УДК 519.246.8+ 517.518.454

ВЫБОР ПАРАМЕТРОВ ПРИ АВТОМАТИЧЕСКОМ ВЫДЕЛЕНИИ ТРЕНДОВЫХ И ПЕРИОДИЧЕСКИХ СОСТАВЛЯЮЩИХ ВРЕМЕННОГО РЯДА В РАМКАХ ПОДХОДА «ГУСЕНИЦА»-SSA

Ф.И. Александров
Санкт-Петербургский государственный университет,
Математико-механический факультет
Россия, 198504, Санкт-Петербург, Петродворец, Университетский пр., 28
E-mail: theo@pdmi.ras.ru

Н.Э. Голяндина
Санкт-Петербургский государственный университет,
Математико-механический факультет
Россия, 198504, Санкт-Петербург, Петродворец, Университетский пр., 28
E-mail: nina@ng1174.spb.edu

Ключевые слова: «Гусеница»-SSA, анализ сингулярного спектра, выделение тренда, выделение периодических компонент, идентификация компонент временного ряда, периодограмма
Key words: “Caterpillar”-SSA, Singular Spectrum Analysis, trend extraction, periodical components extraction, identification of time series components, periodogram

В данной работе рассматривается задача выделения тренда и периодической составляющей с помощью метода анализа временных рядов «Гусеница»-SSA. Исследуются методы автоматической идентификации, которые управляются заданием пороговых значений и позволяют автоматизировать процесс выделения. Изучается модель экспоненциального тренда и экпоненциально-модулированной гармоники, и для них с помощью средств статистического моделирования ищутся оптимальные пороговые значения параметров методов идентификации. Приводятся рекомендации по выбору пороговых значений в условиях отсутствия полной информации о параметрах модели ряда.
THRESHOLDS FOR METHODS OF AUTOMATIC EXTRACTION OF TIME SERIES TREND AND PERIODICAL COMPONENTS WITH THE HELP OF THE “CATERPILLAR”-SSA APPROACH / Th. Alexandrov (Mathematical Department, St.Petersburg State University, Universitetskij pr. 28, St.Petersburg Petrodvorets 198504, Russia, E-mail: theo@pdmi.ras.ru), N. Golyandina (Mathematical Department, St.Petersburg State University, Universitetskij pr. 28, St.Petersburg Petrodvorets 198504, Russia, E-mail: nina@ng1174.spb.edu). The problem of time series trend and periodical components extraction with the help of the “Caterpillar”-SSA approach is considered. Methods of automatic identification for automation of the extraction process are investigated. These methods are managed by thresholds setting. We work with models of exponential trend and exponential-modulated harmonic and use statistical simulation to find optimal thresholds values. Pieces of advice how to choose thresholds in the case of lack of information about time series model parameters are given.

1. Введение

В данной работе мы рассмотрим задачу выделения из временного ряда тренда, а также экспоненциально-модулированной гармоники с помощью подхода «Гусеница»-SSA. Этот подход зародился в 70х-80х годах прошлого столетия. В его основе лежит трансформация ряда в матрицу и ее сингулярное разложение. После идентификации компонент сингулярного разложения происходит их группировка, приводящая к разложению исходного ряда на аддитивные компоненты, такие как тренд, колебания (периодики) и шум. В зарубежной литературе метод наиболее известен под названием SSA (Singular Spectrum Analysis), он возник из теории динамических систем [1]. В России метод получил название «Гусеница» [2] и первоначально был основан на статистических аналогиях с методом главных компонент.

Достоинством метода «Гусеница»-SSA является отсутствие требования априорного знания модели ряда, но при этом сравнение этого метода с «модельными» методами показывает хорошие результаты. К преимуществам метода можно также отнести возможность работы с модулированными гармониками, что выгодно отличает его от методов, в основе которых лежит метод Фурье.

Ссылки на основную литературу по методу «Гусеница»-SSA можно найти в работах [2,3,4,5,6]. За время своего существования метод расширился, возникли его обобщения для анализа многомерных временных рядов, анализа изображений, поиска точек разладки в структуре временного ряда. Появились примеры его применения в широком круге областей: гидрологии, медицине, геофизике, экономике и пр.

Одним из направлений развития метода является автоматизация процедуры идентификации/группировки [5,6], так как используемый визуальный способ идентификации хоть и самый гибкий, но в ряде задач возникает необходимость в автоматизации процесса выделения компонент ряда, возможно с некоторой потерей качества. Однако применение автоматических методов переносит проблему с задачи интерактивной идентификации на задачу интерактивного выбора параметров. Поэтому целью данной статьи является выработка рекомендаций по выбору параметров автоматической идентификации на модельных примерах, имитирующих экспоненциальный тренд и 12-месячную (годовую) модулированную периодичность.

Безусловно, автоматическая идентификация должна опираться на описание некоторой модели компонент ряда. При выдаче рекомендаций мы старались не пользоваться точным заданием модели, а лишь использовать те характеристики ряда, которые могут быть приблизительно определены визуально, или часто бывают примерно известны для рассматриваемого класса рядов. Например, во сколько примерно раз увеличивается ряд/амплитуда периодики за рассматриваемый период.

Для автоматизации идентификации в данной работе использовались методы, основанные на периодограммном (частотном) анализе компонент разложения ряда.

Работа состоит из введения и двух разделов. В первом разделе мы коротко описываем алгоритм метода «Гусеница»-SSA и рассматриваемые методы автоматического выделения тренда и периодических компонент, построенные на его основе.

Во втором разделе приведены результаты численных экспериментов для определения оптимальных пороговых значений предлагаемых методов выделе-

ния тренда/периодики. Основным критерия оптимальности было среднеквадратическое отклонение восстановленного сигнала от истинного. В качестве модельных примеров рассматривались экспоненциальный тренд плюс нормальный белый шум, а также экспоненциально-модулированный синус плюс шум. Для нулевого шума существуют теоретические формулы для оптимальных значений параметров. Они также приведены в работе и проверено совпадение численных экспериментов по отношению к этим теоретическим значениям. Так как для использования оптимальных параметров необходимо точно знать модель ряда, что не всегда возможно, то в разделе даны рекомендации по выбору параметров в условиях неполного задания модели. Анализ поведения среднеквадратического отклонения относительно изменяющихся параметров методов идентификации показывает несимметричный характер ошибок относительно оптимального значения. Это означает, что, например, при идентификации гармоники лучше взять пороговое значение поменьше, с запасом. Чтобы показать, насколько отклонение от оптимальных значений увеличивает среднюю ошибку восстановления сигнала, приведены соответствующие таблицы.

2. Описания методов выделения тренда и периодических составляющих

2.1. Алгоритм «Гусеница»-SSA

Приведем вкратце алгоритм метода «Гусеница»-SSA (более подробно он описан в [3, разделы 1.1, 1.2], [7, раздел 1]). Рассмотрим вещественнозначный временной ряд $F_N = (f_0, \mathrm{K}, f_{N-1})$ длины N, $N > 2$.

Алгоритм можно разбить на четыре шага: вложение, сингулярное разложение, группировка и диагональное усреднение. Первые два в совокупности называются разложением, последние — восстановлением. Основным параметром алгоритма является так называемая длина окна L, $1 < L < N$. Результатом алгоритма является разбиение временного ряда на аддитивные составляющие.

2.1.1. Разложение. Первый шаг, вложение, состоит в формировании из ряда траекторной матрицы $\mathbf{X}$ размером $L \times K$, $K = N - L + 1$, следующим образом. Будем последовательно брать из ряда отрезки длины L и составим из них траекторную матрицу $\mathbf{X} = [X_1 : \mathrm{K} : X_K]$, где $X_j = \left(f_{j-1}, \mathrm{K}, f_{j+L-2}\right)^{\mathrm{T}}$. Далее проводится сингулярное разложение матрицы $\mathbf{X}$:

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \mathrm{K} + \mathbf{X}_d, \quad \mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}},$$

где $\lambda_1 \ge \lambda_2 \ge \mathrm{K} \ge \lambda_d > 0$ — упорядоченные ненулевые собственные числа матрицы $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, $\{U_i\}_{i=1}^d$: $U_i \in \mathbf{R}^L$ — соответствующие им собственные вектора, а $\{V_i\}_{i=1}^d$: $V_i = \lambda_i^{-1/2} \mathbf{X}^{\mathrm{T}} U_i \in \mathbf{R}^K$ будем называть факторными векторами.

2.1.2. Восстановление. На третьем шаге проводится группировка компонент разложения. Разбив $\{1, \mathrm{K}, d\}$ на m непересекающихся подмножеств I_j, получим

$$\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \mathrm{K} + \mathbf{X}_{I_m}, \quad \mathbf{X}_{I_j} = \sum_{k \in I_j} \mathbf{X}_k.$$

Последним шагом является восстановление рядов $F_N^{(j)}$ по сгруппированным матрицам $\mathbf{X}_{I_j}$. Элемент ряда $f_n^{(j)}$ получается с помощью усреднения вдоль антидиагонали элементов матрицы $\mathbf{X}_{I_j}$ с индексами a, b такими, что $a+b=n+2$. Таким образом, получаем разбиение ряда

$$F_N = F_N^{(1)} + \mathrm{K} + F_N^{(m)}.$$

Самым неформализуемым шагом является шаг группировки. Вся информация о каждой из компонент $\mathbf{X}_i$ содержится в собственном числе λ_i, а также в собственном U_i и факторном V_i векторах. Собственный и факторный вектора называют сингулярными векторами, а совокупность $\left(\sqrt{\lambda_i}, U_i, V_i\right)$ — собственной тройкой. Поиск компонент для требуемой группировки, главным образом на основе анализа собственных троек, будем называть процедурой идентификации.

Соответственно, для того, чтобы выделить какую-то составляющую ряда или отделить сигнал от шума, необходимо найти соответствующие искомой составляющей компоненты разложения, сгруппировать их и восстановлением получить искомый ряд.

Условия, при которых такое разложение и выделение возможно, а также принципы и теоретические постулаты, на которые опирается идентификация, описаны в работах [3, раздел 1.5], [6], [7, раздел 2].

2.2. Метод автоматической идентификации компонент, соответствующих тренду

Опишем вкратце методы автоматической идентификации сингулярных компонент для выделения тренда или экспоненциально-модулированной гармонической составляющей. Описанные далее методы применяются к сингулярным векторам и разбираются более подробно в работе [6].

В основание метода идентификации тренда положим следующую идею: сингулярные вектора компонент, соответствующих тренду, ведут себя подобно самому тренду (см. [3], [7, раздел 3.2]). Поэтому достаточно сформулировать метод в применении к произвольному ряду.

Метод низких частот основан на частотном представлении ряда. Для его изложения введем понятие периодограммы. Рассмотрим разложение Фурье вещественного временного ряда $G_M = (g_0, \mathrm{K}, g_{M-1})$:

$$g_n = c_0 + \sum_{1 \le k \le (M-1)/2} \left(c_k \cos(2n\pi k / M) + s_k \sin(2n\pi k / M)\right) + (-1)^n c_{M/2},$$

где $0 \le n \le M-1$, $k \in \mathbf{Z}$ и $c_{M/2} = 0$, если M — нечетное. Тогда периодограммой $\Pi_G^M(\omega)$ ряда G_M назовем функцию, определенную следующим образом при $\omega \in \{k/M\}_{k=0}^{\lfloor M/2 \rfloor}$:

$$\Pi_G^M(k/M) = \frac{M}{2} \begin{cases} 2c_0^2, & k = 0, \\ c_k^2 + s_k^2, & 1 \le k \le \dfrac{M-1}{2}, \\ 2c_{M/2}^2, & \text{если } M - \text{четное и } k = M/2. \end{cases}$$

Видно, что значение $\Pi_G^M(\omega)$ отражает вклад в разложение ряда G_M гармоники с частотой ω. Будем считать, что ряд является трендом, если гармониче-

ские составляющие с низкими частотами дают большой вклад в его разложение Фурье. Задав параметр ω_0, $0 < \omega_0 < 0.5$, будем считать областью низких частот интервал $[0, \omega_0]$. Посчитаем для ряда G_M отношение

$$\mathbf{C}(G_M) = \frac{\sum_{M\omega_0 < k \le M/2} \Pi_G^M(k/M)}{\sum_{0 \le k \le M/2} \Pi_G^M(k/M)}.$$

Величину $\mathbf{C}(G_M)$ можно интерпретировать как вклад гармоник со средними и высокими частотами в разложение Фурье последовательности g_0, K, g_{M-1}. Будем считать, что ряд G_M содержит трендовую составляющую, если $\mathbf{C}(G_M) \le \mathbf{C}_0$ для заданного порогового уровня $\mathbf{C}_0$.

2.3. Метод автоматической идентификации компонент гармоники

Метод Фурье для автоматической идентификации компонент, соответствующих экспоненциально модулированной (сокращенно — э.-м.) гармонической составляющей, тоже основан на анализе периодограмм сингулярных векторов [5,6]. Воспользуемся тем фактом, что э.-м. гармонике с частотой $\omega < 1/2$ соответствует две компоненты сингулярного разложения, сингулярные вектора которых имеют тоже э.-м. гармонический вид с той же частотой и экспоненциальным показателем (см. [3, раздел 1.6.1], [7, раздел 3.2.1]). Алгоритм метода Фурье можно поделить на две части.

2.3.1. Метод Фурье, часть 1. Воспользуемся тем, что периодограммы двух сингулярных векторов, соответствующих э.-м. гармонике, должны достигать максимальных значений на одной и той же частоте. Это и будем проверять. Пусть для рассматриваемой пары компонент с номерами i и $i+1$ θ_i и θ_{i+1} — аргументы максимумов периодограмм их сингулярных векторов. Пусть s_0, $s_0 \in \mathbf{Z}$ — пороговое значение метода. Если $M|\theta_i - \theta_{i+1}| \le s_0$, где M — длина сингулярного вектора, то будем считать, что пара $(i, i+1)$ соответствует э.-м. гармонике. Заметим, что θ_i является оценкой частоты найденной э.-м. гармоники. Поиск компоненты, соответствующей э.-м. гармонике с периодом 2, должен проводиться отдельно, так как ей соответствует одна компонента. В этом случае используется критерий $M|\theta_i - 0.5| \le s_0$.

2.3.2. Метод Фурье, часть 2. В первой части метода мы использовали только одно свойство периодограммы — аргумент ее максимума. Этого недостаточно, метод может ошибочно идентифицировать пары компонент, вовсе не соответствующие э.-м. гармонике. Учтем тот факт, что два гармонических сингулярных вектора (собственных или факторных), соответствующие гармонике, не только имеют такой же период, как и сама гармоника, но также имеют разницу в фазе, примерно равную $\pi/2$.

Зададим величину $\rho_{(a,b)}$, где a и b — номера двух сингулярных векторов $Y_a, Y_b \in \mathbf{R}^M$, формулой

$$\rho_{(a,b)} = \frac{1}{2} \max_{0 \le k \le M/2} \left(\Pi_{Y_a}^M(k/M) + \Pi_{Y_b}^M(k/M) \right).$$

Нетрудно увидеть, что если элементы векторов Y_a и Y_b образуют гармонические ряды с одной той же частотой ω и сдвигом фазы на $\pi/2$, а $M\omega$ — целое число, то $\rho_{(a,b)} = 1$.

Воспользуемся этим для усовершенствования метода Фурье. Рассмотрим пары компонент, уже идентифицированные в первой части метода, и будем считать, что пара компонент с номерами a и b соответствует гармонике, только если выполняется $\rho_{(a,b)} \geq \rho_0$, где $\rho_0 \in (0,1)$ — заранее заданное пороговое значение. Ясно, что чем больше ρ_0, тем строже условие. Похожим образом формулируется критерий и для гармоники с периодом 2.

Поскольку первая часть метода Фурье используется как подготовительная перед второй частью, можно зафиксировать значение s_0, установив его равным 1, что вполне достаточно для учета дискретности области определения периодограммы. Управление методом тогда будет совершаться только варьированием значения ρ_0.

3. Оптимальные пороговые значения для методов идентификации

Описанные выше методы автоматической идентификации требуют задания пороговых значений. Таким образом, задача интерактивной идентификации сводится теперь к задаче выбора пороговых значений. Целью данной работы является получение оптимальных пороговых значений.

Для того чтобы сосчитать оптимальное пороговое значение (а в реальности — выработать инструкции по обработке рядов определенного типа), необходимо задать модель исследуемого ряда. Подразумевается, что в ряде есть тренд и периодическая составляющая. Одним из простых случаев, тем не менее наблюдаемых в реальности, является наличие в ряде экспоненциального тренда и экспоненциально-модулированной гармонической составляющей. Будем рассматривать ряды $F_N = (f_0, \text{K}, f_{N-1})$, $N > 2$, следующего вида:

$$f_n = A_T e^{\alpha_T n} + \varepsilon_n, \quad f_n = A_H e^{\alpha_H n} \cos(2n\pi\omega) + \varepsilon_n, \quad n = 1, \text{K}, N,$$

где ε_n — нормальный белый шум с нулевым средним и дисперсией σ^2.

Для расчета оптимальных пороговых значений мы будем применять средства статистического моделирования. При проведении исследования необходимо ограничить параметры моделей, задав разумные диапазоны возможных значений. Будем пользоваться при этом такими характеристиками ряда, которые просто оценить или которые зачастую известны при исследовании ряда.

Во-первых, зафиксируем период гармоники, установив его равным 12 (что соответствует годовой периодичности для ежемесячных данных). Заметим, что соотношение сигнал/шум для рассматриваемого ряда можно менять тремя способами: изменяя экспоненциальные показатели, длину ряда, дисперсию шума. Зафиксируем одну из этих характеристик, длину ряда N. Это наиболее удобно для сравнения результатов, которые будут получены. Для достижения лучшей разделимости надо брать N таким, чтобы L и $K = N - L + 1$ делились бы на 12. Пусть $N = 119$, при этом в ряд укладывается 10 периодов гармоники.

Для того чтобы охватить ряды с различным соотношением сигнал/шум, будем проводить исследование при разных экспоненциальных показателях и дисперсиях шума. Экспоненциальные показатели тренда и гармоники будем задавать, основываясь на том, во сколько раз увеличивается ряд (амплитуда ряда для гармоники). Так, будем считать, что для экспоненциально-модулированной гармонической составляющей реального ряда еще приемлемым является увеличение амплитуды примерно в 50 раз за $N = 119,$ что примерно соответствует экспоненциальному показателю $\alpha = 0.033.$ Стандарт шума будем увеличивать до тех пор, пока не появятся слишком большие искажения в результатах, связанные с резким ухудшением качества разделимости сигнала и шума (см. [3, раздел 6.1.2] [7, раздел 2.4]).

3.1. Схема исследования

Исследование было построено следующим образом. Для заданного вида сигнала проводилась серия испытаний с разными пороговыми значениями критерия идентификации (C_0 для метода низких частот, ρ_0 — для метода Фурье). Для каждого фиксированного порогового значения R раз моделировался временной ряд в модели «сигнал плюс белый шум». Для каждой реализации временного ряда с помощью метода идентификации с текущим пороговым значением, примененного к собственным векторам сингулярного разложения траекторной матрицы ряда, строился восстановленный сигнал и вычислялся средний по времени квадрат отклонения его значений от истинных значений сигнала. Затем полученные значения усреднялись по реализациям, из результата извлекался корень, и тем самым мы получали оценку среднеквадратического отклонения (СКО). Оптимальным для заданных параметров модели ряда считалось пороговое значение, при котором достигалось минимальное значение СКО. Это можно интерпретировать следующим образом: при задании оптимального порогового значения выделенный сигнал будет в среднем наиболее близок к исходному.

Число компонент, соответствующих сигналу, в идеале должно равняться рангу сигнала, т.е. числу ненулевых компонент сингулярного разложения траекторной матрицы сигнала. В рассмотренных примерах ранг экспоненциального ряда равен 1, а ранг э.-м. гармонического ряда равен 2. Поэтому для контроля в каждой серии испытаний проводился расчет еще среднего количества идентифицированных компонент. Этот показатель показал хорошую согласованность со статистикой СКО. Было замечено, что минимум значений СКО достигался при количестве компонент, в среднем чуть большем ранга выделяемого сигнала, что объясняется несимметричной формой графика среднеквадратического отклонения.

3.2. Метод Фурье

3.2.1. Расчет оптимальных пороговых значений. Будем проводить выделение гармонического сигнала из ряда, являющегося зашумленной э.-м. гармоникой:

$$F_N : \ f_n = A_H e^{\alpha_H n} \cos(2n\pi\omega) + \varepsilon_n,$$

где $N = 119,\ A_H = 3,\ \omega = 1/12,\ \varepsilon_n$ — нормальный белый шум с нулевым средним и дисперсией σ^2.

Проиллюстрируем процесс поиска оптимального порогового значения. Приведем данные, полученные средствами статистического моделирования при 5000 повторах для $\alpha_H = 0.0136$ и $\sigma = 2$. Таблица 1 в строках содержит посчитанные для заданных ρ_0 оценки следующих характеристик: СКО и среднее количество компонент, идентифицированных как гармонические. В этой таблице ρ_0 изменяется с шагом 0.02, при расчетах же оптимальных пороговых значений для большей точности брался меньший шаг.

Таблица 1. Зависимость характеристик восстановления гармоники от порогового значения для $\alpha_H = 0.0136$ и $\sigma = 2$ при $R = 5000$ повторах

ρ_0	**Оценка СКО**	**Среднее число идентифицир. компонент**
0.8	0.5995	2.92
0.82	0.5774	2.74
0.84	0.5553	2.59
0.86	0.5344	2.46
0.88	0.5093	2.32
0.9	0.4881	2.21
0.92	0.7802	2.11
0.94	3.0434	1.54
0.96	5.6220	0.19

Пороговое значение, при котором достигается минимальное значение СКО, считается оптимальным (точнее, наилучшим из всех рассмотренных). Видно, что в данном случае это значение $\rho_0 = 0.9$, причем количество идентифицированных компонент близко к размерности сигнала, которая равна 2.

Вычислим оптимальные пороговые значения для разных экспоненциальных показателей α_H и стандартов шума σ. Будем рассматривать следующие значения экспоненциального показателя: 0, 0.005, 0.009, 0.0136, 0.02. Для объяснения того, почему были выбраны именно эти числа, приведем таблицу 2, приближенно показывающую для каждого значения, во сколько раз за 119 точек вырастает амплитуда ряда с таким показателем.

Таблица 2. Зависимость между экспоненциальным показателем и амплитудой при длине ряда равной 119

α_H	**0**	**0.005**	**0.009**	**0.0136**	**0.02**
Во сколько раз (прибл.) вырастает амплитуда	1	2	3	5	10

Таблица 3 показывает рассчитанные с помощью статистического моделирования оптимальные пороговые значения. Расчеты производились при длине окна $L = 60$ и на 5000 повторах.

Таблица 3. Оптимальные пороговые значения для метода Фурье

		α_H				
		0	**0.005**	**0.009**	**0.0136**	**0.02**
σ	**0**	1	0.992	0.976	0.948	0.895
	1	0.983	0.97	0.952	0.934	0.882
	2	0.96	0.95	0.934	0.91	0.86
	2.5	0.94	0.93	0.914	0.893	0.84
	3	0.917	0.913	0.9	0.87	0.823

3.2.2. Согласованность полученных результатов с теорией. С помощью непосредственных вычислений можно доказать следующее утверждение.

Предложение 3.1. Рассмотрим ряд $G_M: \ g_n = e^{\alpha n}\cos(2n\pi\omega)$. Пусть $\omega \in (0,0.5)$ и M такие, что $M\omega$ — целое. Если $M \to \infty$ и $\alpha \to 0$, причем $M\alpha \to \gamma > 0,$ то

$$\rho_{\{1,2\}} \to \frac{2}{\gamma}\frac{(e^{\gamma}-1)}{(e^{\gamma}+1)}.$$

Пользуясь этой формулой, рассчитаем ожидаемые оптимальные пороговые значения для метода Фурье в условии отсутствии шума, они приведены в таблице 4. Видно, что посчитанные численно при $\sigma = 0$ значения (первая строка таблицы 3) практически совпадают с теоретическими.

Таблица 4. Посчитанное с помощью предложения 3.1 ожидаемое оптимальное пороговое значение при отсутствии шума

α_H	**0**	**0.005**	**0.009**	**0.0136**	**0.02**
Оптимальное пороговое значение	1	0.99257	0.97639	0.94797	0.89508

Приведем рис. 1, отображающий изменение оптимальных пороговых значений с ростом σ для каждого из рассмотренных α_H. Видно, что соотношение между значениями, задаваемое предложением 3.1, в целом соблюдается и для ненулевого шума (на рисунке это соответствует одинаковым расстояниям между значениями при фиксированном стандарте).

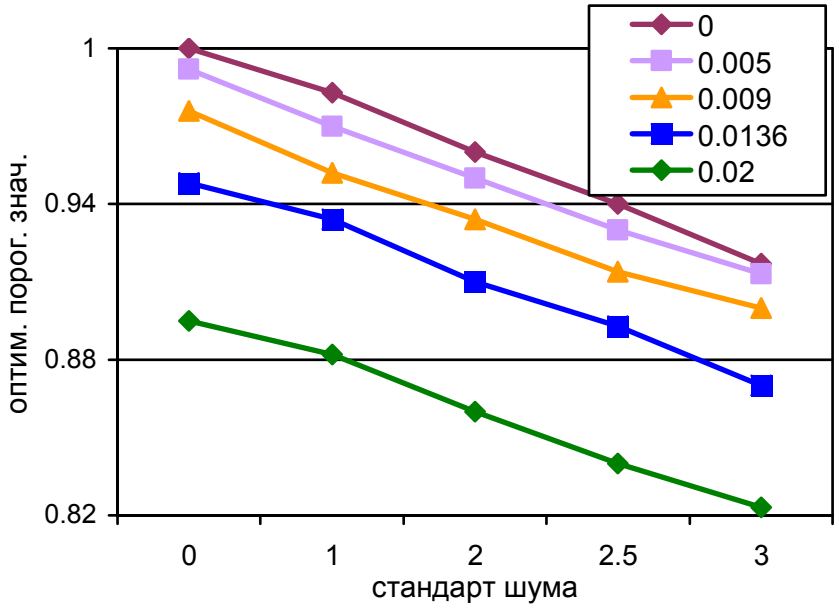

Рис. 1. Зависимость оптимальных пороговых значений метода Фурье от стандарта шума для различных экспоненциальных показателей

3.2.3. Качество выделения э-.м. гармоники. Приведем в таблице 5 для тех же α_H и σ значения минимальных оценок СКО (т.е. оценок СКО при оптимальных пороговых значениях ρ_0). Видно, что значения СКО даже при сравнительно большом шуме достаточно малы (среднеквадратическое отклонение исходного ряда от сигнала равно σ).

Таблица 5. Минимальные значения СКО

		α_H				
		0	**0.005**	**0.009**	**0.0136**	**0.02**
σ	**0**	0	0	0	0	0
	1	0.220	0.221	0.224	0.227	0.257
	2	0.447	0.451	0.471	0.465	0.553
	2.5	0.634	0.593	0.600	0.628	0.676
	3	0.779	0.753	0.729	0.769	0.878

3.2.4. Рекомендации по выбору порогового значения при отсутствии полной информации о параметрах модели ряда. В реальном исследовании параметры ряда, такие как экспоненциальный показатель или соотношение сигнал/шум, известны лишь приблизительно. В этом случае рассчитать оптимальное пороговое значение затруднительно и пороговое значение нужно выбирать с некоторым допуском.

Воспользуемся тем фактом, что функция СКО несимметрична относительно своего минимума, очень медленно изменяясь слева, что подтверждается рис. 2. Раз так, то мы можем выбирать пороговую точку левее, не слишком сильно ухудшая СКО. Это позволит нам в среднем без особого увеличения СКО идентифицировать большее количество компонент, что может быть удобно в случае, когда необходимо с большей надежностью идентифицировать компоненты гармоники, пусть даже ценой ошибочной идентификации негармонических компо-

нент. Сдвиг порогового значения вправо чреват тем, что гармоническая компонента не будет идентифицирована, за счет чего СКО в среднем резко увеличивается. Несимметричным поведением значений СКО объясняется тот факт, что минимум достигается при идентификации в среднем компонент количеством чуть больше ранга ряда (см. таблицу 1, в ней при $\rho_0 = 0.9$ идентифицировалось в среднем 2.21 компонент).

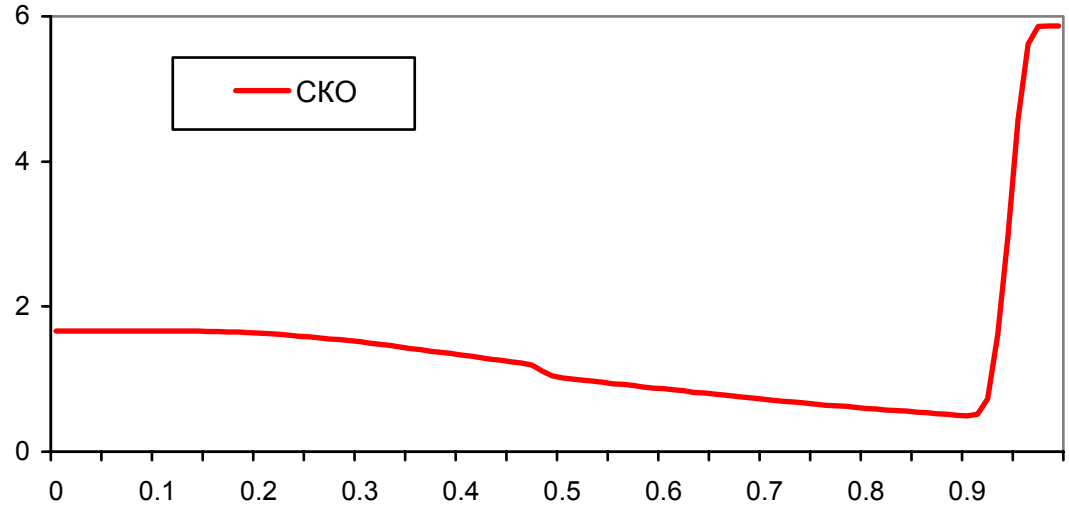

Рис. 2. Зависимость СКО от ρ_0 при $\alpha_T = 0.0136$ и $\sigma = 2$

Второе соображение, которое необходимо принимать во внимание при выборе порогового значения — это то, что с ростом экспоненциального показателя оптимальное пороговое значение уменьшается, что проиллюстрировано в таблице 3. Например, в случае, если известен интервал, в котором находится экспоненциальный показатель (или его значение известно с какой-то точностью), необходимо в качестве порогового выбирать значение, соответствующее оптимальному пороговому значению верхней границы интервала.

В целом, на основе проведенного исследования, а также применений метода к реальным временным рядам, можно принять значение 0.8 за пороговое значение по умолчанию для метода Фурье, так как при этом охватывается достаточное множество э.-м. гармонических рядов с различными экспоненциальными показателями при разном соотношении сигнал/шум. Таблица 6 содержит оценки СКО, полученные при $\rho_0 = 0.8$. Изучение среднего количества идентифицированных компонент показывает, что при этом даже в плохих случаях оно не будет превышать 3, то есть в среднем будет выделяться меньше одной лишней компоненты.

Таблица 6. Значения СКО, соответствующие $\rho_0 = 0.8$

		α_H				
		0	**0.005**	**0.009**	**0.0136**	**0.02**
σ	**0**	0	0	0	0	0
	1	0.30197	0.30015	0.30306	0.30041	0.301422
	2	0.6028	0.61348	0.60214	0.60412	0.606682
	2.5	0.74213	0.76158	0.75653	0.7477	0.748635
	3	0.89193	0.91595	0.90897	0.89697	0.898619

3.3. Метод низких частот

3.3.1. Расчет оптимальных пороговых значений. Проводилось исследование ряда

$$F_N : \ f_n = A_T e^{\alpha_T n} + \varepsilon_n,$$

где $N = 119, \ A_T = 3, \ \varepsilon_n$ — нормальный белый шум с нулевым средним и дисперсией σ^2.

Расчет проводился по той же схеме, что и для метода Фурье: при такой же длине окна $L = 60$ и параметре метода — границе низких частот $\omega_0 = 0.075$. Выбор такого ω_0 был сделан по следующим соображениям. Пусть мы работаем с реальным временным рядом с месячными данными, который содержит сезонную составляющую. Так как наибольший период гармоники, входящей в сезонную компоненту, равен 12, а гармоники не должны быть включены в тренд, значит, область низких частот $[0, \omega_0]$ обязана лежать левее $1/12 \cong 0.083$.

Приведем промежуточные результаты расчета оптимального порогового значения при 3000 повторах для $\alpha_T = 0.0136$ и $\sigma = 2$. В столбцах таблицы 7 стоят те же характеристики, что и в таблице 1 для метода Фурье. Наименьшее СКО достигается здесь при $\mathbf{C}_0 = 0.02,$ для которого среднее количество компонент, идентифицированных как соответствующих тренду, близко к размерности сигнала, равной в данном случае 1.

Таблица 7. Зависимость характеристик восстановления тренда от порогового значения для $\alpha_T = 0.0136$ и $\sigma = 2$ при $R = 3000$ повторах

$\mathbf{C}_0$	**Оценка СКО**	**Среднее число идентифицир. компонент**
0	8.1896	0
0.01	0.4813	1.04
0.02	0.3243	1.11
0.03	0.3357	1.19
0.04	0.3481	1.28
0.05	0.3577	1.37

Приведем таблицу 8, которая, как и таблица 3 для метода Фурье, показывает рассчитанные с помощью статистического моделирования оптимальные пороговые значения для разных значений экспоненциального показателя α_T и стандарта шума σ (количество повторов равно 3000). Во время исследований ре-

альных временных рядов было замечено, что экспоненциальный показатель тренда часто принимает значения большие, чем экспоненциальный показатель гармоники и значение 0.02 не является для него пределом, поэтому добавим к рассмотренным выше значениям показателя предельно возможное значение 0.05, которому соответствует возрастание значений ряда в 380 раз за 119 точек. Видно, что при таких больших значениях ряда шум оказывает минимальное влияние на результат.

Таблица 8. Оптимальные пороговые значения для метода низких частот

		α_T					
		0	**0.005**	**0.009**	**0.0136**	**0.02**	**0.05**
σ	**0**	0	0.001	0.0032	0.007	0.0145	0.0596
	1	0.001	0.003	0.005	0.009	0.0158	0.0598
	2	0.00285	0.005	0.007	0.01055	0.017	0.06
	2.5	0.0052	0.006	0.009	0.012	0.0182	0.06
	3	0.00645	0.0076	0.01	0.014	0.019	0.0602

3.3.2. Согласованность полученных результатов с теорией. Прямое вычисление коэффициентов разложения Фурье для экспоненциального ряда дает следующий результат.

Предложение 3.2. Для ряда G_M с $g_n = e^{\alpha n}$ значения периодограммы $\Pi_G^M(k/M)$ будут задаваться следующими коэффициентами:

$$c_k = C\left(\cos(2\pi k/M) - e^{-\alpha}\right), \qquad s_k = -C\sin(2\pi k/M),$$

$$\text{где } C = \frac{2}{M}\frac{e^{\alpha}(e^{\alpha M}-1)}{e^{2\alpha}+1-2e^{\alpha}\cos(2\pi k/M)},$$

$$c_0 = \frac{e^{\alpha M}-1}{M(e^{\alpha}-1)}, \qquad c_{M/2} = \frac{e^{\alpha M}-1}{M(-e^{\alpha}-1)}.$$

Пользуясь этим утверждением, можно сосчитать значения $\mathbf{C}(G_M)$ для ряда G_M : $g_n = e^{\alpha n}$, с известной длиной M при известном ω_0. Фиксирование M и ω_0 дает информацию о том, сколько точек решетки $\{k/M\}_{k=1}^{\lfloor M/2 \rfloor}$ попадет в интервал $[0, \omega_0]$. Посчитанные таким образом пороговые значения можно считать оптимальными при значении стандарта шума $\sigma = 0$. В таблице 9 приведены результаты для $M = 60$ (так как мы в методе низких частот считаем периодограммы собственных векторов, длина которых равна длине окна $L = 60$). Видно, что они совпадают со значениями, посчитанными на стадии статистического моделирования.

Таблица 9. Теоретические оптимальные пороговые значения при отсутствии шума

α_T	**0**	**0.005**	**0.009**	**0.0136**	**0.02**	**0.05**
Оптимальное пороговое значение	0	0.00098	0.00313	0.00694	0.01417	0.05953

Как и для метода Фурье, приведем рис. 3, отображающий изменение оптимальных пороговых значений для каждого из рассмотренных значений α_T (кроме 0.05) с ростом σ. Видно, что соотношение между значениями, задаваемое теоретически посчитанными значениями (см. таблицу 6), в целом соблюдается и для ненулевого шума.

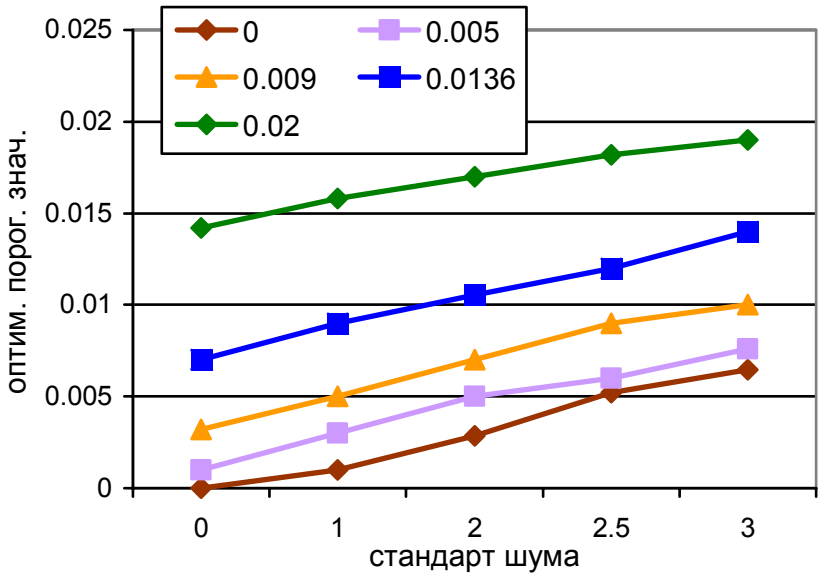

Рис. 3. Зависимость оптимальных пороговых значений метода низких частот от стандарта шума для различных экспоненциальных показателей

3.3.3. Качество выделения экспоненциального тренда. Приведем в таблице 10 для тех же α_T и σ минимальные значения СКО (т.е. оценки СКО при оптимальных пороговых значениях $\mathbf{C}_0$). Видно, что значения СКО даже при сравнительно большом шуме достаточно малы.

Таблица 10. Минимальные значения СКО

		α_T					
		0	**0.005**	**0.009**	**0.0136**	**0.02**	**0.05**
σ	**0**	0	0	0	0	0	0
	1	0.15449	0.15698	0.15527	0.1535	0.16033	0.18355
	2	0.3097	0.30937	0.31842	0.3106	0.31799	0.37421
	2.5	0.3914	0.39652	0.39621	0.39571	0.4008	0.45856
	3	0.46768	0.4677	0.47263	0.48228	0.48311	0.55446

3.3.4. Рекомендации по выбору порогового значения при отсутствии полной информации о параметрах модели ряда. Выбирая оптимальное пороговое значение для метода низких частот, нужно руководствоваться соображениями, подобными тем, которые мы принимали во внимание для метода Фурье. Во-первых, график значений СКО несимметричен относительно минимума, причем справа от него он возрастает очень медленно. Во-вторых, с ростом экспоненциального показателя растет оптимальное пороговое значение. Исходя из этого, можно рекомендовать в случае, когда экспоненциальный показатель α_T

не известен точно, брать в качестве порогового то значение, которое является оптимальным для наибольшего возможного значения α_T, так как для остальных возможных экспоненциальных показателей данное значение будет превышать их оптимальное и тем самым идентификация будет осуществляться устойчиво и с достаточно хорошим СКО.

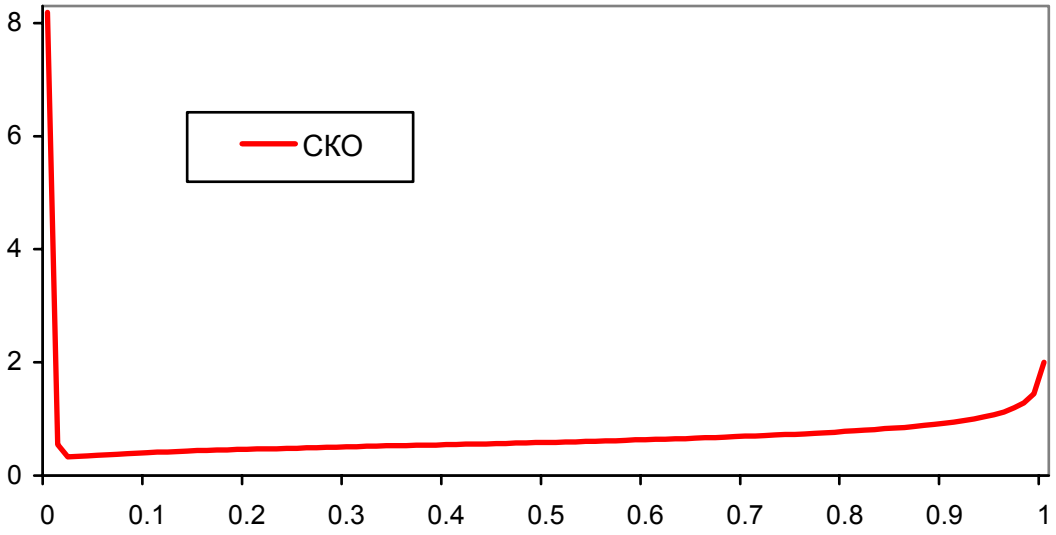

Рис. 3. Зависимость значений СКО от $\mathbf{C}_0$ при $\alpha_T = 0.0136$ и $\sigma = 2$

Исходя из приведенных оптимальных пороговых значений для рассматриваемых возможных экспоненциальных показателей и стандартов шума, можно считать 0.1 пороговым значением по умолчанию для метода низких частот.

Таблица 11 содержит оценки СКО, соответствующие выбранному пороговому значению по умолчанию, равному 0.1. Они достаточно невелики (и, как следовало ожидать, не сильно отличаются от величин СКО, достигаемых в оптимальных точках). Среднее количество компонент, идентифицированных при $C_0 = 0.1,$ не превышает 1.8, т.е. превышает ранг ряда в среднем меньше, чем на 1.

Таблица 11. Оценки СКО, соответствующие пороговому значению 0.1

		α_T					
		0	**0.005**	**0.009**	**0.0136**	**0.02**	**0.05**
σ	**0**	0	0	0	0	0	0
	1	0.19656	0.19828	0.20184	0.19782	0.2016	0.20313
	2	0.40455	0.40099	0.40037	0.40109	0.40395	0.412357
	2.5	0.51182	0.50031	0.49616	0.50321	0.5065	0.500372
	3	0.60826	0.5985	0.5962	0.60384	0.60801	0.600448

Список литературы

1. Broomhead D.S., King G.P. Extracting qualitative dynamics from experimental data // Physica D. 1986. Vol. 20. C. 217-236.
2. Главные компоненты временных рядов: метод «Гусеница» // Под. ред. Д.Л. Данилова, А.А. Жиглявского. Санкт-Петербург: Изд-во СПбГУ, 1997. 307 с. http://www.gistatgroup.com/gus/.
3. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. Boca Raton: Chapman & Hall/CRC, 2001. 305 p.
4. Elsner J., Tsonis A. Singular Spectrum Analysis. A New Tool in Time Series Analysis. New York: Plenum Press, 1996. 163 p.
5. Vautard R., Yiou P., Chil M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals // Physica D. 1992. Vol. 58. P. 95-126.
6. Выделение аддитивных компонент временного ряда на основе метода «Гусеница» http://www.pdmi.ras.ru/~theo/AutoSSA.html.
7. Голяндина Н.Э. Метод “Гусеница”-SSA: анализ временных рядов: Учеб. пособие. Санкт-Петербург: Изд-во СПбГУ, 2004. 76 с.