbf{D}_s a_1 = \frac{1}{R} \sum_{r=1}^R (\hat{a}_1^{(r)})^2 - (\bar{a}_1)^2.$$

### 4.2.1 Значения параметров

Сравним методы при обработке рядов различной длины  $N$  и для разных стандартов шума  $\sigma$ . Для получения более полной картины при одном фиксированном показателе (например, при фиксированном  $N$ ) другой изменялся с некоторым шагом в установленных пределах.

Отношение сигнал/шум в модели 4.2 определяют три показателя: амплитуда  $A$ , длина ряда  $N$  и стандарт шума  $\sigma$ . Зафиксируем  $A$  с тем, чтобы интерпретация результатов стала возможной,  $A = 1$ , и пусть  $\omega = 1/6$ .

Для моделирования используется  $R = 10^4$  реализаций.

В качестве длины окна  $L$  в методе SSALRF возьмём  $L = \lfloor N/2 \rfloor$ .

**Замечание 4.3** Рассматривались также другие значения для длины окна. Например, при  $L = 3$  траекторное пространство исходного ряда является подпространством  $\mathbb{R}^3$ , что соответствует ЛРФ порядка 2. Так как искомая ЛРФ имеет порядок 2, соответствующий характеристический полином не имеет лишних корней. Но при  $L = \lfloor N/2 \rfloor$  методы показали наилучшие результаты, видимо из-за асимптотического по  $\min(L, K)$  характера разделимости синуса и шума (см. [10, раздел 2.4], [41, раздел 6.1.3]).

**Замечание 4.4** Фаза  $\phi$  в рассмотренной модели зашумлённого гармонического ряда (4.2) не влияет на результаты оценки  $a_1$ ,  $a_1$  и  $\omega$ , поэтому при моделировании можно считать  $\phi = 0$ .

### 4.2.2 Результаты

**Качество модифицированного SSALRF** Исследование метода, являющегося модификацией метода SSALRF (в нем проводится сначала выделение сигнала и уже к нему применяется SSALRF) показало, что его качество при оценке  $a_1$  и  $\omega$  очень незначительно отличается от качества собственно метода SSALRF (результаты для  $\hat{\omega}$  при  $N = 60$  изображены на рис. 4.1). Принимая во внима-

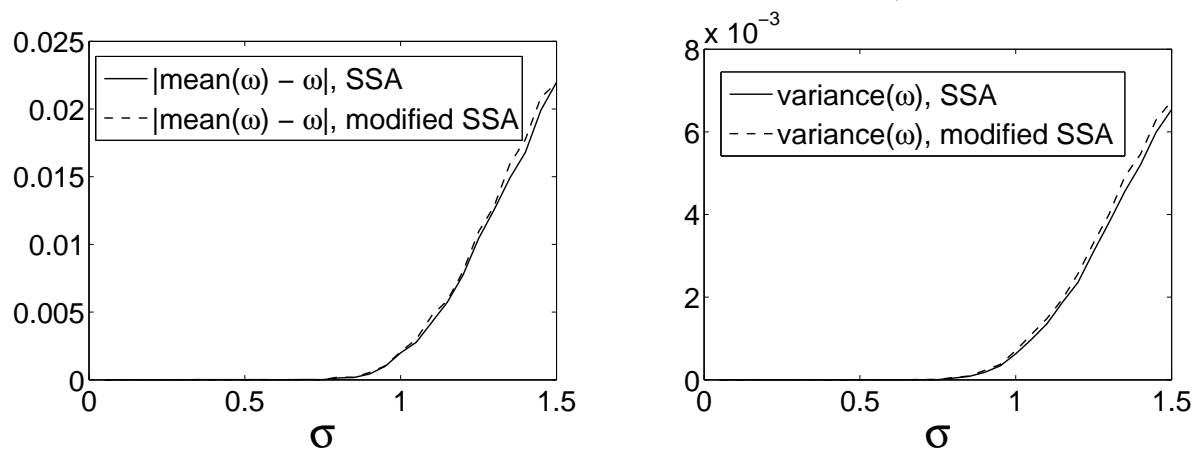


Рис. 4.1: Зависимость  $|\bar{\omega} - \omega|$ ,  $\mathbf{D}_s \omega$  от  $\sigma$  для методов SSALRF и модифицированного SSALRF (содержащего шаг восстановления сигнала).

ние то, что в первой части модифицированного метода совершаются трудоемкие вычисления, далее будем рассматривать только немодифицированный вариант метода SSALRF (алгоритм 4.1).

**SSALRF лучше REGR при не слишком большом уровне шума** Исследование, проведенное при различных  $N$  и  $\sigma$  показало, что нет однозначного ответа на вопрос, какой из методов лучше. Метод SSALRF лучше метода REGR

как по смещению, так и по дисперсии при достаточно малом шуме или при достаточно большой длине ряда.

Это видно на рис. 4.2, 4.3, на которых изображены результаты при  $N = 60$  ( $N = 10/\omega$ ), где  $\sigma$  изменяется с шагом 0.05 в интервале  $[0.05, 1.5]$  для  $\omega$  и в интервале  $[0.05, 1]$  для  $a_1$ . На этих графиках видно, что и смещение, и

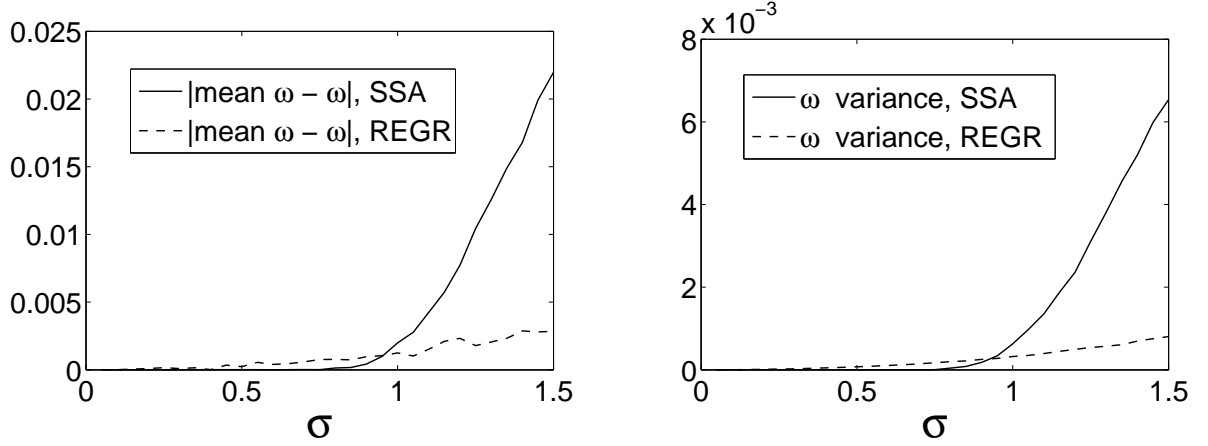


Рис. 4.2: Зависимость  $|\bar{\omega} - \omega|$ ,  $\mathbf{D}_s \omega$  от  $\sigma$  для методов SSALRF и REGR.

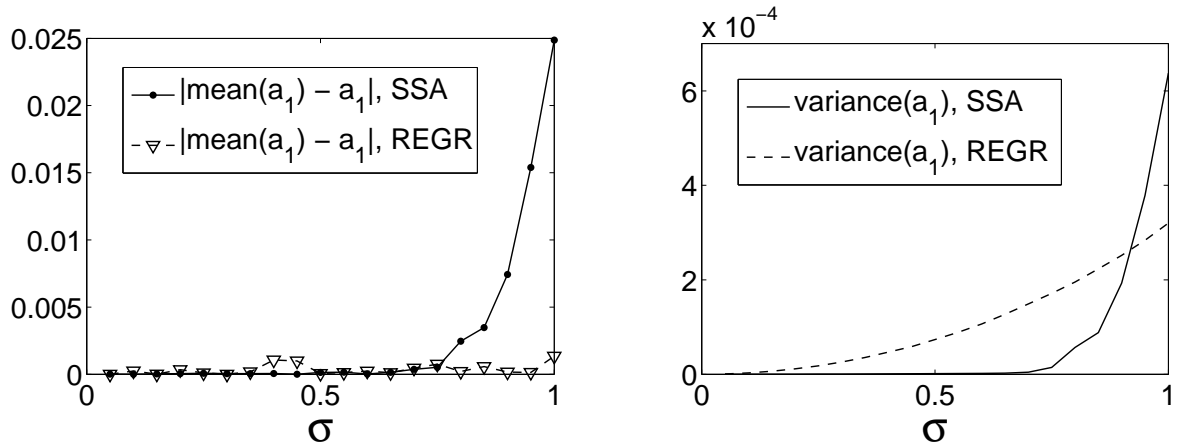


Рис. 4.3: Зависимость  $|\bar{a}_1 - a_1|$ ,  $\mathbf{D}_s a_1$  от  $\sigma$  для методов SSALRF и REGR.

дисперсия оценок для метода SSALRF до какого-то уровня  $\sigma$  меньше, чем для REGR. Это означает, что при таких  $\sigma$  метод SSALRF лучше, чем REGR. Заметим, что область, в которой SSALRF лучше, довольно велика:  $\sigma \lesssim 1$  для  $\hat{\omega}$  и  $\sigma \lesssim 0.8$  для  $\hat{a}_1$ . Для того чтобы продемонстрировать величину такого уровня шума, приведем на рис. 4.4 графики примеров зашумлённых гармонических рядов при  $N = 60$ ,  $\omega = 1/6$ ,  $A = 1$  и  $\sigma = 1$ .

Исследуем зависимость качества методов от  $N$  (длины ряда). Приведём результаты только для  $\hat{\omega}$ , которые проще интерпретируются, в отличие от результатов для  $\hat{a}_1$ . Графики на рис. 4.5 изображают отклонение и дисперсию оценки

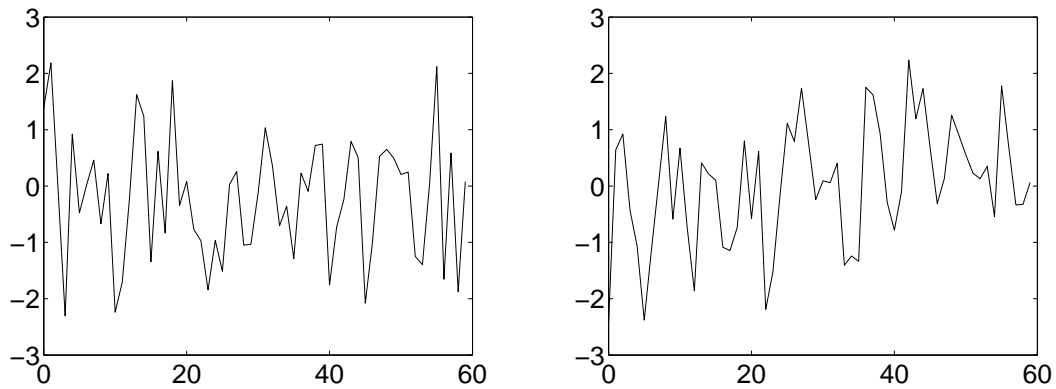


Рис. 4.4: Две реализации ряда (4.2) при  $N = 60$ ,  $\omega = 1/6$ ,  $A = 1$ ,  $\sigma = 1$ .

$\hat{\omega}$ , которые были получены при фиксированном  $\sigma = 1$  для  $N = 20 \dots 100$ , где  $N$  увеличивалось с шагом 5. Поскольку увеличение  $N$  соответствует уменьшению  $\sigma$  (в смысле влияния на результаты), то графики выглядят подобно приведенным выше графикам зависимости от  $\sigma$ , с точностью до замены направления оси абсцисс.

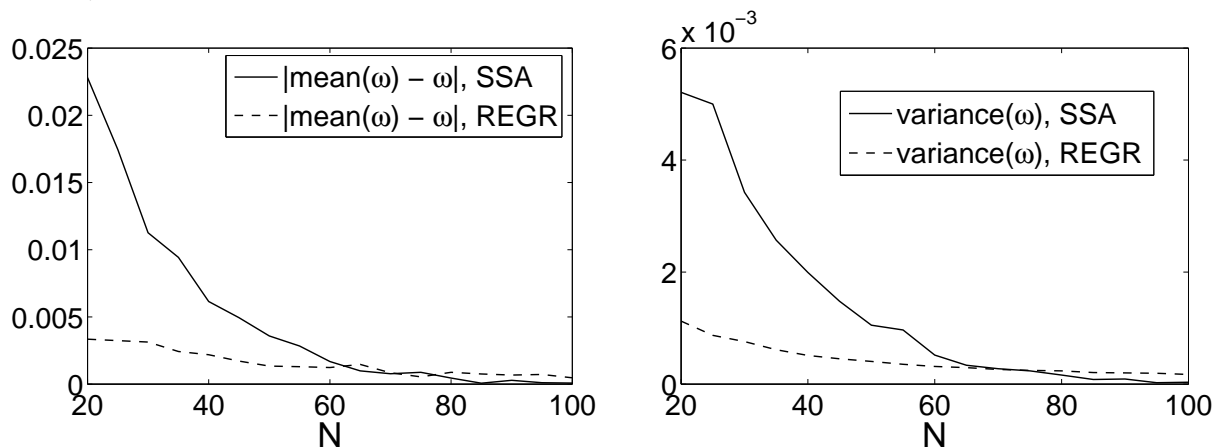


Рис. 4.5: Зависимость  $|\bar{\omega} - \omega|$ ,  $\mathbf{D}_s \omega$  от  $N$  для методов SSALRF и REGR.

**Замечание 4.5** При оценке с помощью метода SSALRF ожидалось, что его качество будет лучше при  $L$ :  $L\omega \in \mathbb{N}$  (см. описание условий отделимости гармоника в разделе 1.2). Однако при изменении  $N$  с шагом 5 бралось  $L = \lfloor N/2 \rfloor$  (т. е. значение  $L$  попадало и на плохие значения, когда  $L\omega \notin \mathbb{N}$ ) и результаты показали, что если такая зависимость и есть, то незначительная.

**Выводы** При не слишком большом уровне шума метод SSALRF, основанный на подходе “Гусеница”-SSA, продемонстрировал лучшее качество, чем метод REGR, использующий оценку МНК. Этот результат был зарегистрирован в достаточно большой области параметров, задающих отношение сигнал/шум. Это

хорошо характеризует метод SSALRF и подход “Гусеница”-SSA вообще, ведь он не настроен на работу с рядами одной модели. Единственная информация о ряде, которая использовалась в методе SSALRF —это его ранг, в то время как метод REGR представляет собой оценку параметров заданной модели.

Этой особенностью метода “Гусеница”-SSA объясняется и тот факт, что при достижении некоторого порога отношения сигнал/шум наблюдается резкое ухудшение качества метода SSALRF. При слишком большом уровне шума метод SSALRF выделяет сигнал, всё более непохожий на гармонику. Зависимость качества метода SSALRF от изменения длины ряда, описанная выше, также объясняется свойствами метода “Гусеница”-SSA.

## Заключение

В диссертационной работе рассмотрены задачи выделения и прогноза таких аддитивных компонент одномерного временного ряда как тренд и периодическая составляющая. Предложены новые математические методы их решения в рамках подхода “Тусеница”-SSA и построены алгоритмы, которые исследованы с помощью статистического моделирования. Данные алгоритмы управляются заданием параметров и могут быть использованы как для полностью автоматической, так и для автоматизированной обработки данных.

Исследование продемонстрировало, что качество предложенных алгоритмов во многих случаях близко к качеству существующих в рамках подхода “Тусеница”-SSA визуальных способов интерактивной обработки рядов.

**Метод выделения тренда** позволяет получать тренды различного разрешения при наличии в ряде как периодической составляющей, так и шума сложной структуры. Для улучшения качества результата может быть использована существующая информация о составляющих ряда. Алгоритм, использующий предложенный метод, выделяет тренд с хорошим качеством как в средней части ряда, так и на его концах.

**Метод выделения периодических составляющих** решает задачу выделения экспоненциально-модулированных (э-м) гармонических компонент ряда с неизвестными параметрами (уровнем модуляции и частотой). С его помощью можно выделять э-м гармонические компоненты ряда в присутствии тренда и шума сложной структуры. Для оценки параметров выделенной э-м гармонической разработаны новые методы. Алгоритм, построенный на основе данных методов, может быть использован для выделения э-м гармонической с известной или неизвестной частотой и для выделения периодической составляющей с известным периодом.

Статистическое исследование предложенного метода оценки частоты зашумлённой гармонической компоненты показало, что его качество при не слишком большом уровне шума лучше, чем качество другого метода, использующего формулу МНК для линейной регрессии.

На основе алгоритмов выделения компонент ряда построены алгоритмы прогноза этих компонент. Показано, что для выбора параметров алгоритмов выделения и прогноза могут быть использованы сходные принципы.

Все алгоритмы реализованы в виде комплекса программ. Представлены результаты обработки с их помощью временных и пространственных рядов.



**Кратко сформулируем выводы проделанной работы:**

- Разработаны математические методы автоматической идентификации собственных троек, соответствующих тренду или периодическим составляющим в рамках подхода “Гусеница”-SSA;
- Созданы и реализованы в виде программного комплекса алгоритмы выделения тренда и периодических составляющих на основе предложенных методов;
- Рассмотрены способы выбора параметров и пороговых значений алгоритмов;
- Проведено статистическое исследование качества работы алгоритмов при наилучших в среднем пороговых значениях;
- Построена математическая модель применения алгоритма выделения тренда к множеству рядов;
- Представлены результаты обработки временных и пространственных рядов с помощью предложенных алгоритмов;
- Выполнено статистическое сравнение подхода “Гусеница”-SSA и регрессионного метода при оценке коэффициентов ЛРФ порядка два для периодического сигнала.

Таким образом, в диссертации нашли своё отражение теоретические и методологические вопросы математического моделирования и разработки устойчивых алгоритмов, которые были реализованы в виде комплекса программ, пригодных для решения научных и практических задач.

# Библиография

- [1] **Айвазян, С. Ф.** Прикладная статистика. Классификация и снижение размерности / С. Ф. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин. — М.: Финансы и статистика, 1989. — 607 С.
- [2] **Александров, Ф. И.** Выделение аддитивных компонент временного ряда на основе метода “Гусеница”: дипломная работа / Александров Фёдор Игоревич, СПбГУ, 2003. — 37 С.
- [3] **Александров, Ф. И.** Оценка коэффициентов линейной рекуррентной формулы порядка 2, управляющей сигналом / Ф. И. Александров // Мат. модели. Теория и приложения. — 2004. — вып. 5. — С. 50–61. — ISBN 5-9651-0082-5.
- [4] **Александров, Ф. И.** Выделение аддитивных компонент временного ряда при пакетной обработке методом “Гусеница”-SSA / Ф. И. Александров // Вестник СПбГУ, Сер. 1. Математика. Механика. Астрономия. — 2006. — № 2. — С. 71–74. — ISSN 1025-3106.
- [5] **Александров, Ф. И.** Автоматизация выделения трендовых и периодических составляющих временного ряда в рамках метода “Гусеница”-SSA / Ф. И. Александров, Н. Э. Голяндина // Exponenta Pro. Математика в приложениях. — 2004. — вып. 3–4. — С. 54–61.
- [6] **Александров, Ф. И.** Выбор параметров при автоматическом выделении трендовых и периодических составляющих временного ряда в рамках подхода “Гусеница”-SSA / Ф. И. Александров, Н. Э. Голяндина, IV Международная конференция “Идентификация систем и задачи управления” SICPRO’05, Москва, январь 2005 // Труды IV Международной конференции “Идентификация систем и задачи управления”. — М.: 2005. — С. 1849–1864. — ISBN 5-201-14975-8.
- [7] **Андерсон, Т.** Статистический анализ временных рядов / Т. Андерсон. — М.: Мир, 1976. — 755 С.
- [8] **Белонин, М. Д.** Факторный анализ в нефтяной геологии / М. Д. Белонин, И. В. Татаринев, О. М. Калинин, В. К. Шиманский, О. В. Бескровная, В. В. Гранский, Т. Е. Похитонова — М.: ВИЭМС, 1971.
- [9] **Бриллинджер, Д.** Временные ряды. Обработка данных и теория / Д. Бриллинджер. — М.: Мир, 1980. — 536 С.

- [10] **Голяндина, Н. Э.** Метод “Гусеница”-SSA: анализ временных рядов: учеб. пособие / Н. Э. Голяндина. — СПб.: ВВМ, 2004. — 76 С. — ISBN 5-9651-0019-1.
- [11] **Голяндина, Н. Э.** Метод “Гусеница”-SSA: прогноз временных рядов: учеб. пособие / Н. Э. Голяндина. — СПб.: ВВМ, 2004. — 52 С. — ISBN 5-9651-0020-5.
- [12] **Голяндина, Н. Э.** Варианты метода “Гусеница”-SSA для анализа многомерных временных рядов / Н. Э. Голяндина, В. В. Некруткин, Д. В. Степанов, II Международная конференция “Идентификация систем и задачи управления” SICPRO’03, январь 2003 // Труды II Международной конференции “Идентификация систем и задачи управления” SICPRO’05. — М.: 2003. — С. 2139–2168.
- [13] **Голяндина, Н. Э.** Метод “Гусеница”-SSA для анализа временных рядов с пропусками / Н. Э. Голяндина, Е. В. Осипов // Мат. модели. Теория и приложения — 2005 — вып. 6. — С. 50–61.
- [14] **Градштейн, И. С.** Таблицы интегралов, сумм, рядов и произведений / И. С. Градштейн, И. М. Рыжик, изд-е 4-ое, переработанное при участии Ю. В. Геронимуса и М. Ю. Цейтлина. — М.: Физматгиз, 1963. — 1100 С.
- [15] **Данилов, Д. Л.** Главные компоненты временных рядов: метод “Гусеница” / Д. Л. Данилов, А. А. Жиглявский (ред.). — СПбГУ, изд-во Пресском, 1997. — 308 С.
- [16] **Дженкинс, Г.** Спектральный анализ и его приложения / Г. Дженкинс, Д. Ваттс. — М.: Мир, 1971. — 287 С.
- [17] **Ивченко, Г. И.** Математическая статистика: учеб. пособие для втузов / Г. И. Ивченко, Ю. И. Медведев. — М.: Высш. шк., 1984. — 248 С.
- [18] **Кислицин, М. М.** Многомерная статистика временных рядов наблюдений в авиационной эргономике / М. М. Кислицин // Вопросы кибернетики. Биотехнические системы в авиационной эргономике. — 1978. — вып. 51. — С. 117–126.
- [19] **Колмогоров, А. Н.** Об оценке параметров стационарного гауссовского марковского процесса / А. Н. Колмогоров, М. Арато, Я. Г. Синай. — ДАН СССР, 1962, т. 146 — С. 747–750.
- [20] **Некруткин, В. В.** Разложения временных рядов / В. В. Некруткин // Главные компоненты временных рядов: метод “Гусеница”: сб. науч. тр. / ред. Д. Л. Данилов, А. А. Жиглявский. — СПбГУ, 1997. — С. 194–227.

- [21] **Степанов, Д. В.** Варианты метода “Гусеница”-SSA для прогноза многомерных временных рядов / Д. В. Степанов, Н. Э. Голяндина, IV Международная конференция “Идентификация систем и задачи управления” SICPRO’05, январь 2005 // Труды IV Международной конференции “Идентификация систем и задачи управления” SICPRO’05. — М.: 2005. — С. 1831–1848.
- [22] **Alexandrov, F.** Automatic extraction and forecast of time series cyclic components within the framework of SSA / F. Alexandrov, N. Golyandina, 5th St.Petersburg Workshop on Simulation, St.-Petersburg, June 2005 // Proc. of the 5th St.Petersburg Workshop on Simulation. — SPb.: 2005. — P. 45–50. — ISBN 5-9651-0102-3.
- [23] **Alonso, F. J.** An Automatic Filtering Procedure for Processing Biomechanical Kinematic Signals / F. J. Alonso, J. María Del Castillo, P. Pintado // Eds. Barreiro J.M. et al.: ISBMDA 2004, Lecture notes in computer science, vol. 3337. — Springer-Verlag Berlin Heidelberg, 2004. — P. 281–291.
- [24] **Allen, M. R.** Interactions between the atmosphere and oceans on time scales of weeks to years: Ph.D. thesis, Univ. Of Oxford, / M. R. Allen — Oxford, England, 1992. — 202 P.
- [25] **Allen, M. R.** Monte Carlo SSA: detecting irregular oscillations in the presence of colored noise / M. R. Allen, L. A. Smith // J. Clim. — 1996. — vol. 9. — P. 3373–3404.
- [26] **Antoniou, I.** Principal Component Analysis of Network Traffic: the “Caterpillar”-SSA Approach / I. Antoniou, V. V. Ivanov, V. V. Ivanov, P. V. Zrelov // Abstracts of VIII Int. Workshop on “Advanced Computing and Analysis Techniques in Physics Research”, ACAT’2002. — Moscow, 2002. — P. 176.
- [27] **Baratta, D.** Application of an Ensemble Technique based on Singular Spectrum Analysis to Daily Rainfall Forecasting / D. Baratta, G. Cicioni, F. Masulli, L. Studer // Neural Networks. — 2003. — vol. 16, № 3-4. — P. 375–387.
- [28] **Bertero, M.** Resolution in diffraction-limited imaging, a singular value analysis I. The case of coherent illumination / M. Bertero, E. R. Pike // Optica Acta. — 1982. — vol. 29. — P. 727–746.
- [29] **Broomhead, D. S.** Extracting qualitative dynamics from experimental data / D. S. Broomhead, G. P. King // Physica D. — 1986 — vol. 20. — P. 217–236.

- [30] **Buchstaber, V. M.** Time Series Analysis and Grassmannians / V. M. Buchstaber // Amer. Math. Soc. Transl. — 1994. — vol. 2, № 162. — P. 1–17.
- [31] **Cadzow, J. A.** Signal Enhancement - A Composite Property Mapping Algorithm / J. A. Cadzow // IEEE Transactions on Acoustics, Speech and Signal Processing. — 1988. — v. 36. — P. 49–62.
- [32] **Chatfield, C.** The Analysis of Time Series: An Introduction / Christopher Chatfield. — 2nd ed. — Chapman&Hall, London, 1980 — 268 P.
- [33] **Colebrook, J. M.** Continuous plankton records — zooplankton and environment, northeast Atlantic and North Sea, 1948-1975 / J. M. Colebrook // Oceanol. Acta. — 1978. — vol. 1. — P. 9–23.
- [34] Economagic.com: Economic Time Series Page [Электронный ресурс]. — Режим доступа: <http://www.economagic.com/>.
- [35] **Elsner, J. B.** Singular Spectrum Analysis: A New Tool in Time Series Analysis / J. B. Elsner, A. A. Tsonis — Plenum, 1996. — 164 P.
- [36] **Findley, D. F.** Some Recent Developments and Directions in Seasonal Adjustment / D. F. Findley // Journal of Official Statistics. — 2005. — vol. 21, № 2. — P. 343–365.
- [37] **Fisher, R. A.** Tests of significance in Harmonic Analysis / R. A. Fisher // Proc. of the Royal Society, Ser. A. — 1929. — vol. 125. — P. 54–59.
- [38] **Fraedrich, K.** Estimating dimensions of Weather and Climate Attractors / K. Fraedrich // J. Atmos. Sci. — 1986. — vol. 43. — P. 419–432.
- [39] **Fukuda, K.** Detecting underlying time series structures and change points within a phenological data set using singular spectrum analysis (SSA) / K. Fukuda, I. Hudson, M. Keatley, XXII International Biometric Conference, Cairns, Australia, July 2004 // Abstracts of the XXII International Biometric Conference. — 2004.
- [40] **Ghil, M.** Advanced spectral methods for climatic time series / M. Ghil, R. M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. Robertson, A. Saunders, Y. Tian, F. Varadi, P. Yiou // Rev. Geophys. — 2002. — vol. 40, № 1. — P. 1–41.
- [41] **Golyandina, N.E.** Analysis of Time Series Structure: SSA and Related Techniques / N. E. Golyandina, V. V. Nekrutkin, A. A. Zhigljavsky — Chapman&Hall, 2001 — 305 P. — ISBN 1-58488-194-1.

- [42] **Golyandina, N. E.** The “Caterpillar”-SSA method for analysis of time series with missing values / N. E. Golyandina, E. V. Osipov // Submitted to Journal of Statistical Planning and Inference, 2006.
- [43] **Hodrick, R. J.** Postwar U.S. Business Cycles: An Empirical Investigation / Robert J. Hodrick, Edward C. Prescott // Journal of Money, Credit and Banking. — 1997. — vol. 29(1) — P. 1–16.
- [44] **Holloway, D. M.** Analysis of Pattern Precision Shows That Drosophila Segmentation Develops Substantial Independence From Gradients of Maternal Gene Products / D. M. Holloway, L. G. Harrison, D. Kosman, C. E. Vanario-Alonso, A. V. Spirov // Developmental Dynamics. — 2006. — vol. 235. — P. 2949–2960.
- [45] **Hsieh, W. W.** Nonlinear multichannel singular spectrum analysis of the tropical Pacific climate variability using a neural network approach / W. W. Hsieh, A. Wu // J. Geophys. Res. — 2002. — vol. 107, № C7. — P. 3076.
- [46] **Kaiser, R.**, Combining filter design with model-based filtering (with an application to business-cycle estimation) / R. Kaiser, A. Maravall // International Journal of Forecasting. — 2005. — vol. 21, № 4. — P. 691–710.
- [47] **Kondrashov, D.** Spatio-temporal filling of missing points in geophysical data sets / D. Kondrashov, M. Ghil // Nonl. Proc. Geophys. — 2006. — vol. 13. — P. 151–159.
- [48] **Kuiper, J.** A survey and comparative analysis of various methods of seasonal adjustment / J. Kuiper // Proc. of the NBER/Bureau of the Census Conference on Seasonal Analysis of Economic Time Series / Ed. A. Zellner. — Washington, D.C., 1976. — P. 59–76.
- [49] **Kumaresan, R.** Data-adaptive principal component signal processing / R. Kumaresan, D. W. Tufts // Proc. of IEEE Conference On Decision and Control. — Albuquerque, 1980. — P. 949–954.
- [50] **Makridakis, S.** Forecasting: Methods and Applications / S. Makridakis, S. Wheelwright, R. Hyndman, 3rd ed. — John Wiley & Sons, New York, 1998. — 656 P.
- [51] **Marple, S. L.** Digital spectral analysis: with applications / S. L. Marple. — Prentice-Hall, Inc., 1987. — 492 P.
- [52] **McElroy, T.** An iterated parametric approach to nonstationary signal extraction / T. McElroy, A. Sutcliffe // Computational Statistics & Data Analysis. — 2006. — vol. 50. — P. 2206–2231.

- [53] **Muratani, K.** Watermarking 3D Polygonal Meshes Using the Singular Spectrum Analysis / K. Muratani, K. Sugihara // Proc. of IMA Conference on the Mathematics of Surfaces — 2003 — P. 85–98.
- [54] **Packard, N. H.** Geometry from a time series / N. H. Packard, J. P. Crutchfield, J. D. Farmer, R. S. Shaw // Phys. Rev. Lett. — 1980. — vol. 45. — P. 712–716.
- [55] **Pike, E. R.** Generalized information theory for inverse problems in signal processing / E. R. Pike, J. G. McWhirter, M. Bertero, C. de Mol // IEE Proc. — 1984. — vol. 131. — P. 660–667.
- [56] **Planas, C.** The Analysis of Seasonality in Economic Statistics: A Survey of Recent Developments / C. Planas, EUROSTAT working group document. — 1997.
- [57] **Serita, A.** Principal component analysis and singular spectrum analysis of ULF geomagnetic data associated with earthquakes / A. Serita, K. Hattori, C. Yoshino, M. Hayakawa, N. Isezaki // Natural Hazards and Earth System Sciences. — 2005. — vol. 5. — P. 685–689.
- [58] **Schoellhamer, D. H.** Singular spectrum analysis for time series with missing data / D. H. Schoellhamer // Geophys. Res. Lett. — 2001. — vol. 28, № 16. — P. 3187–3190.
- [59] **Schuster, Sir A.** On the Investigation of Hidden Periodicities with Application to a Supposed Twenty-Six Day Period of Meteorological Phenomena / Sir A. Schuster // Terrestrial Magnetism. — 1898. — vol. 3. — P. 13–41.
- [60] SSA-MTM Toolkit [Электронный ресурс]. — Режим доступа: <http://www.atmos.ucla.edu/tcd/ssa/>.
- [61] **Stepanov, D. V.** SSA-based approaches to analysis and forecast of multidimensional time series / D. V. Stepanov, N. E. Golyandina // Proc. of the 5th St.Petersburg Workshop on Simulation, St.-Petersburg, June 2005 — SPb.: VVM com. Ltd., 2005. — P. 293–298.
- [62] **Stewart, G.** On the early history of the singular value decomposition [Электронный ресурс] / G. Stewart, Technical Report TR-92-31, University of Maryland, Institute for Advanced Computer Studies. — 1992. — Режим доступа: <http://citeseer.ist.psu.edu/stewart92early.html>.
- [63] **Takens, F.** Detecting strange attractors in turbulence F. Takens // Lecture Notes in Mathematics, eds. D. A. Rand, L.-S. Young — 1981. — vol. 206. — P. 366–381.
- [64] Time series analysis and forecasting, Caterpillar-SSA method [Электронный ресурс] / Caterpillar. Time series analysis and forecasting,

N. E. Golyandina, V. V. Nekrutkin, K. A. Braulov. — Режим доступа:  
<http://www.gistatgroup.com/>.

- [65] **Varadi, F.** Searching for signal in noise by Random-lag Singular Spectrum Analysis F. Varadi, J. M. Pap, R. K. Ulrich, L. Bertello, C. J. Henny // The Astrophysical Journal. — 1999. — vol. 526. — P. 1052–1061.
- [66] **Varadi, F.** Random-lag Singular Cross-Spectrum Analysis / F. Varadi, R. K Ulrich., L. Bertello, C. J. Henney // The Astrophysical Journal. — 2000. — vol. 528. — P. 53–56.
- [67] **Vautard, M.** Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series / M. Vautard, M. Ghil // Physica D. — 1989. — vol. 35. — P. 395–424.
- [68] **Vautard, R.** Singular-Spectrum Analysis: A toolkit for short, noisy chaotic signals R. Vautard, P. Yiou, M. Ghil // Physica D. — 1992. — vol. 58. — P. 95–126.
- [69] **Whitney, H.** Differentiable Manifolds / H. Whitney // Ann. Math. — 1936. — vol. 37. — P. 645–680.