

Санкт-Петербургский государственный университет

Н. Э. Голяндина

Метод «Гусеница»-SSA:
прогноз временных рядов

У ч е б н о е п о с о б и е

Санкт-Петербург
2003

Санкт-Петербургский государственный университет

Н. Э. Голяндина

Метод «Гусеница»-SSA:
прогноз временных рядов

У ч е б н о е п о с о б и е

Санкт-Петербург
2003

УДК 519.2+681.3

Р е ц е н з е н т ы: доц. В.В. Некруткин (С.-Петерб. гос. ун-т),
с.н.с. В.Н. Солнцев (С.-Петерб. гос. ун-т)

*Печатается по постановлению
Редакционно-издательского совета
С.-Петербургского государственного университета*

Голяндина Н.Э.

Метод «Гусеница»-SSA: прогноз временных рядов: Учеб.
пособие. — СПб., 2003. — 55 с.

В основу учебного пособия положен курс лекций "Главные компоненты временных рядов", читаемый на кафедре статистического моделирования математико-механического факультете СПбГУ. Пособие содержит материалы части курса, посвященной прогнозу временных рядов с помощью метода «Гусеница»-SSA. Кроме описания теории метода приведены указания по его практическому применению, включая выбор параметров, а также подробно разобранный пример применения метода для прогноза реальных данных. Данная работа является продолжением учебного пособия "Метод «Гусеница»-SSA: анализ временных рядов"

Предназначено для студентов и аспирантов математико-механического факультета.

ISBN

© Н. Э. Голяндина, 2003
© С.-Петербургский
государственный
университет, 2003

Содержание

Введение	4
1 Базовый алгоритм метода «Гусеница»-SSA	7
2 Временные ряды конечного ранга, линейные пространства и управляющие линейные рекуррентные формулы	10
2.1 Ряды конечного ранга и ЛРФ	10
2.2 Линейные пространства и ЛРФ	13
2.3 Рекуррентное продолжение и L -продолжение	16
2.4 ЛРФ и характеристический полином	19
2.5 Главные и побочные корни характеристического полинома	22
3 Алгоритмы SSA-прогноза	24
3.1 Рекуррентное SSA-прогнозирование	24
3.2 Векторное SSA-прогнозирование	26
4 Доверительные интервалы	31
4.1 Эмпирические доверительные интервалы для прогноза исходного временного ряда	32
4.2 Бутстреп-доверительные интервалы для прогноза сигнала	34
5 Выводы и рекомендации	36
6 Приложение: пример прогнозирования временного ряда	44
6.1 Описание временного ряда линейной рекуррентной формулой	44
6.2 Прогноз и доверительные интервалы	47
Список литературы	55

Учебное пособие посвящено прогнозированию временных рядов на основе достаточно нового и быстро развивающегося метода анализа временных рядов, метода «Гусеница»-SSA (в русскоязычной литературе он появился под именем «Гусеница», а в англоязычной — под именем SSA, Singular Spectrum Analysis, отсюда и полное название метода). Описание метода, его применений и ссылки на литературу можно найти в книгах [1, 2]. Данное учебное пособие является продолжением учебного пособия [3] и существенно использует изложенный там материал. При изложении мы будем придерживаться обозначений и понятий книги [2], а также будем приводить без ссылок и доказательств содержащиеся там результаты.

Учебное пособие основано на курсе лекций, читаемых на математико-механическом факультете СПбГУ. Материал лекций используется на спецсеминаре, в рамках которого рассматривается практическое применение метода для прогноза временных рядов.

Введение

Прогнозирование, в отличие от анализа, возможно только при наличии математической модели временного ряда. Модель должна либо порождаться самими данными, либо, по крайней мере, тестироваться на имеющихся данных. Мы будем рассматривать модель, согласованную с методом «Гусеница»-SSA.

В методе «Гусеница»-SSA временному ряду $F_N = (f_0, \dots, f_{N-1})$ ставится в соответствие набор векторов, составленных из скользящих отрезков ряда выбранной длины L . Эти вектора порождают траекторное подпространство $\mathcal{L}_d \subset \mathbb{R}^L$ размерности $d \leq L$.

В работах [4, 5] содержатся результаты, связывающие подпространства $\mathbb{R}^L$ с линейными рекуррентными уравнениями и порождаемыми ими временными рядами. В [6, 7, 2] идеи работ [4, 5] доведены до алгоритмов продолжения и прогноза временных рядов.

Таким образом, при прогнозировании с помощью метода SSA мы рассматриваем множество временных рядов, описываемых с помощью линейных рекуррентных формул (ЛРФ). При этом по-

рядок ЛРФ может быть, вообще говоря, заранее неизвестен, поэтому класс рядов, управляемых такими ЛРФ, достаточно широк и важен для практических применений. Например, бесконечный временной ряд управляется ЛРФ тогда и только тогда, когда он является линейной комбинацией произведений экспонент, полиномов и гармоник.

Ряд, управляемый ЛРФ, естественным образом порождает *рекуррентное продолжение*, так как каждый его член равен линейной комбинации некоторого количества предыдущих. Поэтому коэффициенты этой линейной формулы (если они известны) могут быть использованы и для продолжения временного ряда. Важно отметить, что нам необязательно искать ЛРФ минимальной размерности, так как любая ЛРФ, управляющая рядом, приводит к одному и тому же продолжению. Тем самым, главной задачей является нахождение порядка и коэффициентов некоторой ЛРФ, управляющей рядом. Общая идея нахождения ЛРФ состоит в следующем.

Пусть d — минимальная размерность всех ЛРФ, управляющих рядом. Можно доказать, что если длина окна L больше, чем d , и длина временного ряда достаточно велика, то размерность траекторного пространства ряда F равняется d . Траекторное пространство порождает (при слабых и естественных ограничениях) ЛРФ размерности $L-1$, которая управляет временным рядом. Базовый метод SSA порождает естественный базис траекторного пространства, на основе которого конструктивно строится управляющая ЛРФ. Если мы применим эту ЛРФ к последним точкам ряда, то получим продолжение ряда F .

Та же самая идея работает и в случае продолжения аддитивной компоненты $F^{(1)}$ ряда $F = F^{(1)} + F^{(2)}$. Здесь мы предполагаем, что компонента $F^{(1)}$ управляется некоторой ЛРФ и является separable с остаточным рядом $F^{(2)} = F - F^{(1)}$ при выбранном значении длины окна L . Тогда она может быть выделена с помощью метода SSA и продолжена по соответствующей линейной рекуррентной формуле.

Предположим, что рассматриваемый ряд F может быть пред-

ставлен в виде суммы ряда $F^{(1)}$, допускающего рекуррентное продолжение, и остаточного ряда $F^{(2)}$. Если второй ряд является шумом, то мы приходим к задаче прогнозирования сигнала $F^{(1)}$ при наличии шума $F^{(2)}$. Также можно ставить задачу прогнозирования ряда $F^{(1)}$, если это тренд или сезонная компонента ряда F .

В реальных задачах условия, при которых ряд или компонента ряда допускает точное продолжение, как правило, не выполняются. В этом случае главным предположением для возможности прогнозирования является предположение о том, что при некоторой длине ряда компоненты ряда $F^{(1)}$ и $F^{(2)}$ приближенно сильно разделимы с помощью метода SSA. Тогда с помощью SSA мы можем приближенно восстановить ряд $F^{(1)}$ и получить тем самым аппроксимацию как самого ряда $F^{(1)}$, так и его траекторного пространства. Другими словами, мы получаем сразу и ЛРФ, приближенно управляющую рядом $F^{(1)}$, и начальные данные для этой формулы. Следовательно, мы имеем все, чтобы построить *приближенное рекуррентное продолжение* ряда $F^{(1)}$ (т. е. *прогноз*).

Остановимся кратко на содержании учебного пособия. В разделе 1 включен алгоритм метода «Гусеница»-SSA. Следующий раздел содержит описание рядов, которые могут быть точно продолжены с помощью информации, предоставляемой методом «Гусеница»-SSA, а также способ построения такого продолжения. В разделе 2 приведены два метода прогноза (приближенного продолжения) временных рядов на основе метода «Гусеница»-SSA. В четвертом разделе находится описание построения доверительных интервалов для точек прогноза. В разделе 6 приведены рекомендации по выбору параметров, проверке качества прогноза. Материалы этого раздела важны для практического применения метода и суммируют приведенные выше теоретические результаты. Чтобы продемонстрировать, как описанный метод может быть применен для прогнозирования реальных временных рядов, а также для того, чтобы проиллюстрировать теоретические результаты, в приложение к работе включен подробно разобранный пример прогноза временного ряда методом «Гусеница»-SSA.

1 Базовый алгоритм метода «Гусеница»-SSA

Для того, чтобы ввести обозначения, используемые в дальнейшем, приведем в сокращенном виде алгоритм базового метода SSA.

Пусть $N > 2$. Рассмотрим вещественнозначный временной ряд $F = (f_0, \dots, f_{N-1})$ длины N . Будем предполагать, что ряд F — ненулевой, т. е. существует по крайней мере одно i , такое что $f_i \neq 0$.

Базовый алгоритм состоит из двух дополняющих друг друга этапов, разложения и восстановления.

Первый этап: разложение

Шаг 1: Вложение

Процедура вложения переводит исходный временной ряд в последовательность многомерных векторов.

Пусть L — некоторое целое число (*длина окна*), $1 < L < N$. Процедура вложения образует $K = N - L + 1$ *векторов вложения*

$$X_i = (f_{i-1}, \dots, f_{i+L-2})^T, \quad 1 \leq i \leq K,$$

имеющих размерность L .

L -Траекторная матрица (или просто *траекторная матрица*) ряда F

$$\mathbf{X} = [X_1 : \dots : X_K]$$

состоит из векторов вложения как столбцов. Другими словами, траекторная матрица имеет вид

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{K-1} \\ f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \dots & f_{N-1} \end{pmatrix}.$$

Очевидно, что $x_{ij} = f_{i+j-2}$ и матрица $\mathbf{X}$ имеет одинаковые элементы на «диагоналях» $i + j = \text{const}$. (Таким образом, траекторная матрица является *ганкелевой*.) Существует взаимно-однозначное соответствие между ганкелевыми матрицами размерности $L \times K$ и рядами длины $N = L + K - 1$.

Шаг 2: Сингулярное разложение

Результатом этого шага является сингулярное разложение (SVD = Singular Value Decomposition) траекторной матрицы ряда.

Пусть $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. Обозначим $\lambda_1, \dots, \lambda_L$ *собственные числа* матрицы $\mathbf{S}$, взятые в неубывающем порядке ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$) и $U_1, \dots, U_L$ — ортонормированную систему *собственных векторов* матрицы $\mathbf{S}$, соответствующих собственным числам. Пусть $d = \max\{i, \text{таких что } \lambda_i > 0\}$.

Если обозначить $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$, $i = 1, \dots, d$, то сингулярное разложение матрицы $\mathbf{X}$ может быть записано как

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d, \quad (1)$$

где $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$. Каждая из матриц $\mathbf{X}_i$ имеет ранг 1. Поэтому их можно назвать *элементарными матрицами*. Набор $(\sqrt{\lambda_i}, U_i, V_i)$ мы будем называть *i-й собственной тройкой* сингулярного разложения (1).

Второй этап: восстановление

Шаг 3. Группировка

На основе разложения (1) процедура группировки делит все множество индексов $\{1, \dots, d\}$ на m непересекающихся подмножеств $I_1, \dots, I_m$.

Пусть $I = \{i_1, \dots, i_p\}$. Тогда *результатирующая матрица* $\mathbf{X}_I$, соответствующая группе I , определяется как

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}.$$

Такие матрицы вычисляются для $I = I_1, \dots, I_m$, тем самым разложение (1) может быть записано в сгруппированном виде

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}. \quad (2)$$

Процедура выбора множеств $I_1, \dots, I_m$ и называется *группировкой собственных троек*.

Шаг 4: Диагональное усреднение

На последнем шаге базового алгоритма каждая матрица сгруппированного разложения (2) переводится в новый ряд длины N .

Пусть $\mathbf{Y} — L \times K$ матрица с элементами y_{ij} , $1 \leq i \leq L$, $1 \leq j \leq K$.

Процедура *диагонального усреднения* переводит матрицу $\mathbf{Y}$ в ряд $(g_0, \dots, g_{N-1})$, сопоставляя каждой k -й диагонали матрицы (значениям y_{ij} с $i + j = k$) число g_{k-2} как среднее арифметическое соответствующих значений.

Заметим, что если матрица $\mathbf{Y}$ является траекторной матрицей некоторого ряда $(h_0, \dots, h_{N-1})$ (другими словами, если матрица $\mathbf{Y}$ является ганкелевой), то $g_i = h_i$ для всех i .

Применив диагональное усреднение к результирующим матрицам $\mathbf{X}_{I_k}$, мы получаем ряды $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$ и, следовательно, исходный ряд $(f_0, \dots, f_{N-1})$ раскладывается в сумму m рядов:

$$f_n = \sum_{k=1}^m \tilde{f}_n^{(k)}.$$

2 Временные ряды конечного ранга, линейные пространства и управляющие линейные рекуррентные формулы

В этом разделе мы опишем класс временных рядов, которые могут быть продолжены с помощью *SSA*-прогноза, и приведем результаты о связи между пространствами, порожденными самими рядами, и линейными рекуррентными формулами.

2.1 Ряды конечного ранга и ЛРФ

Рассмотрим временной ряд $F_N = (f_0, \dots, f_{N-1})$ с $N \geq 3$ и зафиксируем длину окна L ($1 < L < N$).

В результате процедуры вложения мы получаем последовательность векторов вложения:

$$X_i^{(L)} = X_i = (f_{i-1}, \dots, f_{i+L-2})^T, \quad i = 1, \dots, K,$$

$\mathfrak{L}^{(L)} = \mathfrak{L}^{(L)}(F_N) \stackrel{\text{def}}{=} \text{span}(X_1, \dots, X_K)$ — *траекторное пространство* ряда F_N .

Определение 2.1 Пусть $0 \leq d \leq L$. Если $\dim \mathfrak{L}^{(L)} = d$, то мы будем говорить, что *ряд F_N имеет L -ранг d* и записывать это как $\text{rank}_L(F_N) = d$. Для нулевого ряда положим $\dim \mathfrak{L}^{(L)} = 0$.

Очевидно, что равенство $\text{rank}_L(F_N) = d$ может быть справедливым только если

$$d \leq \min(L, K). \quad (3)$$

При фиксированном d будем называть L , для которого выполнено условие (3), допустимым. Заметим, что если длина окна, равная L , является допустимой, то и длина окна, равная $N - L + 1$, также допустима.

Определение 2.2 Если равенство $\text{rank}_L(F_N) = d < N/2$ имеет место для любого допустимого L , то будем говорить, что *ряд F_N имеет ранг d* ($\text{rank}(F_N) = d$).

Если такое d существует, то ряд F_N будем называть *рядом конечного ранга*.

Предложение 2.1 1.

$$\text{rank}_L(F_N) = \text{rank } \mathbf{X} = \text{rank } \mathbf{X}\mathbf{X}^T = \text{rank } \mathbf{X}^T\mathbf{X}.$$

2. Ортонормированная система собственных векторов $U_1, \dots, U_d$, соответствующих положительным собственным числам $\lambda_1 \geq \dots \geq \lambda_d$ матрицы $\mathbf{X}\mathbf{X}^T$, составляет ортонормированный базис пространства $\mathfrak{L}^{(L)}$.

3. Число $\text{rank}_L(F_N)$ равняется порядку сингулярного разложения траекторной матрицы $\mathbf{X}$.

Следующее утверждение показывает, что множество рядов конечного ранга достаточно большое.

Предложение 2.2 Любой ряд, являющийся линейной комбинацией произведений полиномов, экспонент и косинусов, является рядом конечного ранга.

Определение 2.3 Будем говорить, что временной ряд F_N имеет *конечно-разностную размерность не большую, чем d* ($\text{fdim}(F_N) \leq d$), если $1 \leq d < N - 1$ и найдутся числа $a_1, \dots, a_d$ такие, что

$$f_{i+d} = \sum_{k=1}^d a_k f_{i+d-k}, \quad 0 \leq i \leq N - d - 1, \quad a_d \neq 0. \quad (4)$$

Легко увидеть, что при $d < N - 2$ из неравенства $\text{fdim}(F_N) \leq d$ следует $\text{fdim}(F_N) \leq d + 1$.

Определение 2.4 Число $d = \min\{k : \text{fdim}(F_N) \leq k\}$ называется *конечно-разностной размерностью* (сокращенно — просто *размерностью* ряда F_N) и записывается $\text{fdim}(F_N) = d$. Для нулевого ряда F_N с $f_n \equiv 0$ положим $\text{fdim}(F_N) = 0$.

Как правило, мы будем предполагать, что $\text{fdim}(F_N) > 0$.

Определение 2.5 Формула (4) называется *линейной рекуррентной формулой (ЛРФ)*. ЛРФ (4) с $d = \text{fdim}(F_N)$ называется *минимальной* ЛРФ. Если (4) справедливо, то будем говорить, что ряд F_N управляется линейной рекуррентной формулой (4).

Заметим, что минимальная ЛРФ определяется однозначно.

Приведем примеры ЛРФ, управляющих некоторыми модельными рядами:

a) степенная последовательность $f_n = a^n$: $f_{i+1} = af_i$, $d = 1$;

b) экспоненциально-гармоническая последовательность

$$f_n = Ae^{\alpha n} \cos(2\pi\omega n + \phi),$$

где $\omega \in (0, 1/2)$, управляется ЛРФ

$$f_{i+2} = 2e^\alpha \cos(2\pi\omega) f_{i+1} - e^{2\alpha} f_i, \quad d = 2;$$

c) линейная последовательность $f_n = an + b$ с $a \neq 0$:

$$f_{i+2} = 2f_{i+1} - f_i, \quad d = 2;$$

d) квадратичная последовательность $f_n = n^2$:

$$f_{i+3} = 3f_{i+2} - 3f_{i+1} + f_i, \quad d = 3.$$

Следующее утверждение показывает эквивалентность класса рядов конечного ранга и рядов, управляемых ЛРФ, для рядов, бесконечных в обе стороны. Рассмотрим бесконечный в обе стороны ряд $F = (\dots, f_{-1}, f_0, f_1, \dots)$. Для него можно рассмотреть аналогичное определение ранга как ранга траекторной матрицы с бесконечным числом столбцов. Заметим, что для бесконечного ряда ограничения на i и d в определении конечно-разностной размерности ряда пропадают и можно уже говорить не о том, существует

или нет конечно-разностная размерность, а о том, является ли она конечной или бесконечной. Если размерность конечна, то существует ЛРФ, управляющая рядом; иначе такой линейной рекуррентной формулы не существует.

Предложение 2.3 *Любой бесконечный ряд является рядом конечного ранга d тогда и только тогда, когда он имеет размерность d .*

Отсюда, в частности, следует, что все приведенные выше примеры рекуррентных формул задают минимальные ЛРФ.

Заметим, что для временных рядов конечной длины ряд конечного ранга не обязательно имеет конечно-разностную размерность. В качестве контрпримера можно привести ряд с $f_n = 1$ для $0 \leq n \leq N - 2$ и $f_{N-1} = 2$. Очевидно, что $\text{rank}_L(F_N) = 2$ для $2 \leq L \leq N - 1$. В то же время, при $d < N - 1$ равенство (4) не может выполняться ни для какого набора коэффициентов a_k .

Следующая теорема (см. [4]; в данной формулировке содержится в [2]) показывает, что получить из ряда конечного ранга ряд, управляемый ЛРФ, можно, отбросив несколько первых и несколько последних элементов ряда. Обозначим $F_{i,j} = (f_{i-1}, \dots, f_{j-1})$.

Теорема 2.1 *Пусть $1 \leq \text{rank}_L(F_N) = d < L$. Тогда найдутся целые числа d_0 и M , такие что $0 \leq d_0 \leq d$, $0 \leq M \leq d - d_0$ и $\text{fdim}(F_{M+1, M+K+d_0}) = d_0$.*

2.2 Линейные пространства и ЛРФ

Пусть $\mathfrak{L}_d \subset \mathbf{R}^L$ — некоторое линейное подпространство размерности d . Для любого вектора $X \in \mathbf{R}^L$ будем обозначать $X^\nabla \in \mathbf{R}^{L-1}$ вектор, состоящий из первых $L - 1$ компонент вектора X . Обозначим $P_1, \dots, P_d$ некоторый ортонормированный базис пространства $\mathfrak{L}_d$ и рассмотрим линейное векторное пространство $\mathfrak{L}_d^\nabla$, натянутое на вектора $P_1^\nabla, \dots, P_d^\nabla$. Пусть π_i — последняя координата вектора P_i .

Предполагая, что $e_L = (0, \dots, 0, 1) \notin \mathfrak{L}_d$, получим неравенство $\nu^2 \stackrel{\text{def}}{=} \pi_1^2 + \dots + \pi_d^2 < 1$. Естественно назвать ν^2 *коэффициентом вертикальности* пространства $\mathfrak{L}_d$. Заметим, что коэффициент вертикальности равняется квадрату косинуса угла между вектором e_L и пространством $\mathfrak{L}_d$, а следовательно, эта характеристика не зависит от выбора базиса пространства.

Из $e_L \notin \mathfrak{L}_d$ следует, что $d < L$. Положим

$$\mathcal{R} = \frac{1}{1 - \nu^2} \sum_{i=1}^d \pi_i P_i^\nabla, \quad \mathcal{R} = (a_{L-1}, \dots, a_1)^\text{T}. \quad (5)$$

Теорема 2.2 Пусть $X = (x_1, \dots, x_L)^\text{T} \in \mathfrak{L}_d$ и $e_L \notin \mathfrak{L}_d$. Тогда $x_L = \sum_{k=1}^{L-1} a_k x_{L-k}$, где коэффициенты a_k определены в (5).

Следствие из этой теоремы дает нам вид одной из управляющих рядом ЛРФ размерности $L - 1$.

Следствие 2.1 Пусть $\mathfrak{L}_d = \mathfrak{L}^{(L)}(F_N)$ и $e_L \notin \mathfrak{L}_d$. Тогда $f_{i+L-1} = \sum_{k=1}^{L-1} a_k f_{i+L-1-k}$, $0 \leq i \leq K - 1$, где коэффициенты a_k определены в (5).

Таким образом, вектор $\mathcal{R}$ задает коэффициенты управляющей рядом ЛРФ. Посмотрим теперь на вектор $\mathcal{R}$ с другой точки зрения. Положим

$$\mathcal{Q} = (1 - \nu^2) \begin{pmatrix} -\mathcal{R} \\ 1 \end{pmatrix}.$$

Предложение 2.4 Вектор $\mathcal{Q}$ является ортогональной проекцией вектора e_L на $\mathfrak{L}_d^\perp$. Более того, $\|\mathcal{Q}\|^2 = 1 - \nu^2$ и $\|\mathcal{R}\|^2 = \nu^2 / (1 - \nu^2)$.

Заметим, что если $\nu^2 = 1$, то $\mathcal{Q} = \mathbf{0}_L$ (вектор $\mathcal{R}$ не определен в этом случае). Обратно, когда $\nu^2 = 0$, мы имеем $\mathcal{Q} = e_L$ и $\mathcal{R} = \mathbf{0}_{L-1}$.

Из предложения 2.3 следует, что ряд, управляемый ЛРФ размерности d , при $L \geq d$ порождает траекторное пространство размерности d . Также, в силу следствия 2.1, подпространство, являющееся траекторным и не содержащее единичного орта, порождает ряд, управляемый ЛРФ.

Предположим теперь, что $\mathfrak{L}_d \subset \mathbb{R}^L$ — некоторое произвольное линейное пространство размерности $d < L$ и $e_L \notin \mathfrak{L}_d$. Сформулируем вопрос следующим образом: «Существует ли ненулевой ряд такой, что вектора его L -вложения принадлежат $\mathfrak{L}_d$?» Следующее утверждение показывает, что для $d < L - 1$ ответ на этот вопрос, как правило, отрицателен.

Зададим пространство $\mathfrak{L}_d$ с помощью системы из $L - d$ линейных уравнений:

$$x_L + c_{j1}x_{L-1} + \dots + c_{j,L-2}x_2 + c_{j,L-1}x_1 = 0$$

($1 \leq j \leq L - d$), где вектора $(c_{j,L-1}, \dots, c_{j1}, 1)^T$ линейно независимы (это возможно в силу предположения $e_L \notin \mathfrak{L}_d$).

Предложение 2.5 *Бесконечный ненулевой ряд $\Phi = (\varphi_0, \varphi_1, \dots)$ такой, что вектора его L -вложения принадлежат $\mathfrak{L}_d$, существует тогда и только тогда, когда полиномы*

$$P_j(\lambda) = \lambda^{L-1} + c_{j1}\lambda^{L-2} + \dots + c_{j,L-2}\lambda + c_{j,L-1}$$

($j = 1, \dots, L - d$) имеют по крайней мере один общий корень.

Следствие 2.2 *Пусть $d = L - 1$ и $e_L \notin \mathfrak{L}_d$. Тогда ЛРФ*

$$\varphi_n = a_{11}\varphi_{n-1} + \dots + a_{1,L-2}\varphi_{n-L+2} + a_{1,L-1}\varphi_{n-L+1}, \quad n \geq L,$$

с $a_{1k} = -c_{1k}$ порождает все ряды, чьи вектора L -вложения принадлежат $\mathfrak{L}_d$.

Таким образом, если $d < L - 1$, то по крайней мере два полинома с произвольными коэффициентами должны иметь общий корень, что является вырожденной ситуацией.

2.3 Рекуррентное продолжение и L -продолжение

Если временной ряд F_N управляется ЛРФ (4) размерности $d < N$, то существует естественное *рекуррентное продолжение* такого ряда, порожденное этой же формулой (4).

Переформулируем теперь понятие рекуррентного продолжения на геометрическом языке. Начнем с определения.

Рассмотрим временной ряд $F_N = (f_0, \dots, f_{N-1})$ и зафиксируем длину окна $1 < L < N$. Обозначим $X_1, \dots, X_K$ соответствующие вектора L -вложения и положим $\mathfrak{L}^{(L)} = \text{span}(X_1, \dots, X_K)$. Пусть $d = \dim \mathfrak{L}^{(L)}$ (другими словами, L -ранг ряда F_N равняется d). Очевидно, что $d \leq \min(L, K)$.

Определение 2.6 Будем говорить, что ряд F_N *допускает продолжение в подпространстве $\mathfrak{L}^{(L)}$* (или, кратко, *допускает L -продолжение*), если существует единственное число $\tilde{f}_N$ такое, что все вектора L -вложения ряда $\tilde{F}_{N+1} = (f_0, \dots, f_{N-1}, \tilde{f}_N)$ принадлежат $\mathfrak{L}^{(L)}$. В этом случае, ряд $\tilde{F}_{N+1}$ (также, как и само число $\tilde{f}_N$) мы будем называть *одношаговым L -продолжением* ряда F_N .

Предложение 2.6 1. Если $e_L \in \mathfrak{L}^{(L)}$, то F_N не допускает L -продолжения. В частности, если $d = L$, то временной ряд не может быть L -продолжен, так как условие единственности в этом случае нарушается.

2. Если $d < L \leq K$ и $e_L \notin \mathfrak{L}^{(L)}$, то ряд F_N допускает L -продолжение.

Начиная с этого момента мы будем предполагать, что условия пункта 2 предложения 2.6 относительно подпространства $\mathfrak{L}^{(L)}$ выполнены.

Предложение 2.7 1. Одношаговое L -продолжение ряда F_N может быть выполнено по формуле

$$\tilde{f}_N = \sum_{k=1}^{L-1} a_k f_{N-k}, \quad (6)$$

где вектор $\mathcal{R} = (a_{L-1}, \dots, a_1)^T$ определен в формуле (5) с $\mathfrak{L}_d = \mathfrak{L}^{(L)}$.

2. Ряд F_N управляется той же самой ЛРФ (6), т. е.

$$f_{i+L-1} = \sum_{k=1}^{L-1} a_k f_{i+L-1-k}, \quad 0 \leq i \leq N-L.$$

3. Если ряд F_N допускает одношаговое L -продолжение, то он может быть L -продолжен на любое число шагов вперед. Следовательно, существует бесконечный ряд, являющийся L -продолжением ряда F_N .

4. Пусть временной ряд F_N удовлетворяет ЛРФ

$$f_{i+d_0} = \sum_{k=1}^{d_0} b_k f_{i+d_0-k}, \quad 0 \leq i \leq N-d_0-1, \quad (7)$$

и $d_0 \leq \min(L-1, K)$. Тогда $d \leq d_0$, $e_L \notin \mathfrak{L}^{(L)}$ и ряд допускает L -продолжение, которое может быть произведено с помощью той же формулы (7).

Таким образом, из сформулированных выше предложений следует, что при выполнении условий $\text{rank}_L(F_N) < L \leq K$ и $e_L \notin \mathfrak{L}^{(L)}$ понятия рекуррентной продолжимости и L -продолжимости совпадают.

Несмотря на то, что в определении L -продолжимости ряда F_N L -ранга d все элементы этого ряда формально присутствуют, продолженное значение $\tilde{f}_N$ определяется (при выполнении условий $e_L \notin \mathfrak{L}_d$ и $d < L \leq K$) только последним вектором L -вложения X_K . Более точно, вектор вложения $\tilde{X}_{K+1}$ L -продолженного ряда $\tilde{F}_{N+1}$ имеет вид

$$\tilde{X}_{K+1} = \mathbf{D}_L X_K, \quad (8)$$

где

$$\mathbf{D}_L = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & a_{L-1} & a_{L-2} & \dots & a_1 \end{pmatrix} \quad (9)$$

и вектор $\mathcal{R} = (a_{L-1}, \dots, a_1)$ определен в (5). По определению, $\tilde{X}_{K+1} \in \mathfrak{L}_d$. Также заметим, что вектор $\tilde{X}_{K+1}^\nabla$, состоящий из первых $L-1$ компонент вектора $\tilde{X}_{K+1}$, совпадает с вектором $(X_K)_\Delta$, состоящим из последних $L-1$ компонент вектора X_K .

Рассмотрим некоторый вектор $X \in \mathfrak{L}_d$, по-прежнему предполагая, что $e_L \notin \mathfrak{L}_d$ (до сих пор мы рассматривали не произвольные вектора из $\mathfrak{L}_d$, а только вектора вложения). Поставим задачу нахождения вектора $Y \in \mathfrak{L}_d$ такого, что расстояние между векторами X_Δ и Y^∇ минимально. Заметим, что $Y = \mathbf{D}_L X$ тогда и только тогда, когда $X_\Delta \in \mathfrak{L}_d^\nabla$, где пространство $\mathfrak{L}_d^\nabla$ состоит из всех векторов Z^∇ с $Z \in \mathfrak{L}_d$. В этом случае $Y^\nabla = X_\Delta$, и последняя компонента y вектора Y имеет форму $y = (\mathcal{R}, X_\Delta)$. Следующее предложение дает решение этой задачи в общем случае.

Предложение 2.8 Пусть $\mathbf{P}$ — оператор, соответствующий ортогональной проекции $\mathbf{R}^{L-1} \mapsto \mathfrak{L}_d^\nabla$. Тогда $Y^\nabla = \mathbf{P}X_\Delta$ и $y = (\mathcal{R}, Y^\nabla)$.

Как уже обсуждалось выше, если $e_L \notin \mathfrak{L}_d$, $d < \min(L, K)$ и вектор $X = X_K$ совпадает с последним столбцом L -траекторной матрицы временного ряда F_N , то вектор X_Δ принадлежит $\mathfrak{L}_d^\nabla$ и $\mathbf{P}X_\Delta = X_\Delta$. Следовательно, предложение 2.8 описывает то же самое продолжение ряда F_N с помощью формул (8) и (9). Заметим, что для $d = L-1$ оператор $\mathbf{P}$ является тождественным (равенство $\mathbf{P}X_\Delta = X_\Delta$ справедливо для любого X).

Предложение 2.9 Пусть $d = L-1$, $e_L \notin \mathfrak{L}_{L-1}$ и $X \in \mathfrak{L}_{L-1}$. Тогда $\mathbf{D}_L X \in \mathfrak{L}_{L-1}$.

Обсудим теперь свойства оператора Π . Пусть $P_1, \dots, P_d$ — ортонормированный базис пространства $\mathfrak{L}_d$. Положим $\mathbf{V}^\nabla = [P_1^\nabla : \dots : P_d^\nabla]$ и $W = (\pi_1, \dots, \pi_d)^\top$, где π_i — последняя компонента вектора P_i . Как и прежде, пусть $\nu^2 = \pi_1^2 + \dots + \pi_d^2$. Введем новую матрицу $\mathbf{A} = (\mathbf{V}^\nabla)^\top \mathbf{V}^\nabla$.

Предложение 2.10 *Если $e_L \notin \mathfrak{L}_d$, то справедливо следующее.*

1. Матрица Π линейного проекционного оператора Π имеет форму

$$\Pi = \mathbf{V}^\nabla (\mathbf{V}^\nabla)^\top + (1 - \nu^2) \mathcal{R} \mathcal{R}^\top.$$

2. В обозначениях предложения 2.8 мы имеем равенства $y = \mathcal{R}^\top X_\Delta$ и

$$Y^\nabla = \Pi X_\Delta = \mathbf{V}^\nabla (\mathbf{V}^\nabla)^\top X_\Delta + y(1 - \nu^2) \mathcal{R}.$$

2.4 ЛРФ и характеристический полином

В разделе 2.2 был продемонстрирован метод нахождения коэффициентов ЛРФ. Однако представляет также интерес описание поведения ряда, управляемого полученной ЛРФ. Покажем, как это поведение может быть выражено в терминах корней характеристического полинома.

Пусть ряд $F = (f_0, \dots, f_n, \dots)$ управляется ЛРФ (4) с $a_d \neq 0$ и $i \geq 0$. Рассмотрим *характеристический полином* линейной рекуррентной формулы (4):

$$P_d(\lambda) = \lambda^d - \sum_{k=1}^d a_k \lambda^{d-k}$$

и обозначим k_m ($1 \leq m \leq p$) кратности его (в общем случае, комплексных) несовпадающих корней $\lambda_1, \dots, \lambda_p$ ($k_1 + \dots + k_p = d$). Заметим, что так как $a_d \neq 0$, все корни λ_m отличны от нуля.

Теорема 2.3 *Вещественнозначный ряд F удовлетворяет ЛРФ (4) тогда и только тогда, когда*

$$f_n = \sum_{m=1}^p \left(\sum_{j=0}^{k_m-1} c_{mj} n^j \right) \lambda_m^n, \quad (10)$$

где комплексные коэффициенты c_{mj} зависят от первых d элементов ряда $f_0, \dots, f_{d-1}$.

Замечание 2.1 Если корень λ_m вещественен, то его вклад в сумму (10) имеет форму полинома по n , умноженного на λ_m^n (т. е. на экспоненту). Более того, соответствующие коэффициенты c_{mj} также вещественны. Если корень λ_m является комплексным, то характеристический полином имеет также сопряженный к нему комплексный корень той же кратности. Переписав один из этих корней как $\rho e^{i2\pi\omega}$ с $\omega \in (0, 1/2)$, мы получим, что совместный вклад обоих корней в сумму (10) имеет вид полинома по n , умноженного на $\rho^n \cos(2\pi\omega n + \phi)$. Более того, в этом случае соответствующие коэффициенты c_{mj} и c_{lj} также являются сопряженными комплексными числами. В обоих случаях, если кратность корня λ_m равна 1, то полином по n становится константой.

Таким образом, любой бесконечный ряд, удовлетворяющий ЛРФ (4) для любого i , обязательно является суммой произведений экспонент, полиномов и гармоник.

Рассмотрим подробно случаи $d = 1$ и $d = 2$.

$d = 1$ В этом случае у характеристического полинома $p_1(\lambda) = \lambda + b_1$ может быть только один вещественный корень λ_1 и $f_n = c\lambda_1^n$. Обозначим $\alpha = \ln(|\lambda_1|)$. Если $\lambda_1 > 1$, то ряд F с $f_n = ce^{\alpha n}$, является возрастающей экспонентой. При $0 < \lambda_1 < 1$ ряд F с $f_n = ce^{\alpha n}$ — убывающая экспонента. При $\lambda_1 = 1$ члены ряда F равны одной и той же константе: $f_n = c$. При $\lambda_1 < 0$ ряд F с $f_n = c(-1)^n e^{\alpha n}$ имеет пилообразный вид. Заметим, что выражение $(-1)^n$ можно трактовать как косинус с периодом, равным двум.

- $d = 2$ — **Два различных вещественных корня.** Этот случай сводится к случаю с $d = 1$, и ряд F является суммой двух экспонент с различными показателями.
- **Один вещественный корень кратности два.** Здесь $f_n = (an + b)\lambda_1^n$. Интерес представляет случай $\lambda_1 = 1$, когда F является линейной последовательностью.
- **Два комплексно-сопряженных корня.** Если два корня являются комплексно-сопряженными, то их можно представить в виде $\lambda_1 = a + ib$ и $\lambda_2 = a - ib$, т. е. оба корня задаются точкой (a, b) на комплексной плоскости. Записав эту точку в полярных координатах, мы получим $\lambda_1 = \rho e^{i2\pi\omega}$ и $\lambda_2 = \rho e^{-i2\pi\omega}$, $\omega \in (0, 1/2)$, где ρ — модуль корня, а ω — полярный угол. Оба эти корня порождают два слагаемых в (10), которые при суммировании приводят к ряду с общим членом

$$f_n = A\rho^n \cos(2\pi\omega n + \phi),$$

причем параметры A и ϕ определяются начальными данными, к которым применяется ЛРФ. При $\rho = 1$ мы получаем чисто гармоническую последовательность. В этом случае можно сказать, что соответствующие комплексно-сопряженные корни полинома лежат на единичной окружности (в координатах комплексной плоскости).

Асимптотическое поведение компоненты ряда, соответствующей вещественному корню (или паре комплексно-сопряженных корней), существенно зависит от $\rho = |\lambda_m|$. Продемонстрируем влияние отклонения ρ от единицы на простейшем случае $d = 1$, что соответствует вещественному корню кратности 1. Тогда, если к примеру $\lambda_1 = \rho = 0.8$ и $|c_{10}| \neq 0$, то $|f_n|$ за 10 шагов уменьшается примерно в 10 раз, а за 100 шагов — в $5 \cdot 10^9$ раз. Наоборот, если $\lambda_m = \rho = 1.2$ (и $|c_{m0}| \neq 0$), то $|f_n|$ возрастает в 6 раз за 10 шагов и в $8 \cdot 10^7$ раз за 100 шагов.

2.5 Главные и побочные корни характеристического полинома

Если ряд F_N имеет размерность, равную d , то характеристический полином его минимальной линейной рекуррентной формулы имеет ровно d корней. Но один и тот же ряд может подчиняться многим другим ЛРФ порядка $r > d$. Рассмотрим такую ЛРФ

$$f_{r+i} = b_1 f_{r+i-1} + b_2 f_{r+i-2} + \dots + b_r f_i. \quad (11)$$

Характеристический полином $P_r(\lambda)$ ЛРФ (11) имеет r корней, d из которых (мы будем называть их *главными корнями*) совпадают с корнями минимальной ЛРФ. Другие $r - d$ корней являются *побочными*: в силу единственности представления (10) коэффициенты c_{mj} , соответствующие этим корням, равны нулю. Однако сама ЛРФ (11) управляет более широким классом рядов по сравнению с минимальной.

Так как корни характеристического полинома определяют его коэффициенты однозначно, то тем самым корни полинома задают сам полином. Следовательно, удалив (обнулив) все побочные корни характеристического полинома, мы получим минимальную ЛРФ временного ряда.

Пример 2.1 Годовая сезонность

Пусть ряд F_N имеет период 12 (например, этот ряд описывает некоторое сезонное поведение). Тогда его можно представить в виде суммы константы и шести гармоник:

$$f_n = c_0 + \sum_{k=1}^5 c_k \cos(2\pi nk/12 + \phi_k) + c_6 \cos(\pi n). \quad (12)$$

При условии, что $c_k \neq 0$ для $k = 0, \dots, 6$, ряд имеет размерность 12. Другими словами, характеристический полином ряда (12) имеет 12 корней. Все эти корни имеют единичные модули. Два

вещественных корня (+1 и -1) соответствуют первому и последнему членам представления (12). Гармоническая компонента с частотой $\omega_k = k/12$ порождает два комплексно-сопряженных корня $\exp(\pm i2\pi k/12)$, соответствующих полярному углу $\pm 2\pi k/12$.

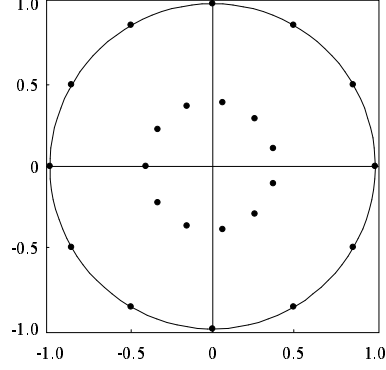


Рис. 1: Годовая сезонность: главные и побочные корни.

Рассмотрим теперь ЛРФ, которая не является минимальной. Пусть длина ряда N достаточно большая. Если мы выберем некоторую длину окна $L > 13$ и возьмем $r = 12$, $\mathfrak{L}_r = \mathfrak{L}^{(L)}(F_N)$, то вектор $\mathcal{R} = (a_{L-1}, \dots, a_1)^T$, определенный в (5), порождает ЛРФ

$$f_{i+L-1} = a_1 f_{i+L-2} + \dots + a_{L-1} f_i, \quad (13)$$

которая не минимальна, но которая в то же время управляет рядом (12).

Возьмем $c_0 = \dots = c_6 = 1$, $\phi_1 = \dots = \phi_5 = 0$ и $L = 24$. Корни характеристического полинома формулы (13) изображены на рис. 1. Можно заметить, что главные 12 корней полинома образуют правильный 12-угольник с вершинами, лежащими на единичном круге комплексной плоскости. Одиннадцать побочных корней расположены внутри единичного круга и невелики по модулю.

3 Алгоритмы SSA-прогноза

3.1 Рекуррентное SSA-прогнозирование

Опишем формально алгоритм рекуррентного прогнозирования. При этом для замкнутости алгоритма мы повторим некоторые определения.

Входные данные алгоритма:

- (a) Временной ряд $F_N = (f_0, \dots, f_{N-1})$, $N > 2$.
- (b) Длина окна L , $1 < L < N$.
- (c) Линейное пространство $\mathfrak{L}_r \subset \mathbf{R}^L$ размерности $r < L$. Предполагается, что $e_L \notin \mathfrak{L}_r$, где $e_L = (0, 0, \dots, 0, 1)^T \in \mathbf{R}^L$. Другими словами, $\mathfrak{L}_r$ не является «вертикальным» пространством. Обычно пространство $\mathfrak{L}_r$ задается некоторым ортонормированным базисом, но результаты прогнозирования не зависят от вида конкретного базиса.
- (d) Число M точек прогноза.

Обозначения и комментарии:

- (a) $\mathbf{X} = [X_1 : \dots : X_K]$ (где $K = N - L + 1$) — траекторная матрица временного ряда F_N .
- (b) $P_1, \dots, P_r$ — ортонормированный базис пространства $\mathfrak{L}_r$.
- (c) $\hat{\mathbf{X}} \stackrel{\text{def}}{=} [\hat{X}_1 : \dots : \hat{X}_K] = \sum_{i=1}^r P_i P_i^T \mathbf{X}$. Вектор $\hat{X}_i$ является ортогональной проекцией X_i на пространство $\mathfrak{L}_r$.
- (d) $\tilde{\mathbf{X}} = \mathcal{H} \hat{\mathbf{X}} = [\tilde{X}_1 : \dots : \tilde{X}_K]$ — результат ганкелизации матрицы $\hat{\mathbf{X}}$. Матрица $\tilde{\mathbf{X}}$ — траекторная матрица некоторого ряда $\tilde{F}_N = (\tilde{f}_0, \dots, \tilde{f}_{N-1})$.
- (e) Для любого вектора $Y \in \mathbf{R}^L$ обозначим $Y_\Delta \in \mathbf{R}^{L-1}$ вектор, состоящий из последних $L - 1$ компонент вектора Y , а $Y^\nabla \in \mathbf{R}^{L-1}$ — вектор, состоящий из первых $L - 1$ компонент вектора Y .
- (f) Положим $\nu^2 = \pi_1^2 + \dots + \pi_r^2$, где π_i — это последняя компонента вектора P_i ($i = 1, \dots, r$). Так как ν^2 равняется квадрату косинуса угла между вектором e_L и линейным пространством $\mathfrak{L}_r$,

то его можно назвать *коэффициентом вертикальности* пространства $\mathfrak{L}_r$.

(g) Пусть $e_L \notin \mathfrak{L}_r$ (другими словами, мы предполагаем, что пространство $\mathfrak{L}_r$ не является вертикальным). Тогда $\nu^2 < 1$. Как следует из теоремы 2.2, последняя координата y_L любого вектора $Y = (y_1, \dots, y_L)^T \in \mathfrak{L}_r$ является линейной комбинацией его первых компонент $y_1, \dots, y_{L-1}$:

$$y_L = a_1 y_{L-1} + a_2 y_{L-2} + \dots + a_{L-1} y_1.$$

Вектор $\mathcal{R} = (a_{L-1}, \dots, a_1)^T$ может быть представлен в виде

$$\mathcal{R} = \frac{1}{1 - \nu^2} \sum_{i=1}^r \pi_i P_i^\nabla \quad (14)$$

и не зависит от выбора базиса $P_1, \dots, P_r$ линейного пространства $\mathfrak{L}_r$.

Алгоритм рекуррентного прогноза:

Используя введенные выше обозначения, определим временной ряд $G_{N+M} = (g_0, \dots, g_{N+M-1})$ по формулам

$$g_i = \begin{cases} \tilde{f}_i & \text{для } i = 0, \dots, N-1, \\ \sum_{j=1}^{L-1} a_j g_{i-j} & \text{для } i = N, \dots, N+M-1. \end{cases} \quad (15)$$

Числа $g_N, \dots, g_{N+M-1}$ образуют M членов рекуррентного прогноза. Для краткости будем называть рекуррентный прогноз *R-прогнозом*.

Замечание 3.1 Определим линейный оператор $\mathcal{P}^{(r)} : \mathfrak{L}_r \mapsto \mathbf{R}^L$ по формуле

$$\mathcal{P}^{(r)} Y = \begin{pmatrix} Y_\Delta \\ \mathcal{R}^T Y_\Delta \end{pmatrix}, \quad Y \in \mathfrak{L}_r.$$

Если положить

$$Z_i = \begin{cases} \tilde{X}_i & \text{для } i = 1, \dots, K, \\ \mathcal{P}^{(r)} Z_{i-1} & \text{для } i = K + 1, \dots, K + M, \end{cases} \quad (16)$$

то матрица $\mathbf{Z} = [Z_1 : \dots : Z_{K+M}]$ является траекторной матрицей временного ряда G_{N+M} . Следовательно, формула (16) может рассматриваться как векторная запись рекуррентного прогноза, определенного в (15).

Если пространство $\mathfrak{L}_r$ является линейной оболочкой некоторого набора собственных векторов из сингулярного разложения траекторной матрицы временного ряда F_N , то соответствующий алгоритм R -прогнозирования мы будем называть *R –SSA-прогнозированием*.

Замечание 3.2 Обозначим $\mathfrak{L}^{(L)} = \text{span}(X_1, \dots, X_K)$ траекторное пространство ряда F_N . Предположим, что $\dim \mathfrak{L}^{(L)} = r < L$ и $e_L \notin \mathfrak{L}^{(L)}$. Если мы используем алгоритм R –SSA-прогнозирования с $\mathfrak{L}_r = \mathfrak{L}^{(L)}$, то $\mathbf{X} = \hat{\mathbf{X}} = \tilde{\mathbf{X}}$ и, следовательно, $\tilde{F}_N = F_N$. Это означает, что начальные точки $g_{N-L+1}, \dots, g_{N-1}$ для прогнозирующей рекуррентной формулы (15) совпадают с последними $L-1$ элементами ряда F_N .

3.2 Векторное SSA-прогнозирование

Алгоритм R –SSA-прогнозирования является основным прогнозирующим алгоритмом в силу его непосредственной связи с линейными рекуррентными формулами. Здесь мы приведем одну его модификацию, которая в ряде случаев может давать более точные результаты прогноза.

Сначала объясним происхождение этого алгоритма. Вернемся к базовому методу SSA как методу анализа временных рядов и предположим, что нашей целью является извлечение некоторой аддитивной компоненты $F_N^{(1)}$ из ряда F_N (в предположении, что она

приближенно разделима с остатком). Тогда, выбрав подходящую длину окна L , мы получим сингулярное разложение траекторной матрицы временного ряда F_N и можем выбрать собственные тройки $(\sqrt{\lambda_i}, U_i, V_i)$, $i \in I = (j_1, \dots, j_r)$, соответствующие ряду $F_N^{(1)}$. Результирующая матрица будет иметь вид

$$\mathbf{X}_I = \sum_{i \in I} \sqrt{\lambda_i} U_i V_i^T$$

и, после диагонального усреднения, мы придем к восстановленному ряду $\tilde{F}_N^{(1)}$, который является приближением к $F_N^{(1)}$.

Заметим, что столбцы $\hat{X}_1, \dots, \hat{X}_K$ результирующей матрицы $\mathbf{X}_I$ принадлежат линейному пространству $\mathfrak{L}_r = \text{span}(U_i, i \in I)$. Если ряд $F_N^{(1)}$ сильно отделим от ряда $F_N^{(2)} \stackrel{\text{def}}{=} F_N - F_N^{(1)}$, то $\mathfrak{L}_r$ совпадает с $\mathfrak{L}^{(L,1)}$ (траекторным пространством ряда $F_N^{(1)}$) и матрица $\mathbf{X}_I$ является ганкелевой (в этом случае $\mathbf{X}_I$ — траекторная матрица ряда $F_N^{(1)}$). Если $F_N^{(1)}$ и $F_N^{(2)}$ приближенно сильно разделимы, то пространство $\mathfrak{L}_r$ будет близко к пространству $\mathfrak{L}^{(L,1)}$ и матрица $\mathbf{X}_I$ будет приближенно ганкелевой.

Коротко идея «векторного прогноза» может быть описана следующим образом. Представим себе, что мы можем продолжить последовательность векторов $\hat{X}_1, \dots, \hat{X}_K$ на M шагов таким образом:

1. вектора продолжения Z_m ($K < m \leq K + M$) принадлежат тому же самому линейному пространству $\mathfrak{L}_r$;
2. матрица $\mathbf{X}_M = [\hat{X}_1 : \dots : \hat{X}_K : Z_{K+1} : \dots : Z_{K+M}]$ является приближенно ганкелевой.

Заметим, что, в силу предложения 2.5, при выполнении первого условия вообще говоря невозможно сделать так, чтобы матрица $\mathbf{X}_M$ была в точности ганкелевой.

Имея матрицу $\mathbf{X}_M$, мы можем получить ряд G_{N+M} с помощью диагонального усреднения. Так как первые элементы восстановленного ряда $\tilde{F}_N^{(1)}$ совпадают с элементами ряда G_{N+M} , то последний может рассматриваться как прогноз ряда $F_N^{(1)}$.

Опишем теперь формально алгоритм векторного прогноза в том же стиле, как это было сделано для рекуррентного.

Предварительные замечания:

- Алгоритм векторного прогноза имеет те же самые входные данные, как и алгоритм R -прогнозирования
- Обозначения в пунктах (а)-(g) раздела 3.1 остаются в силе. Введем еще несколько обозначений.

Рассмотрим матрицу

$$\Pi = \mathbf{V}^\nabla (\mathbf{V}^\nabla)^\mathrm{T} + (1 - \nu^2) \mathcal{R} \mathcal{R}^\mathrm{T},$$

где $\mathbf{V}^\nabla = [P_1^\nabla : \dots : P_r^\nabla]$. Матрица Π является матрицей линейного оператора, задающего ортогональную проекцию $\mathbf{R}^{L-1} \mapsto \mathfrak{L}_r^\nabla$ (см. предложение 2.10), где $\mathfrak{L}_r^\nabla = \text{span}(P_1^\nabla, \dots, P_r^\nabla)$.

Определим линейный оператор $\mathcal{P}^{(v)} : \mathfrak{L}_r \mapsto \mathbf{R}^L$ по формуле

$$\mathcal{P}^{(v)} Y = \begin{pmatrix} \Pi Y_\Delta \\ \mathcal{R}^\mathrm{T} Y_\Delta \end{pmatrix}, \quad Y \in \mathfrak{L}_r.$$

Алгоритм векторного прогноза:

1. В обозначениях, введенных выше, определим вектора Z_i следующим образом:

$$Z_i = \begin{cases} \widehat{X}_i & \text{для } i = 1, \dots, K \\ \mathcal{P}^{(v)} Z_{i-1} & \text{для } i = K + 1, \dots, K + M + L - 1. \end{cases}$$

2. Образовав матрицу $\mathbf{Z} = [Z_1 : \dots : Z_{K+M+L-1}]$ и сделав ее диагональное усреднение, мы получим ряд $g_0, \dots, g_{N+M+L-1}$.
3. Числа $g_N, \dots, g_{N+M-1}$ образуют M членов *векторного прогноза*.

Если пространство $\mathfrak{L}_r$ является линейной оболочкой некоторого набора собственных векторов, полученного с помощью алгоритма SSA, то мы будем называть соответствующий алгоритм как *векторный SSA-алгоритм* или, короче, *алгоритм $V - SSA$ -прогноза*.

Прокомментируем свойства векторного прогноза. Во-первых, если $\mathfrak{L}_r$ является траекторным пространством ряда F_N , то результаты векторного и рекуррентного прогноза совпадают. Хотя результаты и не отличаются, суть у рекуррентного и векторного прогнозов разная. Рекуррентный прогноз выполняет продолжение ряда непосредственно (с помощью ЛРФ), в то время как векторный прогноз основан на L -продолжимости в подпространстве. В случае приближенного продолжения эти два метода прогноза дают обычно разные результаты. Причем, чем хуже качество приближения, тем больше разница между результатами двух прогнозов.

В типичной ситуации (в реальных задачах) редко встречаются временные ряды такие, чтобы их траекторные пространства имели размерность $r < L - 1$ и, следовательно, мы имеем дело с пространством $\mathfrak{L}_r$, которое вообще говоря не является траекторным пространством прогнозируемого ряда (см. предложение 2.5).

Рекуррентный прогноз использует пространство $\mathfrak{L}_r$, чтобы получить прогнозирующую ЛРФ, не заботясь о том, чтобы это пространство являлось траекторным также и для ряда прогноза.

Процедура векторного прогноза пытается выполнить L -продолжение временного ряда в $\mathfrak{L}_r$; любой вектор $Z_{i+1} = \mathcal{P}^{(v)} Z_i$ принадлежит пространству $\mathfrak{L}_r$, а вектора Z_{i+1}^∇ и $(Z_i)_\Delta$ настолько близки, насколько это возможно. Последняя координата вектора Z_{i+1} получена из Z_{i+1}^∇ также с помощью ЛРФ, используемой при рекуррентном прогнозировании. Однако, так как матрица $\mathbf{Z}$ не является ганкелевой, то последующее диагональное усреднение, к сожалению, портит соотношение между точками прогноза, задаваемое этой ЛРФ.

Оба прогнозирующих метода имеют два этапа: диагональное усреднение и продолжение. При рекуррентном прогнозировании диагональное усреднение используется для получения восстанов-

ленного ряда, к последним точкам которого затем применяется ЛРФ. Для векторного метода ситуация обратная: сначала строится векторное продолжение, и только потом, для получения точек прогноза, используется диагональное усреднение.

Замечание 3.3 Чтобы получить M точек прогноза алгоритм векторного прогноза выполняет $M+L-1$ шагов. Целью этого является сохранение результатов прогноза при изменении числа шагов прогноза: прогноз на M шагов должен совпадать с первыми M значениями прогноза на $M+1$ или более шагов. Из-за свойств диагонального усреднения приходится производить $L-1$ лишних шагов.

В случае приближенной разделимости тяжело сравнивать результаты рекуррентного и векторного прогнозов теоретически. Вообще говоря, близость результатов двух методов прогноза говорят в пользу его устойчивости (достоверности).

Рекуррентный прогноз проще для интерпретации в силу удобного описания ЛРФ в терминах ее характеристического полинома. С другой стороны, результаты анализа данных показывают, что векторный прогноз обычно более «консервативен» (или менее «радикален») в тех случаях, когда рекуррентный прогноз демонстрирует резкое возрастание или убывание ряда.

4 Доверительные интервалы

Согласно основным предположениям для *SSA*-прогнозирования, аддитивная компонента $F_N^{(1)}$ ряда F_N должна управляться ЛРФ относительно небольшой размерности и при этом ряд остатков $F_N^{(2)} = F_N - F_N^{(1)}$ должен приближенно сильно разделяться с $F_N^{(1)}$ для некоторой длины окна L . В частности, предполагается, что ряд $F_N^{(1)}$ — это конечный подряд бесконечного ряда $F^{(1)}$, который является рекуррентным продолжением ряда $F_N^{(1)}$. Эти предположения не могут быть проигнорированы, но, к счастью, они справедливы для достаточно широкого класса практических задач.

Чтобы ввести доверительные границы для точек прогноза, мы должны еще более усилить наши предположения, причем не только по отношению к $F_N^{(1)}$, но также и для ряда $F_N^{(2)}$. Во-первых, будем рассматривать ряд остатков $F_N^{(2)}$ как конечный подряд некоторого бесконечного случайного (шумового) ряда $F^{(2)}$, который «портит» сигнал $F^{(1)}$. Другие предположения связаны главным образом с рядом остатков $\tilde{F}_N^{(2)} = F_N - \tilde{F}_N^{(1)}$, где $\tilde{F}_N^{(1)}$ — восстановленная компонента ряда F_N . Так как $\tilde{F}_N^{(1)} \approx F_N^{(1)}$, то свойства ряда $\tilde{F}_N^{(2)}$ сильно связаны со свойствами $F_N^{(2)}$. Более точная формулировка дополнительных предположений зависит от решаемой задачи и применяемого метода.

Здесь мы рассмотрим следующие две задачи, связанные с построением доверительных границ для прогноза. Первая задача связана с построением доверительных интервалов для прогнозирования значения всего ряда $F = F^{(1)} + F^{(2)}$ целиком в некоторый будущий момент времени $N + M$. Вторая задача формулируется как построение доверительных границ для прогноза значения сигнала $F^{(1)}$ также в некоторый будущий момент времени.

Эти две задачи решаются различными способами. Первый из них использует информацию, непосредственно полученную в процессе обработки временного ряда. Этот вариант будем называть эмпирическим. Второй требует дополнительной информации о мо-

дели ряда $\tilde{F}_N^{(2)}$ для того, чтобы можно было выполнить бутстреп-моделирование ряда F_N .

Опишем коротко обе задачи построения доверительных границ для R - SSA -прогнозирования. Для V -прогнозирования все построения полностью аналогичны.

4.1 Эмпирические доверительные интервалы для прогноза исходного временного ряда

Предположим, что мы уже получили значение прогноза $\tilde{f}_{N+M-1}^{(1)}$, т. е. мы уже выполнили M шагов процедуры R - SSA -прогнозирования. По определению, мы используем $\tilde{f}_{N+M-1}^{(1)}$ как прогноз будущего значения $f_{N+M-1}^{(1)}$ сигнала $F^{(1)}$. Как уже упоминалось, нашей задачей является построение доверительного интервала для (будущего) значения f_{N+M-1} всего ряда F .

Рассмотрим процедуру *мульти-стартового M -шагового рекуррентного продолжения*. Возьмем относительно небольшое целое M и проведем M шагов рекуррентного продолжения с помощью прогнозирующей ЛРФ, беря в качестве начальных данных скользящие отрезки восстановленного ряда длины $L-1$ от $(\tilde{f}_0^{(1)}, \dots, \tilde{f}_{L-2}^{(1)})$ до $(\tilde{f}_{K-M}^{(1)}, \dots, \tilde{f}_{N-M-1}^{(1)})$, $K = N - L + 1$.

Значения последних точек $g_{j+M+L-1}$ этих продолжений можно сравнить со значениями $f_{j+M+L-1}$ исходного ряда F_N . Таким образом, мы получаем *ряд мульти-стартовых M -шаговых остатков* H_{K-M+1} с

$$h_j^{(M)} = f_{j+M+L-2} - g_{j+M+L-2}, \quad j = 0, \dots, K - M.$$

Если мы предположим, что восстановленный ряд $\tilde{F}_N^{(1)}$ совпадает с рядом $F_N^{(1)}$, который управляется ЛРФ, то мы получим, что $g_k = f_k^{(1)}$ и ряд мульти-стартовых M -шаговых остатков совпадает с последними $K - M + 1$ значениями стационарного шумового ряда $F_N^{(2)}$.

Если эти предположения не верны, то $h_j^{(M)}$ не совпадают с $f_{j+M+L-2}^{(2)}$. Но даже если совпадения нет, предположим, что ряд мульти-стартовых M -шаговых остатков является стационарным и эргодическим в том смысле, что его эмпирическая функция распределения сходится к некоторой теоретической при $N \rightarrow \infty$. Тогда, имея в распоряжении ряд H_{K-M+1} , мы можем оценить квантили теоретического распределения (например, верхнюю и нижнюю 2.5% квантили).

Заметим, что значения $g_{j+M+L-2}$ получены за то же число шагов с помощью той же ЛРФ, что и точки прогноза $\tilde{f}_{N+M-1}^{(1)}$, при этом начальные значения взяты из одного и того же восстановленного ряда. Так как прогнозирование требует предположения о сохранении структуры ряда в будущем, то полученная эмпирическая функция распределения ряда мульти-стартовых M -шаговых остатков может быть использована для построения доверительного интервала для значения f_{N+M-1} .

Более формально, зафиксируем некоторый доверительный уровень γ ($0 < \gamma < 1$), и пусть $\alpha = 1 - \gamma$. Если обозначить $c_{\alpha/2}^-$ и $c_{\alpha/2}^+$ нижнюю и верхнюю $\alpha/2$ -квантили, вычисленные на основе эмпирической функции распределения ряда мульти-стартовых M -шаговых остатков, то мы получим *эмпирический доверительный интервал*

$$\left(\tilde{f}_{N+M-1}^{(1)} + c_{\alpha/2}^-, \tilde{f}_{N+M-1}^{(1)} + c_{\alpha/2}^+ \right),$$

который содержит f_{N+M-1} с вероятностью, близкой к γ . Очевидно, число K должно быть достаточно большим, чтобы эмпирическая функция была устойчива.

Если ряд мульти-стартовых M -шаговых остатков можно рассматривать как белый шум, то можно предложить другую модификацию эмпирического доверительного интервала. Предполагая, что справедлива гипотеза о гауссовском белом шуме, можно построить стандартный симметричный доверительный интервал для f_{N+M-1} с помощью выборочных среднего и дисперсии ряда мульти-

стартовых M -шаговых остатков. Конечно, гипотеза о белом шуме должна быть проверена с помощью стандартных статистических процедур.

4.2 Бутстреп-доверительные интервалы для прогноза сигнала

Рассмотрим метод построения доверительных границ для сигнала $F^{(1)}$ в момент времени $N + M - 1$. В нереализуемой на практике ситуации, когда известны как сам сигнал, так и истинная модель шума, с помощью моделирования и метода Монте-Карло можно определять статистические свойства значения прогноза по отношению к истинному значению.

Действительно, предположим, что правило выбора набора собственных троек фиксировано. Тогда мы можем промоделировать S независимых реализаций $F_{N,i}^{(2)}$ процесса $F_N^{(2)}$ и затем применить процедуру прогноза к S независимым временным рядам $F_{N,i} \stackrel{\text{def}}{=} F_N^{(1)} + F_{N,i}^{(2)}$. Результаты прогноза образуют выборку $\tilde{f}_{N+M-1,i}^{(1)}$ ($1 \leq i \leq S$), которая и дает информацию о $f_{N+M-1}^{(1)}$. Таким образом могут быть построены доверительные интервалы по методу Монте-Карло.

Так как на практике мы не знаем вид и значения сигнала $F_N^{(1)}$, то мы не можем применить такую процедуру. Опишем бутстреп-вариант моделирования для построения доверительных интервалов для точек прогноза.

При подходящем выборе длины окна L и набора собственных троек, соответствующих сигналу, мы получаем представление $F_N = \tilde{F}_N^{(1)} + \tilde{F}_N^{(2)}$, где $\tilde{F}_N^{(1)}$ (восстановленный ряд) аппроксимирует сигнал $F_N^{(1)}$, а $\tilde{F}_N^{(2)}$ — ряд остатков. Предположим теперь, что (стохастическая) модель остатков $\tilde{F}_N^{(2)}$ известна. Например, мы постулируем некоторую модель для $F_N^{(2)}$ и, так как $\tilde{F}_N^{(1)} \approx F_N^{(1)}$, применяем ту же самую модель для $\tilde{F}_N^{(2)}$ с оцениваемыми параметрами.

Тогда, моделируя S независимых реализаций $\tilde{F}_{N,i}^{(2)}$ ряда $F_N^{(2)}$, мы

получаем S временных рядов $F_{N,i} \stackrel{\text{def}}{=} \tilde{F}_N^{(1)} + \tilde{F}_{N,i}^{(2)}$ и S раз проводим прогнозирование, получая $\tilde{f}_{N+M-1,i}^{(1)}$ также, как в варианте метода Монте-Карло.

Более точно, каждый временной ряд $F_{N,i}$ порождает свой восстановленный ряд $\tilde{F}_{N,i}^{(1)}$ и свою линейную рекуррентную формулу при выборе одной и той же длины окна L и одного и того же набора собственных троек. Поэтому для каждого ряда на основе последних $L - 1$ точек ряда $\tilde{F}_{N,i}^{(1)}$ в качестве начальных значений для рекуррентного прогноза по соответствующей ЛРФ, выполняется M шагов прогноза, результатом чего являются значения $\tilde{f}_{N+M-1,i}^{(1)}$.

Как только выборка $\tilde{f}_{N+M-1,i}^{(1)}$ ($1 \leq i \leq S$) из спрогнозированных значений получена, мы можем вычислить (эмпирические) нижнюю и верхнюю квантили по заданному доверительному уровню γ и получить соответствующий доверительный интервал для прогноза сигнала. Этот интервал называется *бутстреп-доверительным интервалом*.

Простейшей моделью для $\tilde{F}_N^{(2)}$ является модель гауссовского белого шума. Соответствующая гипотеза может быть проверена с помощью стандартных тестов на случайность и нормальность.

5 Выводы и рекомендации

Суммируем материал предыдущих разделов, взяв за основу R – SSA-прогнозирование.

1. Постановка задачи

Есть ряд $F_N = F_N^{(1)} + F_N^{(2)}$ и стоит задача прогноза компоненты ряда $F_N^{(1)}$. Если $F_N^{(2)}$ может рассматриваться как шумовая компонента, то задача состоит в прогнозировании сигнала $F_N^{(1)}$ в присутствии шума $F_N^{(2)}$.

2. Основные предположения

- Ряд $F_N^{(1)}$ допускает рекуррентное продолжение с помощью ЛРФ относительно небольшой размерности d .
- Существует такое число L , что ряды $F_N^{(1)}$ и $F_N^{(2)}$ приближенно разделимы при выборе длины окна, равной L . Это важное предположение, так как любой временной ряд $F_N^{(1)}$ можно рассматривать как аддитивную компоненту ряда F_N в том смысле, что $F_N = F_N^{(1)} + F_N^{(2)}$ с $F_N^{(2)} = F_N - F_N^{(1)}$. Условие о (приближенной) разделимости означает, что $F_N^{(1)}$ является естественной аддитивной компонентой ряда F_N с точки зрения метода SSA.

3. Правильный выбор параметров

Так как мы должны выбрать длину окна L , обеспечивающую достаточно хорошее качество разделимости, и выбрать набор собственных троек, соответствующих ряду $F_N^{(1)}$, то все основные правила для SSA как метода анализа временного ряда остаются в силе. Заметим, что в данном случае нам необходимо отделить ряд $F_N^{(1)}$ от $F_N^{(2)}$, но мы не нуждаемся в полном разложении всего ряда $F_N = F_N^{(1)} + F_N^{(2)}$.

4. Особенности и опасности

Задача прогнозирования с помощью SSA имеет свои особенности по сравнению с задачей анализа.

- Так как выбранная длина окна L (вместе с набором собственных троек) порождает ЛРФ размерности $L - 1$, которая используется в качестве рекуррентной прогнозирующей формулы, то проблема побочных корней характеристического полинома становится довольно важной. Выбор $L = d + 1$, где под d подразумевается размерность минимальной ЛРФ, является оптимальным. К сожалению, на практике маленькие значения L не обеспечивают достаточного качества разделимости. Поэтому необходимо выбрать наименьшую длину окна, большую чем d и приводящую к достаточно хорошей разделимости.
- Линейное пространство $\mathcal{L}_r$ размерности r , определяющее прогнозирующую ЛРФ, является линейной оболочкой собственных векторов из выбранных собственных троек. Так как условие $r \geq d$ должно быть выполнено, число собственных выбираемых троек, соответствующих ряду $F_N^{(1)}$, должно быть не меньше, чем d .
- При анализе ряда с помощью SSA, если мы дополним множество правильно выбранных собственных троек некоторыми лишними собственными тройками с маленькими сингулярными значениями, то результат восстановления существенно не изменится. Если же мы имеем дело с прогнозом, то такое действие может вызвать сильное изменение пространства $\mathcal{L}_r$; его размерность увеличится и, как следствие, ЛРФ прогноза также изменится. Все это может привести к кардинальному изменению прогноза. Заметим, что величины дополнительных сингулярных значений в этом случае не важны. Таким образом, к выбору собственных троек, описывающих ряд $F_N^{(1)}$, необходимо относиться очень внимательно.

5. Характеристики прогноза

Перечислим несколько характеристик, которые могут быть полезны при оценке качества прогноза.

- *Характеристики делимости.* Все характеристики делимости, рассмотренные подробно в работе [3], посвященной анализу временных рядов, важны также и для прогнозирования.
- *Корни полинома.* Корни характеристического полинома прогнозирующей ЛРФ могут помочь в описании поведения прогноза. Эти корни полинома могут быть полезны при ответе на следующие вопросы:
 - (а) Пусть мы ожидаем (предполагаем), что прогноз должен иметь некоторую специальную форму (например, ожидается, что значения будут возрастать). Описывают ли корни полинома такую возможность? Например, для экспоненциального роста необходимо, чтобы был вещественный корень единичной кратности, немного больший по значению, чем 1; если мы пытаемся спрогнозировать годовую сезонность, то должна присутствовать пара комплексных корней с частотами $\approx k/12$, и т.д.
 - (б) Возможно ли получить опасно неустойчивый, противоречивый прогноз? В терминах корней полинома, каждый побочный корень увеличивает такой шанс. Если модуль побочного корня существенно меньше единицы, то небольшое изменение начальных данных для применяемой ЛРФ не должно привести к существенным ошибкам долгосрочного прогноза. Однако побочные корни с большими, чем 1, модулями могут привести к проблемам даже для краткосрочного прогноза.
- *Коэффициент вертикальности.* Значение коэффициента вертикальности ν^2 равно квадрату косинуса угла меж-

ду пространством $\mathfrak{L}_r$ и вектором-ортом e_L . Условие $\nu^2 < 1$ является необходимым для возможности прогнозирования. Если ν^2 близко к 1, то, в силу (14), коэффициенты прогнозирующей ЛРФ будут большими и, следовательно, некоторые корни характеристического полинома также будут большими по модулю. Если ожидаемое поведение значений прогноза не предполагает резкого возрастания или убывания, то большие значения коэффициента вертикальности говорят о возможных проблемах, которые могут возникнуть, в частности, о возможной неустойчивости прогноза. Такая ситуация скорее всего означает, что были выбраны лишние собственные тройки (или вообще, предположения о ряде не выполнены).

6. Роль начальных данных

Не считая числа M шагов прогноза, к формальным параметрам алгоритма $R-SSA$ -прогнозирования относятся длина окна L и набор I собственных троек, описывающих ряд $F_N^{(1)}$. Эти параметры определяют как прогнозирующую ЛРФ (14), так и начальные данные для применения этой ЛРФ. Очевидно, что результаты прогноза существенно зависят от начальных данных, особенно если прогнозирующая ЛРФ имеет лишние корни.

Метод $R-SSA$ -прогнозирования в качестве начальных данных использует последние значения $\tilde{f}_{N-L+1}^{(1)}, \dots, \tilde{f}_{N-1}^{(1)}$ восстановленного ряда $\tilde{F}_N^{(1)}$. В силу свойств диагонального усреднения последние (и первые) члены ряда $F_N^{(1)}$ обычно восстанавливаются с меньшей точностью по сравнению с серединой ряда. Этот эффект может вызвать увеличение ошибок прогноза.

Например, любой линейный (и не являющийся константой) ряд с общим членом $f_n = an + b$ управляется минимальной ЛРФ $f_n = 2f_{n-1} - f_{n-2}$, которая не зависит от a и b . Па-

параметры a и b , используемые при прогнозе (в данном случае — продолжении ряда), полностью определяются начальными значениями f_{N-2} и f_{N-1} . Очевидно, что ошибки в начальных данных могут существенно изменить поведение прогноза, даже сменить тенденцию к возрастанию на тенденцию к убыванию.

Таким образом, важно проверять последние точки восстановленного ряда, в частности, сравнивать их с ожидаемым поведением ряда $F_N^{(1)}$.

7. Восстановленный ряд и ЛРФ

В ситуации сильной разделимости рядов $F_N^{(1)}$ и $F_N^{(2)}$, а также правильного выбора собственных троек, восстановленный ряд будет управляться ЛРФ и совпадать с рядом $F_N^{(1)}$. Несогласованность в соответствии между восстановленным рядом и рядом, управляемым найденной ЛРФ, является индикатором возможных ошибок: недостаточного качества разделимости (возможно вызванного неправильным выбором параметров) или общего несоответствия модели. Две характеристики данного соответствия могут оказаться полезными.

- *Глобальная аппроксимация.* Кроме использования ЛРФ для прогнозирования мы можем использовать ее для аппроксимации или всего ряда или его частей. Например, если мы вместо последних возьмем первые значения восстановленного ряда и сделаем $N - L + 1$ шагов прогнозирующей процедуры, то мы можем проверить, насколько хорошо восстановленный ряд глобально описывается (аппроксимируется) данной ЛРФ.
- *Локальная аппроксимация.* Описанная выше процедура соответствует долгосрочному прогнозированию. Чтобы проверить качество краткосрочного прогноза, мы должны проверить локальное соответствие между восстановленным рядом и прогнозирующей ЛРФ. Для этого рас-

смотрим несколько другую процедуру.

Этот метод используется в разделе 4.1 для построения эмпирических доверительных интервалов и называется мульти-стартовым рекуррентным продолжением. Согласно ему, для относительно небольшого Q выполняется Q шагов процедуры мульти-стартового рекуррентного продолжения, т. е. происходит Q -кратное применение прогнозирующей ЛРФ к скользящим начальным данным, меняющимся от $(\tilde{f}_0^{(1)}, \dots, \tilde{f}_{L-2}^{(1)})$ до $(\tilde{f}_{K-Q}^{(1)}, \dots, \tilde{f}_{N-Q-1}^{(1)})$, $K = N - L + 1$. Полученные результаты сравниваются со значениями $f_{L-1}^{(1)}, \dots, f_{N-1}^{(1)}$.

Поскольку как ЛРФ, так и начальные данные содержат некоторые ошибки, то локальное несоответствие для небольших Q обычно более информативно, чем глобальное. Более того, проверяя результаты локальной аппроксимации для различных Q , можно оценить максимальное число шагов M , при прогнозировании на которое можно получить разумные результаты.

8. Устойчивость и надежность прогноза

В то время как правомерность прогнозирования, основанная на требовании сохранения структуры ряда в будущем, не может быть проверена на основе внутренних свойств временного ряда, устойчивость прогноза поддается некоторой проверке. Упомянем несколько способов такой проверки.

- *Различные алгоритмы.* Близость результатов различных вариантов прогноза (например, рекуррентного и векторного прогнозов) говорит в пользу его достоверности.
- *Различные начальные данные.* Так как последние элементы восстановленного ряда могут содержать значительные ошибки, то можно начать прогноз с одной из предыдущих точек. Результаты сравнения прогнозов, начатых с разных точек, также говорят об устойчивости прогноза.

- *Различные длины окна.* Если характеристики делимости устойчивы при небольших изменениях длины окна, то можно сравнивать прогнозы при измененных L .
- *Прогнозирование укороченного временного ряда.* Укоротим исходный ряд F_N удалением нескольких его последних элементов. Если условия делимости устойчивы относительно такой операции, то мы можем сравнивать прогноз укороченного ряда с последними, удаленными, точками ряда. Этот способ более объективен, так как в построении прогноза не участвуют те точки ряда, на которых прогноз тестируется. Однако его недостатком является то, что значения в последних (наиболее «недавних») точках ряда не учитываются при построении прогноза.

9. Доверительные интервалы

Хотя как эмпирические, так и бутстреп-доверительные интервалы не являются теоретическими и носят несколько условный характер, тем не менее, они дают важную дополнительную информацию о точности прогноза и его устойчивости.

- *Общие предположения.* Моделью исходного ряда служит представление $F_N = F_N^{(1)} + F_N^{(2)}$, где $F_N^{(1)}$ — сигнал, а $F_N^{(2)}$ — «шум». Предполагается, что ряд $F_N^{(1)}$ допускает рекуррентное продолжение и поставлена задача его прогноза.
- *Цели.* Эмпирические доверительные интервалы строятся для прогноза всего ряда F_N , про который предполагается, что в будущем его структура сохранится. Бутстреп доверительные интервалы строятся для продолжения сигнала $F_N^{(1)}$.
- *Источник данных и дополнительные предположения.* Источником данных для построения бутстреп-доверительных интервалов служит ряд остатков $\tilde{F}_N^{(2)} = F_N - \tilde{F}_N^{(1)}$, где

$\tilde{F}_N^{(1)}$ — восстановленный ряд. Предполагается, что статистическая модель для ряда остатков построена и ее параметры оценены на основе значений ряда $\tilde{F}_N^{(2)}$.

Данные, используемые для построения эмпирических доверительных интервалов, также имеют форму некоторого ряда остатков, но в данном случае построение ряда остатков основано на мульти-стартовом ряде остатков, описанном в разделе 4.1. Для построения доверительных границ прогноза в точке $N + Q - 1$ длина используемого ряда остатков равняется $N - L - Q + 2$. Поэтому для устойчивости доверительных границ необходимо, чтобы $N - L - Q + 2$ было достаточно большим (превышало несколько десятков).

- *Проверка предположений.* В случае, когда соответствующие ряды остатков являются реализациями белого шума, предположения для обоих вариантов доверительных интервалов выполняются. Соответствующая гипотеза проверяется стандартными статистическими методами. Но даже если эта гипотеза отвергается, доверительные границы все равно имеют смысл, так как создают масштабную шкалу, в которой, в частности, можно сравнивать различные методы и измерять устойчивость.
- *Бутстреп-прогноз.* Результаты бутстреп-моделирования, используемого для получения бутстреп-доверительных интервалов, могут дать дополнительную информацию о значениях прогноза. Например, получив прогноз промоделированных рядов $F_{N,i}$ ($1 \leq i \leq S$), можно рассматривать среднее соответствующих (случайных) прогнозов в качестве еще одного варианта прогноза.

6 Приложение: пример прогнозирования временного ряда

Продemonстрируем методику применения метода «Гусеница»-SSA на примере прогноза временного ряда FORT (объемы месячных продаж крепленых вин в Австралии с января 1984 года по июнь 1994 года, рис. 2). Естественно, любой прогноз временного ряда должен предваряться анализом ряда. В данном случае мы будем опираться на анализ рассматриваемого ряда, проведенный в [3, Приложение В]. Так как мы будем рассматривать как сам ряд длины $N = 174$, так и его подряд, состоящий из первых 120 точек, то там, где это необходимо, будем использовать обозначения FORT174 для всего ряда и FORT120 для укороченного ряда. Приведенные графики будут одновременно являться иллюстрацией понятий, возникающих при изложении теоретических аспектов метода.

Так как прогноз методом «Гусеница»-SSA применяется к рядам, управляемым (может быть, приближенно) линейными рекуррентными формулами, то начнем с исследования ряда FORT с этой точки зрения.

6.1 Описание временного ряда линейной рекуррентной формулой

В [3] мы определили, что ряд FORT может быть представлен в виде «сигнал+шум» и смогли приближенно отделить сигнал от шумовой компоненты, причем качество разделения при длине окна $L = 84$ было достигнуто довольно хорошее (w -корреляция между этими компонентами равна 0.004). При этом сигналу соответствовало 11 собственных троек, что означает, что его траекторное подпространства должно хорошо аппроксимироваться подпространством размерности 11. В свою очередь, подпространство порождает линейную рекуррентную формулу размерности $83 = L - 1$ такую, что восстановленный ряд (в данном случае — сигнал) будет аппроксимироваться рядом, управляемым этой ЛРФ.

В таблице 1 представлена информация о первых 19 из 83 корней характеристического полинома данной ЛРФ. Корни упорядочены по убыванию их модуля. Если в столбце «Тип» таблицы написано «компл.», то это означает, что строка содержит информацию о двух комплексно-сопряженных корнях и указана величина абсолютного значения мнимой части корня. Первые шесть строк таблицы можно легко проинтерпретировать: 1–3 и 5–6 строки соответствуют комплексно-сопряженным корням, порождающим гармоники с периодами, примерно равными 6, 4, 2.4, 12 и 3. Модуль, больший единицы, указывает на возрастание амплитуды, меньший единицы — на убывание. Результаты согласуются с [3, рис. 11]. Четвертая строка таблицы соответствует вещественному корню с модулем 0.997. Так как число главных корней должно быть равно как раз одиннадцати, то больше главных корней нет. Остальные корни являются побочными, по модулю меньшими единицы. Все корни изображены на рис. 3 на комплексной плоскости в координатах (Re, Im).

Таблица 1: Ряд FORT: характеристики корней характеристического полинома полной ЛРФ

N	Re	Im	Модуль	Частота	Период	Тип
1	0.497	0.871	1.003	1.053	5.969	компл.
2	-0.002	1.000	1.000	1.573	3.994	компл.
3	-0.870	0.489	0.998	2.630	2.389	компл.
4	0.997	0.000	0.997	0.000	Нет	вещ.
5	0.861	0.497	0.994	0.524	12.002	компл.
6	-0.478	0.866	0.989	2.075	3.028	компл.
7	-0.094	0.972	0.976	1.667	3.768	компл.
8	-0.391	0.894	0.975	1.983	3.168	компл.
9	0.796	0.563	0.975	0.615	10.212	компл.
10	0.401	0.888	0.975	1.147	5.480	компл.

Проверим, насколько хорошо ряд описывается порожденной вос-

становленным подпространством ЛРФ (для дальнейших ссылок будем называть таким формулы *полными ЛРФ*). Максимальная ошибка глобальной аппроксимации равняется 132, что составляет меньше 10% от значений ряда. Напомним, что так как в качестве начальных значений ЛРФ были взяты первые 83 значений ряда, то ошибка аппроксимации считается, начиная с 84-ой точки.

Удалим теперь все корни полинома кроме одиннадцати главных. Главные корни порождают новую (минимальную) ЛРФ, которая, если взять в качестве начальных точки с 73-ей по 83-ю, то на оставшейся части ряда она будет аппроксимировать ряд даже чуть лучше (максимальная ошибка равна 94). Если взять в качестве начальных первые 11 точек, то максимальная ошибка аппроксимации, сосчитанная на 12–174 точках, увеличится (что не удивительно) до 495. Таким образом, ряд достаточно хорошо аппроксимируется рядом, управляемым минимальной ЛРФ размерности 11 с корнями, изображенными на рис. 4. Поэтому, с точностью до констант, можно выписать аналитический вид этой аппроксимации. Для локальной аппроксимации соотношение результатов примерно такое же (абсолютные значения ошибок меньше).

Так как мы знаем точное значение периода, то можно еще подправить ЛРФ, изменив корни так, чтобы они в точности соответствовали периодам 6, 4, 2.4, 12 и 3 (будем называть эту формулу размерности 11 *исправленной минимальной ЛРФ*). Ошибки локальной аппроксимации при этом даже уменьшатся.

Выпишем вид ряда, управляемого исправленной минимальной ЛРФ:

$$\begin{aligned} f_n = & C_1 0.997^n + C_2 0.994^n \sin(2\pi n/12 + \phi_2) + \\ & + C_3 \sin(2\pi n/4 + \phi_3) + C_4 1.003^n \sin(2\pi n/6 + \phi_4) + \\ & + C_5 0.998^n \sin(2\pi n/2.4 + \phi_5) + C_6 0.989^n \sin(2\pi n/3 + \phi_6). \end{aligned} \quad (17)$$

Здесь коэффициенты C_i и ϕ_i определяются начальными данными. Мы выписали слагаемые в порядке убывания их вкладов в сумму на интервале от 1 до 174, а не в порядке убывания модулей порождающих их корней, который руководит упорядоченностью вкладов

в будущем, построенном по соответствующей ЛРФ.

Таким образом, предварительное исследование показало, что ряд достаточно хорошо соответствует требуемой модели и можно приступать к прогнозированию.

6.2 Прогноз и доверительные интервалы

Устойчивый долгосрочный прогноз. Как было показано выше, ряд FORT оказался почти идеальным для прогноза: он состоит из сигнала, хорошо аппроксимируемого рядом, управляемым ЛРФ, и белого шума. На рис. 5 приведены первые 24 точки рекуррентного прогноза на 60 точек вперед по полной ЛРФ, построенной по подпространству, порожденному первыми одиннадцатью собственными тройками. Значения векторного варианта прогноза, а также рекуррентного прогноза вида (17) по исправленной минимальной ЛРФ почти совпадают с изображенным на рис. 5, поэтому мы их не приводим. Такое совпадение подтверждает устойчивость построенного прогноза.

На рис. 6 изображены бутстреп-доверительные интервалы, соответствующие уровню доверия 0.95. При построении доверительных интервалов мы вначале убедились, что ряд остатков восстановления удовлетворяет необходимым условиям. А именно, можно считать, что ряд $\tilde{F}^{(2)} = F - \tilde{F}^{(1)}$, где $\tilde{F}^{(1)}$ — восстановленный по первым 11 собственным тройкам ряд, состоит из независимых реализацией нормально-распределенных случайных величин. Исследование остатков уже проводилось при анализе ряда. Гипотеза о независимости подтверждается, в то время как критерий Аббе для проверки гипотезы о нормальности дает пороговый уровень значимости 0.01. Некоторые отклонение от нормальности могут немного исказить величину доверительных интервалов, однако качественно им все равно можно доверять. Заметим, что размер доверительных интервалов лишь незначительно растет в течении 5 лет.

Таким образом, для ряда FORT построен прогноз на 60 точек вперед, которому вполне можно доверять при условии, что струк-

тура ряда в дальнейшем не изменится.

Краткосрочный прогноз. Не всегда данные таковы, что по ним возможно построить прогноз на большое число точек вперед. Причинами этого могут быть следующие: 1) недостаточная длина ряда по сравнению с величиной шума и/или сложностью сигнала; 2) нелинейная структура сигнала, в том смысле, что он плохо аппроксимируется рядом, управляемым ЛРФ относительно небольшой размерности. Мы рассмотрим только первый случай.

Укоротим временной ряд FORT до первых 120 точек так же, как это делалось при анализе ряда в [3]. Для того, чтобы было удобно ссылаться, назовем его FORT120, а исходный ряд — FORT174. Проведенный в [3, Приложение В] анализ ряда показал, что при длине окна $L = 60$ к сигналу можно отнести первые 11 собственных троек сингулярного разложения траекторной матрицы ряда FORT120. Возникающее при разложении перемешивание двух гармоник существенно для прогноза только в том смысле, что включать/не включать собственные тройки с номерами 8–11 необходимо одновременно.

Также как и для ряда FORT174, построим прогноз на 60 точек вперед на основе первых 11 собственных троек. Структура порожденного соответствующими собственными векторами пространства для полного и укороченного рядов полностью аналогична. Разница в том, что ряд FORT120 короче, поэтому качество приближенной отделимости сигнала от шума должно быть хуже. Это подтверждает ошибка глобальной аппроксимации, максимальное значение которой равно 312 (вместо 132 для FORT174).

Для укороченного ряда мы имеем большое преимущество при оценке качества прогноза, состоящее в том, что мы можем сравнивать значения прогноза с реальными значениями ряда. На рис. 7 изображены значения рекуррентного прогноза (жирная сплошная линия) по сравнению с точками ряда FORT174. Видно, что на первых 24 точках прогноз качественно описывает будущее поведение, а затем начинаются довольно сильные расхождения. Приведем информацию о первых 19 из 59 корней характеристического полино-

ма полной прогнозирующей ЛРФ для ряда FORT120 (таблица 2). Сравнение с таблицей 1 показывает, что поведение прогноза ряда FORT120 обусловлено большим модулем комплексно-сопряженных корней, соответствующих модулированной гармонике с периодом 6 (первая строка таблицы). На рис. 7 видно, что именно растущая по амплитуде 6-месячная гармоника в основном и отвечает за расхождение прогноза с реальными данными.

Таблица 2: Ряд FORT120: характеристики корней характеристического полинома полной ЛРФ

N	Re	Im	Модуль	Частота	Период	Тип
1	0.505	0.878	1.013	1.049	5.990	компл.
2	-0.885	0.480	1.007	2.644	2.376	компл.
3	0.000	1.000	1.000	1.571	4.001	компл.
4	0.997	0.000	0.997	0.000	нет	вещ.
5	0.862	0.496	0.994	0.522	12.033	компл.
6	-0.490	0.851	0.982	2.093	3.002	компл.
7	0.366	0.896	0.968	1.183	5.311	компл.
8	0.768	0.587	0.966	0.652	9.635	компл.
9	-0.128	0.957	0.966	1.704	3.688	компл.
10	-0.899	0.350	0.965	2.771	2.268	компл.

Для того, чтобы подтвердить меньшую устойчивость прогноза ряда FORT120 по сравнению с прогнозом ряда FORT174, построим доверительные интервалы. На рис. 8 видно, что размер бутстреп-доверительных интервалов сильно увеличивается, даже захватывая область отрицательных значений.

Чтобы окончательно убедиться в том, что плохое качество долгосрочного прогноза является следствием недостаточной разделимости, а не того, что структура первых 120 точек ряда отличается от структуры последующих точек, применим исправленную минимальную ЛРФ, полученную при анализе ряда FORT174 и соответ-

ствующую формуле (17). Результат изображен на рис. 9 и показывает довольно хорошее качество прогноза (жирная сплошная линия соединяет точки прогноза, а реальные значения помечены квадратиками).

Тестируя качество прогноза на укороченном ряде, мы имеем возможность количественно выразить качество прогноза, например, с помощью среднеквадратичного отклонения значений прогноза от реальных значений. Рекуррентный прогноза по полной прогнозирующей ЛРФ дает значение среднеквадратичной ошибки, равное 549, векторный вариант прогноза приводит к чуть меньшей ошибке 535, в то время как прогноз по исправленной минимальной ЛРФ уменьшает ошибку до 314.

Таким образом, исследованный пример подтверждает теоретические результаты о зависимости качества прогноза от точности разделимости.

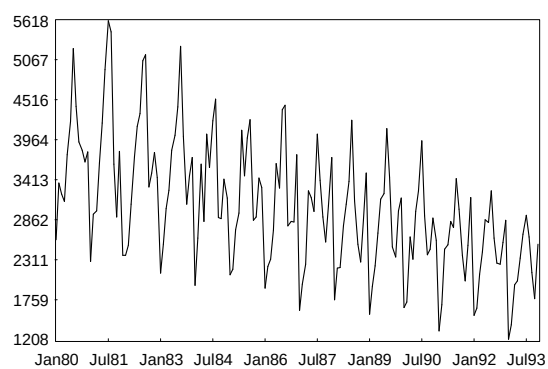


Рис. 2: FORT174: график исходного ряда месячных объемов продаж крепленых вин в Австралии

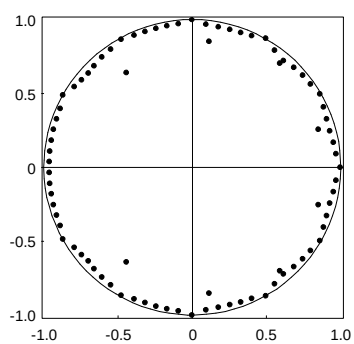


Рис. 3: FORT174: главные и побочные корни характеристического полинома полной ЛРФ

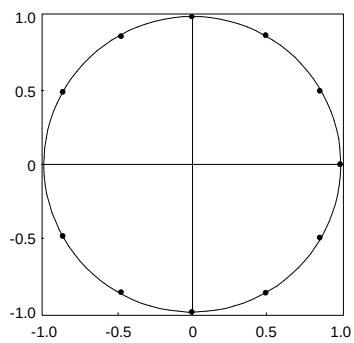


Рис. 4: FORT174: корни характеристического полинома минимальной ЛРФ

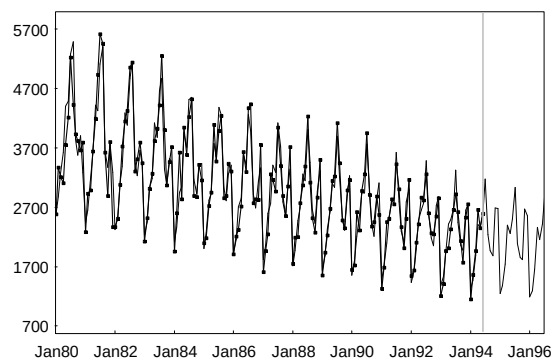


Рис. 5: FORT174: Рекуррентный прогноз на 24 точки вперед

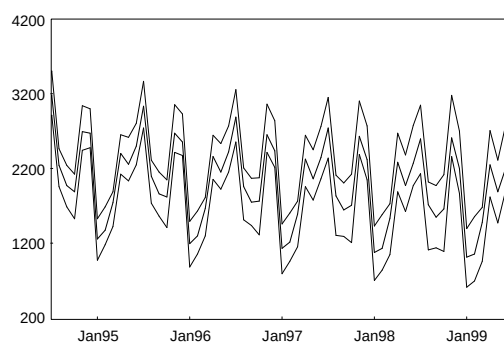


Рис. 6: FORT174: Рекуррентный прогноз на 60 точек и бутстреп-доверительные интервалы

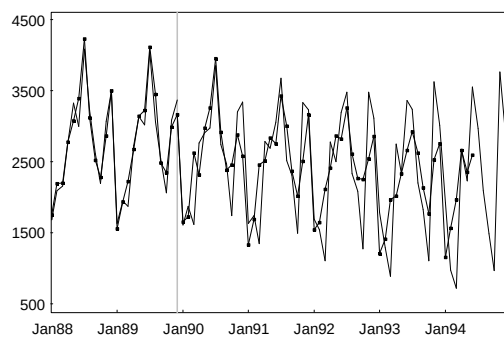


Рис. 7: FORT120: Рекуррентный прогноз по полной ЛРФ на 60 точек на фоне реальных значений

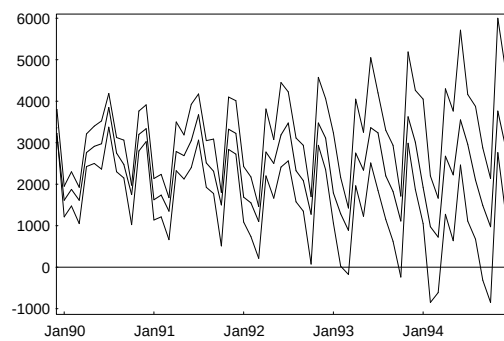


Рис. 8: FORT120: Рекуррентный прогноз на 60 точек и бутстреп-доверительные интервалы

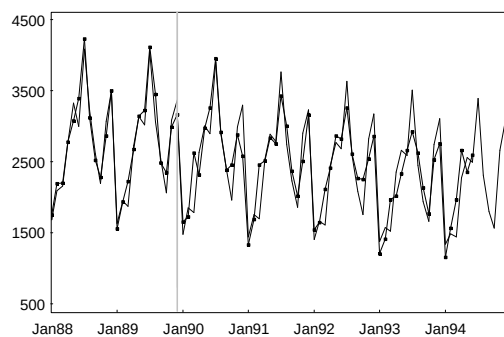


Рис. 9: FORT120: Рекуррентный прогноз на 60 точек по исправленной минимальной ЛРФ (FORT174), на фоне реальных значений

Список литературы

- [1] Данилов Д.Л., Жиглявский А.А. (ред.) *Главные компоненты временных рядов: метод «Гусеница»*, С.-Петербург: Пресском, 1997, 308 с.
- [2] Golyandina N., Nekrutkin V., and Zhigljavsky A. *Analysis of Time Series Structure: SSA and Related Techniques*, London: Chapman & Hall/CRC, 2001, 305 P.
- [3] Голяндина Н.Э. *Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие*, Санкт-Петербург: Изд-во СПбГУ, 2003, 87 с.
- [4] Buchstaber, V.M. *Time series analysis and grassmannians*. — In: Applied Problems of Radon Transform./ Ed. S.Gindikin. Providence, RI: AMS, 1994, p. 1–17 (Ser. AMS Transactions — Series 2, V. 162.)
- [5] Бухштабер В.М. Многомерные развертки временных рядов. Теоретические основы и алгоритмы. *Обзор прикл. промышл. матем., сер. Вероятн. и статист.*, 1997, т. 4, вып. 4, с. 629–645.
- [6] Данилов Д.Л. Метод «Гусеница» для прогнозирования временных рядов, в сб. *Главные компоненты временных рядов: метод «Гусеница»*. Под ред. Д.Л.Данилова и А.А.Жиглявского. С.-Петербург: Пресском, 1997, с.73–104.
- [7] Некруткин В.В. Аппроксимирующие пространства и продолжения временных рядов, в сб.: *Статистические модели с приложениями в эконометрике и смежных областях*. Под редакцией С.М.Ермакова и Ю.Н.Каштанова. Санкт-Петербург: НИИХ СПбГУ, 1999, с.3–32.